



Universidad Central de Venezuela
Facultad de Ciencias
Maestría en Modelos Aleatorios



Clasificación de las series temporales basado en la Aproximación Agregada Simbólica usando el método de Detección del Punto de Cambio Bayesiano (DPCB-AAS).

Liseth Daniela Ramirez Polidor, Tesis de la Universidad Central de Venezuela, Facultad de Ciencias.
liseth.estadistica@gmail.com

Mayo del 2023, Caracas , Venezuela.

Resumen.

El modelo DPCB-AAS permite definir y proceder en forma expedita clasificar series temporales, tal es el caso de los instrumentos financieros. Para ello, proponemos el uso del modelo de Aproximación Agregada Simbólica usando el Método de Detección de Punto de Cambio Bayesiano cuyo contenido y singularidades se presentarán con rigurosidad hasta convertirlo en una herramienta de fácil manejo para el usuario, reduciendo y categorizando el patrón obtenido para cada serie y agrupar por similitud en la cadena de caracteres resultante.

Palabras Claves: Aproximación Agregada simbólica, Detección del Punto de Cambio Bayesiano, DPCB-AAS, ruptura, Cadena de caracteres, Recorrido, longitud, reducción, Distancia de AAS.

Introducción.

En el mundo financiero, en particular en el área del *trading*, se trata de conocer y manejar los instrumentos financieros, vistos como series temporales (los precios de derivados, bonos, acciones, futuros, etc.), y a su vez conocer las relaciones que existen entre ellos además de las actividades comerciales que tienen individualmente, ya que al invertir en ellas, los usuarios tratan de buscar la relación entre mayor ganancia y menor riesgo, lo que hace poco factible la comparación, ya que es inevitable pensar que a medida que transcurre el tiempo va en aumento el número de instrumentos reflejado en históricos con periodicidades pequeñas (nano-segundos, segundos, minutos, entre otros) introduciendo severos problemas computacionales.

Todo esto lleva a buscar una manera de agrupar los instrumentos financieros por medio de su comportamiento histórico, para llegar a la obtención de mayor cantidad de información sobre el mercado y tratar sistemáticamente cada instrumento financiero, utilizando un método que permita un ahorro significativo en el tiempo de cómputo sin sacrificios en la calidad de la predicción. Además de lograr conocer si determinados instrumentos (así sean de grupos económicos diferentes) poseen similitud.

Planteamiento y métodos.

Para poder clasificar grandes cantidades de instrumentos financieros con el fin de estudiarlos y trabajarlos de forma grupal aprovechando el desarrollo computacional contemporáneos se combinaron dos métodos que permiten convertir la serie original en una función de segmentos y tomar el punto medio por segmento para reducirlo y transformarlo en una cadena de caracteres y compilar todas las cadenas generadas.

A continuación en la tabla 1, se muestran a grandes rasgos los pasos y métodos a utilizar para crear el DPCB-AAS.

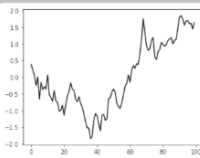
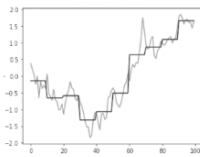
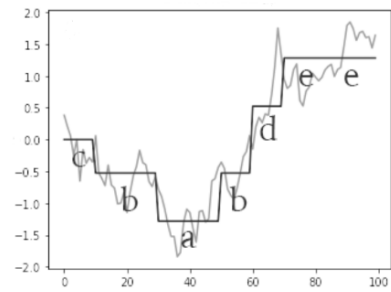
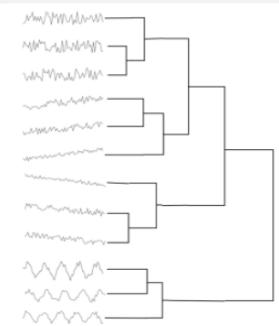
Detección del Punto de Cambio Bayesiano (DPCB)	Aproximación Agregada Simbólica (AAS)	Distancia AAS
Segmentar la serie temporal por el cambio de parámetros de la distribución en un segmento determinado, siendo la detección del punto de ruptura bayesiano nuestro principal interés.	Estandarizar los valores nos permite representar dichos tramos por símbolos. Por convención, la letra 'a' representará el valor mínimo estandarizado de la serie y la 'z' el valor máximo, utilizando AAS.	Es una medida de distancia, con el mismo concepto de la distancia euclidiana, la cual el cálculo se hará en función del orden que tenga la cadena de caracteres.
<p>Detección del Punto de Cambio Bayesiano (DPCB)</p> <p>1 </p> <p>2 </p>	<p>Aproximación Agregada Simbólica (AAS)</p> <p>3 </p>	<p>Distancia de AAS.</p> <p>4 </p>

Tabla 1. Pasos Generales de la aplicación del modelo.

1. Detección del punto de cambio bayesiano (DPCB).

Este método permite detectar los cambios significativos que pueden ocurrir en una variable en el tiempo sobre los parámetros generativos en secuencias de datos. Para ello se calcula a partir de la distribución posterior las ubicaciones de los puntos de cambio y de esta manera predecir si existe una ruptura en el siguiente dato en la secuencia dado el segmento previo.

¹Supongamos una secuencia de observaciones X_1, X_2, \dots, X_T , este se puede particionar en segmentos los cuales estarán en función de los cambios que pueden ocurrir sobre la serie, dichas rupturas se representan con la letra g , siendo la variable aleatoria independiente e idénticamente distribuida (v.a.i.i.d) con una distribución de probabilidad $P(X_i|\eta_g)$, siendo η_g los parámetros generativos en cada partición g , dicha probabilidad, denotada como $P(g)$, a priori indicará si existe un cambio abrupto para los valores posteriores. Cada partición en la serie se representa como $X_{a:b}$, siendo "a" y "b" el largo de una partición que no posee cambios abruptos (no posee cambio en los parámetros).

¹ Adams & MacKay. Bayesian Online Changepoint Detection. Cavendish Laboratory Cambridge. (2007).

La probabilidad posteriori será inferir si existe un cambio en los parámetros de la variable dado la longitud que tenga el segmento previo, el cual se representará como r_t y la notación $X_{t(r)}$ como la muestra de observaciones asociada al recorrido r_t , donde $r_t=0$, si y sólo si ocurre un cambio en los parámetros de la serie, por lo tanto un cambio abrupto sobre el segmento previo.

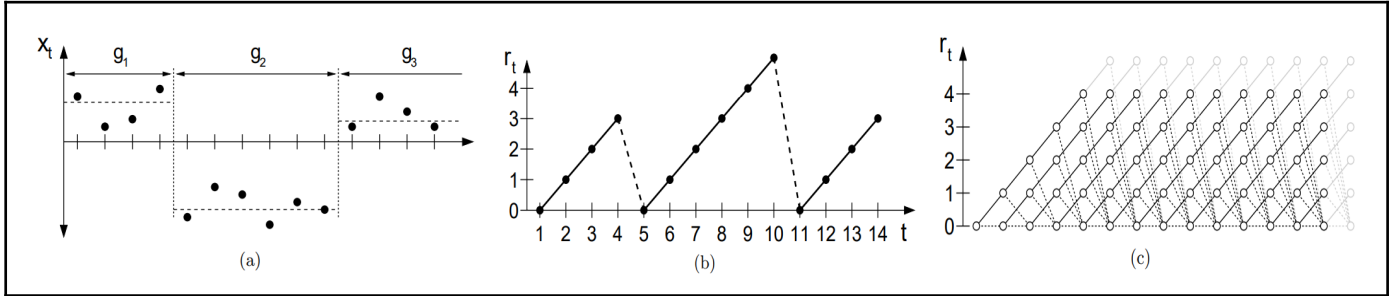


Figura 1. Ejemplo teórico de la detección del punto de cambio bayesiano.

En la figura 1, ²en el gráfico (a) se muestra la serie temporal con los puntos de corte determinados, representando g_1, g_2 y g_3 las particiones donde la v.a.i.i.d, en el (b) se presenta el recorrido, el cual será $r_t = 0$ cuando se ubica el cambio de parámetro en la serie y en el (c) Se visualiza el recorrido punto a punto para determinar si existe un cambio abrupto sobre la serie y que recorrido tiene mayor probabilidad de ocurrencia.

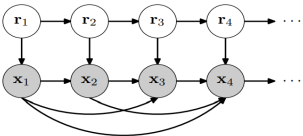
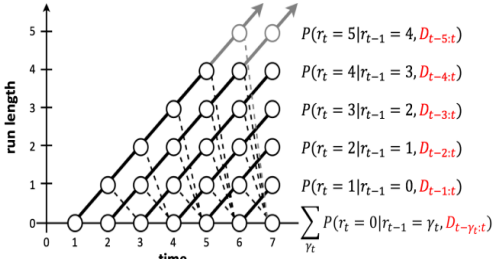
<p>(a) Función a priori del recorrido, llamada Función de riesgo,</p> $P(r_t r_{t-1}) = \begin{cases} H(r_{t-1} + 1), & \text{si } r_t = 0 \\ 1 - H(r_{t-1} + 1), & \text{si } r_t = r_{t-1} + 1 \\ 0, & \text{c.o.o} \end{cases}$ $H(\tau) = \frac{P(g=\tau)}{\sum_{t=\tau}^{\infty} P(g=\tau)}$ <p>La longitud (τ) de ejecución puede aumentar o restablecer a cero, tal como se muestra en la figura 1.b y posible resultado de la función de riesgo.</p>	<p>(b) Para hacer el cálculo del recorrido, este depende de los valores previos de x, y hallar la distribución a posteriori.</p> $P(r_t x_{1:t}) = \frac{P(r_t, x_{1:t})}{P(x_{1:t})}$ $P(x_{t+1} x_{1:t}) = \sum_{r_t} P(x_{t+1} r_t, x_{1:t}^{(r)}) P(r_t x_{1:t})$ 	 <p>(c) Se registran todas las posibles combinaciones recursivas del recorrido y que a su vez tienen mayor probabilidad de ocurrencia. Donde $D = \{x_t\}_{t=1}^T$ (D conjunto de observaciones i.i.d) como se muestra en la figura 1.c.</p>
--	--	---

Tabla 2. Generalización de los pasos a desarrollar con la distribución a priori y posteriori.

A continuación se muestra la notación general del cálculo de la distribución del DPCB

² Adams & MacKay. Bayesian Online Changepoint Detection. Cavendish Laboratory Cambridge. (2007).

1. Inicio $P(r_0) = \{1 \text{ si } t = 1, P(r_0 = \tau) \text{ c. o. c}\}$	Para realizar lo antes expuesto en la tabla 2 se debe tener en cuenta las posibles combinaciones recursivas del recorrido esto permitirá inferir si el punto entrante será un punto de cambio.
2. Se observa un nuevo dato x_t	Una vez iniciado con la probabilidad de la variable dado el recorrido se mostrará el dato nuevo.
3. Evaluar la probabilidad de predicción $\pi_{1:t} = P(x_t r_t, x_{1:t})$	La evaluación del dato nuevo estará condicionada por el recorrido en conjunto con los datos previos.
4. Cálculo de la probabilidad de recurrencia y de cambio. $\begin{cases} P(r_t = r_{t-1} + 1, x_{1:t}) = P(r_{t-1}, x_{1:t-1})\pi_{1:t}(1 - H(r_{t-1})) \\ P(r_t = 0, x_{1:t}) = \sum_{r_{t-1}} P(r_{t-1}, x_{1:t-1})\pi_{1:t}H(r_{t-1}) \end{cases}$	Ambas probabilidades del conjunto entre el recorrido y el segmento se calculan y dependiendo de cual posea la mayor probabilidad se determinará la base para inferir si el siguiente dato es una ruptura.
5. Cálculo de la probabilidad del segmento de la variable x. $P(x_{1:t}) = \sum_{r_t} P(r_t, x_{1:t})$	Esta es la probabilidad total de la probabilidad conjunta entre el recorrido y la variable hasta t.
6. Determinar la distribución del largo del segmento. $P(r_t) = P(r_t, x_{1:t})/P(x_{1:t})$	Teniendo definido 4 y 5, podremos determinar $P(r_t)$.
7. Realizar predicción. $P(x_{t+1} x_{1:t}) = \sum_{r_t} P(x_{t+1} x_{1:t}, r_t)P(r_t x_{1:t})$	Ya teniendo una base robusta con el punto 6, se predice el dato posterior x_{t+1} .
8. Retornar en el punto 2.	Se iteran los pasos con el dato entrante.

Tabla 3. ³ Pasos Generales de la DPCB.

³ Bayesian Online Changepoint Detection. 13 August 2019. Gregory Gundersen. <https://gregorygundersen.com/blog/2019/08/13/bocd/>



Distribución asociada.

En los apartados previos se ha podido observar la estructura general del cálculo asociado a la probabilidad y distribución para hallar la existencia significativa o no de una ruptura por segmento, pero no se menciona que distribución se asocia para inferir el cambio en la serie, para ello es necesario tener claro la conjunción y la familia exponencial, esta será base para inferir por medio de la distribución normal.

⁴Para ello es necesario entender que para poder determinar los segmentos dentro de una serie temporal se necesita definir sus respectivos parámetros, en este caso asociado a la distribución normal, la cual parte de la premisa de que cada segmento es localmente estacionario y posee una media y varianza desconocida. Para estimarlos en inferencia bayesiana es necesario conocer su previa conjugada, ya que la conjunta permitirá evaluar el valor actual respecto a los valores previos y normalizar, ya que se ⁵presta bien en análisis secuencial, dado que:

A medida que el modelo ve más datos, el posterior coloca más densidad alrededor del verdadero sesgo del proceso generativo. Y después de cada punto de datos, el posterior se convierte en el nuevo anterior.

Tal como se puede observar el uso de la probabilidad conjunta en el paso 3 y 4 en la DPCB (tabla 3), y la participación de la familia exponencial, dará la condición suficiente de pertenencia del dato actual sobre los parámetros asociados al segmento previo.

Donde $D = \{x_t\}_{t=1}^T$, con μ y $\lambda = \frac{1}{\sigma^2}$ desconocidos $\rightarrow N(D|\mu, \lambda)$

Siendo la probabilidad a priori y posterior, la siguiente expresión:

$$\begin{aligned} p(D|\mu, \lambda) &= \prod_{n=1}^N p(x_n|\mu, \lambda) \\ &= \prod_{n=1}^N \left[\left(\frac{\lambda}{2\pi} \right)^{\frac{1}{2}} e^{-\frac{\lambda}{2}(x_n-\mu)^2} \right] \\ &= \left(\frac{\lambda}{2\pi} \right)^{\frac{1}{2}} e^{-\frac{\lambda}{2} \sum_{n=1}^N (x_n-\mu)^2} \propto \lambda^{\frac{1}{2}} e^{-\frac{\lambda}{2} \sum_{n=1}^N (x_n-\mu)^2} \end{aligned}$$

Tabla 4. Tratamiento de la distribución normal sobre la data secuencial.

⁴ Machine Learning A Probabilistic Perspective. 2012. Kevin P. Murphy

⁵ Bayesian Online Changepoint Detection. 13 August 2019. Gregory Gundersen. <https://gregorygundersen.com/blog/2019/08/13/bocd/>

2. Agregación Simbólica Aproximada (AAS).

Es una representación simbólica basada en datos de series de tiempo que codifica distribuciones típicas de puntos de datos. Para ello se necesita algún otro método que será la base para detectar los puntos de cambio en el tiempo, supóngase una serie de tiempo, tal como se muestra en la siguiente representación gráfica y como se puede observar a simple vista la serie posee un comportamiento completamente aleatorio y que no son distinguibles puntos de corte. Ahora en base a un modelo de detección de puntos de corte el AAS permitirá reducir la serie a una cadena de caracteres, siendo valioso para comparar con otras series y a su vez detectar puntos anómalos sobre ella. Ahora suponga el hecho de que se detectaron los segmentos y se aplica el AAS y arroja la siguiente gráfica.

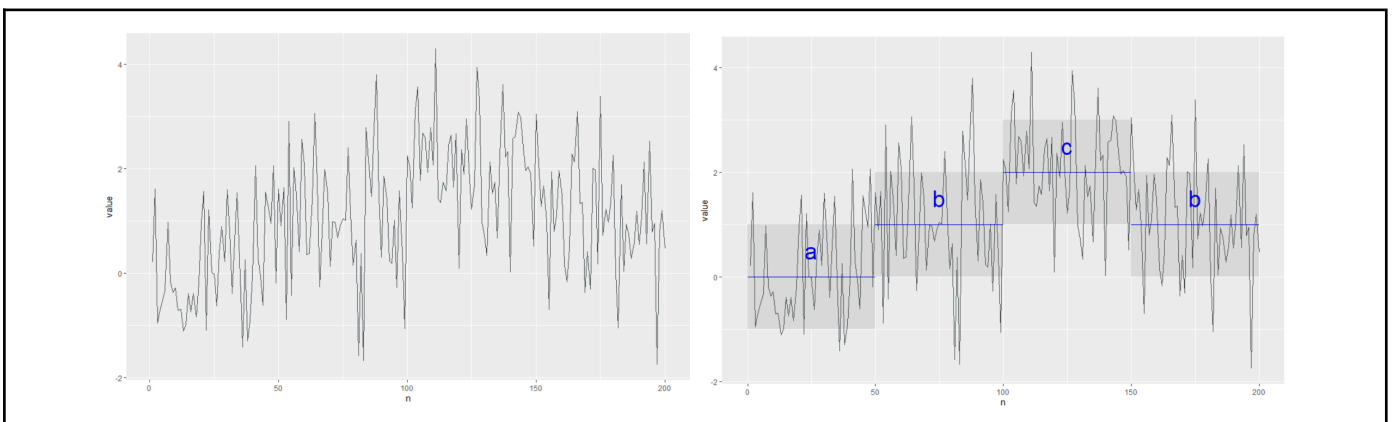


Figura 2. Ejemplo Teórico del AAS.

En base al ejemplo que se presenta en la figura 2, ahora la serie estará representada por cuatro valores categóricos, por lo cual se puede observar la reducción de la serie de 200 observaciones a cuatro caracteres “abcb”.

Tabla 5. Notación general del AAS

Notación	Definición
C	Los valores que puede tomar la serie de tiempo $C = c_1, c_2, c_3, \dots, c_n$
\bar{C}	Será el valor que represente la reducción de la serie, la cual será explicado con detalle más adelante en el presente documento. $\bar{C} = \bar{c}_1, \bar{c}_2, \bar{c}_3, \dots, \bar{c}_n$
\hat{C}	Va a ser la representación simbólica de cada valor que arroja \bar{C} , siendo $C = c_1, c_2, \dots, c_n$
w	Es el número de segmentos obtenidos por medio de \bar{C} .
α	Tamaño del alfabeto (letras a utilizar para representar la cadena de caracteres de cada serie de tiempo).

Ahora (considerando la información en la tabla 1) hay que tener en cuenta que $w < n$, siendo ‘n’ la cantidad de observaciones que tiene la serie de tiempo.

⁶ El modelo base para realizar la reducción de la serie será detección de puntos de cambio bayesianos en series temporales.

Es necesario que los caracteres que representa cada segmento sean equiprobables, esto se obtiene ya que las series de tiempo están normalizadas (tienen una distribución gaussiana), para ello se define:

Punto de ruptura: Es el punto de corte que representa la diferencia entre caracteres o símbolos, el cual se deriva de la separación de cuantiles partiendo de la normalización de las series, este se representará por β_i . Ahora para poder hallar β es necesario conocer la cantidad de particiones que se realizará sobre la data, estas se representan con la letra α .

Ejemplo. Se tiene un trimestre de data con periodicidad diaria y se particiona la data mes a mes, implica que de 365 observaciones la reducción de la serie de tiempo estará representada por 3 observaciones (implica $\alpha = 3$). Hay que tener en cuenta que al ser tres particiones la cantidad de β_i donde $\forall i = 1, 2, \dots, n - 1$, implica dos β_i

	β_1	β_2
--	-----------	-----------

Donde $\beta_1 = \frac{1}{3} = 0,33 \Rightarrow \int_{-\infty}^{\beta_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 0,33 \Rightarrow \beta_1 = -0,43$. El valor de -0.43 se puede visualizar desde la tabla de la distribución normal. Para el caso de β_2 sería 0.43, ya que al provenir de una serie con distribución normal y al ser una distribución simétrica $\beta_2 = 0,43$.

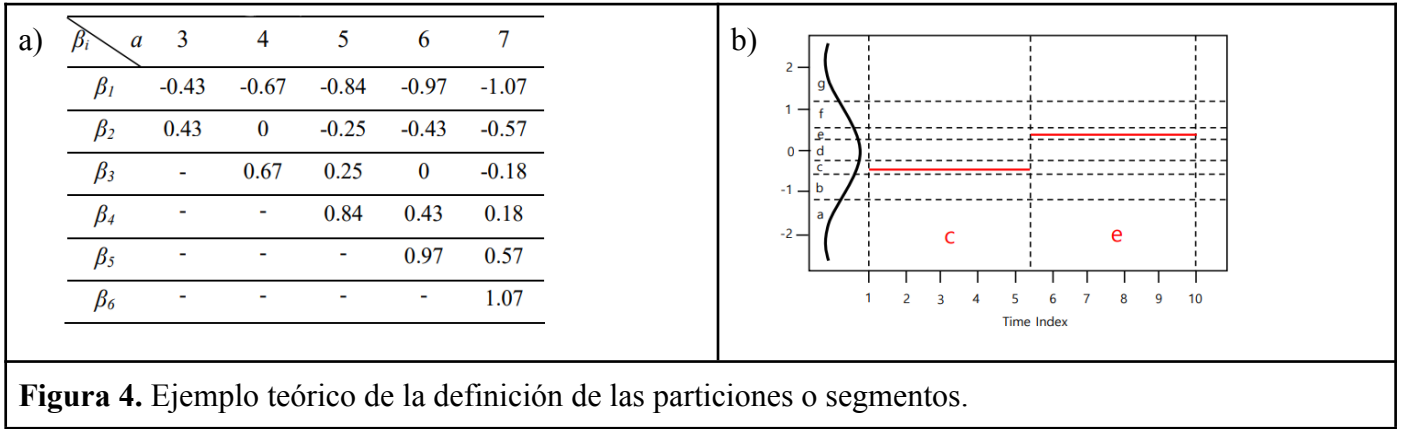
z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.0	0.0013	0.0010	0.0007	0.0005	0.0003	0.0002	0.0002	0.0001	0.0001	0.0000
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0046	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3373	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483

Tal como se muestra en la figura asociada a la distribución normal y dependiendo de la cantidad de letras que estarán asociada a las series temporales, por lo que independientemente de los valores asociados a la serie, los valores obtenidos de la partición se mantendrán siempre iguales.

Figura 3. Tabla de la distribución normal acumulada.

Para el caso de más puntos de ruptura el cálculo se realiza de la siguiente manera:

$$\beta_1 = \frac{1}{\alpha}, \beta_2 = \frac{2}{\alpha}, \beta_3 = \frac{3}{\alpha}, \beta_4 = \frac{4}{\alpha}, \dots$$



Ahora como se muestra en la tabla a, se puede observar cuales serian los valores si se tienen entre 3 a 7 letras o caracteres asociada a la reducción de la serie por medio del AAS, por ejemplo para el caso de tres letras quedaría expresado de la siguiente manera: $a \in (-\infty, -0.43]$, $b \in (-0.43, 0.43]$ y $c \in [0.43, +\infty)$. Una vez estén definidas las particiones, tal como se muestra en el gráfico b, los segmentos se verían de la siguiente forma, logrando aclarar por cuál región pertenece cada valor de la serie temporal.

2.1 Clasificación- Distancia de AAS.

Para poder clasificar las cadenas de caracteres producida por cada instrumento financiero se aplica una medida muy similar a la distancia Euclidiana, la cual dada dos series de tiempo Q y C de la misma longitud n ahora si transformamos los segmentos obtenidas por medio del DPCB de la variable como \bar{Q} y \bar{C} . Si se transforman más los datos en la representación simbólica se puede definir una función de una mínima distancia entre la serie de tiempo original de dos palabras, esta se expresaría como:

$$D(Q, C) \equiv \sqrt{\sum_{i=1}^n (q_i - c_i)^2} \rightarrow DR(\bar{Q}, \bar{C}) \equiv \sqrt{\frac{n}{w}} \sqrt{\sum_{i=1}^w (\bar{q}_i - \bar{c}_i)^2} \rightarrow MINDIST(\hat{Q}, \hat{C}) \equiv \sqrt{\frac{n}{w}} \sqrt{\sum_{i=1}^w (dist(\hat{q}_i, \hat{c}_i))^2}$$

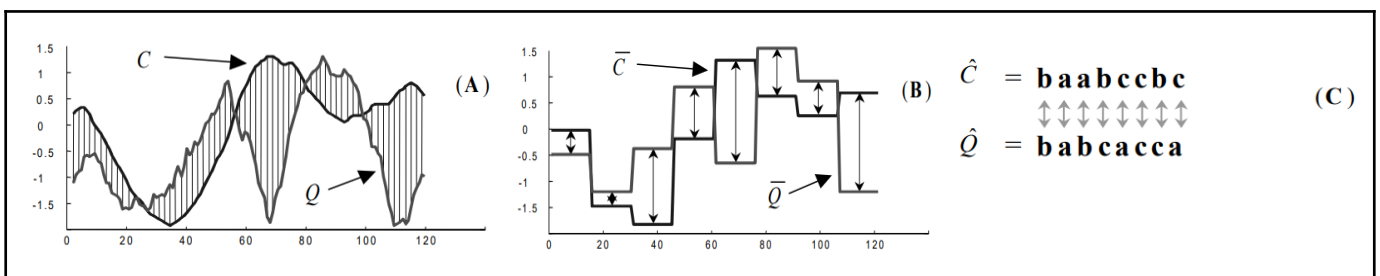
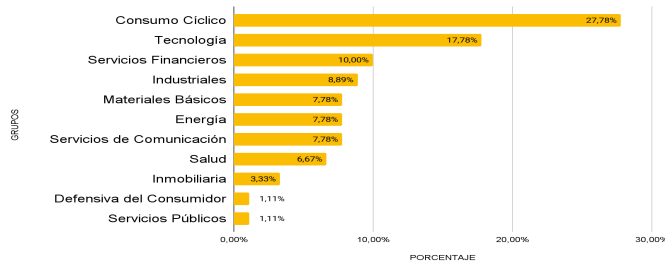


Figura 5. Asociación entre la distancia euclidiana y la distancia de AAS. (A) Como se observa el cálculo de la distancia euclidiana entre dos series temporales. (B) representación del AAS de dos series temporales y (C) Representación en caracteres de las dos series temporales.

A continuación se presentan los resultados del ejemplo práctico aplicado, el cual se trabajó con las series temporales de los precios de una muestra de instrumentos financieros pertenecientes a diferentes sectores económicos, pasando de tener data con periodicidad diaria a quincenal (de 730 observaciones a 104 aproximadamente), como reducción de la serie y base para hallar la cadena de caracteres por instrumento sobre la aplicación del método AAS.

Contexto de la data y Resultados experimentales obtenidos.

Recuento de Sector Económico



Se trabajó con una muestra de 91 instrumentos financieros asociados a diferentes sectores económicos tal como se encuentra distribuido en el gráfico.

La data de dichos instrumentos el periodo de referencia que se trabajó fue desde inicio del 2021 hasta finales del 2022, obtenidas desde <https://finance.yahoo.com/>.

Instrumentos Clasificados según su patrón.

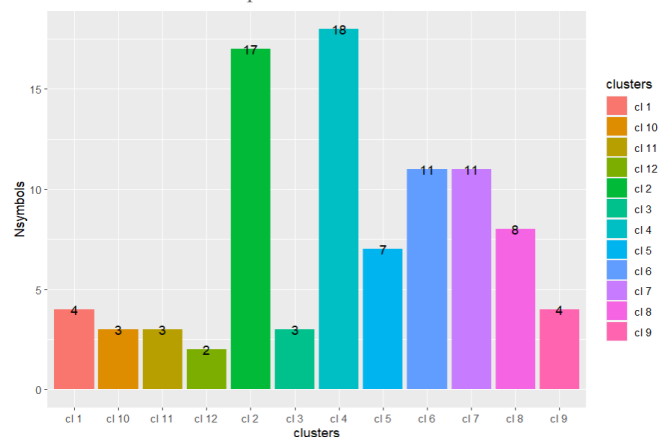


Los patrones de las series entre clusters son diferentes entre ellos, y en términos generales a pesar de poseer diferencia entre los clusters, en ellas se puede apreciar casos de instrumentos con patrones ligeramente diferentes entre ellos, es posible que sea por la volatilidad sobre el instrumentos, siendo la coincidencia entre otros instrumentos interno al cluster los cambios más significativos o abruptos en la línea temporal.

En total se tiene 91 instrumentos financieros en el grupo de acciones, los cluster con mayor número de instrumentos son el cluster 4, 2, 6 y 17, estos cuatro clusters abarcan el 62% de los instrumentos.

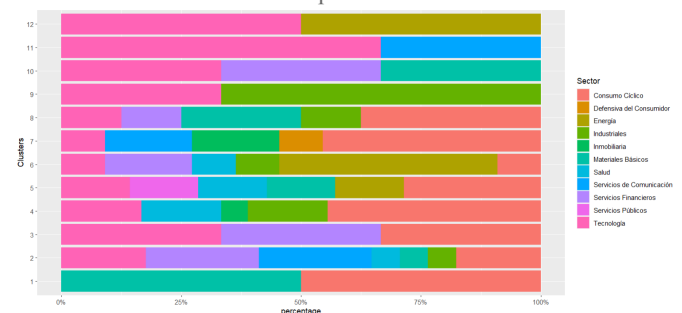
Dichos cluster 2, 4 y 7 tendencias decrecientes, tal como se muestra en el gráfico anterior. Es de destacar que en general el 58% de los cluster poseen una tendencia decreciente, implica que para este periodo la mayoría de las acciones han tendido perdida.

Número de instrumentos por cluster.



En términos generales, todos los clusters poseen instrumentos financieros pertenecientes a diferentes sectores económicos, y a pesar de ello, dichos instrumentos poseen interno a cada cluster un patrón muy similar entre ellos.

Recuento del sector económico por cluster





Conclusiones.

- Se puede visualizar una diferencia entre los patrones obtenidos en los diferentes clusters.
- El 50% de los clusters contienen más de 7 instrumentos con patrones similares.
- El 58% de los grupos de clasificación de patrones, posee un comportamiento decreciente.
- Se puede observar diferencias dentro del cluster, a pesar de ello los instrumentos en cada cluster mantienen la forma que posee el patrón de los mismos.
- De los 11 sectores económicos 4 de ellos abarcan el 65% de los instrumentos financieros de los 91 que abarca la muestra.
- El 91% de los Clusters posee al menos un instrumento asociado a la categoría de tecnología, siendo la categoría con mayor presencia en los diferentes grupos de clasificación, y el segundo sector económico con presencia en los grupos de clasificación de patrones es Consumidor Cíclico conformando el 67% de ellos.

Recomendaciones.

- Es recomendable aplicar estrategias que aborden técnicas diferentes de tal manera de lograr buscar cual se ajusta mejor a determinado tipo de patrones y con ello profundizar o detallar mayor cantidad de indicadores que permita evaluar mejor el desempeño de una estrategia sobre los diferentes cluster (o tipo de patrones).
- Detallar que tipo de relación poseen los instrumentos por cluster y cual es la razón del por qué instrumentos que son miembros de sectores económicos diferentes el patrón es similar a otros instrumentos de otros ramos. Esto va a permitir obtener mejores análisis del mercado debido a que al tener una gran cantidad de instrumentos financieros se hace poco factible estudiarla de forma individual y detallada. Al tener acceso directo a que un determinado instrumento que posee un patrón similar a otros se puede analizar su relación y qué características hacen que posea ese comportamiento.
- Profundizar los parámetros internos al DPCB-AAS, ya que hay instrumentos sobre los clúster que no poseen patrones similares al comportamiento generalizado de los instrumentos que se ubican internos al clúster, al igual que la aplicación de indicadores que proporcionen una visión cuantitativa de la clasificación obtenida.



Bibliografía.

- Ahlame Douzal-Chouakria · José A. Vilar Pierre-François Marteau. *Advanced Analysis and Learning on Temporal Data*. First ECML PKDD Workshop, AALTD 2015, Porto, Portugal, September 11, 2015, Revised Selected Papers (Lecture Notes in Computer Science, 9785). (2015).
- Adams & MacKay. Bayesian Online Changepoint Detection. Cavendish Laboratory Cambridge. (2007).
- Bondu, M. Boullé, B. Grossin. SAXO : An Optimized Data-driven Symbolic Representation of Time Series. Computer Science, The 2013 International Joint Conference on Neural Networks (IJCNN). (2013).
- Boullé, M.: Data grid models for preparation and modeling in supervised learning. In: Guyon, I., Cawley, G., Dror, G., Saffari, A. (eds.) Hands on Pattern Recognition. Microtome, Brookline (2010).
- Chiu, B., Keogh, E., Lonardi, S.: Probabilistic discovery of time series motifs. In: 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, pp. 493–498 (2003).
- Cortes, C., Fisher, K., Pregibon, D., Rogers, A., Smith, F.: Hancock: a language for extracting signatures from data streams. In: 6th International Conference on Knowledge Discovery and Data Mining, Boston, pp. 9–17 (2000).
- Robert, Christian P.; Casella, George. Monte Carlo Statistical Methods. Published by Springer. (1999).
- Lin, J., Keogh, E., Lonardi, S., Chiu, B.: A symbolic representation of time series, with implications for streaming algorithms. In: 8th Workshop on Research Issues in Data Mining and Knowledge Discovery, San Diego (2003).
- Thach Le Nguyen · Severin Gsponer · Iulia Ilie · Georgiana Ifrim. Interpretable Time Series Classification using All-Subsequence Learning and Symbolic Representations in Time and Frequency Domains. Insight Centre for Data Analytics, University College Dublin, Ireland. (2018).
- Ratanamahatana, C., Keogh, E., Bagnall, T., Lonardi, S.: A novel bit level time series representation with implications for similarity search and clustering. In: PAKDD (2005).
- Keogh, E., Lin, J., Fu, A.: HOT SAX: efficiently finding the most unusual time series subsequence. In: 5th IEEE International Conference on Data Mining, Houston, Texas, pp. 226–233 (2005).
- Keogh, E., Wei, L., Xi, X., Lonardi, S., Shieh, J., Sirowy, S.: IntelligentIcons: integrating lite-weight data mining and visualization into GUI operating systems. In: International Conference on Data Mining (2006).
- Lelys Bravo de Guenni, Isabel Llatas Salvador, Maria Eglee Perez Hernandez. Juegos de reglas Bayesianas. Isabel Llatas. XXI ESCUELA VENEZOLANA DE MATEMATICAS (IVIC). (2008).
- Hewa Anuradha Priyadarshani. Bayesian Minimum Description Length Techniques for Multiple Changepoint Detection. Clemson University. (2015).