

I JORNADAS DE DOCTORADO EN INFORMÁTICA

26 DE FEBRERO DE 2018,
UNIVERSIDAD DE ALMERÍA

Luis Iribarne,
Ester Martín
y Manuel Berenguel
(eds.)

I Jornadas de Doctorado en Informática

© del texto:
los autores

Libros Electrónicos n.º 117

© de la edición:

Editorial Universidad de Almería, 2021

editorial@ual.es

www.ual.es/editorial

Telf/Fax: 950 015459

α

ISBN: 978-84-1351-082-8

Colabora:

Comisión académica de doctorado en informática

<https://sites.google.com/ual.es/jdi2018/>

@doctoradoINFUAL

doctorado.informatica@Ual

Financiado por:

Escuela Internacional de Doctorado
y Departamento de Informática



En este libro puede volver al índice
pulsando el pie de la página

I Jornadas de Doctorado en Informática

15 de febrero de 2018, Universidad de Almería

Las I Jornadas de Doctorado en Informática de la Universidad de Almería (JDI2018@UAL) es un evento que se realiza en el marco del Programa de Doctorado en Informática de la UAL cuyo objetivo principal es la realización del seguimiento de la investigación anual de los estudiantes de doctorado, y que sirve, a la vez, como punto de encuentro anual para estudiantes, tutores, directores y miembros de la comunidad universitaria con intereses en la investigación científica en el ámbito de la Informática. Así mismo, estas jornadas pretenden servir como foco para la difusión y divulgación de los resultados de investigación de la Informática que se están desarrollando en las tesis doctorales soportadas por los actuales proyectos de I+D de los Grupos de Investigación del Departamento de Informática.

© 2018

Editores: Luis Iribarne, Ester Martín y Manuel Berenguel
Programa de Doctorado en Informática
Departamento de Informática
Escuela Internacional de Doctorado de la Universidad de Almería
Universidad de Almería

Contenido

SECCIÓN I: Trabajos presentados

1. Calvo Cruz Nicolás: "Computación de altas prestaciones en el diseño óptimo y control de plantas solares de torre"	2
2. Carballo López José Antonio: "Modelado y optimización para una gestión eficiente de recursos en desalación solar"	12
3. Fernández García Antonio Jesús: "Técnicas híbridas de machine learning y transformación de modelos para la evolución de interfaces mashup"	22
4. García García Francisco José: "Procesamiento de consultas espaciales en entornos escalables y distribuidos"	30
5. García Salmerón José Manuel: "Computación de altas prestaciones para la resolución de problemas de optimización global"	40
6. Gil Vergel Juan Diego: "Estrategias de control jerárquico y optimización aplicadas a plantas de destilación por membranas alimentadas con energía solar"	50
7. Medina López Cristóbal: "Eficiencia y seguridad en protocolos de streaming peer-to-peer"	59
8. Ojeda Castelo Juan Jesús: "Un modelo inteligente de interacción natural autoadaptativo basado en visión artificial"	69
9. Orts Gómez Francisco José: "Acelerando la computación científica basada en estructuras de datos irregulares"	79
10. Puertas Martín Savins: "Computación de altas prestaciones para la resolución de problemas de optimización en bioinformática"	89
11. Ramos Teodoro Jerónimo: "Estrategias de gestión energética en entornos productivos con apoyo de energía solar"	99
12. Wang Hui: "Contribution of models in the disease and energy management in intensive growing system"	107

SECCIÓN II: Otros trabajos

13. Alulema Flores Darwin Omar: "Una metodología basada en modelos y servicios para la interoperabilidad de aplicaciones heterogéneas en televisión digital"	114
14. Barrionuevo Rosales Gabriel: "Computación de altas prestaciones para la modelización del sistema eco-evolutivos basados en agentes"	123
15. Gómez Navarro Francisco José: "Modelado y simulación de un nuevo concepto de vehículo urbano electrónico ligero basado en la optimización del uso de energías renovables y la reducción de CO2"	129
16. González Revuelta María Esther: "El acceso seguro del ciudadano a su historia digital permite avanzar en la gestión de procesos asistenciales"	135
17. Loaiza Loayza Mónica Cecibel: "Resolución de problemas de localización competitiva mediante procedimientos basados en GIS"	145
18. Molero Pérez José Manuel: "Computación de altas prestaciones aplicadas a la detección de anomalías en imágenes hiperespectrales"	151
19. Montoya Ríos Ana Paola: "Aportaciones al modelado y control climático de invernaderos"	161
20. Ortega López Luis: "Procesado eficiente de imágenes en el campo de la agricultura de precisión"	171
21. Rodríguez Gracia Diego: "Arquitecturas adaptativas en tiempo de ejecución mediante transformaciones inteligentes basadas en técnicas de soft computing y desarrollo de software dirigido por modelos"	177
22. Ruiz Ferrández Miriam: "Modelado y optimización de problemas en sanidad vía computación de altas prestaciones"	189
23. Santamaría López Teresa Miriam: "Distribución de carga en sistemas globales de streaming de video en directo"	199

SECCIÓN I: Trabajos presentados

Computación de Altas Prestaciones en el Diseño Óptimo y Control de Plantas Solares de Torre

Optimización Sectorial Continua

N.C. Cruz¹

Universidad de Almería, ncalvocruz@ual.es

Resumen El campo de helióstatos de las centrales solares de torre tiene un gran impacto tanto en su coste como en su rendimiento. Concretamente, puede suponer hasta un 50% de la inversión inicial y ocasionar hasta un 40% de pérdida energética. Por este motivo es muy importante optimizar su diseño. Sin embargo, se trata de un problema de optimización extremadamente complejo y aún no resuelto. En general, se suele confiar en distribuir los helióstatos siguiendo patrones conocidos. Desgraciadamente, esta técnica es muy rígida y se sabe que sus resultados son mejorables. En este trabajo se formula el problema como un caso de optimización continua con restricciones. Se diseña entonces un algoritmo genético e intrínsecamente paralelo para su resolución. Finalmente se plantea una meta-heurística para descomponer el problema de optimización en subproblemas de más fácil resolución. La asociabilidad del optimizador genético con la meta-heurística es directa. Este documento informativo incluye además valoraciones sobre los resultados obtenidos. Sin embargo, se obvian ciertos detalles que actualmente se encuentran bajo gestión de varias revistas científicas.

1. Introducción

Las centrales solares de torre (CST) destacan por su eficiencia termodinámica y capacidad de operar con reserva de calor [2,6]. Según la temática de este trabajo, las CST pueden definirse como un conjunto de espejos altamente reflectantes, llamados helióstatos, y un receptor de radiación. Los helióstatos siguen el movimiento aparente del Sol y concentran la radiación en el receptor. Se alcanza por tanto una alta densidad de radiación sobre su superficie. Esta energía se transfiere a un fluido caloportador que circula en su interior. Éste es utilizado finalmente en un ciclo de turbina para generar electricidad. Estas ideas se ilustran en la Figura 1. El lector puede encontrar más información sobre este tipo de centrales en [1].

El campo de helióstatos supone hasta el 50% de los costes de inversión iniciales y puede llegar a causar hasta un 40% de pérdida energética en operación [3,6]. Por consiguiente, su diseño debe ser cuidadosamente optimizado (capacidad de concentración, coste, etc.). Sin embargo, es un problema complejo en el que [5]: i) las coordenadas de cada helióstato deberían considerarse directamente

Computación de Altas Prestaciones en el Diseño Óptimo y Control de Plantas Solares de Torre

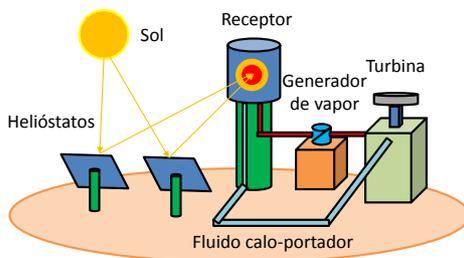


Figura 1. Esquema de una planta solar de recepción central.

(y los campos comerciales tienen cientos de helióstatos) y ii) la función objetivo es computacionalmente costosa, multi-modal y de estructura no explotable. Por este motivo, la mayoría de estrategias de diseño se apoyan de un modo u otro en a) patrones de distribución conocidos y con un conjunto reducido de parámetros y b) selección de posiciones disponibles [5,12]. Además, estas estrategias suelen combinarse. Un patrón de distribución conocido es el clásico ‘radial-staggered’ (usado en [6,14,15]). Una propuesta más reciente es la que se hace en [14], donde se estudia el uso de un innovador esquema bio-inspirado con forma de espiral. En relación a la selección directa de posiciones, el método de [17], en el que se forma una malla sobre la superficie disponible, es bastante conocido. Los trabajos de [3,14,15] son ejemplos de combinar uso de patrones y selección de posiciones. Es una buena forma de atenuar la rigidez de los patrones. Estrategias especialmente interesantes por su flexibilidad son las de [4,12]. En [4] se refina con precisión la posición de los helióstatos a partir de un campo preliminar. En [12] se aborda el problema directamente con un método de gradiente.

Este trabajo se enmarca en la línea de optimización directa del campo solar. Para tal fin se ha diseñado un algoritmo genético con una estructura robusta y paralela. Aunque no es una idea totalmente innovadora, el uso de este tipo de algoritmos ha sido mayormente vinculado a patrones de distribución [3,4,15]. Al contar en este caso con un espacio de búsqueda mucho más complejo, es vital poder explorar muchas soluciones por unidad de tiempo. El método incluye además un punto para inyectarle conocimiento específico. También se ha desarrollado una nueva meta-heurística para el diseño del campo de helióstatos. Aspira a facilitar la resolución directa del problema a cualquier optimizador incluido el previamente mencionado. Para hacerlo, se va alterando la dimensionalidad y el espacio de búsqueda del problema que se le presenta al optimizador seleccionado.

2. Avances

2.1. Planteamiento del problema

El problema planteado consiste en ubicar un cierto número de helióstatos, sobre una superficie plana, de forma que concentren la máxima potencia posible

Computación de Altas Prestaciones en el Diseño Óptimo y Control de Plantas Solares de Torre

en su receptor. Sea H el total de helióstatos a desplegar. Todos se asumen con el mismo tamaño y especificaciones (como se suele hacer, para aprovechar las ventajas de su producción a gran escala). Su superficie reflectante es rectangular y tiene un tamaño de $l \times w$ (alto y ancho respectivamente). Cada helióstato h puede identificarse en el campo por su punto central, $C_h = (x_h, y_h)$ (coordenadas Cartesianas). Por lo tanto, un campo de H helióstatos puede definirse como un vector $F = (C_1, \dots, C_H) = (x_1, y_1, \dots, x_H, y_H)$ en \mathbb{R}^{2H} .

Según las definiciones previas, el problema tendrá $2H$ dimensiones, dos coordenadas por helióstato. Sea $P_T(F)$ la potencia total reflejada sobre el receptor por un campo F a lo largo de un conjunto T de instantes de interés (posiciones solares con la densidad de radiación incidente asociada), $T = \{t_1, \dots, t_T\}$. Dependiendo del caso, T puede ser desde un único instante (optimización en punto de diseño) hasta muchos para abarcar, por ejemplo, todo un año. Ésto varía la complejidad del problema y, sobre todo, su coste computacional. $P_T(F)$, que será la función objetivo del problema, puede ser analíticamente expresada como se muestra en la Expresión (1) [12,13]:

$$P_T(F) = A \sum_{t=t_1}^T I_t \left(\sum_{h=h_1}^H \eta_h(t) \right) \quad (1)$$

donde A , es el área reflectante del modelo de helióstato ($\approx l \times w$ (m²)) e I_t es la densidad de radiación incidente en el instante t (kW/m²). En relación a $\eta_h(t)$, se trata del factor de eficiencia instantáneo del helióstato h en el instante t (de 0 a 1, mínimo y máximo respectivamente). Éste depende de i) el instante, ii) la posición del helióstato h en relación al receptor y iii) los otros helióstatos por la interacción indirecta entre ellos. Su cálculo implica simular el comportamiento del campo candidato. Como se explica en [3,14,18], este factor modela y abarca diferentes fuentes de pérdida energética en operación. De hecho, se puede descomponer en diferentes sub-factores. La definición escogida en este trabajo se muestra en la Expresión (2), y es la misma seleccionada en [3,13,14].

$$\eta_h = \eta_{cos} \eta_{sb} \eta_{itc} \eta_{aa} \eta_{ref} \quad (2)$$

A continuación se resumen los factores incluidos en la Expresión (2):

- η_{cos} (**Pérdida coseno**): El área útil de un helióstato se reduce por el coseno del ángulo de incidencia de la radiación solar. Se calcula según [18].
- η_{sb} (**Pérdida por bloqueo y sombreado**): Cada helióstato puede obstruir parcialmente la radiación incidente (sombreado) o reflejada (bloqueo) por cualquier otro. Se calcula según el método de [16] con el filtrado de candidatos usado en [14]. Es el factor computacionalmente más costoso.
- η_{itc} (**Pérdida de intercepción**): El mapa de flujo reflejado por los helióstatos puede caer parcialmente fuera del receptor. Se calcula con el modelo propuesto en [13] que no depende del tiempo y se obtiene más rápidamente.
- η_{aa} (**Pérdida por atenuación atmosférica**): La atmósfera atenúa la radiación reflejada de los helióstatos en su trayectoria. Se estima con el mismo modelo usado en [14] (también independiente del tiempo).

- η_{ref} (**Pérdida por reflectividad**): Los helióstatos no puede garantizar una reflexión sin pérdida. Se considera una constante de fabricación como en [14].

Debe también considerarse que: i) la base del receptor es el origen de coordenadas, ii) los helióstatos no pueden estar a una distancia menor que R_{min} ni mayor que R_{max} de dicho punto y iii) los helióstatos deben poder mover su superficie reflectante sin colisionar con otros, pudiendo calcularse la distancia de seguridad como $d = \sqrt{l^2 + w^2}$ (véase la Figura 2). El problema de optimización puede formalizarse entonces como se incluye en la Expresión (3). Por lo tanto, hay un total de $H(H - 1)/2 + 2H$ restricciones que cumplir (aprovechando que la distancia entre un punto a y otro b es la misma en ambos sentidos). La dimensionalidad del problema es pues, potencialmente muy grande.

$$\begin{aligned}
 & \underset{F}{\text{maximizar}} && P_T(F) \\
 & \text{sujeto a} && \sqrt{x^2 + y^2} \geq R_{min} + \frac{d}{2}, \forall C = (x, y) \\
 & && \sqrt{x^2 + y^2} \leq R_{max} - \frac{d}{2}, \forall C = (x, y) \\
 & && \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \geq d, \forall C_i = (x_i, y_i), C_j = (x_j, y_j) : (i \neq j)
 \end{aligned} \tag{3}$$

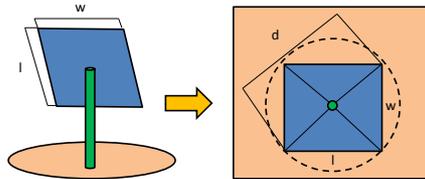


Figura 2. Distancia de seguridad de un helióstato.

2.2. Optimizador genético desarrollado

Los algoritmos genéticos (AG) [11] se suelen usar para resolver problemas complejos de optimización. Ésto se debe a que su base teórica no está ligada a ningún problema en particular sino a la propia evolución de las especies. Generan una población de soluciones candidatas ('individuos') y se simula su evolución hasta que se cumple una condición de parada. Sin embargo, por su propia abstracción, estos métodos se adaptan en última instancia al problema objetivo, lo que suele implicar seleccionar sus operadores y alcance.

Computación de Altas Prestaciones en el Diseño Óptimo y Control de Plantas Solares de Torre

Para el problema planteado se ha diseñado el AG que se describió en [10]. Por el coste computacional del problema, su objetivo de diseño era estar adaptado a una ejecución en paralelo directamente. Esta estrategia adapta todo el proceso para explotar los entornos de computación de alto rendimiento. Así, se atenúa el coste computacional del problema y se pueden abordar mayores instancias en mejores condiciones.

La estructura de los individuos consiste en vectores de longitud $2H$ con un campo adicional para registrar la calidad del campo que representan, es decir, la potencia total reflejada sobre el receptor tras su evaluación (valor de la función objetivo, Expresión (1)). En la Figura 3 se muestra un esquema ilustrativo destacando que cada par de coordenadas corresponde a un cierto helióstato. Sin embargo, los AG están más orientados a optimización sin restricciones [19], lo que requiere una adaptación en este caso. El problema se va a abordar como uno sin restricciones. Sin embargo, cualquier solución que viole restricciones (infactible) tendrá un valor de calidad muy bajo. La penalización dependerá del grado de violación. Esta estrategia, común en este ámbito, se conoce como ‘penalización estática’ [19]. En consecuencia, el algoritmo ignora las restricciones de la Expresión (3), pero altera la evaluación de cualquier solución candidata que las viole según la Expresión (4). En dicha fórmula, m_c es la distancia entre el helióstato c y la base de la torre, es decir, $m_c = \sqrt{x_c^2 + y_c^2}$, e I_T es la suma total de densidad de radiación incidente en cada instante de T . Finalmente, V es un conjunto abstracto que contiene un registro con cada helióstato c en F y las restricciones que viola (para calcular sólo los factores implicados). Cada helióstato en V tiene además un conjunto especial asociado, V_c , en el que se lista cualquier

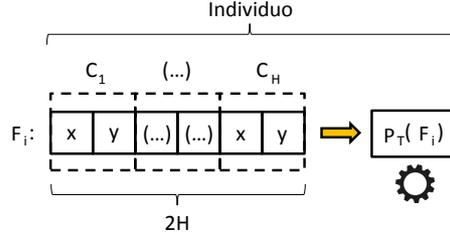


Figura 3. Diagrama de un individuo.

$$eval(F) = \begin{cases} P_T(F), & \text{si } V = \emptyset \\ 0 - AI_T \left(\frac{(R_{min}+d/2)-m_c}{(R_{min}+d/2)} + \frac{m_c-(R_{max}-d/2)}{m_c} + \frac{d-dist(C_c, C_z)}{d} \right); \forall c \in V, \forall z \in V_c & \text{si no} \end{cases} \quad (4)$$

Computación de Altas Prestaciones en el Diseño Óptimo y Control de Plantas Solares de Torre

En relación al procedimiento del algoritmo, los parámetros que espera recibir son los siguientes:

- *pop_size*: tamaño de la población, que se mantiene constante durante la búsqueda.
- *num_pairs*: número de parejas a formar durante el cruce.
- *tourn_size*: tamaño del torneo para la selección (tanto para cruce como para reemplazo).
- *ov_mut*: probabilidad de mutación de un descendiente.
- *per_mut*: probabilidad de alterar cada helióstato de un descendiente una vez que la mutación ha comenzado.
- *cycles*: número de generaciones o ciclos a ejecutar.

Nótese además que, aunque no se destaca, es necesario contar con información contextual (p. ej. tamaño de los helióstatos, del receptor, etc.) para evaluar las soluciones candidatas. El procedimiento ejecuta los pasos mostrados en el Algoritmo 1. Como se puede apreciar, es un proceso evolutivo tradicional en el que las partes que requieren evaluar la función objetivo se distribuyen entre hilos concurrentes (en un entorno de memoria compartida). De hecho, todos ellos están forzados a esperar a que el hilo principal defina la población del siguiente ciclo antes de avanzar en la línea 12. De esta forma, los hilos comparten el coste de evaluar la función objetivo (que hace falta en la inicialización (línea 5), reproducción (línea 9) y mutación (línea 10)) mientras se mantiene una población común. Las funciones *createThreads*, *runInParallel* y *getChunkSize* hacen referencia a la creación de un conjunto de hilos para operar en distintos rangos de la matriz de población. De forma similar, la etiqueta *synchr* indica que la actualización de una cierta variable debe ser síncrona. La etiqueta *barrier_master_do* declara que la operación debe ser ejecutada por el hilo principal mientras el resto le espera. Los detalles de cada operador del algoritmo se describen a continuación.

La función *GenerateInitialPop* (línea 5) es responsable de crear tantas soluciones candidatas como se requiera por el tamaño de la población. Sin embargo, también es necesaria su evaluación según la Expresión (4) para formar individuos completos. De hecho, los helióstatos en este punto sólo están forzados a respetar las restricciones ligadas a R_{min} y R_{max} para limitar su posición, pero sin prestar atención a posibles colisiones. Por tanto, la existencia de soluciones parcialmente infactibles es posible desde el principio. Sin embargo, la etiqueta especial ‘*INJECT?*’ declara la posibilidad de incluir algunos ‘individuos especiales’. Este concepto hace referencia a añadir algún individuo obtenido a partir de un patrón de distribución como la espiral biomimética de [14]. No obstante, ésto es un arma de doble filo porque permite añadir conocimiento fiable a la población desde su origen, pero también marca una fuerte influencia en la misma ya que la aptitud de esos individuos será mucho mayor que la de los otros. Podría por tanto inducirse una deriva genética importante que diera lugar a una convergencia prematura. Esta opción debería evitarse o, al menos minimizarse, en la medida de lo posible.

La función *SelectProgenitors* (línea 8) simplemente busca dos progenitores para cada pareja. Para hacerlo, se aplica selección por torneo. Este método es

Computación de Altas Prestaciones en el Diseño Óptimo y Control de Plantas Solares de Torre

```

Input: Int  $pop_{size}, num_{pairs}, tourn_{size}, cycles$ , Real  $ov_{mut}, per_{mut}$ 
Output: Vector  $F$  in  $\mathbb{R}^{2H}$ 
IndividualSet  $pop$ , Individual  $bestIndividual$ ; /* Compartido entre hilos */
ThreadTeam  $threads = createThreads()$ ; /* Crear equipo */
 $threads.runInParallel()$ ; /* Dentro de cada hilo: */
Integers  $range = getChunkSize()$ ; /* Calcular mi trabajo como hilo */
 $pop = GenerateInitialPop(pop_{size}, range, INJECT?)$ ;
 $bestIndividual < synchr > = UpdateBest(pop)$ ;
for  $i = 1$  to  $cycles$  do
  IndividualSet
  |  $progs = SelectProgenitors(pop, range < num_{pairs} >, tourn_{size})$ ;
  | IndividualSet  $desc = Reproduce(progs)$ ;
  | IndividualSet  $descMut = Mutate(desc, ov_{mut}, per_{mut})$ ;
  |  $bestIndividual < synchr > = UpdateBest(desc, descMut)$ ;
  |  $pop < barrier\_master\_do > = Replace(pop, descMut, KEEP\_BEST)$ ;
end
return  $bestIndividual.F$ 

```

Algoritmo 1: Optimizador genético del campo de helióstatos.

uno de los más populares por combinar uniformidad de exploración con presión selectiva ajustable. Por consiguiente, todo progenitor de cada pareja se selecciona como el mejor de una muestra aleatoria de $tourn_{size}$ participantes. Esta estrategia se sigue igualmente en la función *Replace* (línea 12), en la que el hilo principal forma la población del ciclo siguiente a partir de $pop_{size} - 1$ torneos. El término ‘-1’ se debe a que la mejor solución conocida se pasa directamente a la población siguiente. En caso contrario, no habría garantía de que participara en ningún torneo y podría perderse. Este enfoque se llama ‘elitismo’ en el ámbito de los AG, y se basa en la idea de que la estructura de una buena solución puede orientar a las otras a mejores zonas del espacio de búsqueda.

La función *Reproduce* (línea 9) coge las parejas y obtiene de cada una de ellas dos descendientes. Para hacerlo, se aplica cruce uniforme. Este método permite un alto grado de hibridación, que propicia exploraciones más completas del espacio de búsqueda. Su procedimiento consiste en los siguientes pasos: Primero, se forma una máscara auxiliar de reproducción al azar. Se trata de una cadena binaria de longitud H (un bit por cada helióstato o pareja de coordenadas C), en la que cada bit tiene la misma probabilidad de ser un 0 o un 1. En segundo lugar, un primer descendiente se forma tomando los helióstatos (coordenadas x e y) de su progenitor i para cada posición en la que la máscara auxiliar tiene un 1, mientras que se toman del progenitor $i + 1$ en caso contrario. En tercer lugar, la máscara se invierte y un segundo descendiente se obtiene aplicando las mismas reglas. Cualquier nuevo descendiente deberá ser evaluado según la Expresión (4).

La función *Mutate* (línea 10) está destinada a permitir que la población pueda alcanzar zonas completamente nuevas del espacio de búsqueda. Para hacerlo, cada descendiente tiene una probabilidad de ov_{mut} de sufrir algún tipo de mutación. En concreto, su conjunto de helióstatos es recorrido y cada uno tiene una probabilidad de per_{mut} de ser aleatoriamente reposicionado. Debe ten-

Computación de Altas Prestaciones en el Diseño Óptimo y Control de Plantas Solares de Torre

erse en cuenta que los individuos alterados suponen nuevas soluciones y, como tales, deben ser evaluadas de nuevo. Además, cuando se aplica, el proceso de mutación puede sobrescribir soluciones prometedoras y hacerlas peores (para admitir variabilidad). Por este motivo, se mantiene una copia previa de los no mutados para actualizar la mejor solución conocida en caso de ser necesario.

Finalmente, el método devuelve el mejor vector de coordenadas que definen un campo que se haya encontrado a lo largo de la búsqueda. Tras implementar el método propuesto se ha comprobado que se logran soluciones aceptables con una aceleración lineal con el número de hilos. Se ha estimado además que para un T anual (al menos 192 instantes de interés), la versión paralela resolvería el problema en poco más de un mes mientras que, secuencialmente, necesitaría más de un año. El trabajo completo está sometido a revisión en la revista Journal of Supercomputing [8].

2.3. Meta-heurística sectorial

En la sección 2.1 se plantea un problema de optimización continua de $2H$ variables y $(H^2 - H)/2 + 2H$ restricciones. Hay que colocar, en una superficie, H heliostatos sin que colisionen entre sí y maximizando su eficiencia óptica. Esta depende de los factores coseno, de intercepción, de atenuación, de reflectividad y de bloqueo y sombreado. Afortunadamente, obviando el bloqueo y sombreado, el resto de factores pueden considerarse exclusivamente ligados a cada heliostato y su posición. Por lo tanto, si se divide el espacio disponible en sectores circulares independientes como se muestra en la Figura 4, la complejidad del problema se reduce drásticamente. Estas regiones se presentan al optimizador como

mc
po:
[7]

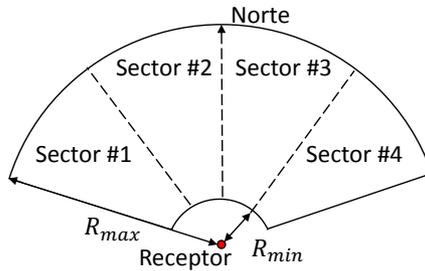


Figura 4. Distancia de seguridad de un heliostato.

Evidentemente, las evaluaciones de la función objetivo serían algo menos precisas. Sin embargo, se parte de la idea de que el problema completo es tan complejo que su resolución óptima es inviable actualmente. Se espera entonces

que un mismo optimizador logre un mejor resultado final abordando partes independientes y relativamente sencillas que tratando de abarcar el problema en toda su extensión. Además, las trayectorias desde cada helióstato al receptor estarían contenidas en cada sector. Por este motivo, el bloqueo podría incluso estimarse con un error despreciable y la única fuente de imprecisión real sería el sombreado. Estos fenómenos, de hecho, suelen registrarse en el entorno inmediato de cada helióstato [3]. Esta consideración acotaría aún más el error de aplicar esta meta-heurística con cualquier optimizador.

Habiendo combinado esta meta-heurística con el algoritmo anterior, se ha comprobado que el mismo optimizador logra mejores soluciones a través de ella que independientemente. El artículo donde se detalla está sometido a revisión en la revista [9].

3. Conclusiones

En este trabajo se ha descrito formalmente el problema de optimización del campo de helióstatos. No se aborda desde un patrón de distribución conocido ni dado un conjunto de posiciones disponibles sino que se hace directamente sobre las coordenadas de cada helióstato. Este método tiene potencial para superar los diseños que se apoyan en los otros métodos y que hacen una búsqueda, por tanto, más limitada. Una vez formulado el problema completo y orientado a maximizar la potencia concentrada por el receptor, se han diseñado un algoritmo genético adaptado y una meta-heurística con la que sería combinable. El algoritmo tiene una estructura minimalista y paralela que le permite explorar muchas soluciones por unidad de tiempo. El diseño de sus operadores le lleva también hacia una convergencia adecuada. Se apoya en i) elitismo, ii) cruce uniforme, iii) penalización estática y iv) selección por torneo. Incluye además la posibilidad de inyectar conocimiento mediante individuos con patrones conocidos. La meta-heurística reduce la complejidad que el optimizador escogido percibe del problema creando subproblemas independientes en sectores circulares. Se puede mantener entonces una estrategia directa para casos complejos aunque se sacrifique algo de precisión en evaluación. Este trabajo ha dado lugar a dos publicaciones en actas de congreso internacional. Además, las versiones detalladas de ambos, y que incluyen experimentación exitosa, están en revisión en dos revistas científicas internacionales.

Referencias

1. S. Alexopoulos and B. Hoffschmidt. Advances in solar tower technology. *WIREs Energy Environ.*, 6(1):1–19, 2017.
2. A. L. Avila-Marin, J. Fernandez-Reche, and F. M. Tellez. Evaluation of the potential of central receiver solar power plants: Configuration, optimization and trends. *Appl. Energ.*, 112:274–288, 2013.
3. S. M. Besarati and D. Y. Goswami. A computationally efficient method for the design of the heliostat field for solar power tower plant. *Renewable Energy*, 69:226–232, 2014.

Computación de Altas Prestaciones en el Diseño Óptimo y Control de Plantas Solares de Torre

4. R. Buck, A. Pfahl, and T. H. Roos. Target aligned heliostat field layout for non-flat terrain. *Proceedings of 1st Southern African Solar Energy Conference*, 2012.
5. E. Carrizosa, C. Domínguez-Bravo, E. Fernández-Cara, and M. Quero. An optimization approach to the design of multi-size heliostat fields. Technical report, Technical report IMUS. 2014, 2014.
6. F. J. Collado and J. Guallar. Campo: Generation of regular heliostat fields. *Renew. Energ.*, 46:49–59, 2012.
7. N. Cruz, J. Redondo, J. Álvarez, M. Berenguel, and P. Ortigosa. Sector-based division for continuous heliostat field optimization. In *Proceedings of the International Workshop on Optimization and Learning: Challenges and Applications (OLA2018)*, Alicante, Spain, February 2018.
8. N. Cruz, S. Salhi, J. Redondo, J. Álvarez, M. Berenguel, and P. Ortigosa. Design of a parallel genetic algorithm for continuous and pattern-free heliostat field optimization. *Preprint submitted to The Journal of Supercomputing*, 2017.
9. N. Cruz, S. Salhi, J. Redondo, J. Álvarez, M. Berenguel, and P. Ortigosa. Hector: Sector-based heliostat field optimization. *Preprint submitted to Advanced Engineering Informatics*, 2017.
10. N. Cruz, S. Salhi, J. Redondo, J. Álvarez, M. Berenguel, and P. Ortigosa. A parallel genetic algorithm for continuous and pattern-free heliostat field optimization. In *Proceedings of the 17th International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE 2017*, volume 2, pages 684–694, Rota, Spain, July 2017. ISBN 978-84-617-8694-7.
11. J. H. Holland. Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence. *Michigan Press*, 1975.
12. S. Lutchman, A. Groenwold, P. Gauché, and S. Bode. On using a gradient-based method for heliostat field layout optimization. *Energy Procedia*, 49:1429–1438, 2014.
13. S. L. Lutchman. Heliostat field layout optimization for a central receiver. Master's thesis, Stellenbosch University, 2014.
14. C. J. Noone, M. Torrilhon, and A. Mitsos. Heliostat field optimization: A new computationally efficient model and biomimetic layout. *Solar Energy*, 86(2):792–803, 2012.
15. R. Pitz-Paal, N. B. Botero, and A. Steinfeld. Heliostat field layout optimization for high-temperature solar thermochemical processing. *Solar Energy*, 85(2):334–343, 2011.
16. A. Ramos and F. Ramos. Heliostat blocking and shadowing efficiency in the videogame era. *arXiv preprint arXiv:1402.1690*, 2014.
17. M. Sánchez and M. Romero. Methodology for generation of heliostat field layout in central receiver systems based on yearly normalized energy surfaces. *Solar Energy*, 80(7):861–874, 2006.
18. W. B. Stine and M. Geyer. *Power from the Sun*. Public website: <http://www.powerfromthesun.net/book.html> (Último acceso: Mayo, 2017), 2001.
19. Ö. Yeniay. Penalty function methods for constrained optimization with genetic algorithms. *Mathematical and Computational Applications*, 10(1):45–56, 2005.

Modelado y Optimización para una Gestión eficiente de Recursos en Desalación Solar

Modelado y Optimización Campo NEP

José Carballo

Universidad de Almería UAL

Abstract. El trabajo presenta la motivación de la tesis *Modelado y Optimización para una Gestión eficiente de Recursos en Desalación Solar*, muestra los objetivos principales que se plantean y el estado de desarrollo o consecución de los mismos. Además se muestran los principales resultados y conclusiones obtenidos durante el último año gracias a los trabajos desarrollados en el marco de la tesis.

1 Introducción.

Algunos de los principales problemas que debe afrontar la humanidad en las próximas décadas derivan de la escasez de agua dulce, debido al cambio climático, la salinización o contaminación de las fuentes (1). Actualmente, las principales fuentes de consumo de agua dulce actualmente son de origen industrial, humana, agrícola y energética. En lo referente a la demanda industrial de agua dulce, se prevé que esta aumente en todos los sectores de la producción (2). La demanda por consumo humano apunta hacia un aumento dado que la población mundial está creciendo a un ritmo de 80 millones de personas por año, pudiéndose alcanzar los 9,1 mil millones en el año 2050 (3). La demanda de agua en agricultura y producción de energía, representa el 15 % y el 70% del total respectivamente, previéndose un aumento de la demanda total en ambos sectores. Con estas predicciones, las Naciones Unidas pronostican que entre 2-7 mil millones de personas deban enfrentarse a escasez de agua a mitad de siglo (4).

Por ello, la gestión racional de los recursos naturales como el agua dulce y la promoción de las energías limpias y renovables en sustitución de las energías fósiles que contaminan y esquilman a su vez estos recursos naturales se ha convertido en tema principal para numerosos organismos que advierten de las ventajas ambientales, estratégicas y socioeconómicas del uso de las energías renovables frente a las energías fósiles y la gestión eficiente de los recursos.

La Unión Europea por ejemplo, considera estos hechos en gran parte de sus actividades de investigación e innovación en el Programa Marco denominado Horizonte 2020 (*H2020*). España por su parte los contempla en las líneas estratégicas de investigación del Plan Nacional de Investigación (Estrategia Española para la Ciencia, Tecnología e Innovación 2013-2020), en el que se encuadra el

Modelado y Optimización para una Gestión reciente de Recursos en Desalación Solar Modelado...

proyecto “Estrategias de control y gestión energética en entornos productivos con apoyo de energías renovables” (*ENERPRO*) del que depende la presente investigación.

Este proyecto trata de analizar, diseñar y aplicar técnicas de modelado, control y optimización (en el ámbito del control jerárquico y control predictivo basado en modelo MPC) para conseguir una gestión eficiente de energía (electricidad y calor/frío de proceso), agua y CO_2 , en sistemas productivos apoyados en energías renovables y sistemas de almacenamiento. Mediante una gestión óptima de estos recursos y la adaptación de la generación a la demanda, se pretende demostrar que el control automático permite conseguir ahorros económicos y reducir el impacto medioambiental en la explotación de procesos complejos.

ENERPRO es un proyecto coordinado entre la Universidad de Almería y la Plataforma Solar de Almería a través de *CIESOL* formado por dos subproyectos. El subproyecto desarrollado por la Plataforma Solar de Almería, en el cual se encuadra la tesis, tiene como objetivos específicos: Modelado dinámico de una planta híbrida solar-gas de desalación térmica, desarrollo de modelos de predicción de variables de contorno, perturbaciones, fuentes y cargas, análisis y modelado de sistemas de almacenamiento y otros sistemas auxiliares para la reducción de costes, obtención de modelos simplificados para tareas de control, desarrollo de estrategias *MPC* (jerárquicas, híbridas y económicas) para plantas de desalación, acoplamiento de plantas de desalación como fuentes de agua y energía a invernaderos y edificios, prueba de los algoritmos en simulación y condiciones reales, transferencia tecnológica a compañías e instituciones interesadas en el proyecto, diseminación de los resultados en diferentes foros nacionales e internacionales y a diferentes niveles.

Como se puede apreciar en la lista de objetivos específicos, este subproyecto se centra en el campo de la desalación, que cobra especial relevancia si se tiene en cuenta que el 50 % de la población está viviendo en territorios costeros y que los avances en el campo de las renovables hacen posible la obtención de agua dulce mediante desalación con energía solar. En la actualidad una de las principales tecnologías de desalación es la destilación térmica (5). Las principales ventajas de los procesos de destilación térmica son su capacidad para demandar una fuente térmica de baja temperatura, su fiabilidad, operación y mantenimiento sencillos, obtención de agua dulce de alta pureza y su capacidad de operar con todo tipo de aguas de alimentación, a alta temperatura, muy salobres o incluso contaminadas.

Las técnicas de destilación térmicas son consideradas energéticamente ineficientes, puesto que en cualquier proceso de destilación el coste energético es al menos el 50% del coste total. La destilación térmica demanda dos formas diferentes de energía: energía térmica que representa la mayor parte de la energía consumida y que puede provenir de diferentes fuentes (combustibles fósiles, energía residual (6), energía solar (8)), y energía eléctrica consumida principalmente por los actuadores y sistemas de bombeo.

Por todo lo anterior, en esta tesis se va a tratar la aplicación de técnicas de modelado, control y optimización para lograr una gestión eficiente de energía y agua dulce mediante el uso de energía solar.

1.1 Objetivos de la tesis.

Se pretende realizar contribuciones en el campo del modelado, control y optimización de sistemas termo solares con aplicaciones industriales. De manera general se pretende estudiar, modelar y optimizar, a partir de los estudios previos que muestran el gran potencial, la viabilidad técnica y económica del sistema *AQUASOL* (Fig. 1) en la configuración formada por la planta de destilación térmica acoplada a una bomba de calor de doble efecto cuya fuente de energía térmica es un campo solar y una caldera de gas.

Dentro de los objetivos de la tesis se encuentra, en primer lugar, realizar un análisis bibliográfico completo de toda la temática involucrada. En segundo lugar, el modelado de la *MED*, análisis y estudio de los puntos de funcionamiento óptimos y criterios de optimización. En tercer lugar, modelado dinámico y análisis de los puntos de funcionamiento óptimos de la *DEAHP*. Este sistema precisa de control automático en las bombas que impulsan el bromuro de litio, que es fundamental para asegurar la estabilidad.

Ampliados y validados los modelos de la *MED* y *DEAHP*, desarrollar un modelo conjunto y estudiar las condiciones más favorables de funcionamiento respecto a los criterios anteriormente mencionados para el sistema integrado por ambos sistemas. Posteriormente desarrollar y validar modelos dinámicos del generador de vapor, la caldera de gas y campo *NEP*, para hacer posible el estudio del comportamiento de de los sistemas y la obtención de los parámetros óptimos de funcionamiento.

Finalmente acoplamiento de modelos, que requerirá previsiblemente del diseño de lazos de control de bajo nivel y el modelado de subsistemas como válvulas, bombas etc... Acoplados todos los sistemas se procederá a la validación y la optimización del sistema completo junto con el pertinente análisis económico.

1.2 Instalación experimental: Sistema Aquasol.

El trabajo experimental necesario para la elaboración de esta tesis doctoral será realizado en la instalación *AQUASOL*. Esta instalación data de comienzos de 1987 cuando una unidad desoladora *MED* fabricada por *ENTROPIE* fue instalada en la Plataforma Solar de Almería, con el fin de probar y desarrollar el proceso de destilación acoplado con energía solar térmica. Posteriormente esta unidad ha sufrido varias modificaciones, las más importantes dentro del marco del proyecto *AQUASOL* (9), que finalmente ha dado su nombre a la instalación. Con el fin de cumplir con los objetivos del proyecto *AQUASOL* (Enhanced Zero Discharge Seawater Desalination using Hybrid Solar Technology), en 2002 el sistema se remodeló con los que se esperaba reducir el coste de producción de agua desalada en sistemas solares *MED*.

Modelado y Optimización para una Gestión reciente de Recursos en Desalación Solar Modelado...

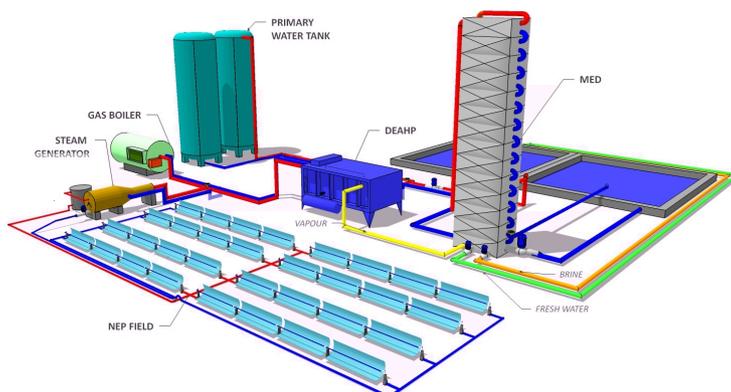


Fig. 1: Aquasol

El principal objetivo era la incorporación eficiente de la energía solar al proceso (10), para ello se añadieron los siguientes equipos: un campo solar de 44 captadores de placa plana de gran apertura, almacenamiento térmico, una nueva bomba de calor de absorción de doble efecto, un secador solar avanzado, un generador de vapor y un pequeño campo solar de colectores cilindro parabólicos (11).

La actual configuración Fig. 1 hace que el sistema sea flexible respecto a la fuente energética y lo dota de tres modos de operación (12): Modo solar: El campo de colectores provee toda la energía térmica del sistema. La planta *MED* es alimentada a través de los tanques. Modo fósil: La caldera de gas propano a través de la bomba de calor proporciona toda la energía térmica del sistema. En este modo, la *DEAHP* (Double effect absorption heat pump) puede conectarse directamente a la planta *MED* o indirectamente a través de los tanques (13). Modo híbrido: La energía térmica proviene de ambas fuentes, de la bomba de calor y del campo solar. La *DEAHP* permite la operación a carga parcial que puede variar del 30% al 100% estrangulando el flujo de vapor que circula entre la caldera y la bomba. En este modo, la alimentación de la planta *MED* se realiza a través de los tanques. f:thermalmodel

2 Avances y resultados.

2.1 Modelado.

Para el modelado de la planta *MED* (Fig. 2), se ha adaptado un modelo basado en primeros principios de la planta, implementado en el lenguaje de modelado *Modelica* (15) y previamente desarrollado, calibrado y validado (16). Dicho modelo fue modificado y completado con el desarrollo matemático que permite el análisis energético y exergético de la planta, además permite determinar nuevos índices de rendimiento que toman en cuenta tanto la cantidad como la calidad de la energía empleada. Aunque no muy extendido el análisis exergético es un

Modelado y Optimización para una Gestión reciente de Recursos en Desalación Solar Modelado...

método clave para la mejora del rendimiento de este tipo de procesos (17). Los datos obtenidos a través del modelo muestran un buen ajuste con los datos reales, resultando un modelo preciso.

Para el modelado de la *DEAHP*, se partió del modelo desarrollado por de la Calle et al (16; 20), basado en primeros principios e implementado en el lenguaje de modelado *Modelica*. Al igual que en el caso anterior, el modelo fue modificado y completado con todo el planteamiento matemático que permite el análisis energético y exergético de la planta.

Posteriormente los modelos de la planta *MED* y la *DEAHP* fueron acoplados (Fig. 2) y validados con datos experimentales medidos en las instalaciones reales. El modelo resultante aun siendo bastante preciso es demasiado complejo e inestable, las simulaciones llevadas a cabo demandan un alto esfuerzo computacional.

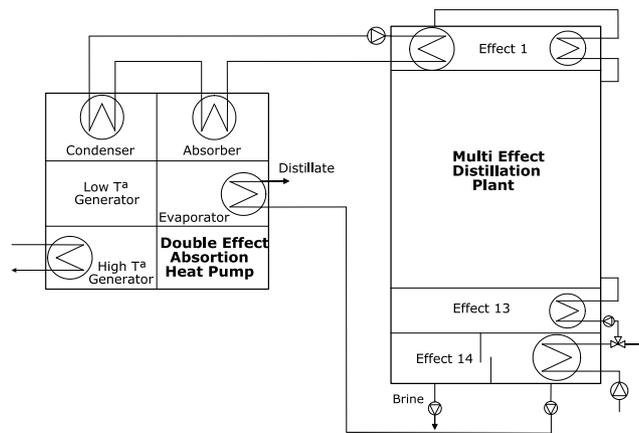


Fig. 2: DEAHP-MED

Finalmente se ha desarrollado un modelo flexible y dinámico que permite simular el comportamiento térmico transitorio del campo *NEP* (Fig. 3a), para su uso en tareas de optimización, evaluación del rendimiento y control. Este modelo de la planta *NEP* se basa en los principios físicos detallados descritos en el trabajo R.Forristal (21) Fig. 3b, el modelo se ha adaptado, validado y mejorado con el desarrollo exergético necesario para poder realizar análisis exergéticos, como recomienda el autor. Se ha desarrollado utilizando la librería *ThermoCycle* (22; 23) y se ha implementado en el lenguaje de modelado orientado a objetos basado en ecuaciones *Modelica*. Este modelo resulta ser flexible, preciso y altamente configurable, lo que permite adaptarlo fácilmente y le otorga un amplio campo de aplicación.

Modelado y Optimización para una Gestión reciente de Recursos en Desalación Solar Modelado...

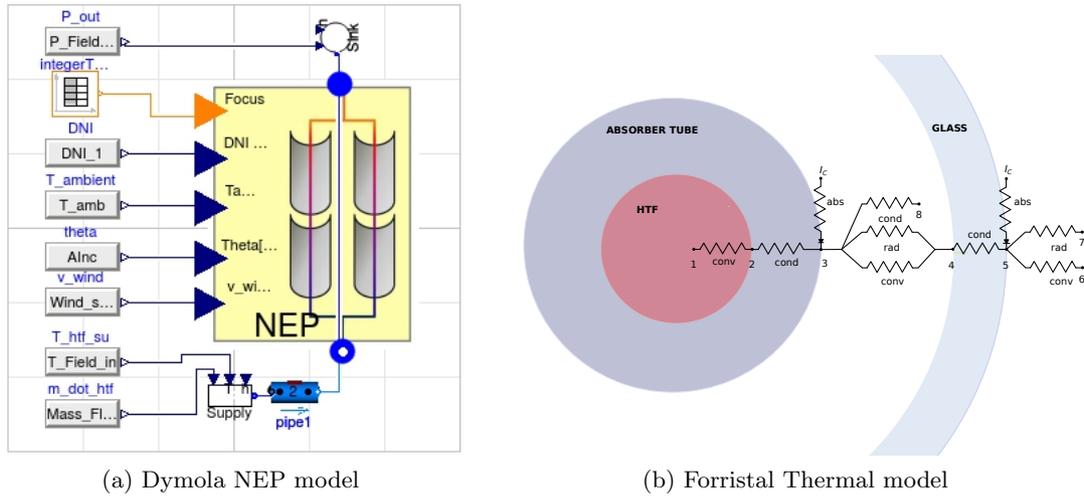


Fig. 3: NEP model

2.2 Optimización.

Con el objetivo de reducir los costes energéticos y mejorar la operación del proceso de desalación térmica, se realizó un estudio (14) sobre la influencia de los parámetros operacionales y las condiciones óptimas de operación de la planta *MED*. El estudio consta de tres partes: análisis energético y exergético de la planta en condiciones nominales, estudio de sobre la influencia de cada uno de los parámetros de operación de la planta y una optimización multicriterio mediante algoritmos genéticos y nuevos criterios de optimización exergética.

Todo el proceso de optimización multicriterio de los parámetros de funcionamiento se realizó mediante el empleo de las herramientas de la librería de optimización implementada en *DYMOLA* (18). El análisis energético y exergético del proceso en condiciones nominales, reveló las principales fuentes de consumo energético y exergético. Por un lado, la energía térmica contenida en los flujos de agua salientes son la principal fuente responsable del alto consumo de energía, por otro lado la generación de entropía y el calor intercambiado con el ambiente son los responsables de la degradación de la exergía. En segundo lugar, el estudio sobre la influencia de los parámetros de operación de la planta, muestra que la producción de la planta real en condiciones nominales está lejos de los valores máximos simulados y que algunos parámetros operativos no tienen una gran influencia en el proceso con respecto al resto.

Finalmente, mediante el empleo del algoritmo de optimización multicriterio genético implementado en la herramienta de modelado, se llevó a cabo un proceso de optimización teniendo en cuenta los diferentes criterios de rendimiento energético y exergéticos, estableciendo puntos de operación óptimos respecto de los diferentes criterios empleados.

Modelado y Optimización para una Gestión reciente de Recursos en Desalación Solar Modelado...

Como se comenta anteriormente, una forma de reducir el consumo de energía es acoplar las plantas de destilación con bombas de calor. Continuando con los trabajos iniciados por Carballo et al. (19; 14), se realizó un nuevo estudio de optimización del acoplamiento de la bomba de absorción de calor de doble efecto (*DEAHP*) con la planta *MED*. La optimización se realizó mediante el empleo de algoritmos genéticos de optimización multicriterio. El objetivo principal del problema de optimización era disminuir el consumo de energía de todo el sistema tratando de mantener los parámetros de operación nominales considerados para la planta *MED* de acuerdo con los criterios energéticos y exergéticos. Se obtuvieron dos puntos de funcionamiento. El primero, cuyo objetivo era reducir el esfuerzo de control sobre los niveles de contenedores muestra que se puede lograr un estado estable disminuyendo la presión de vapor y las referencias de nivel de todos los contenedores además del generador de baja temperatura. Aunque este punto de ajuste garantiza la estabilidad del sistema, disminuye el rendimiento térmico y exergético de la *DEAHP*.

El segundo punto de optimización, cuyo objetivo principal es aumentar el rendimiento térmico, muestra que al aumentar la presión del vapor se puede lograr un aumento de rendimiento térmico significativo aunque disminuye el rendimiento exergético. Además, el alcance de un estado estable no está garantizado.

Hoy en día, una gran parte de los esfuerzos de la comunidad científica se centran en mejorar la eficiencia de los sistemas, el proceso de operación y reducir los costos de las tecnologías de aprovechamiento de energía solar. Con el objetivo de encontrar ineficiencias o posibles fuentes de mejora, se realizó un análisis energético y exergético del campo de colectores solares *NEP* a partir del modelo desarrollado previamente. Dicho estudio está en revisión. El estudio muestra los principios en que se basa el modelo, la validación y un análisis energético y exergético de la planta basados en el modelo. Los resultados sobre la validación del modelo muestran que el modelo predice los valores de las variables configuradas como salida con precisión, tanto en estado estacionario como transitorio. Nótese que la propiedad de acausalidad del lenguaje de modelado hace que todas las variables se pueden configurar como entradas o salidas siempre que se establezca un sistema que se pueda resolver, convirtiendo el modelo en una herramienta flexible y personalizable, que puede emplearse para la optimización, el análisis térmico o exergético, la evaluación de rendimiento en estacionario o transitorio o para tareas de control.

Por otro lado, el análisis energético y exergético muestra que niveles elevados de radiación solar directa (*DNI*) mejoran los índices de rendimiento térmico y exergético del campo *NEP*. Por el contrario, los cambios rápidos en los valores de *DNI* causan una reducción en ambos rendimientos. Los índices de rendimiento térmico muestran valores más altos que los índices de rendimiento exergéticos, aunque muestran un comportamiento muy similar durante la simulación. El análisis térmico revela que las principales pérdidas ocurren en la reflexión y concentración de la radiación solar, mientras que el análisis exergético revela que la destrucción de la exergía en la superficie del tubo absorbente debido a

la baja temperatura de la superficie del tubo es la fuente más importante de ineficiencias exergeticas.

3 Conclusiones y trabajos futuros.

Todos los trabajos realizados llevan implícito un análisis bibliográfico sobre el sistema objeto del modelado y análisis. Por lo que el primer objetivo de la tesis se ha cubierto en los sistemas estudiados. El primero de estos sistemas ha sido la planta *MED*, sobre la cual se ha desarrollado un modelo que ha sido empleado posteriormente para realizar el análisis y la optimización de la planta. El estudio fue publicado en la revista *Desalination* (14). Como conclusiones principales del trabajo se puede extraer que el modelo es muy preciso y cumple a la perfección con los objetivos de diseño, además el análisis energético se ha revelado como una potente herramienta para detectar ineficiencias, determinar nuevos criterios de rendimiento y puntos óptimos de funcionamiento.

El segundo sistema estudiado (*DEAHP*), ha sido modelado y analizado. El modelo desarrollado ha sido acoplado al modelo de la *MED* dando como resultado un modelo del sistema compuesto que se ha empleado en el análisis y optimización del sistema conjunto. Los resultados fueron presentados en el congreso *EuroMed 2017, Desalination for Clean Water and Energy*, justificando de este modo la consecución del tercero y el cuarto de los objetivos marcados. Como conclusiones principales del trabajo se puede remarcar que aunque el modelo compuesto permitió tanto el análisis como la optimización, no resultó muy operativo debido a la gran complejidad y la inestabilidad por los altos requerimientos computacionales.

El siguiente sistema bajo estudio ha sido el campo *NEP*, del que se ha desarrollado y validado un modelo que se ha empleado posteriormente para realizar un estudio de caracterización del comportamiento térmico y exergetico del sistema tanto en estado estacionario como transitorio. El estudio se encuentra bajo revisión a la espera de ser publicado. Las conclusiones principales del trabajo son que el modelo se ajusta a los requerimientos de diseño y que el análisis térmico y exergetico de la instalación apunta a causas diferentes de ineficiencias.

Actualmente se está trabajando en el desarrollo del modelo del generador de vapor para integrarlo con el modelo del campo *NEP*. Finalizado este trabajo se comenzará con el modelo de la caldera de gas, disponiendo de este modo de todos los modelos necesarios de los subsistemas que conforman el sistema *AQUASOL*. Finalmente se integraran todos los sistemas y se procederá a la validación y la optimización del sistema completo junto con el estudio de las técnicas de control necesarias y el pertinente análisis económico.

Bibliography

- [1] Van Weert, F., Van Der Gun, J.: Saline and brackish groundwater at shallow and intermediate depths: genesis and world-wide occurrence. 39th International Association of Hydrologists, Niagara Falls: International Association of Hydrologists (2012)
- [2] Water, U.: Managing water under uncertainty and risk, the united nations world water development report 4, un water reports, world water assessment programme (2012)
- [3] Le Blanc, D., Roehri, R.: Back to our common future sustainable development in the 21st century (sd21) project summary for policymakers (2012)
- [4] Connor, R.: The United Nations world water development report 2015: water for a sustainable world. Volume 1. UNESCO Publishing (2015)
- [5] Li, C., Goswami, Y., Stefanakos, E.: Solar assisted sea water desalination: A review. *Renewable and Sustainable Energy Reviews* **19** (2013) 136–163
- [6] Palenzuela, P., Zaragoza, G., Alarcón-Padilla, D.C., Blanco, J.: Evaluation of cooling technologies of concentrated solar power plants and their combination with desalination in the mediterranean area. *Applied Thermal Engineering* **50**(2) (2013) 1514–1521
- [7] Zarza, E., Ajona, J., León, J., Gregorzewski, A., Genthner, K.: Solar thermal desalination project at the plataforma solar de almeria. *Solar energy materials* **24**(1-4) (1991) 608–622
- [8] Al-Karaghoul, A., Kazmerski, L.L.: Energy consumption and water production cost of conventional and renewable-energy-powered desalination processes. *Renewable and Sustainable Energy Reviews* **24** (2013) 343–356
- [9] Alarcon-Padilla, D.C., Blanco-Galvez, J., Garcia-Rodriguez, L., Gernjak, W., Malato-Rodriguez, S.: First experimental results of a new hybrid solar/gas multi-effect distillation system: the AQUASOL project. *Desalination* **220**(1-3) (2008) 619–625
- [10] Alarcon-Padilla, D.C., Garcia-Rodriguez, L., Blanco-Galvez, J.: Assessment of an absorption heat pump coupled to a multi-effect distillation unit within AQUASOL project. *Desalination* **212**(1-3) (2007) 303–310
- [11] Ibarra, M., Alarcón-Padilla, D., Blanco-Gálvez, J., Zaragoza, G., Palenzuela, P.: Performance of small parabolic through collector as thermal energy supply to steam generation. In: Proc. SolarPACES Conference, Marrakech, Morocco. (2012) 24997
- [12] Palenzuela, P., Hassan, A.S., Zaragoza, G., Alarcón-Padilla, D.C.: Steady state model for multi-effect distillation case study: Plataforma solar de almería med pilot plant. *Desalination* **337** (2014) 31–42
- [13] Fernández-Izquierdo, P., García-Rodríguez, L., Alarcón-Padilla, D.C., Palenzuela, P., Martín-Mateos, I.: Experimental analysis of a multi-effect distillation unit operated out of nominal conditions. *Desalination* **284** (2012) 233–237

- [14] Carballo, J.A., Bonilla, J., Roca, L., De la Calle, A., Palenzuela, P., Alarcón-Padilla, D.C.: Optimal operating conditions analysis for a multi-effect distillation plant according to energetic and exergetic criteria. *Desalination* (2017)
- [15] Fritzson, P.: Principles of object-oriented modeling and simulation with Modelica 3.3: a cyber-physical approach. John Wiley & Sons (2014)
- [16] de la Calle, A., Bonilla, J., Roca, L., Palenzuela, P.: Dynamic modeling and simulation of a solar-assisted multi-effect distillation plant. *Desalination* **357** (2015) 65–76
- [17] Gomri, R.: Energy and exergy analyses of seawater desalination system integrated in a solar heat transformer. *Desalination* **249**(1) (2009) 188–196
- [18] Pfeiffer, A.: Optimization Library for Interactive Multi-Criteria Optimization Tasks. In: Proc. 9th International Modelica Conference, Munich, Germany (nov 2012) 669–680
- [19] Carballo, J.A., Bonilla, J., Roca, L., de la Calle, A., Palenzuela, P., Berenguel, M.: Optimal operating conditions analysis of a multi-effect distillation plant. *Desalination and Water Treatment* **69** (2017) 229–235
- [20] de la Calle, A., Roca, L., Bonilla, J., Palenzuela, P.: Dynamic modeling and simulation of a double-effect absorption heat pump. *International Journal of Refrigeration* **72** (2016) 171–191
- [21] Russell, F.: 27–Heat Transfer Analysis and Modeling of a Parabolic Trough Solar Receiver Implemented in Engineering Equation Solver. *National Renewable Energy Laboratory* (October) (2003) 164
- [22] Quoilin, S., Desideri, A., Wronski, J., Bell, I., Lemort, V.: ThermoCycle: A Modelica library for the simulation of thermodynamic systems. (2014) 683–692
- [23] Desideri, A., Hernandez, A., Gusev, S., van den Broek, M., Lemort, V., Quoilin, S.: Steady-state and dynamic validation of a small-scale waste heat recovery system using the thermocycle modelica library. *Energy* **115** (2016) 684–696
- [24] Bonilla, J., Carballo, J.A., Roca, L., Berenguel, M.: Development of an open source multi-platform software tool for parameter estimation studies in fmi models. In: Proceedings of the 12th International Modelica Conference, Prague, Czech Republic, May 15–17, 2017. Number 132, Linköping University Electronic Press (2017) 683–692

Técnicas Híbridas de Machine Learning y Transformación de Modelos para la Evolución de Interfaces Mashup

Antonio Jesus Fernández-García

Applied Computing Group. University of Almeria, Spain.
ajfernandez@ual.es

Abstract. El uso de interfaces basadas en componentes (o mashup) es muy frecuente en la actualidad dada la versatilidad que ofrecen este tipo de interfaces a la hora de permitir que los usuarios la personalicen de forma acorde a sus necesidades. El éxito de este tipo de aplicaciones en el mercado radica, en gran medida, en la capacidad de los usuarios para encontrar los componentes que necesitan usar dentro del catálogo que ofrecen las aplicaciones mashup, que en muchos casos es muy amplia, lo que puede dificultar algunas tareas de búsqueda y reconfiguración. En esta tesis, se propone una metodología para permitir que las interfaces mashup sean inteligentes y adapten su estructura dinámicamente a las necesidades de los usuarios. La metodología propuesta en esta tesis, cubre todos los aspectos desde la adquisición de los datos de las interfaces mashup hasta el uso de técnicas computacionales *Machine Learning* sobre grandes cantidades de datos heterogéneos recogidos de los usuarios para generar nuevas formas de adaptar una interfaz mashup a sus usuarios. Se pretende que esta metodología sea capaz de evolucionar en el tiempo mediante el despliegue de un sistema autónomo que actualice las formas de adaptación de la aplicación mashup.

1 Introducción

Los usuarios, a día de hoy, requieren acceso a información de forma instantánea. Muchas veces esta información no consiste solamente en datos en bruto, sino que la mayoría de las veces, y cada vez más, se trata de información elaborada que debe ser procesada para poder ser ofrecida. Por estos motivos, las aplicaciones tipo mashup están creciendo en popularidad. En general, las mashups administran una gran cantidad de datos heterogéneos de diferentes fuentes y manejan diferentes tipos de usuarios. Pueden ser diseñadas para ser accesibles mediante todo tipo de dispositivos (*Smartphones, Tablets, Wearables...*), admitiendo múltiples formas de interacción (táctil, gestual, voz...) y su despliegue en entornos *cloud* puede asegurar su alta disponibilidad y accesibilidad. Sin embargo, la forma en la que los usuarios acceden a la información también puede ir evolucionando debido a diferentes factores a lo largo del tiempo.

Esta posible evolución, motiva la creación de nuevas interfaces mashup capaces de dar respuesta al uso requerido por los usuarios en el momento que accedan a ellas. Tradicionalmente, los sistemas informáticos se definen en tiempo de diseño acorde a una arquitectura y a unos requisitos iniciales. En sistemas informáticos dinámicos y evolutivos, en los que se pretende que la adaptación de la interfaz se ajuste a las necesidades del usuario analizadas anteriormente, puede ser de utilidad que la arquitectura evolucione en “tiempo de ejecución” para adaptarse a las nuevas necesidades del entorno.

Para desarrollar un marco de adaptación dinámica de interfaces de usuario tipo mashup es necesario aplicar un amplio rango de técnicas como pueden ser Ingeniería Dirigida por Modelos, Computación en la Nube e Inteligencia Computacional. El trabajo realizado en [4, 5], ya soluciona la adaptación dinámica en tiempo de ejecución de las interfaces de usuario haciendo uso de las técnicas anteriormente comentadas. No obstante, hay un aspecto que no está contemplado en el trabajo, la evolución en el tiempo de dichas interfaces.

El objetivo de esta tesis es proponer un método que permita a las interfaces mashup basadas en componentes ser inteligentes y evolucionar con el tiempo. Esto se logra mediante el estudio, análisis y la aplicación de diferentes algoritmos *Machine Learning*, técnicas *Big Data* para procesar grandes volúmenes de datos heterogéneos y una metodología impulsada por la transformación de modelos que sea capaz de generar las nuevas formas de adaptación de la interfaz para hacerla autónoma. Merece la pena destacar que esta metodología para construir interfaces de usuario evolutivas no solo es aplicable a las interfaces gráficas de usuario, sino que se puede extender a todos los tipos de interfaces hombre-máquina tipo mashup.

La Figura 1 muestra el proceso general de la metodología propuesta, mostrando un escenario marco donde, mediante la adquisición de la interacción del usuario en la interfaz y la información del contexto de dicha interacción, se pueden crear interfaces inteligentes y dinámicas que modifican su estructura en tiempo de ejecución.

La metodología puede descomponerse en cuatro etapas bien diferenciadas:

- **Almacenamiento del comportamiento del usuario.** Esta etapa registra toda la actividad que genera el usuario en su interacción con la interfaz mashup, así como toda la información del entorno durante el momento en el que se produce dicha interacción y la almacena en una base de datos relacional. La Figura 2 muestra las partes del sistema de adquisición de datos, que son:
 - **#1** Creación de un servicio web llamado *getInteraction* que recibe todas las interacciones que se producen en la interfaz mashup, la cual es responsable de llamar a este servicio web cada vez que se produce una interacción, enviando todos los datos que se necesitan almacenar.
 - **#2** En el caso de que se necesiten más datos sobre un usuario, un servicio o cualquier otro aspecto de la interfaz mashup, el siguiente paso es conseguir esa información, que puede ser consultada directamente a la mashup (si dispone de algún sistema de acceso).

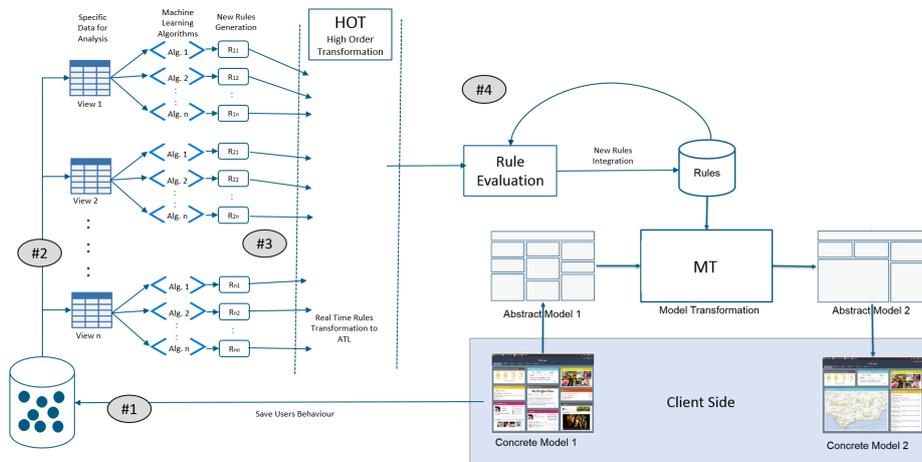


Fig. 1. Metodología para la creación de interfaces mashup dinámicas y evolutivas

- **#3** Estudio del entorno de la interfaz mashup e identificar qué datos valdría la pena obtener del contexto. Si hay datos interesantes que obtener, y se identifican los servicios que proporcionan dicha información de contexto, debe implementarse la conexión con estos servicios web externos para poder recabar la información.
 - **#4** Cuando se recopilan todos los datos necesarios, son almacenados en la base de datos. Se requiere serializar las operaciones para evitar conflictos de integración al agregar relaciones a los datos.
- **Creación de *datasets*.** El objetivo que se persigue al crear *datasets* es organizar los datos, es decir, estructurarlos para que puedan servir como elementos de entrada para los algoritmos utilizados en el aprendizaje automático. El hecho de que haya múltiples *datasets* es debido a que, con cada uno de ellos, se pretende estudiar una adaptación específica de la mashup (cambiar su estructura, sugerir un componente...). Esta parte comprende un amplio conjunto de procesos de minería de datos y *feature engineering* como pueden ser eliminar atributos, filtrar instancias, cambiar tipos de datos, discretizar valores en atributos, crear nuevos atributos a partir de existentes, dividir un atributo en varios, reducir el número de clases, etc.
- **Aplicación de técnicas y algoritmos *machine learning*.** Varios tipos de algoritmos de aprendizaje automático presentes en la literatura son analizados y aplicados. Esta es la sección central de la tesis doctoral y en la que se ha producido el mayor avance este año. La siguiente sección (*Avances*) presenta el trabajo realizado en este ámbito.
- **Aplicación del conocimiento generado en las aplicaciones mashup.** El último paso, tras generar el conocimiento que permita una adaptación dinámica en tiempo real de la interfaz mashup a las necesidades del usuario de forma evolutiva, trata de aplicar dicho conocimiento en la propia interfaz

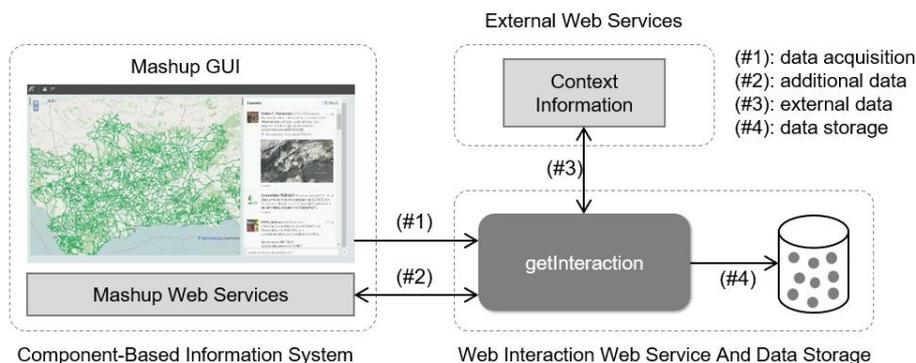


Fig. 2. Sistema de Adquisición de Datos

para sacar partido a dicho conocimiento generado. En los trabajos mencionados [4, 5], ya se realiza una adaptación dinámica de las aplicaciones mashup a través de una serie de reglas almacenadas en un repositorio de forma estática. En estos casos, las reglas inferidas generadas por los modelos creados con algoritmos de aprendizaje automático tienen su propio formato que se adapta a la identidad de los datos procesados y al tipo de algoritmo utilizado. Usando transformaciones de modelos [8] las salidas de los algoritmos se transforman al formato adecuado para la interfaz donde se aplican. Debe existir un módulo de evaluación de reglas que defina si una nueva regla generada debe agregarse al repositorio de reglas o no. Es posible que nuevas reglas generadas contradigan algunas de las reglas ya existentes en el repositorio. En estos casos, las reglas generadas pueden ser incorporadas a la interfaz mashup utilizando otros medios como sistemas de recomendación no invasivos o sistemas de adaptación automática invasiva a través de la transformación de la interfaz en tiempo real utilizando el conocimiento generado por modelos de predicción. La forma de aplicar el conocimiento generado por los modelos es muy heterogénea, quedando dicho conocimiento a disposición de los desarrolladores de la aplicación para su explotación en las circunstancias más adecuadas acorde a sus necesidades y a los requerimientos del mercado de la aplicación software.

2 Avances

Los avances de este último año se han centrado en crear los modelos tras aplicar los algoritmos de aprendizaje automático y en analizar sus resultados. Tal y como comentábamos en la sección anterior, tras el almacenamiento de todos los datos de interacción de una aplicación mashup se crean varios *dataset*, cada uno ellos con un objetivo, y se optimizan como paso previo a la aplicación de los algoritmos de aprendizaje automático.

Una vez optimizado un *dataset* tras la aplicación de técnicas de *feature engineering* nos debemos asegurar que su conjunto de atributos debe ser lo más significativo posible para describir de la mejor forma la realidad que representan. Puede parecer que un gran número de atributos puede describir mejor un problema y por tanto, se pueden construir mejores modelos predictivos, pero esto no es del todo cierto. Los métodos de selección de atributos pueden ayudar a reducir el *overfitting* [3], evitar problemas derivados de espacios con muchas dimensiones, o provocados por la escasa cantidad de datos [7] que pueden tener los *dataset* con un gran conjunto de atributos. Además, un *dataset* con un subconjunto de atributos implica tiempos de entrenamiento más cortos al construir los modelos, pues eliminan atributos redundantes o irrelevantes.

En nuestro trabajo, hacemos uso de métodos de selección de atributos que admiten todo tipo de atributos con independencia de sus tipos de datos. En concreto, los métodos utilizados son: *Mutual Information Score* y *Chi-Squared Statistics*.

El método *Mutual Information Score* se define como:

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p(x) p(y)} \right); \quad (1)$$

donde $p(x, y)$ es la función de probabilidad conjunta de X e Y y $p(x)$ y $p(y)$ son las funciones de distribución de probabilidad marginal de X e Y respectivamente [6].

El método *Chi-Squared Statistic* se define como:

$$X^2 = \sum_{i=1}^n \sum_{j=1}^m \frac{(O_{(i,j)} - E_{(i,j)})^2}{E_{(i,j)}}; \quad (2)$$

donde $O_{(i,j)}$ es el valor observado de dos variables y $E_{(i,j)}$ es el valor esperado de dos variables. El valor esperado puede ser calculado mediante la siguiente fórmula:

$$E_{(i,j)} = \frac{\sum_{i=1}^c O_{(i,j)} \sum_{k=1}^c O_{(k,j)}}{N}; \quad (3)$$

donde $\sum_{i=1}^c O_{(i,j)}$ es la suma de la i_{th} columna y $\sum_{k=1}^c O_{(k,j)}$ es la suma de la k_{th} columna [9].

Una vez definido un nuevo *dataset* con los atributos del *dataset* original que mayor puntuación han obtenido en los métodos de selección de características comentados, el siguiente paso es crear los modelos de recomendación.

En nuestra caso, estamos trabajando haciendo uso de una selección de cuatro algoritmos de clasificación disponibles en Microsoft Azure Machine Learning

Studio [11]. Hacemos uso de la implementación de los algoritmos de AzureML debido a la facilidad que ofrece esta plataforma para configurar servicios web a través de modelos entrenados para desplegarlos. De la misma manera, también valoramos positivamente que los lectores pueden acceder fácilmente a estos algoritmos como invitados anónimos de forma gratuita en la plataforma y realizar cómodamente los experimentos¹. Los algoritmos que usamos son bien conocidos en la literatura: *Decision Forest*, *Decision Jungle*, *Redes Neuronales* y *Regresión Logística*.

Los algoritmos de árboles de decisión crean una estructura tipo árbol donde cada nodo representa una pregunta sobre un atributo. Las respuestas a esa pregunta crean nuevas ramas para expandir la estructura hasta que se llega al final del árbol, siendo el nodo de hoja el que indica la clase predicha [12]. El algoritmo *Decision Forest*, crea varios árboles de decisión y selecciona la salida más popular de ellos.

El algoritmo *Decision Jungle* es una extensión del algoritmo *Decision Forest* donde cada árbol es reemplazado por un grafo acíclico dirigido. Es más eficiente en cuanto al uso de memoria porque elimina la necesidad de repetir nodos hoja y permite que las ramas se fusionen. Como desventaja necesita más tiempo de computación.

El algoritmo de *Redes Neuronales* crea un conjunto de niveles interconectados, donde cada nivel consiste en un conjunto de nodos (neuronas) que recibe entradas y produce salidas ponderadas. Los nodos de la primera capa son las entradas, los nodos de la última capa son la salida y a los nodos intermedios se les conoce como nodos ocultos. Una red neuronal puede verse como un gráfico acíclico dirigido ponderado [13].

El algoritmo de *Regresión Logística* predice la probabilidad de salida de una clase ajustando las instancias a una función logística. Tiene como objetivo construir una función que describa las relaciones entre los atributos de entrada y la etiqueta de clase [10].

Una vez aplicados los algoritmos de aprendizaje automático sobre los *dataset* después de seleccionar sus atributos más relevantes es necesario evaluar los modelos generados por dichos algoritmos. Los resultados de los modelos pueden ser analizados a través de matrices de confusión. Una matriz de confusión es un diseño de tabla donde cada fila representa el número de instancias de cada clase y cada columna representa la clase que ha sido predicha por el modelo. En esta tabla, podemos conocer el rendimiento de un modelo más allá de la precisión global que ofrece.

Para validar la efectividad del sistema de recomendación y los métodos de análisis de datos aplicados, realizamos experimentos sobre una aplicación web tipo-mashup real, de la cual disponemos de datos de interacción de sus usuarios. El nombre de la aplicación es ENIA (Environmental Information Agent) [1], una interfaz de usuario basada en componentes para la gestión ambiental utilizada por la Red Andaluza de Información Medioambiental (REDIAM) [2].

¹ <https://studio.azureml.net/>

3 Conclusiones

Nuestro trabajo presenta una metodología práctica, que obtiene un *dataset* sin procesar de una aplicación tipo mashup basada en componentes. Una vez obtenido el *dataset*, aplica técnicas de *feature engineering* como eliminar características, filtrar instancias, transformar tipos de datos, o fusionar y dividir atributos, entre otros, para optimizar dicho *dataset*. Este proceso es fundamental para tener éxito a la hora de aplicar algoritmos de aprendizaje automático, ya que la representación de los atributos tiene un gran impacto en el rendimiento de modelos predictivos. Tras las técnicas de *feature engineering*, los métodos *Mutual Information Score* y *Chi-Squared Statistics* de selección de atributos se aplican para obtener un subconjunto del *dataset* original. Estos métodos seleccionan automáticamente los atributos más relevantes del *dataset* y, al mismo tiempo, junto con las técnicas de *feature engineering*, eliminan problemas como el overfitting, la alta dimensionalidad de los atributos o derivados por la escasez de los datos.

Una vez optimizado el *dataset*, se crean varios sistemas de recomendación utilizando conocidos algoritmos de clasificación como *Decision Forest*, *Decision Jungle*, *Regresión Logística* y *Redes Neuronales*. Los modelos creados se evalúan y comparan en términos de precisión general y precisión promedio relativa a la predicción de cada clase, así como analizando la matriz de confusión que muestra la precisión específica de la desviación que predice una clase en comparación con la clase etiquetada.

Como conclusión, en las pruebas realizadas se obtienen de forma general buenos resultados, alcanzando altos porcentajes de precisión en muchos casos. Esto, sin duda, contribuye al objetivo planteado en esta tesis: la evolución de las interfaces mashup aplicando técnicas híbridas de *machine learning*.

References

1. ENIA. Environmental Information Agent Project. <http://acg.ual.es/projects/enia/ui/>. Online; last accessed 18 December 2017.
2. The Andalusian Environmental Information Network (REDIAM). <http://www.juntadeandalucia.es/medioambiente/site/rediam/>. Online; last accessed 18 December 2017.
3. Salem Alelyani, Jiliang Tang, and Huan Liu. *Feature Selection for Clustering: A Review*. Chapman & Hall/CRC, 2013.
4. Javier Criado, Diego Rodríguez-Gracia, Luis Iribarne, and Nicolás Padilla. Toward the adaptation of component-based architectures by model transformation: behind smart user interfaces. *Software: Practice and Experience*, 45(12):1677–1718, 2015.
5. Javier Criado, Cristina Vicente-Chicote, Nicolás Padilla, and Luis Iribarne. A model-driven approach to graphical user interface runtime adaptation. In *5th International Workshop on Models*, pages 49–59, 2010.
6. Girish Chandrashekar and Ferat Sahin. A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1):16 – 28, 2014. 40th-year commemorative issue.

7. Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The elements of statistical learning: Data mining, inference, and prediction. *International Statistical Review*, 77(3):482–482, 2009.
8. Ivan Kurtev, Klaas van den Berg, and Frédéric Jouault. Rule-based modularization in model transformation languages illustrated with atl. *Science of Computer Programming*, 68(3):138 – 154, 2007. Special Issue on Model Transformation.
9. Nathan Mantel. Chi-square tests with one degree of freedom; extensions of the mantel-haenszel procedure. *Journal of the American Statistical Association*, 58(303):690–700, 1963.
10. P. McCullagh and J.A. Nelder. *Generalized Linear Models, Second Edition*. Chapman and Hall/CRC Monographs on Statistics and Applied Probability Series. Chapman & Hall, 1989.
11. Microsoft Corporation. Microsoft Azure Machine Learning Studios.
12. John Ross Quinlan. Induction of decision trees. *Mach. Learn.*, 1(1):81–106, March 1986.
13. Raúl Rojas. *Neural Networks: A Systematic Introduction*. Springer-Verlag New York, Inc., New York, NY, USA, 1996.

Procesamiento de Consultas Espaciales en Entornos Escalables y Distribuidos

Francisco José García García

E-mail: paco.garcia@ual.es

Universidad de Almería

Abstract. Debido al omnipresente uso de aplicaciones de datos espaciales y las grandes cantidades de datos que éstas generan, el procesamiento de consultas a gran escala en sistemas distribuidos es cada vez más popular. Por eso, se han realizado algoritmos paralelos y distribuidos para dos de las consultas más estudiadas, K Closest Pair Query ($KCPQ$) y ε Distance Join Query (εDJQ). También se han presentado mejoras que utilizan técnicas de muestreo y aproximación para mejorar su rendimiento. Además presentamos el trabajo realizado sobre una nueva consulta denominada Reverse k -Nearest Neighbor ($RkNN$). Por último, hemos comparado dos de los sistemas de gestión de datos espaciales distribuidos más recientes y punteros, a saber, SpatialHadoop y LocationSpark, mediante la evaluación del rendimiento de los anteriores algoritmos en diferentes situaciones con grandes conjuntos de datos del mundo real.

Keywords: Big Data, MapReduce, Spark, Spatial Data Processing, Spatial Query Evaluation, SpatialHadoop, LocationSpark, Distance Joins, RNNQ

1 Introducción

El procesamiento de consultas espaciales ha recibido una considerable atención por parte de la comunidad de base de datos. Muchos trabajos de investigación se han centrado en la mejora del rendimiento de estas consultas proponiendo algoritmos eficientes o diseñando nuevas estructuras de índices complejos. Sin embargo, todas estas aproximaciones se centran en métodos de acceso que son ejecutados en un entorno centralizado. Con el rápido crecimiento en el tamaño de grandes conjuntos de datos, el procesamiento de estos de forma paralela y distribuida se está convirtiendo en una práctica habitual. Un conjunto de algoritmos paralelos espaciales como k Nearest Neighbor (kNN) query [25], kNN Join ($kNNJ$) [14] y similarity join [16] en MapReduce [2] han sido desarrollados e implementados recientemente.

MapReduce [2] es un framework para el procesamiento y manejo de conjuntos de datos a gran escala en un cluster distribuido, que ha sido utilizado para diferentes aplicaciones, como la generación de índices de búsqueda, clustering de documentos, análisis de logs de acceso, y otras formas de análisis de datos. MapReduce fue introducido con el objetivo de proporcionar un paradigma de

computación paralelo y distribuido que fuese sencillo a la vez que potente, proporcionando buena escalabilidad y mecanismos de tolerancia de fallos.

Sin embargo, como se indica en [3], MapReduce tiene debilidades relacionadas con la eficiencia cuando tiene que ser aplicado a datos espaciales. Una deficiencia de este framework es la falta de mecanismos de indexación que pudieran permitir el acceso selectivo a regiones específicas de los datos espaciales, lo cual proporcionaría algoritmos de procesamiento de consultas más eficientes. Una solución reciente a este problema es la extensión de Hadoop, llamada SpatialHadoop [4], que es un framework que soporta de forma inherente la indexación espacial en un entorno de Hadoop.

En SpatialHadoop, los datos espaciales se particionan y distribuyen a los nodos, de tal manera que los datos con proximidad espacial se encuentran en la misma partición. Además, las particiones generadas están indexadas, facilitando el diseño de algoritmos de consulta eficientes que acceden solo a una parte de los datos y que siguen devolviendo el resultado correcto. Como se demuestra en [4], varios algoritmos son propuestos para consultas espaciales, como rango, kNN, spatial joins y skyline queries. El procesamiento eficiente de las consultas más representativas y estudiadas para grandes volúmenes de datos espaciales es un desafío, y es el principal objetivo de esta tesis.

Lo que se pretende en esta tesis doctoral es, a partir del estudio y uso de diferentes entornos paralelos y distribuidos, el desarrollo de nuevos algoritmos eficientes que aprovechen las características que proporcionan. También se presentarán los resultados de la ejecución de un extenso conjunto de experimentos que demuestren la eficiencia y escalabilidad de nuestras propuestas con el uso tanto de datos sintéticos como de datos reales.

La tesis se enmarca dentro del proyecto TIN2013-41579-R del Ministerio de Economía y Competitividad (MINECO) del Gobierno de España.

El resto de este artículo continua con la presentación de los avances realizados durante el último año en el desarrollo de la tesis. Finalmente, se muestran las conclusiones obtenidas hasta el momento y se muestra una descripción de posibles trabajos futuros a realizar.

2 Avances

Durante el último año, la investigación se ha centrado principalmente en tres objetivos: la mejora de algoritmos presentados en trabajos anteriores, el desarrollo de nuevas consultas espaciales y la comparativa con otros sistemas disintos de SpatialHadoop.

2.1 Efficient Large-scale Distance-Based Join Queries in SpatialHadoop

La investigación recogida en [10] extiende substancialmente nuestros anteriores trabajos [7, 9], siendo estos los cimientos sobre los que se sustenta. Las Distance-Based Join Queries (*DBJQ*) han recibido considerable atención de la comunidad

de bases de datos, debido a su importancia en numerosas aplicaciones. Las *DBJQs* más representativas y conocidas son la *K Closest Pairs Query (KCPQ)* y la ε Distance Join Query (εDJQ).

Para la consulta *KCPQ*, dados dos conjuntos de puntos (P y Q) y un valor positivo K como el número de pares del resultado final, esta consulta espacial encuentra los K pares de puntos más cercanos de $P \times Q$ de acuerdo a una determinada métrica de distancia (ej., Manhattan, Euclidea, Chebyshev, etc.). La implementación de *KCPQ* en MapReduce [9] consiste en adoptar la metodología top- K de MapReduce. La idea básica es tener P y Q particionados mediante algún método (ej., Grid) en n y m celdas de puntos y generar $n \times m$ posibles pares de celdas a combinar. Entonces, cada par de celdas (una de P y otra de Q) es enviado como entrada de la fase *Map*. Cada *mapper* lee su par de celdas y realiza un algoritmo de barrido del plano (*Clásico* o *Reverse Run*) *KCPQ (PSKCPQ)* entre los P y Q locales a cada par. Los K resultados de todos los *mappers* son enviados a un único *reducer* que obtendrá el top- K global a todos los *mappers*. Finalmente, los resultados se escriben en ficheros del Hadoop Distributed File System (HDFS), almacenando solo las coordenadas de los puntos y la distancia entre ellos.

Se puede observar claramente que el rendimiento de la solución propuesta con este algoritmo *KCPQ MapReduce* dependerá del número de celdas en las que los 2 conjuntos de puntos estén particionados. Por otro lado, sabemos que los algoritmos *KCPQ* basados en barrido del plano utilizan una distancia de poda, que es el valor de la distancia al K -ésimo par más cercano que se haya encontrado por el momento, para descartar aquellas combinaciones de pares de puntos que no se necesitan considerar como candidatos del resultado final. Como se sugiere en [9], necesitamos encontrar previamente un límite superior del valor de la distancia del K -ésimo par más cercano de los conjuntos de datos que se unen, denominado β . Este cálculo de β se puede obtener muestreando globalmente ambos conjuntos de datos y ejecutando un algoritmo *PSKCPQ* sobre ambos muestreos.

Utilizando el método anterior se observa claramente que el mayor sobrecoste ocurre en la ejecución del muestreo global, ya que es un trabajo MapReduce completo. Con el objetivo de reducir este tiempo, se ha propuesto el uso de los índices globales de SpatialHadoop, que proporcionan el tamaño de cada celda del índice, así como el número de elementos que contienen, de manera que se pueda obtener una idea de como los datos están distribuidos dentro de cada celda. Para simplificar el proceso de muestreo, encontraremos un par de celdas, que por sus características, puedan contener los K pares más cercanos con el valor de β más pequeño posible. Entonces, podemos muestrear localmente en esas celdas sin tener que ejecutar un trabajo MapReduce.

Nuestro primer método, llamado β local processing, se basa en realizar un muestreo local sobre el par de celdas con mayor densidad y mayor área de intersección. Por otro lado, y de acuerdo a [1], podemos aplicar técnicas de aproximación basadas en α -allowance en algoritmos *KCPQ* basados en barrido del plano (*PSKCPQ*) [15] y en sus tres variantes (Strip, Window y Semi-Circle) para obtener un resultado aproximado de forma rápida sobre el total de elementos

de las celdas anteriores. Éste método puede ser llevado a cabo multiplicando δ , que es el valor de poda actual, por $(1 - \alpha)$, dando lugar a $\alpha PSKCPQ$. Debido a que es una técnica de aproximación basada en distancias, cuanto mayor sea α , menor será el límite superior del valor de δ (es decir, menos puntos serán considerados y más puntos serán descartados).

Para el caso de la ε Distance Join Query (εDJQ), dados dos conjuntos de puntos (P y Q) y un límite de distancia ε , esta consulta encuentra todas las posibles parejas de puntos de $P \times Q$, que se encuentran al menos a una distancia ε entre ellos.

El algoritmo propuesto para εDJQ en MapReduce adopta la metodología de combinación en la fase *Map* de MapReduce. Ésta consiste en particionar P y Q mediante algún método (ej., Grid) en n y m celdas de puntos, respectivamente. Entonces, cada posible par de celdas (una de P y una de Q) es enviado como entrada de la fase *Filter*. Esta fase, descarta los pares de celdas que tengan una distancia mínima superior a ε . En la fase *Map* cada *mapper* lee su par de celdas y realiza un algoritmo de barrido del plano (*Clásico* o *Reverse Run*) εDJQ ($PS\varepsilon DJQ$) entre los P y Q locales en ese par (es una variación del algoritmo basado en barrido del plano para $KCPQ$ [15]). Los resultados de todos los *mappers* son combinados en la fase *Reduce* y son escritos en ficheros HDFS, almacenando solo los pares de puntos con distancia menor a ε .

Además, se ha realizado otra propuesta que hace uso de los índices locales proporcionados por SpatialHadoop para obtener mejoras de rendimiento sobre el algoritmo εDJQ anterior en MapReduce. Este algoritmo consiste en realizar un recorrido en profundidad sobre árboles-R. Esto es, para cada par de nodos internos, uno de cada índice, la mínima distancia entre ellos es calculada y si es superior a ε , el par es descartado. De otra manera, los hijos de los nodos se chequean en el siguiente paso siguiendo un orden de profundidad. Cuando se alcanzan los nodos hoja, se aplica el mismo algoritmo de barrido del plano que se utiliza cuando no hay índices.

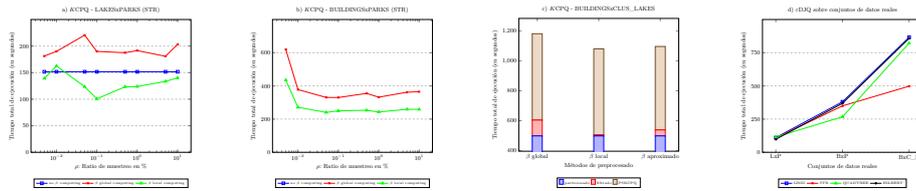


Fig. 1. Coste de $KCPQ$ (tiempo de ejecución) sin y con computación β (gráficas a,b,c). Coste de εDJQ (tiempo de ejecución) vs diferentes conjuntos de datos reales (gráfica d).

Una comparativa detallada de rendimiento de los algoritmos y mejoras propuestas en diferentes escenarios con grandes conjuntos de datos sintéticos y reales ha sido realizada (Figura 1). Para $KCPQ$, se ha demostrado experimentalmente que la eficiencia de las mejoras sobre los algoritmos presentados en [9], con

el cálculo de un límite superior (β) del valor de la distancia del K -ésimo par más cercano, utilizando una fase de preprocesado local basada en técnicas de muestreo y de aproximación. Y para εDJQ , se ha implementado el algoritmo basado en barrido del plano *Reverse Run* [15] siguiendo un esquema similar a al del algoritmo *KCPQ*, y que ha demostrado obtener resultados finales en tiempos de respuesta adecuados.

2.2 A Comparison of Distributed Spatial Data Management Systems for Processing Distance Join Queries

El trabajo realizado en [6] busca comparar los algoritmos y mejoras realizados en [10], que utilizan SpatialHadoop, con otros sistemas de gestión de datos espaciales distribuidos (DSDMS). Estos pueden clasificarse en sistemas basados en disco [13] y basados en memoria [24]. Los DSDMS basados en disco se caracterizan por ser sistemas basados en Hadoop y uno de los más representativos es el propio SpatialHadoop [4]. Por otro lado, los DSDMS en memoria se caracterizan por ser sistemas basados en Apache Spark², siendo LocationSpark [19] uno de los más representativos.

Apache Spark es un marco de procesamiento de datos a gran escala, rápido, confiable y distribuido en memoria. Aprovecha la estructura de datos Resilient Distributed Dataset (RDD), que permite almacenar datos de forma transparente en la memoria y pasarlos a disco solo si es necesario [23]. Por lo tanto, puede reducir una gran cantidad de operaciones de escritura y lectura de disco para superar en rendimiento al entorno Hadoop. Como Spark mantiene el estado de los recursos asignados hasta que se completa un trabajo, reduce el consumo de tiempo en la preparación y recolección de recursos.

LocationSpark [19] es una biblioteca de Spark que proporciona una API para el procesamiento y optimización de consultas espaciales basada en los operadores de flujo de datos estándar de Spark. Es un sistema eficiente de procesamiento de consultas espaciales distribuidas en memoria. LocationSpark proporciona varias optimizaciones para mejorar Spark en la gestión de datos espaciales y están organizadas por capas: gestión de memoria, índices espaciales, ejecutor de consultas, planificador de consultas, operadores espaciales y analítica espacial. Dado que nuestro objetivo principal es incluir las DJQ (*KCPQ* y ε DJQ) en LocationSpark, nos interesa la capa de *operadores espaciales*, que es donde se han implementado. Para esta capa, LocationSpark admite consultas espaciales y actualizaciones de datos espaciales. Proporciona un amplio conjunto de consultas espaciales, incluida la consulta de rango espacial, *KNNQ*, combinación espacial y *KNNJQ*.

En la literatura, hasta ahora, hay pocos estudios comparativos entre Hadoop y sistemas basados en Spark. Motivados por este hecho, el trabajo realizado se ha basado en comparar SpatialHadoop y LocationSpark para el procesamiento

² disponible en <https://spark.apache.org/>

de consultas de combinación basado en distancia, en particular para $KCPQ$ y εDJQ , y proporcionar criterios para adoptar uno u otro DSDMS.

Suponiendo que \mathbb{P} es el conjunto de datos más grande que se combinará y \mathbb{Q} es el más pequeño, y siguiendo las ideas presentadas en [19], podemos describir el *Plan de ejecución* para $KCPQ$ en LocationSpark de la siguiente manera. En la Etapa 1, los dos conjuntos de datos se dividen en particiones según un índice espacial dado. En la Etapa 2, los datos estadísticos se agregan a cada partición, $S_{\mathbb{P}}$ y $S_{\mathbb{Q}}$, y se combinan por pares, $S_{\mathbb{P}\mathbb{Q}}$. En la Etapa 3, las particiones de \mathbb{P} y \mathbb{Q} con la mayor densidad de puntos, \mathbb{P}_{β} y \mathbb{Q}_{β} , se seleccionan para combinarse usando un algoritmo $KCPQ$ basado en barrido del plano [15] para calcular un límite superior del valor de distancia del par K -ésimo más cercano (β). En la Etapa 4, usando *Query Plan Scheduler*, \mathbb{P} se divide en $\mathbb{P}_{\mathbb{S}}$ y $\mathbb{P}_{\mathbb{NS}}$ son las particiones que presentan y no presentan sesgo de datos, respectivamente. El mismo particionado se usa para \mathbb{Q} . En la Etapa 5, se aplica un algoritmo $KCPQ$ entre puntos de $\mathbb{P}_{\mathbb{S}}$ y $\mathbb{Q}_{\mathbb{S}}$ que están en la misma partición y también para $\mathbb{P}_{\mathbb{NS}}$ y $\mathbb{Q}_{\mathbb{NS}}$ en la Etapa 6. Estas dos etapas se ejecutan de forma independiente y los resultados se combinan en la Etapa 7. Finalmente, aún es necesario calcular si hay algún candidato presente para cada partición que se encuentre en los límites de esa misma partición en el otro conjunto de datos. Para hacer esto, usamos β' que es la distancia máxima del conjunto actual de candidatos como el radio de una consulta de rango con centro en cada partición para obtener posibles nuevos candidatos en esos límites. El cálculo de $KCPQ$ de cada partición con sus candidatos se ejecuta en las Etapas 8 y 9, y estos resultados se combinan en la Etapa 10 para obtener la respuesta final.

El *Plan de ejecución* para εDJQ en LocationSpark es una variación del desarrollado para $KCPQ$, donde se eliminan las etapas de filtrado, ya que $S_{\mathbb{P}\mathbb{Q}}$ está filtrado por ε (es decir, $\beta = \beta' = \varepsilon$), que es la distancia umbral conocida de antemano.

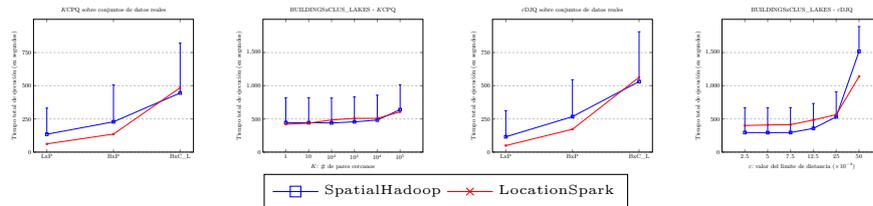


Fig. 2. Tiempos de ejecución de $KCPQ$ and εDJQ considerando diferentes conjuntos (izquierda) y considerando diferentes valores de k y ε (derecha).

La ejecución de un conjunto de experimentos (Figura 2) ha demostrado que LocationSpark es el ganador general teniendo en cuenta el tiempo de ejecución, debido a la eficiencia del procesamiento en memoria proporcionado por Spark y las mejoras adicionales como el *Query Plan Scheduler*. Sin embargo, SpatialHadoop es un DSDMS más maduro y robusto debido al tiempo dedicado a investigarlo y

desarrollarlo (varios años) y proporciona más operaciones espaciales y técnicas de particionado espacial.

2.3 RkNN Query Processing in Distributed Spatial Infrastructures: A Performance Study

En [8] presentamos el trabajo realizado sobre una nueva consulta denominada Reverse k -Nearest Neighbor ($RkNN$) query [12, 18], la cual ha sido implementada tanto SpatialHadoop como en LocationSpark. Una consulta de ($RkNN$) devuelve los objetos que tienen el objeto de consulta en el conjunto de sus k vecinos más cercanos. Es el problema complementario al de encontrar los k -Nearest Neighbors (kNN) de un objeto de consulta. El objetivo de una consulta $RkNN$ ($RkNNQ$) es identificar la *influencia* de un objeto de consulta sobre todo el conjunto de datos, y varios ejemplos reales se mencionan en [12]. Aunque el problema de $RkNN$ es el complemento del problema de k -Nearest Neighbour, la relación entre kNN y $RkNN$ no es simétrica y el número de vecinos inversos de k de un objeto de consulta no se conoce de antemano. Una solución rápida del problema $RkNN$ requiere un tiempo de orden $O(n^2)$, ya que los k vecinos más cercanos de todos los objetos n en el conjunto de datos deben encontrarse [12]. Con el rápido aumento en la escala de los grandes conjuntos de datos de entrada, se diseñaron e implementaron algoritmos paralelos y distribuidos para $RkNNQ$ en MapReduce [2] [11], pero no existen *implementaciones de RkNNQ* en Spark [23].

En general, nuestro algoritmo $RkNNQ$ paralelo y distribuido en SpatialHadoop se basa en el algoritmo SFT [17] y consiste en una serie de trabajos MapReduce. Suponiendo que \mathbb{P} es el conjunto de datos que se procesará y q es el punto de consulta, la idea básica es tener \mathbb{P} particionado por algún método (por ejemplo, Grid) en b bloques o celdas de puntos. Luego, se ejecuta una consulta $kNNQ$ basada en MapReduce para encontrar todos los posibles puntos candidatos de \mathbb{P} . Para llevarlo a cabo, encontramos la partición de \mathbb{P} donde se encuentra q . Se obtiene una primera respuesta para $kNN(\mathbb{P}, q, K)$ y usamos la distancia del punto K -ésimo a q para encontrar si hay posibles candidatos en otras particiones cercanas a q . Para garantizar un resultado exacto, el valor de K debe ser mayor que k ($K \gg k$) como se propone en [20], con una magnitud de al menos $K = 10 \times d \times k$, donde d es la dimensionalidad del conjunto de datos que se está examinando (por ejemplo, para puntos 2d, $K = 20 \times k$). A continuación, se ejecuta una consulta de rango con un círculo centrado en q con esa distancia como radio para finalmente obtener el resultado de la consulta $kNNQ$. Los candidatos con su distancia al punto de consulta q se escriben en archivos HDFS para ser la entrada de los próximos trabajos. En este momento, se verifica a cada candidato para ver si es parte de la respuesta final. Es decir, se calcula la cantidad de puntos que forman parte de la consulta de rango centrada en el punto candidato y con el radio la distancia a q . Si este *número* es menor que k , se verifica que el punto sea $RkNN$ de q . Finalmente, los resultados se escriben en archivos HDFS, almacenando solo las coordenadas de puntos y la distancia con q .

La implementación en LocationSpark usa las funciones, que éste proporciona, *knnfilter* y *rangefilter* [19] y es muy similar a la implementada para Spatial-

Hadoop. Cabe señalar que la diferencia más importante al implementar $RkNNQ$ en LocationSpark y SpatialHadoop es el hecho de que el primero no necesita almacenar resultados intermedios en el disco, ya que es un $DSDMS$ en memoria.

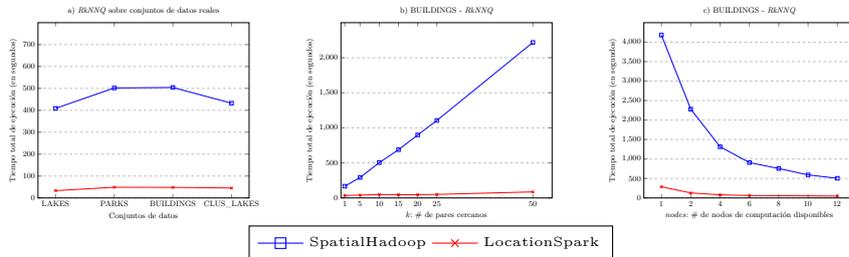


Fig. 3. Tiempos de ejecución de $RkNNQ$ considerando diferentes conjuntos de datos (gráfica a), diferentes valores de k (gráfica b) y diferente número de nodos de computación (gráfica c).

En los resultados de diferentes experimentos (Figura 3), LocationSpark vuelve a presentar mejores resultados teniendo en cuenta el tiempo de ejecución, debido a la eficiencia del procesamiento en memoria proporcionado por Spark. Sin embargo, SpatialHadoop muestra tendencias de rendimiento interesantes debido a la naturaleza del algoritmo propuesto, ya que el uso de múltiples trabajos MapReduce en un $DSDMS$ basado en disco necesita múltiples accesos de disco a conjuntos de datos.

3 Conclusiones y Trabajos futuros

El procesamiento de consultas espaciales se ha estudiado activamente en entornos centralizados, sin embargo, para marcos paralelos y distribuidos no ha conseguido una atención similar. Por esta razón, durante el último año hemos trabajado principalmente con sistemas de gestión de datos espaciales en la mejora de algoritmos espaciales, el estudio y la implementación de nuevas consultas y la comparación de diferentes sistemas. Podemos destacar las siguientes contribuciones:

- Hemos mejorado el algoritmo $KCPQ$ basado en barrido del plano para MapReduce [9] mediante el uso de nuevas técnicas de muestreo y aproximación que se aprovechan del particionado de SpatialHadoop [10].
- Hemos propuesto un nuevo algoritmo MapReduce para $\varepsilon DJQs$ en SpatialHadoop [10], basado en técnicas de barrido de plano, de forma similar a nuestro algoritmo $KCPQ$ para MapReduce.
- Hemos presentado los resultados de un extenso estudio experimental [10] que compara el rendimiento de los algoritmos MapReduce propuestos y sus mejoras en términos de eficiencia y escalabilidad.
- Se han creado nuevos algoritmos en LocationSpark [6] (los primeros en la literatura) para realizar, de manera eficiente, paralela y distribuida, $KCPQ$ y εDJQ en grandes conjuntos de datos espaciales del mundo real.

- Hemos propuesto nuevos algoritmos en SpatialHadoop y LocationSpark [8] (los primeros en la literatura) para ejecutar algoritmos eficientes paralelos y distribuidos para *RkNNQ* en grandes conjuntos de datos espaciales.
- La ejecución de un conjunto de experimentos [6, 8] para comparar el rendimiento de los dos DSDMS (SpatialHadoop y LocationSpark) ha demostrado que LocationSpark es el ganador general en cuanto a tiempo de ejecución, debido a la eficiencia del procesamiento en memoria proporcionado por Spark. Sin embargo, hay que tener en cuenta que SpatialHadoop es un DSDMS más maduro y robusto debido al tiempo dedicado a investigarlo y desarrollarlo (varios años) y proporciona más operaciones espaciales y técnicas de particionado espacial.

Trabajos futuros podrían incluir las siguiente tareas:

- Implementar otras consultas espaciales en DSDMS, como el framework KNNJ query [14] o basadas en la consulta *RkNNJ* [5].
- Estudiar otros DSDMSs como *Simba* [21] para ver si se pueden aprovechar las características que estos presentan para la mejora de las consultas ya implementadas.
- Nuestra propuesta actual de *RkNNQ* es una buena base para el desarrollo de nuevas mejoras, en las que el número de candidatos pueda ser reducido mediante la adaptación de algoritmos *RkNNQ* recientes [22] a la metodología MapReduce.
- Implementar otras técnicas de particionado, porque es un factor fundamental en el rendimiento de las consultas espaciales.

Referencias

1. Corral, A., Vassilakopoulos, M.: On approximate algorithms for distance-based queries using r-trees. *Computer Journal* 48(2), 220–238 (2005)
2. Dean, J., Ghemawat, S.: Mapreduce: Simplified data processing on large clusters. In: *OSDI Conference*. pp. 137–150 (December 2004)
3. Doulkeridis, C., Nørnvåg, K.: A survey of large-scale analytical query processing in mapreduce. *VLDB Journal* 23(3), 355–380 (2014)
4. Eldawy, A., Mokbel, M.F.: Spatialhadoop: A mapreduce framework for spatial data. In: *ICDE Conference*. pp. 1352–1363 (April 2015)
5. Emrich, T., Kriegel, H., Kröger, P., Niedermayer, J., Renz, M., Züfle, A.: On reverse-k-nearest-neighbor joins. *GeoInformatica* 19(2), 299–330 (2015)
6. García-García, F., Corral, A., Iribarne, L., Mavrommatis, G., Vassilakopoulos, M.: A comparison of distributed spatial data management systems for processing distance join queries. In: *ADBIS Conference*. pp. 214–228 (2017)
7. García-García, F., Corral, A., Iribarne, L., Vassilakopoulos, M.: Distance range queries in spatialhadoop. In: *JISBD Conference*. pp. 1–14 (2016)
8. García-García, F., Corral, A., Iribarne, L., Vassilakopoulos, M.: Rknn query processing in distributed spatial infrastructures: A performance study. In: *MEDI Conference*. pp. 200–207 (2017)

9. García-García, F., Corral, A., Iribarne, L., Vassilakopoulos, M., Manolopoulos, Y.: Enhancing spatialhadoop with closest pair queries. In: ADBIS Conference. pp. 212–225 (2016)
10. García-García, F., Corral, A., Iribarne, L., Vassilakopoulos, M., Manolopoulos, Y.: Efficient large-scale distance-based join queries in spatialhadoop. *GeoInformatica* (2017) <https://doi.org/10.1007/s10707-017-0309-y> (2017)
11. Ji, C., Qu, W., Li, Z., Xu, Y., Li, Y., Wu, J.: Scalable multi-dimensional rnn query processing. *Concurrency and Computation: Practice and Experience* 27(16), 4156–4171 (2015)
12. Korn, F., Muthukrishnan, S.: Influence sets based on reverse nearest neighbor queries. In: SIGMOD Conference. pp. 201–212 (May 2000)
13. Li, F., Ooi, B.C., Özsu, M.T., Wu, S.: Distributed data management using mapreduce. *ACM Comput. Surv.* 46(3), 31:1–31:42 (2014)
14. Nodarakis, N., Pitoura, E., Sioutas, S., Tsakalidis, A.K., Tsoumakos, D., Tzimas, G.: kdann+: A rapid aknn classifier for big data. *Trans. Large-Scale Data- and Knowledge-Centered Systems* 24, 139–168 (2016)
15. Roumelis, G., Vassilakopoulos, M., Corral, A., Manolopoulos, Y.: New plane-sweep algorithms for distance-based join queries in spatial databases. *GeoInformatica* 20(4), 571–628 (2016)
16. Silva, Y.N., Reed, J.M.: Exploiting mapreduce-based similarity joins. In: SIGMOD Conference. pp. 693–696 (May 2012)
17. Singh, A., Ferhatosmanoglu, H., Tosun, A.S.: High dimensional reverse nearest neighbor queries. In: Proceedings of the twelfth international conference on Information and knowledge management. pp. 91–98. ACM (2003)
18. Stanoi, I., Agrawal, D., El Abbadi, A.: Reverse nearest neighbor queries for dynamic databases. In: ACM SIGMOD workshop on research issues in data mining and knowledge discovery. pp. 44–53 (2000)
19. Tang, M., Yu, Y., Malluhi, Q.M., Ouzzani, M., Aref, W.G.: Locationspark: A distributed in-memory data management system for big spatial data. *PVLDB* 9(13), 1565–1568 (2016)
20. Tao, Y., Papadias, D., Lian, X.: Reverse knn search in arbitrary dimensionality. In: Proceedings of the Thirtieth international conference on Very large data bases-Volume 30. pp. 744–755. VLDB Endowment (2004)
21. Xie, D., Li, F., Yao, B., Li, G., Zhou, L., Guo, M.: Simba: Efficient in-memory spatial analytics. In: SIGMOD Conference. pp. 1071–1085 (June 2016)
22. Yang, S., Cheema, M.A., Lin, X., Wang, W.: Reverse k nearest neighbors query processing: Experiments and analysis. *PVLDB* 8(5), 605–616 (2015)
23. Zaharia, M., Chowdhury, M., Das, T., Dave, A., Ma, J., McCauley, M., Franklin, M.J., Shenker, S., Stoica, I.: Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing. In: NSDI Conference. pp. 15–28 (2012)
24. Zhang, H., Chen, G., Ooi, B.C., Tan, K.L., Zhang, M.: In-memory big data management and processing: A survey. *IEEE Transactions on Knowledge and Data Engineering* 27(7), 1920–1948 (2015)
25. Zhang, S., Han, J., Liu, Z., Wang, K., Feng, S.: Spatial queries evaluation with MapReduce. In: GCC Conference. pp. 287–292 (August 2009)

Computación de altas prestación para la resolución de problemas de optimización global

Jose Manuel García Salmerón

Grupo Supercomputación-Algoritmos, Departamento de Informática, Universidad de Almería, josemanuel@ual.es

Resumen En Optimización Global mediante Ramificación y Acotación cuando el espacio de búsqueda es un símplice es habitual usar la bisección por lado mayor como método de refinamiento. En este trabajo se han estudiados diferentes algoritmos, tanto paralelos como secuenciales, para abordar instancias grandes de este problema. Por otra parte, se están estudiando métodos de división del símplice alternativos a partir de símplices regulares. Todos los algoritmos desarrollados han sido probados en problemas reales.

Keywords: Símplice, Bisección, División regular, Branch and bound, Árbol Binario, Memoria Compartida, Multicore, Manycore, PThreads, Intel Xeon Phi, Matriz copositiva, Blending.

1. Introducción

Los algoritmos de Ramificación y Acotación (del inglés *Branch and Bound*, B&B) se aplican a la resolución de problemas de Optimización Global mediante la realización de una búsqueda exhaustiva del mínimo de una función objetivo en un dominio dado.

$$f^* = f(x^*) = \min_{x \in S} f(x). \quad (1)$$

En este trabajo estamos interesados en problemas donde el espacio de búsqueda está definido por un n -símplice regular, S , definido en un espacio n -dimensional [7].

$$S_1 = \{x \in \mathbb{R}^n \mid \sum_{i=1}^n x_i = 1; x_i \geq 0, i = 1, \dots, n\}. \quad (2)$$

Los algoritmos B&B están caracterizados por las reglas de Acotación, Selección, División, Rechazo y Terminación [5]. En la primera parte del trabajo se pretende estudiar cómo conseguir el árbol de búsqueda generado con tamaño mínimo cuando solo se tienen en cuenta las reglas de División y Terminación, es decir, ningún nodo del árbol es eliminado. Se usará como regla de división la Bisección del Lado Mayor (BLM, ó *Longest Edge Bisection*, LEB, en inglés) (véase [1,9]) y como regla de terminación la longitud del lado mayor de un subproblema o nodo del árbol. Para ello investigamos el efecto de la BLM sobre el número de símplices generados. El problema de obtener el árbol binario de

Computación de altas prestación para la resolución de problemas de optimización global

tamaño menor usando como refinamiento el método de bisección por lado mayor fue introducido en [13] como un problema de Optimización Combinatoria. El problema combinatorio aparece con las múltiples opciones de selección de lado a dividir (véase Figura 1b dónde tras realizar la primera bisección aparecen tres opciones de lado mayor a dividir). El objetivo final de nuestro estudio es poder establecer una regla de división que genere el menor árbol, permitiendo así reducir el coste computacional de los algoritmos B&B aplicados a un espacio inicial determinado por un n -símplice regular [10]. Las arquitecturas y algoritmos paralelos permiten abordar instancias de este problema de optimización combinatoria, que por su dimensión o tamaño del árbol generado requieren mucho tiempo de computación en su versión secuencial. Para obtener una solución en estas instancias, se hace necesario el desarrollo de algoritmos paralelos [2, 12, 14]. En esta parte del trabajo se han estudiado y desarrollado diferentes versiones paralelas del algoritmo sobre diversas arquitecturas de memoria compartida que nos permiten determinar el tamaño del árbol mínimo en un tiempo razonable. La sección 2.1 describe los diferentes algoritmos desarrollados y compara los resultados obtenidos en distintas arquitecturas.

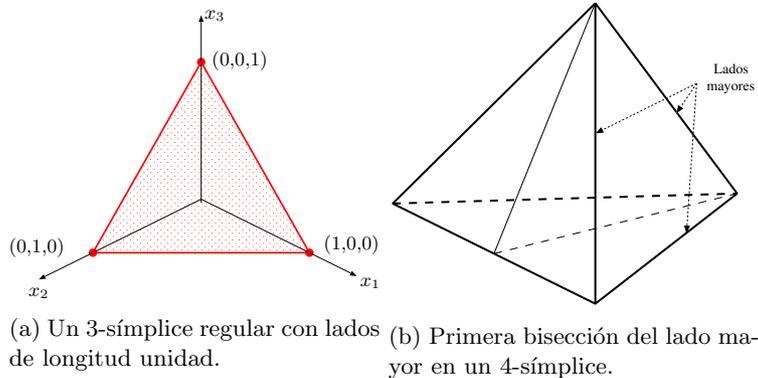


Figura 1: Ejemplos símplices.

El esquema básico de un algoritmos de B&B se muestra en el algoritmo 1. El comportamiento, y por tanto la eficiencia del algoritmo dependerá de como se establezcan las reglas de Selección, Acotación, Rechazo, División y Terminación. Como hemos mencionado, en este estudio no estudiaremos las reglas de Acotación, Rechazo ni de Selección ya que solo estamos interesados en encontrar la regla de División basada en la bisección del lado mayor que genere un árbol de menor tamaño. Por lo tanto, no existe poda y el árbol se generaría completamente. Su tamaño no solo dependerá del lado mayor elegido para ser dividido, sino también de la regla de terminación usada. La regla de terminación no procesa más un simplex S cuando $w(S) \leq \epsilon \cdot w(S_1)$, donde $w(S)$ es el tamaño del simplex que se define como la longitud de su lado mayor.

Computación de altas prestación para la resolución de problemas de optimización global

Algorithm 1 Ramificación y Acotación

Require: S_1 : símlice inicial, ϵ : precisión

```

1:  $\Lambda := \{S_1\}$  // Conjunto de trabajo
2:  $\Omega := \{\}$  // Conjunto final
3:  $ns := 1$  // Número de símlices
4: while  $\Lambda \neq \emptyset$  do
5:   Selecciona  $S_i$  de  $\Lambda$  // Regla de selección
6:   Evalúa  $S_i$  // Regla de acotación
7:   if  $S_i$  no puede ser eliminado then // Regla de rechazo
8:     if  $w(S_i) \leq \epsilon$  then // Regla de terminación
9:       Almacena  $S_i$  in  $\Omega$ 
10:    else
11:       $\{S_{2i}, S_{2i+1}\} := \text{BLM}(S_i)$  // Regla de división
12:      Almacena  $S_{2i}, S_{2i+1}$  en  $\Lambda$ 
13:       $ns := ns + 2$ 
14: return  $\Omega$  y  $f(x^*)$ 

```

Como se ha mencionado, habitualmente se emplea BLM en el proceso de refinamiento de un símlice porque es un método sencillo que garantiza la convergencia del algoritmo en una solución. Sin embargo con este método puede ocurrir que más de un lado sea candidato para la división. Por lo tanto, una forma alternativa de realizar la división consiste en emplear símlices regulares [8]. Un simplex regular Δ está definido por su centro c y su radio r de forma que sus vértices se definen como

$$V = c\mathbf{1}^T + \rho D, \quad (3)$$

donde $\mathbf{1}$ es el vector de unos y $D = (d_1, \dots, d_n) = E - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ es una matriz simétrica con las direcciones desde el centro a los vértices del símlice. Señalar que el centro del símlice unidad S es $c = \frac{1}{n}\mathbf{1}$, en cambio su radio es 1.

Las principales ventajas de emplear símlices regulares para el proceso de refinamiento son su facilidad para almacenarnos con dos valores (centro y radio), la regularidad de las divisiones, la efectividad de algunos test de rechazo debido a que todos los vértices están a la misma distancia entre si. Sin embargo, y como se puede observar en la figura 2, para dividir un símlice regular en sub-símlices regulares es necesario que se produzca un solape entre los sub-símlices. Las zonas solapadas deben ser evaluadas solo una vez con el fin de no repetir trabajo, por lo que controlar el solape es fundamental para esta forma de división.

Todo lo introducido anteriormente será descrito con más detalle en la siguiente sección, en la que además se presentarán los resultados obtenidos al aplicar estos métodos de refinamiento del símlice en problemas de diseño de mezclas y detección de matrices copositivas.

Computación de altas prestación para la resolución de problemas de optimización global

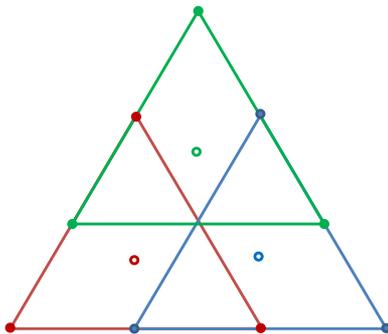


Figura 2: División regular en 3 dimensiones. Los puntos representan los vértices y centros de los sub-símplices.

2. Avances

Se describen los principales desarrollos en cada una de las líneas de trabajo del proyecto de tesis. Posteriormente en la sección 3 se describen los resultados obtenidos hasta la fecha y las posibles líneas de trabajo futuro.

2.1. Tamaño del árbol mínimo

En esta sección se pretende dar una visión global del trabajo realizado para determinar el tamaño del árbol mínimo empleando BLM, así como el trabajo realizado en [10] y [13] en el que se busca una heurística de selección del lado a dividir que genere el árbol mínimo, estudio que comenzó en [3].

Algoritmo secuencial para el cálculo del tamaño del árbol binario mínimo

El algoritmo secuencial debe comprobar todas las opciones de división de lado mayor, y tener en cuenta solo aquellas que generan un árbol binario de menor tamaño. Esta acción se puede realizar mediante recursión, como muestra el algoritmo 2, el cual explora todas las opciones de división realizando una búsqueda en profundidad. Se realiza una exploración en profundidad porque presenta unos requerimientos de memoria menores que otros tipos de búsqueda. Los parámetros iniciales del algoritmo 2 son el n -simplex regular inicial S_1 , uno de sus lados L (todos son iguales) y la precisión requerida ϵ .

La figura 3 muestra la ejecución del algoritmo 2 para un 2-simplex de 3 dimensiones y $\epsilon = 0,5$. La figura permite resaltar algunos aspectos que, para simplificar, no se han incluido en el algoritmo:

- Si los símplexes hermanos S_2 y S_3 son simétricos, ambos generan sub-árboles de igual tamaño. Por lo tanto, solo es necesario procesar uno de ellos. En ese caso se devuelve como resultado el doble del tamaño del sub-árbol procesado.

Algorithm 2 TamañoÁrbolMínimo(S, L, ϵ)

Require: S : simplex, L : lado mayor, ϵ : precisión

```

1: if  $w(S) \leq \epsilon$  then
2:   return 1
3:  $\{S_l, S_r\} := \text{DivideSímplice}(S, L)$ 
4: for cada lado mayor  $L_i$  de  $S_l$  do
5:    $rl_i := \text{TamañoÁrbolMínimo}(S_l, L_i, \epsilon)$ 
6: for cada lado mayor  $L_i$  de  $S_r$  do
7:    $rr_i := \text{TamañoÁrbolMínimo}(S_r, L_i, \epsilon)$ 
8:  $r_l := \text{mín}_i\{rl_i\}$ 
9:  $r_r := \text{mín}_i\{rr_i\}$ 
10: return  $1 + r_l + r_r$ 

```

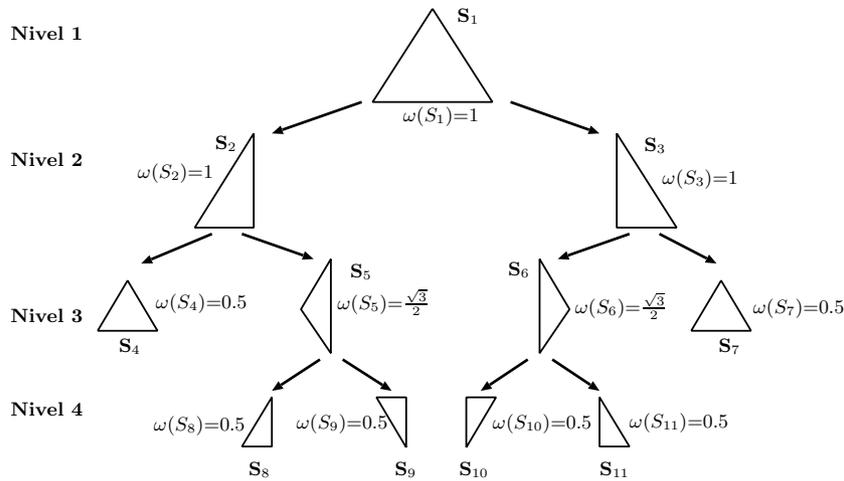


Figura 3: Árbol binario para un simplex de dimensión 3 y $\epsilon = 0,5$.

Por ejemplo, los simplexes S_2 y S_3 o S_8 y S_9 de la figura 3 son hermanos simétricos.

- Si el simplex es regular, solo hay que procesar un lado, ya que todos los lados son de igual tamaño, produciendo además hermanos simétricos. Por ejemplo, los simplexes S_1 , S_4 y S_7 de la figura 3 son regulares.

En el algoritmo 2, sería conveniente poder determinar la existencia de parejas simétricas de hermanos, pero esto es computacionalmente más costoso que determinar si dos simplexes hermanos son simétricos y se abordará en trabajos futuros.

Algoritmos paralelos para el cálculo del tamaño del árbol binario mínimo

El paralelismo es fácilmente aplicable en la construcción de árboles binarios ya que las ramas del árbol pueden visitarse en paralelo. Uno de los inconvenientes de los árboles que se crean en este trabajo es que existen ramas con distintos niveles, tal como muestra la figura 3. La diferencia de niveles aumenta conforme lo hace n . En los trabajos [2, 12, 14] se han desarrollado distintos algoritmos paralelos con los que obtener el tamaño del árbol mínimo. Ha forma de resumen de estos trabajos las versiones paralelas más relevantes son:

B2F : Basada en dos fases. En la primera fase secuencial se generan los posibles sub-símplices hasta que su número sea mayor o igual al número de hebras determinado por el usuario. En la segunda fase se ejecuta el algoritmo secuencial para cada sub-símplice en paralelo. Finalmente, una vez todas las hebras han terminado su trabajo se recogen los resultados parciales de cada una para obtener el tamaño del árbol mínimo. Las dos principales desventajas de este algoritmo son, por un lado, la cantidad de tiempo necesaria para la primera fase hasta generar suficiente trabajo para todas las hebras, y por el otro, el desbalanceo de la carga que se produce por asignar el trabajo completamente desde el inicio.

CDH : Mediante la creación dinámica de hebras se crea una hebra en cualquier momento que se tenga trabajo pendiente hasta alcanzar el número máximo de hebras definido por el usuario, obtenido un balanceo implícito de la carga. Por otra parte se utiliza un *Cut-off*, que en símplexes próximos a ϵ ejecuta el algoritmo secuencial para evitar la creación de hebras en niveles muy bajos del árbol que generarían un bajo grado de paralelismo. La principal limitación de este algoritmo radica en los tiempos de creación y destrucción de las hebras, así como en los tiempo de espera de resultados entre la hebra y todas las hebras nuevas que dependen de ella.

IPTH : Con la idea de resolver los problemas presentados por los dos algoritmos anteriores se realizó una implementación, que en lugar de seguir el clásico esquema recursivo de los algoritmos de B&B, emplea un planteamiento iterativo. En IPTH una hebra generada para el calcular el tamaño mínimo de un sub-árbol, almacena ese tamaño en el nodo correspondiente, explora otros sub-árboles pendiente en el nodo actual, termina el nodo actual si todos los sub-árboles han sido calculados, o busca trabajo en ramas superiores del árbol si el trabajo pendiente ya está asignado a otra hebras. De esta forma las hebras no están constantemente creándose y destruyéndose sino que una vez creadas continúan trabajando mientras encuentren trabajo por hacer y solo se destruyen cuando, durante la subida por el árbol en busca de trabajo, llegan al nodo raíz del árbol sin haber encontrado trabajo pendiente, aunque este puede existir en ramas que están siendo procesadas. Las hebras que alcanzan la raíz del árbol, al morir permiten la creación de nuevas hebras que ayuden a las que están procesando niveles inferiores del árbol general.

Estos algoritmos han sido evaluados en diferentes arquitecturas de memoria compartida tradicionales y en una tarjeta Intel Xeon Phi. Teniendo en cuenta,

Computación de altas prestación para la resolución de problemas de optimización global

no solo el tiempo de ejecución, sino también el consumo de memoria y energía realizado por cada uno de ellos.

La comparativa de las maquinas de memoria compartida pueden encontrarse en [2], donde el algoritmo IPTH implementado en C/C++ y una implementación similar a CDH, implementada mediante tareas de Intel TBB, obtienen resultados muy ajustados en cada instancia evaluada. Hay que destacar IPTH en instancias grandes, donde el consumo de memoria es determinante.

Tras los buenos resultados respecto al uso de la memoria del algoritmo IPTH se quiso evaluar este algoritmo en una arquitectura con una limitada cantidad de memoria como es la Intel Xeon Phi [12]. La comparación entre el equipo con CPUs Xeon y tarjeta Xeon Phi muestra que la Phi, ni usando todos sus cores, es competitiva frente a una CPU tradicional. Si bien es cierto que el precio de la Phi es considerablemente inferior a la de los procesaros Xeon, la relación entre rendimiento, consumo y coste no hacen a la Phi como una buena opción para resolver este tipo de problemas.

Selección del lado mayor para la generación del menor árbol binario

En [13] se continuó el trabajo comenzado en [3], desarrollando un algoritmo determinístico que analiza el árbol de búsqueda general en busca de patrones o similitudes que puedan conformar un conjunto de reglas de selección del lado a dividir y que generen el árbol binario mínimo. Este trabajo demostró que la secuencia presentada para $n = 4$ en [3] es única y mostró la complejidad del problema combinatorio abordado, haciendo patente el desarrollo de algoritmos paralelos para resolver instancias mayores. Posteriormente y ante la dificultad presentada por el algoritmo determinista de [13] se ha comenzado a trabajar en un nuevo algoritmo metaheurístico basado en las mismas ideas pero que incluye elementos metaheurísticos que permiten obtener resultados con un tiempo y recursos menores a los del algoritmo determinístico. Este nuevo algoritmo fue presentado en [10] y continua en desarrollo.

2.2. Problema de diseño de mezclas (Blending)

Una vez presentados los distintos métodos de refinamiento de símplexes, tanto la habitual bisección del lado mayor como la nueva división por símplexes regulares, en [11] se busca evaluar cada uno de ellos a través de un problema de diseño de mezclas. Un problema de diseño de mezclas (conocido como *Blending* en inglés) consiste en identificar los productos de una mezcla (materiales o ingredientes), cada uno representado por el vector $x \in \mathbb{R}^n$, con restricciones lineales o cuadráticas en su diseño. El conjunto de posibles mezclas se define matemáticamente por el símplex unidad, donde el valor de cada coordenada x_j , según la definición de símplex de la ecuación 2, representa la fracción de cada material en un producto x .

En problemas de diseño de mezclas, el objetivo es minimizar el coste del producto (minimizando el coste de cada materia, $f(x) = c^T x$, donde el vector c proporciona el coste del material) manteniendo las expectativas de calidad. Las variables con las que se trabaja no son continuas debido a la mínima dosis md

Computación de altas prestación para la resolución de problemas de optimización global

que el problema presenta en la practica, $x_j = 0$ or $x_j \geq md$. La figura 4 muestra de forma gráfica esta dosis mínima para un espacio de búsqueda 2-dimensional (figura de la izquierda) y para uno 3-dimensional (figura de la derecha). Está representación consiste en un símplex unidad al cual se le ha quitado el espacio correspondiente a la dosis mínima permitida.

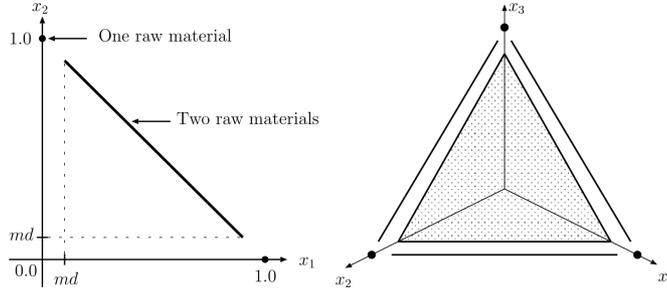


Figura 4: Símplexes 2D y 3D teniendo en cuenta la dosis mínima.

El algoritmo desarrollado para resolver el problema planteado sigue el esquema de un algoritmo de B&B como el algoritmo 1 y sus reglas fueron presentadas en [6]. Para probar el algoritmo nos hemos centrado en 5 tests de prueba. Cada uno de estos test presenta diferencias en cuanto a la dimensión y el numero de restricciones (lineales y cuadráticas), algunos han sido tomados directamente de la industria. Se han comparado la división por lado mayor con la división usando símplexes regulares.

Los resultados obtenidos muestran que para problemas pequeños la división empleado símplexes regulares en satisfactoria, reduciendo la cantidad de evaluaciones necesarias, permitiendo además obtener soluciones para precisiones que la división por lado mayor no es capaz de resolver. Sin embargo, en los problemas de mayor dimensión la división regular divide el espacio de búsqueda muy poco a poco, lo que incrementa notablemente el número de evaluaciones.

2.3. Detección de matrices copositivas (Copositivity)

La copositividad juega un importante rol en Optimización Combinatoria y Cuadratica. Durante las últimas décadas la detección de matrices copositivas mediante la división de símplexes ha hecho un gran avance. Recientemente ha demostrado que el uso de división mediante símplexes regulares puede presentar algunas ventajas sobre la bisección. La detección de matrices copositivas consiste en determinar si la matriz A en cada punto de \mathbb{R}^+ satisface que $\forall x \in S, x^T Ax \geq 0$.

Al igual que en la sección anterior sobre le diseño de mezclas, esta sección pretende introducir los desarrollos que se están llevado acabo actualmente sobre

el problema de las matrices copositivas. En este caso, se busca evaluar la división regular de símlices en la detección de matrices copositivas (*Copositivity detection* en inglés) frente a un método de bisección adaptado a este problema en particular [4]. Para hacer frente a esta bisección específica se está trabajando en adaptar nuestra división regular a este problema en particular.

Esta línea de trabajo es la más reciente que estamos llevando a cabo por lo que los detalles y resultados aún no están publicados. Se mostrarán algunos resultados preliminares durante la exposición.

3. Conclusiones

En este informe se han resumido las diferentes líneas de trabajo abiertas dentro de proyecto de tesis que se está desarrollando. Destacando el desarrollo de algoritmo paralelos que han permitido obtener resultados para el problema de determinar el árbol binario de tamaño mínimo en menor tiempo y abordando instancias mayores del problema. Junto al desarrollo de estos algoritmos se han realizado evaluaciones de los mismos en diversas arquitecturas para comprobar el rendimiento de cada uno de ellos según la instancia abordada. Con el mismo fin, determinar el tamaño del árbol mínimo, se ha desarrollado un algoritmo de búsqueda de patrones cuyo objetivo es obtener un conjunto de reglas de selección del lado a dividir que permitan llegar al árbol mínimo sin necesidad de evaluar todas las opciones.

Por otra parte, destacar todo el trabajo que se está llevando a cabo en la división regular de símlices. Trabajando en métodos que reducen el solape y aprovechan las características propias de cada problema. Estos nuevos métodos han sido probados en problemas de diseño de mezclas y en la detección de matrices copositivas. En cuanto a la línea de trabajo en detección de matrices copositivas, su estudio está avanzado y se espera realizar una publicación con los desarrollos realizados próximamente. Relativo a esto y a la división regular de símlices, se han descubierto formas de aprovechar los símlices regulares para este problema concreto y se espera que los resultados mejoren a los métodos actuales empleados en este problema.

Referencias

1. Andrew Adler. On the bisection method for triangles. *Mathematics of Computation*, 40(162):571–574, 1983. DOI:10.1090/S0025-5718-1983-0689473-5.
2. G. Aparicio, J.M.G. Salmerón, L. G. Casado, R. Asenjo, and E.M.T. Hendrix. Parallel algorithms for computing the smallest binary tree size in unit simplex refinement. *Journal of Parallel and Distributed Computing*, 112(2):166–178, 2017. DOI:10.1016/j.jpdc.2017.05.016.
3. Guillermo Aparicio, Leocadio G. Casado, Boglárka G-Tóth, Eligius M. T. Hendrix, and Inmaculada García. On the minimum number of simplex shapes in longest edge bisection refinement of a regular n-simplex. *Informatica*, 26(1):17–32, 2015.

Computación de altas prestación para la resolución de problemas de optimización global

4. S. Bundfuss and M. Dür. Algorithmic copositivity detection by simplicial partition. *Linear Algebra and its Applications*, 428(7):1511–1523, 2008. DOI:10.1016/j.laa.2007.09.035.
5. L. G. Casado, E.M.T. Hendrix, and I. García. Infeasibility spheres for finding robust solutions of blending problems with quadratic constraints. *Journal of Global Optimization*, 39(2):215–236, 2007. DOI:/10.1007/s10898-006-9072-6.
6. Leocadio G. Casado, Eligius M. T. Hendrix, and Inmaculada García. Infeasibility spheres for finding robust solutions of blending problems with quadratic constraints. *Journal of Global Optimization*, 39(4):577–593, Dec 2007.
7. L.G. Casado, I. García, B.G. Tóth, and E.M.T. Hendrix. On determining the cover of a simplex by spheres centered at its vertices. *Journal of Global Optimization*, 50(4):645–655, 2011. DOI:10.1007/s10898-010-9524-x.
8. B. G.-Tóth, E. M. T. Hendrix, L. G. Casado, and I. García. On refinement of the unit simplex using regular simplices. *Journal of Global Optimization*, 64(2):305–323, 2016.
9. R. Horst. On generalized bisection of n -simplices. *Mathematics of Computation*, 66(218):691–698, 1997. DOI:10.1090/S0025-5718-97-00809-0.
10. E.M.T. Hendrix J.M.G. Salmerón, J.L. Redondo and L.G. Casado. On metaheuristics for longest edge bisection of the unit simplex to minimise the search tree size. pages 41–44, September 2016. *Proceedings of the XIII Global Optimization Workshop*.
11. E.M.T. Hendrix J.M.G. Salmerón, L.G. Casado and J.F.R Herrera. On regular simplex division in solving blending problems. pages 45–48, September 2016. *Proceedings of the XIII Global Optimization Workshop*.
12. L.G. Casado J.M.G. Salmerón, J.J. Moreno and E.M.T. Hendrix. Tamaño del árbol mínimo en el refinamiento de un simplex regular sobre arquitecturas mic. Septiembre 2017. *Jornadas de Paralelismo 2017*, DOI:10.5281/zenodo.894031.
13. J.M.G. Salmerón, G. Aparicio, L.G. Casado, I. García, E.M.T. Hendrix, and B.G. Tóth. Generating a smallest binary tree by proper selection of the longest edges to bisect in a unit simplex refinement. *Journal of Combinatorial Optimization*, pages 1–14, 2017. DOI:10.1007/s10878-015-9970-y.
14. J.M.G. Salmerón, J.F.R. Herrera, G. Aparicio, L.G. Casado, I. García, and E.M.T. Hendrix. Estrategias paralelas para obtener el tamaño del árbol mínimo en la división por el lado mayor de un símplice regular. pages 29–35, Septiembre 2014. *Actas XXV Jornadas de Paralelismo*.

Estrategias de control jerárquico y optimización aplicadas a plantas de destilación por membranas alimentadas con energía solar

Doctorando: Juan D. Gil

Centro Mixto CIESOL, ceiA3, Universidad de Almería.
{juandiego.gil }@ual.es

Abstract. La destilación por membranas alimentadas con energía solar (*Solar Membrane Distillation*, SMD) es una tecnología de desalación en fase de investigación, adecuada para el desarrollo de plantas autónomas capaces de cubrir requerimientos medios de demanda de agua. Esta tesis tiene como objetivo la aportación de contribuciones desde el punto de vista de control y optimización que supongan un avance hacia la comercialización de la tecnología de destilación por membranas. Así pues, se pretende demostrar que el control automático y los algoritmos de optimización pueden contribuir a reducir los costes económicos y a mejorar la eficiencia de la operación, mediante una gestión óptima de la energía solar, combinada con la realización de diseños óptimos de las instalaciones que se ajusten a las necesidades específicas de cada escenario.

1. Introducción

En las últimas décadas, la escasez de agua se está convirtiendo en uno de los principales retos que debe afrontar la humanidad. El crecimiento de la población junto con el incremento de las actividades agrícolas e industriales, han contribuido a una sobreexplotación de las reservas de agua dulce, sobrepasando el límite de renovación de dicho recurso. De este modo, las tecnologías de desalación se están convirtiendo progresivamente en un elemento necesario fundamental, especialmente en las zonas áridas o semiáridas con escasez de agua. Estas tecnologías requieren intensivos sistemas de generación de energía para su funcionamiento, por lo que deben ser asociadas con fuentes de energía renovables para su sostenibilidad económica. El uso de fuentes de energía renovables en los procesos de desalación no solo reduce los costes económicos de dichos procesos, sino que también reemplaza el uso de fuentes tradicionales como los combustibles fósiles, contribuyendo así a un desarrollo medioambiental sostenible y eficiente.



Figura 1. Instalación piloto de la PSA.

En este contexto, la Destilación por Membranas (*Membrane Distillation*, MD) con apoyo de energía solar es una tecnología apropiada para el desarrollo de pequeñas plantas autosuficientes de desalación, que pueden ser implantadas en zonas aisladas con unas buenas condiciones de irradiancia solar [1]. Esta tecnología destaca por su baja temperatura de operación, que permite que sea fácilmente acoplable a tecnologías solares. Además, tiene una serie de características que hacen que la planta pueda ser completamente automatizada, como son la sencillez del proceso, la fiabilidad y los bajos requerimientos de mantenimiento. No obstante, esta tecnología se encuentra actualmente en fase de investigación y todavía no ha sido comercializada a escala industrial, debido principalmente a problemas técnicos en el diseño del módulo, problemas de humectación en las membranas, a la baja producción de destilado y a la incertidumbre asociada a los costes económicos [2]. Una de las pocas plantas (ver Fig. 1) descrita totalmente en la literatura [1] se encuentra en la Plataforma Solar de Almería (PSA, www.psa.es).

MD es un proceso de separación impulsado térmicamente que se produce en una membrana hidrófoba y microporosa. La fuerza impulsora del proceso es el gradiente de presión que se genera a ambos lados de la membrana, como resultado de una diferencia de temperatura. De esta forma, las moléculas de agua se evaporan y pasan a través de la membrana, mientras que los componentes no volátiles son rechazados. Como todas las tecnologías de destilación térmica, los procesos MD tienen la capacidad de tratar agua con concentraciones de sal elevadas, pero sin necesidad de laboriosos procesos de pretratamiento químico del agua de alimentación, obteniendo permeados de alta pureza. Los sistemas MD suelen ser clasificados dependiendo del lugar donde se produzca la condensación del permeado [3]. Las configuraciones más empleadas son: Destilación por Contacto Directo (*Direct Contact MD*, DCMD) y Destilación por Membranas con Espacio de Aire (*Air Gap MD*, AGMD) en los cuales el proceso de condensación se produce dentro del módulo, y Destilación por Membranas con Barrido de Gas (*Sweeping Gas MD*, SGMD), Destilación por Membrana Líquida (*Permeate Gap MD*, PGMD) y Destilación en Condiciones de Vacío (*Vacuum MD*, VMD) en los que la condensación tiene lugar en un condensador externo al módulo.

Como se ha mencionado anteriormente, en la PSA se encuentra una de las pocas instalaciones piloto MD que existen en la actualidad, en la cual se llevan

Estrategias de control jerárquico y optimización aplicadas a plantas de destilación...

a cabo investigaciones dirigidas a la caracterización y evaluación de diferentes tipos de módulos y configuraciones [4]. En la planta MD de la PSA (ver Fig. 2), la energía térmica necesaria para el proceso de destilación la proporciona un campo solar formado por captadores planos dispuestos en dos filas de cinco captadores cada una. La potencia nominal del campo es de 7 kW a unos 90 °C. A la salida del campo solar hay instalado un aerotermo, que se utiliza para evitar excesos de temperatura que puedan dañar las membranas. Posteriormente, se dispone un tanque de almacenamiento térmicamente aislado (1500 L), que se emplea como buffer o almacenamiento energético y que dota a la instalación de un carácter híbrido, permitiendo la producción de destilado en varios modos de operación tal y como se presentó en [5]. A continuación, se encuentra el sistema de distribución, el cual posibilita la conexión de varios módulos simultáneamente al sistema de generación de energía. Cada módulo se conecta al sistema de distribución a través de su propio intercambiador.

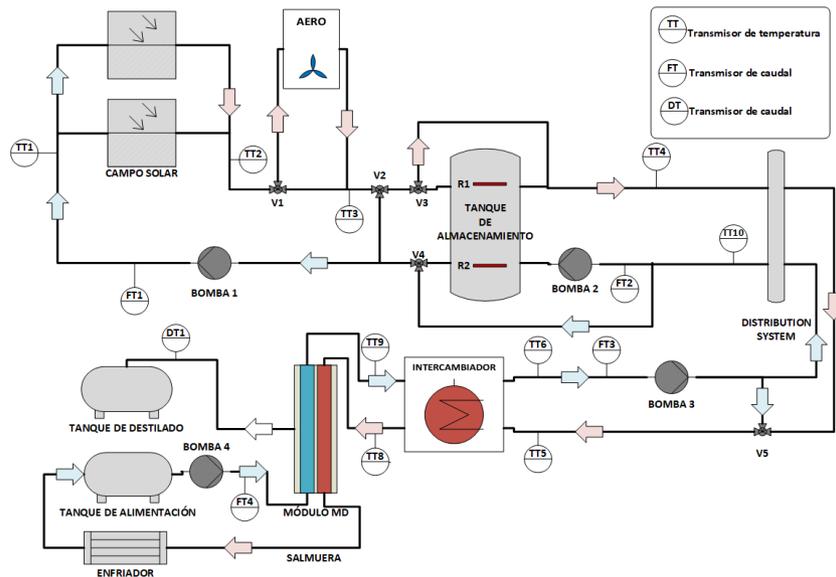


Figura 2. Diagrama esquemático de la instalación.

Dentro del módulo (ver Fig. 3), la bomba 4 impulsa la solución de alimentación (agua de mar) hacia el canal de condensación del módulo. Cuando dicha solución llega al intercambiador de calor, es calentada con el fluido proveniente del sistema de generación de energía. A continuación, la solución de alimentación caliente es dirigida hacia el canal de evaporación del módulo, de modo que las moléculas volátiles de la solución se evaporan y pasan a través de la membrana, mientras que los componentes no volátiles se rechazan en forma de salmuera y son reconducidos al tanque de alimentación. El permeado es posteriormente

Estrategias de control jerárquico y optimización aplicadas a plantas de destilación...

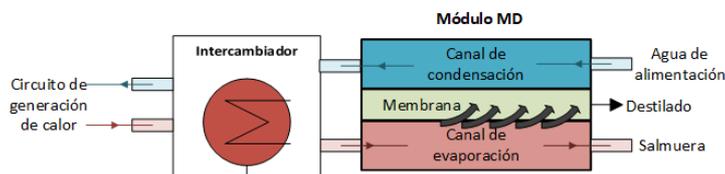


Figura 3. Diagrama esquemático del módulo MD.

condensado y vertido en el tanque de destilado. Se debe tener en cuenta que durante una operación, en el tanque de alimentación se va incrementando la temperatura y la salinidad de la solución inicial, debido a la recirculación de la salmuera. Para mantener las condiciones deseadas se utiliza un tanque auxiliar y un enfriador. Toda la planta está completamente monitorizada por medio de un PLC (*Programmable Logic Controller*) y un sistema de supervisión SCADA (*Supervisory Control And Data Acquisition*) con un tiempo de muestreo de 1 segundo.

En la literatura, hay muy pocos trabajos relacionados con el control y optimización de procesos MD. Desde el punto de vista del control automático casi todos los trabajos han sido realizados en simulación, y proponen estructuras de control básicas formadas por lazos individuales. En [6], se desarrolla un modelo de una planta MD y se realiza una optimización fuera de línea de su comportamiento. Posteriormente se propone un sistema de control de temperatura basado en controladores ON/OFF que intenta seguir los puntos óptimos, obtenidos por el algoritmo de optimización, durante una operación real simulada. En [7], nuevamente se emplean controladores del tipo ON/OFF para el control de la diferencia de temperatura que se produce entre el tanque y el campo solar, de modo que se desarrollan dos modos de operación: diario y nocturno. En [8], se prueban en simulación dos bucles de control destinados a controlar la temperatura del campo solar de alimentación. En [9], se desarrolla un generador de consignas que pretende mantener una diferencia de temperatura predefinida entre las dos partes de membranas. Un enfoque de control más completo es el presentado en [10], en el cual se propone un algoritmo de control óptimo en tiempo real para regular el caudal de alimentación acorde a las condiciones de operación, intentando maximizar la producción de destilado. Por último, la principal contribución de control hasta la fecha es la presentada en [11], donde se desarrolla un modelo de red neuronal que se utiliza para el análisis del comportamiento del sistema bajo diferentes condiciones de operación, y que posteriormente es empleado para implementar un sistema de control que optimiza la producción de destilado en base a las condiciones de operación.

La mayoría de trabajos de optimización propuestos hasta la fecha, se centran en la búsqueda de condiciones de operación óptimas para módulos concretos [12,13]. La metodología empleada en todos los trabajos suele ser la misma. En primer lugar, se elabora un modelo del módulo MD basado en datos experimentales. Posteriormente, se aplica un algoritmo de optimización que tiene como

objetivo maximizar la producción o la eficiencia térmica del módulo. Las variables de decisión de los problemas de optimización planteados suele ser la temperatura de alimentación, el caudal de alimentación y la temperatura que debe proporcionar la fuente de energía. El único trabajo que propone un algoritmo de optimización destinado al diseño óptimo de la planta es [14]. En este trabajo se intenta minimizar los costes económicos anuales, planteando un problema de optimización que usa como variables de decisión el área de la membrana del módulo, el área efectiva del campo solar, el caudal de recirculación por el campo solar y el caudal de alimentación.

El objetivo de esta tesis consistirá en realizar aportaciones que supongan un avance hacia la comercialización a gran escala de la tecnología MD. Así pues, se pretende demostrar que el control automático y los algoritmos de optimización pueden contribuir a reducir los costes económicos y a mejorar la eficiencia energética de la operación diaria, mediante una gestión óptima de la energía solar. Por otro lado, se pretenden desarrollar algoritmos de optimización que permitan la realización de diseños óptimos de las instalaciones, ajustándose a las demandas de agua específicas de cada escenario.

2. Objetivos

Los objetivos de la presente tesis se pueden dividir en dos grandes bloques. Por un lado, el desarrollo de estrategias de control jerárquicos para la optimización en tiempo real de la operación en base a diferentes criterios (disminución de costes de operación, maximización de la eficiencia térmica etc.), y por otro, el desarrollo de algoritmos de optimización para el dimensionamiento óptimo de la planta.

En primer lugar, la estrategia de control jerárquico se fundamenta en las técnicas de control predictivo basado en modelo (*Model Predictive Control*, MPC) [15]. Un sistema de control jerárquico se encuentra dividido en varias capas. El nivel más bajo de la estructura estaría formado por el control de las principales variables que intervienen en el proceso, que aplicado al caso de una planta de destilación por membranas sería el control de las temperaturas y caudales de los principales subsistemas de la instalación. El nivel superior lo constituyen algoritmos dedicados a la generación automática de referencias en base a diferentes criterios, como pueden ser la reducción de costes o la eficiencia térmica del proceso en sí. Por tanto, los objetivos que se plantean para el desarrollo de la arquitectura de control jerárquica son:

1. Desarrollo de un modelo completo que permita representar con exactitud el comportamiento de la instalación piloto MD de la PSA. Este modelo será empleado en el proceso de diseño, prueba y validación del sistema de control, pero además servirá como modelo de predicción para la capa superior de la estructura jerárquica.
2. Implementación de un sistema de control directo que sea capaz de mantener la temperatura y el caudal del sistema de generación de energía en unos va-

lores deseados. Para ello se realizará un completo desarrollo de la estructura preliminar presentada en [5,16,17].

3. Desarrollo de modelos de predicción de perturbaciones que ayuden a mejorar la actuación de la capa superior de la estructura jerárquica.
4. Diseño, implementación y prueba de la capa superior basada en control MPC.

En segundo lugar, para el desarrollo del algoritmo de optimización para el diseño óptimo, se plantean los siguientes objetivos:

1. Estudio de las variables de decisión del problema de optimización.
2. Estudio y planteamiento de diferentes casos y escenarios con demanda de agua variable.
3. Diseño, implementación y prueba del algoritmo de optimización.

3. Estado de desarrollo

En este apartado se presenta el estado de desarrollo de la tesis doctoral. Actualmente se han realizado avances y contribuciones en el primero de los objetivos mencionados en el apartado anterior: desarrollo de una estrategia de control jerárquico para la operación óptima de la instalación. El estado de consecución de cada uno de los puntos presentados en el apartado anterior para dicho objetivo se presenta a continuación:

1. *Desarrollo de un modelo completo que permita representar con exactitud el comportamiento de la instalación piloto MD de la PSA. Este modelo será empleado en el proceso de diseño, prueba y validación del sistema de control, pero además servirá como modelo de predicción para la capa superior de la estructura jerárquica.*

Se ha desarrollado un modelo completo de la instalación, utilizando tanto modelos basadas en primeros principios como modelos basados en datos experimentales. Todo lo relacionado con la obtención del modelo y la validación del mismo se presentó en [18].

2. *Implementación de un sistema de control directo que sea capaz de mantener la temperatura y el caudal del sistema de generación de energía en unos valores deseados. Para ello se realizará un completo desarrollo de la estructura preliminar presentada en [5,16,17].*

Se han diseñado cuatro bucles de control para el sistema de generación de energía de la instalación. Los cuatro bucles se gestionan a través de un generador de consignas que calcula las referencias para cada uno de ellos en base a la temperatura deseada a la entrada del módulo. Todo lo relacionado con el diseño, pruebas en simulación y pruebas en la instalación real se presentó en [18].

3. *Desarrollo de modelos de predicción de perturbaciones que ayuden a mejorar la actuación de la capa superior de la estructura jerárquica.*

Siguiendo las ideas presentadas en [19], se han desarrollado modelos de predicción basados en series temporales, los cuales fueron presentados en [20,21].

4. *Diseño, implementación y prueba de la capa superior basada en control MPC.*

En lo respectivo a la estrategia de control jerárquico se han realizado tres contribuciones:

- En primer lugar, en [20] se presentó el desarrollo del algoritmo de control jerárquico, en el cual la capa superior de la estructura de control jerárquica estaba basada en el algoritmo PNMPC (*Practical Nonlinear Model Predictive Control*) [22], mientras que la capa inferior estaba compuesta por el sistema de control directo. De este modo, la capa de control superior calcula las referencias óptimas para el sistema de control directo, tratando de maximizar la eficiencia térmica y la producción de destilado del módulo MD. En este trabajo se presentaron resultados preliminares en simulación, los cuales fueron comparados con los obtenidos mediante una operación manual.
- En segundo lugar, en [21] se presentaron los resultados obtenidos mediante la aplicación del algoritmo de control jerárquico en la instalación real, planteando dos funciones objetivo, una dedicada a optimizar la eficiencia térmica y maximizar la producción de destilado, y la otra dedicada a minimizar los costes de operación. Los resultados obtenidos se compararon con una operación sin la estrategia de control jerárquica, con el fin de evidenciar los beneficios obtenidos en términos económicos, de eficiencia energética y de costes de operación. Además, se plantearon dos casos potenciales de aplicación, presentando los ahorros económicos y energéticos que se pueden llegar a conseguir con la aplicación de la técnica de control jerárquica.
- Por último, en [23] se planteó un caso de estudio en simulación de una planta formada por una instalación SMD y un invernadero. En este caso, la estrategia de control jerárquico se utilizó para la gestión óptima de la planta, utilizando una función objetivo que trata de maximizar la eficiencia energética de la operación al mismo tiempo que se satisface la demanda de agua del invernadero, objetivos que requieren condiciones de operación contrapuestas.

4. Conclusiones

En la presente tesis doctoral se pretende realizar contribuciones en el campo del control jerárquico y optimización aplicadas a los sistemas MD alimentados con energía solar, con el fin de ayudar al desarrollo comercial de dicha tecnología. Hasta el momento se han realizado contribuciones destinadas a optimizar

la operación de la instalación en tiempo real, con objetivos como maximizar la producción de destilado y la eficiencia térmica, y minimizar los costes de operación. Los resultados obtenidos muestran como las técnicas de control desarrollados pueden mejorar significativamente la operación diaria de la instalación [18,20,21,23].

En trabajos futuros se seguirán desarrollando técnicas de control para mejorar la operación de la planta en base a métricas diferentes. Además, se desarrollarán nuevos modelos del módulo MD los cuales tengan en cuenta su operación a altas salinidades. Por otra parte, se abordará el desarrollo de algoritmos de optimización para el diseño óptimo de las instalaciones en base a una demanda de agua determinada.

Referencias

1. G. Zaragoza, A. Ruiz-Aguirre, and E. Guillén-Burrieza, "Efficiency in the use of solar thermal energy of small membrane desalination systems for decentralized water production," *Applied Energy*, vol. 130, pp. 491–499, 2014.
2. M. Khayet, "Solar desalination by membrane distillation: Dispersion in energy consumption analysis and water production costs (a review)," *Desalination*, vol. 308, pp. 89–101, 2013.
3. A. Alkhudhiri, N. Darwish, and N. Hilal, "Membrane distillation: a comprehensive review," *Desalination*, vol. 287, pp. 2–18, 2012.
4. A. Ruiz-Aguirre, J. Andres-Manas, J. Fernández-Sevilla, and G. Zaragoza, "Modeling and optimization of a commercial permeate gap spiral wound membrane distillation module for seawater desalination," *Desalination*, vol. 419, pp. 160–168, 2017.
5. J. D. Gil, A. Ruiz-Aguirre, L. Roca, G. Zaragoza, and M. Berenguel, "Solar membrane distillation: A control perspective," in *23th Mediterranean Conference on Control and Automation (MED 2015)*. Torremolinos, Málaga, Spain, 2015, pp. 796–802.
6. H. Chang, G.-B. Wang, Y.-H. Chen, C.-C. Li, and C.-L. Chang, "Modeling and optimization of a solar driven membrane distillation desalination system," *Renewable Energy*, vol. 35, no. 12, pp. 2714–2722, 2010.
7. H. Chang, S.-G. Lyu, C.-M. Tsai, Y.-H. Chen, T.-W. Cheng, and Y.-H. Chou, "Experimental and simulation study of a solar thermal driven membrane distillation desalination process," *Desalination*, vol. 286, pp. 400–411, 2012.
8. J.-S. Lin, H. Chang, and G. B. Wang, "Modelling and control of the solar powered membrane distillation system," in *AIChE Annual Meeting*. Minneapolis, MN, USA, 2011.
9. F. Eleiwi, I. N'Doye, and T.-M. Laleg-Kirati, "Feedback control for distributed heat transfer mechanisms in direct-contact membrane distillation system," in *2015 IEEE Conference on Control Applications (CCA)*. Sydney, Australia, 2015, pp. 1624–1629.
10. A. M. Karam and T. M. Laleg-Kirati, "Real time optimization of solar powered direct contact membrane distillation based on multivariable extremum seeking," in *Control Applications (CCA), 2015 IEEE Conference on*. IEEE, 2015, pp. 1618–1623.

Estrategias de control jerárquico y optimización aplicadas a plantas de destilación...

11. R. Porrazzo, A. Cipollina, M. Galluzzo, and G. Micale, "A neural network-based optimizing control system for a seawater-desalination solar-powered membrane distillation unit," *Computers & Chemical Engineering*, vol. 54, pp. 79–96, 2013.
12. M. R. Qtaishat and F. Banat, "Desalination by solar powered membrane distillation systems," *Desalination*, vol. 308, pp. 186–197, 2013.
13. M. Khayet and C. Cojocaru, "Air gap membrane distillation: Desalination, modeling and optimization," *Desalination*, vol. 287, pp. 138–145, 2012.
14. Y.-H. Chen, Y.-W. Li, and H. Chang, "Optimal design and control of solar driven air gap membrane distillation desalination systems," *Applied energy*, vol. 100, pp. 193–204, 2012.
15. E. F. Camacho and C. B. Alba, *Model predictive control*. Springer Science & Business Media, 2013.
16. J. D. Gil, A. Ruiz-Aguirre, L. Roca, G. Zaragoza, M. Berenguel, and J. L. Guzmán, "Control de plantas de destilación por membranas con apoyo de energía solar—parte 1: Esquemas," in *XXXVI Jornadas Automática, Bilbao, España*, 2015.
17. —, "Control de plantas de destilación por membranas con apoyo de energía solar—parte 2: Resultados," in *XXXVI Jornadas Automática, Bilbao, España*, 2015.
18. J. D. Gil, L. Roca, G. Zaragoza, and M. Berenguel, "A feedback control system with reference governor for a solar membrane distillation pilot facility," *Renewable Energy*, vol. 120, pp. 536–549, 2018.
19. A. Pawlowski, J. L. Guzmán, F. Rodríguez, M. Berenguel, and J. E. Normey-Rico, "Predictive control with disturbance forecasting for greenhouse diurnal temperature control," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 1779–1784, 2011.
20. J. D. Gil, L. Roca, A. Ruiz-Aguirre, G. Zaragoza, J. L. Guzmán, and M. Berenguel, "Using a nonlinear model predictive control strategy for the efficient operation of a solar-powered membrane distillation system," in *25th Mediterranean Conference on Control and Automation (MED 2017)*. Valleta, Malta, 2017.
21. J. D. Gil, L. Roca, A. Ruiz-Aguirre, G. Zaragoza, and M. Berenguel, "Optimal operation of a solar membrane distillation pilot plant via nonlinear model predictive control," *Computers & Chemical Engineering*, vol. 109, pp. 151–165, 2018.
22. A. Plucenio, D. Pagano, A. Bruciapaglia, and J. Normey-Rico, "A practical approach to predictive control for nonlinear processes," *IFAC Proceedings Volumes*, vol. 40, no. 12, pp. 210–215, 2007.
23. J. D. Gil, L. Roca, M. Berenguel, A. Ruiz, G. Zaragoza, and A. Gimenez, "Control predictivo para la operación eficiente de una planta formada por un sistema de desalación solar y un invernadero," in *XXXVIII Jornadas Automática, Gijón, España*, 2017.

Eficiencia y seguridad en protocolos de streaming peer-to-peer

Cristóbal Medina-López

Universidad de Almería, Sacramento S/N, Almería, España,
cristobalmedina@ual.es

Resumen Existen dos grandes retos que deben afrontarse para conseguir que la tecnología P2P se aplique con éxito a la transmisión de contenido en directo a través de Internet: (1) la simplificación y adecuación de los protocolos de streaming P2P a dispositivos de bajos recursos usados en la electrónica de consumo de los hogares. (2) maximizar la seguridad y privacidad del contenido que viaja por este tipo de redes. En este trabajo se profundiza en ambos problemas y se resumen las técnicas aplicadas que suponen un avance en el estado del arte.

Keywords: Redes P2P, Streaming, WebRTC, Pollution Attacks, Multimedia

1. Introducción

En la actualidad, los sistemas de transmisión de contenido multimedia se usan de forma masiva. Según un estudio realizado por Cisco System, más de la mitad de los datos transmitidos a través de Internet están relacionados con contenido de audio o vídeo [7]. Esto ha motivado que aparezcan multitud de arquitecturas de streaming que básicamente tratan de transmitir varias señales multimedia a los usuarios, o sistemas finales, con la máxima calidad posible, es decir, baja latencia, baja pérdida de datos, alta interactividad, alta resolución, y garantías de seguridad.

La solución para streaming de vídeo en directo más usada hoy en día es el modelo C/S (Cliente/Servidor), que se basa en la idea de que el stream es generado en un servidor y, bajo demanda (pull model) o de forma automática (push model), el stream es transmitido tratando de minimizar la latencia al conjunto de clientes. Sin embargo, los sistemas puros C/S están limitados por el ancho de banda del servidor. Por este motivo, la mayoría de los portales de streaming C/S utilizan una red de entrega de contenidos (Content Delivery Network, CDN, en inglés) en la que el contenido multimedia se encuentra replicado en un conjunto de servidores [5]. Podría afirmarse que dicha arquitectura es la que actualmente tiene más éxito ya que es usada por los sistemas de distribución multimedia más populares. Sin embargo, la arquitectura P2P es más escalable debido a que el proveedor de contenido (*source*) solo envía una única copia del stream o un pequeño número de copias. Por su parte, los peers dedican su ancho de banda

de subida a reenviar el contenido tantas veces como sea necesario. Además, esto convierte a las soluciones P2P en la mejor opción si se busca privacidad, ya que el contenido no tiene que atravesar servidores de terceras partes sino que se transmite directamente entre clientes.

En este contexto, se desarrolló el protocolo P2PSP¹. Un protocolo de la capa de aplicación diseñado para la transmisión de contenido en tiempo real en Internet sobre una red P2P. Una red P2PSP básica consiste en un Splitter (S) y un conjunto de peers (P_i) denominado team. El splitter recibe el contenido multimedia desde el source(O), divide el stream en trozos de datos del mismo tamaño (chunks) y los envía al team siguiendo un esquema Round-Robin. Cada peer se encarga de reenviar los chunks recibidos desde el splitter al resto de peers (los chunks recibidos desde otros peers no son reenviados). Finalmente, cada peer reconstruye el stream y lo envía al reproductor (L). Cuando el splitter ha enviado un chunk a cada uno de los peers, se dice que ha completado un round.

Debido a que los peers obtienen la mayoría del contenido de otros peers y no desde el splitter, es fácil para un peer malicioso (MP) cambiar el contenido. Este ataque se denomina de contaminación (pollution attack). Existen diferentes formas de realizar un ataque de contaminación del contenido: ataques persistentes, cuando el atacante siempre envenena los chunks de sus víctimas, o ataques on-off, que ocurren cuando el atacante solo envenena algunos chunks (por ejemplo el 10%) con la intención de engañar al posible sistema de reputación existente. Los peers que no reenvían o free-riders, afectan de forma similar a nuestro sistema, por lo que son tratados del mismo modo.

Los ataques anteriores pueden ser modificados incluyendo el modo selectivo y el colaborativo. En un ataque selectivo, se seleccionan los peers a los que enviar los chunks envenenados. En un ataque colaborativo, varios atacantes colaboran con el objetivo de incrementar el daño del ataque y evitar ser descubiertos. En presencia de peers policía (TPs), los atacantes podrían colaborar para descubrirlos. En este estudio, proponemos una estrategia basada en la técnica de compartición de secretos de Shamir para forzar a todos los peers a colaborar con parte del resto del team.

En cuanto a los dispositivos, hoy en día la mayoría de los dispositivos que se usan para consumir contenido multimedia son móviles. Sin embargo, en ocasiones, el usuario experimenta una baja calidad de experiencia (*QoE*) debido a el tamaño reducido de las pantallas, especialmente si hablamos de contenido de alta definición. Como una solución a este problema, Google lanzó al mercado *Google Chromecast* (GC). Un dispositivo que se conecta al televisor y permite reproducir un stream desde un dispositivo auxiliar (un smartphone, tablet o PC). Tiene dos modos de funcionamiento, ambos utilizando el modelo C/S: (1) recibe el contenido directamente del dispositivo auxiliar, y (2) descarga el contenido desde Internet siguiendo las indicaciones del dispositivo auxiliar. En este trabajo, se pretende mostrar como GC es capaz de ejecutar el protocolo P2P gracias al navegador Web que incluye.

¹ p2psp.org

2. Avances

2.1. Streaming P2P en dispositivos con pocos recursos

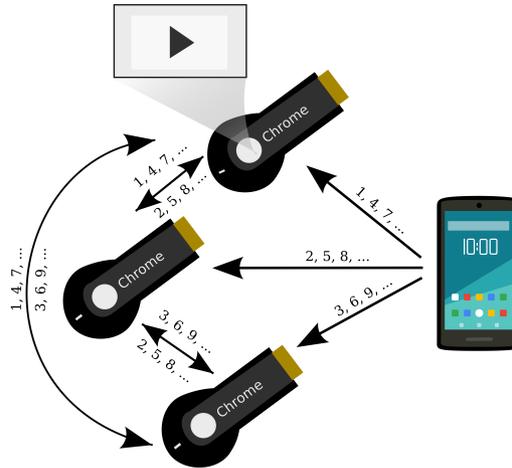


Figura 1: Streaming desde un smartphone usando un enfoque P2P sin servidores multimedia intermedios. La fuente envía una única copia del vídeo. Los dispositivos Chromecast (peers) son los encargados de compartirse el vídeo entre ellos.

Uno de los factores más interesantes en los smartphones y en dispositivos de recursos limitados como GC, es la inclusión del navegador Web Google Chrome en su conjunto de aplicaciones. La constante evolución de los navegadores Web hacen a estos dispositivos capaces de ejecutar códigos más complejos que un simple renderizado HTML. Gracias a la evolución de HTML y de sus APIs, existen numerosas aplicaciones Web que se asemejan a aplicaciones de escritorio, como pueden ser los reproductores multimedia, el software de finanzas o las herramientas de modelado. De hecho, las aplicaciones Web se están volviendo cada vez más populares que las de escritorio. Algunas de sus ventajas son: que son multiplataforma, no necesitan instalación, y son fáciles de usar. No obstante, aún existen limitaciones que impiden que todas las aplicaciones puedan migrar al navegador, como es el caso de las multihebradas.

Nuestro objetivo es implementar el protocolo P2PSP para que pueda ser ejecutado en estos dispositivos (ver Figura 1), medir su rendimiento y mostrar las ventajas de la arquitectura P2P frente a la arquitectura C/S.

Implementación

Uno de los elementos básicos y necesarios para llevar a cabo una implementación

del P2PSP es la entidad *splitter*, ya que será la encargada de recibir el stream, dividirlo en chunks del mismo tamaño y enviarlo al team. Además, el *splitter* tiene la función de presentar cada peer entrante al resto del team. Esta última capacidad es similar a la que debe llevar a cabo el Signaling Server en WebRTC durante el proceso de señalización. Los dos procesos son obligatorios, el primero, para cumplir la especificación del protocolo P2PSP y el segundo, para iniciar una *PeerConnection* en WebRTC. Por tanto, parece lógico que el mismo dispositivo realice ambas funciones. Para realizar una comunicación bidireccional entre el navegador y el Signaling Server se puede usar *WebSockets*, el cual se encuentra disponible en *HTML5*. Normalmente, los navegadores que soportan WebRTC también soportan *WebSocket*. Usar *JSON* (un formato ligero para el intercambio de datos) junto a *WebSocket* es una buena opción para realizar el intercambio del objeto *session description* del protocolo *SDP* en WebRTC.

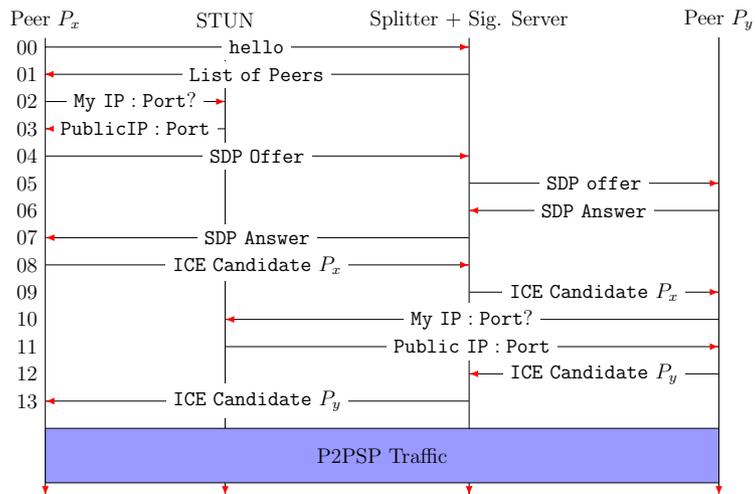


Figura 2: Secuencia de intercambio de mensajes a la llegada de un peer a un team P2PSP.

La Figura 2 muestra el proceso completo desde que un peer entra en el team hasta que comienza a recibir chunks. En primer lugar se procede siguiendo el protocolo P2PSP, es decir, el peer que quiere entrar en el team (*A* en este caso) envía un mensaje *Hello* al *splitter*, que le contesta enviándole la lista de peers y la cabecera del vídeo. A continuación se inicia el proceso de señalización. Para ello, el peer *A* pregunta al servidor *STUN* cuál es su *IP_publica:Puerto* y *STUN* responde con la información solicitada. Entonces, *A* envía un mensaje de tipo *SDP Offer* al que *B* tiene que dar respuesta con uno de tipo *SDP Answer* para concluir la descripción de las sesiones de comunicación multimedia. Además, *A* y *B* comparten las *IPs* por las que podrían ser alcanzados usando mensajes del tipo *ICE candidate*. Hasta este momento, *A* y *B* se han comunicado a través

del splitter+signaling server. Una vez realizados estos intercambios de mensajes, los peers ya pueden realizar una comunicación directa entre ellos, y por tanto, pueden intercambiar los chunks.

Este proceso es posible si A no se encuentra detrás de un NAT simétrico. En caso contrario, todas las comunicaciones pasarían a través de un servidor TURN.

Con respecto a la parte del cliente, el usuario sólo tiene que lanzar una instancia del peer P2PSP. El peer está compuesto por código JavaScript, HTML5 y WebRTC y se ejecuta en el navegador Chrome del Google Chromecast. Esto permite disfrutar del contenido multimedia directamente en la televisión, gracias a que el navegador ya dispone de un reproductor multimedia. Cuando el buffer del peer P2PSP se ha llenado a la mitad de su capacidad, se comienza a enviar su contenido al buffer del reproductor que se encuentra en el objeto *MediaSource*.

Para comunicar peers en distintos GC se usa WebRTC, ya que permite crear un canal de comunicación directo entre los distintos navegadores (peers) usando un protocolo de transporte ligero conocido como SCTP (Stream Control Transmission Protocol).

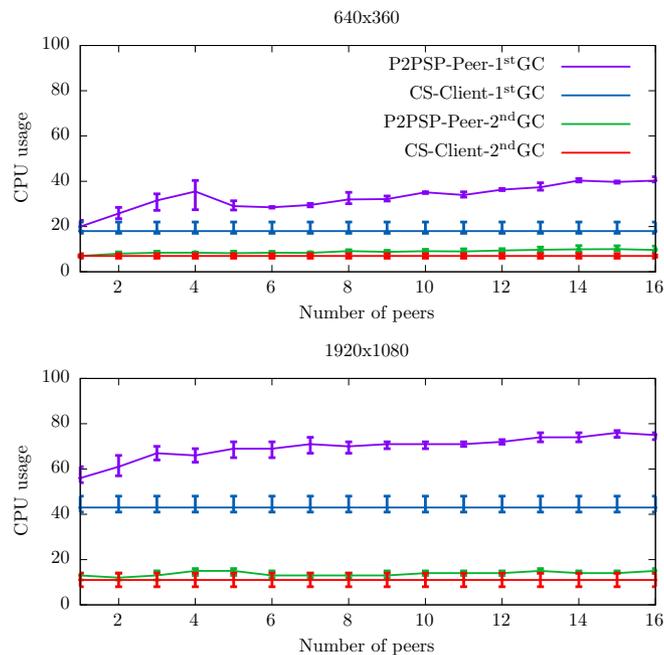


Figura 3: Comparación del uso de CPU entre un peer P2PSP y un cliente CS ejecutándose en la primera y segunda generación de GC.

Experimentación

Hemos llevado a cabo un conjunto de experimentos para comparar los recursos computacionales requeridos para ejecutar un peer P2PSP y un cliente en dispositivos GC. El peer y el cliente en GC reproducen el mismo vídeo tras solicitarlo vía WebRTC y WebSocket respectivamente. Ambos envían los chunks al reproductor del mismo modo (etiqueta de vídeo HTML + Media Source Extensions).

La Figura 3 muestra el uso de CPU para un peer y un cliente ejecutándose en las dos primeras generaciones de GC transmitiendo dos vídeos con distintas resoluciones (640x360 y 1920x1080). Para la primera generación de GC los peers muestran un uso de CPU superior al de los clientes debido a que deben reenviar chunks en lugar de sólo recibirlos. Las diferencias son mínimas en la segunda generación de GC.

Puede encontrar más información sobre este estudio en el artículo *Media streaming on resource-constrained devices without third-party servers: A case study on a spartphone and Chromecast using P2PSP* [3], donde se estudia detalladamente la implementación y se analizan los resultados experimentales obtenidos.

2.2. Estrategia de defensa basada en Shamir's Secret Sharing

Las redes completamente conectadas, como P2PSP, son una alternativa eficiente a las redes en forma de árbol. En el P2PSP, como en muchos otros sistemas P2P, los peers maliciosos (MPs) pueden llevar a cabo ataques colaborativos para contaminar el stream. En este trabajo proponemos usar el sistema de compartición de secretos de Shamir y la monitorización usando peers de confianza (TPs) para forzar a los peers a reenviar el contenido sin contaminar a parte del team con el objetivo de no ser expulsados del mismo.

Propuesta

La idea principal es muy simple “si quieres mantenerte en el team debes tener un buen comportamiento”, sin embargo, es muy difícil para el splitter determinar quién está teniendo un buen comportamiento con los demás peers. Por tanto, relajamos este requisito añadiendo la siguiente frase “con al menos t peers”.

El sistema de compartición de secretos de Shamir [6] se ha usado anteriormente en redes P2P con un objetivo diferente al nuestro [1]. Principalmente se ha usado para mantener el contenido privado y seguro. En lugar de eso, nosotros lo vamos a usar para forzar la colaboración entre peers.

La función de Shamir $SSS(t, n)$ genera un secreto k y un conjunto n de shares. Un peer que tiene t de n shares puede derivar k . La probabilidad de dos funciones SSS que produzcan un mismo secreto debería ser muy baja.

La idea es cifrar la firma de cada mensaje que contiene un chunk de S a P_i con una clave k_i . Los mensajes reenviados entre los peer deben tener una firma válida no cifrada. Por tanto, el peer que reenvía el mensaje tiene que tener al menos t shares para derivar k_i en la ronda anterior. A continuación se muestra como los shares para cada k_i son repartidos por S a los peers. Definimos un round r como:

- Para S : Las n iteraciones necesarios para enviar un mensaje a cada peer P_i , $i = 1, \dots, n$.
- Para un peer: Las $|L|$ iteraciones para reenviar un mensaje recibido de S . $|L| < n$ es posible porque un peer puede eliminar a los MPs de su lista de peers.

En un round r , un peer P_q reenviará un mensaje válido, y junto a el, el share necesario para el peer P_i para derivar k_i^{r+1} solo si:

- P_q puede obtener al menos t_{r-1} shares en el round $r - 1$ de otros peers para descifrar la firma (derivando k_q^r) y
- P_i siempre envía un mensaje válido a P_q .

Por definición,

$$n_r = \#MP_r + \#WIP_r. \quad (1)$$

EL número de TPs ($\#TPs$) está incluido en el número de peer bienintencionados en un round r ($\#WIP_r$).

Bajo las siguientes condiciones:

$$\#MP_r < t_r \leq \#WIP_r, \quad (2)$$

- Cualquier peer que se comporte bien puede obtener al menos t_r shares de los WIPs.
- Los MPs no son capaces de derivar la clave por ellos mismos. Incluso si los MPs se comunican los shares recibidos desde S entre ellos, cada MP aún necesitará al menos, $t_r - \#MP_r$ shares. Por tanto, cada MP está forzado a enviar un mensaje válido a un número $t_r - \#MP_r$ de WIPs. Por otra parte, si no envían un mensaje válido, los TPs se quejarán al S , y serán expulsados del team.

El mejor valor para t_r es $t_r = \#WIP_r$, pero este valor es difícil de conocer. De las ecuaciones (1) y (2) podemos deducir:

$$\#MP_r < \left\lceil \frac{n_r}{2} \right\rceil. \quad (3)$$

Por tanto, para un sistema que satisfaga la ecuación (3),

$$\#MP_r < \left\lceil \frac{n_r}{2} \right\rceil \leq t_r \leq \#WIP_r. \quad (4)$$

De ahora en adelante, se usará $t_r = \left\lceil \frac{n_r}{2} \right\rceil$.

Ahora veremos cómo se generan los shares, se envían a los peer desde el splitter y se comparten en el team.

Al comienzo de una ronda r , todos los MPs expulsados en la ronda $r - 1$, los peers que han solicitado unirse al team en la ronda $r - 1$, y los WIPs peers que han solicitado abandonar el team en la ronda $r - 1$ son conocidos. Así pues, n_r es un valor conocido. Conociendo esta información, S genera un vector $K_{(n_r)}^{r+1}$ de

n_r claves y una matriz $A_{(n_r \times n_r)}^{r+1}$ de shares haciendo n_r llamadas a $\text{SSS}(t_r, n_r)$. Cada llamada da como resultado K_i^{r+1} y una fila $A_{i,*}^{r+1}$ con n_r shares útiles para derivar K_i^{r+1} .

En el round r , S envía al peer P_i un mensaje $eSP_{j,i}^r$

$$eSP_{j,i}^r = \{eCH_{j,i}^r, SH_i^{r+1}\}. \quad (5)$$

$$eCH_{j,i}^r = \{C_j, j, P_i, r, E_{K_i^r}[S_{pr}(H(C_j||j||P_i||r))]\} \quad (6)$$

$eCH_{j,i}^r$ tiene el siguiente contenido:

- El chunk C_j .
- El índice j del chunk en el stream.
- El peer P_i encargado de reenviar C_j .
- El round r .
- La firma de todos los campos anteriores cifrada con la clave secreta de Shamir para P_i en el round r (K_i^r).

La segunda parte de $eSP_{j,i}^r$ (ver ecuación (5)) es:

$$SH_i^{r+1} = \{\{SH_i^{r+1}\}_q, q = 1, \dots, n_r\}, \quad (7)$$

con

$$\{SH_i^{r+1}\}_q = \{P_i, P_q, r + 1, A_{q,i}^{r+1}, S_{pr}(H(P_i||P_q||r + 1||A_{q,i}^{r+1}))\} \quad (8)$$

Tenemos que, SH_i^{r+1} es un vector con n_r elementos. Cada elemento $\{SH_i^{r+1}\}_q$ contiene:

- El peer P_i que tendrá que enviar el share $A_{q,i}^{r+1}$ al peer P_q .
- El peer P_q que usará el share anterior para derivar K_q^{r+1} .
- El round ($r + 1$) para el que el share es creado.
- El share.
- La firma de los campos anteriores usando la clave privada del Splitter, lo cual permitirá a P_q validar el contenido.

Una vez que P_i ha recibido $eSP_{j,i}^r$ de S , para que P_i no sea eliminado de la lista de peers de P_q , tiene que enviar a P_q un mensaje válido.

$$PP_{i,q}^r = \{CH_{j,i}^r, \{SH_i^{r+1}\}_q\}. \quad (9)$$

Donde $CH_{j,i}^r$ es igual a $eCH_{j,i}^r$ pero con la firma del Splitter sin cifrar con K_i^r :

$$CH_{j,i}^r = \{C_j, j, P_i, r, S_{pr}(H(C_j||j||P_i||r))\} \quad (10)$$

De este modo, P_q puede validar el contenido de ambos $CH_{j,i}^r$ y $\{SH_i^{r+1}\}_q$ tras la recepción de $PP_{i,q}^r$ desde P_i .

Sin embargo, para conseguir una firma válida que convierta $eCH_{j,i}^r$ en $CH_{j,i}^r$, P_i tiene que enviar al menos t_{r-1} mensajes $PP_{i,q}^{r-1}$ válidos a P_q , $q = 1, \dots, t_{r-1}$ en el round $r - 1$ para conseguir al menos t_{r-1} $PP_{q,i}^{r-1}$ shares que permitan derivar K_i^r (y t_{r-1} chunks no contaminados). Este proceso fuerza a los peers a compartir sus chunks.

La figura 4 muestra varios ejemplos de ataques. El ataque 4a es el más agresivo, pero es el ataque donde más rápido se expulsan a los MPs. Expulsar a los MPs realizando los ataques 4b y 4c es más difícil, pero el impacto del ataque es mucho menor. En estos casos, establecer $t_{r+1} = \#WIPs_{r+1}$ no permite expulsar a los maliciosos, pero es un ataque lento que carece de interés si el objetivo es destruir el team rápidamente.

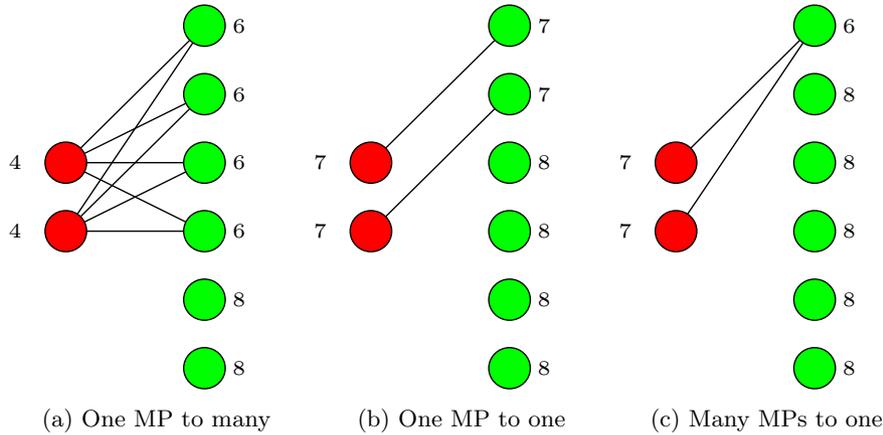


Figura 4: Tres ejemplos de posibles ataques de contaminación. Los círculos verdes son los WIPs y los rojos los MPs. Las líneas muestran los ataques y el número junto a los nodos el número de shares conseguido.

Sería interesante definir una heurística que expulse a los peer con una probabilidad baja si son WIPs basándose en el comportamiento del team durante toda la transmisión.

Puede encontrar más información sobre este estudio en los artículos *On pollution attacks in fully connected P2P networks using trusted peers* [4] y *On Mitigating Pollution and Free-Riding Attacks by Shamir's Secret Sharing in Fully Connected P2P Systems* [2] donde se analizan esta y otras estrategias para afrontar los ataques de contaminación del stream.

3. Conclusiones

Por un lado, se ha mostrado que es posible la implementación del protocolo P2PSP en dispositivos de bajos recursos como Google Chromecast en términos

de uso de CPU. Pero no solo en GC, sino que las tecnologías Web permiten desarrollar una red P2PSP heterogénea donde pueden interaccionar distintos tipos de dispositivos para compartir vídeo en tiempo real, usando como elemento común el navegador Web. Por otro lado, se ha propuesto un nuevo mecanismo de defensa para forzar a los peers a colaborar con el resto del team basado en el algoritmo de compartición de secretos de Shamir y el uso de Trusted Peers. Tras analizar la propuesta, concluimos que los ataques más agresivos pueden ser completamente mitigados. Para los demás ataques, se podría incrementar el número de TPS. Complementariamente, establecer de forma dinámica el valor t de Shamir podría mejorar la efectividad del mismo.

Referencias

1. Jinsong Han, Yanmin Zhu, Yunhao Liu, Jianfeng Cai, and Lei Hu. Provide privacy for mobile p2p systems. In *Distributed Computing Systems Workshops, 2005. 25th IEEE International Conference on*, pages 829–834. IEEE, 2005.
2. Cristóbal Medina-López, Vicente González-Ruiz, and L. G. Casado. On Mitigating Pollution and Free-Riding Attacks by Shamir’s Secret Sharing in Fully Connected P2P Systems. In *International Wireless Communications and Mobile Computing Conference*, pages 711–716, Valencia, Spain, 2017. IEEE.
3. Cristóbal Medina-López, Max Bastian Mertens, Vicente González Ruiz, and L.G. Casado. Media streaming on resource-constrained devices without third-party servers: A case study on a smartphone and Chromecast using P2PSP. *IEEE Consumer Electronics Magazine*, 2018.
4. Cristóbal Medina-López, Ilshat Shakirov, L.G. Casado, and Vicente González-Ruiz. On pollution attacks in fully connected P2P networks using trusted peers. In *Intelligent Systems Design and Applications*, Porto, 2017. Springer International Publishing.
5. George Pallis and Athena Vakali. Insight and perspectives for content delivery networks. *Communications of the ACM*, 49(1):101–106, 2006.
6. Adi Shamir. How to share a secret. *Commun. ACM*, 22(11):612–613, November 1979.
7. Cisco System. Cisco visual networking index: global mobile data traffic forecast update, 2014–2019. *Cisco Public Information*, 2015.

Un Modelo Inteligente de Interacción Natural Autoadaptativo basado en Visión Artificial

Juan Jesús Ojeda Castelo¹

Applied Computing Group (TIC-211), Departamento de Informática, Universidad de Almería¹

juanje.ojeda@ual.es

Abstract. La interacción natural ha sido integrada en varias áreas de conocimiento pero especialmente como una técnica de experimentación que no ha conseguido afianzarse hasta ahora. Los métodos tradicionales de interacción siguen siendo el medio más popular para controlar un sistema informático. Esta situación se debe a que principalmente no se ha conseguido que este tipo de interacción sea totalmente fiable y ofrezca un tiempo de respuesta óptimo. En esta Tesis se tiene el objetivo principal de investigar sobre un modelo de interacción natural que sea portable, inteligente, adaptable y low cost. En este informe se describe el procedimiento que se ha seguido hasta este momento para alcanzar dicho objetivo en un futuro próximo.

1 Introducción

En los últimos años las Tecnologías Informáticas (TIN) han ido avanzando de manera rápida y constante, tanto a nivel hardware como software. Sin embargo, hay una disciplina importante dentro de las TIN que se ha desarrollado más lentamente que las demás: la Interacción del usuario. Desde que en el año 1967, el ingeniero Douglas Engelbart inventó el ratón [1], se ha estado utilizando este periférico junto con el teclado para interactuar con la Interfaz Gráfica de Usuario (GUI, Graphical User Interface) de los ordenadores personales y sistemas informáticos. Actualmente, este vacío se enmarca como reto dentro de la disciplina de la Informática conocida como Interacción Hombre-Máquina (más conocido por su término en inglés como HCI, Human Computer-Interaction). A su vez, la interacción del usuario ha sido motivo de gran interés en la literatura HCI llegando a crear una comunidad científica específica en el ámbito de la Interacción Natural del usuario, o también conocido como Natural User Interaction (NUI) [2, 3].

La Interacción Natural del usuario (NUI) se refiere a la comunicación que se realiza entre el usuario y el sistema informático, mediante acciones que el usuario realiza habitualmente con alguna parte de su cuerpo. Esta interacción natural debe conocer las capacidades del individuo, sus intereses y tenerlas en cuenta para ir un paso por delante del usuario o de lo contrario nunca será realmente natural, sino sólo una aproximación de lo que se entiende como interacción natural [4]. Esta interacción natural se divide principalmente en cuatro

Un Modelo Inteligente de Interacción Natural Autoadaptativo basado en Visión Artificial

tipos: interacción táctil (touch-based), reconocimiento de gestos, reconocimiento de voz e interfaz cerebro-computadora (brain-computer interface).

La interacción táctil se basa en el contacto entre el usuario y el dispositivo, principalmente con los dedos del mismo [5]. Esta interacción es la más conocida por los usuarios, debido a que en la actualidad los dispositivos móviles se basan en este tipo de interacción. Sin embargo, esta interacción se ha utilizado anteriormente con las pantallas táctiles y PDAs [6]. Este tipo de interacción no sólo consiste en tocar iconos o trabajar con la interfaz, sino que permite reconocer gestos como los que se utilizan para ampliar una fotografía en un dispositivo móvil [7].

El reconocimiento de gestos consiste esencialmente en validar una serie de patrones de movimiento [8]. Estos movimientos se pueden realizar con el cuerpo [9], las manos [10] o incluso la cara [11]. Los gestos se clasifican en estáticos y dinámicos [12]. Los gestos estáticos son aquellos en los que su posición no varía respecto al tiempo, mientras que en los dinámicos el tiempo es una variable fundamental para determinar si un gesto es reconocido.

El reconocimiento de voz tiene el objetivo de identificar la voz del usuario y validar las palabras que pronuncia [13]. Este método se utiliza para biometría con el fin de reconocer ciertas características de la voz como el timbre o el género y poder autenticarte en un sistema de una forma más segura. En la actualidad, el reconocimiento de voz está muy presente en la vida cotidiana de las personas, debido a que las compañías han incorporado este sistema para controlar acciones del sistema operativo. Los ejemplos más conocidos son Ok Google de Google, Siri de Apple y Cortana de Microsoft.

2 Objetivos

El objetivo principal es la realización de un sistema de interacción natural mediante reconocimiento de gestos con determinadas partes del cuerpo. Este sistema se caracterizará por ser **low cost, portable, inteligente y adaptable**. En el proceso de reconocimiento se utilizará una cámara estándar para el análisis y procesamiento de las imágenes.

Los objetivos específicos necesarios para cumplir el propósito de esta tesis son:

- **Realizar reconocimiento con motricidad fina:** En primer lugar se va a realizar un reconocimiento de gestos con motricidad fina, donde el objetivo es detectar y reconocer gestos con las manos.
 - Segmentar manos: La detección de las manos requiere de un proceso de segmentación inicial para distinguir el objetivo de reconocimiento (las manos) del fondo de la imagen obtenida.
 - Extraer características: La longitud de los dedos o el contorno de la mano pueden ser características a tener en cuenta para realizar un reconocimiento efectivo. Decidir qué características convienen más en el proceso de reconocimiento es un paso importante y decisivo para el resto del proceso.

Un Modelo Inteligente de Interacción Natural Autoadaptativo basado en Visión Artificial

- Realizar un seguimiento de la mano (tracking): En el proceso de reconocimiento es necesario saber la posición de las manos en todo momento para identificar cada uno de los movimientos que el usuario está realizando en cada instante de tiempo. En este paso será necesario utilizar uno de los algoritmos existentes o diseñar uno nuevo.
- Reconocer gestos con las manos: El reconocimiento de gestos puede ser estático o dinámico. El tipo de reconocimiento de gestos que se va a desarrollar en esta propuesta va a ser dinámico, con lo cual el tiempo será un factor relevante en el proceso de reconocimiento. En esta etapa es importante decidir qué tipo de técnica de Inteligencia Artificial se va a utilizar para reconocer los gestos, ya que habrá métodos que se ajusten más a las necesidades del sistema. Se ha optado por implementar uno de estos métodos porque se necesita un reconocimiento de gestos que sea fluido y preciso para que la interacción del usuario sea natural.
- **Crear nuevos gestos:** En los procesos de reconocimiento que se han comentado anteriormente, el sistema inicialmente reconocerá una serie de gestos predefinidos. La creación de nuevos gestos definidos por el usuario hará que el sistema sea personalizado para éste y los gestos se adaptarán mejor a las características del individuo.
 - Obtener las características del gesto: Es necesario definir qué características se van a requerir para definir un nuevo gesto.
 - Almacenar las características del gesto: Los valores relevantes que determinarán un gesto y será comprobado por el proceso de reconocimiento de gestos para validar un gesto, tienen que ser almacenados en una base de datos.
 - Analizar gestos similares: Cuando el usuario crea un nuevo gesto existe la posibilidad de que el gesto tenga muchas similitudes con un gesto que ya haya almacenado. Esta comprobación es imprescindible porque es más fácil añadir las nuevas características de un gesto a uno ya existente que crear uno nuevo, aumentando de este modo la eficiencia del procedimiento.
 - Entrenar el sistema: El usuario tendrá que repetir el gesto más de una vez para entrenar el sistema y que el reconocimiento de este nuevo gesto sea fiable y preciso.
- **Diseñar un sistema adaptable:** Este proceso tiene que ser flexible para que pueda ser utilizado por el mayor número de usuarios y a su vez que el reconocimiento de gestos vaya mejorando con los distintos cambios que presentará el sistema a lo largo del tiempo.
 - Obtener los datos del usuario: Se tendrá que decidir los datos del usuario que son relevantes para el proceso de reconocimiento de gestos y definir el método para obtenerlos.
 - Elaborar un modelo de usuario: La realización de un sistema adaptable requiere de un modelo de usuario que estará formado por las características que han sido definidas en el paso previo.
 - Aprendizaje del sistema: Se integrarán algoritmos de Machine Learning para que el sistema aprenda de la interacción de los usuarios con el sistema.

Un Modelo Inteligente de Interacción Natural Autoadaptativo basado en Visión Artificial

- **Diseñar un sistema portable:** Se propone crear un sistema que permita trasladar el reconocimiento de gestos a diferentes escenarios como por ejemplo, una casa inteligente, un coche, una Raspberry Pi, etc.
 - Crear un módulo portable: Se requiere la creación de un módulo para que sea fácil utilizar este procedimiento en cualquier ambiente que sea válido para incluir un proceso de reconocimiento gestual.
 - Asociar los gestos a una acción del sistema: Cada uno de los ámbitos donde se va a utilizar este sistema tendrá su propia metodología para llevar a cabo una acción en el propio sistema, por eso es debido encontrar una forma de conectar la respuesta final del proceso de validación de un gesto con la comunicación del sistema residente, que sea universal.
- **Evaluación del sistema:** La resolución que determinará si el sistema es válido y ha cumplido las expectativas que se plasmaron en un principio en el proyecto de tesis, será llevada a cabo mediante un proceso de evaluación con distintos perfiles de usuario.
 - Definir el método de evaluación: En la literatura hay diversos métodos de evaluación respecto a sistemas interactivos y genéricos con la participación de usuarios. En primera instancia se tendrá que revisar cada uno de los métodos y elegir el más adecuado según las características de este estudio.
 - Probar el prototipo con usuarios reales: El prototipo desarrollado será probado con usuarios finales utilizando el método de evaluación considerado en el paso anterior.
 - Analizar los resultados obtenidos: Una vez obtenidos los resultados del proceso de evaluación se procederá a analizarlos con diversas herramientas estadísticas para elaborar una conclusión del proceso de investigación.
 - Obtener unas conclusiones: Con ayuda de los resultados obtenidos en esta evaluación se determinará si la hipótesis se ha cumplido y los datos son favorables para avalar la investigación realizada.

3 Progreso de la Investigación

En primer lugar, se diseñó un proceso de reconocimiento de gestos para detectar gestos estáticos (posturas) con las manos. Este proceso de reconocimiento se componía de tres fases:

1. Eliminación del fondo (Background subtraction)
2. Proceso de calibración
3. Reconocimiento de gestos estáticos

3.1 Eliminación del fondo (Background subtraction)

La eliminación del fondo es un método que se suele utilizar para diferenciar los objetos de primer plano respecto la imagen de fondo. Esta técnica almacena ejemplos de imágenes previas en memoria y genera un modelo de fondo basado en propiedades estadísticas de estas muestras. Además, se ha aplicado un filtro Gaussiano para evitar cambios bruscos en la imagen.

3.2 Proceso de calibración

En esta etapa se utilizó, en primera instancia, el método de Clasificador en Cascada para detectar la cara del usuario. Esta técnica fue integrada en el proceso por dos razones:

1. Detectar la posición inicial de las manos respecto de la cara.
2. Identificar áreas de las imágenes donde no es necesario buscar las manos.

Cuando se ha identificado la cara del usuario se procede a hacer una segmentación a través del color de la piel (skin color detection) y de este modo omitir información que no es relevante en la imagen, como por ejemplo, la ropa del usuario o el pelo. Este método de segmentación funciona correctamente en los experimentos con un valor entre 109 - 133.

A continuación, se procede a realizar la extracción geométrica de características (Geometric Features Extraction). En este proceso se empieza extrayendo los puntos del contorno que forma la mano y se genera una imagen binaria que actúa como máscara, con el fin de obtener solo el perímetro de la mano. Al utilizar Convex Hull surge el inconveniente de que algunos puntos convexos que detecta pueden proceder del brazo, por este motivo se aplica el método convexityDefects para detectar solo los puntos de la mano.

El siguiente paso consiste en hacer un seguimiento (tracking) de la mano para tenerla localizada en todo momento. Para este fin, se aplica un algoritmo de flujo óptico basado en Lucas-Kanade, el cual permite hacer un seguimiento de los puntos de interés entre dos imágenes consecutivas, debido a que sería muy costoso hacerlo de todos los píxeles que componen dichas imágenes. La información procedente de este proceso de seguimiento, permite realizar una estimación de la trayectoria de cada uno de los puntos seleccionados y como consecuencia, preveer el desplazamiento de la imagen.

3.3 Reconocimiento de gestos estáticos

La secuencia de métodos descritos en los apartados anteriores permite que se haga un reconocimiento de los dedos de la mano y sea posible saber los dedos que están extendidos en cualquier momento. El reconocimiento de gestos consiste en identificar si la mano del usuario está cerrada o abierta dependiendo del número de dedos que el proceso ha detectado.

Este sistema de reconocimiento presenta el principal inconveniente de que se necesita un fondo uniforme como una pared blanca para que la detección de la mano se haga correctamente. Además, es preciso un proceso de calibración previo.

Esta metodología ha sido desarrollada porque para ofrecer un medio de interacción natural se necesita un proceso que no sea muy costoso computacionalmente y que ofrezca un tiempo de respuesta lo más bajo posible.

3.4 Algoritmos de Características

El algoritmo de extracción de características es un proceso que utiliza una imagen como entrada para localizar información específica en ella, como por ejemplo el color, la textura o la forma. Después de encontrar la información necesaria, ésta se almacena en un *vector de características*. Después de esta etapa se suele aplicar un método de clasificación para que se pueda determinar la clase a la que pertenece una imagen a partir del vector de características.

Este tipo de algoritmos han sido utilizados con el fin de obtener información sobre la mano que pueda resultar de interés en el reconocimiento de gestos. Existe una gran variedad de algoritmos de extracción que se utilizan para detectar objetos especialmente a partir de su contorno, más concretamente analizando las esquinas. Sin embargo, no hay definido ningún algoritmo específico para la extracción de características de las manos.

Los algoritmos de extracción de características que han sido probados son:

- SIFT
- SURF
- ORB
- FAST & BriefDescriptorExtractor
- FAST & DAISY
- AKAZE & BoostDesc
- AKAZE
- FAST & LUCID
- FREAK
- FAST & FREAK
- MSDDetector
- StarDetector & FREAK
- FAST & VGG
- AKAZE & VGG
- SURF & VGG
- AGAST & SURF
- BRISK
- GFTT & FREAK
- GFTT & SURF
- KAZE
- MSER & SURF
- MSER & FREAK

Los candidatos que se han seleccionado debido a que han detectado un número aceptable de puntos en la mano, así como la localización significativa de los mismo (punta de los dedos, zonas convexa de los dedos, palma de las manos y su contorno), han sido (Ver Figura 1):

- AGAST & SURF
- AKAZE & BoostDesc
- FAST & FREAK

Un Modelo Inteligente de Interacción Natural Autoadaptativo basado en Visión Artificial

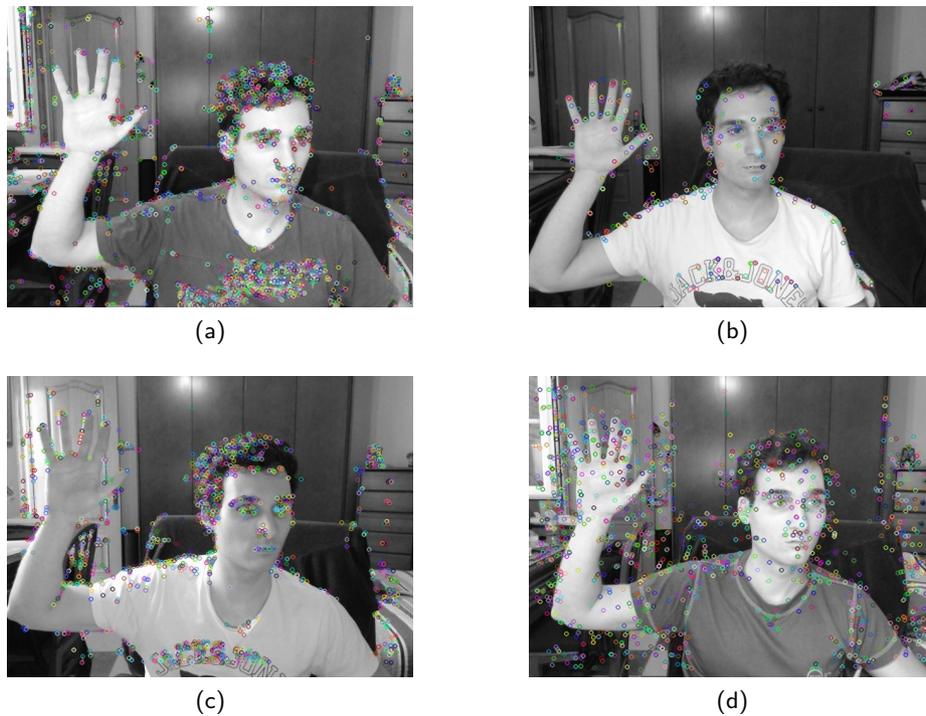


Fig. 1. Candidatos de los Algoritmos de Extracción

– SURF

Sin embargo, estos algoritmos no solo detectan características en la mano del usuario, sino que también detectan otros puntos de interés en el resto de la imagen. Por esta razón, se le aplicó una técnica de eliminación de fondo para que los algoritmos se centraran exclusivamente en el usuario (Ver Figura 2).

3.5 Clasificador en Cascada basado en características Haar

Este método es un enfoque basado en Machine Learning donde una función es entrenada a partir de imágenes positivas (imágenes con el objeto a detectar) y negativas (imágenes sin el objeto). De esta forma, es capaz de detectar el objeto en otras imágenes diferentes.

Cuando el entrenamiento ha sido finalizado, es posible generar un archivo XML con la finalidad de detectar el objeto sin necesidad de realizar el entrenamiento nuevamente (Ver Figura 3). En algunos repositorios open source se encuentran archivos XML con el entrenamiento de esta técnica orientado a la detección de manos. Estos archivos han sido testeados para comprobar si este método podría ser útil en el reconocimiento de gestos con las manos, obteniendo los siguientes resultados:

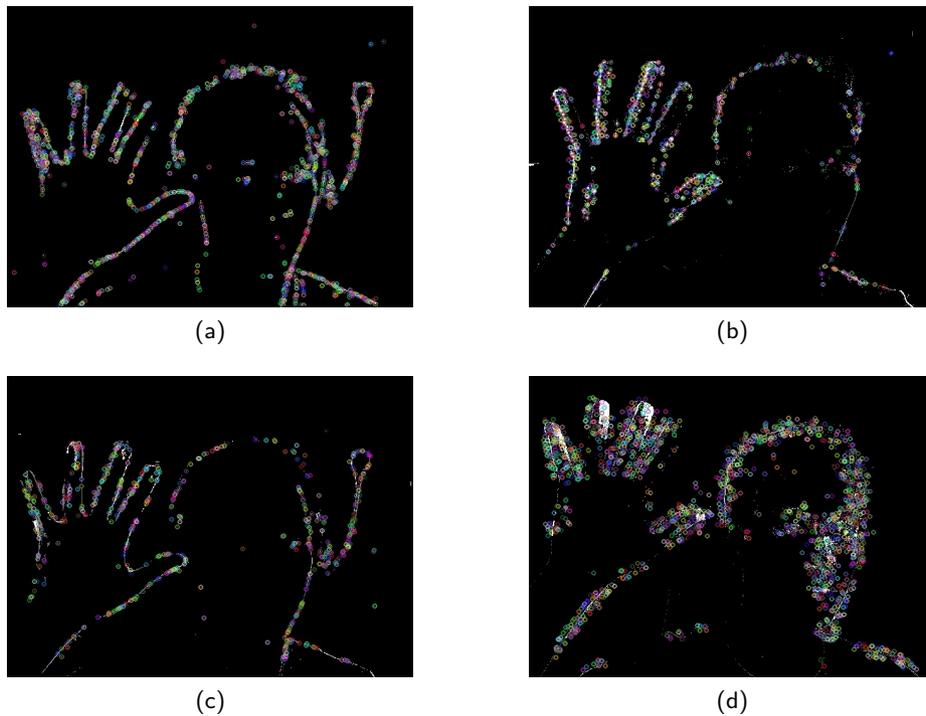


Fig. 2. Candidatos de los Algoritmos de Extracción con Eliminación de Fondo

- Se han probado 7 archivos XML, de los cuales 3 de ellos no han detectado la mano.
- Los restantes archivos han obtenido las siguientes tasa de aciertos:
 - Archivo 1 → $51/351 \approx 14.5\%$
 - Archivo 2 → $287/448 \approx 64.1\%$
 - Archivo 3 → $103/118 \approx 87.3\%$
 - Archivo 4 → $330/541 \approx 60.9\%$

3.6 Profundidad

En una primera instancia se pensó en hacer uso de la profundidad como un medio alternativo porque una de las desventajas de trabajar con la visión es que la luminosidad puede afectar al rendimiento de la misma. La idea consistía en tener sistema robusto donde se pudiera trabajar con ambas técnicas, en el cual la profundidad sería un sistema complementario, en el caso de que el sistema de visión fallara o como otra alternativa a utilizar y tener más opciones. Sin embargo, utilizar la profundidad requería de otra cámara más, puesto que habría que utilizar visión estereoscópica y no interesa este enfoque por el momento.

Un Modelo Inteligente de Interacción Natural Autoadaptativo basado en Visión Artificial

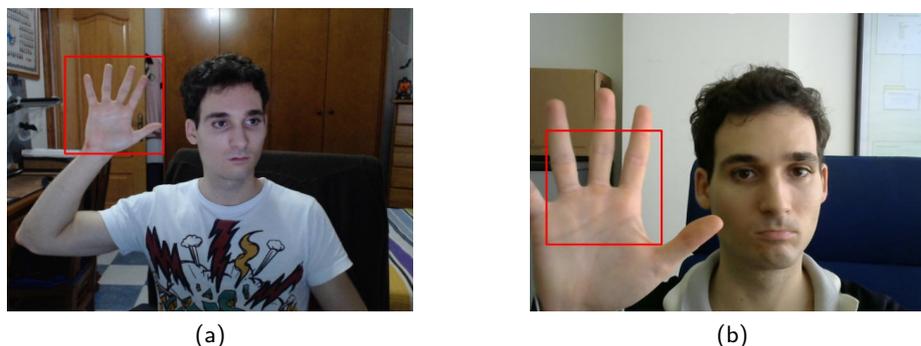


Fig. 3. Ejemplos del Clasificador en Cascada Haar para el Reconocimiento de Manos

4 Conclusiones

La propuesta presentada al inicio de la Sección 3 tiene un coste computacional bajo y el tiempo de respuesta es aceptable. Sin embargo, es un método que tiene en desventaja la luminosidad y necesita un modelo uniforme de fondo para funcionar correctamente mediante un proceso de calibración previo. Estas circunstancias hacen que esta técnica esté muy limitada para un sistema de interacción natural como se propone en la Tesis, no obstante, es un primer paso hacia este tipo de interacción.

Los algoritmos de extracción de característica se suelen utilizar para detectar objetos en una imagen pero afortunadamente también han detectado ciertas características en las manos del usuario con lo cual puede ser útil para incorporarlos en el proceso de interacción natural. Es necesario utilizar una eliminación del background para que omita aquellos atributos que no tienen relación con el usuario y se deberían seleccionar exclusivamente las características correspondientes a la mano.

El Clasificador en Cascada basado en características Haar ha sido desestimado como una posible técnica para el reconocimiento de gestos con las manos debido a que su tasa de aciertos ha sido muy baja, teniendo muchos falsos positivos.

No se ha desarrollado todavía un algoritmo de características cuyo objetivo sea la definición de la mano. Por ello, se está diseñando un algoritmo de extracción de características que detecte una serie de atributos característicos de la mano y la información que pueda extraerse de uno de los algoritmos candidatos con los que se ha experimentado.

Por último, se está trabajando también en un algoritmo que permita elegir el modo de interacción más conveniente en función de diversos factores del entorno, como por ejemplo, la movilidad del usuario.

References

- [1] Myers, B.A.: A brief history of human-computer interaction technology. *interactions* **5**(2) (1998) 44–54
- [2] Falcao, C., Lemos, A.C., Soares, M.: Evaluation of natural user interface: A usability study based on the leap motion device. *Procedia Manufacturing* **3** (2015) 5490–5495
- [3] Wigdor, D., Wixon, D.: *Brave NUI world: designing natural user interfaces for touch and gesture*. Elsevier (2011)
- [4] Norman, D.A.: Natural user interfaces are not natural. *interactions* **17**(3) (2010) 6–10
- [5] Chêne, D., Pillot, V., Chaumon, M.É.B.: Tactile interaction for novice user. In: *International Conference on Human Aspects of IT for the Aged Population*, Springer (2016) 412–423
- [6] Gupta, A., Pietrzak, T., Roussel, N., Balakrishnan, R.: Direct manipulation in tactile displays. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM (2016) 3683–3693
- [7] Villamor, C., Willis, D., Wroblewski, L.: *Touch gesture reference guide*. Touch Gesture Reference Guide (2010)
- [8] Ohn-Bar, E., Trivedi, M.M.: Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations. *IEEE transactions on intelligent transportation systems* **15**(6) (2014) 2368–2377
- [9] Song, Y., Demirdjian, D., Davis, R.: Continuous body and hand gesture recognition for natural human-computer interaction. *ACM Transactions on Interactive Intelligent Systems (TiiS)* **2**(1) (2012) 5
- [10] Marin, G., Dominio, F., Zanuttigh, P.: Hand gesture recognition with leap motion and kinect devices. In: *Image Processing (ICIP), 2014 IEEE International Conference on*, IEEE (2014) 1565–1569
- [11] Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2015) 815–823
- [12] Arjunlal, M.: A survey on hand gesture recognition and hand tracking. *International Journal of Scientific Engineering and Applied Science* **2**(1) (2016) 4
- [13] Graves, A., Mohamed, A.r., Hinton, G.: Speech recognition with deep recurrent neural networks. In: *Acoustics, speech and signal processing (icassp), 2013 IEEE international conference on*, IEEE (2013) 6645–6649

Acelerando la computación científica basada en estructuras de datos irregulares

Algoritmos genéticos aplicados a problemas de planificación en arquitecturas heterogéneas

Francisco Orts¹

Universidad de Almería, francisco.orts@ual.es

Resumen En este trabajo se estudia un Modelo de Microreología Activa donde una partícula trazadora Browniana se introduce en un fluido en el que hay localizadas más esferas pequeñas, que también sufren el movimiento Browniano. La trazadora es empujada con una fuerza constante y su trayectoria es supervisada con la finalidad de extraer las propiedades mecánicas del fluido. En particular, el observable es el coeficiente de fricción efectivo, que puede ser obtenido a partir de la velocidad media de la trazadora. Sin embargo, debido a los enormes requerimientos computacionales de este modelo su aplicabilidad es limitada. El objetivo de este trabajo es el de acelerar estos modelos mediante la explotación óptima de los modernos clústeres de Multi-GPUs. En la bibliografía se pueden encontrar trabajos anteriores basados en implementaciones GPU de las simulaciones incluidas en el modelo. Dichas implementaciones explotan el paralelismo de estos modelos a un bajo nivel. Sin embargo, con este trabajo proponemos el diseño de un planificador de tareas basado en un algoritmo genético para acelerar estos modelos sobre los modernos clústeres de Multi-GPU. Esto se consigue gracias a que los cores de las CPUs y las GPUs pueden ser explotados de forma óptima a la vez. Este diseño se ha evaluado con varias configuraciones de un clúster de Multi-GPU y ha conseguido reducir significativamente tiempo de ejecución de otras implementaciones similares.

1. Introducción

La reología es el estudio de las propiedades de flujo de la materia, particularmente de la materia blanda, que muestra un comportamiento complejo, identificado por dependencias no lineales de la tensión frente a la velocidad de deformación. Ejemplos típicos son: las pinturas, el almidón de maíz, etc. Por lo tanto, es un tema interdisciplinar, que es estudiado por diferentes comunidades, incluyendo ingenieros químicos, físicos, ingenieros de materiales y químicos [1]. Los modelos de reología teórica tratan de proporcionar ecuaciones constitutivas (macroscópicas) para un material particular basado en principios fundamentales a nivel microscópico, relacionando así el comportamiento macroscópico con la dinámica microscópica. La microreología, por otra parte, fue propuesta hace

casi veinte años como una técnica alternativa para obtener propiedades reológicas en muestras de especial dificultad o coste. En la microreología, se utiliza el movimiento microscópico de los trazadores coloidales utilizados como sondas, ya sea libremente en el medio, o sometidos a fuerzas externas [2], para evaluar directamente la dinámica microscópica.

En este trabajo, el Modelo de Microreología Activo (AMM) descrito en [4,5] se considera como un prototipo en este campo. Un conjunto de esferas Brownianas representan el baño cuya reología se calcula mediante el estudio de la dinámica de una partícula trazadora mayor que las del baño.

El gran tamaño de la partícula trazadora implica grandes efectos de tamaño finito, que en este trabajo corregimos usando un modelo teórico basado en la ecuación de Navier-Stokes. Por lo tanto, un número grande de trayectorias para distintos tamaños de sistema tiene que ser analizado. Este gran número de trayectorias hace el modelo computacionalmente costoso y lento. Las enormes exigencias computacionales han limitado el número de partículas del baño y, como consecuencia, un estudio detallado de la aplicabilidad del modelo.

La computación GPU ha sido utilizada en la literatura para acelerar simulaciones de sistemas coloidales para la consideración explícita de las interacciones hidrodinámicas de largo alcance [6,7]. Recientemente, se ha desarrollado una implementación GPU para acelerar el cálculo de las trayectorias de la trazadora asumiendo una fricción constante con el solvente sin interacciones hidrodinámicas, que es el objetivo de este trabajo [8,9]. Como diferentes sistemas son simulados, en este artículo se busca una optimización del problema que use los diferentes elementos de procesado que hay en los modernos clústeres de Multi-GPUs.

El mapeo de tareas sobre un clúster es un desafío debido a la heterogeneidad de la plataforma y de las tareas. Una propuesta de planificación para las tareas paralelas involucradas en AMM es la principal contribución de este trabajo. Por tanto, el punto de inicio de la propuesta paralela asume que los códigos que simulan las ejecuciones en CPUs y GPUs están disponibles y, además, las GPUs pueden explotar un paralelismo de bajo nivel para el cómputo de cada trayectoria.

Podría considerarse una planificación dinámica pero las características particulares de AMM permiten desarrollar un preprocesado que define la cola de tareas para cada CPU y GPU (de aquí en adelante llamados Unidades de Procesado, PUs), evitando así la sobrecarga que supone una planificación dinámica. Para justificar los detalles de este preprocesado, en primer lugar, la planificación se define formalmente como un problema de optimización global y después, se considera un algoritmo genético general para solucionarlo. De este modo, la aproximación paralela incluye tres fases: (1) profiling experimental del tiempo de ejecución consumido por cada trayectoria sobre cada tipo de PU de acuerdo con el número de las partículas del baño; (2) definición de la planificación propuesta por el Algoritmo Genético (GA) sobre todas las PUs; y (3) ejecución paralela considerando la planificación definida en la etapa anterior. El resultado es una versión paralela de AMM que reduce de forma considerable el tiempo de ejecución.

2. Avances

2.1. Modelo de microreología

El problema físico que se pretende estudiar, llamado, Microreología Activa, se ha descrito en la literatura [2,10]. Por lo tanto, en esta sección sólo se describe brevemente. Una partícula intrusa (trazadora), típicamente de tamaño coloidal, se introduce en un medio (normalmente un fluido viscoelástico) y una fuerza externa actúa sobre ella. La dinámica de esta partícula trazadora se monitoriza, proporcionando información de la reología del fluido. El fluido se modela como un sistema de esferas Brownianas, imitando coloides duros. El movimiento Browniano se describe por la ecuación de Langevin [11], que para la partícula j se escribe:

$$\frac{d^2 \mathbf{r}_j}{dt^2} = \sum_i \mathbf{F}_{ij} - \gamma_0 \frac{d\mathbf{r}_j}{dt} + \boldsymbol{\eta}_j(t) + \mathbf{F}_{\text{ext}} \delta_{j1} \quad (1)$$

donde los términos en el lado derecho son las fuerzas de interacción, fricción con el solvente, fuerza aleatoria y fuerza externa, respectivamente; γ_0 es el coeficiente de fricción con el solvente, el cual está relacionado con la fuerza aleatoria de acuerdo con el teorema FDT [11]. La fuerza externa, \mathbf{F}_{ext} , actúa sólo sobre la trazadora, etiquetada como $j = 1$, que es constante en nuestro modelo, este hecho es expresado mediante la conocida delta de Kronecker, denotada por δ_{j1} .

La partícula trazadora se empuja con una fuerza constante, y su trayectoria se graba. El coeficiente de fricción efectivo de la trazadora con el baño, γ_{eff} , se obtiene a partir de la velocidad media de la trazadora utilizando la relación de estado estacionaria: $\mathbf{F}_{\text{ext}} = \gamma_{\text{eff}} \langle \mathbf{v} \rangle$. Se necesita un gran número de trayectorias para obtener valores precisos de γ_{eff} . Sin embargo, la trazadora deforma el baño cuando se desplaza, ya que su tamaño es mucho mayor que el de las partículas del baño, y se pueden presentar efectos de tamaño finito, debido a las condiciones periódicas de contorno. Teniendo en cuenta la mecánica de fluidos, Hasimoto [13] mostró que el coeficiente de fricción efectivo medido mediante un array de partículas depende del espaciado del entramado, L , ya que:

$$\frac{1}{\gamma_{\text{eff}}} = \frac{c}{L} + \frac{1}{\gamma_{\infty}} \quad (2)$$

donde c es una constante que depende de la estructura del array, y γ_{∞} es el coeficiente de fricción efectivo medido para una partícula aislada [13]. Siguiendo este resultado teórico, las simulaciones se hacen usando distintos tamaños de problema, L , para obtener $\gamma_{\text{eff}}(L)$. Como se muestra en la Fig. 1, el coeficiente de fricción inverso es lineal con el inverso del tamaño del sistema, y el coeficiente de fricción de un sistema infinito puede extraerse de la extrapolación a $1/L \rightarrow 0$. Debido a que la fricción de volumen es constante, cambiar el sistema de tamaño implica cambiar el número de partículas. Por tanto, nuestro problema necesita un gran número de simulaciones de *i)* sistemas con distinto número de partículas, y *ii)* un gran número de trayectorias para cada tamaño de sistema, requiriendo *iii)* la resolución de N ecuaciones de movimiento para cada trayectoria.

Acelerando la computación científica basada en estructuras de datos irregulares

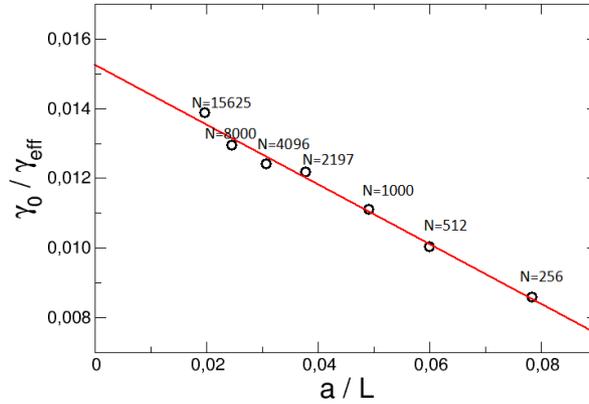


Figura 1. Coeficiente de fricción inverso vs. longitud inversa de la caja de simulación para un sistema con una fracción de volumen de $\phi = 0,50$, y una trazadora de tamaño $a_t = 3a_b$ empujada con una fuerza $F = 5 k_B T / a_b$.

2.2. Implementación paralela del problema de Microreología

Trabajos anteriores se centraron en acelerar la computación de una sólo trayectoria de la partícula trazadora (primer nivel) en la GPU [8,9]. Sin embargo, para permitir avances en este tipo de modelos, es necesario explotar de forma eficiente el otro nivel de paralelismo sobre clústeres multi-GPU.

El modelo define un conjunto de M tareas que computan cada trayectoria de la trazadora. Por lo tanto, un conjunto de trayectorias de la trazadora puede computarse de forma paralela en los cores de las CPU y en las GPUs del clúster. De este modo, cada CPU-core (GPU) puede ejecutar el código secuencial (código CUDA) para computar una única trayectoria, y el conjunto de las tareas puede computarse con la colaboración de los elementos de cómputo del clúster (referidos como PUs). Sólo la GPU aprovecha el primer nivel de paralelismo.

Es necesario definir un planificador de tareas adecuado con el fin de obtener el rendimiento óptimo de la versión paralela. Esto puede verse como un planificador de máquinas paralelas no relacionadas [14]. Formalmente, puede definirse como sigue: Sea $\mathcal{M}^k \in \mathcal{M}$ la PU k del clúster multi-GPU \mathcal{M} , con $k = 1, \dots, K$ donde K es el número de CPU-cores y GPUs disponibles. Sea $\{R_m\}$ el conjunto de trayectorias de la trazadora que define el modelo donde R_m identifica cada trayectoria particular, con $m = 1, \dots, M$ y $M = I \times q$ representa el número total de las trayectorias a computar, donde I denota los distintos tamaños del sistema (N_i con $1 \leq i \leq I$) y q representa el número de trayectorias para cada tamaño. El propósito es encontrar un mapeo que minimice el makespan C_{max} :

Acelerando la computación científica basada en estructuras de datos irregulares

minimizar:

$$C_{max}$$

sujeto a:

$$\sum_{i=1}^s x_{k,i} t_k^i \leq C_{max}, \quad 1 \leq k \leq K$$

$$\sum_{k=1}^K x_{k,i} = q, \quad 1 \leq i \leq I$$

$$x_{k,i} \in \{0, 1 \dots q\}, \quad 1 \leq k \leq K; \quad 1 \leq i \leq I$$

donde $x_{k,i}$ es un elemento de la matriz, X , que define la asignación de las trayectorias (trabajos) a PUs (máquinas), su valor (entero) representa el número de trayectorias de tamaño N_i asignado a la k -th PU; t_k^i representa el tiempo de ejecución de cada trayectoria. Las restricciones para X significan que cada trayectoria se computa en una máquina y cada conjunto de q trayectorias con el mismo tamaño se distribuye en todas las PUs. De este modo, una asignación de trabajos a máquinas se define por la matriz X . Nuestro propósito es la identificación de la matriz que optimiza el makespan. Este tipo de problema de optimización es NP-completo, ya que el número de los posibles valores para los elementos de X es muy alto incluso para valores reducidos de I , q y K .

Hay una gran variedad de trabajos previos donde los algoritmos genéticos (GA) se utilizan para resolver problemas de planificación [15]. En este artículo, se adapta un GA clásico para resolver el problema de planificación de las tareas involucradas en un modelo de microreología sobre plataformas Multi-GPU.

El GA trabaja con un conjunto de individuos que representan cada posible solución del problema de planificación (*la población*). Esto se traduce en un procedimiento iterativo que comienza con un conjunto arbitrario de individuos, $P(0)$, y en cada iteración t se aplican la selección, así como operadores genéticos, a la población, $P(t)$. De este modo, la población evoluciona constantemente. El mecanismo de selección permite que los individuos de las nuevas poblaciones estén más cerca del óptimo (ver Algoritmo 1).

Cada individuo se representa por la matriz X de dimensión $K \times I$, X define la asignación de trayectorias a las PUs según la definición dada anteriormente. Es bien conocido que en los problemas de planificación C_{max} alcanza el valor mínimo cuando todas las máquinas completen sus trabajos a la vez; esto significa que la carga de trabajo estaría equilibrada entre las máquinas de acuerdo a su capacidad computacional. El desequilibrio puede ser evaluado por el parámetro $UB = t_{Max} - t_{Min}$, donde t_{Max} (t_{Min}) es el máximo (el mínimo) tiempo de ejecución entre todos los elementos del sistema paralelo. UB puede ser considerado como la función objetivo de la planificación. $T_k = (t_k^1, \dots, t_k^I)$ sería el vector de los tiempos de ejecución para cada trayectoria de tamaño N_i sobre la PU, k , y $X_k = (x_{k,1}, \dots, x_{k,I})$ sería el vector que define la asignación de trayectorias de

Algorithm 1 Algoritmo Genético

```

1:  $t \leftarrow 0$ 
2: Inicializar la población aleatoriamente  $P(0)$ 
3: Evaluar la función objetivo (fitness) para la población  $P(0)$ 
4: while condición de finalización no es verdadera do
5:    $t \leftarrow t + 1$ 
6:   Seleccionar  $P(t)$  de  $P(t - 1)$ 
7:   Aplicar operadores genéticos (cruce y mutación) a  $P(t)$ 
8:   Evaluar fitness para  $P(t)$ 
9: end

```

cada tamaño en el mismo elemento de computación. Por tanto, el tiempo de ejecución de la PU k se computa por $t_k = X_k \cdot T_k$ y UB se evalúa como la diferencia entre el max y el mín de t_k para cada individuo X . Tras la evaluación de la población entera, todos los individuos se ordenan de acuerdo a su función objetivo. Por tanto, los individuos con menor UB serán seleccionados cuando el GA avance.

En cada iteración del Algoritmo 1 se aplican dos operaciones a la población para promover la evolución. Primeramente, se define un conjunto de pares de individuos (padres) y entonces, se producen nuevos individuos (hijos) por el operador cruce. Para definir el cruce, las matrices relacionadas con cada individuo se consideran como vectores de I columnas. Se aplica la operación de cruce en un punto, descrita en la Figura 2. Una columna aleatoria se selecciona para dividir las matrices de ambos padres y generar nuevos individuos intercambiando los cuatro conjuntos de columnas. Por tanto, con este esquema, los hijos pueden ser considerados soluciones válidas ya que se verifican las restricciones para las columnas de sus matrices. Tras la recombinación, actúa el operador mutación sobre cada descendiente y éste puede alterar la distribución de cada columna (con una probabilidad de 1%). Ésto consiste en un intercambio aleatorio de trayectorias con el mismo tamaño entre un par de PUs, p. e. elementos en la misma columna de la correspondiente matriz intercambian sus trabajos de forma parcial. La población mantiene un tamaño constante (S). La fase de selección sólo consiste en la selección de los mejores S individuos ya que la población ha sido previamente ordenada de acuerdo a su fitness, UB . El proceso iterativo acaba cuando el mejor individuo en poblaciones sucesivas tiene valores similares de UB .

Por tanto, si el conjunto de tiempos de ejecución $\{t_k^i\}$ con $1 \leq k \leq K$ y $1 \leq i \leq I$ es conocido, el GA descrito anteriormente puede identificar un mapeo de las trayectorias entre las PUs cercano al óptimo, X^* .

Consecuentemente, para cada instancia de datos de entrada del modelo de microreología se definen tres etapas para optimizar su tiempo paralelo en un clúster heterogéneo: (1) etapa de profiling, que estima el conjunto de tiempos de ejecución $\{t_k^i\}$ de acuerdo a los distintos tamaños de las trayectorias simuladas en el modelo y el número de tipos de elementos de proceso de la plataforma; (2) cálculo del mapeo cercano al óptimo por el GA, que identifica el número de

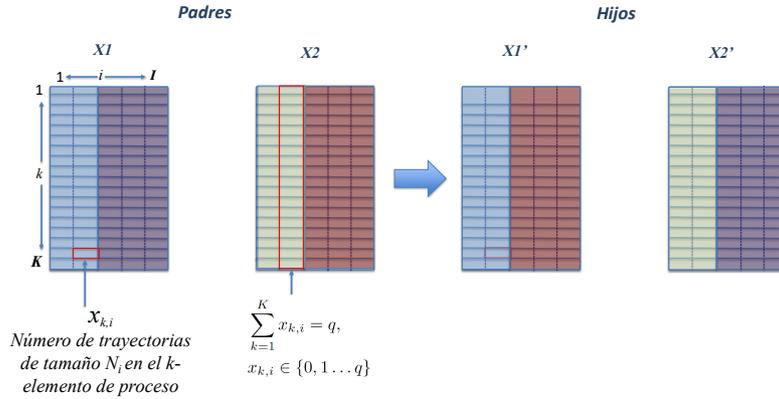


Figura 2. Recombinación para producir nuevos individuos de la población.

Tabla 1. Características de cada una de las 8 GPUs.

	M2070
Peak performance (double prec.) (TFlops)	0.51
Peak performance (simple prec.) (TFlops)	1.03
Device memory (GB)	5.2
Clock rate (GHz)	1.2
Memory bandwidth (GBytes/s)	150
Multiprocessors	14
CUDA cores	448
Compute Capability	2.0

trayectorias de cada tamaño para cada PU y (3) ejecución paralela de todas las simulaciones en las PUs del clúster. Un software específico basado en Python se ha desarrollado para distribuir las tareas que ejecuta cada PU en cada instante de tiempo de acuerdo al mapeo proporcionado por la etapa anterior.

2.3. Evaluación del planificador

Hemos testeado problemas de microreología con los siguientes tamaños: $N_i = 216, 512, 1000, 2197, 4096, 8000$ y 15625 partículas, con 250 trayectorias de 500 pasos temporales (correspondientes a 10^6 pasos).

Se ha considerado un clúster Multi-GPU que se compone de 4 nodos (Bullx R424-E3 Intel Xeon E5 2650) con 16 CPU-cores, 8GB RAM y 2 NVIDIA Tesla M2070 GPUs (Fermi) por nodo. Las características de las GPUs se dan en la Tabla 1. Las instancias del modelo de Microreología han sido mapeadas en dicho clúster Multi-GPU utilizando 3 configuraciones: (1) 64 CPU-cores; (2) 8 GPUs; y (3) 56 CPU-cores y 8 GPUs.

La versión CPU del problema de Microreología es un código secuencial codificado en Fortran. La versión GPU se ha implementado utilizando ANSI C y

CUDA (8) [16] como interfaz. Hemos utilizado un módulo de Phyton (Python’s Multiprocessing module) para implementar el planificador. Además, hemos considerado aritmética en doble precisión.

En primer lugar, se realiza la etapa de profiling. Es decir, se realiza una estimación del conjunto de tiempos de ejecución $\{t_k^i\}$ según los tamaños de sistema involucrados en el modelo y las dos clases de PUs (CPU-cores y GPUS) (ver Tabla 2). Nótese que AF es el factor de aceleración del tiempo de ejecución de una GPU frente al tiempo de ejecución de una CPU. El tiempo de ejecución aumenta con N_i . Además, el uso de la computación GPU no es beneficioso para acelerar problemas de microreología cuando N_i es bajo. Sin embargo, cuando $N_i \geq 1000$, el empleo de GPUs muestra ventajas en términos de rendimiento.

Tabla 2. Tiempo total de ejecución (en segundos) para 7 tamaños de sistema (N_i). Las columnas t_{GPU}^i y t_{CPU}^i identifican el tiempo de ejecución para cada trayectoria en una GPU y en un CPU-core, respectivamente, y AF es el factor de aceleración del tiempo de ejecución de la GPU con respecto al de la CPU.

N_i	t_{GPU}^i	t_{CPU}^i	AF
216	1580	790	0.5
512	1785	1860	1.0
1000	2240	3715	1.7
2197	2930	8710	3.0
4096	4450	18065	4.0
8000	7650	43080	5.6
15625	12050	113940	9.5

Tabla 3. Media de las estimaciones del tiempo total de ejecución y del parámetro de desbalanceo UB , en segundos, para 10 ejecuciones del GA con los parámetros de la Tabla 2 y las 3 configuraciones del clúster Multi-GPU.

	64 CPU-cores	8 GPU cards	56 CPU-cores + 8 GPUs
<i>Runtime</i>	742856 ± 7	1021551 ± 15	369299 ± 162
<i>UB</i>	51 ± 7	74 ± 22	2612 ± 225

En segundo lugar, se ejecuta el GA utilizando las tres configuraciones del clúster multi-GPU (64 CPU-cores; 8 GPUs; y 56 CPU-cores + 8 GPUs) y los datos de la Tabla 2. El tiempo medio de 10 ejecuciones del GA se muestra junto con \pm sus intervalos de confianza en la Tabla 3. La columna UB muestra el parámetro de desequilibrio (la función objetivo de la planificación), \pm su intervalo de confianza. Nótese que los intervalos de confianza de los tiempos de ejecución son más pequeños en el caso de ejecución homogénea (es decir, para 64 CPU-cores, es 7 con respecto a 742856) que en el caso de la configuración heterogénea. Esto es debido al hecho de que es más complicado mantener el

Tabla 4. Tiempo promedio, en segundos, de una ejecución del algoritmo genético (“Test GA”) y 2 ejecuciones utilizando las configuraciones heterogéneas (“ Test Experimental 1 y 2”) de la Tabla 3 y los parámetros de entrada de la Tabla 2.

	Test GA	Test Experimental 1	Test Experimental 2
t_{Max}	369222.5	379611.9	379961.6
t_{Min}	366873.6	376234.4	376456.5
UB	2348.9	3377.5	3505.0

equilibrio entre unidades de proceso heterogéneas. Además, podemos observar que la configuración más rápida es la heterogénea (compuesta por 56 CPU-cores y 8 GPUs) ya que tanto las CPUs como GPUs pueden contribuir para acelerar todas las instancias del problema de Microreología. De acuerdo a la Tabla 2, el GA distribuirá la ejecución de los tamaños pequeños del problema a CPU-cores y los grandes a las GPUs.

Finalmente, hemos centrado nuestra atención en la configuración más rápida según el GA (56 CPU-cores + 8 GPUS), para planificar todos los tamaños de sistema sobre el clúster multi-GPU. Por lo tanto, la Tabla 4 muestra el tiempo de ejecución de una simulación que usa el GA en comparación con dos ejecuciones paralelas reales que usan la planificación propuesta por el GA. Las implementaciones reales son las columnas “Test Experimental 1 y Test Experimental 2” de la Tabla 4. Analizando los tiempos de ejecución, podemos concluir que la estimación de nuestro GA es muy cercana al tiempo de ejecución de la experimentación real, teniendo en cuenta que el GA no tiene en cuenta ni el carácter aleatorio del modelo, ni la contención de cada nodo multicore. En resumen, nuestro GA para planificar las tareas del modelo de microreología sobre un clúster multi-GPU ha sido capaz de reducir de forma considerable el tiempo de ejecución. El tiempo de ejecución de las versiones anteriores las hacía inabordables, pero con este trabajo se ha pasando de ejecuciones de meses a solamente días.

3. Conclusiones

El trabajo se ha centrado en el desarrollo de una estrategia de planificación para distribuir las tareas paralelas de diferentes tamaños entre los cores CPU y GPU de un clúster Multi-GPU. En trabajos anteriores se habían implementado códigos CPU y GPU que eran eficientes para computar las simulaciones de las trayectorias de la trazadora. En esta propuesta se ha analizado y evaluado una aproximación que orquesta de forma óptima las tareas sobre los CPU-cores y las GPUS de un clúster moderno. Nos hemos basado en un GA, para definir a priori cual sería la mejor configuración del conjunto de tareas sobre los recursos CPUs y GPUs disponibles en la plataforma. Experimentalmente, hemos comprobado que los tiempos de ejecución obtenidos por el GA son muy similares a los tiempos de ejecución de los experimentos reales. Como resultado, hemos conseguido reducir el tiempo de ejecución de los modelos de microreología. De este modo, tanto la estadística como el tamaño del sistema a resolver han podido ampliar-

se. Actualmente, estamos trabajando en nuevas implementaciones que tengan en cuenta la contención de las arquitecturas multicore en la estimación del AG. Además, tenemos interés en evaluar nuestra estrategia en otras plataformas.

Referencias

1. P. Cicuta and A.M. Donald, “Microrheology: a review of the method and applications,” *Soft Matter*, vol. 3, pp. 1449–1455, 2007.
2. A.M. Puertas and T. Voigtmann, “Microrheology of colloidal systems,” *J. Phys. Condens. Matter*, vol. 26, no. 24, pp. 243101, 2014.
3. J. L. Hennessy and D. A. Patterson, *Computer Architecture: A Quantitative Approach*, Morgan Kaufmann, 2011.
4. I. Gazuz, A.M. Puertas, T. Voigtmann, and M. Fuchs, “Active and nonlinear microrheology in dense colloidal suspensions,” *Phys. Rev. Lett.*, vol. 102, no. 24, pp. 248302, 2009.
5. M.V. Gnann, I. Gazuz, A.M. Puertas, M. Fuchs, and T. Voigtmann, “Schematic models for active nonlinear microrheology,” *Soft Matter*, vol. 7, no. 4, pp. 1390–1396, 2011.
6. D.S. Bolintineanu, Gary S. Grest, Jeremy B. Lechman, Flint Pierce, Steven J. Plimpton, and P. Randall Schunk, “Particle dynamics modeling methods for colloid suspensions,” *Computational Particle Mechanics*, vol. 1, no. 3, pp. 321–356, 2014.
7. M. Kopp and F. Höfling, “GPU-accelerated simulation of colloidal suspensions with direct hydrodynamic interactions,” *EPJ ST*, vol. 210, no. 1, pp. 101–117, 2012.
8. G. Ortega, A.M. Puertas, and E.M. Garzón, “Accelerating the problem of microrheology in colloidal systems on a GPU,” *J. Supercomput.*, vol. 73, no. 1, pp. 370–383, 2017.
9. G. Ortega, A.M. Puertas, F.J. de Las Nieves, and E.M. Garzón, *GPU Computing to Speed-Up the Resolution of Microrheology Models*, pp. 457–466, Springer International Publishing, Cham, 2016.
10. T.A. Waigh, “Advances in the microrheology of complex fluids,” *Rep. Prog. Phys.*, vol. 79, no. 7, pp. 074601, 2016.
11. J.K.G. Dhont, *An Introduction to Dynamics of Colloids*, Studies in Interface Science. Elsevier Science, 1996.
12. W. Paul and D. Y. Yoon, “Stochastic phase space dynamics with constraints for molecular systems,” *Phys. Rev. E*, vol. 52, pp. 2076–2083, Aug 1995.
13. H. Hasimoto, “On the periodic fundamental solutions of the Stokes equations and their application to viscous flow past a cubic array of spheres,” *J. Fluid Mech.*, vol. 5, pp. 317–328, 1959.
14. D.B. Shmoys and E. Tardos, “An approximation algorithm for the generalized assignment problem,” *Math. Program.*, vol. 62, no. 3, pp. 461–474, Dec. 1993.
15. V. Sels, J. Coelho, A.M. Dias, and M. Vanhoucke, “Hybrid tabu search and a truncated branch-and-bound for the unrelated parallel machine scheduling problem,” *Computers & Operations Research*, vol. 53, pp. 107 – 117, 2015.
16. NVIDIA Corporation, “CUDA C PROGRAMMING GUIDE PG-02829-001_v7.5,” 2015.

Computación de altas prestaciones para la resolución de problemas de optimización en bioinformática*

OptiPharm: un algoritmo evolutivo para comparar la forma molecular

Savíns Puertas Martín
savinspm@ual.es

Grupo de Investigación Supercomputación - Algoritmos (SAL), Universidad de Almería, Campus de Excelencia Internacional Agroalimentario (ceiA3) Almería, España

Resumen Los métodos de Cribado Virtual (VS) pueden acelerar drásticamente el proceso de descubrimiento de fármacos. Entre los enfoques de VS más utilizados se encuentran los Métodos de Similitud de Forma (SSM) que consisten en comparar en detalle la forma global de una molécula frente a una base de datos de compuestos farmacológicos. Si bien existen ya varios métodos de SSM, en este trabajo presentamos *OptiPharm*, que incluye un esquema de metaheurística parametrizable que mejora la precisión de la predicción y ofrece un mayor rendimiento computacional frente a otros métodos SSM muy usados actualmente como puede ser WEGA.

1. Introducción

El descubrimiento de nuevos fármacos es un proceso muy costoso. Se requiere de alrededor de 15 años y aún así, las tasas de éxito son generalmente

* Este trabajo ha sido financiado por el Ministerio de Economía y Competitividad de España (TIN2015-66680-C2-1-R, CTQ2017-87974-R), la Junta de Andalucía (P12-TIC301), la Fundación Séneca-Agencia de Ciencia y Tecnología de la Región de Murcia bajo los proyectos 19419/PI/14 y 18946/JLI/13. Powered@NLHPC: Esta investigación ha sido parcialmente soportada por la infraestructura de supercomputación del NLHPC (ECM-02). Los autores también agradecen los recursos informáticos y el soporte técnico proporcionado por la Plataforma Andaluza de Bioinformática de la Universidad de Málaga. Este trabajo ha sido parcialmente apoyado por las instalaciones informáticas del Centro Extremeño de Tecnologías Avanzadas (CETA-CIEMAT), fundado por el Fondo Europeo de Desarrollo Regional (FEDER). CETA-CIEMAT pertenece a CIEMAT y al Gobierno de España. Juana López Redondo es una beneficiaria del programa español “Ramón y Cajal”, cofinanciado por el European Social Fund. Savíns Puertas Martín es un beneficiario del programa español “Formación de profesorado universitario”, financiado por el Ministerio de Educación, Cultura y Deporte de España.

muy bajas [1]. Se han utilizado muchos enfoques experimentales para descubrir nuevos compuestos con las propiedades farmacológicas deseadas que van desde la medicina tradicional [2] hasta las infraestructuras de cribado de alto rendimiento (High Throughput Screening, HTS) [3]. Usando este último enfoque, se comenzó a usar una metodología para filtrar bases de datos de compuestos en subconjuntos mucho más pequeños de compuestos que podrían caracterizarse experimentalmente. Esta idea se denominó Cribado Virtual (Virtual Screening, VS) y redujo el tiempo y los costes necesarios en el descubrimiento de fármacos [4].

Los métodos de VS se pueden dividir en métodos basados en estructura (Structure Based Virtual Screening, SBVS) y basados en ligandos (Ligand Based Virtual Screening, LBVS). Cuando se conoce la estructura de la proteína objetivo, se puede aplicar SBVS. Sin embargo, los métodos de SBVS no siempre se pueden aplicar, especialmente cuando el número de estructuras conocidas de la proteína objetivo es muy bajo [5]. En estos casos se usan los métodos LBVS, donde solo se necesitan datos sobre los compuestos conocidos para derivar en otros nuevos y mejorados. Este estudio se centra principalmente en los métodos LBVS y en concreto en una de sus técnicas, los métodos de similitud de forma (Shape Similarity Methods, SSM). En los SSM, una gran base de datos de compuestos se evalúa contra una molécula de consulta, para proporcionar información sobre cuál de las moléculas de la base de datos es geoméricamente similar, en términos de forma molecular global, a la molécula consultada (*query*).

En este trabajo, presentamos un nuevo método llamado *OptiPharm*, que supera el estado del arte de WEGA [6] en términos de rendimiento, manteniendo una precisión similar en la predicción, e incorporando un nuevo esquema de optimización que puede adaptarse a los distintos conjuntos de datos moleculares por medio de parámetros.

2. OptiPharm

OptiPharm encuentra sus orígenes en UEGO [7], en consecuencia, es un algoritmo evolutivo multimodal que es capaz de resolver problemas donde la función objetivo tiene múltiples óptimos locales y descubrir la estructura de estos óptimos [8–10]. La solución de *OptiPharm* implica una técnica de *enfriamiento* que permite centrarse en las regiones prometedoras del espacio, comenzando con un radio relativamente grande que disminuye a medida que la búsqueda avanza.

En los algoritmos evolutivos multimodales, el concepto de subpoblación, el cual es fundamental en *OptiPharm*, es sustituido por especie. Tanto es así que la gestión de las especies es una de las partes centrales de *OptiPharm*. De hecho, consiste en un conjunto de procedimientos para crear, seleccionar y eliminar especies durante todo el proceso de optimización.

Una especie puede ser considerada como una *ventana* en todo el espacio de búsqueda (ver Figura 1). Esta ventana está definida por su centro y su radio. El centro es una solución, y el radio es un número positivo. Además, el radio de la especie no es constante a lo largo de la ejecución del algoritmo ni es el mismo

para cada especie, sino que coincide con el diámetro del dominio inicial en el primer nivel y converge a cero cuando el número de niveles o iteraciones tiende hacia el infinito

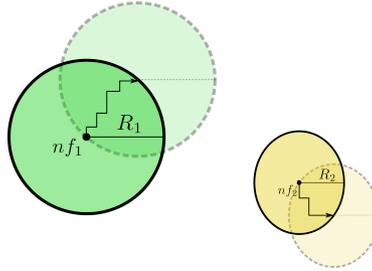


Figura 1. Proceso de optimización de especies.

2.1. Parámetros de OptiPharm

En *OptiPharm*, al igual que en UEGO, los parámetros más importantes son los que se definen en cada nivel: el radio (R_i), el número máximo de evaluaciones para la creación de especies (new_i) y para la optimización (n_i). Estos parámetros se calculan a partir de los siguientes parámetros de entrada proporcionados por el usuario [8]:

- *evals* (N): El número máximo de evaluaciones para todo el proceso de optimización.
- *levels* (L): El número máximo de niveles (es decir, etapas de enfriamiento).
- *max_spec_num* (M): Longitud máxima del listado de especies.
- *min_r* (R_L): El radio asociado con el nivel máximo.

2.2. Descripción algorítmica de OptiPharm

En la Figura 2 se da una descripción global de *OptiPharm*. A continuación se describen las diferentes etapas clave del algoritmo:

- *Inicialización*: Se crea una nueva lista de especies que consiste en cinco especies en el nivel 1. Teniendo en cuenta que el radio asociado en el nivel 1 es el diámetro del espacio de búsqueda, estas especies siempre cubren todo el espacio de búsqueda.
- *Reproducción*: Para cada especie en la lista, se crean soluciones candidatas aleatorias en la ventana de la especie, y por cada par de soluciones candidatas, se computa el punto medio del segmento que las conecta. Se evalúan tanto el punto medio como los extremos y si alguna de estas soluciones tiene un mejor valor de aptitud que el centro de la especie, entonces ese punto será

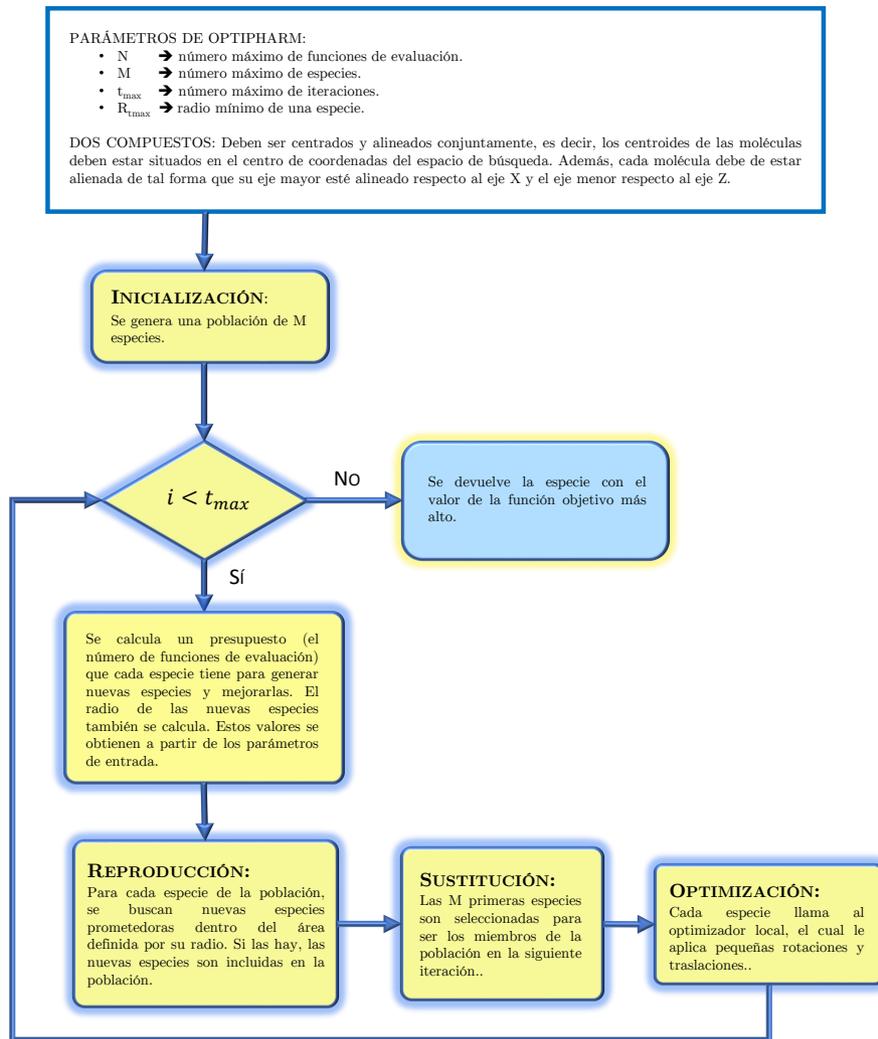


Figura 2. Estructura general de *OptiPharm*.

el nuevo centro, manteniendo el mismo valor de radio. Además, si el valor de la aptitud en un punto medio es peor que en los miembros correspondientes de su par, entonces los miembros del par se insertan en la lista de especies.

- *Sustitución*: se eliminan especies para reducir la longitud de la lista al valor max_spec_num . Se ordenan las especies de mayor a menor valor de aptitud y se eliminan las especies menos aptas hasta que el número de éstas sea inferior o igual al del valor max_spec_num .
- *Optimización*: Ejecuta el optimizador para cada especie con un número determinado de evaluaciones (ver Figura 1). Este presupuesto se calcula dividiendo el número máximo de evaluaciones asignadas al procedimiento de optimización en el nivel i , por el número máximo de especies o el tamaño máximo de la población. En este trabajo se ha utilizado el optimizador SASS [11].

2.3. Función a optimizar: similitud de forma

El valor de similitud entre dos moléculas A y B se calcula como el volumen superpuesto de sus átomos. En particular, para comparar los resultados obtenidos por *OptiPharm* con los logrados por WEGA, la función de similitud se ha implementa como en WEGA [12]. A modo de resumen, esta función se escribe de la siguiente forma:

$$V_{AB}^g = \sum_{i \in A, j \in B} w_i w_j v_{ij}^g \quad (1)$$

donde w_i y w_j son los pesos asociados con los átomos i y j , respectivamente. Este peso se calcula al resolver la siguiente expresión matemática:

$$w_i = \frac{v_i^g}{v_i^g + k \sum_{j \neq i} v_{ij}^g} \quad (2)$$

donde k es una constante universal, que se establece en 0.8665, como se indica en [12], v_i es el volumen del átomo i , cuyo valor se calcula como $v_i = \frac{4\pi\sigma^3}{3}$, de manera similar a como se hizo en [12].

Finalmente, la superposición v_{ij}^g se representa como un producto de funciones gaussianas:

$$v_{ij}^g = \int g_i(r) g_j(r) dr = \int p e^{-\left(\frac{3p\pi^{1/2}}{4\sigma_i^3}\right)^{2/3} (r-r_i)^2} p e^{-\left(\frac{3p\pi^{1/2}}{4\sigma_j^3}\right)^{2/3} (r-r_j)^2} dr \quad (3)$$

donde p es un parámetro que controla la suavidad de las esferas gaussianas, es decir, la altura de la función original de Gauss, y σ es el radio del átomo. De forma más específica, el radio representa el conocido radio de van der Waals. Los valores asociados a esos dos parámetros se obtienen por conocimiento empírico. Para este problema, se consideran los mismos valores que los propuestos en [12].

Observe que el valor obtenido en la Ecuación 1 depende del número de átomos de las dos moléculas comparadas, es decir, cuanto mayor sea este número, mayor

será el valor de V_{AB}^g . Se encuentra en el intervalo $[0, +\infty)$. Para poder medir el grado de similitud entre los compuestos, independientemente del número de átomos que los componen, se calcula el valor de Similitud de Tanimoto [13]:

$$Tc = \frac{V_{AB}}{V_{AA} + V_{BB} - V_{AB}} \quad (4)$$

donde V_{AA} y V_{BB} es el volumen de superposición de las moléculas A y B consigo mismas, respectivamente. Tiene un valor en el rango $[0, 1]$, donde 0 significa que no hay superposición, y 1 significa que la forma de ambas moléculas es la misma.

2.4. Características específicas

Adicionalmente, para adaptar *OptiPharm* a la casuística del problema, se han implementando otras características, las cuales se describen brevemente a continuación:

- **Dimensiones del problema.** El número de parámetros a optimizar por *OptiPharm* es 10: 1 para definir el ángulo de rotación de la molécula (θ), 6 para definir los dos puntos que forman el eje ($x1, y1, z1, x2, y2$ y $z2$) y 3 más para definir la traslación en el espacio (Δ_x, Δ_y y Δ_z).
- **Especies iniciales.** Cinco son las especies iniciales de *OptiPharm*. Estas especies representan la molécula variable en su posición inicial, rotada 180° en el eje X, rotada 180° en el eje Y, rotada 180° en el eje Z y una aleatoria para evitar caer en óptimos locales. La elección de estas cinco especies se encuentra fundamentada en los resultados empíricos que se han obtenido con cientos de miles de comparaciones. La experimentación realizada fue la siguiente: se fijaron los ejes y se optimizó el ángulo y la traslación.
- **Limitación de movimiento.** *OptiPharm* es un algoritmo evolutivo global. Esto significa que puede encontrar todos los óptimos locales en todo el espacio de búsqueda. Debido a esto, se ha considerado un espacio de movimiento variable y calculado de forma específica para cada configuración. Con ello se consigue el objetivo de reducir el espacio de búsqueda y por otro lado, se eliminan zonas del espacio físico que obtendrían malas soluciones por estar demasiado lejos del centro.
- **Puntos del eje de rotación.** Para formar un eje son necesarios dos puntos, no siendo posible que coincidan en la misma posición. *OptiPharm* implementa dos técnicas distintas:
 - Considerar un punto en el centro de masas de la molécula y permitir que el otro se genere en cualquier lugar del espacio, siempre que esté dentro de los límites precalculados.
 - Ambos puntos se generan de forma aleatoria dentro del espacio de los límites precalculados.
- **Reconfiguración de límites.** No se puede considerar unos límites constantes para todo par de moléculas a ser comparadas ya que los tamaños pueden ser muy dispares. Es por ello que antes de comenzar cada proceso

de optimización, unos límites específicos son establecidos para cada par de moléculas.

- **Radios de Van del Waals.** Se han considerado las dos opciones: para experimentos con mismo radio para todos los elementos, el valor elegido ha sido 1.8 Å mientras que para experimentos con radios distintos, los valores utilizados pueden ser consultados en [14].
- **Hidrógenos.** Se pueden o no considerar los hidrógenos.

3. Resultados y discusiones

Se diseñaron dos estudios computacionales considerando las bases de datos FDA [15] y DUD [16].

Dada la naturaleza de *OptiPharm*, cada experimento ha sido ejecutado 100 veces mientras que para WEGA, solo una, ya que es determinista (utiliza un método de descenso por gradientes) y diferentes ejecuciones siempre producen el mismo resultado. Además, *OptiPharm*, como es parametrizable, ha sido ejecutado con dos versiones distintas. Cada una persigue un objetivo:

- (i) *OptiPharm Robusto* (OpR). En este caso, el conjunto de parámetros de entrada se elige para hacer que *OptiPharm* sea exhaustivo y robusto. En particular, se consideraron los siguientes valores: $N = 200000$ evaluaciones de funciones, $M = 5$ conformaciones moleculares, $t_{max} = 5$ iteraciones y $R_{t_{max}} = 1$ como el radio más pequeño posible.
- (ii) *OptiPharm Rápido* (OpF). En esta ocasión, los parámetros se ajustan para que los tiempos de ejecución sean más bajos o similares a los de WEGA, permitiendo una comparación justa entre ambos algoritmos. Se consideraron los siguientes valores: $N = 1000$ evaluaciones de funciones, $M = 5$ conformaciones moleculares, $t_{max} = 5$ iteraciones y un radio mínimo de $R_{t_{max}} = 5$.

3.1. Resultados obtenidos para FDA

Es importante mencionar que para todos los algoritmos y todas las instancias, se obtiene una puntuación igual a 1 cuando se compara una molécula consigo misma. Por lo tanto, de aquí en adelante, cuando mencionamos “la molécula con la forma más parecida a una consultada”, excluimos el caso donde la molécula objetivo y la *query* son iguales.

La tabla 1 muestra, para cada *query*, el otro compuesto de la base de datos de la FDA con la similitud de forma más alta (*BestComp*) y la puntuación de función asociada (Tc), de acuerdo con *OptiPharm Robusto*, *OptiPharm Rápido* y WEGA. Como se puede ver, el algoritmo *OptiPharm Robusto* proporciona los valores de similitud de forma más altos Tc , aunque también es el método que consume más tiempo (columna *Time*). Esto significa que se pueden conseguir mejores predicciones usando *OptiPharm Robusto* cuando no hay restricciones de tiempo. Sin embargo, si se requieren tiempos de ejecución inferiores, se deben considerar algoritmos como *OptiPharm Rápido* o WEGA. Aun así, los valores

de tiempo de *OptiPharm Robusto* no son extremadamente altos cuando se comparan con los otros dos métodos. De hecho, teniendo en cuenta la posibilidad de utilizar la computación de alto rendimiento para acelerarla (consulte la sección de Trabajo Futuro), sería perfectamente justificable utilizar el modo Robusto para aumentar el porcentaje de éxito en las predicciones. Por su parte, *OptiPharm Rápido* es el algoritmo más rápido, reduciendo en promedio el coste computacional de WEGA casi 2,5 veces.

Tabla 1. Resultados obtenidos para 7 *queries* de 1751 de la base de datos FDA. Para cada *query*, la molécula de la base de datos (*BestComp*) con el valor de similitud más alto (*Tc*) es motrado, para *OptiPharm Robusto*, *OptiPharm Rápido* y WEGA.

query	OP Robusto			OP Rápido			WEGA		
	BestComp	Tc	Time	BestComp	Tc	Time	BestComp	Tc	Time
DB00529	DB00828	0.921	61	DB00828	0.920	5	DB00828	0.921	16
DB00331	DB01189	0.940	77	DB01189	0.936	6	DB01189	0.940	18
DB01352	DB00306	0.891	117	DB00306	0.884	9	DB00237	0.872	20
DB01365	DB00191	0.944	97	DB00191	0.943	7	DB00191	0.944	17
DB00380	DB00816	0.842	165	DB00816	0.822	11	DB00816	0.842	20
DB06216	DB00370	0.905	169	DB00370	0.902	12	DB09304	0.856	25
DB00693	DB01619	0.841	215	DB01619	0.793	15	DB01068	0.825	24

3.2. Resultados obtenidos para DUD

Para esta base de datos se calculó el área bajo una curva ROC (AUC) para comparar los resultados, como se hizo previamente en otros documentos relacionados [6]. Consulte [17] para obtener una descripción detallada de dicho cálculo.

La tabla 2 muestra los resultados de las pruebas del rendimiento de cribado virtual basado en la forma de *OptiPharm* (en sus dos versiones) y WEGA frente a la base de datos DUD para los 7 primeros blancos farmacológicos en orden alfabético. En el caso de *OptiPharm Robusto* y *OptiPharm Rápido*, se proporcionan valores medios de AUC; mientras que para WEGA, solo se incluye el valor AUC obtenido en cada instancia. Como se puede ver, los valores de AUC obtenidos por los algoritmos son similares aunque algo superiores en las versiones de *OptiPharm*. Por lo tanto, en términos de precisión de las predicciones, todos los algoritmos son equivalentes. Sin embargo, *OptiPharm Rápido* es 2 veces más rápido que WEGA y casi 15 veces más rápido que *OptiPharm Robusto*.

4. Conclusiones y trabajo futuro

En este trabajo se ha presentado el SSM *OptiPharm* que está basado en nuevos enfoques metaheurísticos. Se ha mostrado su rendimiento en términos de precisión de la predicción y en tiempo de ejecución con las bases de compuestos

Tabla 2. Resultados de AUC y tiempo obtenido por las dos versiones de *OptiPharm* y la de WEGA.

name	AUC			Time		
	OpR	OpF	WEGA	OpR	OpF	WEGA
ace	0.39	0.44	0.33	279	15	31
ache	0.71	0.71	0.72	646	36	67
ada	0.67	0.71	0.66	68	5	13
ahr2	0.24	0.28	0.22	87	7	14
ampc	0.70	0.75	0.71	69	5	11
ar	0.73	0.73	0.72	209	18	41
cdk2	0.60	0.58	0.59	184	12	29
mean	0.58	0.60	0.56	220	14	29

DUD y FDA. La comparación realizada con WEGA muestra que *OptiPharm* ofrece la misma precisión predictiva pero con un costo computacional mucho más bajo (la aceleración promedio es de 2x). Otra de las ventajas de *OptiPharm* en comparación con WEGA es que es fácilmente parametrizable por lo que pueden probarse diferentes esquemas heurísticos para adaptarlos a distintas topologías y tamaños moleculares.

OptiPharm se ha diseñado teniendo en cuenta el paralelismo. Obsérvese que cada conformación molecular en la población puede generar una nueva descendencia sin la participación de las conformaciones restantes en la población, lo que significa que el método de *Reproducción* se puede paralelizar fácilmente dividiendo las conformaciones de la población entre las unidades de procesamiento disponibles. Del mismo modo, las conformaciones también se pueden mejorar distribuyéndolas en el procedimiento de *Optimización*. En el futuro, se implementarán y analizarán varios paradigmas de programación basados en arquitecturas de memoria compartida y distribuida.

Referencias

1. Jürgen Drews. Drug discovery: a historical perspective. *Science*, 287(5460):1960–1964, 2000.
2. Xianjun Fu, Lewis H. Mervin, Xuebo Li, Huayun Yu, Jiaoyang Li, Siti Zuraidah Mohamad Zobir, Azedine Zoufir, Yang Zhou, Yongmei Song, Zhenguo Wang, and Andreas Bender. Toward understanding the cold, hot, and neutral nature of chinese medicines using in silico mode-of-action analysis. *Journal of Chemical Information and Modeling*, 57(3):468–483, 2017.
3. Meir Glick, Jeremy L Jenkins, James H Nettles, Hamilton Hitchings, and John W Davies. Enrichment of high-throughput screening data with increasing levels of noise using support vector machines, recursive partitioning, and laplacian-modified naive bayesian classifiers. *Journal of chemical information and modeling*, 46(1):193–200, 2006.
4. Hanna Geppert, Martin Vogt, and Jürgen Bajorath. Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application

- areas, and performance evaluation. *Journal of chemical information and modeling*, 50(2):205–216, 2010.
5. Christopher A Lipinski. Rule of five in 2015 and beyond: Target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions. *Advanced drug delivery reviews*, 101:34–41, 2016.
 6. Xin Yan, Jiabo Li, Zhihong Liu, Minghao Zheng, Hu Ge, and Jun Xu. Enhancing molecular shape comparison by weighted gaussian functions. *Journal of chemical information and modeling*, 53(8):1967–1978, 2013.
 7. Márk Jelasity, Pilar M. Ortigosa, and Inmaculada García. Uego, an abstract clustering technique for multimodal global optimization. *Journal of Heuristics*, 7(3):215–233, 2001.
 8. Márk Jelasity. *The shape of evolutionary search: Discovering and representing search space structure*. PhD thesis, Leiden University, 2001.
 9. Pilar M. Ortigosa. *Métodos Estocásticos de Optimización Global. Procesamiento Paralelo*. PhD thesis, Universidad de Málaga, 1999.
 10. Pilar M. Ortigosa, Juana L. Redondo, Inmaculada García, and José Fernández. A population global optimization algorithm to solve the image alignment problem in electron crystallography. *Journal of Global Optimization*, 37(4):527–539, 2007.
 11. Francisco J. Solis and Roger J.-B. Wets. Minimization by Random Search Techniques. *Mathematics of Operations Research*, 6(1):19–30, 1981.
 12. Xin Yan, Jiabo Li, Zhihong Liu, Minghao Zheng, Hu Ge, and Jun Xu. Enhancing Molecular Shape Comparison by Weighted Gaussian Functions. *Journal of Chemical Information and Modeling*, 53(8):1967–1978, 2013.
 13. Paul Jaccard. Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37:241–272, 1901.
 14. A. Bondi. van der waals volumes and radii. *The Journal of Physical Chemistry*, 68(3):441–451, mar 1964.
 15. David S Wishart, Craig Knox, An Chi Guo, Savita Shrivastava, Murtaza Hassani, Paul Stothard, Zhan Chang, and Jennifer Woolsey. Drugbank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids research*, 34:D668–D672, 2006.
 16. Niu Huang, Brian K. Shoichet, and John J. Irwin. Benchmarking Sets for Molecular Docking. *Journal of Medicinal Chemistry*, 49(23):6789–6801, 2006.
 17. Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

Estrategias de gestión energética en entornos productivos con apoyo de energía solar

Gestión económica de energy hubs con recursos heterogéneos mediante MINLP

Jerónimo Ramos Teodoro (doctorando),
Francisco Rodríguez Díaz (director), Manuel Berenguel Soria (co-director)

Centro mixto CIESOL, Campus de Excelencia Internacional Agroalimentario (ceiA3),
Universidad de Almería, Ctra. Sacramento s/n, Almería 04120, España
{jeronimo.rt,frrodrig,beren}@ual.es

Resumen En los últimos años, el cambio climático y la preocupación por el futuro agotamiento de los combustibles fósiles han favorecido el uso de tecnologías respetuosas con el medio ambiente y el desarrollo sostenible. En este sentido, el aprovechamiento eficiente de los recursos a escala local junto con el uso de energías renovables son factores importantes que contribuyen a reducir el impacto medioambiental del ser humano, por lo que resultan de gran interés desde el punto de vista científico.

El trabajo realizado hasta ahora en el ámbito de la tesis trata del modelado y gestión de un sistema de recursos heterogéneos que incluye agua, dióxido de carbono, electricidad y energía térmica bajo la perspectiva de los energy hubs. El sistema integra una desaladora solar, un invernadero, un edificio bioclimático y un parking fotovoltaico que pueden demandar, proporcionar y/o almacenar los recursos mencionados anteriormente. Así pues, el problema que se plantea consiste en determinar en cada instante de tiempo la producción y almacenamiento de cada recurso que minimizan el coste de operación.

1. Introducción

En la actualidad la mayoría de los países están promoviendo el uso eficiente de energía y agua (especialmente en los sectores industrial, transporte y vivienda) debido a la creciente demanda y a la necesidad de usar fuentes renovables de energía para reducir costes y mejorar la sostenibilidad medioambiental. Generalmente, la eficiencia no se consigue únicamente con mejoras tecnológicas, sino a través de adecuadas estrategias de control y gestión. Este es el ámbito en el que se desarrolla el proyecto ENERPRO [1], dentro del cual se lleva a cabo la actividad investigadora relacionada con la tesis doctoral bajo la concesión de una beca FPI. Se trata de un proyecto coordinado entre el grupo TEP-197 Automática, Robótica y Mecatrónica (ARM) de la Universidad de Almería y el grupo de desalación del CIEMAT-Plataforma Solar de Almería (PSA). El proyecto abarca el análisis, diseño y aplicación de técnicas de modelado, predicción, estimación,

control y optimización para conseguir una gestión óptima de energía (electricidad y calor/frío de proceso), agua y CO₂ en entornos productivos que hacen uso de energías renovables y de sistemas de almacenamiento de energía. A través del control óptimo de estos recursos y mediante la adaptación de la generación a la demanda, se demostrará cómo el control automático permite alcanzar reducción de costes y de impacto medioambiental en la operación de sistemas complejos.

En el futuro de la energía se plantean dos conceptos por los que se ha de caracterizar la red: la producción distribuida y la combinación de múltiples fuentes de energía en una misma planta. Con esto los objetivos que se persiguen son: incrementar la eficiencia en los procesos de conversión de energía (reduciendo, por ejemplo, el número de conversiones desde las fuentes primarias hasta los puntos de consumo), optimizar la producción energética y dotar al sistema completo de flexibilidad y estabilidad. Con el fin de establecer un marco de trabajo que englobe a todos ellos, algunos autores han propuesto conceptos como la multi-generación distribuida (DMG) [2] y, con posterioridad, los denominados sistemas multi-energía (MES) [3], con los que desde una perspectiva más genérica se pretende englobar a todas las metodologías de modelado que comúnmente se aplican en la gestión de recursos energéticos. Aunque en la ref. [3] puede encontrarse una extensa discusión acerca de las características de los MES, no existe aún una definición unánime del concepto. No obstante, algunos autores coinciden en que se trata de la «extensión de la Smart Grid, más allá de la energía eléctrica», es decir, incluyendo otros sectores de la energía.

Originariamente, algunos de los términos surgidos en este ámbito como generación distribuida (DG), cogeneración (CHP), trigeneración (CHCP), poligeneración, se referían casi exclusivamente a la gestión de recursos energéticos. Sin embargo, los MES y la DMG también resultan de interés cuando se habla de gestión de recursos heterogéneos (lo cual incluye recursos tanto materiales como energéticos), dado que son de aplicación las mismas metodologías. Sobre éstos todavía no existen demasiadas publicaciones y por ello es un tema de interés científico, pudiendo constituir en el futuro la evolución lógica del término MES a sistemas multi-recursos (MRS).

Los MES se caracterizan, entre otras cosas, por su aplicabilidad a diferentes escalas y por la integración de sistemas de escala menor en otros de mayor escala. De esta forma, podrían modelarse desde elementos como viviendas individuales hasta regiones o países enteros, pasando por términos intermedios como edificios, industrias, distritos o ciudades. Uno de los enfoques adoptados en este ámbito, junto con el de la smart grid [4], es el de los concentradores de energía, o energy hubs [5,6], término utilizado para designar sistemas que producen, convierten y almacenan diferentes tipos de energía y recursos, sirviendo como interfaz entre productores, consumidores y la propia infraestructura de transporte. La clave de este enfoque reside en su capacidad para modelar sistemas de diversa naturaleza y su aplicación a problemas como diseño y dimensionado de sistemas, planificación de recursos o diseño de redes de distribución [7].

1.1. Objetivos de la tesis y estado actual

Los principales objetivos que persigue esta tesis doctoral son el diseño de técnicas de gestión óptima de recursos (fundamentalmente energía, agua y CO₂) teniendo en cuenta aspectos económicos y medioambientales, en entornos productivos caracterizados por el uso de energías renovables. El desarrollo de estas técnicas de gestión óptima requiere de modelos estáticos y dinámicos con distintos niveles de abstracción y la implementación de técnicas de control y optimización (en el ámbito del control predictivo basado en modelo) para conseguir adecuar la producción a la demanda. Para llevar a cabo esta investigación se determinan varios objetivos particulares:

1. Desarrollo de modelos físicos de sistemas generadores, consumidores y almacenadores de recursos con una filosofía modular y distintos niveles de abstracción que permitan la simulación dinámica a corto, medio y largo plazo a nivel de sistema y su uso en la optimización. Se hará especial énfasis en paradigmas de modelado tipo energy hub.
2. Caracterización de la generación, la demanda y las perturbaciones. Diseño de predictores y estimadores.
3. Validación experimental de los modelos desarrollados en condiciones estacionarias y dinámicas.
4. Aplicación de los modelos diseñados en el desarrollo de estrategias de optimización y coordinación de la producción desde los puntos de vista energético, económico y de seguridad.
5. Desarrollo de estudios de análisis de sensibilidad e incertidumbre usando indicadores basados en la energía y en aspectos económicos en la instalación productiva de referencia. Desarrollo de estrategias de control robustos.
6. Aplicación a las instalaciones del proyecto ENERPRO descritas en el siguiente epígrafe.

Actualmente se dispone de un modelo global de la planta productiva, basado en el enfoque de los energy hub, con el que analizar diferentes escenarios en simulación del reparto de recursos de manera que se satisfaga la demanda de los mismos. Algunos elementos de la planta como las instalaciones de energía solar cuentan con modelos en detalle mientras que para el resto se emplean modelos simplificados o basados en datos (autoregresivos y series temporales) en el caso de las previsiones para la demanda. Asimismo, sobre dicho modelo se ha implementado un algoritmo de optimización basado en programación lineal en enteros mixta (linear mixed integer programming) que determina el reparto de recursos basándose en criterios económicos obteniéndose resultados coherentes.

1.2. Breve descripción de la planta de ensayos de ENERPRO

A modo de planta de pruebas se dispone de un sistema formado por el Centro de Investigación en Energía Solar (CIESOL), un invernadero, un parking fotovoltaico y una desaladora solar que demandan y producen diferentes recursos conforme representa la figura 1.

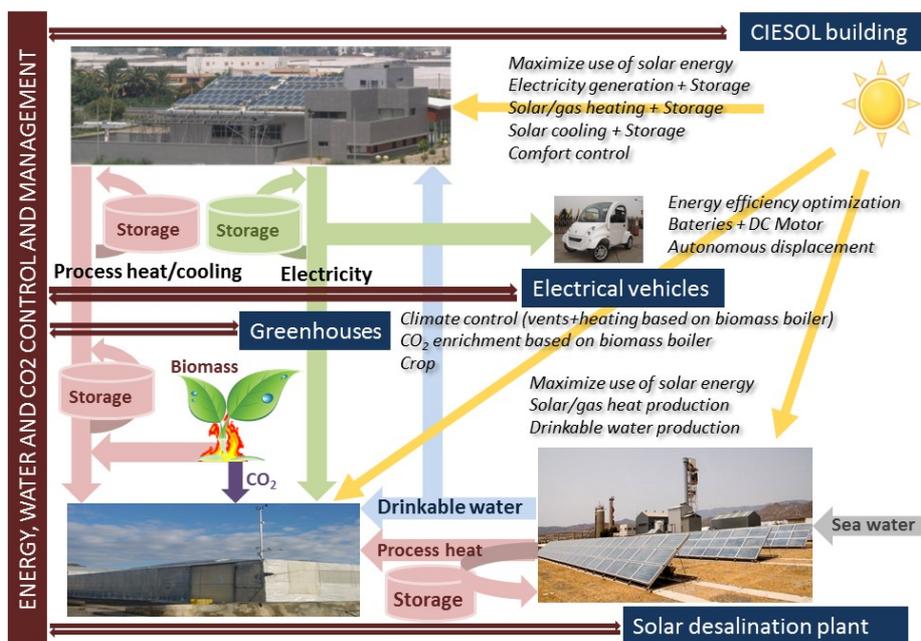


Figura 1. Enfoque y diagrama funcional de la planta de pruebas

Por un lado, CIESOL provee de energía eléctrica fotovoltaica al resto de elementos y de energía térmica mediante el uso de captadores solares. Por otro, tanto en la caldera del invernadero como en la planta desaladora se produce calor a partir de energía renovable y agua potable, en esta última, para abastecer al resto del sistema. Aunque se trata de un conjunto autosostenible, cuenta con suministro hídrico y eléctrico desde la red para ejercer de apoyo. En los siguientes párrafos se detallan los elementos de interés para el propósito de este trabajo, partiendo de la recopilación realizada en trabajos anteriores [8].

El sistema de producción de CIESOL se compone de 42 módulos fotovoltaicos Atersa A-222P, 80 captadores planos Solaris CP1, una máquina de absorción Yazaki WFC SC20 y una bomba de calor reversible Ciatesa Hidripack WE 360. Una descripción detallada del edificio entero puede encontrarse en [9,10].

El sistema AQUASOL de la Plataforma Solar de Almería [11] consiste en una planta desaladora multiefecto (MED) manufacturada por ENTROPIE, que funciona a partir de un campo de 252 captadores CPC Ao Sol 1.12x y el apoyo de una caldera ATTSU RL200 de gas propano.

El invernadero del proyecto ENERPRO [12] está instalado en la Estación Experimental de la Fundación Cajamar (El Ejido, Almería). Se trata de un invernadero tipo parral con una superficie de 877 m² cuyo sistema de calefacción está constituido por un calefactor GP 95 de propano y por una caldera Missouri 150 000 de biomasa.

El parking fotovoltaico de la Universidad de Almería cuenta con 483 paneles CONERGY PA 264P, 24 paneles CONERGY POWER PLUS 240M y 72 paneles FIRST SOLAR FS-380, con una potencia pico total de la instalación de 1176,48 kW y una potencia nominal de 1015 kW.

2. Avances y desarrollo de la tesis

Como parte del desarrollo de la tesis y siguiendo el plan de trabajo propuesto inicialmente, durante los primeros seis meses se ha llevado a cabo la revisión bibliográfica (punto 1, estudios preliminares) del estado actual en gestión de recursos heterogéneos sirviéndose del acceso a bases de datos científicas (Scopus, Elsevier, IEEE entre otras) a través de los servidores de la Universidad de Almería. No obstante, en los meses sucesivos se ha continuado consultando periódicamente dichas bases de datos para mantener la revisión actualizada. Según ese mismo ítem, se ha recopilado la información necesaria de los elementos que conforman la planta (registros de datos u hojas de características).

Con respecto punto 2 del plan de trabajo, a lo largo de la primera anualidad (meses 4 a 20) se ha elaborado un modelo global de la planta de pruebas del proyecto ENERPRO basado en el enfoque de los energy hub 2. Dicho modelo establece las ecuaciones matemáticas presentes entre las entradas y las salidas consideradas que permiten determinar el reparto óptimo de recursos, satisfaciendo las necesidades de demanda. Como ya se ha comentado, para cada uno de

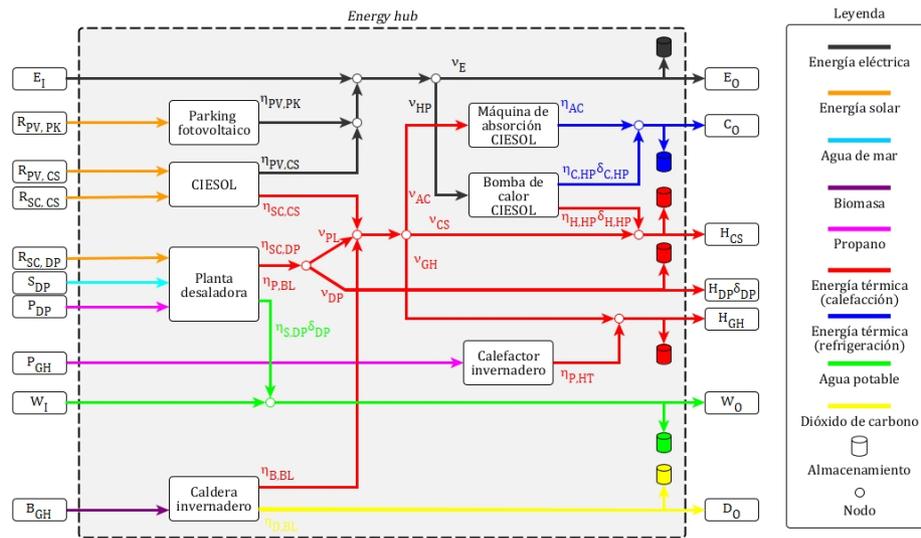


Figura 2. Modelo energy hub propuesto inicialmente para la gestión de la planta ENERPRO

los elementos que conforma el energy hub se han empleado modelos de mayor o menor complejidad para los ensayos iniciales, siendo posible sustituir los modelos ya disponibles por otros que se desarrollen en el futuro. El modelo completo se ha implementado en el entorno de programación MATLAB, aplicando métodos de optimización para obtener el valor de los parámetros y la cantidad de recursos de entrada que, satisfaciendo la demanda, minimicen el coste de operación de la planta en diferentes escenarios de simulación.

El modelado y validación (punto 4, meses 15 a 20) de los componentes primarios está basado en primeros principios, tales como balances de masa, energía y momento, acoplados con correlaciones obtenidas de la literatura científica y curvas de caracterización experimental. Las excelentes capacidades de extrapolación y flexibilidad que ofrecen los modelos basados en primeros principios los convierten en una elección fundamental para lograr los objetivos de optimización del diseño y de la explotación de instalaciones que hacen uso de energías renovables, dado que se requerirá una exploración extensiva del espacio de operación de la planta. Se ha tomado especial atención en el desarrollo, implementación y validación de modelos de las instalaciones con las que se cuenta en el Centro Mixto CIESOL con la finalidad de obtener un conocimiento detallado de los fenómenos físicos que intervienen, así como un alto grado de control del código desarrollado. Los modelos de componentes secundarios se basan en principios fundamentales de la termodinámica de conservación de masa y energía. La solución del sistema resultante de ecuaciones se define matemáticamente como un problema de simulación por computador y consiste en la resolución de ecuaciones algebraico diferenciales acopladas para un conjunto determinado de condiciones de contorno.

En primer lugar, se realizó el planteamiento como un problema MINLP (Mixed Integer Nonlinear Programming) o de programación no lineal en enteros mixta del acrónimo inglés, puntos 3 y 6 de forma preliminar, meses 6 a 12) y se analizaron cuatro escenarios diferentes (punto 5 de forma preliminar, meses 6 a 12) a corto plazo (1 día) distinguiendo entre el tipo de día (frío o cálido) y entre una estrategia basada en MPC en la que se aplica la función de optimización a intervalos de tiempo determinados por un horizonte de predicción deslizante y otra en la que se realiza la optimización del conjunto de datos completo, de la cual ha derivado la realización y defensa del Trabajo Fin de Máster «Gestión energética de un sistema de producción heterogéneo bajo el paradigma energy hub» [8] y ha motivado la asistencia y presentación de resultados en el IV Simposio CEA de Modelado, Simulación y Optimización [13]. Debido a la naturaleza de este planteamiento, los resultados obtenidos no han sido lo suficientemente satisfactorios por dos razones fundamentales: no es posible asegurar que se trabaje con óptimos globales y los tiempos de computación requeridos son demasiados elevados, siendo necesario simplificar algunas restricciones del problema. Por otro lado, el segundo de los factores ha retrasado la evolución de la tesis debido a que para realizar cualquier depuración de código tras contrastar los resultados cada simulación completa conllevaba un tiempo de espera del orden de días en algunas ocasiones, ralentizando la obtención del modelo global.

Debido a estas razones, durante la segunda anualidad, se optó por replantear el problema para formularlo mediante MILP (Mixed Integer Linear Programming) o de programación lineal en enteros mixta (puntos 3 y 6, meses 16 a 20) siguiendo las mismas estrategias de control (pero asegurando que los resultados se correspondan con óptimos globales) y teniendo en cuenta la posibilidad de operar en el mercado eléctrico como productor y consumidor directo (punto 5 con mayor detalle, meses 18 a 20), lo cual supone un enfoque diferente desde el punto de vista económico. De este segundo planteamiento se espera la publicación de un artículo que será enviado a alguna de las revistas IEEE Transactions on Smart Grid o IEEE Transactions on Industrial Electronics, ambas del primer cuartil del JCR y una comunicación (ya aceptada) con el título «Modelado basado en el paradigma de los energy hubs de una explotación agraria bajo invernadero con apoyo de energías renovables» en el I Symposium Ibérico de Ingeniería Hortícola [14].

3. Conclusiones

La contribución más importante que se desprende del trabajo presentado es que se dispone de los elementos suficientes para simular diferentes escenarios de operación. Aunque por razones de tiempo aquí se ha reducido a unos pocos casos concretos el número de pruebas realizadas, podría hacerse funcionar el modelo con conjuntos de datos reales diferentes de los empleados, incorporando modelos de predicción que hagan variar las curvas de demanda en el tiempo conforme se realizan actualizaciones más precisas de los datos, lo cual implicaría utilizar estrategias con horizonte deslizante.

Otra de las particularidades del sistema propuesto es que, pese haberse realizado simplificaciones como la asunción de algunos coeficientes de conversión de los sistemas de almacenamiento o los sistemas de producción por combustión, se contempla la posibilidad de sustituir dichos valores estáticos por modelos de proceso que proporcionen datos variables de manera análoga a como ocurre con el modelo empleado para los paneles fotovoltaicos.

Finalmente, en relación con la programación del modelo y el algoritmo de optimización, un aspecto a tener en cuenta es que al tratarse de un problema de programación no lineal en enteros mixta, es posible que no pueda obtenerse el óptimo global, lo cual justificaría que en algunos casos los sistemas de almacenamiento no se descarguen completamente o que se activen procesos de producción en días en los que la demanda se ha supuesto nula. Así pues, sería interesante realizar el planteamiento del problema mediante programación lineal en enteros mixta y su comparación con los resultados obtenidos, además de tener en cuenta la incertidumbre en las variables de predicción.

Referencias

1. Automatic, Robotics and Mechatronics Research Group (ARM-TEP197): ENERPRO project, <http://www2.ual.es/enerpro/>

2. Chicco, G., Mancarella, P.: Distributed multi-generation: A comprehensive view. *Renewable and Sustainable Energy Reviews* **13**(3) (2009) 535–551
3. Mancarella, P.: MES (multi-energy systems): An overview of concepts and evaluation models. *Energy* **65** (2014) 1–17
4. Farhangi, H.: The path of the smart grid. *IEEE Power and Energy Magazine* **8**(1) (2010) 18–28
5. Geidl, M., Koeppel, G., Favre-Perrod, P., Klöckl, B., Andersson, G., Fröhlich, K.: Energy Hubs for the Future. *IEEE Power and Energy Magazine* **5**(1) (2007) 24–30
6. Geidl, M., Andersson, G.: Optimal Power Flow of Multiple Energy Carriers. *IEEE Transactions on Power Systems* **22**(1) (2007) 145–155
7. Geidl, M., Koeppel, G., Favre-Perrod, P., Klöckl, B., Andersson, G., Fröhlich, K.: The energy hub—a powerful concept for future energy systems. In: *Third annual Carnegie mellon conference on the electricity industry*. Volume 13. (2007) 14
8. Ramos-Teodoro, J.: Gestión energética de un sistema de producción heterogéneo bajo el paradigma energy hub. Trabajo Fin de Máster, Universidad Carlos III de Madrid (2017)
9. Rosiek, S., Batlles, F.J.: Renewable energy solutions for building cooling, heating and power system installed in an institutional building: Case study in southern Spain. *Renewable and Sustainable Energy Reviews* **26** (2013) 147–168
10. Castilla, M.d.M., Álvarez, J.D., Rodríguez, F., Berenguel, M.: *Comfort Control in Buildings*. Springer (2014)
11. Alarcón Padilla, D.C., Blanco Gálvez, J., García Rodríguez, L., Gernjak, W., Malato Rodríguez, S.: First experimental results of a new hybrid solar/gas multi-effect distillation system: the AQUASOL project. *Desalination* **220**(1-3) (2008) 619–625
12. Sánchez, J.A., Rodríguez, F., Guzmán, J.L., Arahál, M.R.: Virtual sensors for designing irrigation controllers in greenhouses. *Sensors* **12**(11) (2012) 15244–66
13. Ramos-Teodoro, J., Álvarez, J.D., Rodríguez, F., Berenguel, M.: Gestión económica de energy hubs con recursos heterogéneos mediante MINLP. In: *IV Simposio CEA de Modelado, Simulación y Optimización*, Universidad de Valladolid (2018) 78–83
14. Ramos-Teodoro, J., Rodríguez, F., Berenguel, M.: Modelado basado en el paradigma de los energy hubs de una explotación agraria bajo invernadero con apoyo de energías renovables. In: *I Symposium Ibérico de Ingeniería Hortícola*, Universidad de Santiago de Compostela - Campus Terra (2018)

Contribution of Models in the Disease and Energy Management in Intensive Growing System

Ph.D. candidate: Hui Wang

Centro Mixto CIESOL, University of Almeria
jorgesanchez@ual.es
hw646@ual.es

Abstract. Early-warning model for plant diseases is valuable when the systems provide timely forecast and warning. Using information about weather, crop, and/or pathogen, warning systems advice growers or other crop managers when they need to take an action to apply a fungicide or bactericide spray to prevent disease outbreaks and make sure quality of vegetable. This article shows the modelling input parameters from a sensor-based tomato disease warning model and a disease incidence forecast system, as well as performance of the models was evaluated using multi simple criteria. Finally, a risk management for disease provide precision agriculture (PA) services with large-scale, long-term, scalable and real-time data collection capabilities for solar greenhouse vegetables.

Keywords: Air temperature, relative humidity, leaf wetness duration, transpiration, pathogen, infection.

1 Introduction

High quality crops require much care in disease management, yields and high visual inspection standards are giving priority to for production. Disease breakout is main reason of low yield and below market standards caused by warm, humid and wind-free environment. Furthermore, complex interaction of crop, soil and environment makes the control of pests and diseases a great challenging, with many important problems unresolved and new ones arising as the industrial changes in production systems. Immoderate spraying of pesticides becomes a straw for grower trying to solve the heavy yield loss. Along with pesticides costs increase, a strong demand for non-chemical control methods is emerging in many countries.

A well-known model was mentioned in 1978 as a tomato blight disease forecast system published to assess disease severity and supply an efficient fungicide application schedule. Environment conditions had been proved useful to identify disease

severity, such as average air temperature, relative humidity and rainfall. Although successful feedback of the model effectively was got from commercial growers to control early blight as well as a weekly fungicide program but with substantially fewer sprays, a poor response from local farms was awkward that were short of sensors and meteorological station. To be acceptable to growers, a simple model was developed based on modified the program, called TOM-CAST could be used control leaf spot [1]. Cylindrical wetness sensor invention and the TOM-CAST model developed realized the disease warning system popularization in large area. However, some of limitations are existent. 1) The forecasting result is not outstanding for some pathogens. 2) It cannot forecast disease spread effectiveness. 3) It ignores disease development.

A fungal population consists of individuals at various stages of their life-cycle, which pass from one stage to the next can be very fast, usually depending on environmental conditions (Figure.1). Temperature and leaf wetness duration are important cases of disease onset. The effect of temperature (T) on spore liberation is probably largely associated with its influence on spore maturation. Elliott [2] has demonstrated that, conidia discharge can occur over the range 5°C-37°C. Leaf wetness duration (LWD) is a key for researching plant disease due to its significance on pathogen surviving and spreading. Under semi-arid conditions of climate, the duration of dew may be one of the most important source of LWD [3]. Based on these principles of wetting formation, transpiration is also a factor caused by increasing air humidity, which was ignored.

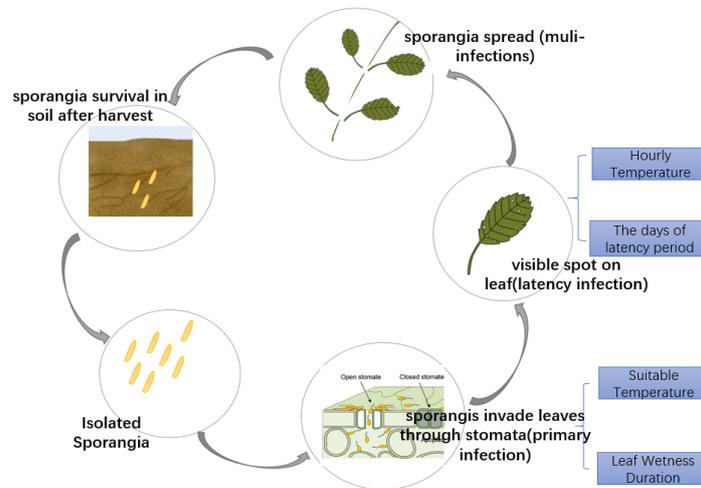


Figure.1 Disease infection flow diagram

Excepting environmental threshold conditions of disease onset, some growth curves are commonly used to describe disease pathogen progress. Those curves can be classified under two broad headings: those that assume that an asymptotic final size exists and those that do not [4]. Inoculum population may approach an asymptote because of limiting plant resources or climate change. Choosing between asymptotic

and non-asymptotic functional forms depends in part upon the response variable of inoculum progress.

Reliable, precise and accurate estimates of disease incidence are important for predicting yield loss, monitoring and forecasting epidemics, for assessing crop germplasm for disease resistance, and for understanding fundamental biological processes including coevolution [5]. The crop growth monitoring system of the European Commission (CGMS, <https://ec.europa.eu/jrc/en/mars>) relies on the WARM crop model [6] to perform in season rice yield forecasting at country level since years. The CGMS currently is based on the Biophysical Models Applications framework (BioMA, <https://en.wikipedia.org/wiki/BioMA>), which allows to develop modelling solutions grounded on software independent components simulating different domains of the agricultural systems [7]. This technology maximizes the reusability and favour the integration of domain-specific research for custom applications. Recently, a new set of components targeting the simulation of the yield losses due to fungal plant diseases was released (Disease, [8]) allowing the coupling with crop models to dynamically simulate the interactions between the fungal.

The objective of this thesis would contribute on constructing a propagable early disease warning and disease incidence modelling system with a smart disease management scheme (Figure.2). Using the measure data from greenhouse and optimize algorithm, training and improving performance of the models to be realize accurate warn the date of disease onset and forecast disease incidence. The parameters of disease warning model to be controlling set-point for adjustment environment in greenhouse. Reappraise risk of disease onset and incidence using data feedback from controlling system. The information provide precision agriculture (PA) services with reference of pesticide schedule for greenhouse management.

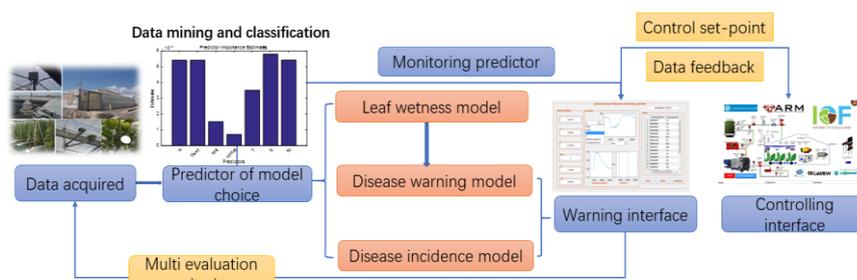


Fig.2 a design approach of disease warning system

2 Objective

The objectives of the present thesis can be divided into two parts. On the one hand, development and construction concept model using optimized input parameters by data mining and classification method (leaf wetness model, disease warning model and disease incidence model), as well as improving preferment of the models by train-

ing acquired data. On the other hand, coupling warning information and control scheme to be realize risk management for disease.

Firstly, the most successfully used disease cycle component in forecasting is infection conditions. Temperature and leaf wetness duration are the main factors lead to pathogen infection. Thus, we have to finish some work are described as below.

1. Development an accuracy leaf wetness model to be the input parameters of disease warning model.
2. Training and Improving performance of the disease warning model through evaluated dataset acquired in greenhouse, as well as researching the correlation of disease incidence and environmental condition. Disease incidence was simulated using curve fitting method.

Next step is developing a control scheme considered the pathogen infection condition to decrease probability and delay disease onset. The objectives we have to do:

1. Study the regular of effect on variable by control device.
2. Optimize control strategy to realize adjustment of pathogen infection condition.
3. Test and evaluate the effectiveness of controller on disease onset.

3 Research progress

This part introduces how is going for the objectives. For the first objective, some work we had done as below:

1. *Leaf wetness model was constructed using neural network with optimized input parameters by data classification and mining.*

We used a machine-learning algorithm based on decision tree for selecting input parameters of model. Decision tree learning is a method commonly used in data mining and classification (Breiman, 9). A decision tree algorithm works by splitting a data set in order to train a model through a recursive partitioning process, and then the model is used to predict the value of a target variable based on the independent variables. Predictor importance for target variable was estimated by summing these estimates over all weak learners in the ensemble. A high value indicates that this predictor is important for the target. In this case, relative humidity, dew temperature, transpiration and globe radiation were the input parameters of leaf wetness model.

Artificial neural networks (ANN) are a form of computing for predicting a set of outputs from a set of input patterns (Zhu, 10). It is similar with central nerves system that could be learn task by considering examples. Leaf wetness to be as the task we want to learn using ANN with greenhouse climate conditions including relative humidity (H), dew temperature (Tdew), transpiration (E) and solar radiation (Ra) which were selected by decision tree learning. Through dealing with amount of data, a self-adaption model was used to estimate LWD.

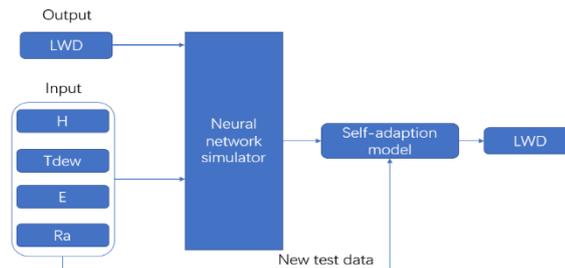


Fig.3 Diagram of MISO neural network modelling

2. *Finish disease warning model construction using climate threshold method and logistic curve. Temperature and leaf wetness duration could be used input parameters of the model. Verify effectiveness of disease warning model based on Bayesian theory. At the same time, building disease incidence model which the relationships of disease incidence and temperature response curve of pathogen was considered.*

Realization of the modelling solution is to get alarm and disease incidence. In this case, the simulation starts on September 23th (bed out date) with threshold of infection conditions and accumulation of the temperature contribution needed to trigger the disease onset, which in line with environmental conditions of tomato in Almeria, Spain. The starting in this model is realized to get data needed of experimental infection of susceptible crop leaves under controlling conditions. The first coupling point between two models starts to simulate impact of leaf spot on leaves of host, by the number of lesion increasing with accumulation of effective temperature. The two models are driven by the agrometeorological variables which are LWD, hourly mean temperature in leaf wetness duration (T_{LWD}), hourly mean temperature (T_h), minimum temperature for infection (T_{min}), maximum temperature for infection (T_{max}) and optimum temperature for infection (T_{opt}). Temperature contribute to crop growth, it drives pathogen infection as well as. Once alarm date was got from the disease warning model, disease incidence was started simulated to evaluate the impact level of leaf spot, which could be guidance of pesticide dose and schedule to loss minimization of yield.

In the second part, we try to do some test of ventilation, humidification, dehumidification and heating devices which is to get the best control strategy was used to ad-

just the conditions of disease onset. Multi-variable control is an ideal we want to choose to realize the controlling of multi inputs (ventilation, humidification, dehumidification and heating devices) and multi outputs (temperature and leaf wetness duration). At the moment, it is initial phase.

4 Conclusion

In the present doctoral thesis is intending to contribute on in the field of disease onset warning and risk management, with the purpose of helping greenhouse manager make a decision timely to avoid loss economic.

So far, the contribution has been done that an optimized leaf wetness model, an evaluated disease warning model, a trained and verified disease infection model using the data of observed date of disease onset, recorded disease incidence and climate parameters from 2016 to 2017. The results are completed in a disease warning system which was constructed by MATLAB GUI tool to provide precision agriculture (PA) services with real time, long time and accurate disease data.

References

1. Gleason, M. L. Disease warning system. *Plant Disease*, 79(2), 113 (1995).
2. Elliott V.J. Response models for conidiospore germination and germ tube elongation of *Mycosphaerella fragariae* as influenced by temperature and moisture. *Phytopathology*, 78, 645–650. (1988).
3. Lomas, J., & Shashoua, Y. The performance of three types of leaf-wetness recorders. *Agricultural Meteorology*, 7, 159–166. (1970).
4. Paine, C. E. T., Marthews, T. R., Vogt, D. R., Purves, D., Rees, M., Hector, A., & Turnbull, L. A. How to fit nonlinear plant growth models and calculate growth rates: An update for ecologists. *Methods in Ecology and Evolution*, 3(2), 245–256. (2012).
5. Bock, C. H., Poole, G. H., Parker, P. E., & Gottwald, T. R. Plant Disease Severity Estimated Visually, by Digital Photography and Image Analysis, and by Hyperspectral Imaging. *Critical Reviews in Plant Sciences*. (2010).
6. Confalonieri, R., Bellocchi, G., Donatelli, M. A software component to compute agro-meteorological indicators. *Environ. Model. Softw*, 25, 1485–1486 (2010).
7. Confalonieri, R., Rosenmund, A.S., Baruth, B. An improved model to simulate rice yield. *Agron. Sustain. Dev*, 29, 463–474 (2009).
8. Bregaglio, S., & Donatelli, M. A set of software components for the simulation of plant airborne diseases. *Environmental Modelling and Software*, 72, 426–444 (2015).
9. Breiman, L., Friedman, J., Stone, C.J., Olshen, R. A. *Classification and Regression Trees*. CRC press. (1984).
10. Zhu, A. Artificial Neural Networks. The. *International Encyclopedia of Geography*. (2017).

SECCIÓN II: Otros trabajos

Una metodología basada en modelos y servicios para la interoperabilidad de aplicaciones heterogéneas en televisión digital

Darwin Alulema¹

¹ Grupo de Investigación de Informática Aplicada (TIC-211), Departamento de Informática, Campus de Excelencia Ceia3 y CEIMAR, Universidad de Almería
doalulema@espe.edu.ec

Abstract. En un mundo interconectado, donde las personas y las cosas interactúan cada vez más, resulta muy importante desarrollar tecnologías que puedan hacer uso de todos los medios ya desplegados en las ciudades, para proveer servicios que faciliten la vida de las personas, como lo está haciendo Amazon, al desplegar sus servicios de ventas online asociado a dispositivos como Alexa o su Botón de compra. En este trabajo se establece una metodología que facilite la etapa inicial de desarrollo de software, al determinar de una forma general los actores y sus relaciones. Se ha considerado la TV digital, al ser una tecnología que se encuentra difundida en la mayoría de hogares, pero que no ha tenido acogida sus nuevas características de conectividad, las cuales pueden ser de mucha utilidad para el despliegue de servicios masivos de uso público. Para esto se ha realizado un análisis de los estándares de TV digital, para determinar nuevos elementos no contemplados en las normativas que permiten esta interconexión. Además, como entorno de pruebas se consideran escenarios de T-Health, debido a que es un entorno que interesa no solo a las personas sino que a las instancias públicas, ya que a partir de esta información se pueden proponer políticas de prevención.

Keywords: MDE, TV Digital, IoT, T-Health.

1 Introducción

El rápido desarrollo de la tecnología ha permitido que esta se inserte en la vida cotidiana de las personas, impulsando el desarrollo de dispositivos inteligentes y plataformas de gestión que habilitan colectivamente el “Mundo inteligente”, en el que muchos pequeños e inteligentes objetos heterogéneos están conectados. Uno de los principales problemas para el desarrollo de un mundo interconectado es la creación sencilla de aplicaciones interactivas que combinen varios dispositivos heterogéneos, y que puedan trabajar a través de diferentes tecnologías de conexión. Por esta razón muchos esfuerzos se centran en establecer conectividad e integración entre objetos inteligentes, pero no hay todavía ninguna arquitectura capaz de proporcionar un acce-

so flexible, fácil y seguro a sus recursos (Mainetti, Mighali, & Patrono, 2015) (5G Américas, 2016).

Debido a este desarrollo han surgido dos problemas: la complejidad del desarrollo de software y la fragmentación de las plataformas. Una solución podría ser aplicar los conceptos de multiplataforma (Ribeiro, 2014) y MDA para permitir reducir la diferencia en el alcance del problema y la implementación de software mediante el uso de tecnologías para la transformación de modelos (Benouda, Essbai, & Moussaoui, 2016).

Para resolver el problema de la comunicación de aplicaciones heterogéneas en un entorno convergente, (Gonzales, Pascual, Nuñez, & Garciun, 2014) incorpora la idea de Internet of Things (IoT) e identifica dos problemas: 1) la interconexión de objetos heterogéneos requiere un protocolo estándar para la comunicación, y 2) para la creación de aplicaciones debería considerar la plataforma, los protocolos, las API, los lenguajes de programación y el alcance del problema. Además, para la comunicación de plataformas y la interacción con los usuarios, se requiere inteligencia, conectividad e integración integradas.

Para construir software se debe observar al sistema como un producto complejo y a su construcción como un proceso ingenieril. Tom Demarco, introdujo el concepto de desarrollo de software basado en modelos o MBD (Model Based Development) en el que destaca que la construcción de un sistema de software debe ser precedida por la construcción de un modelo, tal como se realiza en otras ingenierías (Pons, Giandini, & Pérez, 2010). Es por esto que describir los sistemas a un nivel que permitan la reutilización de sus datos y procesos de negocio (de forma independiente de los lenguajes de programación, la tecnología subyacente y las plataformas), facilita la interoperabilidad con otros sistemas externos (Durán, Troya, & Vallecillo, 2017).

Además, el alto grado de desarrollo de los dispositivos wearables destinados al bienestar físico y cuidado de la salud, como Monitores de presión arterial, Monitores de glucosa, Audífonos, Desfibriladores, Monitores de electrocardiograma (ECG), Bombas de insulina, Pulsioximetría, etc., los cuales son de precios accesibles y de fácil adquisición. Esto permite que se puedan utilizar de forma masiva para recopilar información muy valiosa para el cuidado de la salud (5G Américas, 2016).

Un entorno que tiene especial atención es el cuidado de la salud y un caso particular es el T-Health que incluye al Televisor en el ámbito de la salud, con la ventaja de ser el televisor, el dispositivo central de interacción y comunicación, y que está presente en la mayoría de los hogares en países desarrollados países (Rivas, Anido, & Fernández, 2015) por esta razón se convierte en un elemento de gran importancia para el desarrollo de los sistemas destinados a los cuidados de la salud, ya que no solo permite imagen de alta calidad sino que también incorpora la posibilidad de interactividad.

Se propone en esta tesis doctoral una metodología basada en modelos y servicios para la interoperabilidad de aplicaciones heterogéneas en televisión digital, permitiendo que la TV, los teléfonos móviles, la web y los dispositivos wearables (Nodos IoT) se integren, de modo que los dispositivos físicos no sólo interactúan con su propia interfaz sino que puedan hacer uso de las fortalezas que cada una de las plataformas posee. Además, se pretende realizar propuestas que complementen la norma técnica del estándar ISDB-Tb de TV digital, para el desarrollo de software.

Una metodología basada en modelos y servicios para la interoperabilidad de aplicaciones...

El resto del documento está estructurado de la siguiente manera. La sección 2 describe los avances realizados al momento en lo referente al escenario de prueba y el modelo. Finalmente, la Sección 3 presenta conclusiones.

2 Avances

2.1 Modelo y Arquitectura propuestos

Para el desarrollo de la propuesta se propone dos fases:

Fase 1.- Identificación de puntos de referencia.- Para identificar las partes de congruencia es necesario identificar los requisitos que se recogen en cada una de sus tecnologías. Para esta tarea se establecen dos objetivos: 1) Realizar una comparativa a nivel de características, y 2) Verificar el cumplimiento de las características. Para esto, es necesario revisar las normativas y recomendaciones emitidas por las entidades reguladoras, así como de las empresas.

Para este trabajo se consideró a la TV Digital, Smart Phone, SmartTV y los dispositivos Wearables, de tal manera que para identificar las características de cada una de las plataformas se ha tomado los siguientes puntos de referencia:

- Comunicación entre sistemas.- En este punto se ha analizado las tecnologías que la plataforma dispone para conseguir Interoperabilidad. Aquí se ha considera el modelo OSI, para identificar que protocolo pueden emplear los sistemas para su comunicación.
- Seguridad.- Hay tres pilares fundamentales para proteger los dispositivos, en los cuales se deben emplear esquemas de cifrado y autenticación: 1) Seguridad del dispositivo, 2) Seguridad de la nube y 3) Gestión del ciclo de vida de la seguridad. El primero le corresponde al usuario, mientras que el segundo y tercero dependen del sistema, en virtud de que, es en esta etapa donde se deben implementar las políticas de autenticación e implementar los mecanismos de cifrado de la información, de esta manera que la información no circule en texto plano.
- Desarrollo de aplicaciones.- Aun cuando existen una gran variedad de Herramientas de desarrollo para cada plataforma y sistemas operativos. Se considera esencial emplear Tecnologías híbridas, como HTML5 o JavaScript.
- Arquitectura de Capas.- La arquitectura más amplia que cubre la mayoría de los ámbitos de cada una de las aplicaciones es el modelo OSI.

Fase 2.- Definición de una arquitectura y un modelo.- Para definir la arquitectura y el modelo, se considera un escenario en un nivel superior de abstracción, de tal manera que se defina el comportamiento general que pueden tener los actores en un escenario. De esta manera se crea un modelo que para este caso se ha considerado el UML y un esquema de arquitectura modular.

Una metodología basada en modelos y servicios para la interoperabilidad de aplicaciones...

Para agrupar las distintas arquitecturas y características de plataformas se propone una arquitectura de tres capas, que se aprecia en la Figura 1, que puede ser empleada para el despliegue del IoT, al abarcar tanto sistemas de hardware como de software.

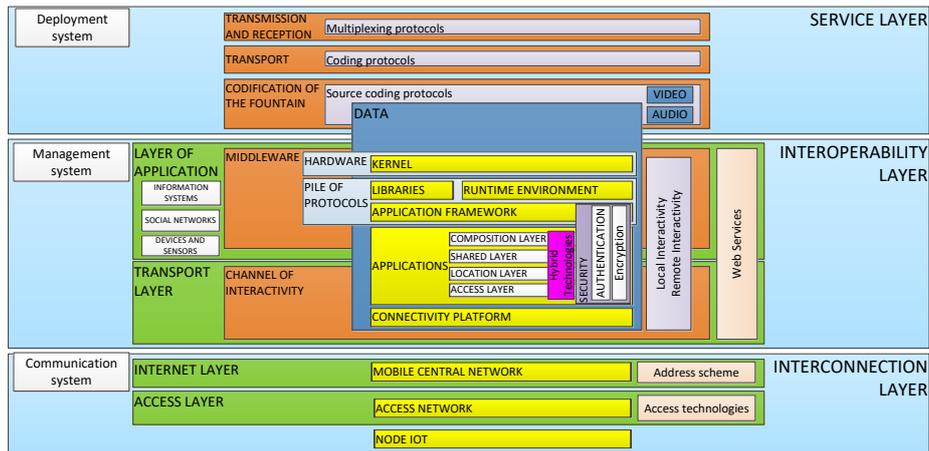


Fig. 1. Vista general de la Arquitectura de integración

Capa de Servicio.- Considera principalmente lo referente a sistemas de broadcasting como la TV, abarcando todo el proceso de codificación, multiplexación y construcción de carrusel de datos e incorporación a la etapa de radio frecuencia y transmisión. Se aplican protocolos como OFDM (Orthogonal Frequency Division Multiplex) para codificar la señal al medio de transmisión. Aquí es importante considerar que también es susceptible a ataques (Bachy et al, 2015) por medio de una transmisión ilegal sobre la frecuencia portadora asignada a la estación de televisión.

Capa de Interoperabilidad.- Considera los elementos comunes a la mayoría de plataformas, para su acceso a sistemas que emplean protocolos HTTP, TCP y UDP. De esta manera es factible emplear Arquitecturas basadas en servicios para el desarrollo de aplicaciones, con lo que se logra independencia de la plataforma. Además se puede implementar esquemas de cifrado como SSL, TLS y HTTPS, que permitan asegurar la información que se envía principalmente cuando se tienen aplicaciones que requieren interactividad remota. Adicionalmente en esta etapa se debe establecer los mecanismos de autenticación.

Capa de Interconexión.- Es equivalente a las dos primeras capas de modelo OSI, lo que permite el empleo de todas las tecnologías de Acceso y de Red. Esta capa se encuentran protocolos propios para el despliegue del IoT, como es WiFi, ZigBee, Lora, Bluetooth Low Energy, LowPAN, COAP y MQTT. En la actualidad aun cuando los sistemas de acceso inalámbrico son los más comunes, estos también son los más inseguros y susceptibles a ataques, por lo que es necesario protocolos que permitan garantizar la integridad de la información, como el protocolos 802.11i.

Para definir cada uno de los sistemas de la arquitectura se ha procurado separar en diferentes áreas. A continuación se describe los elementos relevantes del sistema y las interacciones de los usuarios:

Una metodología basada en modelos y servicios para la interoperabilidad de aplicaciones...

Subsistema de Codificación y Transporte.- Cubre todo el proceso por medio del cual una estación de Televisión y el proveedor del servicio, codifican la señal de audio, video y datos, a una señal que pueda ser procesada. A continuación se construye el carrusel de datos. En esta etapa se incorporan todas las tramas de control a la señal digital, por medio de un servidor Payout. Por último la señal digital es modulada a la frecuencia del canal por el que será transmitida la señal multiplexada.

Subsistema de recepción.- Depende del middleware que disponga el Hardware del receptor de Televisión, de acuerdo a este, se podrá únicamente observar la señal de video y audio, o se podrá ejecutar la aplicación. Además, si se dispone de conexión a la red de Internet se podrá ejecutar la aplicación con interactividad remota, al consumir los servicios de la aplicación.

Subsistema de interactividad.- Este subsistema describe todo el proceso que permitirá el consumo de un Web Service, que actúa como interlocutor con el servidor del proveedor de la aplicación. La información que aquí se recupera es enviada por medio del canal de interactividad para la visualización únicamente al usuario televidente que la solicite.

Subsistema de acceso a la red.- Es este subsistema se describe la arquitectura de red y los nodos intervinientes. En este caso podrán presentarse redes de área personal (PAN), redes de área metropolitana (MAN) o redes de área extendida (WAN). Además, se determina el mecanismo de acceso para que los nodos puedan enviar y recibir información.

Subsistema de seguridad.- Permite que se validen las credenciales de los usuarios, para que de esta manera se pueda discriminar y asegurar la información del sistema. Para que la autenticación se efectúa es recomendable tres fases secuenciales: 1) Fase de generación de claves, 2) Fase de establecimiento: esta fase implica establecer ella identidad del dispositivo después de generar la clave secreta, y 3) Fase de implementación: esta fase final otorga control de acceso a dispositivos autenticados para comunicarse entre ellos. Aunque el modelo no previene completamente los ataques DoS, reduce el riesgo porque el acceso a los recursos se concede a una sola ID en un tiempo (Ammar, Russello, & Crispo, 2018).

Para el diseño del modelo se considera las recomendaciones de la W3C, que menciona el modelo orientado a servicios (SOM), el cual se centra en los aspectos de la arquitectura que se relacionan con el servicio y la acción. SOM explica las relaciones entre un agente y los servicios que proporciona y solicita (W3C, 2004). En la Figura 2, se presenta el modelo propuesto, enfocado a los componentes esenciales que se pueden encontrar en una implementación de IoT.

Una metodología basada en modelos y servicios para la interoperabilidad de aplicaciones...

- Proveedor de Aplicaciones: Desarrolla servicios finales para los usuarios, que los pueden consumir en alguna de las plataformas propuestas. Además, determina los esquema de seguridad que se implementarán.
- Proveedor de Plataforma: Desarrolla servicios intermedios los cuales pueden ser consumidos por otros proveedores para desarrollar aplicaciones finales. Además las plataformas que puede ofertar son dos:
 - Hardware: Constituye toda la infraestructura de hardware y Nodos IoT, desplegados por el proveedor para proveer un servicio, la cual puede ser accedida, si es necesario por el usuario.
 - Software: Constituye todos sistemas de software desplegados por el proveedor para proveer un servicio, los cuales pueden ser accedidos, si es necesario por el usuario.

Nodo IoT.- Es un actor no humano que puede encontrarse del lado del usuario, cuando este consume un servicio final y aporta con su información, para tener una experiencia más enriquecida, o del lado del proveedor cuando este despliega una serie de estos para ofertar un servicio que pueda ser consumido por el usuario. Puede cumplir el rol de sensor o actuador, o los dos a la vez. El Nodo IoT, puede conectarse a la red del sistema de dos formas:

- Red PAN: es el escenario en el cual el nodo IoT se conecta por medio de un intermediario “Portal”, el cual es el encargado de conectarse al servicio. Establece una Red de Área Personal.
- Red LAN/WAN: es el escenario en el cual el nodo IoT se conectar directamente al servicio, pero con las credenciales del usuario propietario. Puede establecer un Red de Área Local o de área Extendida.

2.2 Sistema de prueba

Las principales áreas de la salud en las que se utilizan las TIC y que puede presentar mejores prestaciones con el uso de la TV , se pueden clasificar de la siguiente manera (Mainetti, Mighali, & Patrono, 2015): Alarmas, Consulta remota, Monitoreo de signos vitales, Vigilancia de las actividades de la vida cotidiana, Detección de caídas, Detección Temprana de Enfermedades, Ejercicios en casa, Apoyo a los cuidadores humanos, Acompañamiento y entretenimiento, Robots en el hogar, La TV como puerta de acceso a los Servicios Sociales y de Salud, y Análisis de métricas. Todas estas constituyen ámbitos de aplicación a los cuales el IoT puede hacer un gran aporte. Para el sistema de prueba se han considerado únicamente la Consulta remota, monitoreo de signos vitales, y análisis de métricas.

La Figura 3, ilustra la arquitectura de pruebas, en la que se crea un circuito cerrado de TV, donde el servidor VillageFlow versión 2, instalado en un servidor de Payout, actúa como la estación de televisión, combinando la señal que se envía por cable coaxial a un decodificador de TV digital que emplea el estándar ISDB-Tb. Este decodificador tiene embebido el Middleware Ginga para la construcción y ejecución de la aplicación, además, por medio de un cable UTP se conecta al router para el acceso a Internet. En este caso el Nodo Iot del modelo está encapsulado físicamente dentro del móvil, el cual actúa como “Portal”, para que el sensor pueda enviar los datos. El comportamiento del sistema considera las acciones que el usuario puede realizar al acce-

Una metodología basada en modelos y servicios para la interoperabilidad de aplicaciones...

der desde el Televisor a la programación y si dispone de conexión a Internet puede acceder a información personalizada. Esta información se presenta en forma resumida en la interfaz de televisión y de forma amplia en la interfaz nativa del móvil.

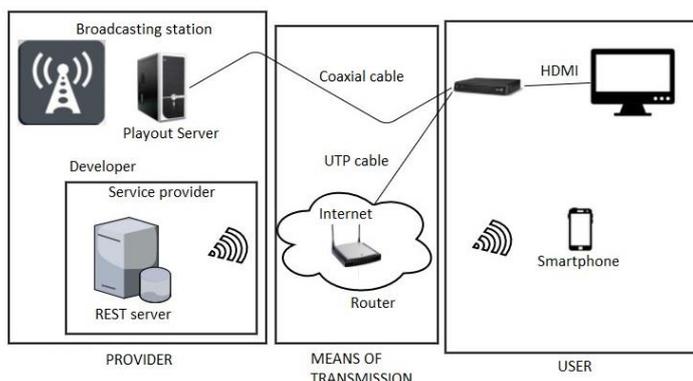


Fig. 3. Arquitectura de pruebas

La autenticación se realiza por medio de la combinación de usuario y password, o por medio de un código Token. Considerando que la interfaz de TV es menos amigable con respecto al ingreso de caracteres alfanuméricos, el código Token es utilizado para la autenticación. Para la etapa de recepción se utilizó un decodificador marca EITV, que posee compatibilidad completa con los estándares declarativos (Ginga-NCL / Luna) y el imperativo (Ginga-J / JavaDTV) que se han especificado para el Sistema Brasileño de Televisión.

En la Figura 4 se muestra la ejecución del sistema en una TV digital, y el despliegue del esquema de autenticación por Token, además la ejecución en un teléfono móvil, que muestra los datos almacenados y la interfaz de captura de los datos del sensor de movimiento, que cuenta los pasos y la distancia caminada.

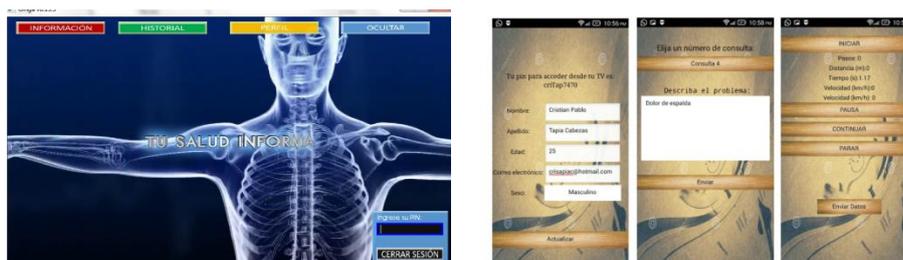


Fig.4. Interfaz del sistema: a) Tv digital, y b) teléfono móvil

3 Conclusiones

Se ha realizado un análisis de las tecnologías mayormente intervinientes en nuestras vidas y que están siendo aplicadas para el despliegue del Internet de las cosas, y además se ha observado la necesidad de tener procesos que permitan el desarrollo de

Una metodología basada en modelos y servicios para la interoperabilidad de aplicaciones...

sistemas que empleen diferentes plataformas y tecnologías. Es así que en este trabajo se propone una metodología que facilite el desarrollo de aplicaciones para el Internet de las cosas. En este sentido en esta primera instancia se propone una primera aproximación de una arquitectura y un modelo, que agrupa estas tecnologías en una nueva categorización, relacionada con un modelo presenta a los actores y sus relaciones. Además se ha creado un escenario de prueba en el entorno del T-Health que sirve para realizar un proceso de retroalimentación, con lo que se pretende depurar el modelo.

Referencias

- 5G Américas. (2016). *Internet de las Cosas en América Latina*.
- Alulema, Iribarne, & Criado. (2017). A DSL for the Development of Heterogeneous Applications. *2017 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*.
- Ammar, Russello, & Crispo. (2018). Internet of Things: A survey on the security of IoT frameworks. *Journal of Information Security and Applications*, 8–27.
- Bachy et al. (2015). Smart-TV Security Analysis: Practical Experiments. *Proceedings of the International Conference on Dependable Systems and Networks*, 497–504.
- Benouda, Essbai, & Moussaoui. (2016). Modeling and Code Generation of Android Applications Using Acceleo. *International Journal of Software Engineering and Its Applications*, 83-94.
- Durán, Troya, & Vallecillo. (2017). *Desarrollo de software dirigido por modelos*. Barcelona: Universitat Oberta de Catalunya.
- Gonzales, Pascual, Nuñez, & Garciun. (2014). Midgar: Domain-Specific Language to generate Smart Objects for an Internet of Things platform. *Eighth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*.
- Mainetti, Mighali, & Patrono. (2015). A Software Architecture Enabling the Web of Things. *IEEE INTERNET OF THINGS JOURNAL*, 2(6), 445-453.
- Pons, Giandini, & Pérez. (2010). *Desarrollo de software dirigido por modelos. Conceptos teóricos y su aplicación práctica*. (1 ed.). La Plata: Universidad Nacional de la Plata.
- Ribeiro, R. d. (2014). XIS-Mobile: A DSL for Mobile Applications. *29th Symposium On Applied Computing*.
- Rivas, Anido, & Fernández. (2015). An Open Architecture to Support Social and Health Services in a Smart TV Environment. *IEEE Journal of Biomedical and Health Informatics*.
- W3C. (2004). *Web Services Architecture Usage Scenarios*. Obtenido de <https://www.w3.org/TR/ws-arch-scenarios/>

High Performance Computing applied to model eco-evolutionary agent based systems

Recent improvements on Weaver

Gabriel Barrionuevo Rosales

Abstract. This paper presents the recent improvements on Weaver, an agent based simulation[5] tool specifically designed to study ecological networks. Each simulation includes hundreds of thousands of individuals (or agents) moving and interacting both with the environment and with other agents. Each individual from each species contains its own genetics and moves along a spatially explicit domain. The configuration of the environmental factors such as temperature, space and humidity is also flexible for each simulation. Thus, the tool provides the means to investigate ecological, evolutionary and population dynamics. The model is based on theoretical and empirical data collected from the literature and from both external and in-house experiments. Written in C++, Weaver is meant to allow efficient simulations with a high number of individuals. Although the first version of Weaver had already delivered some interesting insights, this paper here focuses on explaining the latest improvements that have been added to the new version of Weaver.

Keywords: agent based models, eco-evolutionary dynamics, high performance computing, genetics, animal traits, food webs, abiotic factors

1 Introduction

First we introduce here a general overview of the past development of Weaver. At the very beginning, the simulation framework was purely coded in the R programming language and it only contained a smaller set of functionalities. More details of the former R version, called "miniakira", can be found in its paper [1].

Even though "miniakira" contained only a small set of functionalities, the version was not able to run simulations with a high number of individuals. The main reason for this low performance was that "miniakira" was written in R, and this programming language is not meant for high performance computing. Although R excels when it is used for statistics and it is easy to use, it is an interpreted language and thus it yields low performance in general, if we compare it with any compiled code. At that time, we decided to port the "miniakira" version to C++ and also add several functionalities that were missing. This was a big improvement and the results can be read in the paper published in *Advances on Ecological Research* [3].

Based on the latter paper, key referents on the field of agent based models compared the capabilities of Weaver with those of several other frameworks in

terms of performance, functionality and usability, ranking our platform very high among them[7].

Later on, Weaver was presented in the Highlights of Practical Applications of Agents, Multi-Agent Systems, and Sustainability (PAAMS) in Salamanca [6].

Here we present a general overview of the improvements and features that have been designed, developed and added to the tool since then.

2 Last year progress

The subsequent sections describe the improvements and addition of new functionalities that have been included in Weaver since the last year. First, we discuss the addition of omnivore species. Then we describe the modelling and implementation of new, fine tuned growth curves for each animal species. Later, we present the addition of the reproductive mode and the energy balance during reproduction events, along with the new mating algorithm for adult individuals. We focus then on the implementation of the genetic dominance. We continue with the improvements of the animal interactions with the space. Then we mention the inclusion of a predate-prey ratio and different sizes for the starting generation on each simulation. Finally, we present the new "switching" algorithm that contributes to the stability of the trophic levels.

2.1 Omnivore species

In previous versions of Weaver, it was only possible to parametrize and include either predator or prey animal species. However, omnivorous species are widely present in the soil fauna, so this feature is a must if we want to be able to design realistic food webs.

For several reasons, the inclusion of omnivorous species was not a straightforward task. One of those reasons is that digestion times and metabolic processes are quite different in predator and prey species. Adding omnivorous species means we had to balance out the metabolic models for predator and preys. To put it down in detail: we had to re-model the metabolic gains and loses in a way that the functionality is now standardized for every animal. The former entities (animal and fungus) now share a common property that makes them both "edible". The energy gain from consuming any of the different edibles is now coded directly on the target edible, rather than on the consumer.

2.2 Improved growth curves

Growth curves in the previous versions of Weaver only incurred in the growth of the animal once the individual reached the next instar. Thus, the energy cycle was rather discontinuous. In this version of Weaver, growth curves have been readjusted following a new algorithm. Each individual now experiments a more parsimonious and continuous growth, where the amount of energy needed each time step is computed from the predicted metabolic loss and the distance to the

High Performance Computing applied to model eco-evolutionary agent based systems

expected growth curve. This expected growth curve is defined by the genetics of each individual. An example of an animal's growth curve is depicted and explained in Figure 1 .

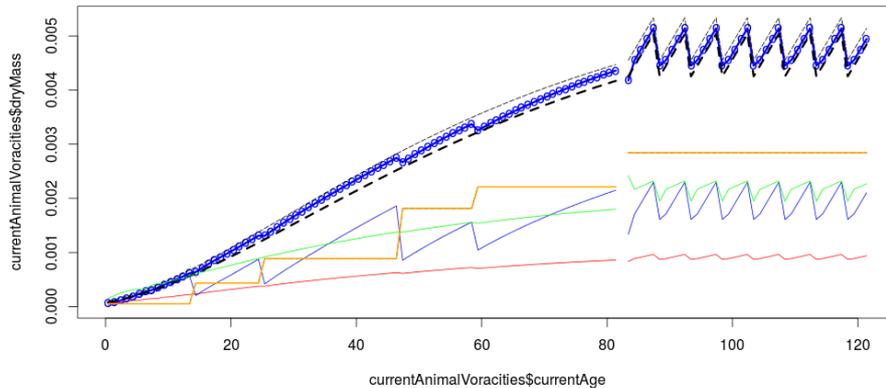


Fig. 1. Example of growth curve in one of the simulations. The animal growth (blue dots) now follows the expected growth curve defined by genetics (discontinuous black line). The orange line represents the size of the animal. The thin blue line represents the current energy.

2.3 Reproductive mode and mating selection

The animals in this version now grow until they achieve the mature status (last instar in Figure 1). They then change their mode to reproductive: the males start searching for a partner to mate with and the females begin consuming more resources in order to gather the needed energy to lay eggs. When they lay eggs, they spend all the energy and the cycle begins again.

The mating algorithm has been improved as well. Males now look for uncoupled females as priority, since they are more likely to mate with them. Females now can gather and store sperm from several males before she lays the eggs. This allows for better selection, since the female can now select only the best sperm in her bank.

2.4 Temporal series

All of the processes mentioned above are highly dependant on the environmental factors, especially on temperature. Temporal series for temperature and humidity

have been added in this version, so both abiotic factors can now follow user-defined series that are repeated again and again. This allows for more diverse interactions inside of the same simulation.

2.5 Revised genetics and dominance

The reproductive events present in the simulations follow the general type of cell division present in nature. Each new individual is created by performing the meiosis process on the parents' genome. In this version, the code for this algorithm has been revised and the offspring individuals now segregate better, resembling even more the populations that can be found in nature.

The Figure 2 represents how the genetics are modelled in Weaver. Each individual has a certain amount of chromosomes (defined by user), and the traits of the individual are derived from the sum of alleles, by using linear interpolation.

The novelty feature that has been added in this version is the genetic dominance. Each allele now has a defined dominance value. The sum of the alleles is now performed by only taking into account the higher dominant allele for each locus of the pair of chromosomes.

2.6 Improved animal interactions with space

Individuals now interact with the space differently. They follow better their preferences for prey (detailed in next subsection) and they also avoid obstacles better. The animals are now able to follow walls and find gaps in those walls in order to capture the prey (or find the fungus, in the case of preys).

2.7 Switching algorithm

One of the key features that has been added recently is the ability of an animal to change its main source of nourishment. The "switching" algorithm, as it is called, now takes into consideration all the species the animal can eat, and applies a preference defined by user.

Each individual also stores now a vector of abundance, based on how many resources of each species it has found throughout its life. The abundance vector is modified every step, so the amount of time steps the animal can "remember" is finite.

These two attributes together provide the final decision of the animal to chase and prey on different resources (be it preys or fungus). This behaviour promotes stability in the system, since the species with the lowest abundance will rarely be predated upon.

2.8 Predator-prey abundance and starting generation

The species ratio now follows the predator-prey power law shown in this paper[8] and that generally appears through the literature. This contributes to the persistence of the simulations, so the system will not crash due to lack of resources in any trophic level.

High Performance Computing applied to model eco-evolutionary agent based systems

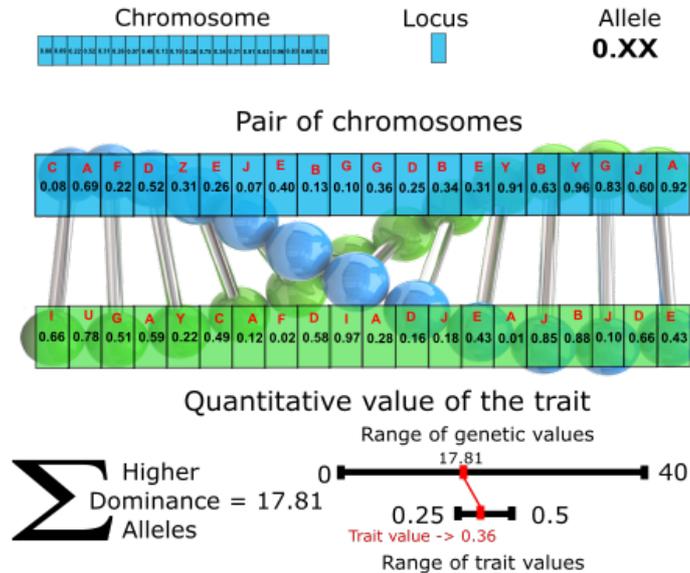


Fig. 2. Simulated genetics and quantitative traits obtained by linear interpolation. Each of the alleles shown in the image now have a genetic dominance value attached. The resulting sum is now computed only taking into account higher dominance alleles for each loci.

3 Conclusions

Weaver is a tool that, as we have seen in this paper, can be improved to almost infinite extent. This platform is the basis in which we can include either the most descriptive models for the processes in ecology, or the most simplified functions for the features that are less trivially unravelled.

To this date, the framework can simulate the most complex ecosystems, including a variety of processes that resemble those found in nature. However, the addition of such features is slowly incurring in the performance of the simulations. After finishing the implementation of the current planned features, the next focus will be on optimizing the framework in terms of computational performance.

High Performance Computing applied to model eco-evolutionary agent based systems

Although the input module in Weaver is already very flexible, the integration with R as input interface could greatly improve the user experience while also providing the tools to directly analyse the output data.

References

1. Moya-Laraño, J., Verdeny-Vilalta, O., Rowntree, J., Melguizo-Ruiz N., Montserrat, M., Laiolo, P.: Climate Change and Eco-Evolutionary Dynamics in Food Webs. *Advances in Ecological Research* 47, 1 (2012)
2. Verdeny-Vilalta, O., Moya-Laraño, J.: Seeking water while avoiding predators: moisture gradients can affect predator-prey interactions. *Animal Behaviour* 90, 101–108 (2014)
3. Moya-Laraño, J., Bilbao-Castro, J.R., Barrionuevo, G., Ruiz-Lupi3n, D., Casado, L.G., Montserrat, M., Meli3n, C.J., Magalh3es, S.: Eco-evolutionary spatial dynamics. Rapid evolution and isolation explain food web persistence. *Advances in Ecological Research* 50, 75–143. (2014)
4. Gandrud, C.: *Reproducible Research with R and RStudio*. Chapman & Hall/CRC Press. (2013)
5. De Angelis, D.L., Grimm, V.: *Individual-based models in ecology after four decades*. F1000Prime Reports (2014)
6. Bilbao-Castro, J.R., Barrionuevo, G., Ruiz-Lupi3n, D., Casado, L.G., Moya-Laraño, J.: *Weaver: A Multiagent, Spatial-Explicit and High-Performance Framework to Study Complex Ecological Networks*. Highlights of Practical Applications of Agents, Multi-Agent Systems, and Sustainability. (2015)
7. Grimm, V., Ayll3n, D. and Railsback, S.F.: Next-Generation Individual-Based Models Integrate Biodiversity and Ecosystems: Yes We Can, and Yes We Must *Ecosystems* 20, 229–236 (2017)
8. Hatton, Ian A., McCann, Kevin S., Fryxell, John M., Davies, T. Jonathan, Smerlak, Matteo, Sinclair, Anthony R. E. and Loreau, Michel: The predator-prey power law: Biomass scaling across terrestrial and aquatic biomes *American Association for the Advancement of Science* 349, 6252 (2015)

Modelado y Simulación de un Nuevo Concepto de Vehículo Urbano Eléctrico Ligero Basado en la Optimización del Uso de Energías Renovables y la Reducción de CO₂

Francisco José Gómez Navarro

Universidad de Almería
Carretera Sacramento s/n
04120 La Cañada de San Urbano
Almería - España
<https://www.ual.es/>

Resumen El presente informe tiene por objeto dar a conocer el proyecto de investigación sobre un nuevo concepto de vehículo muy ligero que aprovecha las diversas fuentes de energía renovable disponibles, optimizando su uso al objeto de maximizar la reducción en la emisión de gases de efecto invernadero ligados al transporte urbano de personas y bienes. En la introducción se presentarán los datos más significativos relativos al transporte urbano de personas y bienes, presente y futuro, que desembocan en la motivación e hipótesis para el presente trabajo. En el apartado de avances se describirá el trabajo realizado, los medios utilizados y el estado de desarrollo de los modelos que serán utilizados para la continuación del proyecto.

Keywords: Vehículo Eléctrico, Movilidad Urbana, Energías Renovables, Reducción CO₂, Modelado Orientado a Objetos, Modelica

1. Introducción

La sociedad moderna ha basado su desarrollo en gran medida en la posibilidad de desplazar cantidades suficientes de bienes y personas entre distintas localizaciones de forma eficaz. El transporte consume el 19 % de la energía a nivel mundial y emite el 23 % del dióxido de carbono (CO₂) debido al consumo energético [1]. Con la tendencia actual, el uso de energía para el transporte aumentará un 50 % para 2030 y más de un 80 % para 2050.

El Intergovernmental Panel on Climate Change (IPCC) advierte que para evitar las desastrosas consecuencias del cambio climático, las emisiones globales de CO₂ deben disminuir, al menos, un 50 % de aquí al año 2050 [2]. El transporte juega un papel decisivo en la consecución de este objetivo, resulta indispensable por tanto incidir en la necesaria adaptación de los modos actuales de desplazamiento. En Europa, se han establecido las bases para una política de transportes competitiva y sostenible[3], entre las que cabe destacar:

Modelado y Simulación de un Nuevo Concepto de Vehículo Urbano Eléctrico Ligeró...

- La eliminación progresiva de los vehículos de «propulsión convencional» en el entorno urbano es una contribución fundamental a una reducción significativa de la dependencia del petróleo, las emisiones de gases de efecto invernadero, la contaminación atmosférica local y la contaminación acústica.
- Debe fomentarse el uso de vehículos de pasajeros más pequeños, más ligeros y más especializados en el transporte por carretera.

El transporte ligero, fundamentalmente orientado a personas, consumió en 2006 el 47 % de la energía dedicada al transporte [1]. El parque mundial de vehículos ligeros (Ligh Duty Vehicle - LDV) es previsible que se triplique para el año 2050, principalmente debido al incremento en países en vías de desarrollo[4]. De lo expuesto anteriormente, se puede concluir que uno de los focos de actuación preferentes para la consecución del objetivo de reducción de la emisión de gases de efecto invernadero está en el **transporte urbano ligero** y en el uso de **fuentes de energía alternativas a los combustibles fósiles**. Estas fuentes deben:

- Ser renovables para evitar su agotamiento.
- Estar disponibles en la zona para evitar la dependencia energética de terceros.
- Ser acumulables para poder disponer de las reservas adecuadas que equilibren la capacidad de producción y la demanda.

1.1. Motivación

El trabajo de investigación propuesto viene justificado por dos motivaciones diferentes y complementarias:

- Necesidad de disponer de modelos adecuados que permitan analizar y anticipar las prestaciones y comportamientos de las distintas alternativas tecnológicas en estudio para el caso de los sistemas de transporte urbano del futuro. La necesidad de disponer de sistemas de propulsión y fuentes energéticas alternativas obliga a disponer de modelos modulares, con capacidad para integrar y simular el comportamiento dinámico del sistema completo, integrando las distintas tecnologías y desde distintos niveles de abstracción. Si bien hay un elevado número de investigadores que han desarrollado e investigado sobre modelos que analizan las posibilidades de estas nuevas tecnologías de alimentación y propulsión, en la mayoría de los casos se basan en la simple adaptación de la concepción tradicional del vehículo a esta nueva realidad.
- Necesidad de avanzar en nuevas propuestas de movilidad urbana, basadas en vehículos más ligeros, mas adaptados al uso específico para transporte ligero de corta distancia, con máximo aprovechamiento de las fuentes de energía renovable disponibles y con esquemas de uso que optimicen su capacidad.

1.2. Hipótesis

La hipótesis principal que se pretende demostrar, es la viabilidad de un nuevo concepto de movilidad urbana basado en el uso intensivo de fuentes de energía renovable, al que podremos llamar Very Light Urban Vehicle (VLUV). Como hipótesis secundaria, se pretende demostrar cómo este concepto consigue una reducción importante en la emisión de CO₂ a la atmósfera y disminuye el consumo de combustibles fósiles y otras fuentes de energía no renovable [5].

2. Avances

El primer paso del trabajo de investigación, tras la revisión bibliográfica, ha consistido en la determinación de un concepto genérico de vehículo ligero para desplazamiento urbano que permita la investigación y desarrollo de los modelos correspondientes a cada uno de los componentes, sus interfaces y la simulación del conjunto completo bajo distintas configuraciones y condiciones de contorno. Para el desarrollo de los distintos modelos se ha optado por el Modelado Orientado a Objetos con ayuda del lenguaje de modelado Modelica. Este estándar abierto tiene las siguientes ventajas desde el punto de vista del modelado y la simulación de sistemas multi-físicos [6][7]:

- Encapsulamiento del conocimiento.
- Capacidad de interconexión topológica.
- Modelado jerárquico.
- Instanciación de objetos.
- Herencia de clases.
- Capacidad de interconexión generalizada.

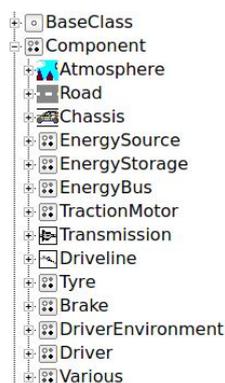


Figura 1. Estructura básica de la librería de componentes

Se utiliza el entorno de desarrollo Dymola© [8], basado en el lenguaje de programación de código abierto Modelica para la programación, simulación y

Modelado y Simulación de un Nuevo Concepto de Vehículo Urbano Eléctrico Liger...

optimización de los modelos. Para la definición de las clases base, los interfaces de conexionado y la estructuración general del modelo se ha tomado como referencia la librería `VehicleInterfaces` de Modelica [9]. Esta librería proporciona una serie de definiciones de interface normalizadas para uso en subsistemas de automoción y modelos de vehículos. Su objetivo es el de promover la compatibilidad entre las distintas librerías de componentes de automoción y proporcionar una estructura flexible y potente para el modelado de vehículos.

La estructura final de los modelos componentes del vehículo se pueden apreciar en la figura 1. Con los diferentes modelos desarrollados de cada uno de los distintos componentes, se pueden conseguir las distintas configuraciones deseadas de vehículos que nos ayudarán en la simulación dinámica, análisis y evaluación de resultados.

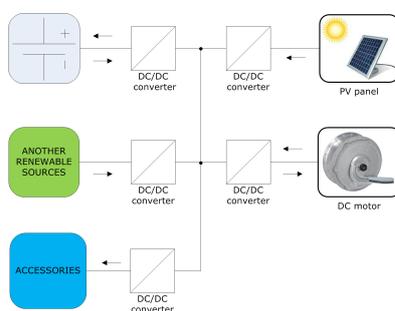


Figura 2. Esquema general del clúster inteligente

2.1. Modelado del vehículo

Al objeto de la simulación dinámica del vehículo completo, de su control y del análisis del balance energético de su actuación, se ha desarrollado un modelo de clúster de energía inteligente, que agrupa los diferentes elementos componentes del vehículo que aportan o consumen energía. Tal como se aprecia en la figura 2, está compuesto principalmente por el panel fotovoltaico (PV), la batería, los convertidores de tensión DC/DC, el motor de accionamiento del vehículo (que puede funcionar como freno regenerativo), los accesorios del vehículo y otras posibles fuentes de energía renovable en estudio. También se aprecia en la figura los sentidos del flujo de energía posibles en cada uno de los componentes. Los componentes que aportan energía son: la batería (en modo descarga), el panel PV, el motor (en modo freno regenerativo) y las otras fuentes de energía renovable instaladas. Los componentes que consumen energía son: la batería (en modo carga), el motor de accionamiento del vehículo y los accesorios. Todos los componentes se interconectan mediante un bus DC a través de los convertidores DC/DC funcionando en modo uni ó bidireccional.

En la figura 3 se muestra el modelo completo del vehículo en Modelica.

Modelado y Simulación de un Nuevo Concepto de Vehículo Urbano Eléctrico...

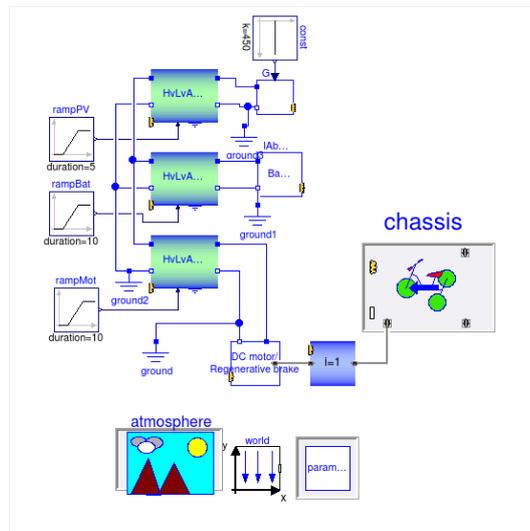


Figura 3. Modelo completo del vehículo en Modelica

3. Publicaciones

La siguiente contribución ha sido aceptada:

”Modelling a Smart-Grid for a Solar Powered Electric Vehicle”

Francisco J. Gomez, Luis J. Yebra, Antonio Gimenez

Submission number: 113

Type of submission: Discussion Contribution

9th International Conference on Mathematical Modelling

Vienna, Feb 21-23, 2018

4. Conclusiones

Se ha realizado la introducción y descripción del proyecto de investigación en curso. Se han revisado las motivaciones y justificaciones del mismo, así como del estado de avance. Los trabajos en curso, una vez definidas las clases base e interfaces a utilizar, están orientados a conseguir el modelo dinámico completo de un clúster de energía inteligente para ser usado en un vehículo animado por energía solar y otras fuentes de energía renovable, con el propósito del diseño y optimización del control del mismo. El modelo será utilizado para realizar las simulaciones en tiempo real que permitan la optimización de los algoritmos de control de los distintos componentes. Se han presentado los principales componentes, su esquema de interconexión, así como los flujos de energía. Se ha presentado un ejemplo del modelo completo del vehículo en Modelica.

Referencias

1. International Energy Agency, “IEA Response System for Oil Supply Emergencies 2012,” International Energy Agency, Paris, Tech. Rep., 2012. [Online]. Available: http://www.iea.org/publications/freepublications/publication/EPPD{_}Brochure{_}English{_}2012{_}02.pdf
2. “Sustainable Transportation Energy Pathways: A Research Summary for Decision Makers,” Institute of Transportation Studies, University of California, Davis, Tech. Rep., 2011. [Online]. Available: http://www.its.ucdavis.edu/?page{_}id=10063{\&}pub{_}id=1499
3. European Commission, *A sustainable future for transport - Towards an integrated, technology-led and user-friendly system*. Luxembourg: Publications Office of the European Union, 2009.
4. United Nations Environment Programme, Nairobi, Tech. Rep.
5. G. Pasaoglu, M. Honselaar, and C. Thiel, “Potential vehicle fleet co2 reductions and cost implications for various vehicle technology deployment scenarios in europe,” *Energy Policy*, vol. 40, no. C, pp. 404–421, 2012. [Online]. Available: <https://EconPapers.repec.org/RePEc:eee:enepol:v:40:y:2012:i:c:p:404-421>
6. Fritzson P., *Principles of Object-Oriented Modeling and Simulation with Modelica 2.1*. Wiley-IEEE Press, 2004. [Online]. Available: <http://www.wileyurope.com/WileyCDA/WileyTitle/productCd-0471471631.html>
7. M. Association, “Modelica ® - A Unified Object-Oriented Language for Physical Systems Modeling Language Specification,” *Interface*, vol. 5, no. 6, p. 250, 2010. [Online]. Available: www.modelica.org
8. Dassault Systèmes AB, Ed., *Dymola - Dynamic Modeling Laboratory - User Manual*. Lund - Sweden: Dassault Systèmes AB, 2012, vol. 1, no. May. [Online]. Available: <http://www.dymola.com>
9. “VehicleInterfaces Library (Version 1.2). Tutorial and Reference,” p. 157, 2011. [Online]. Available: <https://build.openmodelica.org/Documentation/VehicleInterfaces1.2.2/Resources/Documentation/VehicleInterfaces.pdf>

El acceso seguro del ciudadano a su Historia Digital permite avanzar en la gestión de Procesos Asistenciales

González Revuelta, María Esther

Tecnologías de la Información y Comunicación. C.H.Torrecárdenas.Almería

Abstract. El uso de la Historia Clínica Electrónica abre grandes oportunidades de integrar, compartir y disponer de información de forma rápida, mejorando la calidad de la atención recibida y percibida. Es el propio usuario el que cada vez más solicita y demanda acceder a sus datos de salud. Impulsar y dotar al entorno sanitario de herramientas que faciliten el acceso y disponibilidad de la información asistencial nos lleva a mejorar la calidad del servicio al ciudadano, usuario del sistema.

Analizamos la opinión de los usuarios mediante encuestas en cuanto a cuándo y cómo preferiría tener acceso a su estado de salud. Demandan disponer de información acerca del estado de su proceso asistencial, estar permanentemente conectados mediante un sistema de trazabilidad de puntos de control, además de una interacción digital médico-enfermera-paciente a través de las nuevas tecnologías, mejorando la accesibilidad, minimizando tiempos de espera innecesarios y potenciando el papel del usuario mediante el uso de una vía de comunicación más directa con sus resultados y con el estado de su proceso. Avanzamos en esta propuesta aplicada a unos procesos asistenciales integrados (PAIs) concretos, velando en todo momento la protección, seguridad, confidencialidad e integridad de la información tratada..

Keywords: Historia Clínica Electrónica (HCE), acceso HCE, portal pacientes

1 Introducción

El desarrollo de la Historia Clínica Electrónica (EHR) en el ámbito sanitario es un objetivo sobre el que se viene trabajando y avanzando de forma más significativa en los últimos años. Las tecnologías de la información han transformado la forma de atención, la forma de comunicarse e interrelacionarse el profesional sanitario con el paciente/usuario/ciudadano y la forma de gestionar en este ámbito [1]

El acceso seguro del ciudadano a su Historia Digital permite avanzar en la gestión de Procesos Asistenciales

Avanzar y evolucionar en esta línea aporta una serie de beneficios evidentes tanto para el profesional sanitario como para el paciente y abre nuevas líneas de trabajo a desarrollar e investigar[2].

La realidad es que en el momento actual los Sistemas de Salud, con la Historia Clínica Digital, son muy heterogéneos y se encuentran en diferentes fases de desarrollo [3]. Probablemente, en general no se ha avanzado al ritmo que se esperaba, por la propia complejidad técnica que tienen los diferentes sistemas y elementos que hay que integrar e incluso por las propias reticencias de los profesionales a los que les supone, en algunos casos, un cambio drástico en su forma de trabajar.

Nos encontramos en un momento donde el avance tecnológico obliga a impulsar y dotar al entorno sanitario de herramientas que faciliten el acceso y disponibilidad de la información asistencial, encaminadas a mejorar la calidad del servicio al ciudadano y usuario del sistema [4].

Desde el punto de vista del clínico y en definitiva del profesional que trabaja en el entorno sanitario, el poder trabajar con una Historia Clínica Electrónica integral aporta en gran medida, unidad, disponibilidad de la información, facilidad de acceso, posibilidad de intercambio de información entre diferentes sistemas, seguridad del paciente, ahorro de costes, eficacia y eficiencia en el manejo de la información y otras muchas ventajas.

Por otro lado, para el usuario del Sistema de Salud se abren unas posibilidades de información, acceso, comunicación e interrelación que supondrán un cambio muy significativo en la relación médico-paciente y en general del papel más activo del paciente en el desarrollo de su historia de salud [5].

Se promueve la interacción digital médico-paciente a través de las nuevas tecnologías, mejorando la accesibilidad y potenciando el papel del usuario mediante el uso de una vía de comunicación más directa con sus resultados.

Nos planteamos en este punto, de qué forma debería ser el acceso a la información para garantizar en todo momento la seguridad y confidencialidad de los datos que componen la Historia Clínica (HHCC)[6]. Estos datos requieren el más alto nivel de protección que requiere la Ley Orgánica de Protección de datos y el reglamento que la desarrolla [7].

Es obvio, que en la medida que se facilita la disponibilidad de la información, se incrementa el riesgo de mantener la confidencialidad, seguridad e integridad de los datos y por tanto los mecanismos de control de acceso deben ser más rígidos y controlados.

No es una tarea fácil de implementar, ya que se requieren diferentes normas y entre ellas las de seguridad, implicando la necesidad de indicadores normalizados para todos los agentes que intervienen en el acceso al EHR [8].

Muchos de los países dotan a los gestores de atención médica mantener la protección de la información del paciente a través de la ética y prácticas profesionales, así como a través del uso de las tecnologías y el cumplimiento de la normativa legal para resolver el paradigma planteado [9].

La mayoría de las políticas de privacidad de los sistemas de EHR no proporcionan una descripción en profundidad de las medidas de seguridad que utilizan. Por otra parte, el cumplimiento de las normas y reglamentos en los sistemas de registros de

El acceso seguro del ciudadano a su Historia Digital permite avanzar en la gestión de Procesos Asistenciales

historia digital de salud es todavía baja [10]. La seguridad de EHR ve amenazada la privacidad del paciente cuando dicha información médica personal y confidencial se pone a disposición de terceros sin consentimiento o sin autorización.

Hemos analizado en este proyecto un sistema automatizado de protección mixto que garantice en tiempo real tanto los derechos del paciente, principal protagonista en cualquier proceso médico, como las necesidades del facultativo para poder realizar su trabajo. Además de dar soporte a situaciones donde se necesite el acceso a la información desde un centro diferente desde el que se generó el/los episodios/s. Todos los procesos se producen de forma transparente tanto para el paciente como para el facultativo y en tiempo real.

El Ministerio de Sanidad y Consumo junto a las Comunidades, están trabajando en un sistema de compatibilidad orientado a garantizar la disponibilidad y el acceso a la historia del paciente, independientemente del lugar donde éste haya sido atendido [11].

Las autoridades sanitarias a nivel nacional están trabajando desde 2006 en un proyecto HCDSNS (Historia Clínica Digital del Sistema Nacional de Salud), que pretende poner a disposición de profesionales y ciudadanos el acceso a los datos relevantes que van generándose en formato digital relativos a la asistencia sanitaria de un paciente, independientemente del lugar donde sea asistido y haya generado su episodio, siempre cumpliendo todos los requisitos de seguridad, confidencialidad, agilidad y sencillez [12].

El desarrollo de los sistemas actuales de acceso a la Historia de Salud Digital por parte del usuario está en pleno proceso de desarrollo y en este marco es en el que nos vamos a centrar.

Han sido los propios usuarios los que, en el seno de unas comisiones de trabajo realizadas en un centro hospitalario, en el que participaron conjuntamente ciudadanos, usuarios del sistema sanitario público y profesionales del mismo, han detectado una falta de información en determinados momentos de su proceso asistencial y nos han trasladado su interés por resolver esta brecha. Aprovechando el desarrollo de la Historia Clínica Electrónica, analizaremos la situación y desarrollaremos un sistema alternativo que ofrezca al usuario la posibilidad de acceder a datos referentes a su proceso, mediante el acceso a la información de su Historia Clínica Electrónica de forma segura, mejorando tiempos de respuesta en general y mejorando en los procesos de atención en general.

Para concretar un área de conocimiento específico en el ámbito sanitario, se pretende pilotar el estudio en dos Procesos Asistenciales Integrados (PAIs) de gran relevancia social: PAI Cáncer de Cérvix y PAI Embarazo [13].

El paciente podrá disponer de información acerca del estado de su proceso mediante los puntos de control de interés definidos. El usuario podrá participar en cada punto de control mediante una valoración personal y aportaciones de sugerencias que se visualizarán en un entorno compartido. Además, estará permanentemente conectado con su proceso mediante un sistema de mensajería bidireccional que le permitirá acortar tiempos de espera y desplazamientos innecesarios, disminuyendo su ansiedad, a la par que se aligera la demanda asistencial y se optimizan circuitos en los PAIs objeto del estudio [14].

El acceso seguro del ciudadano a su Historia Digital permite avanzar en la gestión de Procesos Asistenciales

Finalmente, utilizando reglas de automatización con mensajes predefinidos en base a resultados propios de pruebas diagnósticas, se emitirá información específica adaptada al estado de salud de cada usuario mediante recomendaciones sobre hábitos dietéticos, higiénicos y en general hábitos de vida saludable que obviamente ayudarán a la mejora de su estado de salud. El modelo sería fácilmente extensible a otros procesos asistenciales, así como exportable a otros sistemas sanitarios tanto públicos como privados.

2 Avances

El desarrollo de este proyecto ha avanzado en varias líneas que finalmente confluyen en un mismo punto: acceso seguro del ciudadano a su historia digital de salud.

- **Hemos analizado y propuesto en este proyecto un sistema automatizado de protección mixto** que garantice en tiempo real tanto los derechos del paciente, principal protagonista en cualquier proceso médico, como las necesidades del facultativo para poder realizar su trabajo. Además de dar soporte a situaciones donde se necesite el acceso a la información desde un centro diferente desde el que se generó el/los episodios/s. Todos los procesos se producen de forma transparente tanto para el paciente como para el facultativo y en tiempo real.

Para garantizar la privacidad de la información usaremos algoritmos simétricos. La clave secreta necesaria para cifrar la información y descifrarla es exactamente la misma. Para garantizar la seguridad de la clave secreta K , ésta será cifrada usando algoritmos de clave pública. En este tipo de algoritmos, la clave usada para cifrar es distinta a la usada para descifrar. Una de ellas será una clave pública, es decir, conocida por todos, mientras que la otra clave será privada y sólo conocida por el propietario de la misma. Si usamos la clave pública para cifrar, estaremos buscando la privacidad de la información custodiada. Sin embargo, si ciframos con la clave privada, estaremos buscando la autenticación de la información en cuestión. [15,16]

- El desarrollo de los sistemas actuales de acceso a la Historia de Salud Digital por parte del usuario está en pleno proceso de desarrollo y en este marco es en el que nos movemos. Realizamos un **análisis de situación en cuanto a los modelos actuales implementados que permiten al ciudadano acceder a su información de salud** a través de plataformas digitales. El análisis y revisión bibliográfica se ha realizado de las diferentes comunidades autónomas de España y en un ámbito internacional, países de la Unión Europea y otros como Canadá y Estados Unidos.

La situación y estrategia de eHealth en un escenario nacional internacional, nos refleja como constantes comunes la existencia de una estrategia nacional que plantea objetivos y prioridades a medio y largo plazo y la existencia de un órgano encargado de coordinar la estrategia de eSalud. No obstante el avance en el desa-

El acceso seguro del ciudadano a su Historia Digital permite avanzar en la gestión de Procesos Asistenciales

rrollo es bastante más lento de lo esperado por la complejidad que supone desarrollar mecanismos de interoperabilidad de los sistemas de información, lo cual está suponiendo un retraso en la conclusión de las actuaciones frente a los tiempos marcados originariamente y otro aspecto común muy relacionado con el objeto de nuestro análisis es la preocupación general de los ciudadanos y profesionales sobre las cuestiones de seguridad y las garantías de confidencialidad de los datos personales, aspectos que coinciden en el ámbito Nacional. [17,18]

- **Encuestas a usuarios del Sistema Sanitario implicados en los procesos objeto de estudio.**

A través de las diferentes comisiones de trabajo que se realizan en los centros sanitarios con los ciudadanos, éstos trasladan su opinión en cuanto a la falta de información relativa a su proceso asistencial durante el desarrollo de las diferentes actuaciones, así como la indisponibilidad de acceso a cualquier información clínica que de él se genere.

En muchas ocasiones estas esperas generan ansiedad, incertidumbre y en general un malestar innecesario que podría evitarse permitiendo que en determinados casos el usuario pueda tener información de su proceso antes de que tenga que acudir de nuevo a su consulta de Hospital o de Primaria. Siempre, sería consensuada esta información cómo y cuándo con los profesionales sanitarios.

Nos planteamos profundizar más en detalle en el grado de interés que tiene el usuario en conocer determinada información de su proceso asistencial a través de las TIC (tecnologías de la información y comunicación). Queremos valorar también qué medios tecnológicos consideran más afines para recibir o acceder a este tipo de información.

Para llevar a cabo este estudio observaciones se ha pedido autorización al Comité Ético de Investigación Clínica (CEIC) del Complejo Hospitalario Torrecárdenas, que ha validado y aprobado el proceso.

Se presentó una serie de documentación requerida que se ha utilizado para el desarrollo de las encuestas: Hoja de información al paciente, consentimiento informado, hoja de encuesta tipo likert, análisis del entorno, viabilidad,...

La población objeto de estudio son los usuarios del Sistema Sanitario Público de Andalucía que se realizan la prueba del Screening de Cáncer de Cérnix en la consulta de Ginecología de Atención Especializada y mujeres embarazadas que se realizan una extracción de seguimiento.

Para la selección de la muestra se utiliza una metodología de muestreo consecutivo, por tanto no probabilístico [19]. Abundando en detalles y, teniendo en cuenta

El acceso seguro del ciudadano a su Historia Digital permite avanzar en la gestión de Procesos Asistenciales

un tiempo de realización medio de encuesta de 10 minutos, el criterio utilizado ha sido el de accesibilidad del paciente y aceptabilidad de inclusión por parte del mismo.

La mecánica ha consistido en el abordaje a la salida de consulta o laboratorio de los usuarios, acceso a los mismos, información sobre el objeto y finalidad de la encuesta y, caso de aceptar la realización de la misma, previo consentimiento informado, que siempre ha sido cumplimentado. Una vez finalizada dicha encuesta, el profesional se dirige al siguiente usuario disponible que exista, hasta cumplimentar el tamaño muestral requerido. Se realiza durante diferentes días y semanas.

Previamente a fin de realizar una reducción en la variabilidad de la información obtenida por los encuestadores se formó en una sesión de 45 minutos a cada uno de ellos sobre los siguientes puntos:

- Debe realizar preguntas claras y con un significado que sea el mismo para el entrevistado que para el entrevistador.
- Debe interpretar adecuadamente la respuesta para la cumplimentación, solicitando las aclaraciones oportunas al entrevistado caso de duda. No debe emitir ninguna opinión mientras se conversa. No debe sugerir respuestas al entrevistado.

Actualmente estamos en proceso de análisis estadístico de los datos recogidos.

- **Piloto del proyecto en dos Procesos Asistenciales Integrados (PAIs) de gran relevancia social: PAI Cáncer de Cérvix y PAI Embarazo, parto y puerperio.**

Un PAI es en sí mismo una herramienta para la mejora continua de la asistencia sanitaria [20]. Detalla el itinerario de los pacientes y el conjunto de actuaciones, decisiones, actividades y tareas que se encadenan de forma secuencial para atender un problema asistencial específico con la máxima calidad. Se trata pues de proporcionar a cada paciente el ciclo de atención integral que requiere para la atención de su problema de salud en las mejores condiciones posible.

Es en estos procesos por tanto, donde entendemos totalmente necesario el desarrollo de herramientas informáticas que apoyen el proceso y se promueva la interacción digital médico-paciente a través de las nuevas tecnologías, mejorando la accesibilidad y potenciando el papel del usuario mediante el uso de una vía de comunicación más directa con sus resultados, simplifica el proceso, acorta tiempos de espera y mejora la presión asistencial optimizando los circuitos.

El acceso seguro del ciudadano a su Historia Digital permite avanzar en la gestión de Procesos Asistenciales

El paciente podrá disponer de información acerca del estado de su proceso mediante los puntos de control de interés definidos.

- Extracción: Si ésta ha sido correcta.
- Recepción: Teniendo en cuenta que la extracción se puede realizar en distintos puntos físicos, como centros de salud, hospital, centros de especialidades,... confirmar que la muestra se ha recibido correctamente en el laboratorio Core.
- Análisis: La muestra está en proceso dentro de los analizadores.
- Resultados: La muestra tiene resultados y/o parámetros solicitados.
- Validación: La muestra está siendo validada por los facultativos especialistas de laboratorio. Identificación de parámetros validados.
- Finalización: La analítica ha finalizado todo el procedimiento de laboratorio y los resultados están disponibles en la web.

Además, podrá estar permanentemente conectado con su proceso mediante el acceso a un sistema de monitorización y envío de alertas que le permitirá acortar tiempos de espera y desplazamientos innecesarios, disminuyendo su ansiedad, a la par que se aligera la demanda asistencial y se optimizan circuitos en los PAIs objeto del estudio.

Actualmente estamos en proceso de análisis y definición de reglas de automatización con mensajes predefinidos en base a resultados propios de pruebas diagnósticas. Se podrá tener información en relación con la trazabilidad de la muestra y a los resultados. Y en base a estas reglas se emitirá información específica adaptada al estado de salud de cada usuario mediante recomendaciones sobre hábitos y recomendaciones de vida saludable.

Al consensuar que para determinados diagnósticos y bajo determinadas condiciones se podrían incorporar cambios en el circuito habitual del proceso a fin de mejorar tiempos de respuesta y avanzar en la eficacia de los circuitos, se mejora el propio proceso y redonda positivamente en el paciente.

PAI embarazo, parto y puerperio [21]: Las embarazadas deben realizarse estudios de screening para la detección precoz de diabetes gestacional, serologías infecciosas, detección de enfermedades congénitas,.. que se contemplan. Los resultados son alentadores, ya que una amplia mayoría de casos estos resultados son normales o negativos; sin embargo, la incertidumbre de las personas frente a estos resultados que pueden tener alto impacto en su salud o de los futuros descendientes, generando situaciones de ansiedad que podrían acortarse significativamente utilizando nuestra herramienta.

El acceso seguro del ciudadano a su Historia Digital permite avanzar en la gestión de Procesos Asistenciales

Por otro lado, en la consulta especializada del Cáncer de Cérvix del Hospital, se atienden a usuarias que previamente han sido derivadas desde Atención Primaria con un screening citológico con alguna alteración que debe confirmarse con pruebas moleculares. La espera aquí es igualmente angustiosa para la mujer. Con el acceso a la información clínica por parte de la paciente utilizando la tecnología expuesta tendría todo el proceso de la prueba controlado y el resultado lo tendría accesible en tiempo real, y sin demora alguna que fuese vinculada a una nueva cita con el ginecólogo. Un sistema de mensajería expertizada a partir de los resultados enviará la información interpretada de los casos que se interpretaran negativos y que no requirieran más que un control rutinario aplazado.

El modelo presentado es fácilmente exportable a otros procesos asistenciales.

3 Conclusiones

Cada vez más, la gente quiere estar permanentemente informada y actualizada en todos los aspectos de la vida en los que existan procesos de inicio-fin. La salud es una de las cuestiones más valoradas por la población y deben aplicarse los mismos estándares de comunicación que se emplean para otros aspectos. No obstante, tal vez por la criticidad de los datos que se manejan, no ha evolucionado al mismo ritmo que en otras áreas.

Son los propios usuarios del Sistema Sanitario Público los que demanda esta necesidad. Necesidad, que por otro lado, se ha constatado a través la elaboración de las encuestas a pacientes implicados en varios procesos asistenciales que han sido objeto de análisis y que hemos desarrollado y analizado.

La propuesta de permitir que en tiempo real pueda acceder a la información del estado de su proceso asistencial, así como recibir mensajes y alertas que le informen sobre posibles avances y cambios producidos, así como poder recibir información relativa a su estado de salud y recomendaciones de actuaciones y hábitos de vida saludable, permite avanzar en este ámbito mediante el desarrollo de herramientas que apoyen el proceso y promueva la interacción digital médico-paciente a través de las nuevas tecnologías, mejora la accesibilidad, posibilita acortar tiempos de espera y desplazamientos innecesarios, disminuyendo la ansiedad en la espera, a la par que se aligera la demanda asistencial, potencia el papel del usuario mediante el uso de una vía de comunicación más directa con sus resultados y se optimizan los circuitos en los Procesos Asistenciales Integrados (PAIs) objeto de estudio.

Como marco de evolución a este proyecto, no podemos perder de vista que cada vez más proliferan dispositivos wearables que recogen señales biológicas del paciente (señal de electrocardiograma (ECG), presión sanguínea, pulso, peso, temperatura, etc.)[22,23]. Estos dispositivos representan herramientas de control de Salud personal que permiten generar y enviar grandes cantidades de información de salud en remoto. Estos elementos podrían pasar a formar parte de la Historia de Salud digital del Paciente, aportando información muy valiosa para el profesional sanitario y a la cual podría acceder el paciente de forma integral. Lógicamente ésta es un área que aún tiene mucho campo de desarrollo, como es el uso de estándares que permitan la inter-

El acceso seguro del ciudadano a su Historia Digital permite avanzar en la gestión de Procesos Asistenciales

operabilidad de dispositivos, así como el uso de herramientas de Big Data que analicen, procesen y generen datos de utilidad que en general ofrezcan nuevas oportunidades de monitorización, evolucionando hacia una medicina personalizada, reduciendo factores de riesgo y en general mejorando la calidad de vida del ciudadano.

References

1. J. E. Cahill, M. R. Gilbert, T. S. Armstrong, "Personal health records as portal to the electronic medical record," *J Neurooncol*, vol. 117, no 1, pp. 1-6, 2014. DOI: 10.1007/s11060-013-1333-x.
2. X. Y. Zhang, P. Zhang, "Recent perspectives of electronic medical record systems," *Exp Ther Med*, vol. 11, no 6, pp. 2083-2085, 2016. DOI: 10.3892/etm.2016.3233.
3. Consejerías de Sanidad de todas las Comunidades Autónomas, INGESA (Ceuta y Melilla), Ministerio de Sanidad y Consumo de España. 1996. Las TIC en el Sistema Nacional de Salud *Red.es*. [Online]. Available: <http://www.salud-e.cl/wp-content/uploads/2013/08/Sanidad+en+Linea+-+Espa%C3%B1a.pdf>.
4. M. Cucciniello, I. Lapsley, G. Nasi, "Managing health care in the digital world: A comparative analysis," *Health Serv Manage Res*, vol. 29, no 4, pp. 132-142, 2016.
5. Jefatura del Estado, "Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica," *BOE num. 274 of November 15th*, 2002. Last update September 22, 2015.
6. Jefatura del Estado Español, "Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal," *BOE num. 298 of December 14th*, 1999. Current revision from March 06, 2011. Ministerio de Justicia Español, "Real Decreto 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de desarrollo de la Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal," *BOE num. 17 of January 19th*, 2008. Last update March 08, 2012.
7. The European Parliament and The Council of The European Union, "Regulation (UE) 2016/679 of the European parliament and of the council of 27 April on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)," May 4, «DOUE» num. 119, L119/1-88. [Online]. 2016. Available: <https://www.boe.es/doue/2016/119/L00001-00088.pdf>.
8. L. Harn, C. Hsu, M. Zhang, T. He, M. Zhang, "Realizing secret sharing with general access structure," *Inf Sci*, vol. 367-368, pp. 209-220, 2016.
9. D. L. Anthony, T. Stablein, "Privacy in practice: professional discourse about information control in health care," *J Health Organ Manag*; vol. 30, no 2, pp. 207-226, 2016.
10. I. Carrión, J. L. Fernández-Alemán, A. Toval A, "Are personal health records safe? A review of free web-accessible personal health record privacy policies," *J Med Internet Res*, vol. 14, no 4, pp. e114, 2012.
11. The European Parliament and The Council of The European Union, "Regulation (UE) 2016/679 of the European parliament and of the council of 27 April on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)," May 4,

El acceso seguro del ciudadano a su Historia Digital permite avanzar en la gestión de Procesos Asistenciales

- «DOUE» num. 119, L119/1-88. [Online]. 2016. Available: <https://www.boe.es/doue/2016/119/L00001-00088.pdf>.
12. Ministerio de Sanidad, Servicios Sociales e Igualdad, "ProyectoHCDSNS: La Historia Clínica Digital del Sistema Nacional de Salud," *Ministerio* [Online]. 2017. Available: <http://www.msssi.gob.es/profesionales/hcdsns/>.
 13. http://www.csalud.juntadandalucia.es/salud/sites/csalud/contenidos/Informacion_General/p_3_p_3_procesos_asistenciales_integrados/
 14. Procesos asistenciales integrados: una herramienta para la mejora continua de la asistencia sanitaria Integrated care processes: a tool for the continuing improvement of healthcare M. C.. del Moral Campaña³ *Semergen, medicina de familia* Vol. 35. Núm. 6. Junio 2009
 15. I. Carrión, J. L. Fernández-Alemán, A. Toval A, "Are personal Health records safe? A review of free web-accessible personal health record privacy policies," *J Med Internet Res*, vol. 14, no 4, pp. e114. 2012.
 16. A. Abbas, S. U. Khan, "A Review on the State-of-the-Art Privacy-Preserving Approaches in the e-Health Clouds", *IEEE J. Biomed. Health Inform.*; Vol. 18, no. 4, pp. 1431-1441. 2014 DOI: 10.1109/JBHI.2014.2300846
 17. European Commission, "Commission Recommendation of 2nd July 2008 on cross-border interoperability of electronic health records systems. Explanatory Memorandum Scope of the Recommendation," *Official Journal of the European Union* L190, pp. 37-43. [Online]. 2008. Available: <http://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:32008H0594&from=EN>
 18. Ministry of Health and Social Affairs Sweden, "Vision for eHealth 2025– common starting points for digitization of social services and healthcare," *Government Offices of Sweden* [Online]. 2016. Available: <https://www.ehalsomyndigheten.se/globalassets/dokument/vision/visionfor-ehealth-2025.pdf>.
 19. Silva aycaguerlc .Muestreo para la investigacion en ciencias de la salud .Madrid ,diaz de santos , 1993 , 174p
 20. Consejería de Salud. http://www.csalud.junta-andalucia.es/salud/sites/csalud/contenidos/Informacion_General/p_3_p_3_procesos_asistenciales_integrados/pai/cancer_cervix_v3?perfil=org. PROCESO ASISTENCIAL INTEGRADO CÁNCER DE CÉRVIX. 2ª edición, 2010
 21. http://www.csalud.junta-andalucia.es/salud/sites/csalud/contenidos/Informacion_General/p_3_p_3_procesos_asistenciales_integrados/pai/embarazo_parto_puerperio_v3?perfil=org PROCESO ASISTENCIAL INTEGRADO EMBARAZO, PARTO Y PUERPERIO. 3ª edición, 2014
 22. La eSalud.com. <http://laesalud.com/category/wearables/>
 23. IM Medico. Los wearables médicos que están revolucionando la salud. <http://www.immedicohospitalario.es/noticia/12502/los-wearables-medicos-que-estan-revolucionando-la-salud>. 19 Oct 2017

Resolución de problemas de localización competitiva mediante procedimientos basados en GIS

Mónica Loaiza

Unidad Académica de Ciencias Sociales, Universidad Técnica de Machala, Km 5 ½ Vía a Pasaje, Machala, Ecuador
mloaiza@utmachala.edu.ec

Resumen. En la actualidad se puede emplear actividades en el ámbito geoespacial e intentar resolver los problemas de tipos de localidad para empresas públicas y privadas teniendo en cuenta la realidad del entorno para la toma de decisiones en la optimización del problema integrándose dos áreas importantes de estudio e investigación como son Matemáticas y Sistemas de Información Geográfico. En la primera fase del plan investigación, el uso de los sistemas de información geográfico como recurso tecnológico nos va a permitir experimentar insertar modelos matemáticos realizando cálculos de rutas utilizando las herramientas QGIS, PostGIS,

Palabras Clave: QGIS, PostGIS, geoespacial, SIG, Locación, Modelo matemático

Abstract. Currently, activities can be used in the geospatial area and try to solve the problems of locality types for public and private companies taking into account the reality of the environment for decision making in the optimization of the problem integrating two important areas of study and research such as Mathematics and Geographic Information Systems. In the first phase of the research plan, the use of geographic information systems as a technological resource will allow us to experiment with inserting mathematical models by calculating routes using the QGIS, PostGIS, Keywords: QGIS, PostGIS, geospatial, GIS, Location tools.

Keywords: QGIS, PostGIS, geospatial, GIS, Location, Mathematical Model

1 Introducción

Sistema de Información Geográfica (GIS) es una herramienta geoespacial en la investigación; para las administraciones públicas y privadas puede ser de fácil uso que permite a personas no expertas poder tener información geoespacial pudiendo manipular de acuerdo con la necesidad y requerimientos del usuario a consultar representados los resultados en los mapas. La investigación consiste en diseñar un sistema global eficiente y efectivo de algoritmos de optimización para resolver problemas de localización competitivos utilizando arquitecturas de alto rendimiento y uno de los factores es incorporar el modelo matemático de locación en QGIS. La dificultad es integrar métodos matemáticos, enfoques entre uno de éstos el modelo matemático en QGIS para resolver el problema de localización en la herramienta es lo que permite estructurar y tomar como base el proyecto de las Islas canarias para hacer un análisis

Resolución de problemas de localización competitiva mediante procedimientos basados en GIS

pretendiendo definir la localidad de forma eficiente teniendo en consideración para una organización en que se podrá atender a la demanda de éstos.

En la primera fase de investigación se ha ido con tareas para poder cumplir los objetivos planteados definiendo los problemas de ubicación sector privado, problemas complejos de ubicación sector público y privado teniendo en cuenta los modelos de ubicación siendo factores claves teniendo en cuenta que deben ser determinadas para determinar la ubicación particular del problema.

2. Justificación

En la actualidad se puede encontrar artículos con dificultad de localización y han podido ser resueltos con las herramientas de QGIS, sin embargo, proyectos de insertar modelos matemáticos son escasos por lo que la investigación a realizar se propone utilizar las herramientas geoespaciales principales como QGIS y PostGIS para aportar en el proyecto del plan de investigación a desarrollar, se pretende cumplir con la investigación de forma pertinente en las primeras fases propuestas

3. Definiciones

Sistemas de Información Geográfica (SIG) es un conjunto de software, hardware, datos geográficos y personal idóneo, sirve organizar para analizar, almacenar, presentar todo tipo de información, consultar información que tenga referencia geográfica existiendo SIG como una base de datos espaciales

Existen formatos de almacenamiento de datos espaciales los objetos de trabajo pueden ser ráster que son archivo de imagen y vectorial que son cobertura. El Primer formato es conocido como imagen que está representado por fenómenos geográficos que varían en el espacio teniendo en cuenta la altitud, pendiente del terreno combinan. Cuanto mas pequeño que sea el tamaño del pixel más precisa será la respuesta. El vectorial se crea a partir de las fuentes de información espacial existente.

Investigación Científica

La investigación científica corresponde a “La investigación y el desarrollo experimental, que comprenden el trabajo creativo llevado a cabo de forma sistemática para incrementar el volumen de conocimientos, incluido el conocimiento del hombre, la cultura y la sociedad, y el uso de esos conocimientos para crear nuevas aplicaciones.” El término Investigación y desarrollo experimental engloba tres actividades: investigación básica, investigación aplicada y desarrollo experimental:

Algoritmo de Weiszfeld, Es la técnica más sencilla y más recurrida para resolver el problema de Weber detallando a continuación, La función Algoritmo Matemáticos

Resolución de problemas de localización competitiva mediante procedimientos basados en GIS

para localidad del modelo de ubicación según el mapa en SIG, son tres niveles de instrumentos para organizar la información:

$$W(x, y) = \sum_{i=1}^m w_i \cdot \sqrt{(x - a_{i1})^2 + (y - a_{i2})^2}$$

Se impone la condición de optimalidad de primer orden, obtenemos el siguiente sistema de ecuaciones no lineales:

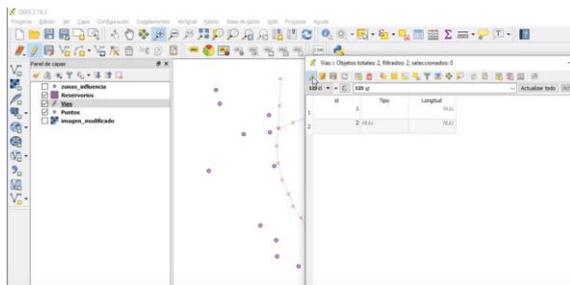
$$\frac{\partial W(x, y)}{\partial x} = \sum_{i=1}^m \frac{w_i \cdot (x - a_{i1})}{\sqrt{(x - a_{i1})^2 + (y - a_{i2})^2}} = 0$$

Existen reglas de nomenclatura:	Los datos espaciales deben ser nombrados según las reglas que detallan al máximo de qué tipo de datos se trata. El nombre del archivo informa sobre su contenido y la extensión informa sobre el formato de los datos.
Uso de metadatos:	El los metadatos describen con detalle los datos espaciales: fecha de creación, organismo responsable, datos de proyección, etc.

QGIS; Un sistema de coordenadas es un sistema de referencia que se utiliza para representar la ubicación de entidades geográficas, imágenes y observaciones como las localizaciones

En Gis utilizando los sistemas de coordenadas globales o esféricos, basados por ejemplo en latitud-longitud

Los sistemas de coordenadas proyectadas que se basan en proyecciones cartográficas como la Mercator transversal, la equivalente de Albers o la de Robinson, que (junto con otros muchos modelos de proyección cartográfica) ofrecen diversos mecanismos para la proyección cartográfica de la superficie esférica del globo sobre un plano de coordenadas cartesianas de dos dimensiones.



La herramienta encargada de la administración de los datos en ArcGIS es ArcCatalog, que

Resolución de problemas de localización competitiva mediante procedimientos basados en GIS

ayuda a organizar y manejar todos los datos, incluyendo herramientas para explorar y encontrar la información geográfica, almacenar y ver los metadatos, visualizando rápidamente cualquier archivo (datos geográficos, tablas, imágenes, etc.).

Muestra de figura de localidad en Argentina



Fig. 2. Mapa de localización del Centro Tecnológico de Desarrollo Regional "Los Reyunos". Está emplazado a 25 km. de la ciudad de San Rafael. Se puede consultar la página www.reyunos.utn.edu.ar para conocer la localización y comodidades del lugar donde se realizará el VI Enidi.

4. Metodología

En una investigación experimental debe ser una metodología cuantitativa. Según Rodríguez Peñuelas (2010, p.32), señala que el método cuantitativo se centra en los hechos o causas del fenómeno social, con escaso interés por los estados subjetivos del individuo. Este método utiliza el cuestionario, inventarios y análisis demográficos que producen números, los cuales pueden ser analizados estadísticamente para verificar, aprobar o rechazar las relaciones entre las variables definidas operacionalmente, además regularmente la presentación de resultados de estudios cuantitativos viene sustentada con tablas estadísticas, gráficas y un análisis numérico.

5. Avances.

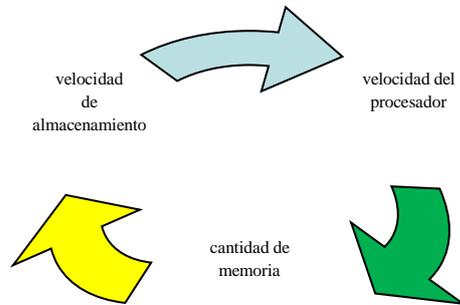
En el desarrollo de la primera fase se descargan las herramientas geoespaciales, utilizando las definiciones básicas de GIS. Las aplicaciones de interés QGIS y PostGIS se han empezado desde cero en el estudio, manejo, introducción e integración de las herramientas a utilizar, integrar el algoritmo propuesto para resolver la mejor localidad de empresa pública o privada.

Durante la investigación se ha obtenido recursos con dificultades dando oportunidad a la investigación empezando la búsqueda de información para la creación de la primera fase y tratar de cubrir los objetivos planteados de esas tareas.

Para el desarrollo de la primera fase se tomó como referencia el modelo matemático integrado a QGIS de las Islas canarias como prototipo.

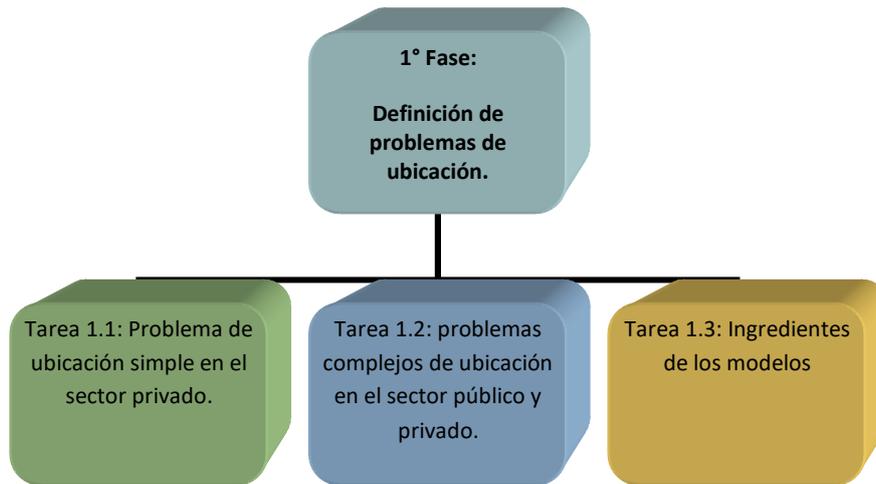
Resolución de problemas de localización competitiva mediante procedimientos basados en GIS

Limitaciones existe con respecto al hardware:



Las tres características son muy importantes determinar el tiempo de procesar las imágenes teniendo en cuenta que influye la efectividad del tipo de algoritmo que se utiliza al integrar el modelo matemático propuesto y empezar a tener resultados.

Organigrama de tarea de la primera fase de la investigación en curso.



La comprobación de la integración del algoritmo es tener el mapa de la Islas Canarias teniendo en cuenta de posibles errores por las imágenes y no tener el mapa completo con sus capas para el trabajo a realizar.

Teniendo como referencia el caso de las Islas Canarias con otros autores que están siendo relevantes y muy parecido al estudio a realizar, en la investigación se tiene como resultado el prototipo, viendo las dificultades que se presentarían en el caso de muestra.

Resolución de problemas de localización competitiva mediante procedimientos basados en GIS

4. Conclusiones.

En el transcurso de la investigación con las aplicaciones libres de GIS y la investigación de estudio QGIS Y PostGIS son herramientas que ayuda a la gestión análisis, y el avance del estudio con herramienta SIG de código abierto, se puede decir que puede estar orientada a la gestión, análisis, aprovechamiento de mapas y bases de datos para almacenar la información que se van presentando, pudiendo registrar las posibles formas de integrar algún modelo matemático propuesto por la localidad competitiva para empresas públicas y privadas.

Al realizar las tareas de las actividades propuesta se puede decir que esta fase se convierte en herramientas accesibles y potentes, los datos territoriales técnicos para el desarrollo y avance de la investigación

Durante el tiempo de investigación se ha revisado y realizado muestras de funcionamiento constantemente más exploración y pruebas con la planificación propuesta en las que se trata de cumplir con el objetivo propuesto de la fase en curso.

Con el tiempo se podrá realizar registros de modificaciones de datos con mayor posibilidad de localidad de empresas públicas o privadas.

5. Referencias

1. Juliana N, Henry L.: Modelo matemático para determinar la ubicación de Centros de Distribución en un contexto real. 386, 387 (2014)
2. Manejo Básico de QGIS para el uso en Sistemas de Información Geográfica en ANP visitado en la web proambiente.org.pe

Computación de altas prestaciones aplicada a la detección de anomalías en imágenes hiperespectrales

José Manuel Molero Pérez

Dpto. de Informática,
jmp384@ual.es

Resumen La detección de anomalías es una tarea importante en el campo de las imágenes hiperespectrales. Uno de los algoritmos más usados y extendidos para detección de anomalías es el algoritmo RX. A pesar de que es aceptado y computacionalmente complejo, existe poca documentación sobre implementaciones optimizadas y paralelizadas de este algoritmo, en particular para arquitecturas multicore y GPU. Se tratan aspectos críticos en la implementación paralela, como el cálculo de la matriz de covarianza, para optimizar el algoritmo. Se proponen optimizaciones computacionales, desarrollando nuevas implementaciones paralelas y estudiando su impacto en términos de precisión y de rendimiento del algoritmo propuesto. Evaluando los resultados haciendo uso de un análisis ROC, para analizar su capacidad de detectar objetivos en imágenes hiperespectrales reales, y del tiempo de ejecución para evaluar el rendimiento.

1. Introducción

En los últimos años, el concepto de imagen hiperespectral ha adquirido una gran relevancia en el campo de la teledetección, abriendo nuevas posibilidades en el estudio detallado de la superficie terrestre [1]. Una imagen hiperespectral está compuesta por un conjunto de decenas o cientos de imágenes de la misma área, que recogen información a distintas longitudes de onda, proporcionando cada imagen información de una longitud de onda diferente. Como podemos ver en la Figura 1, las imágenes hiperespectrales se pueden considerar como una imagen cubica compuesta por tres ejes: (1) muestras o “samples”, que son el conjunto de píxeles que un sensor puede capturar en un instante determinado; (2) líneas, como el número de muestras que un sensor puede recoger en un barrido; y (3) bandas, en cada una de ellas se almacena información sobre una frecuencia espectral concreta para cada píxel de la imagen [2].

De esta manera, cada píxel está formado por un conjunto de valores a través de todas sus bandas, caracterizando de forma única al elemento que representa en la escena. Esto se conoce como firma espectral y permite caracterizar de forma única a los objetos o elementos que han sido capturados por el sensor.

Existen muchas aplicaciones que pueden aprovechar la gran cantidad de información espacial y espectral que nos proporcionan este tipo de imágenes. Así,

Computación de altas prestaciones aplicada a la detección de anomalías en imágenes hiperespectrales

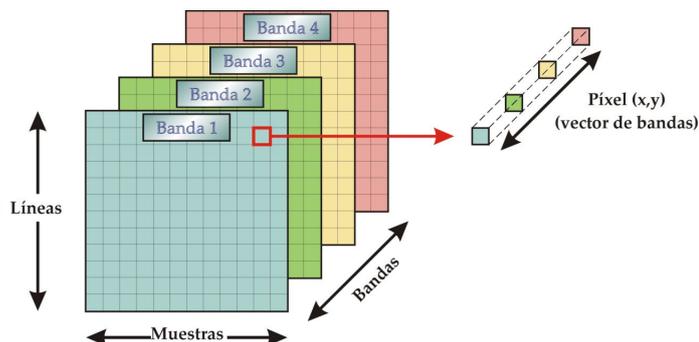


Figura 1. Estructura conceptual de imagen hiperespectral.

la detección de anomalías sobre imágenes hiperespectrales es considerada como uno de estos grandes ámbitos de aplicación, siendo un campo de gran interés y relevancia en el campo del procesamiento de imágenes hiperespectrales. En este contexto, podemos considerar que un algoritmo detector de anomalías es aquel que nos indica, partiendo de una imagen de entrada y sin ninguna otra información a priori, aquellos elementos que son espectralmente diferentes en relación a lo que les rodea espacialmente [3].

Por otra parte, en la actualidad las imágenes hiperespectrales pueden llegar a ser de un gran tamaño, lo que supone manejar datos del orden de GBs. Además del tamaño, su cantidad también aumenta de forma exponencial. Por estas razones, se hace necesario desarrollar estrategias eficientes de procesamiento, postulándose el procesamiento computacional paralelo como un requisito imprescindible en este ámbito. Más aún si lo que se persigue es el procesamiento de la imagen en tiempo real, es decir, procesar la imagen al mismo tiempo que se obtiene directamente desde el sensor [4].

Además, la optimización y el procesamiento paralelo no solo permiten ejecutar en menor tiempo algoritmos de detección de anomalías clásicos. También se hace indispensable su uso cuando se trata de algoritmos cuyo coste y tiempo de ejecución es alto, permitiendo además explorar nuevas variantes y nuevos algoritmos, que de otra forma y sin la aplicación del paralelismo, serían descartados por su elevado tiempo de ejecución y coste computacional [5].

En este ámbito, las arquitecturas de alto rendimiento y procesamiento paralelo actuales, posibilitan la extensión y aceleración de algoritmos de procesamiento. No obstante, es necesario un esfuerzo en la selección de las arquitecturas e interfaces de programación, así como un exhaustivo conocimiento técnico para llevar a cabo una implementación eficiente [6]. En este trabajo, evaluamos plataformas de computación como los procesadores multi-core o CPUs y tarjetas gráficas o GPUs [7].

El presente documento está organizado de la siguiente manera, la Sección 2 describe el concepto y el algoritmo de detección de anomalías, la Sección 3 describe la implementación optimizada y eficiente para plataformas de cómputo de

Computación de altas prestaciones aplicada a la detección de anomalías en imágenes hiperespectrales

los algoritmos descritos, finalmente la Sección 4 muestra los resultados obtenidos durante los experimentos en términos de precisión y rendimiento.

2. Detección de anomalías

Una anomalía estaría formada por un conjunto de firmas espectrales suficientemente representativas y diferentes de los objetos presentes a su alrededor en la imagen.

La detección de anomalías se centra en aquellas técnicas que tratan de detectar elementos en una imagen, que no tienen relación con lo que les rodea, y sin tener información a priori del espectro de los objetivos [3].

Tradicionalmente, los algoritmos de detección de anomalías suelen usar modelos o estrategias para caracterizar el fondo de la imagen. Según sea la estrategia usada durante el proceso de caracterización, podemos diferenciar dos grandes grupos: los que usan una estrategia global, es decir, que consideran toda la imagen como un único conjunto; y los que usan una estrategia local, que solo consideran subconjuntos de la imagen que rodean a la zona de la imagen bajo estudio.

2.1. El algoritmo RX

El algoritmo desarrollado por Reed y Xiaoli, conocido como algoritmo RX, sigue una estrategia global y está basado en la distancia de Mahalanobis. Es uno de los algoritmos de detección de anomalías más extendidos y es usado en la comunidad científica como algoritmo de referencia [8]. Tomando como punto de partida este algoritmo, se proponen optimizaciones basadas en él, que mejoran su calidad de los resultados en términos de detección de anomalías y su rendimiento desde el punto de vista computacional [8].

El algoritmo RX está definido por la siguiente expresión:

$$\delta(\mathbf{x}) = \mathbf{x}^T R^{-1} \mathbf{x}$$

donde \mathbf{x} es un pixel hiperespectral o vector del tamaño de las bandas que contiene la imagen y R es la matriz de correlación correspondiente a la imagen. La matriz de correlación viene dada por:

$$\mathbf{R} = \frac{H^T \cdot H}{n}$$

donde H es la imagen hiperespectral y n el número total de píxeles (muestras por líneas). De esta manera, el algoritmo RX puede ser considerado como un detector global, ya que la matriz de correlación es obtenida usando todos los píxeles de la imagen.

Computación de altas prestaciones aplicada a la detección de anomalías en imágenes hiperespectrales

2.2. El algoritmo RX local

A continuación, se presenta la versión local del algoritmo, en la que por cada pixel de la imagen, se obtiene una matriz de correlación obtenida a partir de un pequeño subconjunto de pixeles que rodean al pixel a estudiar. Así, se propone el algoritmo RX desde una aproximación local o RX local, es decir, aplicando el proceso a una ventana local y deslizante al pixel bajo estudio [9].

De esta manera, el algoritmo RX local viene dado por la siguiente fórmula:

$$\delta(\mathbf{x}) = \mathbf{x}^T \mathbf{R}_\kappa^{-1} \mathbf{x},$$

donde κ determina el tamaño del subconjunto de pixeles que formarán parte del estudio. De la misma manera, la matriz de correlación queda definida de la siguiente manera:

$$\mathbf{R}_\kappa(\mathbf{x}) = \frac{H_\kappa^T \cdot H_\kappa}{n_\kappa}$$

donde κ indica el subconjunto de pixeles incluidos en la ventana local deslizante.

Esta versión puede obtener mejores resultados, pero conlleva un mayor coste computacional, por lo que para abordar este problema es indispensable una implementación eficiente.

2.3. Optimización de los algoritmos RX global y local

Con el objetivo de reducir el coste computacional, se proponen las siguientes optimizaciones [10]. Considerando las siguientes etapas:

- Cálculo de la matriz de correlación R . El coste puede ser reducido aprovechando que la matriz es simétrica. Por esta razón, solo es necesario calcular el conjunto triangular superior o inferior. Además de aplicar las optimizaciones mencionadas en [10].
- Cómputo del filtro $\delta(\mathbf{x})$. En lugar de calcular la inversa o pseudoinversa de la matriz, teniendo en cuenta que la matriz es simétrica y definida positiva, proponemos la siguiente optimización basándonos en la factorización de Cholesky:
 - Descomponer la matriz R en un producto de dos matrices triangulares $R = U^T U$
 - Resolver el sistema triangular $U^T z(x) = x$ para el vector $z(x)$
 - Finalmente, calcular el producto vectorial $\delta(x) = z(x)^T z(x)$ para cada pixel de la imagen.

De esta manera, se consigue reducir el coste computacional en comparación con la aproximación clásica, en etapas que suponen aspectos computacionales críticos en la implementación. Además, esta optimización es especialmente importante en la implementación del algoritmo RX local, permitiendo la escalabilidad del algoritmo y su aplicación práctica.

Computación de altas prestaciones aplicada a la detección de anomalías en imágenes hiperespectrales

3. Implementación y paralelización de los algoritmos RX global y local

En general, las necesidades computacionales del procesado de imágenes hiperespectrales son muy elevadas, ya que las estructuras de datos tridimensionales que se procesan necesitan grandes recursos de memoria y su procesamiento requiere elevadas capacidades de cómputo. Más aún si consideramos aproximaciones locales de los filtros, ya que aumentan la carga computacional y los requisitos de memoria [11].

Además de ser un proceso computacionalmente costoso, también requiere de un tiempo de respuesta bajo para su interés práctico. Por lo que, si se pretende mejorar los resultados en términos de detección, conlleva añadir complejidad computacional y tiempo de cómputo al proceso [12,13].

Sin embargo, los recientes avances en la computación de alto rendimiento y su conocimiento, pueden ayudar a reducir el tiempo de ejecución de este tipo de algoritmos, si se aprovechan de forma adecuada los recursos computacionales actuales, como los multiprocesadores o CPUs y las tarjetas gráficas o GPUs [14,15].

Por estas razones, a continuación se presentan nuevas versiones eficientes y paralelizadas del algoritmo RX global y RX local, para diferentes arquitecturas de altas prestaciones como los procesadores multicore y GPUs [10].

3.1. Implementación sobre CPU

Para esta implementación, la librería OpenMP ha sido usada para desarrollar una versión paralela. Para resolver las diferentes ecuaciones se ha hecho uso de la librería Intel MKL, que está optimizada para aprovechar este tipo de arquitecturas [16]. La correlación (etapa 1) se calcula usando la función `dsyrk`. Para la factorización (etapa 2), usamos la rutina `dpotrf` y `dtrtrs` para obtener el vector intermedio. Finalmente, usamos la operación `ddot` en un bucle para calcular la operación vector-vector para todos los píxeles [14,10].

3.2. Implementación sobre GPU

En el caso de la implementación GPU, también se ha seguido la estrategia definida anteriormente de dos etapas. En la primera etapa, la función `dsyrk` de la librería CUBLAS (incluida en NVIDIA CUDA SDK [17]) calcula el producto matriz-matriz para obtener la correlación. Para la segunda etapa se requiere el uso de un conjunto de funciones a medida si queremos aprovechar y adaptar el cómputo a las características de la plataforma. Cada factorización de Cholesky es realizada por un bloque de la GPU de forma independiente y paralela, con el objetivo de evitar latencias y penalizaciones durante el proceso [18,19]. Finalmente, la resolución de cada sistema lineal y cada pixel de la imagen es asociado a un bloque de cómputo la GPU y cada banda del vector a un thread diferente. Así, los threads pueden cooperar en la resolución del sistema triangular independiente y evitando las costosas comunicaciones entre bloques.

Computación de altas prestaciones aplicada a la detección de anomalías en imágenes hiperespectrales

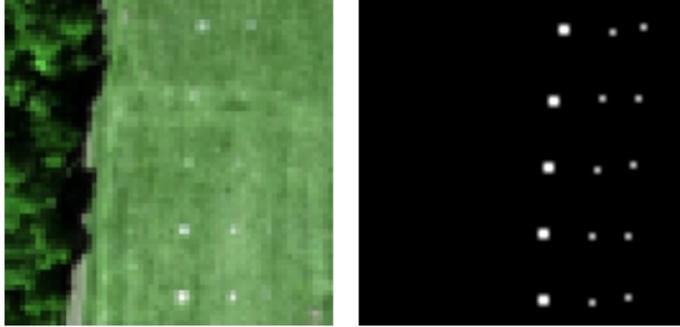


Figura 2. Representación de la imagen HYDICE (izquierda). Localización real de los paneles, imagen usada como verdad terreno (derecha).

4. Evaluación de resultados

Para este trabajo, una imagen hiperespectral real ha sido usada en los diferentes experimentos, con el objetivo de evaluar la precisión del algoritmo y analizar el rendimiento de su implementación. Esta imagen, llamada HYDICE (ver Figura 2), está compuesta por 64 líneas, 64 muestras y 169 bandas espectrales. En ella se puede observar una región de terreno donde 15 paneles fueron colocados en 3 columnas con 5 filas. El tamaño de los paneles es de 3×3 , 2×2 y 1×1 metro(s) respectivamente en cada columna. Siendo la resolución espacial de la imagen de 1,56 metros por pixel [20].

4.1. Análisis ROC

Al no usar información a priori, el detector de anomalías no permite distinguir entre anomalías de interés, y detecciones que no son de nuestro interés, lo que se puede considerar como detecciones falsas, aunque ofrecidas como positivas. Pero también darse el caso contrario, que el detector pase por alto anomalías que si serían de nuestro interés.

Una curva ROC (Receiver Operating Characteristics), se considera un estándar de validación de detectores de anomalías y representa de forma gráfica la probabilidad de detección contra la probabilidad de encontrar un falso positivo [21].

La probabilidad de detección (P_D) y la probabilidad de detectar un falso positivo (P_{FA}) vienen definidos por:

$$P_D = \frac{N_{HIT}}{N_T}; \quad P_{FA} = \frac{N_{MISS}}{N_{TOT}},$$

donde N_{HIT} representa el número de pixeles objetivo detectados dado un cierto umbral; N_T denota el número total de pixeles objetivo en la imagen; N_{MISS} es el número de pixeles del fondo clasificados como objetivo incorrectamente; y N_{TOT} representa el número total de pixeles en la imagen.

Computación de altas prestaciones aplicada a la detección de anomalías en imágenes hiperespectrales

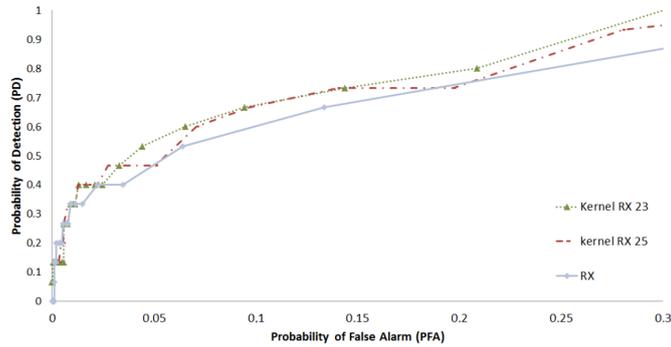


Figura 3. Curvas ROC de la imagen HYDICE usando el algoritmo RX y el RX local con $\kappa = 23$ y $\kappa = 25$.

Cuadro 1. Tiempo de procesamiento de la versión secuencial y paralelizada (en CPU y GPU) del algoritmo RX.

Imagen	Plataforma	Tiempo (en segundos)	
		1 core	Max. cores
HYDICE (64x64x169)	i7 (6)	0.03	0.02
	Tesla k20		0.02

Para generar las curvas ROC, se necesita comparar el resultado obtenido con la imagen verdad terreno (ground-truth), esta imagen contiene la posición real de los paneles mencionados anteriormente. Además, para poder comparar las diferentes curvas ROC, nos basamos en el área bajo la curva (“Area under the curve” or AUC). A mayor área bajo la curva, mejor es la precisión de detección de anomalías del algoritmo seleccionado.

En el Gráfico 3 podemos observar como las versiones locales (kernel) ofrecen una mayor precisión en la detección ya que el área bajo la curva es mayor en todos los casos. Además, a partir de una probabilidad de error de 0,3 la probabilidad de detección es casi 1, lo que significa que todas las anomalías objetivo definidas en la imagen verdad-terreno son localizadas.

4.2. Evaluación de rendimiento

La implementación paralela de los algoritmos RX global y RX local ha sido evaluada en términos de rendimiento en una plataforma CPU multicore y en una plataforma equipada con una GPU. La plataforma CPU está compuesta por un procesador Intel i7 4960 con 6 cores a 3,6 GHz y 16 GB de memoria RAM. La GPU de la segunda plataforma es una NVIDIA Tesla K20 con 2.496 cuda cores a 706 MHz y 5 GB de memoria [22,23].

El Cuadro 1 muestra el tiempo de ejecución durante el procesamiento de la imagen HYDICE, aplicando el algoritmo RX. Evaluando la versión CPU com-

Computación de altas prestaciones aplicada a la detección de anomalías en imágenes hiperespectrales

Cuadro 2. Tiempo de procesamiento de la versión secuencial y paralelizada (en CPU y GPU) del algoritmo RX local.

Imagen	Plataforma	Tiempo (en segundos)					
		1 core			Max. cores		
		Etapa 1	Etapa 2	Total	Etapa 1	Etapa 2	Total
HYDICE (64x64x169)	i7 (6)	6.64	0.65	7.30	0.91	0.47	1.39
	Tesla k20				1.23	0.72	1.95

paramos una versión secuencial y la versión paralela que hace uso de todos los cores disponibles. En cambio, en la versión GPU, solo una versión paralela es evaluable debido a las características de la arquitectura. El resultado obtenido es similar, con una cierta ventaja para la versión CPU, lo que confirma que ambas plataformas son válidas para este tipo de computación y las optimizaciones e implementaciones se adaptan a las arquitecturas seleccionadas.

El Cuadro 2 muestra el tiempo de ejecución durante el procesamiento de la imagen HYDICE, aplicando el algoritmo RX local. Además, en este caso también se detalla el tiempo consumido por cada etapa computacional descrita en las Secciones 2 y 4. Así podemos observar como la etapa 1 conlleva más tiempo de ejecución para ambas arquitecturas y como aun siendo el tiempo de ejecución más alto que para la versión global, el tiempo total es reducido siendo similar al de la versión global, gracias a la optimización e implementación realizada. Igual que anteriormente, los resultados de la arquitectura CPU son ligeramente mejores, confirmando que ambas arquitecturas combinadas con una implementación eficiente, son candidatas válidas en este contexto.

5. Conclusiones

En este informe se recogen avances durante el estudio de un algoritmo de detección de anomalías, y su implementación optimizada y paralelizada en sistemas de computación moderno basados en CPU y GPU. Se ha demostrado como mejorar y optimizar el seleccionado algoritmo RX (y su versión local) con el objetivo de reducir su coste computacional. De esta manera se consigue: (1) reducir el tiempo de ejecución, (2) procesar imágenes de mayor tamaño, o un mayor número de imágenes en menor tiempo y (3) explorar nuevos métodos que permitan mejorar el rendimiento y la precisión del citado algoritmo. Para esto se ha realizado un estudio exhaustivo del algoritmo y de su aplicación en el campo de la detección de anomalías. Por otra parte, también hay que destacar el conocimiento técnico aplicado sobre computación de altas prestaciones y su aplicación práctica a un problema real. En conclusión, se ha revisado un método de detección de anomalías en imágenes hiperespectrales con su implementación eficiente y paralela sobre arquitecturas que permiten reducir el tiempo de ejecución. Como trabajo futuro, se propone estudiar el consumo energético durante

Computación de altas prestaciones aplicada a la detección de anomalías en imágenes hiperespectrales

la computación y así estudiar el impacto de una implementación eficiente sobre la energía requerida por la plataforma.

Agradecimientos

Este trabajo ha sido financiado por el Ministerio Español de Ciencia e Innovación (TIN2015-66680-C2-1-R).

Referencias

1. Plaza, A., Benediktsson, J., Boardman, J., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, J., Marconcini, M., Tilton, J.C., Trianni, G.: Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment* **113** (2009) 110–122
2. Richards, J.A., Jia, X.: *Remote Sensing Digital Image Analysis: An Introduction*. Springer (2006)
3. Chang, C.I., Chiang, S.S.: Anomaly detection and classification for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing* **40**(6) (2002) 1314–1325
4. Plaza, A., Chang, C.I.: *High Performance Computing in Remote Sensing*. CRC Press, Boca Raton (2007)
5. Molero, J.M., Garzón, E.M., García, I., Plaza, A.: Anomaly detection based on a parallel kernel RX algorithm for multicore platforms. *Journal of Applied Remote Sensing* **6** (2012) 12 DOI:0091-3286/2012.
6. Paz, A., Plaza, A., Plaza, J.: Comparative analysis of different implementations of a parallel algorithm for automatic target detection and classification of hyperspectral images. In: *Proc. SPIE. Volume 7455*. (2009) 74550X–74550X–11
7. Paz, A., Plaza, A.: Clusters versus GPUs for parallel target and anomaly detection in hyperspectral images. *EURASIP J. Adv. Signal Process* **2010** (February 2010) 35:1–35:18
8. Reed, I.S., Yu, X.: Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution. *IEEE Trans. Acoustics, Speech and Signal Processing* **38** (1990) 1760–1770
9. Taitano, Y.P., Geier, B.A., Bauer, K.W.: A locally adaptable iterative RX detector. *EURASIP Journal on Advances in Signal Processing* (2010) 10 DOI:10.1155/2010/341908.
10. Molero, J.M., Garzón, E.M., García, I., Plaza, A.: Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **6**(2) (2013) 801–814
11. Lee, C., Gasster, S., Plaza, A., Chang, C.I., Huang, B.: Recent developments in high performance computing for remote sensing: A review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **4**(3) (2011) 508–527
12. Chang, C.I., Ren, H., Chiang, S.S.: Real-time processing algorithms for target detection and classification in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing* **39**(4) (2001) 760–768

Computación de altas prestaciones aplicada a la detección de anomalías en imágenes hiperespectrales

13. Remon, A., Sanchez, S., Paz, A., Quintana-Ortí, E.S., Plaza, A.: Real-time end-member extraction on multi-core processors. *IEEE Geoscience and Remote Sensing Letters* **8** (2011) 924–928
14. Molero, J.M., Paz, A., Garzón, E.M., Martínez, J.A., Plaza, A., García, I.: Fast anomaly detection in hyperspectral images with RX method on heterogeneous clusters. *The Journal of Supercomputing* **58**(3) (2011) 411–419 DOI:10.1007/s11227-011-0598-0.
15. Plaza, A., Chang, C.I.: Special issue on high performance computing for hyperspectral imaging. *International Journal of High Performance Computing Applications* **22**(4) (2008) 363–365
16. Intel: Math kernel library (MKL) documentation (2013)
17. Nvidia: CUBLAS manual (2013)
18. Molero, J.M., Garzón, E.M., García, I., Quintana-Ortí, E.S., Plaza, A.: Accelerating the KRX algorithm for anomaly detection in hyperspectral data on GPUs. In: *Proceedings of the 12th International Conference on Computational and Mathematical Methods in Science and Engineering (CMMSE)*. (2012) 860–863
19. Molero, J.M., Garzón, E.M., García, I., Quintana-Ortí, E.S., Plaza, A.: A batched Cholesky solver for local RX anomaly detection on GPUs. In: *Proceedings of the 13th International Conference on Computational and Mathematical Methods in Science and Engineering, (CMMSE)*. (2013) 1037–1797
20. Rickard, L.J., Basedow, R.W., Zalewski, E.F., Silverglate, P.R., Landers, M.: Hydice: an airborne system for hyperspectral imaging. Volume 1937. (1993) 173–179
21. Ranney, K., Soumekh, M.: Hyperspectral anomaly detection within the signal subspace. *Geoscience and Remote Sensing Letters, IEEE* **3**(3) (July 2006) 312–316
22. Nvidia: Kepler GK110 architecture whitepaper (2013)
23. Farber, R.: *CUDA, Application Design and Development*. Morgan Kaufmann, Waltham (2010)

Aportaciones al modelado y control climático de invernaderos

Ana Paola Montoya Rios

Centro mixto CIESOL, Campus de Excelencia Internacional Agroalimentario (ceiA3), Universidad de Almería, Ctra. Sacramento s/n, Almería 04120, España

apmontoy@gmail.com

Resumen. Se presentan aportaciones al control de la temperatura en invernaderos por medio del empleo de dos estrategias de control modernas. Empleando de las técnicas de identificación de sistemas y reducción de modelos se desarrollaron modelos simplificados de la temperatura de un invernadero y perturbaciones. Se desarrollaron y simularon controladores por adelanto para compensación de perturbaciones antes que estas afecten la dinámica del sistema empleando la técnica *G-H*. Por otra parte se desarrolló un modelo y control híbrido de la temperatura del invernadero durante el invierno, el cual emplea dos sistemas de calefacción complementarios, con dinámicas continuas y discretas. Se empleó una estrategia de control predictivo híbrido que permite optimizar el costo del control de la calefacción, el empleo de los sistemas de calefacción y desviación de la referencia deseada. Ambos controladores han sido probados en simulación con buenos resultados.

Palabras clave: Identificación de sistemas, reducción de modelos, control por Adelanto, Modelos lógicos de dinámica mixta (MLD), Control Híbrido Predictivo Basado en Modelo (HMPC).

1 Introducción

El cultivo bajo invernadero es una técnica de agricultura intensiva que permite obtener productos de excelente calidad fuera de temporada. Dicha técnica es empleada para suplir la necesidad de alimentos de una población creciente. Al interior de los invernaderos se genera un microclima que condiciona el crecimiento y productividad de cultivos [2]. La temperatura del aire interviene en gran cantidad de procesos fisiológicos realizados por las plantas, por lo que se requiere regular el comportamiento de ésta variable y mantenerla dentro de unos rangos establecidos.

El control de temperatura interior de invernaderos es complejo, debido a que [9]:

- Se encuentra correlacionada con otras variables con dinámicas diferentes.
- Los actuadores que se usan para modificar otras variables afectan a la temperatura de forma simultánea.

- Se presentan restricciones en los actuadores con que se controla
- Las temperaturas óptimas para el crecimiento de las plantas varían a lo largo de sus diferentes etapas de desarrollo tanto para periodos diurnos como nocturnos. entre otras complicaciones.

En lugares que presentan estaciones marcadas y en ocasiones en la zona mediterránea se hace necesario refrigerar el invernadero durante el verano, mientras que durante el invierno se requiere emplear un sistema de calefacción para proveer al cultivo las condiciones de temperatura necesarias para su crecimiento y desarrollo.

La técnica más empleada para la climatización estival de invernaderos es la ventilación natural, donde por medio de la apertura de las ventilaciones laterales y cenitales es posible modificar la temperatura del aire al interior del invernadero. Se han empleado diferentes estrategias de control para regular dicha variable y se ha comprobado que la temperatura al interior del invernadero es notablemente afectada por perturbaciones (variables climáticas externas), por lo que es necesario emplear estrategias de control que compensen dicha perturbaciones [14].

El control por adelantado permite generar señales de control basadas en las perturbaciones medibles, de forma que estas sean compensadas en el lazo de control antes que se afecte la dinámica del sistema [6].

En el caso del control de la temperatura nocturna por medio de calefacción también se han empleado diversas técnicas, en donde el sistema de calefacción empleado condiciona en cierta medida la técnica de control a implementar. En algunas ocasiones, cuando la temperatura es muy baja el sistema de calefacción del invernadero no alcanza a suministrar los requerimientos calóricos para mantener la temperatura del aire en los valores deseados, por lo que se hace necesario emplear sistemas de calefacción adicionales [3]. Para el caso de invernaderos que cuenten con calefactores de tuberías y aerotermos, se hace necesario emplear estrategias de control híbridas que permitan modelar y controlar ambos sistemas de calefacción, de forma que se optimice la conmutación entre ambos sistemas, teniendo en cuenta el consumo de combustible, el desgaste de los equipos y la desviación de la consigna de control. Las estrategias de control híbridas permiten controlar sistemas en los que intervienen dinámicas continuas y discretas [8].

Para realizar las simulaciones del comportamiento dinámico de la temperatura del invernadero y de los sistemas de control a diseñar, se requieren de modelos que se ajusten de forma acertada a dicha dinámica. En el caso del control de la temperatura en época estiva se requieren modelos de temperatura simplificados para la obtención de los controladores [13]. Para el desarrollo de controladores por adelantado (*FF*) se hace necesario contar con el modelo de la dinámica de la temperatura del aire al interior del invernadero y con los modelos de cada una de las perturbaciones medibles a ser compensadas [7].

En cuanto al control de híbrido de la calefacción, se requiere de un modelo dinámico de lógica mixta (*MLD*), que permitan mezclar las dinámicas continuas y discretas de los diferentes actuadores, variables y sus respectivas restricciones [5].

1.1 Hipótesis

La tesis doctoral pretende realizar contribuciones en el campo del modelado y control climático de invernaderos en las áreas de control de temperatura diurna por medio de ventilación natural y control de temperatura nocturna por medio de la combinación de dos sistemas de calefacción.

Se parte de la premisa que el control climático de invernaderos que se viene implementando mediante estrategias de control comerciales no tienen en cuenta las especificidades del problema de control multivariable planteado. Dicho problema ha sido resuelto tradicionalmente mediante algoritmos de control clásicos simples que no permiten la inclusión de restricciones, no considerando las perturbaciones medibles y no tiene en cuenta de manera explícita la inclusión de factores económicos en el cálculo de las señales de control. Además dichos algoritmos y los modelos que tradicionalmente se han usado para describir el comportamiento dinámico del invernadero no permiten realizar una representación formal de las interacciones entre las dinámicas continuas y discretas que ocurren en el invernadero, por lo que los algoritmos de control comerciales empleados tradicionalmente solo resuelven en cierta medida el problema de control planteado.

Considerando lo anterior se propone resolver el problema de control de la temperatura interna del invernadero por medio de la obtención de dos controladores avanzados, uno para la regulación diurna y otro para la nocturna. Dichos controladores deberán presentar una mejor relación desempeño-costo que la de los controladores comerciales, a la vez que deben tener en cuenta las restricciones reales del sistema y compensar las perturbaciones.

En primera instancia respecto al control de la temperatura interna durante el día, se considera que la ventilación natural es el mejor método para regular esta variable, debido al bajo costo de utilización que presenta este método de refrigeración. Se quiere demostrar que el uso de controladores por adelantado que compensen las principales perturbaciones medibles del invernadero junto con controladores PI convencionales, presentan un mejor comportamiento que los controladores comerciales.

Para el caso del control de la temperatura nocturna durante la noche en época invernal se pretende demostrar que:

- El uso de sistemas combinados de calefacción proveen alta flexibilidad en el control de la temperatura nocturna durante noches muy frías.
- El sistema invernadero puede ser modelado de manera formal mediante modelos híbridos usando representación MLD (*Mixed Logical Dynamical*) [5].
- La estrategia de control predictivo híbrido es una buena opción para realizar la regulación de la temperatura nocturna del invernadero cuando se usan dos sistemas de calefacción combinados, donde es posible optimizar el desempeño del controlador por medio de una representación formal que permite considerar, condiciones económicas, de conmutación y restricciones reales del invernadero y los diferentes procesos [8].

1.2 Objetivos de la tesis

El principal objetivo de esta tesis doctoral es el de realizar aportaciones significativas en el campo del control climático de invernaderos, concretamente en el control de la temperatura diurna por medio de ventilación natural usando algoritmos de control por adelanto, y el control de la temperatura nocturna en época invernal, por medio del uso combinado de dos sistemas de calefacción usando una estrategia de control híbrida.

Para llevarlo a cabo se requiere el cumplimiento de los siguientes objetivos específicos:

- Estudio y desarrollo de modelos lineales de cada uno de los componentes del sistema invernadero para diferentes épocas del año.
- Validación y simulación de modelos lineales contrastando con datos reales para diferentes épocas del año.
- Diseño de controladores por adelanto para el control de la temperatura interna del invernadero durante el día por medio de ventilación natural.
- Desarrollo del modelado híbrido del sistema invernadero usando la representación formal MLD.
- Obtención y sintonización de controlador híbrido para regulación de la temperatura nocturna del invernadero durante la época invernal, usando dos sistemas de calefacción combinados.
- Comparación del comportamiento de los controladores desarrollados comparando controladores comerciales tradicionales.

2 Materiales y métodos

La aplicación de estrategias de control avanzadas al sistema invernadero requiere contar con un modelo que represente el comportamiento dinámico del sistema para luego desarrollar controladores específicos para cada sistema. A continuación se detallan las técnicas necesarias para adelantar cada una de estas etapas:

2.1 Control por adelanto (FF)

Empleando datos reales recolectados durante el verano y usando la técnica de identificación de sistemas [10], se obtuvieron modelos Autorregresivos de alto orden de la dinámica de la temperatura y diversas perturbaciones de un invernadero. Dicho modelo fue reducido con fines de control usando la técnica propuesta por [13], resultando en un conjunto de modelos de primer orden con retardo de la dinámica de la temperatura y las perturbaciones que la afectan, como se aprecia en la Tabla 1 [12].

Usando dichos modelos se sintonizaron cuatro controladores *FF* por medio del método propuesto en [6], para compensar las perturbaciones mostradas en la Tabla 1. Se desarrollaron controladores tipo lead-lag con y sin retardo, usando un controlador PI, diseñado con el método λ [1], con $K_p = -62.77 \text{ } \%/^{\circ}\text{C}$ y $T_i = 43.51$ segundos, con el fin de controlar la temperatura del aire del invernadero por medio de ventilación natural en época estival.

Tabla 1. Variables, modelos reducidos y controladores por adelantado para el control de la temperatura del invernadero por medio de ventilación natural

Variable	Modelo Reducido	Controlador FF G-H
Temp. externa	$\frac{T_{inv}}{Text} = \frac{1}{42.26s + 1} e^{-s}$	$FF_Tex = -2.979 * \frac{43.51s + 1}{39.31s + 1}$
Velocidad del viento	$\frac{T_{inv}}{Vv} = \frac{-0.218}{17.52s + 1} e^{-s}$	$FF_Vv = 0.694 * \frac{43.51s + 1}{14.579s + 1}$
Radiación	$\frac{T_{inv}}{Rad} = \frac{0.0193}{48.4s + 1} e^{-s}$	$FF_Rad = -0.0573 * \frac{43.51s + 1}{45.449s + 1}$
Humedad	$\frac{T_{inv}}{HR} = \frac{0.0445}{43.7s + 1} e^{-s}$	$FF_HR = -0.133 * \frac{43.51s + 1}{40.759s + 1}$
Ventilación	$\frac{T_{inv}}{Ven} = \frac{-0.014}{43.51s + 1} e^{-6s}$	

2.2 Control híbrido

Se plantea regular la temperatura interna del invernadero usando dos sistemas de calefacción combinados, de forma tal que se minimice el consumo de combustible, el número de activaciones del arotermo y la desviación de la referencia. Este tipo de problemas de control se puede plantear de manera formal mediante *HMPC* (Control Híbrido Predictivo Basado en Modelo [15]). Dicho controlador se basa en un modelo *MLD* del sistema, el cual describe lo describe mediante ecuaciones lineales en diferencias y reglas lógicas.

Se desarrollaron ensayos de identificación de sistemas para obtener modelos reducidos de la temperatura del invernadero, variables asociadas y su relación con cada uno de los calefactores. Para el desarrollo y simulación de los modelos se emplearon las herramientas: Simulink, System Identification Toolbox [10], Hybrid Toolbox [4], e Hysdel [15] del software Matlab [11]. El modelo *MLD* resultante se puede apreciar en la Figura 1.

$$\begin{aligned}
 xt(k+1) &= \begin{cases} 0.9911xt(k) + 0.9955Tout(k) & \text{if } a1(k) \text{ is true} \\ 0 & \text{Otherwise} \end{cases} & c(k) &= \sim c(k) \wedge ((FAH(k) \wedge FAH(k-1)) \vee (FAH \wedge \sim FAH(K-1))) \\
 xc(k+1) &= \begin{cases} 0.9851xc(k) + 0.9925APH(k) & \text{if } (a2(k) \wedge a3(k)) \text{ is true} \\ 0.9851xc(k) & \text{Otherwise} \end{cases} & Y(k) &= 0.007686xt + 0.002667xa + 0.0009049xc + 1.628Ts - 0.006479xs \\
 & & & + 0.008824hr - 0.0002985zh + 0.0000416xr - 0.001594xw \\
 & & & - 0.0001929xv \\
 xa(k+1) &= \begin{cases} 0.9355xa(k) + 0.9674FAH(k) & \text{if } (FAH(k) \wedge a3(k)) \text{ is true} \\ 0.9355xa(k) & \text{Otherwise} \end{cases} \\
 xs(k+1) &= \begin{cases} 0.9948xs(k) + 0.9974Ts(k) & \text{if } a1(k) \text{ is true} \\ 0 & \text{Otherwise} \end{cases} \\
 xh(k+1) &= \begin{cases} 0.9948xh(k) + 0.9974hr(k) & \text{if } a1(k) \text{ is true} \\ 0 & \text{Otherwise} \end{cases} & a1(k) \text{ is true} & \leftrightarrow [Tout(k) > -10] \\
 xr(k+1) &= \begin{cases} 0.9943xr(k) + 0.9971Rad(k) & \text{if } a1(k) \text{ is true} \\ 0 & \text{Otherwise} \end{cases} & a2(k) \text{ is true} & \leftrightarrow [APH(k) > 0.001] \\
 & & & a3(k) \text{ is true} \leftrightarrow [Rad(k) > 5] \\
 xw(k+1) &= \begin{cases} 0.9845xw(k) + 0.9922wind(k) & \text{if } a1(k) \text{ is true} \\ 0 & \text{Otherwise} \end{cases} \\
 xv(k+1) &= \begin{cases} 0.9949xv(k) + 0.9975vent(k) & \text{if } a1(k) \text{ is true} \\ 0 & \text{Otherwise} \end{cases}
 \end{aligned}$$

Figura 1. Modelo MLD de la dinámica de la temperatura del invernadero durante invierno

Donde $(y(k))$ describe la salida del sistema, mientras $x_t, x_c, x_a, x_s, x_h, x_r$ son variables de estado que representan la temperatura externa, el calefactor de tubería, el aerotermo, la temperatura del suelo, la humedad relativa, la radiación solar y la posición de las ventilaciones, con $-100000 \leq x_t, x_c, x_a, x_s, x_h, x_r \leq 100000$. Las entradas no controlables son las variables continuas: Temperatura externa (T_{out}), temperatura del suelo (T_s), humedad relativa (hr), radiación solar global (Rad), velocidad del viento ($wind$) y el porcentaje de apertura de la ventilación ($vent$). Dichas variables se encuentran sujetas a restricciones tales que: $-10^\circ\text{C} \leq T_{out}, T_s \leq 50^\circ\text{C}$, $0\text{W/m}^2 \leq Rad \leq 3000\text{W/m}^2$, $0\% \leq hr \leq 100\%$ and $0\% \leq vent \leq 100\%$. Las entradas manipulables del sistema son el porcentaje de apertura de la válvula de mezcla del calefactor de tuberías (APH) y el estado discreto del aerotermo (FAH), con $0\% \leq APH \leq 100\%$ y $(FAH(k), FAH(k-1)) \in [0,1]$. Las dinámicas del sistema son activadas con base en las condiciones lógicas, a_1, a_2 y a_3 , dadas en desigualdades. Donde a_1 es activada si la temperatura externa se encuentra por debajo de 10°C . La variable a_2 es conmutada si la APH es mayor a 0.001% y a_3 es activada cuando la radiación se encuentra por debajo de 5W/m^2 . Finalmente un estado auxiliar ($c(k)$) $\in [0,1]$ es empleado para monitorear los cambios del FAH .

2.3 Avances y desarrollo de tesis

Para los dos sistemas de control propuestos se ha llegado a la etapa de simulación. Ambos casos se han desarrollado modelos autorregresivos de alto orden por medio de identificación de sistemas. Empleando una técnica gráfica se han obtenido modelos de bajo orden que representen la dinámica del invernadero durante el verano y el invierno. Dichos modelos han sido simulados con datos reales y comparados con el comportamiento real del sistema con buenos resultados. En la Figura 2 se aprecia la comparación de modelos para el caso de la temperatura interna del invernadero durante el verano

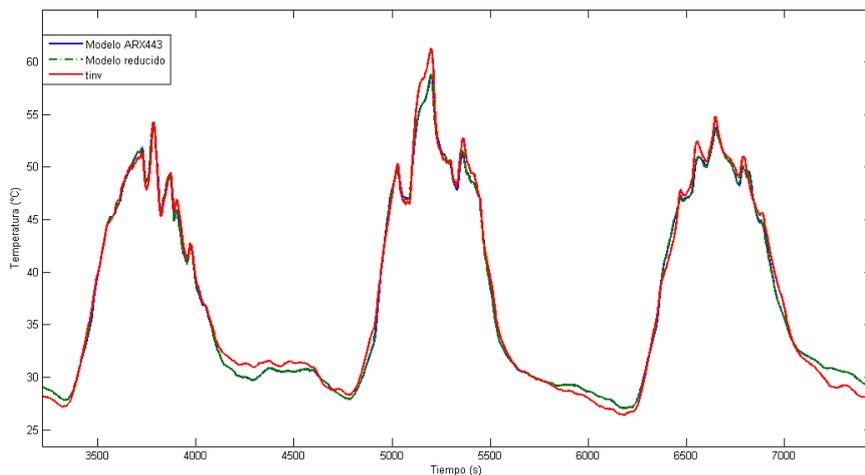


Figura 2. Comparación de temperatura real del invernadero (rojo), temperatura con el modelo de alto orden (azul) y temperatura con el modelo reducido (verde)

Para representación *MLD* de la dinámica del invernadero en invierno se ha empleado el lenguaje de programación Hysdel para obtener el modelo híbrido del sistema invernadero, como se puede apreciar en la Figura 3.

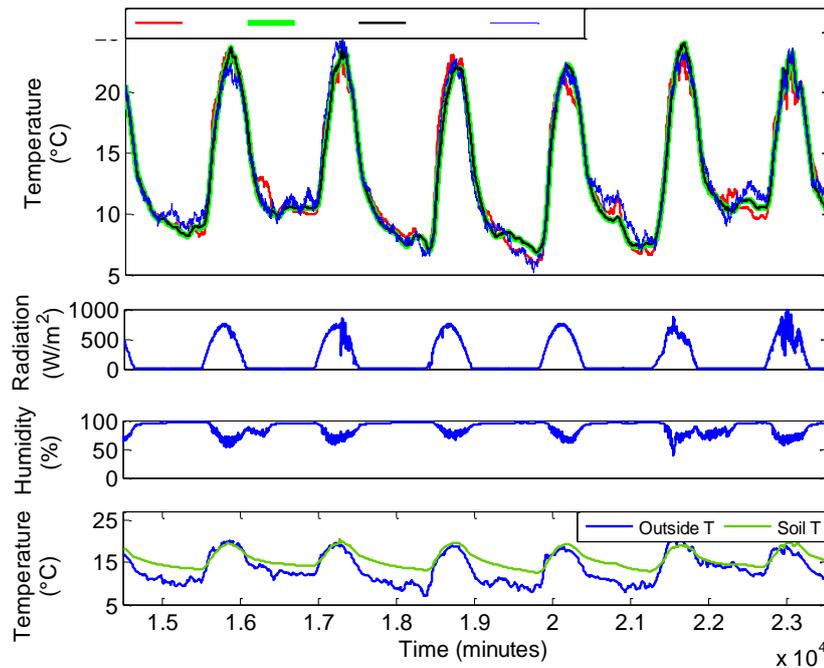


Figura 3. Ajuste de los modelos Híbrido, de alto orden y reducido en el modelamiento de la temperatura del invernadero durante el invierno

Se han desarrollado y comparado el desempeño de controladores tipo *PI*, por adelantado y *FF* empleando el método *G-H* en el control de la temperatura interna del invernadero durante el verano, como se muestra en la Figura 4.

Aportaciones al modelado y control climático de invernaderos

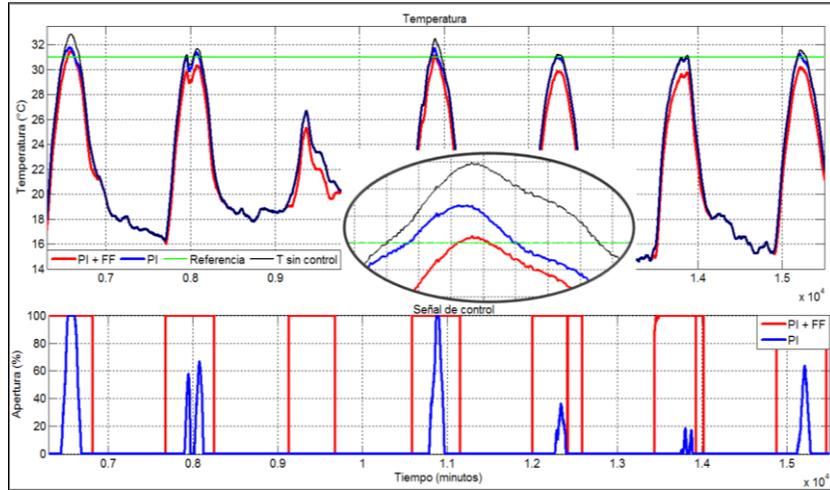


Figura 4. Comparación de Control PI Antiwindup y PI + FF método G-H

Además se desarrolló un controlador híbrido tipo HMPC para el control de la temperatura del aire durante el invierno empleando una combinación de calefactores. El desempeño y costo de implementación de dicho sistema se ha comparado con técnicas de control tradicionales. Los resultados de la simulación se pueden apreciar en la Figura 5.

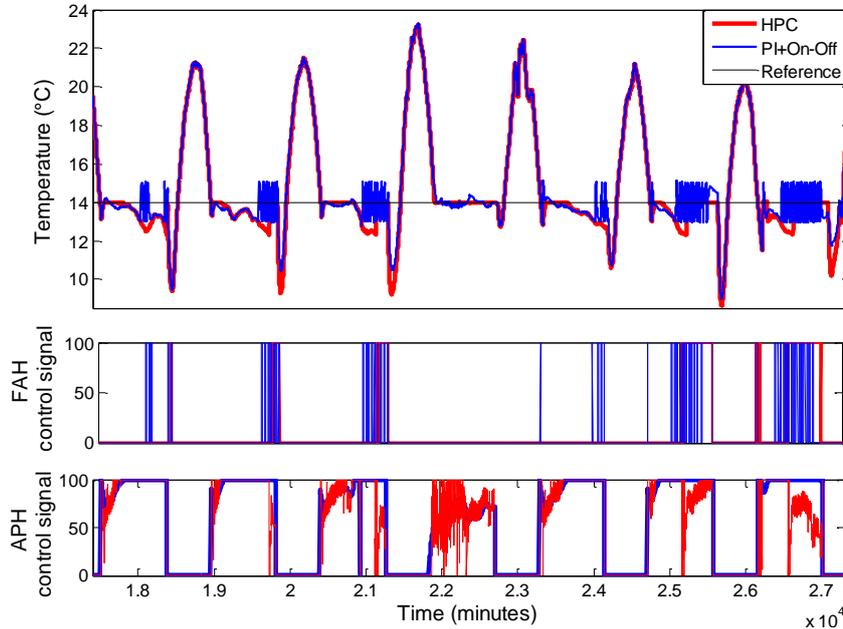


Figura 5. Comparación del desempeño del controlador HMPC, y control PI+On-Off en el control de la temperatura nocturna durante el invierno empleando dos sistemas de calefacción complementarios.

3 Conclusiones

Se han generado aportaciones para el control de la temperatura en invernadero en época estival y durante el invierno. Se ha desarrollado un controlador por adelanto que compense las perturbaciones antes que estas afecten el desempeño del sistema. Mientras que para el control de temperatura durante el invierno se han desarrollado un modelo y sistema de control híbrido que permite emplear dos técnicas de calefacción complementarias teniendo en cuenta los costos de la aplicación y desgaste de los equipos. Ambas técnicas de control han sido probadas en simulación con Buenos resultados comparado con técnicas de control tradicionales.

Referencias

- [1]. Åström, K.J., Hägglund, T. 2005. *Advanced PID Control*. ISA, USA. 461pp
- [2]. Bakker, J., Bot, A., Challa, H., Van de Braak, J., (Editors); 1995, *Greenhouse Climate Control*. Wageningen Pers. Wageningen.
- [3]. Bartzanas, T., Tchamitchian, M., Kittas, C., 2005. Influence of the heating method on greenhouse microclimate and energy consumption. *Biosystems Engineering* 91 (4), 487–499.
- [4]. Bemporad, A. (Ed.), 2010. *Hybrid Toolbox - user's guide*. Italy. Available in <http://cse.lab.imtlucca.it/~bemporad/hybrid/toolbox/>, [01 January of 2015].
- [5]. Borrelli, F., Bemporad, A., Morari, M. (Eds.), 2011. *Predictive control for linear and hybrid systems*. Italy
- [6]. Guzmán, J. L., Hägglund, T. 2011. Simple tuning rules for feedforward compensators. *J. of Process Control*. 21:92-102
- [7]. Guzmán, J. L., Hägglund, T., Veronesi, M. y Visioli, A. 2015. Performance indices for feedforward control. *J. of Process Control*. 26:26-34
- [8]. Guzmán, J.L., Rodríguez, F., Berenguel, M., Arahall, M.R., 2007. A Hybrid control approach to the greenhouse climate control problem. *IFAC/AGRI, 2007. IFAC Agricontrol*. Osijek, Croatia.
- [9]. Kamp, P.G.H., Timmerman, G.J., 2003. *Computerised environmental control in greenhouses. A step by step approach* 2nd edition; PTC (Ed.), The Netherlands. 267pp
- [10]. Ljung, L., 1999. *System Identification, Theory for the user* (2nd ed). Prentice Hall (Ed). New Jersey. 609pp.
- [11]. Mathworks, 2014. *Matlab documentation*. Available in <http://www.mathworks.com/help/matlab/>, [01 January of 2015].
- [12]. Montoya, A., Rodríguez, F., Guzmán, J.L., 2009. Obtención de modelos reducidos para el diseño de controladores de la temperatura interior de invernaderos. V Congreso Nacional y II Congreso Ibérico de Agroingeniería, Lugo, España.
- [13]. Rivera, D.E. 2007. Una metodología para la identificación integrada con el diseño de controladores IMC-PID. *Revista Iberoamericana de automática e informática industrial* 4: 5-18

Aportaciones al modelado y control climático de invernaderos

- [14]. Rodríguez, F., Berenguel, M. y Arahál, M. R. 2001. Feedforward controllers for greenhouse climate control based on physical models. European control conference (ECC), Porto, Portugal, 4-7sep. p 2158-2163.
- [15]. Torrisi, F., Bemporad, A., 2004. Hysdel -a tool for generating computational hybrid models for analysis and synthesis problems. IEEE Transactions on Control Systems Technology 12, 235-249.

Procesado eficiente de imágenes en el campo de la agricultura de precisión

Corrección de imágenes obtenidas con cámaras multiespectrales

Luis Ortega López¹

Universidad de Almería, lol.teleco@gmail.com

Resumen La agricultura de precisión pretende obtener información acerca de los cultivos utilizando herramientas tecnológicas (por ejemplo: GPS, sensores planta-clima-suelo, imágenes multiespectrales, etc.) para monitorizar dichos cultivos y ayudar en la toma de decisiones. En este trabajo hemos considerado una cámara multiespectral ampliamente utilizada en la agricultura de precisión (cámara multiespectral Sequoia) y lo que se pretende es conseguir el registro de las imágenes procedentes de los diferentes sensores de dicha cámara Sequoia con la mayor precisión posible. Para lograr la corrección de las distorsiones geométricas de cada canal se han marcado una serie de subtareas: (1) calibración de cada sensor de la cámara; (2) corrección de la perspectiva; y (3) manipulación de la imagen RGB para poder registrarla junto con el resto de sensores. Este proceso se ha utilizado sobre distintas imágenes (tomadas a distintas distancias con la cámara Sequoia) y como resultado se han obtenido una serie de imágenes corregidas.

1. Introducción

En los últimos años, el procesado de imágenes hiperespectrales se ha perfilado como una técnica analítica prometedora en los procesos de control y automatización de la industria. Una imagen hiperespectral genera mapas espaciales de la variación espectral de la muestra conocidos como datacubos o hipercubos. Estos hipercubos son datos tridimensionales que contienen las dos dimensiones espaciales y la dimensión espectral, esto es, un espectro para cada píxel [1]. En este contexto, también son de interés las imágenes multiespectrales. Dichas imágenes están formadas por relativamente pocas bandas (normalmente entre 3 y 20) y son bandas no necesariamente contiguas unas a otras, mientras que las imágenes hiperespectrales normalmente están formadas por un mayor número de bandas y éstas siempre son contiguas. Con una imagen multiespectral podemos obtener los valores de intensidad en las longitudes de onda discretas en las que el sistema capte radiación mientras que con una imagen hiperespectral lo que obtenemos es el espectro continuo o firma espectral del objeto de análisis.

Uno de los principales retos de la visión hiperespectral y multiespectral es el manejo y análisis de esos datos para la extracción de la información relevante

Procesado eficiente de imágenes en el campo de la agricultura de precisión

contenida en ellas. Dichas imágenes normalmente necesitan de diversas técnicas preprocesado de espectros (normalización, suavizado, centrado, diferenciación, etc.) y análisis multi-variante (técnicas de correlación, análisis de componentes principales, análisis discriminantes, etc).

Las cámaras hiperespectrales y multiespectrales nos permiten detectar cambios en la radiación, como, por ejemplo, en la radiación infrarroja para el caso de la agricultura. En este campo destaca la llamada agricultura de precisión, que consiste en el manejo diferenciado de los cultivos a partir del conocimiento de la variabilidad existente en una explotación agrícola [2]. Para caracterizar esta variabilidad, se utilizan herramientas tecnológicas como los Sistemas de Posicionamiento Global, conocidos popularmente como GPS, sensores planta-clima-suelo e imágenes hiperespectrales. De este modo, la agricultura de precisión trabaja para poder optimizar la gestión de los campos y poder planificar las fechas de siembra, usos de diferentes fertilizantes, frecuencia de riego incluso el conocimiento de su momento óptimo de recogida del fruto.

Los datos captados por todos estos sensores se almacenan digitalmente en forma de tablas y mapas, a partir de los cuales se genera la información que ayuda al agricultor en la toma de decisiones en campo (fertirrigación, podas o aclareos). De hecho, la agricultura de precisión puede ser capaz de poner de manifiesto un gran tipo de incidencias presentes en la cosecha: problemas de riego, variaciones en el tipo de tierra, malas hierbas, que no se ven a nivel del suelo. Además, la captura de datos del espectro infrarrojo se puede utilizar junto con el del visual para crear una imagen de la cosecha que destaque las diferencias entre las plantas sanas y enfermas, que el ojo humano no es capaz de observar. Por otro lado, dicha técnica puede ser capaz de extraer información semanal, diaria e incluso horaria, con el fin de crear una serie temporal, permitiendo observar los cambios en la cosecha, revelando áreas problemáticas, así como las oportunidades que existen para gestionar mejor la cosecha.

Debido a que el objetivo fundamental de la agricultura de precisión es la obtención de mayores rendimientos económicos, medioambientales y sociales, dicha agricultura se ha extendido en los últimos años como un método que consigue una mayor eficacia en las prácticas agrícolas. Sin embargo, todavía es una técnica en desarrollo donde se pueden conseguir avances de interés.

A largo plazo, la idea es utilizar sensores hiperespectrales y multiespectrales capaces de detectar cambios en la radiación, como, por ejemplo, en la radiación infrarroja o en el índice de área foliar de las plantas, con el objetivo de ayudar a los agricultores a mejorar la calidad y la cantidad de su producción agrícola. Por ejemplo, indicándoles si el regadío se está llevando a cabo de forma efectiva a pesar de los desniveles de la tierra, detectando fallos en el sistema de regadío, detectando la aparición de malas hierbas y/o plagas, etc.

En este trabajo nos hemos centrado en un cámara multiespectral, llamada Sequoia, que nos ofrece las imágenes captadas por cada una de las bandas de forma independiente. Es decir, en lugar de ofrecer una imagen con cuatro bandas, nos ofrece una imagen para cada una de las bandas. Para trabajar con dichas imágenes, es necesario corregir la distorsión que existe en cada una de ellas y,

Procesado eficiente de imágenes en el campo de la agricultura de precisión



Figura 1. Características de la cámara Sequoia.

posteriormente, superponer dichas imágenes en un ortomosaico para permitir el reconocimiento digital y el análisis de datos. Un Ortomosaico es una composición de imágenes a la que se le han corregido todos los errores geométricos para que cada punto del terreno sea observado desde una perspectiva perpendicular.

Actualmente, el software PiX4D viene instalado en la cámara Sequoia y se encarga, de forma automática de la generación de mapas indexados y de ortomosaicos. Sin embargo, la superposición de las imágenes de cada sensor la lleva a cabo sin tener en cuenta la posición que existe entre cada uno de los sensores, además de no corregir aberraciones ni la distorsión de la imagen. Por tanto, nos hemos centrado en corregir la distorsión de las imágenes de los sensores de la cámara para que su superposición sea lo más precisa posible.

2. Avances

Actualmente estamos trabajando en el calibrado de la cámara Sequoia, que es una cámara multiespectral que captura imágenes tanto del espectro visible como del no visible, proporcionando datos calibrados para analizar óptimamente la salud y vigor de las cosechas. Sequoia capta diferentes longitudes de onda: Verde, Roja, Red-Edge (Borde rojo) y NIR (Infrarrojo cercano) para destacar la salud de las plantas. En la Figura 1 se pueden observar los sensores de la cámara. Además, Parrot Sequoia viene con 64GB de memoria interna, una ranura para tarjetas SD, conectividad WiFi y sensores de movimientos/localización (GPS).

El sensor multiespectral Parrot Sequoia está equipado con una serie de sistemas integrados (GPS, IMU, sensor de irradianza solar), que la convierten en una

Procesado eficiente de imágenes en el campo de la agricultura de precisión

herramienta autónoma. La información que proporcionan estos sistemas aparece perfectamente integrada en el EXIF de cada imagen, que contiene también los parámetros de calibración geométrica y radiométrica de cada unidad. Estos datos son leídos e interpretados por programas informáticos como Pix4D que utilizan dicha información para generar ortomosaicos.

Para que las tareas de monitorización, reconocimiento de estructuras (por ejemplo zonas cultivadas), etc. tengan éxito, es necesario que las imágenes que vayamos a procesar sean lo más fiables posible. Sin embargo, las lentes de las cámaras pueden presentar diferentes tipos de defectos en algunas distancias focales, distancias de enfoque, etc.

Por todo ello, un preprocesado de las imágenes captadas por los distintos sensores es algo habitual. Dicho preprocesado incluye técnicas como son: normalización, suavizado, centrado, diferenciación, etc., así como análisis multi-variante (técnicas de correlación, análisis de componentes principales, análisis discriminantes, etc.) para eliminar ruidos.

En el caso de la cámara Sequoia, tras analizar las imágenes obtenidas, se percibe una distorsión evidente en los sensores multispectrales. Para poder entender la naturaleza de esta distorsión, a continuación se describirán los elementos que contribuyen a la distorsión de una cámara, su caracterización y algunas estrategias para la eliminación de la misma. Existen tres elementos que contribuyen a la distorsión de una cámara:

- a** Distorsión de la lente. Debido a que los centros de las lentes no están perfectamente alineados y además existen aberraciones en cada una de las lentes. Este efecto se puede dividir en distorsión radial y tangencial, donde la primera es la que tiende a curvar las líneas rectas.
- b** Perspectiva. La conicidad de la perspectiva provoca que las líneas rectas converjan en un punto de fuga.
- c** Relieve del terreno. Las diferencias de distancia en el plano a fotografiar respecto a la lente provocan que aquellos puntos más cercanos se representen más próximos al centro.

Para eliminar los efectos de la perspectiva y también la distorsión de la cámara, se puede utilizar:

- Una rectificación basada en modelos de sensor (modelos físicos o generalizados)
- Una transformación geométrica de homografía [3,4]. Se denomina homografía a toda transformación proyectiva que determina una correspondencia entre dos figuras geométricas planas, de forma que a cada uno de los puntos y las rectas de una de ellas le corresponden, respectivamente, un punto y una recta de la otra.

Como lo que queremos obtener al final es un proceso automatizado de rectificación, la homografía ha sido la técnica escogida en este trabajo. Para poder rectificar cualquier imagen mediante homografía se necesitan tanto los parámetros de calibración de la cámara como la posición de la cámara en el momento de la fotografía, o una referencia en el objeto a fotografiar.

Procesado eficiente de imágenes en el campo de la agricultura de precisión

Los pasos realizados han sido los siguientes:

1. Obtener los parámetros invariables de las cinco lentes de la cámara Sequoia mediante MATLAB. Dichos parámetros son invariantes si no hacemos zoom ni modificamos algunas de las lentes de alguna forma.
2. Corregir la distorsión de las imágenes mediante OpenCV, importando los valores de distorsión y parámetros de cada una de las lentes que se han calculado previamente.
3. Realizar una homografía de la región de interés, que tendremos que definir en el terreno (por ejemplo, colocando cuatro pelotas que delimiten las esquinas de la zona a estudiar y que sean fácilmente diferenciables en todas las bandas espectrales).
4. Las cinco homografías se han superpuesto como una única imagen hiperespectral rectificadas de la región de interés.

Para el proceso de calibración de los parámetros característicos de las cinco lentes se ha utilizado un conjunto de 115 imágenes correspondientes a diferentes espectros (23 por cada uno de los sensores: NIR, RED, REG, GRE y RGB). Una vez realizado el proceso de calibración se han comprobado los resultados obtenidos al corregir dos conjuntos de imágenes (patrones de test y fotos en condiciones reales), observándose una corrección eficiente de las distorsiones de cada una de las imágenes, así como una superposición de las imágenes aceptable para su posterior procesado. Las Figuras 2 y 3 muestran los resultados obtenidos tras aplicar los pasos anteriormente comentados.

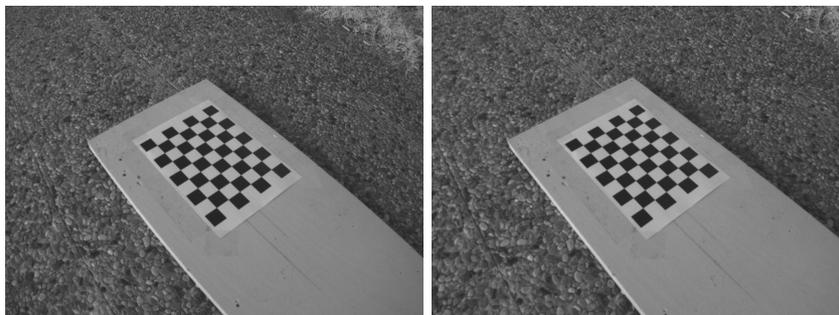


Figura 2. A la izquierda, imagen original obtenida con el sensor NIR. A la derecha, imagen final obtenida tras aplicar las correcciones.

3. Conclusiones

En este trabajo hemos realizado un proceso automático capaz de corregir las distorsiones geométricas de cada canal de una cámara multiespectral, Sequoia. Dicho objetivo se ha conseguido realizando una calibración de cada sensor de

Procesado eficiente de imágenes en el campo de la agricultura de precisión

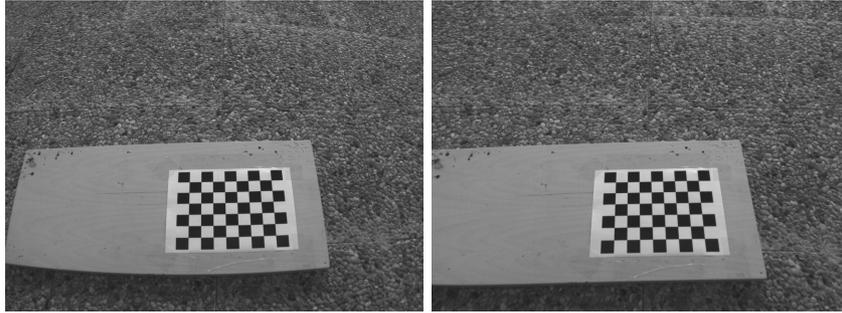


Figura 3. A la izquierda, imagen original obtenida con el sensor RED. A la derecha, imagen final obtenida tras aplicar las correcciones.

la cámara, una corrección de la perspectiva, una manipulación de la imagen RGB, y una superposición de todas estas imágenes corregidas. Los resultados obtenidos se han probado sobre distintas Este proceso se ha utilizado sobre distintas imágenes (tomadas a distintas distancias con la cámara Sequoia) y, como resultado, se han obtenido una serie de imágenes corregidas y con una superposición de las imágenes con una calidad aceptable.

Referencias

1. B. Diezma et al., “La imagen hiperespectral como herramienta de evaluación de la calidad de hortaliza de hoja mínimamente procesada,” in *VI Congreso Ibérico de AgroIngeniería - Comunicaciones finales*, Évora, 2011.
2. S. Fountas et al., “Management strategies and practices for precision agriculture operations,” in *Precision Agriculture 2009 - Papers Presented at the 7th European Conference on Precision Agriculture, ECPA 2009*, 2009, pp. 893–898.
3. J. Heikkila and O. Silven, “A four-step camera calibration procedure with implicit image correction,” in *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, Washington, DC, USA, 1997, CVPR '97, pp. 1106–, IEEE Computer Society.
4. Z. Zhang, “A flexible new technique for camera calibration,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

Arquitecturas adaptativas en tiempo de ejecución mediante transformaciones inteligentes basadas en técnicas de soft computing y desarrollo de software dirigido por modelos

Diego Rodríguez Gracia

Grupo Informática Aplicada (ACG), Departamento de Informática
Universidad de Almería, Ctra Sacramento s/n, 04120 Almería, España
diegorg@ual.es

Resumen. Actualmente el uso de servicios para el desarrollo de sistemas abiertos se ha consolidado y fortalecido. Los avances en la comunidad de la Ingeniería de Sistemas de Servicio promovidos por el refuerzo de las tecnologías de Web Services y Web Semántica y la presencia de nuevas técnicas de Cloud Computing así como la proliferación de soluciones de microservicios, han permitido a los arquitectos de software experimentar y desarrollar nuevas formas de construir nuevos sistemas informáticos abiertos y adaptables en tiempo de ejecución. En esta tesis se desarrollará un esquema para la adaptación de sistemas informáticos dinámicos (es decir, aquellos que modifican su estructura en el tiempo) mediante el uso de técnicas interdisciplinarias de ingeniería dirigida por modelos (MDE) e ingeniería de servicios (SSE) y soft computing. En esta tesis se desarrollará y presentará un esquema basado en microservicios que consigue la adaptación en tiempo de ejecución de sistemas arquitectónicos de software basados en componentes. Este esquema se desarrolla como un proceso de transformación adaptativo de tres capas que se admite en un servicio de toma de decisiones basado en reglas implementado mediante algoritmos de aprendizaje automático (Machine Learning, ML).

Palabras clave: Ciencia de servicios e Ingeniería de Sistemas de Servicio (SSE), sistema domótico adaptativo, smart building. Ingeniería Dirigida por Modelos (MDE), Soft Computing, Máquina de Aprendizaje (ML), sistema de toma de decisiones.

1 Introducción

Algunos sistemas software actuales requieren que su funcionamiento y estructura se adapten a requisitos que no habían sido detectados en tiempo de diseño. La Ingeniería Dirigida por Modelos (MDE) ofrece mecanismos para construir software desde la fase de análisis hasta la de implementación. La ingeniería dirigida por modelos (MDE) se basa en construir modelos en distintos niveles de abstracción que faciliten la especificación del software además de ofrecer mecanismos para la automatización del desarrollo del producto final a través de la utilización de transformaciones de modelos. Desde esta perspectiva, los modelos son tratados como artefactos estáticos que

son construidos en tiempo de diseño donde modelos de mayor nivel de abstracción se transforman en modelos de menor nivel de abstracción hasta generar modelos que puedan ser interpretados por el sistema o llegar a generar código ejecutable. Además, es posible hacer uso de estas técnicas para conseguir la adaptación transformando los modelos del sistema en tiempo de ejecución. Las arquitecturas software basadas en componentes son un objetivo ideal para el uso de estas técnicas ya que la manipulación de los modelos arquitectónicos que representan el sistema pueden provocar la adaptación de su funcionalidad al variar su estructura insertando y eliminando componentes o modificando las relaciones entre ellos.

En esta tesis doctoral se pretende desarrollar un esquema de arquitectura software basada en componentes para la adaptación de sistemas dinámicos en tiempo de ejecución mediante el uso de técnicas interdisciplinarias de la ingeniería dirigida por modelos (MDE) y soft computing y validar dicha arquitectura sobre dos casos de estudio concreto: En primer lugar en interfaces de usuario simples para WIS (sistemas de información basados en Web), de tipo WIMP (Windows, Icons, Menus and Pointers), y basadas en la composición "bottom-up" en tiempo de ejecución de componentes de interfaz COTS del tipo "widgets". En segundo lugar sobre smart buildings.

Actualmente, los sistemas basados en componentes son determinados en tiempo de diseño de acuerdo a una arquitectura de componentes y a los requisitos iniciales. Algunos de los sistemas desarrollados con estas técnicas intentan proporcionar software con la capacidad de adaptación por lo que su comportamiento puede variar dependiendo de las circunstancias que rodean su ejecución, por ejemplo, los cambios en la interacción con el usuario, la variación en los recursos disponibles, cambios de la plataforma donde se ejecuta el software [1], etc. Muchos de estas capacidades de adaptación pueden ser identificadas en la fase de diseño e implementadas en el proceso de desarrollo. Sin embargo, pueden surgir imprevistos a las condiciones originales de manera que fuera necesario que el software del sistema fuera capaz de adaptarse a esas nuevas circunstancias. En estos casos sería interesante dotar al sistema de mecanismos mediante los cuales pudiera adaptar su comportamiento de forma automática, los cuales son conocidos como sistemas autoadaptativos (SAS) [1]. Como consecuencia, la visión estática de los modelos está siendo actualizada por propuestas que tratan de adaptar el software de forma automática mediante la manipulación de los modelos que lo definen [2]. Esta nueva visión hace uso de conceptos MDE y añade funcionalidades para la adaptación de los modelos en tiempo de ejecución.

En el dominio particular de los sistemas de software basado en componentes, el uso de técnicas de MDE puede facilitar el diseño y desarrollo de arquitecturas, por ejemplo, para la definición de su estructura, el comportamiento de sus componentes y sus relaciones, su interacción o propiedades funcionales y no funcionales [3]. Además, la manipulación de los modelos de arquitectura en tiempo de ejecución hace que sea posible la generación de diferentes sistemas de software basados en la misma definición abstracta, la adaptación, por ejemplo, para la interacción del usuario, estado del componente o plataforma de ejecución [4].

Numerosos trabajos de investigación que tratan sobre la adaptación dinámica de sistemas software (DAS) al contexto de ejecución usan aproximaciones basadas en arquitecturas [5], [6]. El framework Rainbow [5] proporciona mecanismos para adaptar y actualizar los modelos de arquitectura a las necesidades del sistema utilizando modelos de arquitectura abstractos para monitorizar, evaluar y adaptar las configura-

ciones que luego se trasladan al sistema en ejecución, pero estas operaciones no las realiza a través de técnicas MDE de transformaciones de modelos. En el contexto de las aplicaciones para móviles se pueden utilizar los modelos de arquitectura para describir la variabilidad, es decir, que los propios modelos contienen la información y el criterio de selección para que el mediado (middleware) pueda derivar la adaptación al contexto en tiempo de ejecución [6]. También existen propuestas donde se definen modelos de variabilidad para describir la lógica de adaptación y separarla así del funcionamiento del sistema [7]. En [28] se desarrolla un experimento de composición de componentes comerciales de un Servicio de Conversión de Imágenes Geográficas (GTS) usado en GIS y UML-RT. Hay trabajos donde los autores proponen implementar el loop de control del sistema adaptativo como un sistema basado en componentes que pueda ser adaptado también [8]. De esta forma el loop de control puede ser reconfigurado en tiempo de ejecución para incorporar nuevo conocimiento de forma dinámica. El objetivo final de esta tesis doctoral es desarrollar un sistema similar, en la que la lógica de adaptación cambie según el conocimiento que se adquiera de la ejecución.

Otro tipo de sistemas adaptativos son las líneas de producto software dinámicas (dynamic software product line, DSPL). Estos sistemas son parecidos a las tradicionales líneas de productos software pero la variabilidad está vinculada al tiempo de ejecución [10], [11]. En [12] los autores aplican su uso en el dominio de la domótica (smart homes). En su propuesta, utilizan modelos de variabilidad para activar o desactivar características en tiempo de ejecución cumpliendo así con las condiciones del contexto. Englobados en las líneas de producto software dinámicas (DSPL) existen propuestas en las cuales se describe una arquitectura para dar soporte a DSPLs controlando el número de variantes que el sistema puede tener [13], [14]. Combinan técnicas dirigidas por modelos (model-driven) y orientadas a aspectos (aspect-oriented) para adaptar los modelos a través de model-weaving. En esta tesis doctoral, frente a esta propuesta, se define la variabilidad del sistema en los procesos de transformación de modelos y en los repositorios que definen las reglas de adaptación.

Existen propuestas sobre sistemas adaptativos que hacen uso de lenguajes de programación de alto nivel para obtener la evolución. En [15], [16], se proponen implementaciones basadas en Java que se ejecuta dentro de una plataforma OSGi [17] para la adaptación de software en tiempo de ejecución. En este caso los programas escritos en lenguajes de programación son artefactos estáticos y no pueden evolucionar en tiempo de ejecución.

Respecto a la composición dinámica de las transformaciones de modelos existen trabajos que proponen una forma incremental para la actualización de los procesos de transformación y su construcción de forma dinámica a partir de un conjunto de reglas [18]. En [19] se propone una aproximación para la composición de transformaciones en ATL mientras que en [20] se describe una propuesta similar haciendo uso del lenguaje de transformación QVT. En [21], [22], los autores presentan un mecanismo para construir transformaciones de modelos a partir de módulos creados previamente y que pueden ser referenciados o importados de un archivo ATL. La fase de transformación de modelos de esta tesis doctoral se basa en estos trabajos para definir un repositorio de reglas ATL que son seleccionadas de forma dinámica para construir una transformación en tiempo de ejecución. Otros trabajos como [23] y [24] proponen una actualización de las transformaciones con el objetivo de realizar una refactorización de los

modelos para adaptarlos en tiempo de ejecución pero no tratan de refactorizar la propia transformación. Uno de los objetivos de refactorizar las transformaciones M2M es la mejora o adaptación de su comportamiento al contexto del sistema por medio de la reestructuración o introducción de nuevas reglas o helpers [25]. Siguiendo una aproximación MDE, estas refactorizaciones se pueden implementar como transformaciones en sí mismas en las que las propias transformaciones participen como entrada y/o salida de las propias transformaciones, en las llamadas Higher Order Transformations (HOT) [26], [27]. En esta tesis doctoral se hace uso de las HOT para generar de forma dinámica las transformaciones que adaptan los modelos de arquitectura en tiempo de ejecución. En [52], partiendo de las definiciones abstractas de la arquitectura que conforma la interfaz de usuario, se propone una arquitectura de transformación en tiempo de ejecución de interfaces de usuario mediante un proceso de transformación modelo-a-modelo (M2M) mediante la definición de reglas de transformación. Dichas reglas de transformación son almacenadas en un repositorio de reglas el cual es construido mediante un metamodelo. Posteriormente el proceso de adaptación se mejora mediante la inserción de una lógica de contexto y sistema de puntuación de reglas del repositorio de reglas para permitir variabilidad en la selección de reglas en la adaptación [53].

Por otro lado, la inteligencia computacional (i.e., “soft computing”) [54] es un conjunto de tecnologías formadas por: redes neuronales artificiales, sistemas difusos, computación evolutiva, métodos bayesianos y probabilísticos, teoría del caos y sistemas “swarm” o inteligencia distribuida. Sobre este conjunto destacamos que la lógica difusa permite tratar la imprecisión, el razonamiento aproximado y definir un lenguaje más cercano al mundo real, las redes neuronales se centran en el aprendizaje, adaptación y clasificación, los métodos probabilísticos se basan en un razonamiento estadístico sobre evidencias. Esto propició el uso de modelos híbridos donde se combinasen las ventajas de la lógica difusa y las redes neuronales por su capacidad de aprendizaje conformando los sistemas neurodifusos [55]. En el grupo de investigación de Informática Aplicada se ha trabajado en el ámbito del “softcomputing” en la recuperación de imágenes basadas en contenido difuso mediante sistemas neurodifusos y redes bayesianas [56] y en la definición de una metodología de selección de características en procesos de aprendizaje inteligente mediante redes bayesianas y sistemas neurodifusos [57]. En el ámbito de la ingeniería del software y la inteligencia computacional podemos encontrar principalmente aplicaciones enfocadas a la evaluación del coste de reutilización de componentes como en [58] y [59] donde se propone modelos neurodifusos para la clasificación de componentes en su reutilización. Sobre la eficiencia en el consumo energético y el nivel de confort de los usuarios podemos encontrar en el campus de la Universidad Politécnica de Bucarest la implementación del control de un edificio inteligente mediante la ejecución de un mecanismo dirigido por un conjunto de políticas según las cuales se activa el sistema de calefacción [37]. Otros estudios se han centrado en la implementación de arquitecturas Human-Robot Cloud (HRC) [38] en el escenario específico de edificios inteligentes para obtener beneficios en eficiencia y ahorro energético. En este trabajo se presenta una propuesta mediante la cual con sensores, procesadores y actuadores, en los que se incluyen identificadores faciales y localizadores de personas, son transformadas a un sistema cognitivo distribuido y reconfigurable el cual puede soportar múltiples aplicaciones en el futuro.

Acerca de las metodologías utilizadas para el control de edificios inteligentes, algunas propuestas se basan ampliamente en la definición de una lógica de inferencia simple usando los valores de diferentes sensores de medición y de sensores de presencia [29], [30], [31]. Otras metodologías utilizan técnicas de inferencia bayesiana (red bayesiana) para predecir patrones de comportamiento de los usuarios [32]. En esta línea, es decir, aplicar redes bayesianas para predecir el comportamiento de los usuarios, en [33] se propone un método capaz de predecir el consumo de energía por habitante. Por otro lado, iDorm [34] se presenta como un sistema adaptativo (capaz de aprender de la interacción con el usuario y así predecir sus futuras necesidades) mediante agentes embebidos. Estos agentes utilizan aprendizaje sincrónico incremental (ISL) basado en lógica difusa. En [35] se propone el uso de cadenas de Markov para establecer la probabilidad de ocupación de áreas determinadas. En [36] se presenta un estudio comparativo de las diferentes metodologías y técnicas usadas para el control de edificios inteligentes.

Como se puede apreciar tras la revisión realizada, existe abundante literatura sobre sistemas adaptativos, pero ninguno de estos trabajos combina técnicas de transformaciones de modelos y soft computing para desarrollar arquitecturas autoadaptativas en tiempo de ejecución.

2 Motivación

Esta tesis doctoral trata de dar solución al problema de la adaptación en tiempo de ejecución de sistemas software. Pero la propuesta a desarrollar en la presente tesis doctoral no será válida para cualquier aplicación sino que se centrará únicamente en sistemas que están basados en componentes. En la metodología a desarrollar, los componentes que existen en una arquitectura en un instante determinado proporcionan la funcionalidad requerida y satisfacen los requisitos del sistema pero, debido a cambios en el contexto de la aplicación, puede ser necesario que estos componentes cambien para satisfacer las nuevas necesidades. Por otro lado, en [9], se trata el importante aspecto de la incertidumbre dentro de los sistemas adaptativos y que pueden tener efecto tanto en las fases de requerimientos como en las de diseño y ejecución. En la propuesta de arquitectura adaptativa a desarrollar en esta tesis el objetivo de construir de forma modular el proceso de adaptación y de separar las reglas de adaptación en un repositorio nos permitirá disponer de un mecanismo de actualización en tiempo de ejecución para mejorar aquellos aspectos que no hayan sido detectados en fases anteriores sin que esta actualización afecte a otros módulos del proceso ni produzca un gran impacto en el sistema.

La propuesta a alcanzar mediante el desarrollo de esta tesis doctoral podría resultar útil en diversos sistemas software donde se requiere la adaptación de arquitecturas basadas en componentes. Ejemplos de sistemas con este tipo de arquitecturas los podemos encontrar en aplicaciones de domótica [41], coches inteligentes [42], edificios inteligentes [43], robótica [44], [45], infraestructuras de redes de comunicación [46], etc. Una ventaja de la propuesta que se intenta alcanzar en esta tesis es que abstrae y aísla el proceso completo de evolución del sistema, ya que es independiente de la plataforma. De esta manera, esta propuesta se puede aplicar directamente en determinados sistemas dinámicos de carácter social, tales como la domótica, edificios inteli-

Arquitecturas adaptativas en tiempo de ejecución mediante transformaciones inteligentes...

gentes, robótica, interfaces gráficas de usuario, etc. En esta tesis doctoral se trabajará en dos escenarios de la vida real para la experimentación del proceso de evolución de sistemas dinámicos. Uno de ellos es en el dominio de interfaces de usuario basadas en componentes, aplicados a la gestión del medio ambiente. Estos interfaces de usuario tienen que estar definidos como modelos arquitectónicos en los que cada componente representa un componente de interfaz de usuario. A diferencia de otras propuestas de interfaz de usuario basada en widgets (como iGoogle, Netvibes o Pageflakes [49]), basada en portlets [50], o mash-up [51], los componentes de IU pueden tener dependencias entre si que afecten al comportamiento del sistema. El proceso de adaptación por tanto podrá cambiar la estructura de las interfaces de usuario a partir de la transformación de los modelos arquitectónicos que las representan. La adaptación viene determinada por cambios producidos en el contexto, como por ejemplo, interacción del usuario, cambio en los requisitos del sistema, eventos temporales, etc. Además, dotando a la arquitectura con un sistema de toma decisiones y mediante una máquina de aprendizaje, que recoge la información de la interacción del usuario con el sistema y genera las correspondientes reglas que definirán el conocimiento del mismo, las cuales serán las utilizadas por el sistema adaptativo para llevar a cabo las correspondientes transformaciones en tiempo de ejecución, la propuesta de arquitectura será capaz de aprender de la interacción con el usuario modificando o incluso generando nuevas reglas de transformación de modelos

Con el fin de comparar y validar la propuesta a desarrollar en esta tesis doctoral, se desarrollará otro escenario en el ámbito del bienestar social aplicado a la domótica para el confort del hogar. En el contexto anteriormente descrito está la propuesta del presente proyecto de tesis. Esta tesis doctoral se enmarca como línea de interés que ha surgido dentro del proyecto del Ministerio de Ciencia e Innovación, ref., TIN2013-41576-R., y como continuación del proyecto finalizado TIN2010-15588, dentro del proyecto de excelencia de la Junta de Andalucía, ref., TIC-6114. Se pretende desarrollar un esquema de arquitectura software basada en componentes para la adaptación de sistemas dinámicos en tiempo de ejecución mediante el uso de técnicas interdisciplinarias de la ingeniería dirigida por modelos (MDE) y soft computing y validar dicha arquitectura sobre dos casos de estudio concreto: en primer lugar en interfaces de usuario simples para WIS (sistemas de información basados en Web), de tipo WIMP (Windows, Icons, Menus and Pointers), y basadas en la composición "bottom-up" en tiempo de ejecución de componentes de interfaz COTS del tipo "widgets". En segundo lugar sobre smart buildings. En nuestra opinión, el conseguir desarrollar un esquema de arquitectura software basada en componentes para la adaptación de sistemas dinámicos en tiempo de ejecución, en la cual, mediante un sistema de toma de decisiones el sistema sea capaz de prever las futuras necesidades del usuario adaptándose al propio usuario y al contexto del sistema, sería un avance significativo en el dominio de los sistemas autoadaptativos. Por otro lado, la propuesta a desarrollar en esta tesis doctoral puede tener una aplicación directa en ciertos sistemas informáticos dinámicos de índole social, como en domótica, edificios inteligentes, robótica, interfaces gráficas de usuario, etc.

3 Tareas realizadas

A continuación se detallan las tareas generales que se han realizado a lo largo del desarrollo de esta tesis doctoral:

- Definir las arquitecturas software a través de una representación basada en componentes.
- Construir un proceso de adaptación para las arquitecturas que no sea estático, es decir, que sea flexible y que cuya lógica de adaptación pueda cambiar dinámicamente de acuerdo a reglas de inferencia determinadas a priori.
- Definir la lógica de adaptación en base a una serie de reglas para conseguir el dinamismo del proceso de adaptación variando las reglas que pueden aplicarse dependiendo del contexto del sistema.
- Establecer una representación para las reglas de adaptación de forma que estas reglas puedan almacenarse en un repositorio, gestionarse y seleccionarse para su aplicación.
- Definir un mecanismo para la selección en tiempo de ejecución de las reglas de adaptación.
- Desarrollar un sistema de toma de decisiones que gestione las reglas de transformación de acuerdo al contexto y a los requisitos del sistema.
- Definir un motor de adaptación que se encargue de aplicar las reglas seleccionadas cambiando así la estructura de componentes de la arquitectura.

4 Conclusiones

En el estado actual de esta tesis, el desarrollo de la arquitectura basada en componentes se ha basado en los cuatro niveles de abstracción inspirado en el marco de referencia Cameleon [39], el cual también se basa en la especificación Object Management Group (OMG), arquitectura dirigida por modelos (MDA) [40]. De esta forma, la adaptación de las arquitecturas se lleva a cabo a partir de procesos ejecutados en los niveles abstracto y concreto de las arquitecturas [47], [48]. En el nivel abstracto se ejecutan procesos de transformación modelo-a-modelo para cambiar y adaptar los modelos de arquitectura abstracta a los cambios producidos en el contexto. El objetivo de esta tesis se centra en el desarrollo de la adaptación realizada en el nivel abstracto de las arquitecturas. En dicho nivel las transformaciones modelo a modelo (M2M) son utilizadas para adaptar los modelos arquitectónicos de forma dinámica. Esta tesis doctoral pretende proporcionar un comportamiento dinámico a las transformaciones para que ellas también cambien adaptándose a los nuevos requisitos y a las variaciones del contexto. De esta forma las transformaciones van a ser construidas en tiempo de ejecución seleccionando, a partir de un repositorio de reglas, aquellas reglas de transformación idóneas para la situación actual. Inspirándonos en la idea de coevolución en la que los cambios en los metamodelos que definen el sistema provocan la adaptación correspondiente en los modelos que definen, la propuesta a desarrollar deberá realizar la refactorización en tiempo de ejecución de las transformaciones M2M que adaptan los modelos arquitectónicos. Este proceso será implementado haciendo uso de Higher Order Transformations (HOT) [26], [27]. Seguidamente se ha versionado la metodología anterior (de adaptación en tiempo de ejecución de arquitec-

Arquitecturas adaptativas en tiempo de ejecución mediante transformaciones inteligentes...

turas) para que sea evolutiva, es decir, que el conocimiento interno, implícito al proceso de adaptación anterior, pueda variar en el tiempo. Esto significa que el sistema debe dotarse de razonamiento inteligente para la inferencia de nuevas reglas de adaptación de la arquitectura, y que no están previamente establecidas en el sistema. La situación de partida es como la que muestra la Figura 1 (Fig.1). En el proceso de adaptación de la arquitectura del sistema intervienen la arquitectura de entrada, y un conjunto de subprocesos y de reglas de adaptación. La adaptación se realiza con una transformación de modelos M2M, considerando las arquitecturas de entrada como modelos.

Las reglas que utiliza el transformador están prefijadas y almacenadas en un repositorio de reglas. El proceso de adaptación selecciona el subconjunto de reglas que aplicará para la transformación. Este subconjunto se genera a través de un proceso que analiza la información de contexto (variables de entorno) y la información procedente de la interacción con el sistema. El objetivo alcanzado en esta fase consiste en haber desarrollado un mecanismo de inferencia de nuevas reglas de adaptación a partir de las variables del entorno y de la interacción del sistema con el usuario (véase de nuevo Fig. 2).

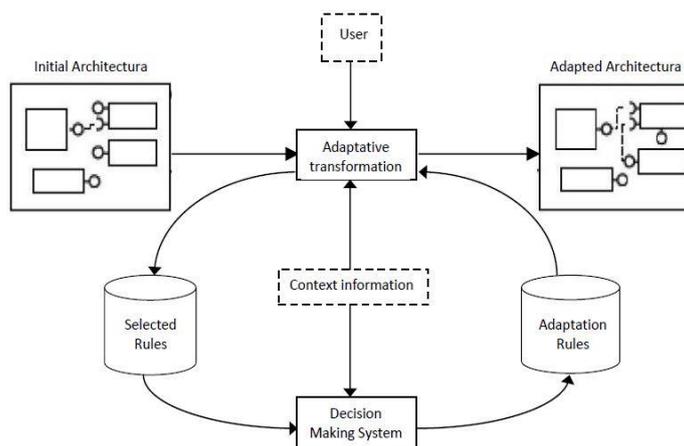


Fig. 1. Esquema de arquitectura autoadaptativa.

Aquí, el proceso de adaptación hace uso de un repositorio de reglas, que deberá ser extendido. En este punto incorpora un nuevo repositorio de trazas que almacene todas las variables del entorno, como las del contexto y de la interacción. La interacción del usuario con el sistema generará grandes volúmenes de datos que podrán evolucionar en el tiempo. Así, una interacción del usuario con la interfaz del sistema (GUI o domótica) puede depender de su perfil de usuario y de diferentes sesiones en el tiempo que, a su vez, acarrearán interacciones que deberán ser detectadas, recogidas y guardadas de la mejor forma para poder (junto con las variables de contexto) inferir nuevas reglas de adaptación mediante el sistema de toma de decisiones. Para este contexto y este problema se ha optado por el uso de bases de datos relacionales para la elaboración de consultas sobre esos datos y la ejecución de algoritmos de análisis para poder realizar aprendizaje automático.

Respecto a este aprendizaje (de nuevas reglas de adaptación) su objetivo es definir una metodología evolutiva donde se contemplen aspectos sobre qué tecnologías aplicar para inferir, consultar e identificar patrones, así como para validar las propias soluciones. Los sistemas basados en inteligencia artificial en general son sistemas estáticos porque una vez creado el conocimiento del sistema se mantiene y no se modifica. Por lo tanto, se necesita dar un paso más en el proceso e introducir el concepto de evolución. En esta tesis se usa una metodología inspirada en [60] que divide el proceso en cuatro etapas: recuperar casos similares, reutilizar el conocimiento existente, revisar la solución propuesta y retener el nuevo conocimiento para el futuro.

Por último, para la experimentación de la metodología de evolución de sistemas dinámicos se trabajará en dos escenarios reales. Uno en el dominio de las interfaces “gráficas” de usuario basadas en componentes. Otro escenario alternativo en el ámbito del bienestar social aplicado a domótica para el confort en el hogar, en concreto en edificios inteligentes.

Referencias

- [1] Cheng BHC, de Lemos R, Geise H, Inverardi P, Magee J. Software Engineering for self-adaptive systems: a research roadmap. In *Software Engineering for Self-Adaptive Systems*, Vol. 5525, Cheng BHC (ed.). LNCS, Springer-Verlang: Berlin, Heidelberg, 2009; 1–26. DOI: 10.1007/978-3-642-02161-9_1.
- [2] G. Blair, N. Bencomo, and R.B. France, “Models@run.time,” *Computer*, vol. 40, no. 10, pp. 22–27, 2009.
- [3] I. Crnkovic, S. Sentilles, A. Vulgarakis, and M.R.V. Chaudron, “A Classification Framework for Software Component Models,” *IEEE Transactions on Software Engineering*, vol. 37, no. 5, pp. 593–615, 2011.
- [4] N. Bencomo, and G. Blair, “Using Architecture Models to Support the Generation and Operation of Component-Based Adaptive Systems,” *Software Engineering for Self-Adaptive Systems*, pp. 183–200, 2009.
- [5] D. Garlan, S.W. Cheng, A.C. Huang, B. Schmerl, and P. Steenkiste, “Rainbow: Architecture-Based Self-Adaptation with Reusable In-frastructure,” *Computer*, vol. 37, no. 10, pp. 46–54, 2004.
- [6] J. Floch, S. Hallsteinsen, E. Stav, F. Eliassen, K. Lund, and E. Gjørven, “Using Architecture Models for Runtime Adaptability,” *IEEE Software*, vol. 23, no. 2, pp. 62–70, 2006.
- [7] F. Fleurey, and A. Solberg, “A Domain Specific Modeling Language Supporting Specification, Simulation and Execution of Dynamic Adaptive Systems,” *Proc. 12th International Conference on Model Driven Engineering Languages and Systems (MoDELS '09)*, pp. 606–621, 2009.
- [8] G. Perrouin, B. Morin, F. Chauvel, F. Fleurey, J. Klein, Y. Le Traon, O. Barais, and J.M. Jezequel, “Towards Flexible Evolution of Dynamically Adaptive Systems,” *Proc. 34th IEEE International Conference on Software Engineering (ICSE '12)*, pp. 1353–1356, 2012.
- [9] A.J. Ramirez, A.C. Jensen, and B.H.C Cheng, “Towards Flexible Evolution of Dynamically Adaptive Systems,” *Proc. 7th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS '12)*, pp. 99–108, 2012.
- [10] S. Hallsteinsen, M. Hinchey, S. Park, and K. Schmid, “Dynamic Software Product Lines,” *Computer*, vol. 41, no. 4, pp. 93–95, 2008.

Arquitecturas adaptativas en tiempo de ejecución mediante transformaciones inteligentes...

- [11] M. Hinchey, S. Park, and K. Schmid, "Building Dynamic Software Product Lines," *Computer*, vol. 45, no. 10, pp. 22–26, 2012.
- [12] C. Cetina, P. Giner, J. Fons, and V. Pelechano, "Autonomic Computing through Reuse of Variability Models at Runtime: The Case of Smart Homes," *Computer*, vol. 42, no. 10, pp. 37–43, 2009.
- [13] C. Cetina, P. Giner, J. Fons, and V. Pelechano, "Autonomic Computing through Reuse of Variability Models at Runtime: The Case of Smart Homes," *Computer*, vol. 42, no. 10, pp. 37–43, 2009.
- [14] B. Morin, O. Barais, J.M. Jezequel, F. Fleurey, and A. Solberg, "Models@run.time to Support Dynamic Adaptation," *Computer*, vol. 42, no. 10, pp. 44–51, 2009.
- [15] B. Morin, O. Barais, G. Nain, and J.M. Jezequel, "Taming Dynamically Adaptive Systems Using Models and Aspects," *Proc. 31st International Conference on Software Engineering (ICSE '09)*, pp. 122-132, 2009.
- [16] F. Imrert, T. Fischer, and K. Meyer-Wegener, "Runtime Adaptation in a Service-Oriented Component Model," *Proc. ICSE 2008 Work-shop on Software Engineering for Adaptive and Self-Managing Systems (SEAMS '08)*, pp. 97-104, May 2008.
- [17] E. Serral, P. Valderas, V. Pelechano, "Supporting runtime system evolution to adapt to user behaviour" *Proc. Advanced Information Systems Engineering*, pp. 378–392, 2010.
- [18] OSGi Alliance, *OSGi Service Platform, Rel. 4.1*, 2007.
- [19] D. Hearnden, M. Lawley, and K. Raymond, "Incremental Model Transformation for the Evolution of Model-Driven Systems," *Proc. 9th International Conference on Model Driven Engineering Languages and Systems (MoDELS '06)*, pp. 321-335, 2006.
- [20] I. Kurtev, K. van den Berg, and F. Jouault, "Rule-based modularization in model transformation languages illustrated with ATL," *Science Computer Programming*, vol. 68, no. 3, pp. 138–154, 2007.
- [21] M. Belaunde, "Transformation Composition in QVT," *Proc. First European Workshop Composition of Model Transformations (CMT '06)*, pp. 39-45, 2006.
- [22] D. Wagelaar, R. Van Der Straeten, and D. Deridder, "Module superimposition: a composition technique for rule-based model transformation languages," *Software and Systems Modeling*, vol. 9, no. 3, pp. 285–309, 2010.
- [23] D. Wagelaar, M. Tisi, J. Cabot, and F. Jouault, "Towards a general composition semantics for rule-based model transformation," *Proc. 14th International Conference on Model Driven Engineering Languages and Systems (MoDELS '11)*, pp. 623-637, 2011.
- [24] I. Porres, "Rule-based update transformations and their application to model refactorings," *Software and Systems Modeling*, vol. 4, no. 4, pp. 368–385, 2005.
- [25] D.S. Kolovos, R.F. Paige, F. Polack, and L.M. Rose, "Update Transformations in the Small with the Epsilon Wizard Language," *Journal of Object Technology*, vol. 6, no. 9, pp. 53–69, 2007.
- [26] M. Wimmer, S. Martínez, F. Jouault, and J. Cabot, "A Catalogue of Refactorings for Model-to-Model Transformations," *Journal of Object Technology*, vol. 11, no. 2, pp. 1–40, 2012.
- [27] M. Tisi, F. Jouault, P. Fraternali, S. Ceri, and J. Bezivin, "On the Use of Higher-Order Model Transformations," *Proc. Fifth European Conference on Model-Driven Architecture Foundations and Applications (ECMDA '09)*, pp. 18-33, 2009.
- [28] M. Tisi, J. Cabot, and F. Jouault, "Improving Higher-Order Transformations Support in ATL," *Proc. 3rd International Conference on Model Transformation (ICMT '10)*, pp. 215-229, 2010.

Arquitecturas adaptativas en tiempo de ejecución mediante transformaciones inteligentes...

- [28] Iribarne L, Troya JM, and Vallecillo A. A Trading Service for COTS Components. *Computer Jour*, 4(3):342-357, 2004.
- [29] T.A. Nguyen, M. Aiello, "Beyond Indoor Presence Monitoring with Simple Sensors", *Proceedings of the 2nd International Conference on Pervasive and Embedded Computing and Communication Systems*, 2012.
- [30] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, T. Weng, "Occupancy driven energy management for smart building automation", *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, BuildSys 10, ACM, New York, NY, USA, 2010, pp. 1–6, 2010.
- [31] A. Marchiori, Q. Han, "Distributed wireless control for building energy management", *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, BuildSys 10, ACM, New York, NY, USA, 2010, pp. 37–42, 2010.
- [32] C. Harris, V. Cahill, "Exploiting user behaviour for context-aware power management", *Wireless And Mobile Computing, Networking And Communications*, 2005. (WiMob2005), *IEEE International Conference*, vol. 4, pp. 122–130, 2005.
- [33] L. Hawarah, S. Ploix, M. Jacomino, "User behavior prediction in energy consumption in housing using Bayesian networks", *Proceedings of the 10th international conference on Artificial intelligence and soft computing: Part I*, ICAISC10, Springer-Verlag, Berlin, Heidelberg, pp. 372–379, 2010.
- [34] H. Hagraas, V. Callaghan, M. Colley, G. Clarke, A. Pounds-Cornish, H. Duman, "Creating an ambient-intelligence environment using embedded agents", *intelligent systems*, *IEEE*, vol. 19, no. 6, pp. 12–20, 2004.
- [35] R.H. Dodier, G.P. Henze, D.K. Tiller, X. Guo, "Building occupancy detection through sensor belief networks", *Energy and Buildings*, vol. 38, no. 9, pp. 1033–1043, 2006
- [36] Tuan Anh Nguyen, Marco Aiello, "Energy intelligent buildings based on user activity: A survey", *Energy and Buildings*, vol. 56, pp. 244–257, 2013.
- [37] Mihail-Bogdan Carutasu, Vladimir Tanasiev, Constantin Ionescu, Alexandra Danu, Horia Necula, Adrian Badea, "Reducing energy consumption in low energy buildings through implementation of a policy system used in automated heating systems", *Energy and Buildings*, vol. 94, no. 1, pp. 227–239, 2015.
- [38] Mavridis, Nikolaos and Pierris, Georgios and BenAbdelkader, Chiraz and Krstikj, Aleksandar and Karaiskos, Christos, "Smart buildings and the human-machine cloud", *GCC Conference and Exhibition (GCCCE)*, 2015 *IEEE 8th*, pp. 1–6, Feb, 2015.
- [39] Calvary G, Coutaz J, Thevenin D, Limbourg Q, Bouillon L, Vanderdonckt J. "A unifying reference framework for multi-target user interfaces". *Interacting with Computers* 2003; 15(3):289–308. DOI: 10.1016/S0953-5438(03) 00010-9.
- [40] OMG. MDA Guide, Version 1.0.1 2003.
- [41] C. Cetina, P. Giner, J. Fons, and V. Pelechano, "Autonomic Computing through Reuse of Variability Models at Runtime: The Case of Smart Homes," *Computer*, vol. 42, no. 10, pp. 37–43, 2009.
- [42] J. Sun, Y.P. Zhang, and J. Fan, "Towards a Context-aware Middle-ware in Smart Car Space," *Proc. 4th International Conference on Genetic and Evolutionary Computing (ICGEC '10)*, pp. 276-279, 2010.
- [43] F. Fouquet, B. Morin, F. Fleurey, O. Barais, N. Plouzeau, and J.M. Jezequel, "A Dynamic Component Model for Cyber Physical Systems," *Proc. 15th Symposium on Component Based Software Engineering (CBSE '12)*, pp. 135-144, 2012.

Arquitecturas adaptativas en tiempo de ejecución mediante transformaciones inteligentes...

- [44] G. Edwards, J. García, H. Tajalli, D. Popescu, N. Medvidovic, G. Sukhatme, and B. Petrus, "Architecture-Driven Self-Adaptation and Self-Management in Robotics Systems," Proc. ICSE 2009 Work-shop on Software Engineering for Adaptive and Self-Managing Systems (SEAMS '09), pp. 142-151, May 2009.
- [45] J.F. Inglés-Romero, C. Vicente-Chicote, B. Morin, and O. Barais, "Towards the Automatic Generation of Self-Adaptive Robotics Software: An Experience Report," Proc. 20th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE '11), pp. 79-86, 2011.
- [46] D. Garlan, S.W. Cheng, A.C. Huang, B. Schmerl, and P. Steenkiste, "Rainbow: Architecture-Based Self-Adaptation with Reusable Infrastructure," Computer, vol. 37, no. 10, pp. 46-54, 2004.
- [47] L. Iribarne, N. Padilla, J. Criado, J. Asensio, and R. Ayala, "A Model Transformation Approach for Automatic Composition of COTS User Interfaces in Web-Based Information Systems," Information Systems Management, vol. 27, no. 3, pp. 207-216, 2010.
- [48] J. Criado, C. Vicente-Chicote, L. Iribarne, and N. Padilla, "A Model-Driven Approach to Graphical User Interface Runtime Adaptation," Proc. 5th International Workshop Models@run.time, pp. 49-59, 2010.
- [49] E. Kroski, Web 2.0 for Librarians and Information Professionals, New York: Neal-Schuman, 2008.
- [50] O. Diaz, and J. Rodriguez, "Portlets as web components: An introduction," Journal of Universal Computer Science, vol. 10, no. 4, pp. 454-472, 2004.
- [51] L. Grammel, and M.A. Storey, "A Survey of Mashup Development Environments," The Smart Internet, pp. 137-151, 2010.
- [52] D. Rodríguez-Gracia, J. Criado, L. Iribarne, N. Padilla, C. Vicente-Chicote, "Adaptive Transformation for Architectural Models", JISBD, pp., 727-740, 2011.
- [53] D. Rodríguez-Gracia, J. Criado, L. Iribarne, N. Padilla, C. Vicente, "Runtime Adaptation of Architectural Models: an approach for adapting User Interfaces", LNCS 7602:16-30, 2012.
- [54] P.P Bonissone, "Soft computing: The convergence of emerging reasoning technologies", Soft Computing, 1 (1):6-18, 1997.
- [55] A. Abraham, "Neuro fuzzy systems: State-of-the-art modeling techniques", In: Connectionist models of neurons, learning processes, and artificial intelligence, pp. 269-276, 2001.
- [56] J.A. Piedra-Fernández, M. Cantón-Garbín, and J. Z. Wang, "Fuzzy Content-Based Image Retrieval for Oceanic Remote Sensing," IEEE Trans. Geosci. Remote Sens., DOI 10.1109/TGRS.2013.2288732, pp. 1-10, Dec. 2013.
- [57] J.A. Piedra-Fernández, M. Cantón-Garbín, and J.Z. Wang, "Feature selection in AVHRR ocean satellite images by means of filter methods," IEEE Trans. Geosci. Remote Sens., vol. 48, no. 12, pp. 4193-4203, Dec. 2010.
- [58] S.S. Parvinder and S. Hardeep "Automatic Reusability Appraisal of Software Components using Neuro Fuzzy Approach", International Journal Of Information Technology, 3(3):209-214, 2006.
- [59] S. Harpreet, K.T. Vishal, "Neuro Fuzzy Logic Model for Component Based Software Engineering", International Journal of Engineering Sciences, 1:303-314, 2011.
- [60] Leake, D. B. Case-based reasoning. John Wiley and Sons Ltd, pp. 196-197, 2003.

Modelado y Optimización de Problemas en Sanidad vía Computación de Altas Prestaciones

Una metodología de decisión aplicada al procesado de alimentos, una versión paralela del algoritmo de optimización y el ajuste de los parámetros de modelos en epidemiología

Miriam R. Ferrández

Dept. de Informática, Universidad de Almería, ceiA3, Ctra. Sacramento, La Cañada de San Urbano, 04120 Almería, España
mferrandez@ual.es

Abstract. Este trabajo refleja los avances en los dos ámbitos incluidos en la tesis doctoral: el procesado de alimentos y la propagación de enfermedades. El nexo común entre ambos es el uso de una metodología de optimización basada en el algoritmo WASF-GA. Esta metodología ha sido analizada en cuanto a eficacia y eficiencia. Además, se ha diseñado una versión paralela, que ha demostrado tener un comportamiento escalable y casi lineal. En cuanto al procesado de alimentos, se propone una herramienta de decisión que permite obtener un conjunto de configuraciones óptimas para el tratamiento en términos de presión y temperatura, de modo que puedan ser utilizadas para satisfacer distintos escenarios de calidad. Respecto a la propagación de enfermedades, se ha abordado el problema inverso consistente en averiguar los parámetros de un determinado modelo para que sus resultados se ajusten a la realidad.

1 Introducción

El proyecto en el que se desarrolla esta tesis doctoral engloba dos problemas en el ámbito de la sanidad: (i) el procesado de alimentos por altas presiones y (ii) la propagación de enfermedades. El marco común es la utilización de técnicas avanzadas de optimización y de computación de altas prestaciones para el tratamiento de dichos problemas.

Nuestra principal contribución al ámbito de la industria alimenticia es una herramienta de decisión que pone a disposición del ingeniero un conjunto de configuraciones para el procesado de un determinado alimento mediante la aplicación de un tratamiento combinado de temperatura y altas presiones. En este sentido, se ha llevado a cabo la ejecución de experimentos computacionales dirigidos a motivar el uso de esta herramienta de decisión y comprobar su funcionamiento utilizando distintos parámetros que afectan al algoritmo de optimización multiobjetivo.

Debido al alto coste en tiempo computacional de la metodología de optimización para el problema de tratamiento de alimentos por altas presiones, se ha abordado la introducción de técnicas de computación de altas prestaciones. En particular, se ha propuesto una versión paralela del algoritmo de optimización multi-objetivo enfocada a repartir las evaluaciones de la función objetivo entre los distintos procesadores utilizando una memoria compartida. Además, se ha realizado una batería de experimentos numéricos con los cuales se ha podido medir la eficiencia del algoritmo paralelo en términos del speedup y que han servido también para motivar el uso de este tipo de estrategias paralelas en problemas computacionalmente costosos como el que nos ocupa.

Por otra parte, en la segunda vía del proyecto dedicada a la propagación de enfermedades, nos hemos centrado en el planteamiento del problema inverso que consiste en determinar los parámetros de ciertos modelos de epidemiología utilizando técnicas multi-objetivo. En este sentido, se han desarrollado tanto pruebas numéricas como demostraciones analíticas.

2 Metodología de optimización

Tras una revisión de los métodos de optimización disponibles en la literatura, hemos optado por utilizar el algoritmo metaheurístico llamado *Weighting Achievement Scalarizing Function Genetic Algorithm* (WASF-GA) [1], puesto que ha demostrado ser eficiente en problemas con más de dos objetivos y, además, permite centrar la búsqueda de soluciones en una región de interés determinada por las preferencias del usuario o tomador de decisiones.

Como su propio nombre indica, la base de WASF-GA es un algoritmo genético que aplica iterativamente las operaciones de cruce y mutación, inspiradas en la teoría de la evolución de Darwin, sobre un conjunto de soluciones denominado población. La característica principal de WASF-GA, que lo distingue de otros métodos, es su mecanismo de selección utilizado para conformar la nueva población tras esa generación de nuevos individuos. Esta selección promueve la supervivencia de aquellos individuos ubicados en la región de interés (ROI) mediante el uso de una función de escalarización (ASF), conocida con el nombre de Wierzbicki y basada en la distancia L_∞ .

De esta forma, para llevar a cabo la ejecución del algoritmo WASF-GA, se han de proporcionar los siguientes parámetros de entrada: (i) un punto de referencia $\mathbf{q} = (q_1, \dots, q_m)$ cuyas componentes son los valores preferidos por el usuario para cada una de las funciones objetivo, este punto determinará la ROI; (ii) un conjunto de N_μ vectores de pesos $\{\mu^1, \dots, \mu^{N_\mu}\}$, cada uno de ellos con un número de componentes igual al número m de funciones objetivo, es decir, $\mu^j = (\mu_1^j, \dots, \mu_m^j)$ y $\mu_i^j > 0$ para $j = 1, \dots, N_\mu$ y $i = 1, \dots, m$; (iii) el número máximo de generaciones h_{\max} ; y (iv) el número N de individuos en la población.

El conjunto de vectores de pesos se utiliza durante el proceso de selección de WASF-GA. La función de escalarización ASF se evalúa para cada uno de los individuos de la población con cada uno de los vectores de pesos μ^j , $j = 1, \dots, N_\mu$. Entonces, teniendo en cuenta los valores obtenidos, se lleva a cabo la

clasificación de los individuos en distintos frentes como se detalla a continuación. En primer lugar, se selecciona el individuo que menor valor ofrece para la ASF con el primero de los vectores de pesos. Este individuo se copia en la primera posición del primer frente y ya no será considerado para las siguientes selecciones. En segundo lugar, de entre los individuos restantes, aquel con menor valor para la ASF con el segundo de los vectores de pesos será el siguiente individuo del primer frente. Este procedimiento se repite hasta que el primer frente contiene un elemento asociado a cada uno de los vectores de pesos y se continúa completando análogamente los siguientes frentes hasta que todos los individuos de la población hayan sido clasificados. Finalmente, se compone la población para la siguiente iteración tomando los N primeros individuos del primer frente. En caso de no ser suficientes, se completa hasta llegar a N individuos con los situados en las primeras posiciones del siguiente frente.

Nuestra contribución en cuanto a la metodología de optimización consiste en una versión paralela de WASF-GA, a la cual hemos llamado ParWASF-GA. La idea fundamental es que la evaluación de cada individuo es independiente del resto de individuos de la población, por lo que es una tarea intrínsecamente paralelizable. ParWASF-GA ha sido diseñado para memoria compartida. En primer lugar, se crean los diferentes hilos y se distribuye el trabajo entre ellos, es decir, se asigna un número de individuos de la población total a cada unidad de procesamiento de forma equitativa. De este modo, para la creación de la población inicial cada unidad genera tantos individuos como le corresponda siguiendo al mismo procedimiento de generación aleatoria que propone WASF-GA. Antes de comenzar el proceso iterativo, se implementa un primer punto de sincronismo que garantice que la etapa de inicialización se ha completado. A continuación, en cada iteración, cada una de las unidades de procesamiento selecciona de entre todos los individuos de la población los padres necesarios para realizar las operaciones de cruce y mutación que den lugar a tantos nuevos individuos como les haya sido asignado en el reparto. Esos nuevos individuos se incluyen en la población compartida por todas las unidades. Para asegurar que todos los individuos estén disponibles, se ha considerado un segundo punto de sincronismo previo al proceso de clasificación en frentes y selección de individuos para la nueva población, ya que esta etapa no es paralelizable. Tal y como hemos descrito anteriormente, es necesario ir conociendo lo que ocurre con cada uno de los individuos a lo largo de la clasificación en los distintos frentes, por tanto, ha de llevarse a cabo en una misma unidad de procesamiento. Sin embargo, esto no repercute negativamente en la eficiencia de la versión paralela. De hecho, su paralelización podría ser contraproducente puesto que el coste computacional de este proceso es insignificante (1200 veces más rápido que una sola evaluación de la función objetivo).

La finalidad de nuestro estudio ha sido evaluar la eficiencia de ParWASF-GA variando los parámetros que influyen directamente en el número de evaluaciones y, con ello, en el tiempo de cómputo y la carga de trabajo: el número de generaciones h_{\max} y número de individuos N .

Modelado y Optimización de Problemas en Sanidad vía Computación de Altas Prestaciones

En primer lugar, hemos observado que las técnicas de paralelismo aplicadas en ParWASF-GA permiten reducir considerablemente los tiempos de cómputo. Para una población de $N = 200$ individuos y $h_{\max} = 30$ iteraciones el tiempo de cálculo de WASF-GA es de 82.87 horas, mientras que ParWASF-GA utilizando 32 cores tarda tan sólo 3.65 horas. En cuanto a los experimentos con número de población fijo $N = 200$, en los que hemos variado el número de generaciones h_{\max} , hemos comprobado que la eficiencia de ParWASF-GA no se ve afectada por los cambios en dicho parámetro. Este resultado es bastante razonable ya que h_{\max} sólo escala el tiempo total de cómputo sin afectar al número de evaluaciones en cada generación. Sin embargo, es importante la comprobación de este hecho, dado que h_{\max} sí afecta a la calidad de la solución: cuantas más iteraciones, más precisas y mejor distribuidas serán las soluciones.

Por otra parte, del estudio con distintos tamaños de población N y un número fijo de $h_{\max} = 30$ generaciones, podemos concluir que, cuanto mayor es el número de individuos, mejor se comporta la versión paralela, esto es, que el algoritmo es escalable (ver Figura 1). Además, como podemos observar también en la Figura 1, el speedup obtenido para ParWASF-GA es casi-lineal hasta 16 cores.

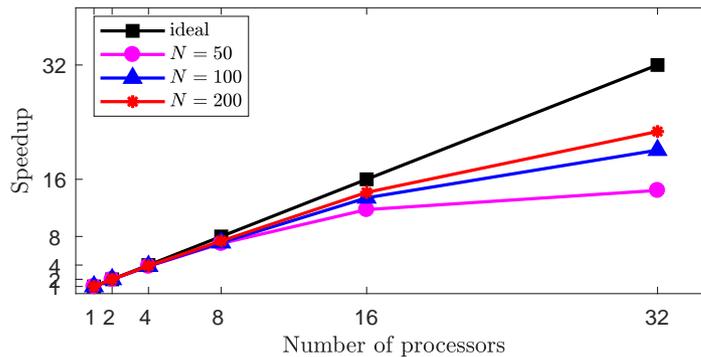


Fig. 1. Speedup variando el número de individuos N en la población para un número fijo de generaciones $h_{\max} = 30$.

El comportamiento de ParWASF-GA continúa siendo bastante bueno cuando se utiliza también paralelismo en los métodos numéricos involucrados en la evaluación de la función objetivo. Sin embargo, hemos podido comprobar que, en el problema que nos ocupa, resulta más eficiente la paralelización del algoritmo de optimización que la de estos métodos numéricos.

Este trabajo está en proceso de *minor revision* para ser publicado en la revista Journal of Supercomputing [2].

3 Procesado de alimentos por altas presiones

El tratamiento de los alimentos para su conservación es un aspecto de vital importancia en las empresas dedicadas a la elaboración y distribución de productos alimenticios. En sus exportaciones a mercados internacionales es fundamental adaptarse a las medidas de calidad impuestas por la legislación vigente en cada país. De este modo, lo más frecuente es que, para un mismo producto, se tengan que diseñar tratamientos diferentes en función del país de destino. Este diseño consiste en determinar la configuración del equipo utilizado para el tratamiento con vistas a cumplir unas restricciones de calidad concretas.

En este trabajo, el tratamiento considerado es el procesado por altas presiones y temperatura, ya que es el más extendido en la industria en las últimas décadas. El equipo de altas presiones está formado por una estructura de acero cilíndrica en el interior de la cual se encuentra el fluido presurizante rodeando al alimento. Su funcionamiento es el siguiente: tras introducirse el alimento envasado, se activan las denominadas bombas presurizantes que aumentan la presión sobre el fluido hasta alcanzar en cada momento la presión indicada por el usuario. Como consecuencia de ello, el alimento sufre también un incremento de su presión y de su temperatura internas que llevarán a la disminución de las actividades de las vitaminas y las enzimas presentes en él. El usuario puede influir en el tratamiento a partir de tres magnitudes de entrada que proporcionará al equipo: la presión a aplicar al fluido que denotaremos P , la temperatura inicial a la que se encuentra el sistema T_0 y la temperatura de refrigeración T_r que servirá para enfriar las paredes de acero cercanas a la zona del alimento y, así, evitar que las temperaturas sean demasiado elevadas.

En este trabajo, el estudio sobre la calidad del alimento se ha centrado en el análisis de la actividad de dos moléculas presentes en el mismo. Por un lado, la enzima llamada Bacillus Subtilis α - Amylase (BSAA) [3], causante de cambios en el sabor e incluso intoxicaciones, cuya actividad se desea reducir. Por otra parte, la vitamina C [4], cuya actividad se quiere preservar.

Teniendo en cuenta el funcionamiento del equipo de altas presiones, cuanta más presión se aplique, mayores serán las temperaturas que se alcancen en el alimento y más decrecerá la actividad enzimática. Sin embargo, también se verá reducida la actividad vitamínica. Además, las altas temperaturas dañan otras propiedades organolépticas del producto ligadas a la calidad. Por tanto, determinar las condiciones de presión y temperatura con las que configurar el equipo para conseguir el resultado deseado sobre el alimento se puede formular en términos de un problema multi-objetivo en el que se plantean distintos objetivos contradictorios para encontrar una solución de compromiso entre ellos. En concreto, para el caso que nos ocupa y que acabamos de detallar, el problema es el siguiente:

Problema 1

$$\begin{cases} \min f_{\text{bsaa}}(T_0, T_r, \Delta P_1, \dots, \Delta P_n), \\ \max f_{\text{vit}}(T_0, T_r, \Delta P_1, \dots, \Delta P_n), \\ \min f_{\text{Tmax}}(T_0, T_r, \Delta P_1, \dots, \Delta P_n). \end{cases} \quad (1)$$

Modelado y Optimización de Problemas en Sanidad vía Computación de Altas Prestaciones

Las dos primeras funciones objetivo se definen como la actividad final, es decir, en el instante t_f , promediada en el dominio Ω_F ocupado por el alimento:

$$f_j(T_0, T_r, \Delta P_1, \dots, \Delta P_n) = \frac{1}{|\Omega_F|} \iint_{\Omega_F} A_j(r, z, t_f) dr dz.$$

En la expresión anterior, denotamos con $A_j(r, z, t)$ la actividad de una molécula ubicada en un punto $(r, z) \in \Omega_F$ en un instante t , y donde el subíndice j se utiliza para distinguir entre la actividad enzimática $j = \text{bsaa}$ o vitamínica $j = \text{vit}$. La última función objetivo $f_{T_{\max}}$ es la temperatura máxima alcanzada en el alimento durante todo el tratamiento, es decir:

$$f_{T_{\max}}(T_0, T_r, \Delta P_1, \dots, \Delta P_n) = \max_{(r,z) \in \Omega_F, t \in [t_0, t_f]} T(r, z, t).$$

Evaluar estas tres funciones objetivo para un tratamiento de HPT determinado por $(T_0, T_r, \Delta P_1, \dots, \Delta P_n)$ conlleva resolver numéricamente un sistema de transferencia de calor acoplado con las ecuaciones cinéticas que describen la evolución de las actividades de la enzima y de la vitamina. Para más detalle sobre las ecuaciones implicadas se remite al lector a la referencia [5].

Hemos de reseñar que, para permitir una presión P variable a lo largo del proceso, hemos considerado que la presión inicial, en $t_0 = 0$, es la atmosférica $P_0 = 0.1$ (MPa) y que en cada intervalo de tiempo $[t_i, t_{i+1}]$, con $t_i = i \cdot \frac{900}{n}$, $i \in \{1, \dots, n\}$ el usuario aplica una variación de presión constante $\Delta P_i \in [\Delta P_{n,\text{dec}}, \Delta P_{n,\text{inc}}]$ (MPa), donde $\Delta P_{n,\text{dec}}$ y $\Delta P_{n,\text{inc}}$ son las variaciones máximas permitidas por el equipo para la disminución y para el aumento de presión, respectivamente, durante 900/ n segundos. En este trabajo, teniendo en cuenta las limitaciones del equipo de altas presiones considerado, asumimos que el rango de temperatura admisible es $[10, 50]$ ($^{\circ}\text{C}$), el rango de presión admisible es $[0.1, 900]$ (MPa) y la gama de variaciones admisibles para la presión es $[\Delta P_{n,\text{dec}}, \Delta P_{n,\text{inc}}] = [-250, 250]$ (MPa). Por lo que todos los problemas tratados se han resuelto considerando estas restricciones.

Para motivar la utilidad de la herramienta de decisión, se han planteado dos problemas específicos, **Problema 2** y **Problema 3**, que corresponden a dos situaciones concretas en las que se define un único objetivo sujeto a unas restricciones sobre la calidad del producto.

Problema 2	Problema 3
$\min f_{\text{bsaa}}(T_0, T_r, \Delta P_1, \dots, \Delta P_n)$	$\max f_{\text{vit}}(T_0, T_r, \Delta P_1, \dots, \Delta P_n)$
s.t. $f_{\text{vit}}(T_0, T_r, \Delta P_1, \dots, \Delta P_n) \geq 0.97$	s.t. $f_{\text{bsaa}}(T_0, T_r, \Delta P_1, \dots, \Delta P_n) \leq 0.4$
$f_{T_{\max}}(T_0, T_r, \Delta P_1, \dots, \Delta P_n) \leq 50$	$f_{T_{\max}}(T_0, T_r, \Delta P_1, \dots, \Delta P_n) \leq 47$

Tanto el **Problema 2** como el **Problema 3** han sido resueltos con el algoritmo mono-objetivo MLS-GA [6], que hemos diseñado combinando un algoritmo genético con un método multicapa [7] y un algoritmo de descenso para mejorar la precisión de la solución. En cuanto al **Problema 1**, se ha resuelto con la metodología multi-objetivo que se detalla en la Sección 2. Dicha metodología ofrece como solución un conjunto de puntos que conforman una aproximación

al denominado frente de Pareto [8], formado por las soluciones que logran un compromiso óptimo e inmejorable entre los distintos objetivos. Los experimentos computacionales realizados han ido dirigidos a comprobar si entre los puntos de esa aproximación obtenida con la resolución del **Problema 1**, podemos encontrar soluciones válidas para problemas mono-objetivo cualesquiera tales como los **Problemas 2** y **3**. Por tanto, tras obtener ese conjunto de puntos que ofrece el algoritmo multi-objetivo, se aplica un mecanismo de decisión que consiste en una primera etapa de filtrado de los individuos que satisfacen las restricciones de un determinado problema mono-objetivo y que, por tanto, se encuentran entre las soluciones admisibles; y una segunda etapa de ordenación o ranking en la que, entre los individuos que han pasado el filtro, se selecciona aquél con el menor valor en el caso de un problema de minimización o mayor valor si estamos ante un problema de maximización.

Comparando los resultados de la metodología de decisión propuesta utilizando WASF-GA y del algoritmo mono-objetivo MLS-GA podemos concluir que, por lo general, WASF-GA parece proporcionar soluciones más precisas para ambos problemas, incluso en aquellos casos en los que WASF-GA tiene menor coste computacional que MLS-GA. De este estudio se deduce la capacidad de la herramienta de decisión basada en WASF-GA para encontrar una solución en el conjunto de puntos de la aproximación al frente de Pareto que parece ser similar o incluso más precisa que la solución de MLS-GA. Por tanto, queda demostrada su utilidad para resolver distintos escenarios de calidad con un mismo conjunto de soluciones. Esto implica un ahorro importante en tiempos de cómputo para la industria alimenticia, ya que utilizando un algoritmo mono-objetivo como MLS-GA, se tendría que ejecutar un nuevo proceso de optimización cada vez que se deseen cambiar las restricciones sobre la calidad, por ejemplo para adaptarlas a la legislación de otro país. Sin embargo, utilizando la metodología propuesta, con una sola ejecución del algoritmo WASF-GA, los ingenieros tienen a su disposición todo un conjunto de puntos de entre los que seleccionar aquel adecuado a cada situación o país.

Además, hemos explorado distintas configuraciones para los parámetros de cruce y mutación del algoritmo y hemos determinado cuál es la que mejores resultados aporta en cuanto a precisión y distribución de las soluciones y en cuanto a resolución de los distintos problemas mono-objetivo [6].

Por otro lado, hemos analizado la sensibilidad de la herramienta de decisión ante pequeños cambios en los parámetros de entrada del modelo $T_0, T_r, \Delta P_1, \dots, \Delta P_n$ y hemos observado que las temperaturas inicial y de refrigeración parecen ser los parámetros más influyentes. Por lo tanto, parece recomendable intentar evitar perturbaciones en los datos de temperatura proporcionados al equipo de altas presiones. En cuanto a la presión, las soluciones parecen ser robustas ante variaciones razonables, es decir, si se comete alguna ligera variación en la entrada de datos, ésta parece que no tendría apenas consecuencias sobre el funcionamiento del proceso.

Este trabajo ha sido enviado a la revista Journal of Food Engineering para su publicación [6].

4 Propagación de enfermedades

Como primera etapa de nuestro trabajo en epidemiología, se ha estudiado el modelo de propagación conocido por las siglas SIRS, ya que es el más sencillo para describir la evolución de enfermedades en las que el individuo se recupera con una inmunidad temporal, de modo que, tras un tiempo, vuelve a ser de nuevo susceptible de ser infectado [9]. Es un modelo de tipo compartimental en el que se consideran tres estados disjuntos para clasificar a los individuos: susceptibles, infectados, y recuperados. Para simplificar, se suele considerar que la población total del país objeto de estudio es constante y se denota por S , I y R al ratio de personas en cada uno de los estados susceptible, infectado y recuperado, respectivamente. En estos términos, SIRS describe la evolución temporal de la enfermedad a partir de las siguientes ecuaciones diferenciales:

$$\frac{dS(t)}{dt} = -\beta I(t)S(t) + \mu R(t), \quad \frac{dI(t)}{dt} = \beta I(t)S(t) - \gamma I(t), \quad \frac{dR(t)}{dt} = \gamma I(t) - \mu R(t),$$

donde se cumple que $S(t) + I(t) + R(t) = 1, \forall t \geq 0$, y los parámetros involucrados son: la tasa de contacto efectivo β (día⁻¹ persona⁻¹), la tasa de pérdida de inmunidad μ (día⁻¹) y la tasa de transición γ del estado I al R (día⁻¹).

A partir de estas ecuaciones, si suponemos conocidas las tasas o parámetros del modelo β , μ y γ , y el estado inicial de la enfermedad (S_0, I_0, R_0) , podemos determinar el estado de la enfermedad $(S(t), I(t), R(t))$ en cada instante de tiempo $t > 0$ resolviendo numéricamente el sistema anterior de ecuaciones diferenciales. Lo cual es de suma utilidad para hacer predicciones sobre el comportamiento de la epidemia. Sin embargo, determinar los parámetros β , μ y γ propios de cada enfermedad a partir de los casos reales que van documentándose no es tarea fácil, debido fundamentalmente a la escasez, e incluso a veces inexactitud, de los datos. Por ello, nos hemos planteado el siguiente problema inverso: conocida la evolución de la enfermedad $(S(t), I(t), R(t)), \forall t \geq 0$, determinar los parámetros (β, μ, γ) que ajustan el modelo SIRS.

A nivel teórico, hemos demostrado analíticamente que tiene solución y que, además, es única. Además, también hemos podido demostrar que dos epidemias descritas mediante el modelo SIRS tienen el mismo punto de equilibrio endémico $P_e = (S_e, I_e, R_e)$ si, y sólo si, sus conjuntos de parámetros, $(\beta_1, \mu_1, \gamma_1)$ y $(\beta_2, \mu_2, \gamma_2)$, son proporcionales, es decir, $\frac{\beta_1}{\beta_2} = \frac{\mu_1}{\mu_2} = \frac{\gamma_1}{\gamma_2}$. A nivel práctico, hemos realizado distintos experimentos cuyos resultados respaldan dichas afirmaciones.

Como segunda etapa, se ha abordado el estudio del modelo más complejo SEIHRD, que, además de los descritos en SIRS, incorpora los siguientes estados: infecciosos E , hospitalizados H y difuntos D . Este modelo ya había sido utilizado con éxito, como núcleo de la herramienta Be-CoDiS, para describir la epidemia de Ébola del año 2015 [10, 11].

Basándonos en los datos reales disponibles sobre la epidemia, se ha utilizado el algoritmo de optimización WASF-GA para obtener los parámetros del modelo SEIHRD. En concreto, se han considerado dos funciones objetivo a minimizar: el error cometido en el número de casos detectados y el error cometido en el número de muertes. Como podemos observar en la Figura 2, esta metodología nos permite hallar un conjunto de valores para los parámetros que dan como resultado del

modelo una evolución simulada de la enfermedad que es prácticamente idéntica a la real.

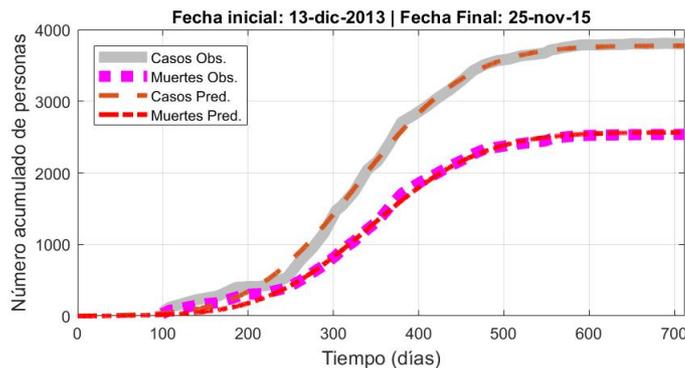


Fig. 2. Casos y muertes para la epidemia de Ébola en Guinea. Comparación de los datos observados con los datos aportados por el modelo SEIHRD.

5 Conclusiones

La versión paralela ParWASF-GA que se propone permite resolver problemas de optimización cuyas funciones objetivo son demasiado costosas de evaluar, consiguiendo una reducción en tiempo computacional casi proporcional al número de unidades de procesamiento utilizadas. Gracias al ahorro en tiempo, utilizando ParWASF-GA, podemos abarcar un mayor número de iteraciones en el proceso de optimización. Este aumento en iteraciones repercute directamente en la calidad de las soluciones que serán más precisas y estarán mejor distribuidas. Además, ParWASF-GA está especialmente recomendado si se requiere trabajar con poblaciones grandes, ya que se obtiene mejor rendimiento del algoritmo. En nuestro caso, tener un mayor número de individuos en la población se traduce en un aumento del número de soluciones que le podemos aportar al ingeniero de alimentos con la metodología de decisión diseñada. Como hemos visto, esto implica cubrir un mayor número de escenarios de calidad con los que satisfacer las necesidades de la industria sin necesidad de ejecutar nuevos procesos de optimización.

Por último, basándonos también en la optimización, hemos planteado una metodología para recuperar los parámetros de los modelos utilizados en epidemiología de modo que éstos se ajusten de manera precisa al comportamiento real de la propagación de la enfermedad. En particular, para el modelo SEIHRD aplicado a la epidemia de Ébola hemos obtenido muy buenos resultados logrando unos parámetros que describen casi de forma exacta el brote de la enfermedad.

Agradecimientos

El presente trabajo ha sido financiado por el Ministerio de Economía y Competitividad de España mediante los proyectos TIN2015-66680-C2-1-R y MTM2015-64865P; por la Junta de Andalucía, a través de los proyectos P11-TIC7176 y P12-TIC301, financiados parcialmente por el Fondo Europeo de Desarrollo Regional (FEDER). Juana López Redondo disfruta de una beca del programa “Ramón y Cajal”, cofinanciado por el Fondo Social Europeo.

References

1. Ruiz, A.B., Saborido, R., Luque, M.: A preference-based evolutionary algorithm for multiobjective optimization: the weighting achievement scalarizing function genetic algorithm. *Journal of Global Optimization* **62**(1) (2015) 101–129
2. Ferrández, M.R., Puertas-Martín, S., Redondo, J.L., Ivorra, B., Ramos, A.M., Ortigosa, P.M.: High-performance computing for the optimization of high-pressure thermal treatments in food industry. To Appear in *Journal of Supercomputing*
3. Denys, S., Ludikhuyze, L.R., Van Loey, A.M., Hendrickx, M.E.: Modeling Conductive Heat Transfer and Process Uniformity during Batch High-Pressure Processing of Foods. *Biotechnology Progress* **16**(1) (2000) 92–101
4. Verbeyst, L., Bogaerts, R., Van der Plancken, I., Hendrickx, M., Van Loey, A.M.: Modelling of vitamin C degradation during thermal and high-pressure treatments of red fruit. *Food and Bioprocess Technology* **6**(4) (2013) 1015–1023
5. Infante, J.A., Ivorra, B., Ramos, A.M., Rey, J.M.: On the Modelling and Simulation of High Pressure Processes and Inactivation of Enzymes in Food Engineering. *Mathematical Models and Methods in Applied Sciences (M3AS)* **19**(12) (2009) 2203–2229
6. Ferrández, M.R., Redondo, J.L., Ivorra, B., Ramos, A.M., Ortigosa, P.M.: A decision tool based on a multi-objective methodology for designing high-pressure thermal treatments in food industry. Preprint version (September 2017)
7. Ivorra, B., Mohammadi, B., Ramos, A.M.: A multi-layer line search method to improve the initialization of optimization algorithms. *European Journal of Operational Research* **247**(3) (2015) 711–720
8. Coello, C.A.C., Lamont, G.B., Van Veldhuizen, D.A.: Evolutionary algorithms for solving multi-objective problems. Second edn. Genetic and evolutionary computation. Springer, New York (2007)
9. Ma, S., Xia, Y.: Mathematical Understanding of Infectious Disease Dynamics. Lecture Notes Series, Institute for Mathematical Sciences. World Scientific (2009)
10. Ivorra, B., Ngom, D., Ramos, A.M.: Stability and sensitivity analysis of be-codis, an epidemiological model to predict the spread of human diseases between countries. *Biomath Communications Supplement* **4**(1) (2017)
11. Ivorra, B., Ngom, D., Ramos, A.M.: Be-codis: A mathematical model to predict the risk of human diseases spread between countries—validation and application to the 2014–2015 ebola virus disease epidemic. *Bulletin of Mathematical Biology* **77**(9) (Sep 2015) 1668–1704

Distribución de carga en sistemas globales de streaming de vídeo en directo

Redes híbridas P2P y CDN usando el Servidor Icecast con un peer como cliente de BitTorrent

Teresa Santamaria-López

Universidad de Almería, Sacramento S/N, Almería, España
tsantamaria710@gmail.com

Resumen: Las redes híbridas P2P y CDN parecen ser la solución más acertada, CDN son todo ventajas, excepto que no escalan de forma barata, cosa que no le pasa a una red P2P. Por este motivo, decidimos estudiar problema de, si hay que desplegar una red de difusión de vídeo, y sabiendo que va a ser una mezcla de CDN y P2P, dado una latencia máxima conocida a priori que debe decidir el proveedor de contenidos, determinar cuántos servidores CDN y con qué capacidad se deberían instalar, teniendo en cuenta que es posible calcular la latencia tanto entre nodos CDN, como en una red P2P. La tecnología Streaming media ha experimentado un crecimiento y actualmente es una popular zona de investigación, la mayoría de los trabajos sobre la tecnología streaming media es asegurar la reproducción de vídeo y la escalabilidad. Existen escasas investigaciones para hacer frente a las limitaciones de la red. En este trabajo investigamos como montar la CDN usando Icecast y como cliente rápido un peer de Bittorrent con posibilidades de recibir el stream en secuencial, además de investigar uno de los principales clientes de BitTorrent.

Palabras claves: Streaming, Redes híbridas P2P y CDN, Servidor Icecast

1 Introducción

El streaming de vídeo es una de las aplicaciones más usadas en Internet y la que, con diferencia, mayor ancho de banda consume. Por dicho motivo, son necesarias estrategias de distribución eficiente de la carga, que básicamente consiste en decidir cómo se distribuye el contenido multimedia para minimizar el gasto de ancho de banda y maximizar la calidad de la experiencia de los usuarios. Por otra parte, la manipulación de contenidos "en directo" (lo que significa que el vídeo se genera a la vez que se transmite), impone restricciones temporales de transmisión que deben tenerse en consideración. Gracias a las tecnologías de información y comunicación, han surgido diferentes maneras para comunicarse, una de las formas que ha ido evolucionando es la mensajería instantánea, y con ella la videoconferencia y el en-

vío de audio y vídeo [2]. Dando lugar a la aparición de Streaming, el cual juntamente con el progreso de las redes sociales, radios online, etc.

En los últimos años, con la proliferación de redes de alta velocidad, hay un crecimiento acelerado del tráfico de datos en internet sobre todo el tráfico relacionado con el vídeo. Peer to Peer (P2P) es un modelo de streaming viable que es capaz de superar el embotellamiento de streaming server centralizado debido a su diseño y arquitectura distribuida.

En términos de distribución de gran cantidad de datos a través de internet, Bit-torrent, un P2P para compartir archivos de protocolo, ha demostrado éxito y logro inspirar el uso del modelo de intercambio de archivos P2P.

En este artículo se pretende investigar el rendimiento P2P para la transmisión de multimedia.

1.1 Red p2p

Se refiere a una red que no tiene clientes ni servidores fijos, sino una serie de nodos que se comportan simultáneamente como clientes y como servidores respecto de los demás nodos de la red. Es una forma legal de compartir archivos de forma similar a como se hace en el email o mensajes.

Las redes de ordenadores Peer- Too- Peer (P2P) son redes que aprovechan, administran y optimizan el uso de banda ancha que acumulan de los demás usuarios en una red por medio de la conectividad entre los mismos usuarios participantes de la red, obteniendo como resultado mucho más rendimiento en las conexiones y transferencias que con algunos métodos centralizados convencionales, donde una cantidad relativamente pequeña de servidores provee el total de banda ancha y recursos compartidos para un servicio o aplicación [3]. Estas redes se conectan en gran parte con otros nodos vía "ad hoc", Los beneficios potenciales de P2P, es la escalabilidad, rendimiento y robustez [4].

1.2 Características

Escalabilidad: Las redes P2P tienen un alcance mundial con cientos de millones de usuarios potenciales. En general, lo deseable es que cuantos más nodos estén conectados a una red P2P mejor será su funcionamiento. Así, cuando los nodos llegan y comparten sus propios recursos, los recursos totales del sistema aumentan.

Robustez: La naturaleza distribuida de las redes Peer – Too-Peer (P2P) también incrementa la robustez en caso de haber fallos en la réplica excesiva de los datos hacia múltiples destinos.

Descentralización: estas redes por definición son descentralizadas y todos los nodos son iguales. No existen nodos como funciones especiales, y por tanto ningún nodo es imprescindible para el funcionamiento de la red.

Anonimato: Es deseable que en estas redes quede anónimo el autor de un contenido, editor, lector, el servicio que lo alberga y la petición para encontrarlo siempre que así lo necesiten los usuarios. Muchas veces el derecho al anonimato y los derechos de autor son incompatibles entre sí.

Seguridad: Es una de las características deseables de las redes P2P menos implementada. Los objetivos de un P2P seguro serían identificar y evitar los nodos maliciosos, evitar el contenido infectado, evitar el espionaje de las comunicaciones entre nodos, creación de grupos seguros de nodos dentro de la red, protección de los recursos de la red, los mecanismos más prometedores son: cifrado multiclave, cajas de arena, gestión de derechos de autor, reputación, comunicaciones seguras, comentarios sobre los ficheros.

1.3 CDN

Es una red de servidores distribuidos por todo el mundo desde la cual el usuario que visita un portal web, dependiendo de su ubicación, se descarga la información proveniente del servidor original donde se subió la web de uno u otro nodo que forma parte de la red CDN. Esto provoca que la descarga de esa información sea mucho más rápida, además de evitar el problema de saturar el servidor donde se aloja el dominio, ya que esta descarga de información se hace desde uno de estos servidores que forman parte de la red en vez de hacerlo desde el servidor original. [1].

Los contenidos que se suelen almacenar en este tipo de redes son de tipo estático, tales como imágenes, documentos, videos, audios entre otros. Todos esos elementos los tiene la red CDN en su cache durante un tiempo determinado, pasado ese tiempo se borra teniéndolos que volver a solicitar al hosting. Cuando un usuario solicita un tipo de información, lo que hace esta red es conectar con el servidor web real para pedirle a aquellos de los que aún no dispone, además de los contenidos dinámicos que no son almacenados en su sistema de cache.

El incremento de la velocidad es la ventaja más importante que ofrece el uso de algunas de las redes CDN. Al estar en cache algunos de los elementos más pesados como las imágenes o ficheros de estilo, la velocidad de carga se ve beneficiada, ya que no será necesario que el servidor sirva toda esa información, lo que agilizará considerablemente la velocidad del sitio.

1.4 Icecast

Icecast es un servidor streaming que permite a las emisoras de radio transmitir con eficacia en audio online, esto se conoce como la difusión por internet. Es potente y estable, lo que significa que puede lanzar todo su tráfico de radio en internet sin tener accidentes u otros problemas, oyentes pueden acceder el audio streaming a través de cualquier reproductor de MP3 en vivo (ver Fig. 1)

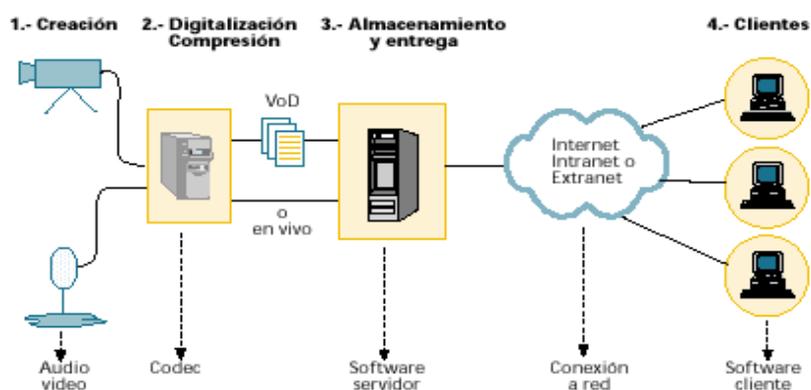


Fig. 1. Instalación de un servicio de video streaming

BitTorrent es un cliente Peer to Peer rapidísimo, con una interfaz que ha mejorado mucho desde la última versión y múltiples opciones que justifican su posicionamiento entre los mayores. Bit Torrent es el cliente oficial de la red con la que comparte nombre. En su instalación incluye una toolbar con buscador incorporado, información meteorológica, acceso a redes sociales, correo, YouTube, la música de Grooveshark integrada y gran variedad de juegos. La caja de búsqueda integrada accede a tu navegador para obtener los feeds que necesites. La información de progreso de la descarga es impresionantemente detallada: peers o usuarios, que se comparten con MuTorrent, ratings, información de avance y velocidad de descarga en gráficos actualizados al Segundo [5].

La configuración es al detalle, muy completa, y cuenta con la opción de instalar aplicaciones con las que, por ejemplo, se puede twittear tus Torrents favoritos para compartirlos, así como programar descargas. Hay varias opciones que marcan la diferencia con otros programas P2P, como la posibilidad de acceder remotamente a tus descargas y controlarlas lejos de tu equipo o convertir tus archivos Torrent para hacerlos compatibles con el dispositivo que prefieras. Y además de todo eso, puedes generar en un par de clicks tus propios Torrents para compartirlos con otros así como crear feeds para recibir las últimas actualizaciones y estar siempre al día. Un P2P potente, rápido y muy completo.

El protocolo BitTorrent se diseñó como una red P2P para compartir archivos de forma lícita, muchas imágenes ISO de distribuciones Linux se pueden descargar usando BitTorrent desde ordenadores y desde dispositivos Android, un cliente BitTorrent para intercambiar archivos de forma rápida entre grandes grupos de personas, o para conseguir imágenes de nuestras distros Linux favoritas, es una de las mejores opciones [6].

2 Estructura y Terminología de una red BitTorrent

.torrent: Ficherito que contiene la información necesaria (codificado) sobre el contenido y el tracker

Index (Índice o catálogo): Listado de ficheritos .torrent manejado por una página web

Tracker: Servidor especial que contiene la información necesaria para que los peers se conecten unos con otros, no contiene nada de contenido solo sabe quién tiene el contenido

Pieces (trozos): el contenido está dividido en trozos más pequeños (256 KB, 512 KB, 1MB)

Peers (Pares): Todos los usuarios que están en la red

Leechers: Usuarios que están descargando el archivo pero que todavía no tienen el archivo completo

Seeds (semillas): Usuarios que tienen el archivo completo

Dos fantásticos clientes de Bittorrent

En este artículo vamos hablar sobre dos de los clientes Bittorrent para sistemas operativos Windows más ligeros y que mejores resultado dan a sus usuarios a día de hoy.

Los clientes de los que hablaremos serán uTorrent y qBittorrent, los dos programas ofrecen un bajo consumo de recursos y una gran velocidad de descarga (dependiendo de los pares).

UTorrent

Ciente torrent por excelencia, caracterizado por su bajo consumo de recursos y su gran velocidad, tiene un sencillo panel de configuración. UTorrent puede descargarse desde su página.

Ciente torrent por excelencia, caracterizado por su bajo consumo de recursos y su gran velocidad, tiene un sencillo panel de configuración. La versión de pago protege contra virus, añade la posibilidad de reproducir archivos mientras se descargan, proporciona acceso remoto seguro a uTorrent y posee un convertidor de ar-

Distribución de carga en sistemas globales de streaming de vídeo en directo

chivos. UTorrent posee cliente para Windows y MacOS, y una especie de servidor para Linux, pero muy difícil de utilizar, y en Linux con las alternativas que existen no merece la pena liarse con este programa, una vez descargado el instalador para Windows, su instalación es similar a la de cualquier otro programa .

Para añadir un torrent podemos utilizar un archivo .torrent, o un magnet link ya que uTorrent asocia automáticamente el programa a los archivos .torrent para que, dando doble click, se abran automáticamente con este programa. Cuando abrimos un archivo, lo primero que nos aparece es una ventana con las opciones principales para guardar el torrent, como los archivos que posee dicho torrent (por si no queremos descargar todos) o la ubicación donde lo guardamos, también podemos importar un magnet link desde archivo>añadir torrent desde enlace.

Principales características:

Archivos: Posee información de los archivos que contiene el archivo descargar.

Información: aquí podemos ver la información sobre el proceso de descarga.

Pares: Datos sobre las conexiones.

Valoraciones: Opiniones de los usuarios sobre el archivo, si es un fake o es bueno.

Velocidad: Una gráfica con la velocidad de descarga del archivo.

Registro

Un log con los acontecimientos ocurridos durante la descarga del torrent. El resto de opciones son para configuraciones avanzadas, y es recomendable no modificarlas ya que pueden empeorar el funcionamiento de uTorrent.

QBittorrent

Es un programa p2p, totalmente gratis no incluye publicidad lo que hace más rápida la descarga. QBittorrent es un cliente de bittorrent libre, se puede instalar en todos los sistemas operativos importantes como: Windows, OS X, Free BSD, OS/2 y las distribuciones de Linux, también tiene a disposición de todos sus códigos fuente, se caracteriza por su facilidad de uso, así como por consumir pocos recursos.

Podemos descargar qBittorrent desde su web oficial para nuestro sistema operativo, la instalación en Windows es como cualquier otra, hacemos click sobre Next y sobre Install, y comienza la instalación, tal como hace uTorrent. Este programa también asocia directamente los archivos .torrent a él, para ejecutarlos con un doble click.

Tras ver ambas aplicaciones, podemos decir que uTorrent tiene más opciones de configuración mientras que qBittorrent tiene mayor sencillez y rendimiento.

Distribución de carga en sistemas globales de streaming de vídeo en directo

Mecánica del funcionamiento

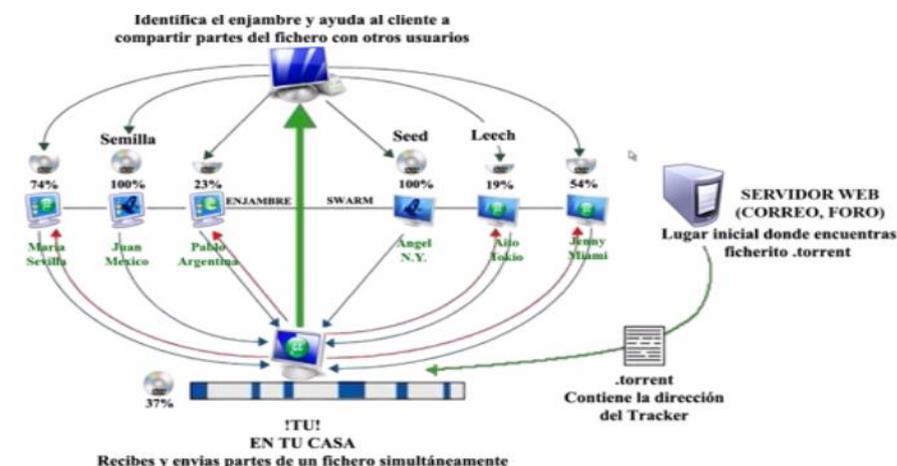


Fig. 2 Mecánica de funcionamiento

El punto de partida es el fichero .torrent, se encuentra en un servidor web, índice, en un foro, incluso se podría enviar por correo, el tamaño es minúsculo. Una vez que tenemos el fichero .torrent, necesitamos un cliente que sepa interpretar la información del fichero, utilizamos uTorrent este se pone en contacto con el tracker.

3 Conclusiones

Las CDN transportan una parte significativa del tráfico de Internet del mundo están en todas partes y mitigan los desafíos más complejos asociados a la distribución de contenido a través de Internet por todo el mundo, desde pequeños y medianos proveedores de contenido hasta empresas globales de mayor tamaño, confía en las CDN para ofrecer una experiencia web impecable a sus usuarios. Las CDN se han convertido en una herramienta esencial para realizar satisfactoriamente actividades empresariales online, principalmente, por una razón: en sus orígenes, Internet no se diseñó para procesar todas las impresionantes tareas que asume en la actualidad, no se desarrolló para gestionar las demandas masivas de datos, vídeo en directo de alta definición, campañas de ventas relámpago y grandes descargas que esperamos hoy en día. Las CDN se han creado específicamente para que Internet funcione mejor, distribuya contenido multimedia a escala y haga posible todas las experiencias conectadas imaginables.

Streaming ofrece grandes ventajas en la difusión y reproducción de contenido multimedia, sea en vivo o bajo demanda, cada petición realizada al servidor no nece-