

Métodos de clústering para series de tiempo

Orlando de Jesus Uc Kantun

Centro de Investigación en Matemáticas. Unidad Monterrey

Email: orlando.uc@cimat.mx

Resumen—En este reporte se muestran el concepto de la distancia DTW y algunos métodos de clústering para series de tiempo, en particular, Euclidian k -means, Soft DTW k -means, DBA k -means y GAK k -means. Se ejemplifican los métodos de clústering en un conjunto de las 35 series de los precios diarios de las acciones al cierre de las compañías que componen el índice S&P/BMV IPC.

I. INTRODUCCIÓN

El análisis de series de tiempo es un área de la estadística, permite modelar datos con una dependencia temporal, y realizar pronósticos precisos y confiables. Los métodos de machine learning han ampliado el alcance de las series de tiempo, al permitir analizar un mayor número de series desde una perspectiva moderna.

El objetivo del presente es introducir y ejemplificar el concepto de distancia DTW y algunos métodos de clústering para series de tiempo, en particular: Euclidian k -means, Soft DTW k -means, DBA k -means y GAK k -means.

II. DYNAMIC TIME WARPING (DTW)

Sean $X_t = \{x_1, x_2, \dots, x_n\}$ y $Y_t = \{y_1, y_2, \dots, y_n\}$ dos series de tiempo. Para poder hacer comparaciones entre ellas primero es necesario definir una medida de distancia adecuada. La primera opción es aplicar la distancia euclidiana definida como:

$$D(X_t, Y_t) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2} \quad (1)$$

Pero esta distancia no permite contemplar la dependencia temporal entre las series. Además, es una distancia rígida, en el sentido de que las comparaciones son entre las observaciones en el mismo tiempo, lo cual solo permite calcular distancias entre series de la misma longitud.

Dynamic Time Warping (Deformación Dinámica del Tiempo) es un algoritmo que busca la alineación no-lineal óptima entre dos series. Fue introducido por Sakoe (1971) con aplicaciones en reconocimiento de voz.

La alineación óptima puede ser calculada recursivamente por:

$$D(X_i, Y_i) = \delta(x_i, y_i) + \min \left\{ \begin{array}{l} D(x_{i-1}, y_{j-1}) \\ D(x_i, y_{j-1}) \\ D(x_{i-1}, y_j) \end{array} \right\} \quad (2)$$

donde X_i es la subsecuencia $\{x_1, \dots, x_i\}$ y Y_i es la subsecuencia $\{y_1, \dots, y_i\}$. La distancia completa está dada por $D(X_t, Y_t) = D(X_n, Y_n)$.

Puede calcular distancias entre series de diferente longitud, lo cual es una ventaja con respecto a la distancia euclidiana.

La implementación directa del algoritmo DTW lleva a un tiempo de cómputo exponencial. Sin embargo, el hecho de que para obtener la distancia completa $D(X_t, Y_m)$ se tenga que realizar una recursión, permite memorizar las distancias parciales, lo que hace que el cálculo de la alineación óptima un proceso de orden $n \times m$, para una serie de longitud n y otra de dimensión m .

A continuación, el algoritmo DTW tomado de (Petitjean, Ketterlin y Gançarski, 2010).

Algorithm 1 DTW

Require: $A = \{a_1, \dots, a_S\}$

Require: $B = \{b_1, \dots, b_T\}$

Let δ be a distance between coordinates of sequences

Let $m[S, T]$ be the matrix of couples (cost, path)

```
1:  $m[1, 1] \leftarrow (\delta(a_1, b_1), (0, 0))$ 
2: for  $i \leftarrow 2$  to  $S$  do
3:    $m[i, 1] \leftarrow (m[i-1, 1] + \delta(a_i, b_1), (i-1, 1))$ 
4: end for
5: for  $j \leftarrow 2$  to  $T$  do
6:    $m[1, j] \leftarrow (m[1, j-1] + \delta(a_1, b_j), (1, j-1))$ 
7: end for
8: for  $i \leftarrow 2$  to  $S$  do
9:   for  $j \leftarrow 2$  to  $T$  do
10:     $minimum \leftarrow \minVal(m[i-1, j], m[i, j-1], m[i-1, j-1])$ 
11:     $m[i, j] \leftarrow (first(minimum) + \delta(a_i, b_j), second(minimum))$ 
12:   end for
13: end for
14: return  $m[S, T]$ 
```

En la Figura 1 se encuentra una comparación entre el cálculo de la distancia euclidiana comparado contra la distancia DTW. Se observa que, al contrario de la distancia euclidiana, la distancia DTW es más flexible, pues encuentra la distancia mínima entre un punto y sus adyacentes. El efecto de la deformación se observa en la flexibilidad de las distancias.

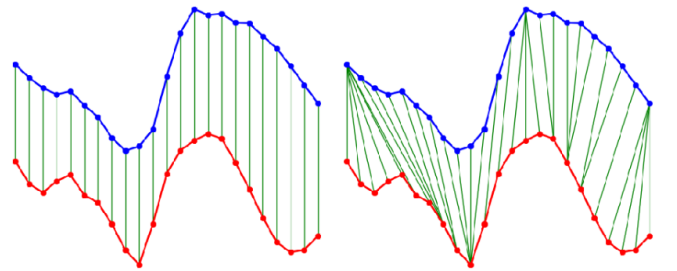


Figura 1. Comparación del cálculo de la distancia euclidiana contra la distancia DTW (Zhang, Tang, Huo y Zhou, 2014).

En la Figura 2 se encuentra un ejemplo de la matriz de cálculo de la distancia DTW. Nótese que se guardan las

distancias parciales, para luego encontrar la alineación mínima óptima.

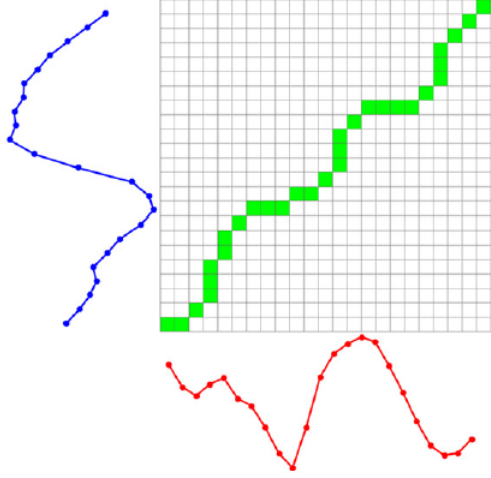


Figura 2. Ejemplo de la matriz de cálculo de la distancia DTW (Zhang, et al., 2014).

III. MÉTODOS DE CLÚSTERING PARA SERIES DE TIEMPO

En general, los métodos de clústering para series de tiempo se pueden agrupar en tres grandes grupos:

1. **Clústering de series de tiempo basado en disimilaridades:** Métodos basados en calcular distancias entre las series.
2. **Clústering de series de tiempo basado en modelos:** Métodos basados en identificar similitudes entre los modelos propuestos para cada serie.
3. **Clústering de series de tiempo basado en dependencias:** Métodos basados en cuantificar las correlaciones entre las series.

(Liao, 2005) y (Aghabozorgi, Shirkhorshidi y Wah, 2015).

III-A. Euclidian k -means

Un método base para clústering de series de tiempo es Euclidian k -means. El objetivo es, dados K centroides elegidos al azar, asignar cada serie al clúster cuya distancia euclidiana sea la mínima.

A continuación, una adaptación del algoritmo tomado de (Bo, Luo y Vo, 2016).

Algorithm 2 Euclidian k -means

Require: Number of clusters K

Require: Data set $T = \{t_1, t_2, \dots, t_n\}$ con $t_i = (x_i, y_i)$

- 1: Choose randomly K instances to be the m_k initial centroids.
- 2: For each instance, assign it to the cluster the centroid of which is the closest to the instance.

$$\min_{1 \leq k \leq K} \|t_i - m_k\|^2$$

- 3: For each cluster, recompute its centroid based on the instances in that cluster.
 - 4: If the convergence criterion is satisfied, then stop; otherwise, go back to Step 2.
-

III-B. Soft DTW k -means

Una variante del algoritmo k -means es Soft (Fuzzy) k -means. fue propuesto por Kaufman y Rousseeuw (1990).

La diferencia radica en que en lugar de minimizar la distancia euclidiana entre cada serie y los centroides, se considera una distancia ponderada por la plausibilidad de que la i -ésima serie pertenezca al k -ésimo clúster. El cálculo de los centroides también es ponderado.

Si en lugar de la distancia euclidiana se considera la distancia DTW el método se llama Soft DTW k -means.

Algorithm 3 Soft DTW k -means

Require: Number of clusters K

Require: Fuzzy parameter q

Require: Data set $T = \{t_1, t_2, \dots, t_n\}$ con $t_i = (x_i, y_i)$

- 1: Choose randomly K instances to be the m_k initial centroids.
- 2: For each instance, assign it to the cluster the centroid of which is the closest to the instance.

$$\min_{1 \leq k \leq K} \gamma_{ik}^q DTW(t_i, m_k)^2$$

with

$$\gamma_{ik}^q = \frac{1}{\sum_{j=1}^K \left(\frac{DTW(t_i, m_k)}{DTW(t_i, m_j)} \right)^{\frac{2}{q}}}$$

$$m_k = \frac{\sum_{i=1}^n \gamma_{ik}^q t_i}{\sum_{i=1}^n \gamma_{ik}^q}$$

$$\sum_k \gamma_{ik} = 1$$

- 3: For each cluster, recompute its centroid based on the instances in that cluster.
 - 4: If the convergence criterion is satisfied, then stop; otherwise, go back to Step 2.
-

III-C. DBA k -means

DTW Barycenter Averaging (DBA) es uno de los métodos más populares para clústering de series de tiempo. Fue propuesto por Petitjean, Ketterlin y Gançarski (2010).

El objetivo es minimizar la suma de cuadrados de las distancias DTW de una serie promedio con respecto al conjunto de series.

DBA sigue un esquema Expectation-Maximization. Se define una serie promedio \bar{T} e iterativamente:

1. Se considera a \bar{T} fija y se calcula la distancia DTW entre cada serie y \bar{T} , para poder encontrar el mejor alineamiento múltiple M .
2. Ahora se considera fijo a M y se actualiza \bar{T} como el mejor baricentro consistente con M .

El concepto de baricentro es en el sentido interpretativo de que el algoritmo calcula un promedio ponderado de las distancias DTW entre las series.

A continuación, el algoritmo DBA propuesto en (Petitjean, Forestier, Webb, Nicholson, Chen y Keogh, 2014).

Algorithm 4 DBA

Require: $T = \{t_1, t_2, \dots, t_n\}$ con $t_i = (x_i, y_i)$, the set of sequences to average.

Require: I , the number of iterations

```

1:  $\bar{T}$  el promedio ponderado inicial.
2: for  $i \leftarrow 1$  to  $I$  do
3:    $\bar{T} = \text{DBA\_update}(\bar{T}, M)$ 
4: end for
5: return  $\bar{T}$ 

```

Algorithm 5 DBA_update

Require: \bar{T}_{init} , the average sequence to refine (of length L).

Require: $T = \{t_1, t_2, \dots, t_n\}$ con $t_i = (x_i, y_i)$, the set of sequences to average.

```

1: // Step #1: compute the multiple alignment for  $\bar{T}_{init}$ 
2: alignment =  $[0, \dots, 0]$  // array of  $L$  empty sets
3: for each  $S$  in  $T$  do
4:   alignment_for_S = DTW_multiple_alignment( $\bar{T}_{init}$ ,  $S$ )
5:   for  $i=1$  to  $L$  do
6:     alignment[i] = alignment[i]  $\cup$  alignment_for_S[i]
7:   end for
8: end for
9: // Step #2: compute the multiple alignment for the alignment
10:  $\bar{T}$  be a sequence of length  $L$ 
11: for  $i=1$  to  $L$  do
12:    $\bar{T}(i) = \text{mean}(\text{alignment}[i])$  // arithmetic mean on the set
13: end for
14: return  $\bar{T}$ 

```

III-D. GAK k -means

La distancia DTW puede ser generalizada al concepto *alineamiento* entre series de tiempo.

Sean $X_t = (x_1, \dots, x_n)$ y $Y_t = (y_1, \dots, y_m)$, entonces un alineamiento π entre las series es un par de vectores (π_1, π_2) de tamaño $p \leq n+m-1$ tal que $1 = \pi_1(1) \leq \dots \leq \pi_1(p) = n$ y $1 = \pi_2(1) \leq \dots \leq \pi_2(p) = m$, con incrementos unitarios y sin repeticiones simultáneas. Las dos coordenadas π_1 y π_2 del alineamiento π son conocidas como funciones deformadoras.

Sea $A(n, m)$ el conjunto de todos los alineamientos entre dos series de longitud n y m , entonces la distancia DTW puede ser definida como:

$$DTW(x, y) = \min_{\pi \in A(n, m)} D_{x, y}(\pi) \quad (3)$$

donde, si se define $|\pi|$ como la longitud de π , entonces:

$$D_{x, y}(\pi) = \sum_{i=1}^{|\pi|} \varphi(x_{\pi_1(i)}, y_{\pi_2(i)}) \quad (4)$$

El Global Alignment Kernel (Cuturi, 2011) es definido como el mínimo suavizado exponencial de todas las distancias de alineamiento

$$K_{GA}(x, y) = \sum_{\pi \in A(n, m)} e^{-D_{x, y}(\pi)} \quad (5)$$

La ecuación (5) se puede reescribir usando la función de similitud k inducida desde la divergencia φ como $k = e^{-\varphi}$:

$$K_{GA}(x, y) = \sum_{\pi \in A(n, m)} \prod_{i=1}^{|\pi|} k(x_{\pi_1(i)}, y_{\pi_2(i)}) \quad (6)$$

Entonces se puede aplicar el GAK kernel para mejorar los resultados obtenidos por k -means, dando pie al algoritmo GAK k -means.

A continuación, una adaptación del algoritmo propuesto en (Dhillon, Guan y Kulis, 2004).

Algorithm 6 GAK k -means

Require: Kernel matrix K

Require: Number of clusters K

Require: Data set $T = \{t_1, t_2, \dots, t_n\}$ con $t_i = (x_i, y_i)$

```

1: Initialize the  $k$  clusters:  $C_1^{(0)}, \dots, C_k^{(0)}$ 
2: Set  $t = 0$ 
3: For each series  $t_i$ , finds its new cluster index as:

```

$$j^*(t_i) = \min_j K_{GA}(t_i, m_j)$$

```

4: Compute the uptades clusters as:

```

$$C_j^{t+1} = \{t_i : j^*(t_i) = j\}$$

```

5: If not converged, set  $t = t + 1$  and go to Step 3; Otherwise, stop.
6: return  $C_1, \dots, C_k$ : partitioning of the points

```

IV. EJEMPLO DE APLICACIÓN: CLÚSTERING SOBRE LOS COMPONENTES DEL IPC

El S&P/BMV IPC es el índice bursátil más importante de la Bolsa Mexicana de Valores (BMV). Es un indicador de la estabilidad del mercado financiero mexicano. Es un índice ponderado, y tiene como componentes a las 35 empresas más grandes de México, ya sea por capitalización, solvencia y/o estabilidad.

Se creó una base de datos con las observaciones diarias de los precios al cierre de las acciones de las 35 empresas que componen el IPC, en el periodo 1/enero/2019 al 8/junio/2020.

Se aplicaron los métodos Euclidian k -means, Soft DTW k -means, DBA k -means y GAK k -means para ajustar clústers, con $k = 3$. El objetivo es poder segmentar a las series de las compañías en 3 grupos: acciones con un buen desempeño, acciones con un desempeño medio y acciones con un mal desempeño.

Se utilizó el módulo `tslearn` de Python 3, que es un módulo en el que se pueden aplicar varios métodos de machine learning para series de tiempo.

GFINBURO	VESTA	GRUMAB
WALMEX	BOLSAA	PEOLES
GFNORTEO	AC	PINFRA
IENOV	ALSEA	OMAB
CUERVO	AMXL	ALFAA
BIMBOA	LIVEPOLC1	ASURB
BSMXB	BBAJIOO	LABB
GMEXICOB	CEMEXCPO	ORBI
GCARSOA1	KIMBERA	ELEKTRA
Q	GAPB	KOFUBL
ENTERA	MEGACPO	TELEVISACPO
FEMSAUBD	RA	

Cuadro I

CLAVE EN PIZARRA DE CADA EMPRESA EN LA BMV.

En el Cuadro I se encuentran las claves en pizarra de cada empresa que aporta información al IPC en la BMV.

En la Figura 3 se encuentran los resultados de aplicar Euclidian k -means. Se puede notar que el primer clúster está formado por las series de las empresas cuyas acciones han tenido un peor desempeño, pues han bajado su precio notoriamente en los últimos 60 días. En el segundo clúster se encuentran las series de las empresas cuyas acciones han tenido un desempeño promedio. En el tercer clúster se encuentran las series de las empresas cuyas acciones han tenido un mejor desempeño, pues a pesar de que tuvieron un pico a la baja se han recuperado rápidamente.

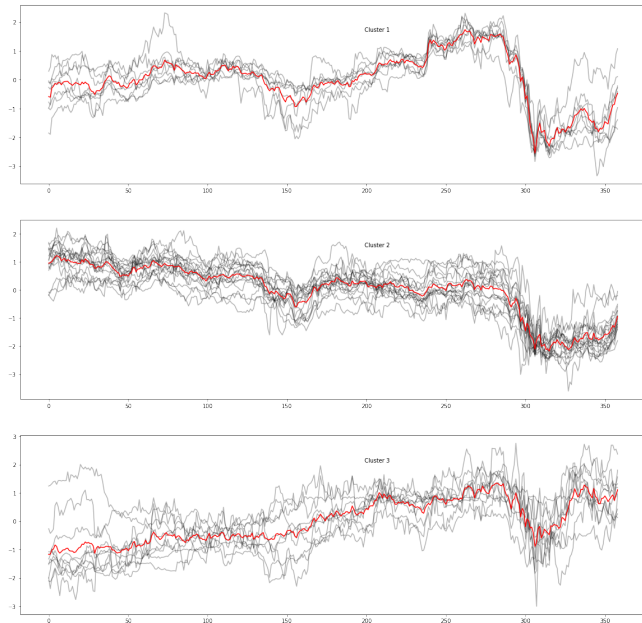


Figura 3. Euclidian k -means

En el Cuadro II se encuentran las empresas con el respectivo clúster asignado, para el caso de Euclidian k -means.

Clave	Clúster	Clave	Clúster
PINFRA	1	CEMEXCPO	2
IENOVA	1	ORBIA	2
OMAB	1	GCARSOA1	2
ASURB	1	KOFUBL	2
GMEXICOB	1	MEGACPO	2
GAPB	1	TELEVISACPO	2
GENTERA	1	FEMSAUBD	2
RA	1	VESTA	3
GFINBURO	2	GRUMAB	3
PEOLES	2	WALMEX	3
GFNORTEO	2	BOLSAA	3
AC	2	CUERVO	3
ALSEA	2	AMXL	3
ALFAA	2	LABB	3
BIMBOA	2	KIMBERA	3
LIVEPOLC1	2	ELEKTRA	3
BSMXB	2	Q	3
BBAJIOO	2		

Cuadro II
RESULTADOS EUCLIDEAN k -MEANS.

En la Figura 4 se encuentran los resultados de aplicar Soft DTW k -means. Nótese que el centroide de cada clúster tiene

menor variabilidad comparado con los centroides encontrados con Euclidian k -means. El primer clúster está formado por las series de las compañías cuyas acciones han tenido un peor desempeño, el segundo clúster está formado por las series de las compañías cuyas acciones han tenido un desempeño medio, y el tercer clúster está formado por las series de las compañías cuyas acciones han tenido un mejor desempeño.

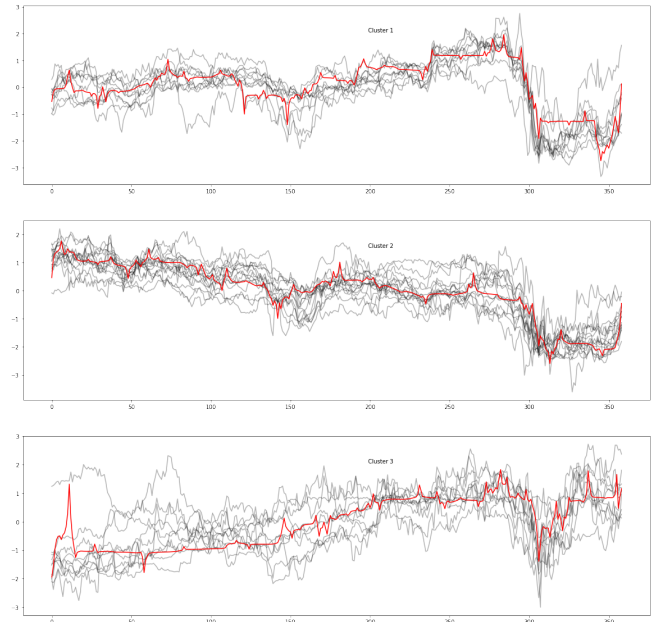


Figura 4. Soft DTW k -means

En el Cuadro III se encuentran las empresas con el respectivo clúster asignado, para el caso de Soft DTW k -means.

Clave	Clúster	Clave	Clúster
GFNORTEO	1	BBAJIOO	2
PINFRA	1	CEMEXCPO	2
IENOVA	1	ORBIA	2
OMAB	1	GCARSOA1	2
AMXL	1	KOFUBL	2
ASURB	1	MEGACPO	2
GAPB	1	TELEVISACPO	2
GENTERA	1	VESTA	3
FEMSAUBD	1	GRUMAB	3
RA	1	WALMEX	3
GFINBURO	2	BOLSAA	3
PEOLES	2	CUERVO	3
AC	2	LABB	3
ALSEA	2	GMEXICOB	3
ALFAA	2	KIMBERA	3
BIMBOA	2	ELEKTRA	3
LIVEPOLC1	2	Q	3
BSMXB	2		

Cuadro III
RESULTADOS SOFT DTW k -MEANS.

En la Figura 5 se encuentran los resultados de aplicar DBA k -means. Los centroides de los clústers son muy parecidos a los encontrados por Soft DTW k -means. El primer clúster está formado por las series de las compañías cuyas acciones han

tenido un peor desempeño, el segundo clúster está formado por las series de las compañías cuyas acciones han tenido un desempeño medio, y el tercer clúster está formado por las series de las compañías cuyas acciones han tenido un mejor desempeño.

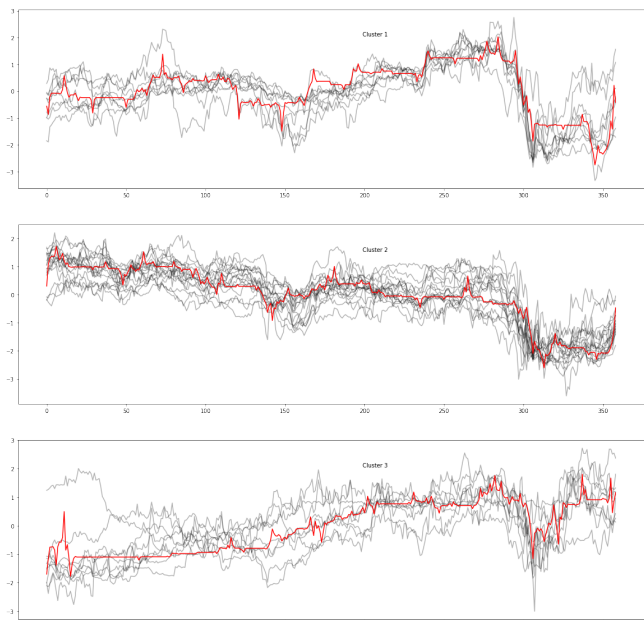


Figura 5. DBA *k*-means

En el Cuadro IV se encuentran las empresas con el respectivo clúster asignado, para el caso de DBA *k*-means.

Clave	Clúster	Clave	Clúster
PINFRA	1	BBAJIOO	2
IENOVA	1	CEMEXCPO	2
OMAB	1	ORBIA	2
AMXL	1	GCARSOA1	2
ASURB	1	KOFUBL	2
GMEXICOB	1	MEGACPO	2
GAPB	1	TELEVISACPO	2
ENTERA	1	FEMSAUBD	3
RA	1	VESTA	3
GFINBURO	2	GRUMAB	3
PEOLES	2	WALMEX	3
GFNORTEO	2	BOLSAA	3
AC	2	CUERVO	3
ALSEA	2	LABB	3
ALFAA	2	KIMBERA	3
BIMBOA	2	ELEKTRA	3
LIVEPOLC1	2	Q	3
BSMXB	2		

Cuadro IV
RESULTADOS DBA *k*-MEANS.

En la Figura 6 se encuentran los resultados de aplicar GAK *k*-means. A diferencia de los métodos de clústering anteriores, en la gráfica no se encuentra los centroides de los clústers. El orden de los clústers también es diferente, pues el primer clúster está formado por las series de las compañías cuyas acciones han tenido un mejor desempeño, el segundo

clúster está formado por las series de las compañías cuyas acciones han tenido un desempeño medio, y el tercer clúster está formado por las series de las compañías cuyas acciones han tenido un peor desempeño.

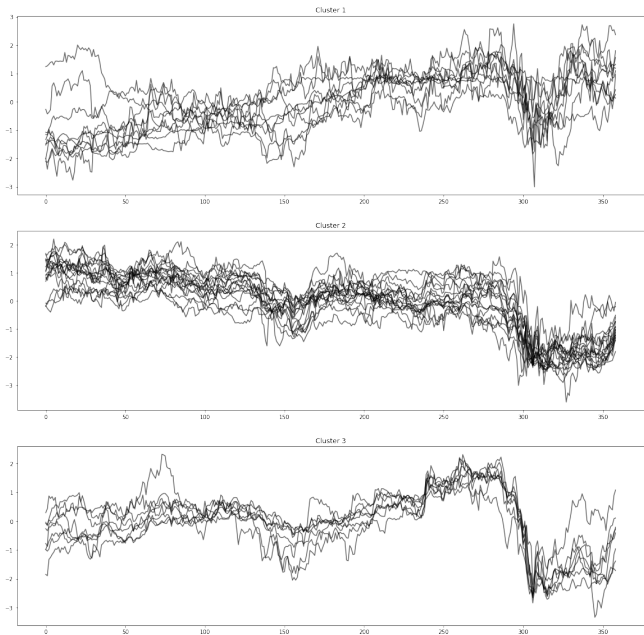


Figura 6. GAK *k*-means

En el Cuadro V se encuentran las empresas con el respectivo clúster asignado, para el caso de GAK *k*-means.

Clave	Clúster	Clave	Clúster
GRUMAB	1	IENOVA	2
WALMEX	1	ALSEA	2
PEOLES	1	AMXL	2
PINFRA	1	KIMBERA	2
CUERVO	1	KOFUBL	2
ALFAA	1	MEGACPO	2
LIVEPOLC1	1	TELEVISACPO	2
BSMXB	1	FEMSAUBD	2
BBAJIOO	1	RA	2
LABB	1	OMAB	3
CEMEXCPO	1	BIMBOA	3
ELEKTRA	1	ASURB	3
Q	1	GMEXICOB	3
GFINBURO	2	ORBIA	3
VESTA	2	GCARSOA1	3
BOLSAA	2	GAPB	3
GFNORTEO	2	ENTERA	3
AC	2		

Cuadro V
RESULTADOS GAK *k*-MEANS.

En general, los resultados encontrados para todos los métodos son muy parecidos, aunque el método GAK *k*-means es el que obtiene resultados un poco más diferentes de los otros tres.

Si se consideran las compañías que aparecen en los mismos clústers para todos los métodos, entonces se pueden agrupar como en el Cuadro VI. Entonces:

- Las compañías mexicanas cuyas acciones han tenido un mejor desempeño son RUMAB(GRUMA), WALMEX(Walmart de México), CUERVO (Casa Cuervo), LABB (GenomaLab), ELEKTRA(Elektra), Q(Quáalitas).
- Las compañías mexicanas cuyas acciones han tenido un desempeño medio son GFINBURO(Grupo Financiero Inbursa), AC (Arca Continental), ALSEA (Alsea), KOFUBL (Coca-Cola Femsa), MEGACPO (Megacable Holdings) y TELVESACPO (Grupo Televisa).
- Las compañías mexicanas cuyas acciones han tenido un peor desempeño son OMAB(Grupo Aeroportuario del Centro Norte), ASURB(Grupo Aeroportuario del Sureste), GAPB(Grupo Aeroportuario del Pacífico), GENTERA(Compartamos Banco).

Bueno	Medio	Malo
GRUMA	GFINBURO	OMAB
WALMEX	AC	ASURB
CUERVO	ALSEA	GAPB
LABB	KOFUBL	GENTERA
ELEKTRA	MEGACPO	
Q	TELVESACPO	

Cuadro VI

COMPAÑÍAS QUE APARECEN EN LOS MISMOS CLÚSTERS PARA TODOS LOS MÉTODOS.

V. CONCLUSIONES

Existen diferentes métodos para clústering de series de tiempo, y en general, se pueden agrupar por basados en disimilaridades, basados en modelos y basados en dependencias.

DTW permite encontrar la distancia entre dos series contemplando la dependencia temporal entre ellas, y a través de una alineación no-lineal óptima mínima.

Los métodos Euclidian k -means y Soft DTW k -means son bastante similares, con la diferencia de que en Soft DTW k -means se utiliza cierta ponderación y la distancia DTW. DBA k -means utiliza la distancia DTW y sigue un esquema Expectation-Maximization para encontrar una serie \bar{T} promedio. GAK k -means utiliza Global Alignment Kernel para la estimación de las distancias.

Los resultados encontrados para todos los métodos son muy parecidos, aunque el método GAK k -means es el que obtiene resultados un poco más diferentes de los otros tres.

Las compañías mexicanas cuyas acciones han tenido mejor desempeño son GRUMAB(GRUMA), WALMEX(Walmart de México), CUERVO (Casa Cuervo), LABB (GenomaLab), ELEKTRA(Elektra), Q(Quáalitas).

Las compañías mexicanas cuyas acciones han tenido un desempeño medio son GFINBURO(Grupo Financiero Inbursa), AC (Arca Continental), ALSEA (Alsea), KOFUBL (Coca-Cola Femsa), MEGACPO (Megacable Holdings) y TELVESACPO (Grupo Televisa).

Las compañías mexicanas cuyas acciones han tenido un peor desempeño son OMAB(Grupo Aeroportuario del Centro Norte), ASURB(Grupo Aeroportuario del Sureste), GAPB(Grupo Aeroportuario del Pacífico), GENTERA(Compartamos Banco).

REFERENCIAS

- [1] Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering—a decade review. *Information Systems*, 53, 16-38.
- [2] Cuturi, M. (2011). Fast global alignment kernels. In *Proceedings of the 28th international conference on machine learning (ICML-11)* (pp. 929-936).
- [3] Dhillon, I. S., Guan, Y., & Kulis, B. (2004, August). Kernel k -means: spectral clustering and normalized cuts. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 551-556).
- [4] Kaufman, L. R., & Rousseeuw, P. PJ (1990) *Finding groups in data: An introduction to cluster analysis*. Hoboken NJ John Wiley & Sons Inc, 725.
- [5] Liao, T. W. (2005). Clustering of time series data—a survey. *Pattern recognition*, 38(11), 1857-1874.
- [6] Petitjean, F., Forestier, G., Webb, G. I., Nicholson, A. E., Chen, Y., & Keogh, E. (2014, December). Dynamic time warping averaging of time series allows faster and more accurate classification. In *2014 IEEE international conference on data mining* (pp. 470-479). IEEE.
- [7] Petitjean, F., Ketterlin, A., & Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44(3), 678-693.
- [8] Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1), 43-49.
- [9] Vo, V., Luo, J., & Vo, B. (2016). Time series trend analysis based on K-means and support vector machine. *Computing and Informatics*, 35(1), 111-127.
- [10] Zhang, Z., Tang, P., Huo, L., & Zhou, Z. (2014). MODIS NDVI time series clustering under dynamic time warping. *International Journal of Wavelets, Multiresolution and Information Processing*, 12(05), 1461011.