Algoritmos de identificación de biclústeres con evolución coherente en microarreglos de ADN*[1]

[Algorithms for discover coherent evolution biclusters on DNA microarrays]

LYDA PEÑA PAZ[2]

Recibo: 20.08.2014 – Aprobación: 04.09.2015

Resumen

El Biclustering es una técnica empleada para el análisis de microarreglos de ADN, con el objetivo de identificar subgrupos de genes y de condiciones que muestren patrones similares de comportamiento. Existen diferentes tipos de biclústeres, siendo los de evolución coherente uno de los más difíciles de identificar debido a que no se consideran los valores exactos de los niveles de expresión sino el sentido en que se mueven. En este trabajo se presenta inicialmente un resumen de los algoritmos utilizados en la identificación de biclústeres con evolución coherente, seguido del análisis de estos y las propuestas para el desarrollo de nuevos algoritmos.

Palabras Claves: *Bioinformática, Microarreglos ADN, Biclustering.*

Abstract

Biclustering is an important technique to microarray data analysis. Biclustering aims to identify a subset of genes that are co-regulated under a subset of conditions. It is possible to identify different types

* Modelo para la citación de este artículo:
PEÑA PAZ, Lyda (2015). Algoritmos de identificación de biclústeres con evolución coherente en microarreglos de ADN. En: Ventana Informática No. 33 (jul-dic). Manizales (Colombia): Facultad de Ciencias e Ingeniería, Universidad de Manizales. p. 11-24. ISSN: 0123-9678

1 Artículo de reflexión proveniente de la tesis *Algoritmo Basado en Colonias de Hormigas para Biclustering de Microarreglos de ADN*, para optar el título de Doctor en Ingeniería Informática en la Universidad Pontificia de Salamanca, campus Madrid, bajo la dirección del Dr. Jesús Alfonso López Sotelo.

2 Ingeniera de Sistemas, Magister en Ciencias Computacionales, PhD(c) en Ingeniería Informática. Docente de Planta, Facultad de Ingeniería, Universidad Autónoma de Occidente (Cali, Valle, Colombia). Correo electrónico: lpena@uao.edu.co

of biclusters in microarrays specifically, constant values, coherent values and coherent evolution. Coherent evolution bicluster is a particular type of bicluster that considers the variability of the expression level but not the exact value, so that the process for finding this kind of bicluster is more difficult because there is not a mathematical model that explains the values of expression levels included in the bicluster. This paper is a review of the different algorithm proposed to do this task and some ideas for new developments are proposed.

Keywords*: Bioinformatics, DNA microarrays, Biclustering.*

Introducción

De acuerdo con la definición propuesta por López, Ibarrola & Martín (2000), los microarreglos (*microarrays)* o *biochips* de ADN, son matrices con miles de espacios microscópicos, cada uno de los cuales es llenado con trozos de una secuencia específica de ADN; consiguiendo así, una elevada densidad de integración de un material biológico inmovilizado sobre una superficie sólida, la cual puede ser empleada para la obtención de información biológica relevante. La bioinformática, junto con la aplicación de algoritmos, surgió por la necesidad de estudiar la cantidad masiva de información biológica que genera, la utilización de microarreglos de ADN para el desarrollo de experimentos ha abierto muchas posibilidades por cuanto permite generar y analizar gran cantidad de datos de forma simultánea. Por esta razón la bioinformática ha tomado un papel protagónico en este campo, debido a la necesidad de generar nuevos algoritmos que permitan realizar ese análisis de datos de una forma eficiente.

Una de las herramientas empleada con mayor frecuencia para el análisis de los datos obtenidos en este tipo de ensayos es el *clustering*, el cual tiene como objetivo formar grupos o conjuntos (*clústeres*) cuyos elementos compartan ciertas características que a su vez los diferencien de los elementos de otros conjuntos. Esta técnica de clasificación no supervisada, implica que los conjuntos de datos no se conocen previamente, haciendo más interesante y difícil esta labor, según indican Rodríguez, Riquelme & Aguilar (2006, 21). Su objetivo aplicado a microarreglos, de acuerdo con Eisen et al. (1998, 14863), es agrupar genes con patrones de expresión similares (co-expresados), promoviendo de esta forma, la comprensión de funcionalidades desconocidas de ciertos genes. Este análisis, según Jiang, Tang & Zhang (2004, 5), se puede realizar en dos formas: *clustering* basado en genes, en el cual se agrupan genes co-expresados de acuerdo con

sus patrones de expresión o *clustering* basado en condiciones, para el cual, se dividen las condiciones en grupos homogéneos, donde cada grupo puede corresponder a algún fenotipo particular como síndromes clínicos o tipos de cáncer.

Si bien el *clustering* resulta muy útil para identificar cierto tipo de patrones en los microarreglos, se encuentra limitado debido a que hace la búsqueda en una sola dimensión[3]; es por ello que surge la propuesta del *Biclustering,* en el cual, se toma en consideración las dos dimensiones, buscando identificar subconjuntos de genes que se regulan de forma similar ante un subconjunto de condiciones.

Uno de los primeros trabajos reconocidos sobre *Biclustering*, fue el de Cheng & Church (2000, 2-6) quienes marcaron la pauta para el desarrollo de diversas propuestas en el tema, convirtiéndose en referente obligado para quienes trabajan en el área. A partir de esta propuesta inicial, se han desarrollado diversos algoritmos para realizar la búsqueda de *biclústeres* más significativos de forma eficiente.

Una forma de clasificar los *biclústeres* es por la relación que guardan los valores de expresión que los conforman, de allí que se hable de *biclústeres* de valores constantes, valores coherentes o evolución coherente. De estos tipos, los más complejos de identificar son aquellos con evolución coherente, ya que buscan un subconjunto de genes que se regulen en el mismo sentido (*up o down*) dentro de un subconjunto de condiciones, para lo cual es irrelevante el valor exacto del nivel de expresión.

En este artículo se presenta una revisión y propuestas alrededor del tema de *biclústeres* con evolución coherente. Inicialmente se explican los términos asociados y se define el problema de la búsqueda de estos, para posteriormente hacer una descripción de los algoritmos existentes y finalizar con un análisis comparativo y propuestas de trabajo en el tema.

1. Biclustering de microarreglos de ADN

1.1 Los microarreglos de ADN

Los microarreglos (*microarrays*) de ADN, también conocidos como biochips, gene chips, DNA chips o *genearray*, son soportes sólidos (láminas de vidrio para microscopio, láminas de silicona o membranas

3 Es decir, la agrupación de genes está basada en toda la dimensión de las condiciones mientras que la agrupación de condiciones se basa en toda la dimensión de los genes. Además, se ha comprobado que en la naturaleza, un subgrupo de genes puede estar co-expresado y co-regulado bajo un conjunto de condiciones pero su comportamiento podría viar bajo otro conjunto distinto.

de nylon) en los que se encuentran inmovilizados en forma ordenada, normalmente en arreglos 80x80. El ADN puede ser impreso, depositado o sintetizado directamente en la superficie sólida. Los ácidos nucleicos en el microarreglo representan todos o parte de los genes de un organismo, ubicando un gen diferente en cada posición del microarreglo. El término microarreglo se emplea comúnmente para designar el dispositivo físico (lámina de vidrio o polímero) al cual se adhieren las moléculas de ADN, pero también para representar el experimento del cual se puede obtener información.

La planificación y desarrollo de experimentos con microarreglos está ligado al tipo de aplicación que se desea realizar y a la tecnología disponible, sin embargo, se distinguen cuatro grandes etapas en estos ensayos, acorde con Biotic (2005): Diseño y fabricación del chip, Preparación de las muestras, Realización de los ensayos sobre la superficie del microarreglo, y Revelado y análisis de resultados. Si bien, existen técnicas y métodos de la bioinformática que pueden ser usados en cada etapa, es en el análisis de resultados donde su aplicabilidad es mayor, debido a que la cantidad y complejidad de los datos obtenidos hace prácticamente imposible su análisis por otros medios.

La gran potencialidad de los microarreglos, señalan Pontes, Giráldez & Aguilar (2013, 2), se debe a su capacidad de crear bases de datos que contengan los niveles de expresión genética de miles de genes simultáneamente, bajo ciertas condiciones experimentales, lo que permite realizar múltiples análisis posteriores, a partir de un único experimento biológico.

Un aspecto importante del análisis de datos provenientes de ensayos con microarreglos, es la capacidad de clasificar los resultados y agruparlos de acuerdo a características comunes, para lo cual se pueden emplear muchos métodos, incluyendo el simple examen visual. Si bien esta agrupación de datos no genera una propuesta definitiva, desde el punto de vista biológico, si constituye una guía para análisis posteriores por parte de los biólogos expertos.

1.2 Biclustering

Este concepto introducido por Hartigan (1972, 125-126), se refiere a una técnica en la que se realiza el *clustering* de genes y condiciones de forma simultánea para realizar la interpretación directa de los resultados. Desde entonces se han propuesto diversos algoritmos aplicados en múltiples campos, uno de los cuales es el análisis de microarreglos.

Kung & Mak (2005, 399) establecen tres aspectos que diferencian las técnicas de *Biclustering* del *clustering* tradicional: la coherencia de los modelos no se conoce previamente, se realiza el *clustering* de genes y

condiciones de forma simultánea y es común que exista solapamiento entre los *biclústeres*, ya que genes con múltiples funciones pueden estar asociados a varios grupos.

De acuerdo con Banka & Mitra (2006, 1) al realizar la clasificación simultánea de genes y condiciones experimentales, el *Biclustering* permite la detección de conjuntos solapados, proporcionando una mejor representación de la realidad biológica que involucra genes con múltiples funciones o aquellos regulados por múltiples factores; de esta forma, es posible descubrir aquellas estructuras locales inherentes a las matrices de expresión de genes.

Se puede definir una matriz de expresión genética como una matriz, donde cada elemento a_{ij} es un valor real que representa el nivel de expresión del gen i bajo la condición experimental j. Si se tiene una matriz A de n*m, con un conjunto de filas $X = \{x_1,\ldots,x_n\}$ y un conjunto de columnas $Y = \{y_1,\ldots,y_m\}$ y además, I y J son subconjuntos de filas y columnas respectivamente, $A_{IJ} = (I,J)$ representa la submatriz de A que contiene los elementos a_{ij} que pertenecen al subconjunto de filas I y al subconjunto de columnas J.

Considerando lo anterior, Madeira & Oliveira (2004, 2-3) definen el problema de los algoritmos de *Biclustering* de la siguiente forma: dada una matriz A se desea identificar un grupo de *biclústeres* $B_k = (I_k,J_k)$, de tal forma que cada *biclúster* B_k satisfaga ciertas características de homogeneidad; la definición de estas características son decisivas en la estructuración del algoritmo.

Aunque la complejidad del problema de *Biclustering* depende de su formulación y más específicamente de la función seleccionada para evaluar la calidad de los *biclústeres* obtenidos, la mayoría de las variantes de este problema tienen complejidad NP-completa, lo que genera un desafío desde el punto de vista algorítmico.

1.3 Tipos de biclústeres

Los *biclústeres* pueden variar en su conformación:

- Con valores constantes: submatrices que tienen los mismos valores de expresión genética para ciertos genes en ciertas condiciones.
- Con valores constantes en filas o columnas: submatrices que mantienen valores constantes para genes bajo cierta condición o para un gen bajo diferentes condiciones.
- Con valores coherentes: para los cuales, sus valores cambian de forma aditiva o multiplicativa a través de las filas y las columnas.
- Con evolución coherente: el problema se enfoca en establecer subconjuntos con datos que reflejen un comportamiento similar, más

que en los valores exactos que presentan; esta evolución puede presentarse en toda la submatriz, en las filas o en las columnas.

1.4 Biclústeres con evolución coherente

Los *biclústeres* con evolución coherente, son subconjuntos de genes que están corregulados a través de un subconjunto de condiciones sin tomar en cuenta su valor exacto, identificando de esta forma, sendas genéticas. Para el algoritmo propuesto, se plantea el problema de identificar *biclústeres*, de la siguiente forma: Sea A una matriz MxN que representa un microarreglo de ADN que tiene M genes (filas) y N condiciones (columnas), cada entrada a_{mn} corresponde al nivel de expresión del gen m bajo la condición n.

El objetivo, es encontrar una submatriz GxT, con $G \subseteq M$, $T \subseteq N$, donde los niveles de expresión de todos los genes en G se mueven en el mismo sentido para todas las condiciones en T. Sea por ejemplo (Figura 1), la matriz A, es posible identificar el *biclúster* compuesto por los genes 2, 3 y 4 y las condiciones 1, 2 y 6, ya que estos conservan la misma tendencia.

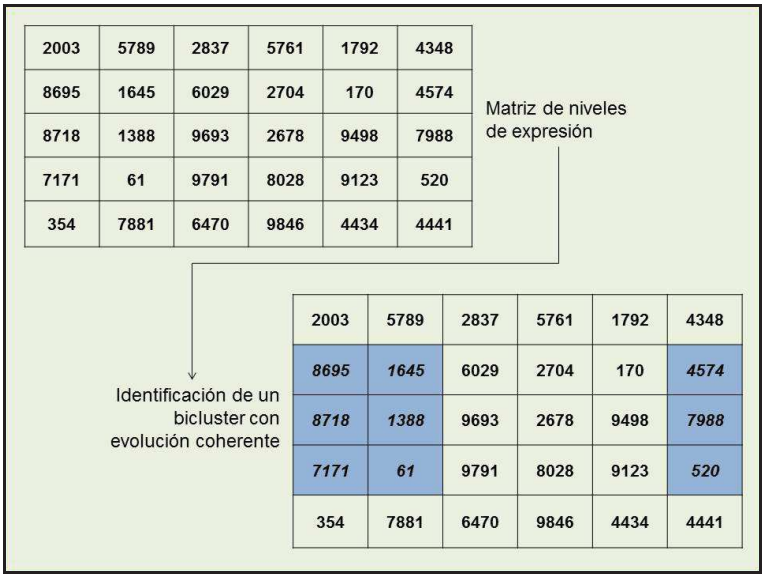

Figura 1. Ejemplo de Identificación de un biclúster con evolución coherente a partir de una matriz de niveles de expresión

El comportamiento de los genes en la matriz original se observa en la Figura 2, en tanto la Figura 3 muestra el *biclúster* identificado; en esta, se puede observar el comportamiento similar del subgrupo de genes bajo el subconjunto de condiciones definidas, aunque no es posible establecer

una escala aditiva o multiplicativa para explicar dicho comportamiento (en cuyo caso, correspondería a un *biclúster* con valores coherentes).

La identificación de biclústeres con evolución coherente permite descubrir sendas genéticas, al identificar genes que se regulan en el mismo sentido para un conjunto específico de condiciones.

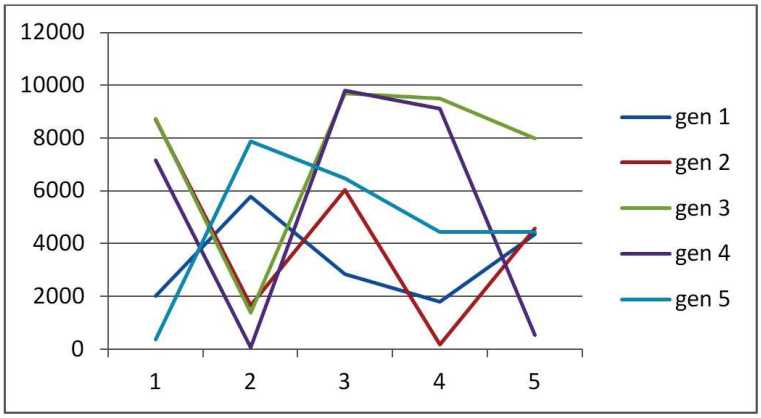

Figura 2. Niveles de expresión génica del microarreglo

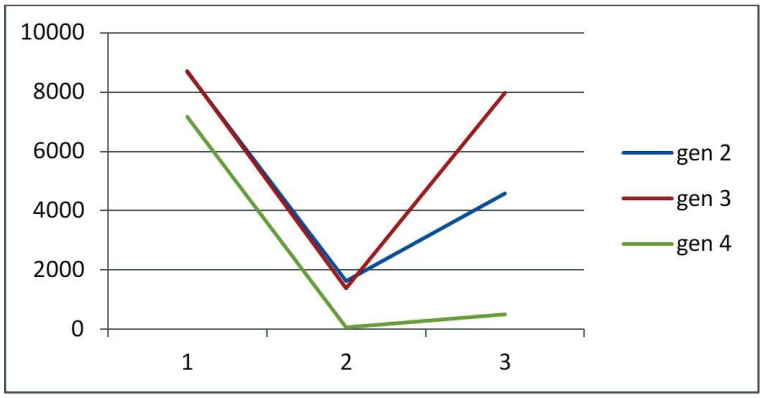

Figura 3. Niveles de expresión génica del biclúster identificado

2. Algoritmos para biclustering con evolución coherente

A lo largo de la historia se han propuesto múltiples algoritmos para la identificación de biclústeres de diferente tipo, por ejemplo *Block Clustering* (Hartigan, 1972) y *Gibbs* (Sheng, Moreau & De Moor, 2013) identifican biclústeres con valores, filas o columnas constantes;

d-biclúster (Cheng & Church, 2000), FLOC (Yang et al. 2005), *pCluste*r (Wang et al. 2002), *Plaid Models* (Lazzeroni & Owen 2002), PRMs (Segal et al. 2001) y *Spectra*l (Kluger et al. 2003) identifican biclústeres con valores coherentes, pero son pocos los algoritmos que tienen como objetivo identificar biclústeres con evolución coherente.

Como se ha mencionado, en los *Biclústeres* con evolución coherente un grupo de genes se encuentra corregulados para un conjunto de condiciones específicas, para lo cual no se toman en cuenta los valores exactos de los niveles de expresión de los genes; en consecuencia las medidas de distancia tradicionalmente empleadas para hallar *biclústeres*, como la distancia Euclídea o Manhattan, no pueden ser usadas. Lo anterior ha llevado a generar algoritmos basados en técnicas no tradicionales para el descubrimiento de esta clase de *Biclústeres*.

2.1 SAMBA (Statistical Algorithmic Method for bicluster analysis)

Consiste en una combinación de la teoría de grafos y el modelo de datos estadístico, propuesta por Tanay, Sharan & Shamir (2002, 137-141), donde se modela la matriz de expresión como un grafo bipartito cuyas partes corresponden a las condiciones y genes, con aristas para los cambios significativos de los niveles de expresión; el objetivo del algoritmo se expresa como la identificación del *subgrafo* de mayor peso, asumiendo que su peso corresponde a su significancia estadística.

Mientras en la primera fase del algoritmo se realiza una normalización de los datos, donde se considera que un gen está regulado hacia arriba o hacia abajo si su nivel de expresión normalizado (con media 0 y varianza 1) es superior a 1 o inferior a -1, en la segunda se aplican técnicas de *hashing* para encontrar los k biciclos más pesados[4] en el grafo que intersectan cada condición o gen.

Por su parte, en la tercera fase del algoritmo se realiza un procedimiento de mejora local sobre los *biclústeres* de cada montículo, donde iterativamente se adiciona o borra un vértice del *biclúster* hasta que no sea posible mejorar la puntuación y se filtran los *biclústers* similares[5].

2.2 xMotif

Murali & Kasif (2003, 82-83) plantean una propuesta para la agrupación de genes que presentan comportamiento similar en conjuntos llamados *xMotif*. Un motivo de expresión de genes conservados o *xMotif* es

4 Es de anotar que para hacerlo más eficiente, se ignoran los genes con un grado mayor a un umbral indicado.

5 Dos *biclústeres* se consideran similares si sus conjuntos de vértices difieren ligeramente, es decir, que el número de condiciones y genes compartidos es superior a un umbral definido.

un subconjunto de genes cuyos niveles de expresión se conservan simultáneamente para un conjunto de condiciones, en cuyo caso se dice que esas muestras encajan en el motivo. Con el fin de evitar que ellos resulten ser demasiado específicos o restrictivos, los autores definen unas características para los motivos que se van a identificar: siendo un Motif un par (C,G) donde C es un subconjunto de condiciones y G un subconjunto de genes, (1) el número de condiciones en C debe ser al menos un α-fracción de todos los ejemplos **(α**> 0), (2) Cada gen en G se conserva a través de todas las condiciones en C y (3) cada gen que no se encuentre en G, está conservado en al menos una β-fracción de todas las condiciones en C (β <1).

Si bien el algoritmo puede encontrar diferentes *xMotif*, se selecciona como respuesta aquel que contiene el mayor número de filas conservadas, para lo cual se emplea el número de filas como función de mérito.

2.3 OPSM (Submatriz de Orden Preservado)

Algoritmo propuesto por Ben-Dor et al. (2003, 2-6) para encontrar un modelo completo máximamente soportado, es decir, un conjunto de columnas con un orden lineal soportado por un número máximo de filas. Los autores definen un modelo completo como el par (J,π), donde *J* es un conjunto de s columnas y $\pi = (j_1, j_2, \ldots, j_s)$ es el ordenamiento lineal de las columnas de *J*. Se dice entonces, que una fila soporta el modelo si los *s* valores correspondientes ordenados de acuerdo a la permutación π son monótonamente crecientes.

Para determinar la calidad de un modelo, se determina la significancia estadística del mismo, mediante la determinación del límite superior de la probabilidad de que una matriz de tamaño n x m contenga un modelo completo de tamaño *s* con *k* o más filas que lo soporten.

2.4 OP-cluster (Order Preserving Cluster)

Este algoritmo presentado por Liu & Wang (2003, 3-6) busca *biclústeres* con orden preservado a partir del cual es posible capturar la tendencia general de los objetos a través de un subconjunto de dimensiones en un espacio multidimensional, en una secuencia de dos pasos:

en el primero se realiza un preprocesamiento de las filas de la matriz, para lo que se ordena de forma ascendente cada fila de acuerdo a sus valores de entrada, posteriormente se agrupan aquellos valores que están cercanos (un valor diferencial menor a un límite establecido), generando para cada fila una secuencia con las etiquetas correspondientes a cada columna, arrojando ara el paso siguiente se emplearán solamente las secuencias de etiquetas y no los valores exactos.

en el segundo se ejecuta minería sobre los conjuntos de filas, con la secuencia de etiquetas (no con los valores exactos), para descubrir todas las subsecuencias frecuentes que se encuentran ocultas en las secuencias generadas. Aunque puede parecer similar al descubrimiento de patrones secuenciales, difiere porque la identificación de las filas asociadas a cada secuencia debe ser recordada para determinar las filas que constituyen un *OP-clúster*, mientras cada columna (o etiqueta) aparece exactamente una vez en cada secuencia.

El algoritmo emplea una estructura de árbol compacta para almacenar la información crucial usada durante la minería de *OP-clústeres*, el descubrimiento de subsecuencias frecuentes y la asociación de filas con estas subsecuencias ocurren simultáneamente.

2.5 Roba (Robust Biclustering Algoritm)

Tewfik, Tchagang & Vertatschitsch (2006, 2409-2414) proponen un algoritmo que se diferencia de propuestas anteriores en tres aspectos: (1) permite identificar todos los *biclústeres* perfectos con evolución coherente que contienen un número de condiciones definidas, (2) usa algebra lineal y herramientas de aritmética evitando la utilización de funciones heurísticas de costo y (3) tiene una complejidad menor que técnicas de *Biclustering* propuestas anteriormente.

El algoritmo incluye dos pasos:

- en la primera parte se realiza el preprocesamiento de datos, empleando rutinas conocidas para manejar los datos faltantes y con ruido que se encuentren en el microarreglo,
- en la segunda parte se hace la identificación de los *biclústeres*, para lo cual se realizan dos subprocesos. Inicialmente se enumeran todas las posibles combinaciones que incluyen k condiciones, donde $K \geq K_{min}$, siendo K_{min} el número mínimo de condiciones definido para un *biclúster* **válido.** Para cada subconjunto de K condiciones, se emplea un procedimiento de ordenamiento por fila. Al finalizar se tiene una matriz que contiene el ranking de las K condiciones para cada fila, cuando el nivel de expresión de las condiciones para el gen se ha ordenado de manera no decreciente.

El segundo subproceso toma como entrada la matriz del paso anterior y a través de un procedimiento de clasificación rápido de filas, identifica los patrones de evolución coherente válidos que involucran todos los genes y un conjunto de K condiciones de forma simultánea. Un paso final de poda elimina los *biclústeres* que están completamente incluidos en otros más grandes.

3. Análisis comparativo

La búsqueda de *biclústeres* en microarreglos de ADN es una labor compleja si se consideran las dimensiones que pueden tener estar matrices y por consiguiente la cantidad de datos a valorar. Dentro de los diferentes tipos de *biclústeres* que pueden llegar a identificarse, aquellos con evolución coherente resultan tener una mayor complejidad por cuanto se busca un subconjunto de genes que varíen de la misma forma ante un subconjunto de condiciones, sin considerar los valores específicos de los niveles de expresión.

La condición antes descrita conlleva a que no sea posible definir un modelo matemático que represente los niveles de expresión del *biclúster* y que por tanto deban emplearse otras aproximaciones para su búsqueda. Los algoritmos que se han propuesto hasta el momento emplean diferentes enfoques y diferentes funciones de mérito para determinar la validez de un *biclúster* respecto a otro; esta diversidad hace que no sea posible una comparación directa de los resultados obtenidos por los diferentes algoritmos.

La mayoría de los algoritmos propuestos realizan la búsqueda de n *biclústeres* de forma simultánea, solamente en el caso de OPSM retorna el *biclúster* que corresponde al modelo máximamente soportado. El enfoque algorítmico empleado, corresponde a Búsqueda Voraz o Enumeración Exhaustiva, lo cual implica gran cantidad de procesamiento, aunque las diferentes aproximaciones emplean estrategias que permiten disminuir el espacio de búsqueda.

Tres de los algoritmos revisados (SAMBA, OP-clúster, ROBA) incluyen una fase de preprocesamiento de los datos, donde se hace limpieza de datos (eliminando datos faltantes y ruido) o algún tipo de normalización necesaria para el proceso posterior.

En general, los algoritmos requieren la definición de unos parámetros iniciales que podrían hacer variar las respuestas obtenidas o el tiempo de ejecución del algoritmo; estos parámetros incluyen número mínimo o máximo de filas o columnas en el *biclúster*, número mínimo de *biclústeres* a buscar o umbrales para determinar la similitud de *biclústeres* al momento de realizar el refinamiento de la solución.

3.1 Propuesta de trabajo futuro

Ha sido probado que la búsqueda de *biclústeres* con evolución coherente no puede abordarse mediante algoritmos secuenciales siendo necesario emplear otro tipo de enfoques. Para evitar realizar una enumeración exhaustiva o un algoritmo voraz, se pueden esquemas de búsqueda colaborativa que permitan la reducción de tiempo de ejecución la vez que

se alcanzan soluciones más significativas. En este sentido se propone la utilización de técnicas de inteligencia artificial, específicamente inteligencia de enjambres, en la cual un grupo de agentes trabajan de manera colaborativa para encontrar soluciones apropiadas en un espacio multidimensional.

Debido a que las técnicas de inteligencia de enjambres, tales como colonias de hormigas han sido empleadas previamente para realizar *clustering* y algunos tipos de *Biclustering*, es posible generar una propuesta que trabaje en la búsqueda de *biclústeres* con evolución coherente, en este caso es importante establecer una función de mérito apropiada para verificar la validez de las soluciones halladas.

Por otra parte, la dificultad para medir los resultados obtenidos de una forma objetiva hace difícil la labor de comparación entre los diferentes algoritmos, por lo cual cobra importancia la definición de una métrica que determine la validez de un *biclúster* de evolución coherente de forma independiente al algoritmo con el que fue desarrollado.

4. Conclusiones

La identificación de *biclústeres* con evolución coherente es una tarea de gran utilidad para la tipificación de posibles caminos genéticos, si bien no se pueden obtener resultados concluyentes, es posible generar una aproximación para que los biólogos expertos realicen una búsqueda más específica.

Debido a la dificultad que implica hallar *biclústeres* con evolución coherente, se han empleado diferentes aproximaciones desde los algoritmos de enumeración exhaustiva y búsqueda voraz, debido a que los elementos incluidos en este tipo particular de *biclústeres* no tienen un comportamiento que pueda ser explicado mediante un modelo matemático.

Una aproximación posible para el problema de la búsqueda de *biclústeres* con evolución coherente es la inteligencia de enjambres, considerando que su enfoque propone la utilización de múltiples agentes que trabajan colaborativamente para la solución de un problema, lo que permitiría una exploración más exhaustiva y rápida del espacio de solución.

La comparación de los algoritmos en términos la efectividad de las soluciones logradas es difícil, por cuanto emplean diferentes enfoques y funciones objetivo, es importante poder establecer una métrica capaz de determinar de forma objetiva la bondad de los *biclústeres* hallados.

5. Referencias bibliográficas

BANKA, H. & MITRA, S. (2006). Evolutionary Biclustering of gene expressions [online]. In: Ubiquity, Vol. 7, No. 42 (oct). New York (NY, USA): ACM. Art. No. 5, p. 1-12. e-ISSN: 1530-2180 <http://ubiquity.acm.org/article.cfm?id=1183082> <doi:10.1145/1183081.1183082> [consult: 31/10/2014] http://ubiquity.acm.org/article.cfm?id=1183082

BEN-DOR, A.; CHOR, B.; KARP, R. & YAKHINI, Z. (2003). Discovering local structure in gene expression data: the order-preserving submatrix problem [online]. In: Journal of Computational Biology : A Journal of Computational Molecular Cell Biology, Vol. 10, No. 3-4, New York (NY, USA): Thomson Reuters p. 373-384. e- ISSN: 1557-8666 <doi:10.1089/10665270360688075> <http://www.cs.technion.ac.il/labs/cbl/publications/order_preserving.pdf> [consult: 15/03/2015]

BIOTIC (2005). Metodología: Técnicas [en línea]. Majadahonda (Madrid, España): Instituto de Salud Carlos III <http://infobiochip.isciii.es/marcos%20navegacion.htm> [consult: 13/04/2015]

CHENG, Y. & CHURCH, G.M. (2000). Biclustering of expression data [online]. In: Eighth International Conference on Intelligent Systems for Molecular Biology, ISMB 2000 (19-23/08/2000), San Diego (CA, USA): International Society for Computational Biology. BOURNE, P.; GRIBSKOV, M.; ALTMAN, R.; JENSEN, N.; HOPE, D.; LENGAUER, T.; MITCHELL, J.; SCHEEFF, E.; SMITH, C.; STRANDE, S. & WEISSIG, H. (ed.) (2000). Proceedings of the ISMB 2000. Palo Alto (CA, USA): The Association for the Advancement of Artificial Intelligence (AAAI). p. 93-103. ISBN: 978-1-57735-115-3. Retrieved from <ftp://ftp.sdsc.edu/pub/sdsc/biology/ISMB00/157.pdf> <https://www.aaai.org/Papers/ISMB/2000/ISMB00-010.pdf> [consult: 11/05/2015]

EISEN, M.B.; SPELLMAN, P.T.; BROWN, P.O. & BOTSTEIN, D. (1998). Cluster analysis and display of genome-wide expression patterns [online]. In: Proceedings of the National Academy of Sciences of the United States of America, PNAS, Vol. 95, No. 25 (dic). Washington, DC (USA): National Academy of Sciences. p. 14863–14868. e-ISSN: 1091-6490 <http://www.pnas.org/content/95/25/14863.full.pdf> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=24541&tool=pmcentrez&rendertype=abstract> [consult: 05/08/2015]

HARTIGAN, J. (1972). Direct clustering of a data matrix [online]. In: Journal of the American Statistical Association, Vol 67, No. 337 (mar). Alexandria (VA, USA): American Statistical Association. p. 123-129. e-ISSN: 1537-274X. <http://amstat.tandfonline.com/doi/full/10.1080/01621459.1972.10481214> <DOI:10.1080/01621459.1972.10481214> [consult: 01/04/2015]

JIANG, D.; TANG, C. & ZHANG, A. (2004). Cluster analysis for gene expression data: a survey [online]. In: IEEE Transactions on Knowledg and Data Engineering, Vol. 16, No. 11 (nov). Washington DC (USA): IEEE Computer Society. p. 1370-1386. ISSN: 1041-4347 <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1339264> [consult: 03/01/2014]

KLUGER, Y.; BASRI, R.; CHANG, J.T. & GERSTEIN, M. (2003). Spectral Biclustering of microarray data: coclustering genes and conditions [online]. In: Genome research, Vol. 13, No. 4 (apr). Cold Spring Harbor (NY, USA): Cold Spring Harbor Laboratory Press. p. 703-716. e-ISSN: 1549-5469 <http://genome.cshlp.org/content/13/4/703.short> <doi:10.1101/gr.648603> [consult: 01/04/2015].

KUNG, S. & MAK, M.W. (2005). A machine learning approach to dna microarray Biclustering analysis [online]. In: International Workshop on Machine Learning for Signal Processing, 2005 (28-29/09/2005). Mystic (CT, USA): IEEE. Proceedings, Piscataway (NJ, USA): IEEE. p. 399-404. ISBN: 0-7803-9517-4. <10.1109/MLSP.2005.1532936> <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1532936> [consult: 31/10/2013].

LAZZERONI, L. & OWEN, A. (2002). Plaid models for gene expression data [online]. In: Statistica Sinica, Vol. 12, No. 1 (jan). Taipei City (Taiwan, USA): Institute of Statistical Science, Academia Sinica & International Chinese Statistical Association. p. 61-86 <http://www3.stat.sinica.edu.tw/statistica/oldpdf/A12n14.pdf> [consult: 01/11/2013]

LIU, J. & WANG, W. (2003). Op-cluster: Clustering by tendency in high dimensional space [online]. In: Third IEEE International Conference on Data Mining, ICDM 2003 (19-22/11/2003). Melbourne (FL, USA): IEEE. WU, X. & TUZHILIN, A. (ed.) (2003). Proceedings of the ICDM-2003. Piscataway (NJ, USA): IEEE. p. 187-194. ISBN: 0-7695-1978-4 <doi: 10.1109/ICDM.2003.1250919> <http://www.cs.ucla.edu/~weiwang/paper/ ICDM03_1.pdf> [consult: 11/10/2014]

LÓPEZ-CAMPOS, G.; IBARROLA, N. & MARTÍN-SÁNCHEZ, F. (2000). Biochips y Bioinformática [en línea]. En: Informática y Salud, I+S, No. 25 (mar-abr). Madrid (España): Sociedad Española de Informática de la Salud. ISSN: 1579-8070 <http://www.seis.es/seis/i_s/i_s25/i_s25_1.htm> <http://82.98.165.8/seis/i_s/i_s25/i_s25_1.htm> [consult: 13/01/2014]

MADEIRA, S.C. & OLIVEIRA, A.L. (2004). Biclustering algorithms for biological data analysis: a survey [online]. In: IEEE/ACM Transactions on computational biology and bioinformatics. Vol. 1, No. 1 (jan-mar). Piscataway (NJ, USA): IEEE Computational Intelligence Society. p. 24-45. ISSN: 1545-5963 IEEE/ACM. <http://www.ncbi.nlm.nih.gov/ pubmed/17048406> [consult: 25/10/2013]. <doi: 10.1109/TCBB.2004.2> [consult: 23/02/2015]

MURALI, T.M. & KASIF, S. (2003). Extracting conserved gene expression motifs from gene expression data [online]. In: Eighth Pacific Symposium on Biocomputing, PSB 2003 (03-07/01/2003), Lihue (HI, USA): International Society for Computational Biology. Proccedings of the PSB 2003. p. 77-88. <http://psb.stanford.edu/psb-online/proceedings/psb03/murali.pdf> [consult: 11/10/2014]

PONTES, B.; GIRÁLDEZ, R. & AGUILAR-RUIZ, J.S. (2013). Configurable pattern-based evolutionary Biclustering of gene expression data [online]. In: Algorithms for Molecular Biology : AMB, Vol. 8, No. 1 (feb), Bethesda (MD, USA). p. 4. <doi:10.1186/1748-7188-8-4> <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3668234/> [consult: 20/10/2014]

RODRÍGUEZ BAENA, D.S.; RIQUELME SANTOS, J.C. & AGUILAR RUIZ, J.S. (2006). Análisis de datos de Expresión Genética mediante técnicas de Biclustering [en línea]. Memoria del periodo de investigación [DEA, Tecnología e Ingeniería del Software, con mención de calidad], Sevilla (España): Universidad de Sevilla, Departamento de Lenguajes y Sistemas Informáticos. <https://www.lsi.us.es/docs/doctorado/memorias/Memoria-v2.pdf> [consulta: 13/01/2014]

SEGAL, E.; TASKAR, B.; GASCH, A.; FRIEDMAN, N. & KOLLER, D. (2001). Rich probabilistic models for gene expression [online]. In: Bioinformatics, Vol. 17, Suppl. 1 (jun). Oxford (England): Oxford University Press. p. S243-252. e-ISSN: 1460-2059 <http://bioinformatics.oxfordjournals.org/content/17/suppl_1/S243.full.pdf+html> [consult: 10/07/2015]

SHENG, Q.; MOREAU, Y. & DE MOOR, B. (2003). Biclustering microarray data by Gibbs sampling. [online]. In: Bioinformatics, Vol. 19, Suppl 2 (oct), Oxford (England): Oxford University Press. p. ii196-205. e-ISSN: 1460-2059. <http://www.ncbi.nlm.nih.gov/pubmed/14534190> <http://bioinformatics.oxfordjournals.org/content/19/suppl_2/ii196.full.pdf+html> [consult: 10/08/2015].

TANAY, A.; SHARAN, R. & SHAMIR, R. (2002). Discovering statistically significant biclústeres in gene expression data [online]. In: Bioinformatics, Vol.18, Suppl. 1 (oct). Oxford (England): Oxford University Press. p. S136-144. e-ISSN: 1460-2059 <http://www.ncbi.nlm.nih.gov/pubmed/12169541> <http://bioinformatics.oxfordjournals.org/content/18/suppl_1/S136.full.pdf+html> [consult: 15/01/2014]

TEWFIK, A.H.; TCHAGANG, A.B. & VERTATSCHITSCH, L. (2006). Parallel identification of gene biclústeres with coherent evolutions [online]. In: IEEE Transactions on Signal Processing, Vol. 54, No. 6 (jun). Piscataway (NJ, USA): IEEE. p. 2408-2417. ISSN: 1053-587X <doi: 10.1109/TSP.2006.873720> <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber =1634843> [consult: 15/04/2014]

WANG, H.; WANG, W.; YANG, J. & YU, P.S. (2002). Clustering by pattern similarity in large data sets [online]. In: ACM SIGMOD international conference on Management of data, SIGMOD '02 (02-07/06/2002). Madison (WI, USA): ACM. Proceedings of the SIGMOD '02, New York (NY, USA): ACM. p. 394-405 ISBN: 1-58113-497-5 <doi:10.1145/564691.564737> <http://portal.acm.org/citation.cfm?id=564691.564737> [consult: 13/03/2015]

YANG J.; WANG, H.; WANG, W. & YU, P.S. (2005) An improved biclustering method for analyzing gene expression profiles [online]. In: International Journal on Artificial Intelligence Tools, Vol. 14, No. 05 (oct). London (UK): World Scientific Publishing Company. p. 771-789. e-ISSN: 1793-6349 <http://www.worldscientific.com/doi/abs/10.1142/S0218213005002387?src=recsys&> <doi:10.1142/s0218213005002387> [consult: 04/04/2015]