

Nueva Metodología Para Clasificar Datos de Series Temporales Usando el Algoritmo Biclustering

Myladis Cogollo Flórez
Departamento de Ciencias Básicas
Universidad EAFIT
Medellín, Colombia
E-mail: mcogollo@eafit.edu.co

Alejandro Palacio Vásquez
Departamento de Ciencias Básicas
Universidad EAFIT
Medellín, Colombia
E-mail: apalac19@eafit.edu.co

Resumen—En este trabajo se propone una metodología de clasificación de series temporales usando la técnica de extracción de características y el algoritmo de clasificación Biclustering, la cual permite trabajar con series temporales de distintas longitudes y con datos faltantes. Los resultados obtenidos con esta metodología son mejores al ser comparados con los obtenidos por medio del método tradicional K -medias usando datos experimentales.

Index Terms—Series de tiempo, Clustering, Biclustering.

I. INTRODUCCIÓN

La minería de datos de series temporales se ha convertido en un tema de interés en años recientes [1]. En algunas investigaciones se tiene un conjunto de datos de series temporales (las cuales pueden tener distintas longitudes), y el interés radica en clasificar dichas series temporales en grupos teniendo en cuenta que las series pertenecientes a un mismo grupo deben tener el mismo patrón de comportamiento y/o características. Para comprender las diferencias en los procesos estimados, sin tener que estudiar y comparar las estimaciones de los parámetros de todas las series (véase por ejemplo [2], [3], [4] entre otros). Este hecho es de particular interés en el control de procesos y el reconocimiento de patrones [5], y en medicina y astronomía [6]. Liao [7] realiza una revisión de la literatura de la clasificación de datos de series temporales. Él muestra la importancia de la clasificación de las series temporales en distintas área de aplicación (véase en la tabla 1 el total de aplicaciones referenciadas por el autor).

Área de aplicación	Total estudios
Negocios y socio-economía	7
Ingeniería	5
Ciencia	4
Medicina	2
Arte y entretenimiento	1

Tabla I: Estudios de aplicación de la clasificación de series temporales hasta el año 2005.

Usualmente, el problema de la clasificación de las series temporales se aborda empleando técnicas de análisis de conglomerados [8], también conocido como algoritmo de agrupamiento o simplemente como “*clustering*”. El cual consiste en agrupar un conjunto de objetos, con base en sus características, de tal modo que los objetos resultantes en el mismo grupo (llamado “*cluster*”) sean más similares entre sí

que con los objetos resultantes de los demás grupos o clusters. Sin embargo, los métodos de conglomerados propuestos se basan en que los datos de entrada son estáticos, es decir, de tipo corte transversal y no cambiantes como lo son las series temporales, lo cual hace que ellos no se puedan aplicar directamente a la serie temporal. Por ello varios investigadores han desarrollado algunos métodos de agrupación específicos para el campo de las series temporales con el fin de dar respuesta a dicho problema y de poder darle un uso adecuado al mismo (véase [9], [10] y [11]). En general, se observa que estos métodos se basan en tres grandes esquemas [7]:

1. Propuestas basadas en datos primarios u originales: Estos métodos trabajan con los datos en bruto, ya sea en el tiempo o dominio de la frecuencia. Esto implica trabajar con bases de datos de altas dimensiones. Las series de tiempo son normalmente muestreadas en el mismo intervalo de tiempo y se usan algoritmos de agrupamiento tradicionales como lo son Fuzzy, K-Means y jerárquicos aglomerativos, lo cual hace que se presenten inconvenientes con la elección de la mejor medida de disimilaridad y número de grupos a formar.
2. Propuestas basadas en características extraídas de los datos primarios: En vez de trabajar con los datos primarios, se realiza la clasificación de la serie con base en algunas de sus características. Sin embargo, las características a extraer en las series dependen de la aplicación. Es decir, que un conjunto de características que funcionan bien en una sola aplicación no podrían ser relevantes en otros estudios. Posteriormente, se aplican los métodos tradicionales de agrupamiento, lo cual conlleva a las dificultades establecidas anteriormente. Adicionalmente, estos métodos sólo son aplicables a series con las mismas longitudes.
3. Propuestas basadas en un modelo ajustado a los datos primarios: Las variables de entrada para los algoritmos de agrupamiento son básicamente las estimaciones de los parámetros o coeficientes de los modelos ajustados (tales como los modelos AR, MA o ARMA). Las series temporales son consideradas similares cuando los modelos que caracterizan las series individuales o los residuales del modelo son similares. La dificultad adicional a la de los métodos de agrupamiento tradicional, radica en seleccionar el modelo adecuado para generar

cada serie temporal.

Adicional a las desventajas presentadas para cada uno de los tres esquemas anteriores, se tiene la elección correcta de la medida de distancia entre dos series temporales. Las medidas de distancia más populares, como lo es por ejemplo la de Euclides, solamente se pueden utilizar en los casos que todos las series tenga igual longitud en el tiempo, en caso contrario, se han propuesto diferentes métodos basados en una medida de distancia conocida como distorsión de tiempo dinámico (*DTW* por su sigla en inglés).

En la literatura se encuentra que adicional a los métodos de agrupación tradicionales ha surgido una nueva modalidad de algoritmo de clasificación que permite agrupar no sólo objetos, sino también sus atributos, dicho algoritmo es llamado *Biclustering* [12]. Éste algoritmo ha sido ampliamente estudiado en la literatura y empleado en áreas como microbiología e ingeniería, sin embargo en la literatura se observa que hay pocos estudios relacionados con el uso del algoritmo *Biclustering* en la clasificación de series temporales [4].

En éste trabajo se propone una metodología para clasificar datos de series temporales usando el algoritmo *Biclustering* y los esquemas de clasificación basado en la extracción de características y ajuste de datos primarios, el cual permite clasificar series temporales de distintas longitudes y/o con datos faltantes. La originalidad e importancia de la investigación propuesta se basa en los siguientes aspectos:

- Las series temporales, usualmente no exhiben una correlación alta a lo largo del tiempo, pero a menudo pueden mostrar un patrón de correlación alta sobre intervalos parciales, por ende el algoritmo *Biclustering* puede encontrar grupos que sean más homogéneos a comparación de los métodos tradicionales para hacer agrupación de datos de series temporales.
- Se encuentra que la técnica de clasificación *Biclustering* ha sido poco explorada en el campo series temporales, especialmente en series de distintas longitudes.
- Los tres enfoques propuestos para clasificar datos de series temporales, finalmente se enfocan en una modificación de las características a emplear en los algoritmos de agrupamiento tradicionales. Lo cual hace que se tengan dificultades en cuanto a la elección de las medidas de disimilaridad y asignación del número de grupos óptimo.

La hoja de ruta está estructurada como sigue: en la sección 2 se presentan los métodos de clasificación de series temporales utilizados hasta la fecha, sus variedades, algoritmos y aplicaciones. En la sección 3 se presenta la técnica o método de *Bicluster*, en que consiste el método, los diferentes tipos de modelos planteados y el enfoque utilizado para poder aplicarlos al análisis de series de tiempo. En la sección 4 se presenta la técnica de extracción de características, como se aplica al análisis de conglomerados con datos de tipo series temporales y cuales son las características más apropiadas y significativas en este tipo de estudio. En la sección 5 se presenta la parte innovadora y propuesta por los autores, *Biclustering* con extracción de características. En la sección 6 se presenta la parte experimental, los datos empleados, los resultados obtenidos y una comparación con los otros

métodos clásicos. En la sección 7 la discusión de los resultados y las conclusiones pertinentes, para finalmente presentar la bibliografía consultada.

II. MÉTODOS DE CLASIFICACIÓN DE SERIES TEMPORALES

Hasta la fecha, varios investigadores han desarrollado algunos métodos de agrupación específicos para el campo de las series temporales. Liao [7] presenta una revisión exhaustiva de la literatura de los métodos de agrupamiento de datos de series de tiempo propuestos hasta el año 2004, así como un recuento de las medidas de distancia o similaridad más comúnmente usadas para este propósito. El autor, encuentra que once algoritmos propuestos se basan directamente en los datos originales, seis indirectamente en las características extraídas de los datos originales, y catorce indirectamente en modelos ajustados a los datos originales. Así mismo, muestra que los trabajos desarrollados, hasta el año 2004, para clasificar datos de series de tiempo son escasos comparados con aquellos focalizados en datos estáticos o de tipo corte transversal.

En los años siguientes, más específicamente hasta el año 2010, los algoritmos planteados para clasificar datos de series temporales, se caracterizan por estimular el uso de medidas de distancia o extracción de características más acordes al comportamiento innato de las series de tiempo (ver [13], [14]).

En [15] proponen una metodología conocida como agrupamiento espectral, la cual se basa en estructuras que permiten identificar patrones en las series de tiempo ligadas a los movimientos secuenciales de una o varias personas captadas en una imagen. En [16] se propone una metodología basada en la densidad de pronóstico de cada una de las series en la muestra, enfocándose en los pronósticos puntuales. Los investigadores utilizan un método de remuestreo en particular combinado con la estimación no paramétrica de Kernel. Otra de las investigaciones propuestas en la literatura, extiende ésta idea al caso de modelos no paramétricos [17]. Finalmente, en [18] se propone un algoritmo para agrupar series de tiempo heterocedásticas, con patrones de volatilidad similares, basados en la descomposición del cuadrado de los residuales de las series.

En los años restantes y en el actual, se han publicado un buen número de artículos e investigaciones relacionadas con la clasificación de datos de series temporales, sin embargo, se observa que dichas investigaciones se basan en las técnicas ya planteadas anteriormente para la clasificación de este tipo de datos y buscan dar un enfoque práctico, más que teórico. En las investigaciones halladas en la literatura, que proponen una nueva metodología, se encuentra que ellas se basan principalmente en la inteligencia artificial, véase por ejemplo [19], [20], [21], [22], [23], pero también en métodos heurísticos como lo son la colonia de hormigas [24]. En [21] se propone una aproximación del algoritmo *k-means* pero en vez de trabajar con la distancia Euclidea, se plantea el uso de la distorsión de tiempo dinámico, (*DTW*) por sus siglas en inglés.

Por otra parte, durante estos últimos años se da el desarrollo de un nuevo método de agrupación llamado el algoritmo *Biclustering*, el cual es un método innovador que propone una

nueva forma de hacer la clasificación permitiendo el solapamiento de los diferentes clusters, así como el agrupamiento no sólo de los objetos si no también de las características o atributos para poder dar pie a encontrar patrones ocultos [12].

En la siguiente sección se presenta una breve descripción de este método de clasificación.

III. MÉTODO BICLUSTERING

Biclustering es una técnica de agrupación relativamente nueva, que detecta grupos informativos compuestos no solo de objetos sino también de atributos, es decir, captura en los diferentes grupos objetos y atributos simultáneamente. Fue propuesta en 1972 por Hartigan [25] bajo el término de “clustering directo” y fue ya en 1996, más de 20 años después, que se le empezó a llamar por su nombre actual [26]. Esta técnica se ha aplicado mucho en los últimos años para estudiar la expresión genética y es por ello que algunos de los términos que se utilizan tienen que ver con dicha disciplina.

Si bien el biclustering se puede tratar como una evolución del clustering [25], debido a que permite agrupar diversos conjuntos de genes (objetos) que tengan un perfil genético (características o atributos) similar en todas las condiciones experimentales analizadas, también es claro que la capacidad del biclustering para el análisis de expresión supera a la de los métodos clustering tradicionales. Esto debido principalmente a las siguientes características:

1. El agrupamiento simultáneo de genes y condiciones.
2. Permite el fenómeno de solapamiento entre los diferentes grupos hallados.

Es por medio de la primera característica que los grupos hallados por el método biclustering, también conocidos como biclusters, representan genes que actúan de manera similar bajo una única serie de condiciones, no necesariamente todas, y es por medio de la segunda característica que se permite tener genes en más de un bicluster a la vez, lo cual resalta la diferencia principal entre Bicluster y clustering clásico.

El objetivo principal de los métodos biclustering es identificar subgrupos de genes y subgrupos de condiciones aplicando clustering sobre las filas (objetos) y columnas (atributos) de forma simultánea, ver figura 1, en lugar de analizar por separado cada una de estas dimensiones [27]. No obstante, los métodos biclustering se enfrentan a los mismos problemas que los métodos clásicos de clustering, problemas que dependen netamente del tipo de datos que se analice, las condiciones de la base de datos entre otras. Pero comparando tiempos de ejecución, el algoritmo biclustering resulta ser más eficiente y eficaz.

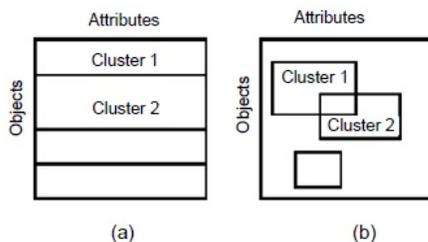


Figura 1: Diferencia entre Clustering (a) y Biclustering (b).

III-A. Tipos y Estructuras de Biclusters

Existen diferentes tipos de biclusters que son útiles para establecer qué clase de algoritmo biclustering es capaz de encontrar dicho tipo de bicluster. Los principales tipos de biclusters se pueden ver en la figura 2 y son, según [27]:

- Biclusters formados por valores constantes (a).
- Biclusters formados por valores constantes en filas (b) o columnas (c).
- Biclusters formados por valores coherentes (d).

Es de resaltar que los algoritmos biclustering más simples existentes, se enfocan en identificar biclusters formados por valores constantes. Cuando el objetivo es encontrar biclusters constantes, se consideran formas de reordenar las filas y las columnas de la matriz de datos para agrupar aquellas que sean similares. Para identificar éstos tipos de biclusters es común normalizar las filas o las columnas de la matriz utilizando para ello la media de fila o de columna. Por medio de lo anterior es que este tipo de biclusters se pueden transformar en biclusters de valores constantes. Por último existen algoritmos que se enfocan en buscar biclusters con valores coherentes tanto en las filas como en las columnas. Este tipo de biclusters no se pueden encontrar con tanta facilidad y no es válido considerar que los valores en el bicluster se obtienen a partir de un modelo aditivo o multiplicativo; se requieren métodos más sofisticados para evaluar la calidad de los resultados

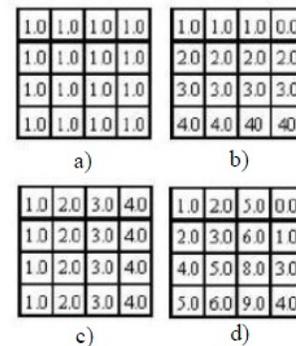


Figura 2: Principales tipos de biclusters

Por otra parte, los algoritmos biclustering pueden considerar dentro de su planteamiento dos tipos de objetivos distintos principalmente. El primero es el de encontrar un único bicluster (ver (a) en figura 3) o encontrar K biclusters, donde K es un número conocido con anterioridad o no. La mayoría de métodos o algoritmos biclustering asumen la existencia de varios biclusters en la matriz de datos pero también existen otras que sólo buscan hallar un único bicluster, es decir, se centran en encontrar un resultado que siga determinado criterio a pesar de que puede encontrar más de un resultado. Es cuando los algoritmos biclustering asumen que existe más de un bicluster en la matriz de datos que se pueden obtener varios tipos de estructuras. Las principales estructuras se pueden ver en la figura 3 y son, según [27]:

- Biclusters con filas y columnas exclusivas (b).
- Biclusters sin solapamiento con estructura de tablero de ajedrez (c).

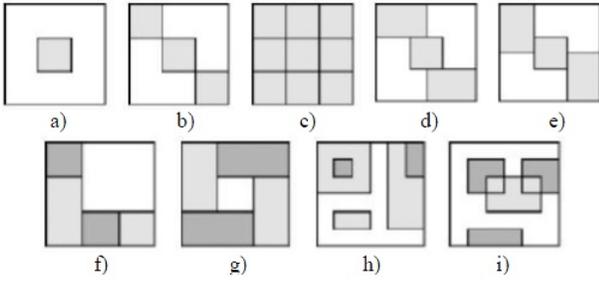


Figura 3: Principales tipos de estructuras de Biclusters

- Biclusters con filas exclusivas (d) y columnas exclusivas (e).
- Biclusters no solapados con estructura de árbol (f).
- Biclusters no solapados y no exclusivos (g).
- Biclusters solapados con estructura jerárquica (h).
- Biclusters arbitrarios (i).

Cabe resaltar que dependiendo de la técnica a utilizar en el momento de hacer el estudio biclustering, se podrá obtener un tipo u otro de estructura.

III-B. Algoritmos Biclustering

La gran mayoría de algoritmos biclustering se han desarrollado desde el año 2000. Existen técnicas muy variadas con características comunes y diferencias, pero no por ende dejan de ser efectivas. Varios modelos y algoritmos Biclustering se han propuesto y han sido empleados con éxito, a continuación se presentan brevemente los dos algoritmos principales de biclustering.

III-B1. El modelo Plaid: Fue propuesto por Lazzeroni y Owen en el 2002 [28]. Es un método Bicluster que asume la estructura de solapamiento y se ajusta a los datos a través de la minimización del error. Sea Z_{ij} un valor de elemento del objeto i -ésimo y j -ésimo atributo, luego el valor de Z_{ij} se modela con la siguiente formula:

$$Z_{ij} = \Psi_{ij0} + \sum_{k=1}^K \Psi_{ijk} \rho_{ik} \kappa_{jk} + \varepsilon_{ij} \quad (1)$$

donde:

- K es el número de Biclusters.
- ε_{ij} es el ruido.
- $\rho_{ik} \kappa_{jk}$ son variables de pertenencia.
- Ψ_{ijk} valor del Bicluster.
- Ψ_{ij0} valor del fondo.

Para estimar el modelo propuesto en la ecuación (1), se propone el modelo de optimización dado por la siguiente ecuación:

$$\min \sum_i \sum_j \left(Z_{ij} - \sum_{n=0}^K \Psi_{ijn} \rho_{in} \kappa_{jn} \right)^2 \quad (2)$$

s.a : $\rho_{ik} \kappa_{jk} = 0, 1$

Dicho problema de optimización es muy difícil de resolver directamente y por ello se propone extraer uno a uno los Biclusters, obteniendo el siguiente problema de optimización:

$$\min \sum_i \sum_j \left(Z_{ij}^{(k)} - \Psi_{ijk} \rho_{ik} \kappa_{jk} \right)^2 \quad s.a : \rho_{ik} \kappa_{jk} = 0, 1 \quad (3)$$

donde:

$$Z_{ij}^{(k)} = Z_{ij} - \sum_{m=0}^{K-1} \Psi_{ijm} \rho_{im} \kappa_{jm}$$

Para resolver el problema de optimización propuesto en la ecuación (3), los autores proponen emplear un método de programación lineal relajada usando los mínimos cuadrados binarios para determinar los valores de ρ_{ik}, κ_{jk} y proponen métodos iterativos para estimar las demás variables.

III-B2. El algoritmo de Cheng y Church: Cheng y Church [29] establecieron un método basado en la búsqueda del resultado final en un problema de optimización. Parten del supuesto de que para que un subgrupo sea un bicluster, sus valores han de evolucionar al unísono y dicha característica se representa por medio del cuadrado medio residuo, (MSR) por sus siglas en inglés. Este valor se considera como un índice de calidad y es la varianza del conjunto de todos los valores del bicluster, más la varianza de la fila y varianza de columna. Dicho valor mide el nivel de coherencia de los valores a lo largo de las filas y las columnas, y a mayor valor del MSR, menos calidad tendrá el bicluster. Partiendo de esta definición, los autores definieron al δ -cálculo como aquel subconjunto de filas y columnas cuyo valor de residuo no supera un determinado valor umbral. Por ende, el objetivo principal del algoritmo de biclustering es el de obtener una sub matriz del mayor tamaño posible y con un residuo no superior a un determinado umbral, el algoritmo se entiende como uno de búsqueda local.

III-C. Biclustering para Clasificar Series Temporales

En la literatura se encuentra que el uso de los métodos biclustering para la clasificación de datos de series temporales es limitado [4]. Solamente se encuentran los trabajos realizados por:

- L. Jeonghwa, L. Youngrok, y J. Chi-Hyucken en 2009 [4]; proponen una modificación del modelo *Plaid* para clasificar datos de series temporales. Se modifica la manera en que se actualizan los parámetros y los efectos en las capas, el método de podamiento o criterio de parada, los valores iniciales y se deben satisfacer unos supuestos claves relacionados con los datos que se van a manejar. Por ejemplo, que las series deben tener igual longitud y no pueden tener datos faltantes ya que trabajan con las series en bruto.
- L. Jeonghwa y J. Chi-Hyuck en 2010 [12]; proponen un algoritmo Biclustering diseñado especialmente para trabajar con series de tiempo que tengan la estructura del modelo *ARMA*. Para el funcionamiento adecuado del algoritmo, los autores plantean tres supuestos fuertes y estos son: los biclusters a encontrar no se pueden solapar, cada objeto tiene un valor constante diferente en el bicluster y por último, todos los biclusters tienen el

mismo orden del modelo *ARMA* y éste es conocido. En consecuencia, estos supuestos restringen principalmente a que cada serie pertenezca a un único bicluster y restringe los biclusters a tener un periodo de tiempo consecutivo. Al igual que la propuesta anterior, el algoritmo obliga a las series temporales a tener una misma longitud y a no tener datos faltantes. Para resaltar, los autores también se basan en el modelo *Plaid*, modifican las técnicas de actualización e inicialización y estiman los parámetros del modelo *ARMA* para cada serie de tiempo.

En general, estas propuestas no consideran la extracción de características y no permiten trabajar con serie de distintas longitudes.

A continuación se muestra la técnica de extracción de características de una serie temporal.

IV. TÉCNICA DE EXTRACCIÓN DE CARACTERÍSTICAS

Como se ha mencionado, en años recientes ha crecido significativamente el deseo de estudiar y aplicar las técnicas de agrupación, a datos que no sean de tipo corte transversal, como lo son las series temporales [30]. No obstante la aplicación directa de los algoritmos clustering tradicionales se puede ver afectada cuando las longitudes de las series a clasificar son muy largas. En estos casos la medida de similaridad planteada entre los datos, tiende a ser dudosa e inexacta debido al espacio de tan alta dimensión. Adicionalmente, la mayoría de los métodos tradicionales no funcionan bien cuando hay datos inexistentes o huecos en las series (lo cual puede pasar con frecuencia cuando los datos son tomados experimentalmente) debido a que se basan en distancias métricas.

Por todo lo anterior surge la necesidad de plantear un método que permita, bajo éstas condiciones, hacer el clustering adecuado de las series sin perder información de las mismas. Al proceso que permite hacer esto, se le conoce como extracción de características y es, como su nombre lo indica, el proceso por el cual se el hace clustering a series de tiempo basado en las características estructurales extraídas. La justificación de dicho proceso es básica, como se mencionó al inicio de esta sección, la mayoría de los métodos clustering requieren de muchas condiciones que suelen fallar dependiendo de las condiciones dadas, por ende, al aplicar extracción de características no se agruparían puntos o valores usando una distancia métrica, se agruparían medidas globales extraídas de cada serie y se eliminarían la mayoría de problemas de los algoritmos, tales como longitudes grandes o datos faltantes, ya que se considera un valor único que representa cada serie y dicho valor es el que alimenta los diferentes algoritmos existentes.

El problema de agrupación se centra ahora en encontrar las mejores características para extraer de cada serie. En [30] argumentan que dependiendo de cómo se esté trabajando la serie, es decir, si se considera la serie cruda o la serie descompuesta las características a extraer varían. En la tabla II se puede ver con mayor claridad lo propuesto por los autores, donde se marca con una *X* a la característica que se recomienda extraer según el tipo de serie.

Característica	Serie Cruda	Serie Descompuesta
Tendencia		X
Estacionalidad		X
Correlación Serial	X	X
No Linealidad	X	X
Asimetría	X	X
Curtosis	X	X
Auto-Similitud	X	
Caos	X	
Periodicidad, Efectos Aleatorios	X	

Tabla II: Características a extraer según tipo de serie propuesta en [30]

El método propuesto en este trabajo está basado en ésta técnica de extracción de características, y el método Biclustering. En la siguiente sección se presentará detalladamente la metodología propuesta.

V. MÉTODO BICLUSTERING CON EXTRACCIÓN DE CARACTERÍSTICAS PARA SERIES TEMPORALES

Con el fin de brindar una alternativa de solución a algunas falencias detectadas en los métodos clustering para series temporales. Se propone una nueva metodología de clasificación aprovechando a su vez, las cualidades del método Biclustering. El objetivo es aplicar la novedosa técnica Biclustering para clasificar series temporales, sin tomar la serie en bruto, como en [11], [12], sino realizando una extracción de características. La metodología propuesta se justifica en lo siguiente:

- Una de las principales fallas de los algoritmos existentes para hacer clustering a datos de series de tiempo se presenta al trabajar con series de mucha longitud. En los casos de aplicación de la vida real, resulta ser bastante común trabajar con series que cuentan con muchas observaciones a través del tiempo, haciendo de ellas series largas y difíciles de manipular para dichos algoritmos. Por ende, al hacer una extracción de características se estaría convirtiendo la serie original de t puntos en un solo dato que logre capturar, a grandes rasgos, la mayor información de la misma. Permitiendo así evadir el problema de grandes longitudes y facilitando el buen funcionamiento de los algoritmos.
- Una segunda falencia de los algoritmos existentes se presenta cuando dos supuestos principales son violados. El supuesto de que todas las series dentro del conjunto de series a ser analizadas deben tener la misma longitud y el supuesto de que ninguna de las series puede tener los denominados huecos o datos faltantes. Nuevamente haciendo referencia a los casos de aplicación reales, resulta ser bastante empírico y escaso poder trabajar con series de tiempo de igual longitud puesto que cada serie representa algún proceso estocástico distinto, que puede tener diferentes mediciones, periodicidades y frecuencias. Lo cual no solo dificulta el hecho de cumplir la igualdad en longitud si no también, como los datos pueden ser tomados experimentalmente, es normal que en algunos puntos del tiempo haya ocurrido algún evento desafortunado y no se haya podido obtener la medida en dicho punto. Estos problemas son fáciles de solucionar si se

hace una extracción de características, si se identifican medidas globales que representen con un alto grado de confiabilidad a cada serie del conjunto de datos. Independiente de la longitud de la serie o de la falta de datos, se podría obtener un único dato que represente dicha serie permitiendo así el funcionamiento adecuado de los algoritmos y evitar entonces los problemas mencionados.

- El algoritmo Biclustering es un método de agrupación relativamente nuevo que ha sido estudiado por varios investigadores y aplicado a otros campos del conocimiento satisfactoriamente. Sin embargo se encuentra que su aplicación ha sido limitada en las series temporales, y en las metodologías propuestas solo se emplean las series brutas lo cual conlleva a un alto costo computacional.
- Como se presentó en la sección 3, la técnica de extracción de características permite eliminar todo tipo de posibles fallas o impedimentos en los algoritmos además de que permite ahorrar en costos computacionales al hacer más corto el tiempo del algoritmo puesto que en vez de trabajar con series de extensa longitud se trabaja con datos únicos que capturan los efectos globales de las series.

La metodología propuesta básicamente consta de tres pasos:

- *Paso 1:* Organización de las series temporales a clasificar. Se construye una matriz, cuyas columnas corresponden a cada serie de tiempo, que pueden o no ser iguales en longitud y pueden presentar datos faltantes.
- *Paso 2:* Extracción de características.
- *Paso 3:* Aplicación del algoritmo Biclustering.

A continuación se explican los pasos 2 y 3:

Paso 2 extracción de características.

Con base en los resultados hallados por [30], se decidió extraer las siguientes características:

- ***Tendencia, estacionalidad y efectos aleatorios:***

Estas características son de las más comunes en una serie de tiempo en cuanto a que se pueden encontrar y extraer con facilidad, además resulta ser bastante natural caracterizar una serie de tiempo por su grado de tendencia y estacionalidad.

En nuestra propuesta, para calcular la tendencia se hace una regresión lineal entre los datos de cada serie y el número de observaciones en el tiempo. Luego, se almacena la pendiente obtenida en dicha regresión y ésta será el dato que representa la tendencia de la serie.

En cuanto a la estacionalidad, se hace una descomposición de los datos de la serie de tiempo en sus tres componentes principales: tendencia, estacionalidad y ruido o efectos aleatorios. Se usa el modelo aditivo dado en la ecuación 4, donde Y_t es la serie de tiempo, T_t es la componente de tendencia, S_t es la componente de estacionalidad y e_t la componente relacionada con el ruido o los efectos aleatorio. Luego se almacenan los datos de S_t para estacionalidad. Sin embargo, al tener tantos datos para S_t como puntos el tiempo haya, se hace una media

de los datos para poder encontrar un único dato que represente dicha estacionalidad de la serie dada.

$$Y_t = T_t + S_t + e_t \quad (4)$$

Finalmente la parte de ruido o efecto aleatorio se obtiene igualmente que la estacionalidad, se descompone la serie en sus tres componentes se almacenan los datos de e_t y al final se hace un promedio para obtener un único dato.

- ***Curtosis:***

La curtosis es una medida que establece si cierto conjunto de datos son puntiagudos o acostados, relativos a una distribución normal. Una base de datos que presente una curtosis alta tiende a tener un pico o un máximo cerca a la media, decrece con rapidez y suele tener colas altas. Una base de datos con una curtosis baja tiende a tener un comportamiento plano cerca a la media en vez de un pico. Es así como, una curtosis positiva indica la existencia de algún pico o máximo en los datos, mientras que una curtosis negativa indica la existencia de una distribución plana y sin picos.

La curtosis K se calcula como se muestra en la ecuación 5, donde Y_t es la media de los datos, σ es la desviación estándar y n es el número de datos u observaciones.

$$K = \frac{1}{n\hat{\sigma}^4} \sum_{t=1}^n (Y_t - \bar{Y}_t)^4 \quad (5)$$

- ***Asimetría:***

La asimetría es como su nombre lo indica una medida encargada de medir la simetría, o más precisamente, la falta de asimetría en un conjunto de datos. Una distribución o conjunto de datos es simétrica si se ve igual tanto a la izquierda o derecha de un punto central. Se utiliza para caracterizar el grado de asimetría de los valores alrededor de la media. Valores negativos en la asimetría indican que los datos son asimétricos hacia la izquierda y valores positivos indican asimetría hacia la derecha. La asimetría indica que tan pesadas son las colas de la distribución de los datos. Para una variable Y_t , el coeficiente de asimetría S se calcula como se muestra en la ecuación 6, donde \bar{Y}_t es la media, σ es la desviación estándar y n es el número de datos u observaciones.

$$S = \frac{1}{n\hat{\sigma}^3} \sum_{t=1}^n (Y_t - \bar{Y}_t)^3 \quad (6)$$

Paso 3 aplicación del algoritmo Biclustering.

Con base en los resultados hallados en la extracción de características, se construye la matriz de entrada del algoritmo Biclustering. Dicha matriz tiene la siguiente estructura:

	<i>Carac. 1</i>	<i>Carac. 2</i>	<i>...</i>	<i>Carac. k</i>
<i>Serie 1</i>	a_{11}	a_{12}	\dots	a_{1k}
<i>Serie 2</i>	a_{21}	a_{22}	\dots	a_{2k}
<i>Serie 3</i>	a_{31}	a_{32}	\dots	a_{3k}
\vdots	\vdots	\vdots	\vdots	\vdots
<i>Serie n</i>	a_{n1}	a_{n2}	\dots	a_{nk}

Figura 4: Matriz diseñada para el nuevo método.

donde a_{ij} es el dato de la j -ésima característica para la i -ésima serie.

El algoritmo Biclustering utilizado para la implementación fue el propuesto por Cheng y Church expuesto en la sección III. El algoritmo recibe el número máximo de biclusters a buscar y tiene dos parámetros, Δ que es el parámetro que indica la puntuación máxima aceptada y α que es el factor de escala.

La viabilidad de la metodología propuesta es examinada experimentalmente en la siguiente sección.

VI. EXPERIMENTACIÓN

En esta sección se presenta la parte de aplicación y experimentación de la metodología propuesta. Los resultados encontrados con ella son comparados con los hallados bajo el método convencional K-medias. Así mismo, se examina el efecto de la elección de las características sobre los resultados.

VI-A. Datos Experimentales

La base de datos considerada para hacer los diferentes experimentos se encuentra disponible en [31], la cual hace referencia a datos de gráficos de control de calidad sintéticos. La base consta de seiscientas series temporales, de frecuencia anual, divididas en seis grupos, cada uno con un total de cien series y cada grupo se caracteriza de alguna forma especial:

- Grupo 1: son las series normales, sin ninguna modificación o caracteriza especial (véase figura 5).
- Grupo 2: es el grupo de las series que presentan una componente cíclica en su comportamiento (véase figura 6).
- Grupo 3: las series se caracterizan por tener una tendencia positiva (véase figura 7).
- Grupo 4: series que poseen la tendencia negativa (véase figura 8).
- Grupo 5: series que tienen un desplazamiento hacia arriba (véase figura 9).
- Grupo 6: series que presentan un desplazamiento hacia abajo (véase figura 10).

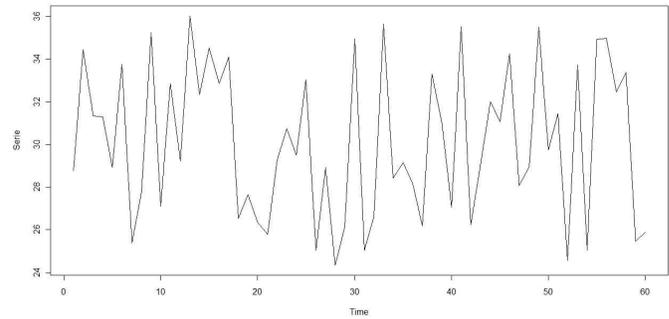


Figura 5: Ejemplo Serie de Tiempo en el Grupo #1

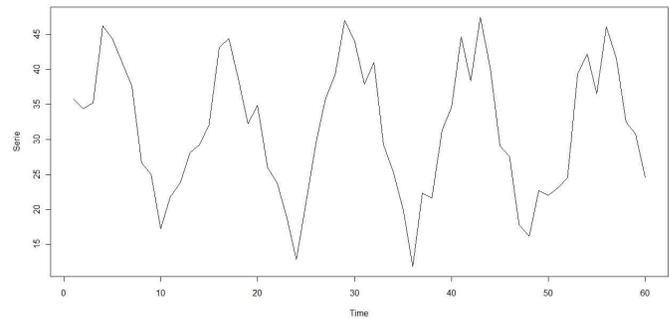


Figura 6: Ejemplo Serie de Tiempo en el Grupo #2

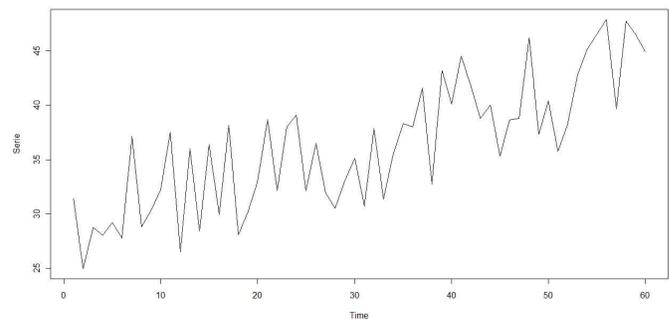


Figura 7: Ejemplo Serie de Tiempo en el Grupo #3

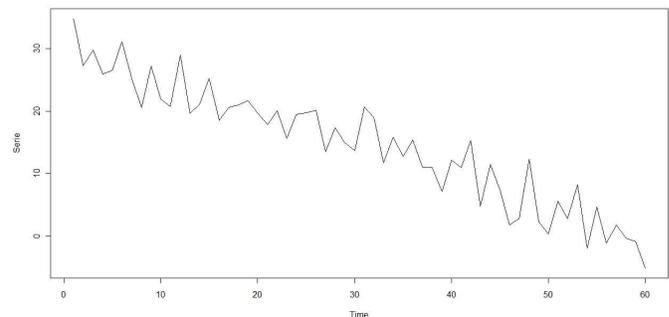


Figura 8: Ejemplo Serie de Tiempo en el Grupo #4

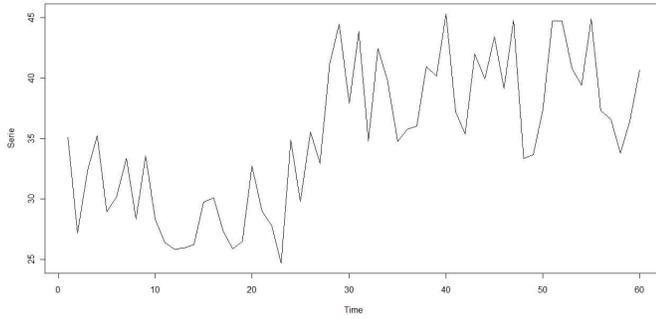


Figura 9: Ejemplo Serie de Tiempo en el Grupo #5

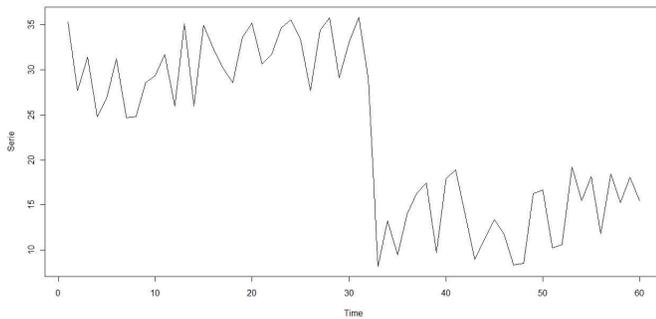


Figura 10: Ejemplo Serie de Tiempo en el Grupo #6

VI-B. Escenarios Evaluados

Se consideran las cinco características presentadas en la sección anterior y se calculan de igual modo. En la tabla III se muestra la notación que se empleará en el resto del documento para referirse a cada una de ellas.

Notación	Característica
B	Tendencia
S	Estacionalidad
R	Ruido
K	Curtosis
A	Asimetría

Tabla III: Notación para cada característica

Con el fin de evaluar el efecto que tiene la elección de las características sobre los resultados del agrupamiento, se consideraron los siguientes escenarios:

Escenario #1	$[B, S, R, K, A]$
Escenario #2	$[B, S, K, A]$
Escenario #3	$[B, K, A]$
Escenario #4	$[B, S, R, A]$
Escenario #5	$[B, S, R, K]$
Escenario #6	$[R, K, A]$
Escenario #7	$[B, S]$
Escenario #8	$[K, A]$
Escenario #9	$[S, K, A]$
Escenario #10	$[B, R]$

Tabla IV: Escenarios propuestos para los experimentos

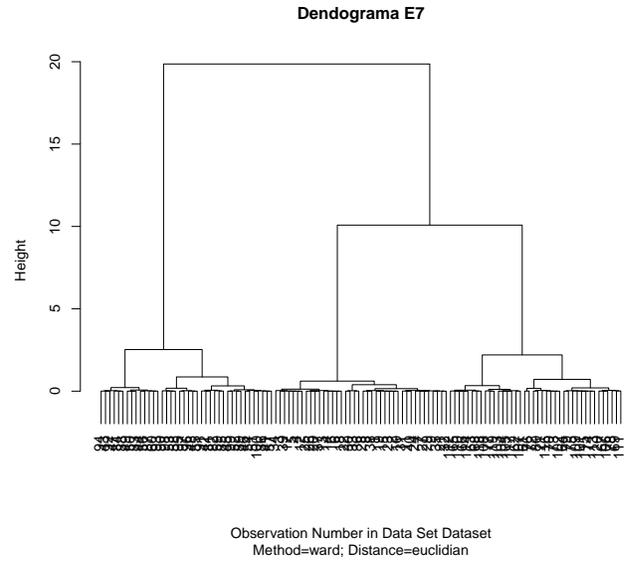


Figura 11: Ejemplo dendrograma escenario #7

Grupos originales	Nuevos grupos					
	1	2	3	4	5	6
1	100	0	0	0	0	0
2	100	0	0	0	0	0
3	0	0	45	20	35	0
4	0	65	0	0	0	35
5	0	0	40	25	35	0
6	0	45	0	0	0	55

Tabla V: Tabla de probabilidades de buena y mala clasificación escenario 1

A continuación se muestran los resultados obtenidos para los escenarios más representativos, bajo el método tradicional de K -medias y posteriormente con la metodología propuesta en este documento, en ambos casos se consideró la misma matriz de datos de entrada del método biclustering. Los escenarios cuyos resultados no son exhibidos se tienen en cuenta en las conclusiones finales del documento.

VI-C. Resultados K -medias

Con el fin de determinar el número de grupos a considerar en el método K -medias, se realizó inicialmente un algoritmo jerárquico aglomerativo. Dicho número es seleccionado a partir del dendrograma, fijando una distancia mínima, véase para el escenario 7 la figura 11 en donde se seleccionaron seis grupos. Una vez se fijó el número de grupos deseados, se ejecutó el algoritmo K -medias. La tabla de probabilidades de buena y mala clasificación para este método en los escenarios 1, 7 y 8 en muestran en las tablas V, VI y VII respectivamente.

En general se esperaría que si el método tiene un buen desempeño, las diagonales de las matrices anteriores deberían estar cercanas al 100%. Sin embargo, se observa que esto no

Grupos originales	Nuevos grupos				
	1	2	3	4	5
1	0	0	0	100	0
2	0	0	0	100	0
3	0	0	45	0	55
4	35	65	0	0	0
5	0	0	45	0	55
6	55	45	0	0	0

Tabla VI: Tabla de probabilidades de buena y mala clasificación escenario 7

Grupos originales	Nuevos grupos					
	1	2	3	4	5	6
1	0	0	10	0	40	50
2	0	5	25	0	15	55
3	0	30	0	40	30	0
4	0	35	0	40	20	5
5	20	20	40	0	10	10
6	5	5	35	15	20	20

Tabla VII: Tabla de probabilidades de buena y mala clasificación escenario 8

ocurre para estos tres escenarios ni para el resto. Adicionalmente, esta técnica requiere que el experimentador defina los números de grupos a formar lo cual puede proporcionar un ruido a los resultados encontrados.

VI-D. Resultados Biclustering

Para aplicar la metodología propuesta, se usó el algoritmo biclustering propuesto por Cheng y Church que se encuentra implementado en el paquete *biclust* del software *R*. En este método, a diferencia del método *K*-medias, el usuario sólo debe especificar un máximo número de grupos a formar, en nuestro caso se consideraron 10. En las tablas VIII, IX y X se muestran los resultados obtenidos en los tres escenarios anteriores. Es de destacar que éstas tablas no corresponden a probabilidades de buena y mala clasificación. Las tablas indican el número de biclusters hallados por el algoritmo, el porcentaje de series pertenecientes a los grupos originales que hayan sido clasificadas en los biclusters, y por último se muestran las características que determinan cada bicluster.

En general para este método, se encontró que el número de biclusters a formar es seleccionado por el mismo algoritmo según un criterio de parada especificado para este método por [29]. Adicionalmente, contrario al método de *K*-medias con el método biclustering se identifica las características más relevantes para cada grupo formado.

La tabla VIII, contiene los resultados obtenidos para el escenario 1, el cual considera todas las características. Se puede evidenciar que el algoritmo bicluster solamente encontró 6 biclusters cuando originalmente se le habían pedido 10. También es claro que de las 120 series trabajadas hubo algunas

Grupos originales	Bicluster					
	1	2	3	4	5	6
1	80	0	0	0	5	0
2	15	0	0	30	0	0
3	0	35	20	0	15	5
4	0	65	10	0	10	0
5	0	15	25	0	0	25
6	5	10	20	0	5	0
Características	BSRK	SRK	SRA	BSRA	SRA	BSRA

Tabla VIII: Resultados obtenidos Biclustering escenario 1

Grupos originales	Bicluster		
	1	2	3
1	100	0	0
2	100	0	0
3	0	100	0
4	20	65	15
5	0	100	0
6	40	60	0
Características	BS	BS	BS

Tabla IX: Resultados obtenidos Biclustering escenario 7

que no logro clasificar, lo cual puede indicar dos aspectos: primero, que el trabajar con las cinco características disminuye la calidad de los resultados obtenidos, o segundo, que hay algunas características que están deteriorando los resultados y deben ser removidas.

En la tabla IX se muestran los resultados obtenidos para el escenario 7 que considera la tendencia y la estacionalidad. El método identificó un total de tres biclusters, y cada uno de ellos determinados por las dos características. Los resultados encontrados dan indicios del buen funcionamiento de la metodología y la elección de estas dos características:

- En el primer bicluster fueron clasificadas, en su totalidad, las series pertenecientes al grupo 1 y al grupo 2 de los datos, es decir, series normales y series con comportamiento

Grupos originales	Bicluster				
	1	2	3	4	5
1	65	25	10	0	0
2	65	25	5	5	0
3	30	35	0	30	0
4	20	65	0	15	0
5	35	15	10	5	30
6	50	15	25	0	10
Características	KA	KA	KA	KA	KA

Tabla X: Resultados obtenidos Biclustering escenario 8

cíclico, que se caracterizan por tener valores próximos en la tendencia y la estacionalidad.

- En el segundo bicluster fueron agrupadas en su totalidad las series pertenecientes a los grupos 3 y 5, es decir, series con tendencia positiva y series con desplazamientos hacia arriba. Lo cual tiene sentido debido a que éstas últimas series tienen una etapa en la cual funcionan netamente como series con tendencia positiva, puesto que crecen prolongadamente.

En el último caso, en la tabla X, se muestran los resultados obtenidos para el escenario 8 que considera la curtosis y la asimetría. Los resultados muestran que fueron 5 los biclusters hallados y que en todos, las dos características fueron determinantes. Sin embargo, se nota una gran dispersión en cuanto a la agrupación de las series, es decir, las series de cada grupo se esparcen casi que uniformemente en los biclusters encontrados y adicionalmente, existen algunas series que no fueron clasificadas. Esto puede indicar que las dos características no son suficientes para este clasificar el conjunto de series temporales.

VII. CONCLUSIONES

Se propuso una metodología para clasificar datos de series temporales basado en el algoritmo Biclustering y la técnica de extracción de características, que permite manipular series con distintas longitudes y datos faltantes.

Se comprobó la viabilidad y el funcionamiento de la metodología propuesta empleando datos experimentales. En el experimento se consideraron diferentes escenarios de las características de las series, que permitieron identificar las características más relevantes para clasificar series de tiempo.

REFERENCIAS

- [1] X. Wang and K. Smith, "Characteristic-based clustering for time series data," *Data Mining and Knowledge Discovery*, vol. 13, pp. 335–364, 2006.
- [2] V. Raja, "Clustering of time series," *Syracuse University*, 2012.
- [3] C. Chu and M. Orozco, "Clasificación de eventos sísmico en el volcán nevado del ruiz empleando reglas fijas de combinación de clasificadores," *Tendencias en Ingeniería de Software e Inteligencia Artificial*, vol. 3, pp. 215–220, 2007.
- [4] L. Jeonghwa, L. Youngrok, and J. Chi-Hyuck, "A biclustering method for time series data analysis," *APIEMS*, 2009.
- [5] C. Faloutsos, M. Ranganathan, and Y. Manolopoulos, "Fast subsequence matching in time-series databases," *In Proc. of the ACM SIGMOD International Conference on Management of Data*, pp. 419–429, 1994.
- [6] J. Scargle, "Timing: New methods for astronomical time series analysis," *Bulletin of the American Astronomical Society*, pp. 32–1438, 2000.
- [7] T. Liao, "Clustering of time series data-a survey," *Pattern Recognition*, vol. 38, pp. 1857–1874, 2005.
- [8] T. Barton, "Cluster analysis of time series data," *Department of Theoretical Computer Science Faculty of Information Technology*, 2012.
- [9] S. Stephan, G. Julia, L. Andreas, D. Ernesto, and A. Sahin, "Pattern recognition and classification for multivariate time series," *SensorKDD*, 2011.
- [10] W. Xiaozhe and S. Kate, "Characteristic-based clustering for time series data," *Data Mining and Knowledge Discovery*, 2005.
- [11] Z. Xiaohang, L. Jiaqi, D. Yu, and L. Tingjie, "A novel clustering method on time series data," *ELSEVIER*, 2011.
- [12] L. Jeonghwa and J. Chi-Hyuck, "Biclustering of arma time series," *Journal of Zhejiang University-SCIENCE A*, 2010.
- [13] M. R. Cogollo and J. D. Velasquez, "Una propuesta para agrupar series de tiempo de distintas longitudes," 2012.
- [14] E. Maharaj and P. D'Urso, "A coherence-based approach for the pattern recognition of time series," *Physica A*, vol. 389, pp. 3516–3537, 2010.
- [15] X. Wang, L. Wang, and A. Wirth, "Pattern discovery in motion time series via structure-based spectral clustering," *IEEE*, 2008.
- [16] A. Alonso, J. Berrendero, A. Hernández, and A. Justel, "Time series clustering based on forecast densities," *Computational Statistics and Data Analysis*, vol. 51, pp. 762–776, 2006.
- [17] J. Vilar, A. Alonso, and J. Vilar, "Non-linear time series clustering based on non-parametric forecast densities," *Computational Statistics and Data Analysis*, vol. 54, pp. 2850–2865, 2010.
- [18] E. Otranto, "Clustering heteroskedastic time series by model-based procedures," *Computational Statistical and Data Analysis*, vol. 52, pp. 4685–4698, 2008.
- [19] V. Kavitha and M. Punithavalli, "An effective performance of fuzzy hierarchical clustering using time series data streams," *Communications in Computer and Information Science*, 2011.
- [20] P. Maji and S. Paul, "Microarray time-series data clustering using rough-fuzzy c-means algorithm," *Proceedings - 2011 IEEE International Conference on Bioinformatics and Biomedicine*, 2011.
- [21] W. Meesrikamolkul, V. Niennattrakul, and C. Ratanamahatana, "Shape-based clustering for time series data," *Lecture Notes in Computer Science*, 2012.
- [22] S. Aghabozorgi, M. Saybani, and T. Wah, "Incremental clustering of time-series by fuzzy clustering," *Journal of Information Science and Engineering*, 2012.
- [23] E. Maharaj and P. D'Urso, "Fuzzy clustering of time series in the frequency domain," *Information Sciences*, 2011.
- [24] K. Pancerz, A. Lewicki, and R. Tadeusiewicz, "Ant based clustering of time series discrete data - a rough set approach," *Lecture Notes in Computer Science*, 2011.
- [25] J. Hartigan, "Direct clustering of a data matrix.," *Journal of the American Statistical Association*, 1972.
- [26] B. Mirkin, "Mathematical classification and clustering.," *Kluwer Academic Publishers.*, 1996.
- [27] S. C. Serrano and P. A. G. López, "Análisis biclustering," 2011.
- [28] L. Lazzeroni and A. Owen, "Plaid models for gene expression data," *Statistica Sinica*, 2002.
- [29] Y. Cheng and G. Church, "Biclustering of expression data.," *Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology*, pp. 93–103, 2000.
- [30] W. Xiaozhe, K. Smith, and R. Hyndman, "Characteristic-based clustering for time series data," *Data Mining and Knowledge Discovery*, pp. 335–364, 2006.
- [31] <http://archive.ics.uci.edu/ml/datasets>