

# **MINERÍA DE DATOS PARA SERIES TEMPORALES**

Mabel Gonzáles Castellanos  
César Soto Valero

Edición: Liset Ravelo Romero  
Corrección: Roberto Suárez Yera

Mabel Gonzáles Castellanos y César Soto Valero, 2014

Editorial Feijóo, 2014

ISBN: 978-959-250-924-5



Editorial Samuel Feijóo, Universidad Central "Marta Abreu" de Las Villas, Carretera a Camajuaní, km 5 ½, Santa Clara, Villa Clara, Cuba. CP 54830

# SÍNTESIS

---

La minería de datos para series temporales es un área de investigación en pleno auge. Esto es debido a la gran cantidad de datos de este tipo que son generados constantemente para fines diversos, y a la importancia que tiene el análisis de series temporales, por su aplicación en la solución de problemas reales. El análisis de series temporales presenta características distintivas que lo diferencian de otros tipos de datos. Entre estas características se encuentran alta numerosidad y dimensionalidad de las series a analizar y la constante actualización de los datos. Los métodos tradicionales de la minería de datos han tenido que lidiar con estos retos dando lugar a una nueva rama reconocida ya por la comunidad científica. Mostrar una recopilación de los principales métodos es el objetivo principal que persigue este trabajo.

# Tabla de Contenidos

---

<b>INTRODUCCIÓN</b>	<b>6</b>
<b>1 MÉTODOS CLÁSICOS PARA EL TRATAMIENTO DE SERIES TEMPORALES</b>	<b>7</b>
1.1 Series temporales	7
1.2 Análisis de series temporales	8
1.3 Modelos clásicos para el análisis de series temporales	11
1.4 Conclusiones del capítulo	13
<b>2 MINERÍA DE DATOS PARA EL ANÁLISIS DE SERIES TEMPORALES</b>	<b>14</b>
2.1 Minería de datos	14
2.2 Minería de datos para series temporales	15
2.3 Tareas de la minería de datos para series temporales	16
2.3.1 Representación e indexado	16
2.3.2 Clasificación	18
2.3.3 Medidas de similitud	19
2.3.4 Emparejamiento de subsecuencias	21
2.3.5 Segmentación	21
2.3.6 Visualización	22
2.3.7 Descubrimiento de patrones y conglomerados	22
2.4 Principales campos de aplicación y algunos problemas representativos	23
2.4.1 ECG200	24
2.4.2 Gun Point	25
2.4.3 Fifty Words	28
2.5 Conclusiones del capítulo	30
<b>3 LA MÉTRICA ELÁSTICA DTW EN EL CONTEXTO DE LA MINERÍA DE DATOS</b>	<b>31</b>
3.1 Definición de DTW	31
3.2 Cálculo del alineamiento óptimo	34
3.3 Reducción del costo computacional al calcular DTW	35

3.4	Restricciones globales	36
3.4.1	Cotas inferiores	38
3.4.2	Reducción de la numerosidad	42
3.5	Eficacia de la métrica DTW en comparación con la distancia euclidiana	44
<b>REFERENCIAS BIBLIOGRÁFICAS</b>		<b>48</b>

# INTRODUCCIÓN

Las series temporales se obtienen mediante la medición de variables a través del tiempo. Resulta difícil imaginar una rama de la ciencia en la que no aparezcan datos que puedan ser considerados como series temporales, por lo que su procedencia abarca los más diversos dominios. El análisis de series temporales se ha extendido a ramas de la ciencia tan diversas como son la estadística, el procesamiento de señales, reconocimiento de patrones, econometría, matemática financiera, pronóstico climático, electroencefalografía, ingeniería y comunicaciones.

Una serie de tiempo está constituida por observaciones históricas de una o varias variables y, por tanto, sus valores son irrepetibles. Los datos almacenados en forma de series temporales son susceptibles a contener información valiosa para su dominio de procedencia. De ahí parte su utilización tradicional en el pronóstico de valores futuros o con la finalidad de interpretar eventos ocurridos.

El surgimiento de la minería de datos, y una rama de la misma que se encarga exclusivamente de las series temporales, han abierto un área de estudio basada en nuevos enfoques con amplias perspectivas de aplicación. Los métodos utilizados en la minería de datos para series temporales son capaces de caracterizar satisfactoriamente series con características complejas. Estos métodos cubren las limitaciones de las técnicas tradicionales utilizadas en el análisis de series temporales ya que adaptan los conceptos de la minería de datos, para tratar este tipo de series como una clase especial de datos. El reconocimiento de la minería de datos para series temporales como una rama consolidada, avalado por el considerable número de investigaciones realizadas durante la última década, ha motivado esta recopilación de los principales aportes acaecidos.

# 1 Métodos clásicos para el tratamiento de series temporales

El presente capítulo aborda los conceptos básicos sobre series temporales, sus principales características, así como los elementos fundamentales que se han de tener en cuenta durante su análisis. Posteriormente se brinda una breve reseña de los enfoques matemáticos y estadísticos clásicos con que se han tratado tradicionalmente las series temporales, destacando sus limitaciones.

## 1.1 Series temporales

El análisis de datos experimentales, que han sido observados en diferentes instantes de tiempo, da lugar a un nuevo tipo de problema con características únicas en el área de los modelos estadísticos. Resulta usual la correlación existente entre observaciones adyacentes en el tiempo y esta particularidad limita el uso de muchos métodos estadísticos tradicionales que suponen la independencia entre las observaciones.

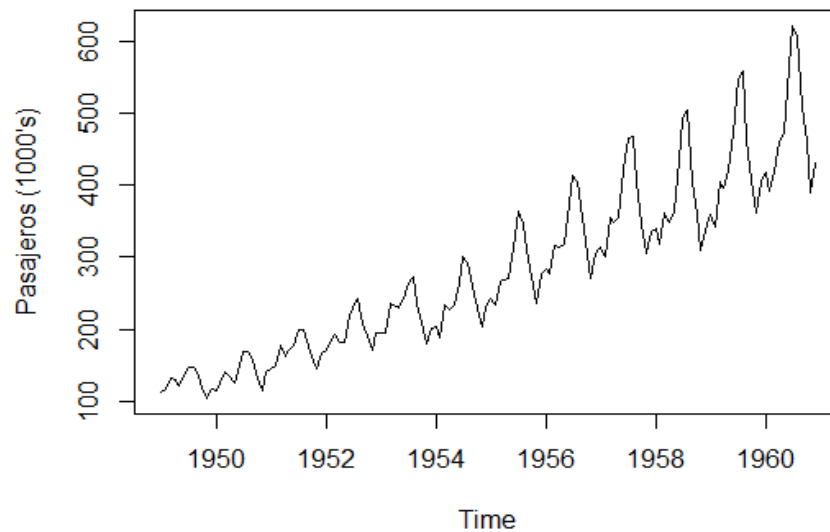
Según Chatfield (2003) una serie temporal consiste en una colección de observaciones realizadas de manera secuencial en el tiempo. Otros autores (Brockwell and Davis, 2002; Wang *et al.*, 2013) ofrecen una definición más rigurosa y la especifican como una secuencia de pares

$$T = [(p_1, t_1), (p_2, t_2), \dots, (p_i, t_i), \dots, (p_n, t_n)]$$

$(t_1 < t_2 < \dots < t_i < \dots < t_n)$ , donde cada  $p_i$  es un punto en el espacio de  $d$ -dimensional al que pertenecen los datos, y cada  $t_i$  es el instante de tiempo al cual se le corresponde la ocurrencia de  $p_i$ .

Una serie de tiempo es considerada continua cuando las observaciones son hechas de forma continua en el tiempo. Si las observaciones tienen lugar solo en momentos específicos entonces es considerada como una serie temporal discreta. La Figura 1.1 muestra la gráfica correspondiente a la cantidad de pasajeros que realizaron vuelos internacionales en una

aerolínea durante el período de tiempo especificado. Como las cantidades son recogidas mensualmente entonces estamos en presencia de una serie de tiempo discreta.



**Figura 1.1. Pasajeros de vuelos internacionales desde los Estados Unidos durante el período de 1949-1960**

## **1.2 Análisis de series temporales**

El análisis de series temporales comprende métodos tanto matemáticos como estadísticos que ayudan a interpretar este tipo de datos teniendo en cuenta las correlaciones temporales existentes en los mismos. La existencia de un orden temporal entre los puntos de datos que forman una serie hace que su análisis sea diferente al de otros problemas que no presentan esta característica.

Existen múltiples objetivos que motivan el análisis de series temporales y están enfocados principalmente tanto a la extracción de información representativa como a la posibilidad de extrapolar y predecir su comportamiento futuro. Estos objetivos en Chatfield (2003), son clasificados como:



- descripción: permite mediante la aplicación de técnicas descriptivas definir las principales propiedades de la serie. La más simple consiste en visualizar gráficamente la serie analizada.
- explicación: en ocasiones es posible usar la variación de unas series de tiempo para explicar la variación en otras. Los modelos de regresión múltiple resultan útiles en esta tarea.
- predicción: es uno de los principales objetivos y consiste en predecir los valores futuros de las series analizadas. Resulta de vital importancia en los análisis económicos e industriales.
- control: este se aplica cuando se desea controlar la calidad de determinado proceso y existen múltiples tipos de procedimientos de control. El control incluye poder tomar medidas oportunas frente al proceso que se está controlando.

El análisis clásico de las series temporales comprende el estudio de cuatro componentes básicos presentes en la serie los cuales son la fuente de su variación. Varios métodos tradicionales están relacionados con la descomposición de la serie temporal en sus componentes aunque se hace notar que este no siempre es el mejor enfoque. Los componentes básicos (Chatfield, 2003) reconocidos en la literatura son:

- tendencia: es una componente de la serie que refleja la evolución a largo plazo del fenómeno observado.
- variación estacional: es el movimiento periódico de corto plazo. Se trata de una componente causal debida a la influencia de ciertos fenómenos que se repiten de manera periódica en el tiempo.
- variación cíclica: además de la variación estacional, debido a alguna otra causa, algunas series exhiben cada cierto período de tiempo variaciones de mayor longitud.
- variación aleatoria: también denominada residuo, no muestra ninguna regularidad y se obtiene una vez eliminadas la tendencia y las variaciones cíclicas de la serie.

Continuando con el ejemplo de la Figura 1.1, en la gráfica se hace visible una tendencia creciente de la serie a lo largo de los años. También es observable la variación estacional

que sufre la serie en el transcurso de los meses pues resulta clara la influencia del período vacacional en el incremento de los vuelos y su disminución en el resto del año.

El objetivo que tradicionalmente ha primado en el análisis de series temporales es el de describir los datos como cierta función en el tiempo que permita analizar con detalles el pasado y hacer predicciones futuras. Esto se logra estableciendo modelos probabilísticos hipotéticos que representen a los datos. En consecuencia, se lleva a cabo el proceso de ajuste, que incluye desde la estimación hasta la predicción, para finalmente determinar un modelo satisfactorio.

Los modelos de series temporales deben considerar la naturaleza del fenómeno que describen y determinar los factores que pueden ser incluidos en cada modelo. Por ejemplo, en muchas series económicas es indispensable considerar los efectos estacionales de la serie. Si esto no se toma en cuenta, los modelos obtenidos no serán los apropiados.

Los métodos utilizados en el análisis de series temporales son típicamente divididos en categorías: dominio de la frecuencia (Brockwell and Davis, 2002) y dominio del tiempo (Shumway and Stoffer, 2011). El primero se basa en la función de densidad espectral y el segundo en la función de autocorrelación. Además, las técnicas de análisis de series temporales pueden ser divididas según sus métodos en paramétricas y no paramétricas (Brockwell and Davis, 2002). Los enfoques paramétricos asumen que la estacionalidad fundamental del proceso estocástico tiene cierta estructura la cual puede ser descrita usando un reducido número de parámetros (por ejemplo, usando autorregresión o corrimiento de medias). En estos enfoques, el objetivo es estimar los parámetros del modelo que mejor describen el proceso estocástico. Por el contrario, los enfoques no paramétricos estiman explícitamente la covarianza o el espectro del proceso sin asumir que este tenga alguna estructura en particular. Adicionalmente, otras clasificaciones han sido creadas para describir los modelos, algunas de ellas son: lineales y no lineales, univariados y multivariados.

### 1.3 Modelos clásicos para el análisis de series temporales

El modelo autorregresivo integrado de media móvil o ARIMA (acrónimo del inglés *Autoregressive Integrated Moving Average*) (Cowpertwait and Metcalfe, 2009) es un modelo paramétrico perteneciente al dominio del tiempo que utiliza variaciones y regresiones de datos estadísticos con el fin de encontrar patrones para efectuar su predicción. Aunque fue desarrollado a finales de los sesenta del pasado siglo, Box y Jenkins (Box and Jenkins, 1976) lo sistematizaron en 1976, convirtiéndolo en una poderosa herramienta para el análisis de series temporales, capaz de modelar una amplia gama de series.

El método ARIMA está limitado por los requerimientos de estacionariedad e invertibilidad del modelo estimado (Camones, 2002), el sistema generador de dicha serie debe ser también invariante y estable. Además, los residuales (las diferencias entre la serie de tiempo y el modelo ARIMA) deben ser independientes y presentar una distribución normal. A pesar de que las técnicas de filtrado pueden ser útiles para convertir las series temporales no estacionarias en estacionarias, no siempre es posible cumplir todos estos requerimientos. Además, la mayoría de ellos involucran cálculos complejos y los resultados que se obtienen no siempre son los mejores.

En resumen, solamente con un modelo adecuado, unido a una correcta identificación de sus parámetros y el supuesto de que la relación entre dichos parámetros es constante en el tiempo, los valores futuros de la serie de tiempo podrán ser pronosticados con un razonable rango de confianza. De no ser así, el modelo ARIMA obtenido resultará inadecuado y los resultados no se corresponderán con la realidad objetiva del fenómeno que se pretende representar.

Por su parte, el análisis espectral (Chatfield, 2003) es en esencia un procedimiento que estima, a partir de un conjunto finito de observaciones, una función en el rango  $(0, \pi)$ . Se considera fundamentalmente no paramétrico ya que no se asume a priori ningún tipo de modelo. Resulta útil para orientar la búsqueda de un modelo paramétrico que se adecue a las características de la serie. La estimación de la función de densidad espectral define un procedimiento más general que la inferencia basada en una clase particular de modelo

paramétrico aunque, por lo general, es menos exacto si el modelo paramétrico resulta realmente apropiado. El análisis espectral obtiene mejores resultados en los casos donde no resulte obvia la existencia de tendencia y variación estacional, principalmente series provenientes de las ciencias físicas. En meteorología y oceanografía el análisis en el dominio de la frecuencia ha resultado particularmente efectivo, no así en otras áreas como las ciencias económicas.

## **1.4 Conclusiones del capítulo**

Las series temporales permiten describir de forma natural gran variedad de fenómenos que transcurren a lo largo del tiempo. Es por ello que su uso se ha extendido a numerosas áreas del conocimiento, especialmente aquellas que requieren predecir el comportamiento de determinadas variables de interés en un momento dado. Existen múltiples modelos matemáticos y estadísticos que permiten describir problemáticas reales satisfactoriamente aunque no se ajustan a todo tipo de problemas. Esto es debido a que la aplicación de estos modelos tradicionales precisa del cumplimiento de ciertos requerimientos, lo cual es una limitación no desdeñable en muchos casos.

## 2 Minería de datos para el análisis de series temporales

La minería de datos para series temporales es una contribución importante a los campos de estudio de la minería de datos y de las series temporales. Los métodos utilizados en la minería de datos para series temporales son capaces de caracterizar satisfactoriamente series con características periódicas, no periódicas, complejas y caóticas. Estos métodos cubren las limitaciones de las técnicas tradicionales utilizadas en el análisis de series temporales ya que adaptan los conceptos de la minería de datos, para tratar este tipo de series como una clase especial de datos. En este capítulo se analizan las principales tendencias de la minería de datos para series temporales en la actualidad.

### 2.1 Minería de datos

Según Larose (2005), se puede definir que: “La minería de datos es el proceso de descubrir nuevas correlaciones significativas, modelos y tendencias, filtrando grandes cantidades de datos guardadas en repositorios, a través del uso de tecnologías de reconocimiento de modelos así como de técnicas estadísticas y matemáticas”.

El objetivo de este proceso es revelar patrones desconocidos a partir de los datos. Su singularidad radica en los tipos de problemas que es capaz de resolver (aquellos con enormes conjuntos de datos y relaciones muy complejas entre ellos).

En ocasiones el descubrimiento de conocimiento en las bases de datos o KDD (acrónimo del inglés *Knowledge Data Discovery*) se trata como sinónimo de minería de datos. Alternativamente, otros ven la minería de datos como simplemente un paso esencial en el proceso de KDD. Por ejemplo, en Han and Kamber (2001) utilizan el término minería de datos para referirse en general al proceso de descubrimiento de conocimiento a partir de grandes bases de datos, almacenes o repositorios de información.

Existen categorías que agrupan los diferentes tipos de tareas existentes para la minería de datos, correspondiendo a los objetivos del análisis y los tipos de problemas que enfrentan.

Dada la naturaleza de los problemas, los podemos agrupar en distintas tareas, tales como (Han and Kamber, 2001):

- Clasificación
- Agrupamiento
- Asociación
- Predicción
- Regresión

## **2.2 Minería de datos para series temporales**

La minería de datos para series temporales requiere tener claramente definidos cuáles serán los eventos que vamos a “minar”. De manera similar es necesario definir las formaciones que apuntan a eventos significativos. En el contexto de la minería de datos para series temporales estas formaciones son llamadas patrones temporales (Povinelli, 1999). Un patrón temporal puede estar asociado a un evento por lo cual es necesario en la predicción de los eventos.

Se supone comúnmente que en las series temporales modeladas con ARIMA, los cambios en el pasado serán aplicados a la predicción del futuro. Por lo que se asume que estos modelos no necesitarán variar a través del tiempo. Desafortunadamente, el sistema generador de una serie de tiempo no tiene por qué ser necesariamente lineal o estacionario. En contraste con lo anterior, los métodos basados en minería de datos son capaces de manipular series temporales no lineales y no estacionarias. Es por este motivo que resultan útiles para predecir eventos imprevistos en la serie como: el alza repentina del precio de algún producto en el mercado o la rotura de alguna clase de motor en una fábrica.

La naturaleza de las series temporales hace que su tratamiento se diferencie de los métodos tradicionales de minería de datos. Entre las características distintivas se encuentran: alta numerosidad, gran número de dimensiones y una constante actualización de sus datos al transcurrir el tiempo.

Considerando su naturaleza continua, es imprescindible considerar una serie de tiempo como un todo en lugar de tratarla como un conjunto de campos numéricos individuales. El cálculo de la similaridad para series temporales se satisface de forma aproximada, a

diferencia de otros tipos de datos donde el concepto de similitud se resuelve de forma exacta, ya que es prácticamente imposible encontrar dos series exactamente iguales. Todas estas características constituyen retos que la minería de datos para series temporales debe resolver.

## **2.3 Tareas de la minería de datos para series temporales**

En los últimos años ha ocurrido un incremento de las investigaciones relacionadas con la minería de datos para series temporales. La diversidad de los trabajos abarcan la mayoría de las líneas que sigue la minería de datos actualmente. En aras de facilitar su análisis varios autores (Fu, 2011) clasifican estas tareas en los campos siguientes:

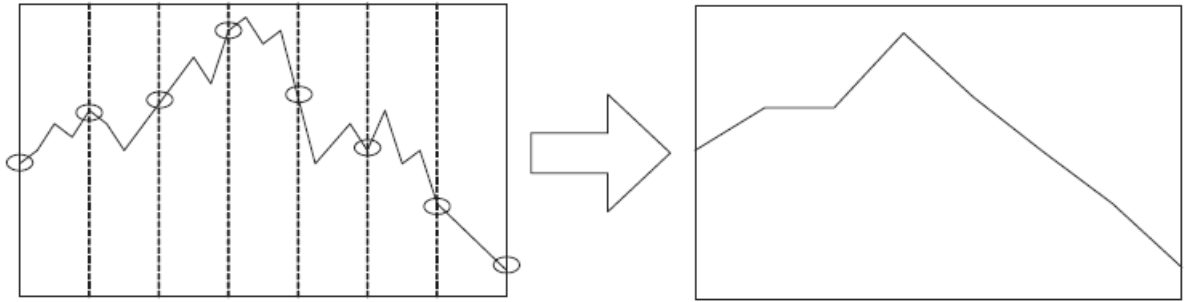
- representación e indexado
- clasificación
- medidas de similitud
- emparejamiento de subsecuencias
- segmentación
- visualización
- descubrimiento de patrones y conglomerados

### **2.3.1 Representación e indexado**

La representación de las series temporales es uno de los principales desafíos que enfrenta la minería de datos debido a la alta dimensionalidad que frecuentemente presenta este tipo de dato. Debido a que la representación influye en el desempeño posterior de los métodos utilizados, una de las principales tareas dentro de la representación de series temporales es la reducción de su dimensión. Se entiende por dimensión el número de puntos de datos de la serie original.

El método más simple para ello es el muestreo (Åström, 1969). En este método, una tasa de  $m/n$  es usada, donde  $m$  es la longitud de la serie de tiempo  $P$  y  $n$  es la dimensión después de la reducción, Figura 2.1. El método de muestreo tiene la inconveniencia de que distorsiona la forma de la serie temporal obtenida si la tasa de muestreo es demasiado pequeña.



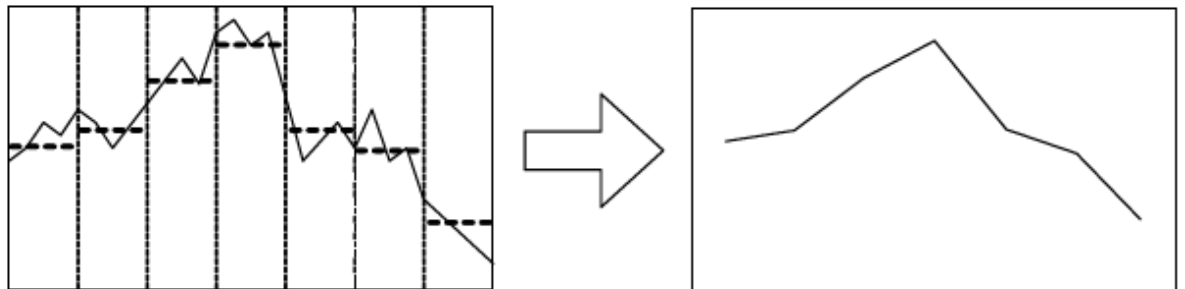


**Figura 2.1. Reducción de la dimensionalidad de una serie de tiempo mediante muestreo. La serie de tiempo de la izquierda es muestreada regularmente (denotado por líneas punteadas) y desplegada a la derecha**

Un método mejorado consiste en utilizar el valor medio de cada segmento para representar el correspondiente conjunto de puntos de datos ( $Y_i$ ). A partir de una serie de tiempo  $P = (p_1, \dots, p_m)$  y la dimensión deseada luego de la reducción  $n$ , la serie de tiempo reducida  $\tilde{P} = (\tilde{p}_1, \dots, \tilde{p}_n)$  es obtenida de la forma siguiente:

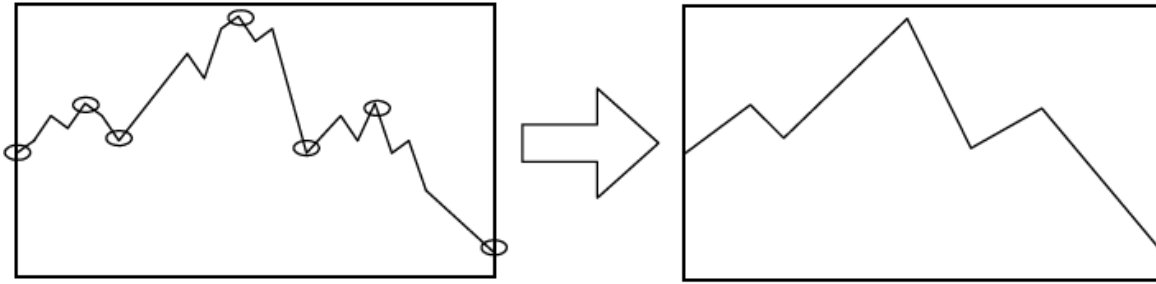
$$p_k = \frac{1}{e_k - s_k + 1} \sum_{i=s_k}^{e_k} p_i$$

donde  $s_k$  y  $e_k$  denotan los puntos donde comienza y finaliza el  $k$ -ésimo segmento respectivamente. De esta forma la serie resultante presenta un mayor parecido con la serie original (Figura 2.2).



**Figura 2.2. Reducción de la dimensionalidad mediante el cálculo de la media en cada segmento**

Otro método, utilizado en la reducción de la dimensionalidad, consiste en aproximar la serie de tiempo usando líneas rectas (Keogh *et al.*, 2001). Reducir la dimensionalidad a partir de la conservación de los puntos extremos de la serie, es también un enfoque promisorio. Estos puntos se denominan puntos perceptualmente importantes (PIP). La figura 2.3 muestra el resultado de aplicar este método.



**Figura 2.3 .Reducción de la dimensionalidad mediante el cálculo de siete PIP**

Diversos métodos han sido propuestos para representar las series temporales. Algunos, como los anteriormente analizados, plantean su representación en el dominio del tiempo directamente. La representación de las series temporales en el dominio de transformación constituye otra larga familia de métodos (Keogh, 1997; Keogh and Pazzani, 2001; Smyth and Keogh, 1997). Otra variante consiste en convertirla a una representación simbólica (Aref *et al.*, 2004; Lin *et al.*, 2005).

### 2.3.2 Clasificación

La clasificación asocia datos entre grupos predefinidos o clases. La mayoría de los algoritmos de clasificación asumen algún conocimiento de los datos o realizan fases de entrenamiento para estas clasificaciones. El problema de la clasificación de series temporales puede ser definido de la forma siguiente:

Dada una base de casos  $D = \{t_1, t_2, \dots, t_n\}$  constituida por series temporales, y un conjunto de clases  $C = \{C_1, C_2, \dots, C_m\}$ , el problema de clasificar dichas series es el de definir una función  $f: D \rightarrow C$  donde cada  $t_i$  es asignada a una clase  $C_j$ . Y una clase  $C_j$  contiene

precisamente a las series asignadas en ella, esto es

$$C_j = \{t_i \mid f(t_i) = c_j, 1 \leq i \leq n, t_i \in D\}.$$

Muchos de los algoritmos de clasificación conocidos se han adaptado para ser utilizados con series temporales. En Rodríguez *et al.* (2000) se propone un método de clasificación de series de tiempo multivariadas mediante la programación lógica inductiva. Se introducen dos tipos de predicados: basado en intervalo y basado en distancia. En Geurts (2001) se propone clasificar los datos de la serie de tiempo basándose en sus propiedades locales o en sus patrones. Por otra parte en Povinelli *et al.* (2004) se presenta una investigación sobre la clasificación de señales basándose en el modelado de un sistema dinámico que captura los datos para la serie usando modelos de texturas Gaussianos.

Los árboles de decisión también han sido aplicados a la clasificación de series temporales como lo denotan los trabajos de Rodríguez y Alonso (2004), y Douzal-Chouakria and Amblard (2012). Las máquinas de soporte vectorial también han como esido utilizadas en esta tarea como se reporta en Orsenigo and Vercellis (2010).

El algoritmo  $k$ -NN es uno de los más populares en la minería de datos y una gran cantidad de los trabajos para series temporales se basan en dicho algoritmo (Zhang *et al.*, 2004; Gonzalez-Abril *et al.* (2009). En el caso de esta última técnica, el algoritmo de los  $k$  vecinos más cercanos ( $1$ -NN), a pesar de su simplicidad, es uno de los que mejores resultados ha ofrecido para la clasificación de series temporales.

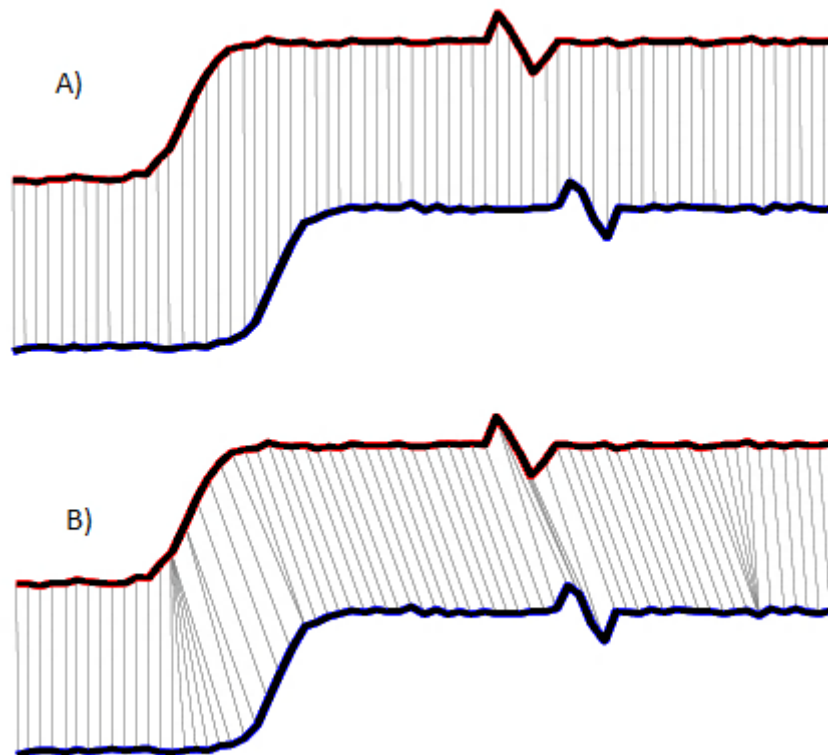
Incluso métodos bioinspirados como los basados en colonias de hormigas (Biswal, 2010] y enjambres de partículas se han adaptado para el trabajo con series temporales. Finalmente, en Wei and Keogh (2006) se proponen clasificadores de series temporales semi-supervisados para los que solo son necesarios un grupo reducido de ejemplos etiquetados para el aprendizaje.

### 2.3.3 Medidas de similitud

Las medidas de similitud tienen una gran importancia para las distintas tareas de la minería de datos pues constituyen el núcleo de la mayoría de los métodos utilizados en análisis de

series temporales. No resulta trivial definir funciones de similitud dada la naturaleza numérica y continua de las series temporales. Existen dos enfoques principales para el cálculo de la similitud: considerar la serie de tiempo en toda su longitud, y la comparación de subsecuencias.

Una de las distancias más usadas es la tradicional distancia euclidiana, que se emplea fundamentalmente en las series temporales transformadas. En esta medida se calcula la diferencia entre cada punto de datos de la serie objetivo respecto a su similar en la serie de referencia. En la Figura 2.4 A) se muestra cómo se efectúa el cálculo de la diferencia entre dos series utilizando la distancia euclidiana.



**Figura 2.4. Cálculo de la diferencia entre dos series temporales. A) Utilizando la distancia euclidiana. B) Utilizando la métrica elástica DTW**

Entre los beneficios de utilizar esta medida se encuentra una complejidad computacional de orden lineal. Como consecuencia de las características particulares que ostentan las

series temporales, varios estudios revelan que no siempre es la distancia indicada para dominios más específicos (Wang *et al.*, 2013).

Una de las medidas de similitud más populares usada actualmente se conoce con el nombre de distorsión dinámica del tiempo o DTW (acrónimo del inglés *Dynamic Time Warping*) (Sakoe and Chiba, 1978). En la Figura 2.4 B se muestra la comparación entre dos series utilizando la métrica DTW, siendo apreciable la diferencia existente entre la distancia euclidiana y DTW. La principal característica de esta última es la elasticidad durante las comparaciones, que hace posible la alineación entre puntos desfasados en el tiempo. Como esta métrica ha originado multitud de trabajos se le dedica un capítulo completo para su análisis.

#### **2.3.4 Emparejamiento de subsecuencias**

Dadas una secuencia de entrada y una serie de tiempo de mayor longitud, la tarea en este caso es hallar las subsecuencias en la serie de tiempo que se “empareje” mejor con la secuencia dada. Los primeros trabajos sobre este tema se pueden revisar en Agrawal *et al.*, (1993) y Faloutsos *et al.* (1994). A partir de estos trabajos, numerosas investigaciones se han llevado a cabo para mejorar el funcionamiento de la búsqueda de subsecuencias. Por ejemplo, los métodos de Moon *et al.* (2001) y Moon *et al.* (2002) proponen el uso de ventanas móviles y en Han *et al.* (2007) se desarrolla un algoritmo de ordenamiento de subsecuencias para reducir el número de las mismas a las cuales se necesita tener acceso durante el emparejamiento.

#### **2.3.5 Segmentación**

La segmentación puede ser vista tanto como un paso de preprocesado para numerosas tareas de la minería de datos o como una técnica de análisis de tendencia. También puede ser considerada como un proceso de discretización. En Das *et al.* (1998) se propone un método simple de discretización. Una ventana de longitud fija es usada para segmentar la serie de tiempo en subsecuencias y de esta forma representarla mediante patrones primitivos. Este proceso depende fundamentalmente de la elección del ancho de la ventana.

Existen al menos dos desventajas significativas. Primero, los patrones fundamentales aparecen típicamente con diferentes longitudes a través de toda la serie. Segundo, como un resultado de la segmentación de cualquier serie de tiempo, los patrones más importantes pueden perderse cuando se separan datos en el tiempo.

Por tanto, es preferible usar enfoques dinámicos, los cuales identifiquen los puntos de datos que podemos dividir en el tiempo antes de proceder al segmentado de la serie. La tarea de segmentación descrita anteriormente puede ser vista también como un problema de optimización. En Chung *et al.* (2004) se propone una solución basada en computación evolutiva.

### **2.3.6 Visualización**

La visualización es un importante mecanismo para presentar la serie de tiempo procesada. De esta forma se facilita su análisis a los usuarios. Es además una poderosa herramienta que hace más factibles las tareas de minería de datos en la serie. Algunas de las herramientas más importantes desarrolladas son: *TimeSearcher* (Hochheiser and Shneiderman, 2004, Keogh et al., 2002a) y *VizTree* (Lin et al., 2005). Ambas incluyen:

- visualización de calendarios y conglomerados (Van Wijk and Van Selow, 1999).
- visualización en espiral (Weber et al., 2001).

### **2.3.7 Descubrimiento de patrones y conglomerados**

El descubrimiento de patrones, también llamado descubrimiento causal de patrones o detección de anomalías, es la tarea no trivial de descubrir patrones (Keogh et al., 2002b) interesantes en una serie de tiempo. Dada su importancia, se ha convertido en una de las tareas fundamentales de la minería de datos para series temporales y puede ser aplicada a numerosos dominios de investigación (Caraça-Valente and López-Chavarrías, 2000; Lerner et al., 2004; Ma and Perkins, 2003). Variadas técnicas se han desarrollado, entre ellas cabe señalar: el algoritmo de Gecko (Chan and Mahoney, 2005), las técnicas basadas en distancia (Das et al., 1998; Oates, 1999, Wang et al., 2002), el modelo basado en cadenas

de Markov para series temporales (Panuccio *et al.*, 2002) y el método de agrupamiento neuronal de conglomerados para el reconocimiento autoorganizado (Lagus *et al.*, 1996).

En Mueen *et al.* (2011) se introduce un algoritmo llamado DAME (del acrónimo en inglés de *Disk Aware Motif Enumeration*) para encontrar patrones en enormes bases de datos, las cuales contienen decenas de millones de series de tiempo. El algoritmo realiza una proyección de las series en una línea ordenada lo cual le permite realizar un número de comparaciones menor que un algoritmo de fuerza bruta.

En Rodpongpun *et al.* (2012) se propone un método para conglomerados de subsecuencias dentro de series temporales. Este algoritmo trabaja de forma selectiva, seleccionando solamente las subsecuencias esenciales. Esto garantiza la obtención de conglomerados con sentido y todo esto se realiza de manera automática.

Otro trabajo novedoso en este campo lo constituye el de Meesrikamolkul *et al.* (2012). En este artículo se propone un método nombrado SCTS (acrónimo del inglés *Shape-based Clustering for Time Series*). Este propone una nueva forma para efectuar el cálculo de la media entre dos series. La propuesta en cuestión consigue un menor tiempo computacional y mejora la eficacia de las propuestas anteriores.

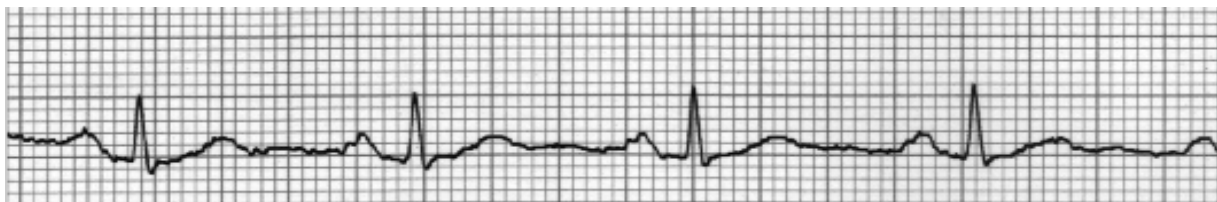
## **2.4 Principales campos de aplicación y algunos problemas representativos**

El repositorio de series temporales conocido como UCR (acrónimo del inglés *University of California Riverside*) (Keogh, 2011) ha sido creado como un servicio público para la comunidad científica que trabaja la minería de datos y el aprendizaje automatizado. Su objetivo es alentar las investigaciones en el campo de la clasificación de series temporales. En este sitio se encuentran a disposición de los investigadores más de 50 conjuntos de datos internacionales de probada fiabilidad, así como información valiosa sobre los mismos (sus creadores, la cantidad de instancias que contienen, sus clases, los mejores resultados obtenidos de cada uno de ellos con diversos algoritmos de clasificación y varias medidas de similitud, etc.).

En UCR se publican además los más novedosos artículos científicos sobre el tema, así como el código fuente de varios algoritmos tradicionales implementados. Durante muchos años se ha invitado a la comunidad científica cuyo campo es la minería de datos a que contribuya con nuevos conjuntos de datos para el sitio. Esto se ha realizado con el objetivo de que la colección existente represente los intereses de grupos cada vez más diversos de investigadores, y no de algunos en particular. Debido a la utilidad que tienen los conjuntos de datos para la validación experimental de las investigaciones, este epígrafe muestra algunos de ellos; como ejemplos además, de la diversidad de aplicaciones que presenta actualmente el campo de la minería de datos para series temporales.

#### 2.4.1 ECG200

La clasificación de enfermedades cardíacas ha recibido una gran atención de la comunidad científica por la gran cantidad de datos disponibles libremente y su potencial aplicación en la medicina, En la Figura 2.5. Ejemplo de un electrocardiograma.



**Figura 2.5.** Ejemplo de un electrocardiograma

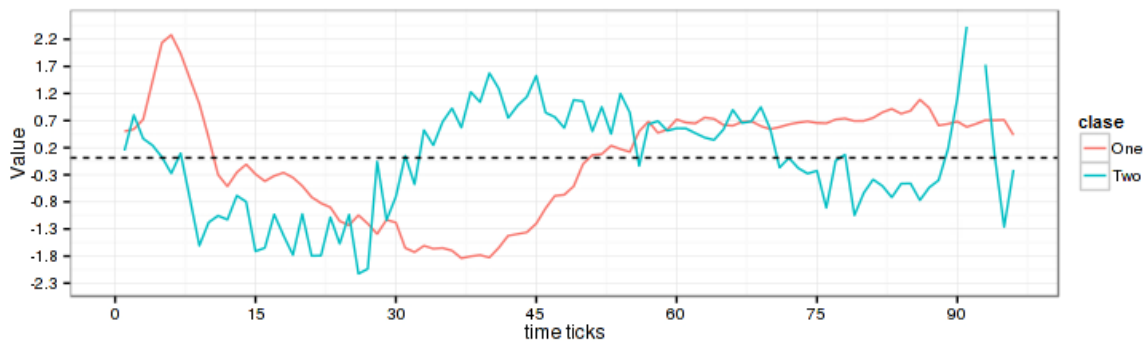
El Instituto Nacional de Metrología de Alemania ha provisto la compilación de un gran conjunto de datos ECG<sup>1</sup> para su investigación con algoritmos de medición computacionales. La información de ECG fue recogida de voluntarios sanos y otros con algún tipo de enfermedad cardíaca. Cada dato almacenado en ECG es una serie de tiempo registrada por un electrodo durante cada pulsación del corazón. Los datos han sido anotados por cardiólogos y dos clases han sido definidas: comportamiento normal y comportamiento

---

<sup>1</sup> El electrocardiograma (ECG/EKG, del alemán *Elektrokardiogramm*) es la representación gráfica de la actividad eléctrica del corazón, que se obtiene con un electrocardiógrafo en forma de cinta continua.



anormal. De las 200 instancias del conjunto de datos, 75 fueron identificadas por los especialistas como anormales, y 125 como normales. Todos los datos han sido normalizados y reescalados para que tengan longitud 95. La Figura 2.6 muestra la comparación entre dos series temporales que representan dos instancias ECG de diferente clase.

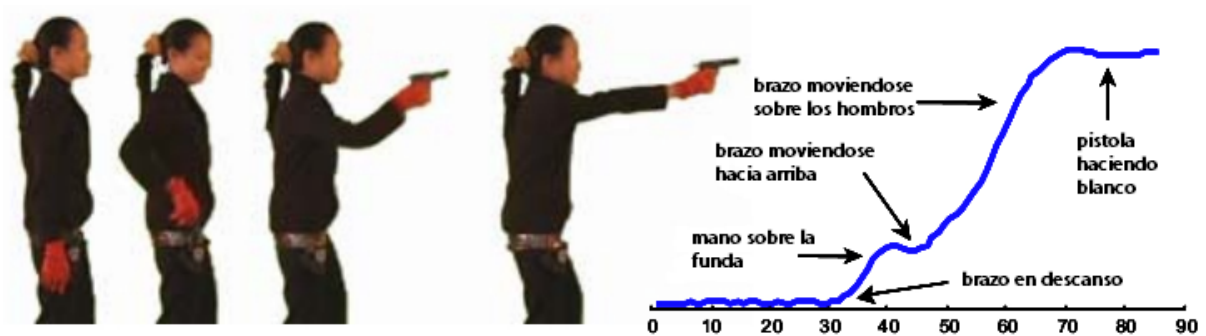


**Figura 2.6.** Ejemplo de dos clases del conjunto de datos ECG200

Debido a que los cardiólogos están más interesados en las ocurrencias anormales de los electrocardiogramas, el objetivo de la minería de datos para este problema es conocer si una instancia dada es clasificada en normal o anormal, dándole mayor importancia a aquellas que son clasificadas como anormales.

### 2.4.2 Gun Point

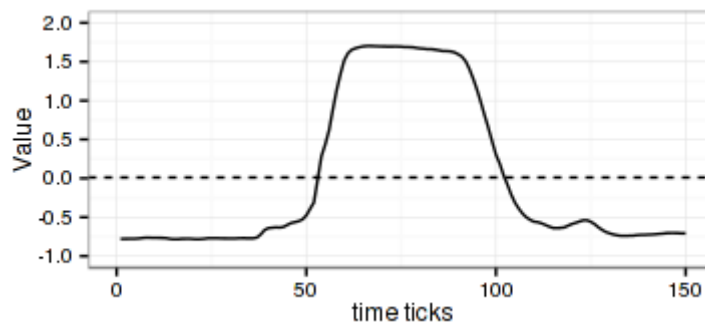
El conjunto de datos Gun Point pertenece al dominio de los videos de vigilancia. Gun Point consta de dos clases, cada una contiene 200 instancias, 100 para cada clase. Todas las instancias fueron creadas usando un actor del sexo masculino y uno del sexo femenino, durante una única sesión de pruebas en la que los actores fueron grabados en un video (a 30 cuadros por segundo), como se muestra en la Figura 2.7.



**Figura 2.7.** Instantáneas de una secuencia de video; el movimiento de la mano derecha es rastreado y convertido en una serie de tiempo

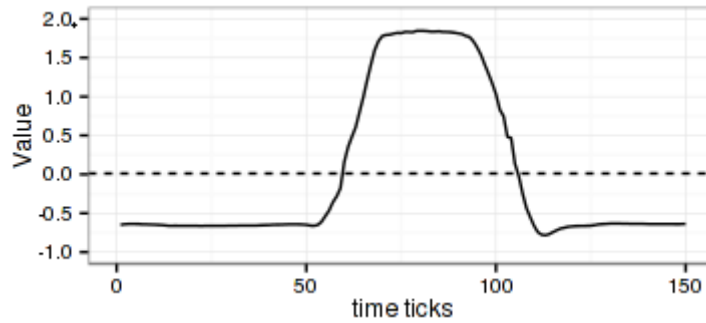
La grabación fue fácilmente segmentada en 150 puntos de datos que representan una instancia. Las dos clases son:

- Gun: Primeramente, los actores sitúan ambas manos a los lados del cuerpo. Luego levantan un arma desde su funda ubicada en su cintura hasta que la pistola se posiciona para hacer blanco (esto ocurre en aproximadamente un segundo). A continuación vuelven a poner el arma en su funda y ambas manos a los lados del cuerpo, Figura 2.2.8.



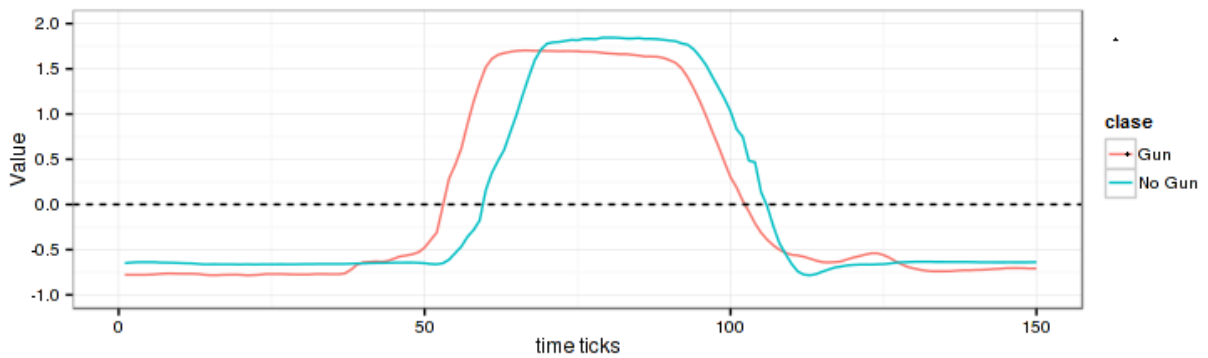
**Figura 2.8.** Ejemplo de una instancia de la clase Gu

- No Gun: Los actores mantienen su pistola en su funda. Apuntan con el dedo índice a un objetivo aproximadamente en un segundo, para más tarde volver a situar sus manos a ambos lados del cuerpo, Figura 2.9.



**Figura 2.9. Ejemplo de una instancia de la clase No Gun**

El centroide de la mano derecha del actor es capturado por la cámara en los ejes **X** y **Y**; los cuales parecen estar muy correlacionados. En este experimento solo se considera el eje **X** por simplicidad. Para más detalles sobre el experimento consultar Ratanamahatana and Keogh (2004a).



**Figura 2.10. Ejemplo de dos clases del conjunto de datos Gun Point**

En este problema el objetivo consiste en conocer si el actor apunta o no con un arma mediante la caracterización de cada uno de sus movimientos. Los eventos considerados relevantes son tomados mediante las observaciones de una cámara y corresponden al movimiento de la mano derecha del actor. Aunque las clases sean muy similares entre sí, Figura 2. y Figura 2., es posible para un humano clasificar visualmente ambas clases con gran precisión después de notar que el actor debe alzar su mano sobre la funda de su pistola para sacarla. Esta acción genera una sutil distinción entre ambas clases la cual se hace visible fácilmente en la Figura 2..

### 2.4.3 Fifty Words

El problema de transcribir e indexar archivos históricos existentes es aún un reto. Incluso para figuras históricas de la talla de Isaac Newton existen una gran cantidad de trabajos escritos a mano que todavía no han sido publicados (los trabajos de Newton exceden el millón de palabras). Para otras muchas personalidades, están recogidos actualmente muchos trabajos escritos a mano, colecciones que tienen un valor incalculable para biógrafos e investigadores, y que todavía no han sido descifrados y traducidos enteramente debido a la complejidad de este problema. Sorprendentemente, es posible transformar texto escrito a mano en series temporales,

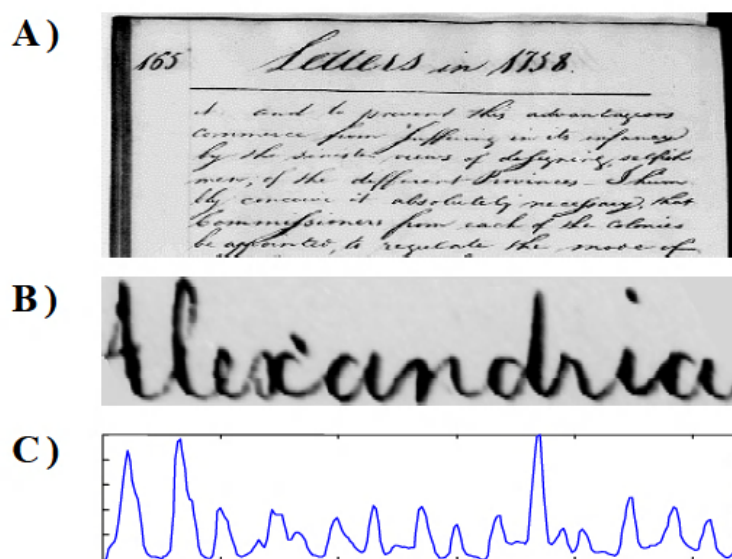
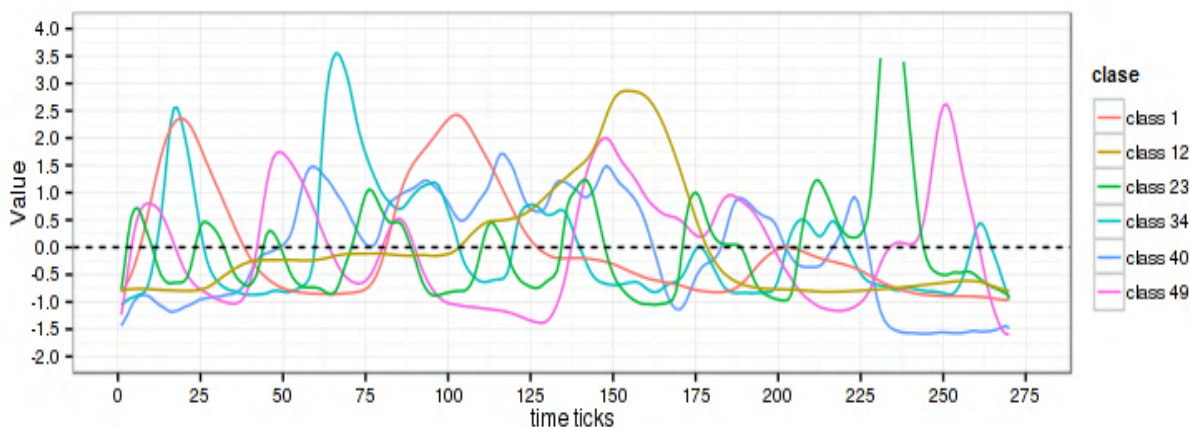


Figura 2.11. A) Ejemplo de un texto escrito por George Washington. B) La palabra “Alexandria” del texto A) luego de haber sido procesada para eliminar su inclinación. C) La palabra de B) convertida en una serie de tiempo

Por ejemplo, consideremos el problema de traducir textos bíblicos a dos lenguajes diferentes (Ratanamahatana and Keogh, 2004a): inglés y español. Para ello, el texto bíblico es convertido por entero en cadenas de bits de acuerdo a las ocurrencias de una palabra seleccionada en el texto. Por ejemplo, un apartado de la Biblia en español que contiene la palabra “Dios” en la frase “En el comienzo Dios creó el cielo y la tierra” será representada

como “0001000000”. Esta cadena de bits es convertida entonces en una serie de tiempo registrando el número de ocurrencias de la palabra dentro de una ventana móvil predefinida para todo el texto. Intuitivamente, por cada aparición de la palabra en idioma inglés, debe estar presente su correspondiente en español. Sin embargo, pueden existir algunas discrepancias en el número de palabras existentes en el texto completo, así como en la posición de la palabra dentro de cada oración para ambos lenguajes. Estas irregularidades pueden ser analizadas en detalle usando técnicas de minería de datos.

El conjunto de datos conocido como Fifty Words, fue creado por Rath y Manmatha para el emparejamiento de imágenes (Rath and Manmatha, 2003). Contiene 2 381 imágenes de palabras, de 10 páginas escritas. Se han tomado imágenes de 50 palabras comunes en idioma inglés como “the”, “and”, etc. obteniéndose 905 instancias en total. Cada imagen de palabra es representada por una serie de tiempo cuatridimensional la cual describe las características de la imagen. Por ejemplo en la Figura 2. se muestra el perfil de la palabra “Alexandria”. Por simplicidad, Fifty Words considera solo la primera dimensión de cada imagen, la cual tiene una longitud promedio de 270.



**Figura 2.12. Ejemplo de seis clases del conjunto de datos Fifty Words**

Luego, el problema para este conjunto de datos se centra en diferenciar la palabra “the” del resto. En total, existen 109 imágenes de la palabra “the” y 796 imágenes de las otras palabras, para un total de 905 imágenes en todo el conjunto de datos. La Figura 2.12

muestra algunos ejemplos representativos de varias clases en el conjunto de datos Fifty Words.

## **2.5 Conclusiones del capítulo**

Las series temporales han despertado el interés de los especialistas en la rama de la minería de datos debido a las limitaciones que presentan los métodos matemáticos y estadísticos para su análisis. Existe un número creciente de problemas con una clara modelación a partir de las series temporales y otros que aunque no resulta obvio también pueden convertirse en series temporales. Los modelos que analizan las series temporales usando técnicas de minería de datos son capaces de resolver problemas donde los modelos estadísticos y matemáticos fallan. Dada la gran cantidad de aplicaciones prácticas que han surgido, actualmente existe un incremento de los estudios que se realizan en el campo de la minería de datos para series temporales; los cuales están divididos en las categorías siguientes: representación e indexado, clasificación, medidas de similitud, emparejamiento de subsecuencias, segmentación, visualización y descubrimiento de patrones y conglomerados. Aunque no resulta fácil enfrentar características de las series temporales como su alta dimensionalidad y numerosidad, los métodos de minería de datos han sido adaptados satisfactoriamente al trabajo con series temporales.

### 3 La métrica elástica DTW en el contexto de la minería de datos

DTW es una técnica muy popular para la comparación de series temporales. Con esta técnica no solo se consigue el valor de la disimilitud entre dos series sino que además se obtiene el alineamiento óptimo entre ellas, emparejándolas de forma no lineal mediante contracciones y dilataciones de las series en el eje temporal. Por consiguiente, este emparejamiento permite encontrar regiones equivalentes entre las series y hallar su similitud

DTW ha encontrado aplicación en varias disciplinas como son: minería de datos, reconocimiento de gestos, robótica, en procesos fabriles o en medicina (Keogh and Pazzani, 2001). En minería de datos para series temporales DTW es usada comúnmente para calcular la diferencia entre dos series temporales.

En reconocimiento del lenguaje oral, donde tuvo su primera aplicación, esta medida de disimilitud resulta útil para determinar si dos ondas sonoras representan la misma frase en una conversación interpersonal cualquiera. Esto se debe a que la duración del sonido de cada letra puede variar, pero la onda sonora en general debe tener la misma forma para la misma frase. Una gran cantidad de algoritmos y restricciones se han discutido en la literatura a partir del surgimiento de DTW y es por esto que dedicamos todo un capítulo para su análisis.

#### 3.1 Definición de DTW

Para alinear dos series temporales  $P = (p_1, \dots, p_n)$  y  $Q = (q_1, \dots, q_m)$  usando DTW, primeramente se construye una matriz  $M_{n \times m}$  la cual se utiliza para almacenar las distancias parciales entre las subsecuencias de las series originales. Cada elemento  $m_{ij}$  de la matriz

$M$ , contiene la distancia  $d(q_i, p_j)$  entre dos puntos  $q_i$  y  $p_j$  de cada serie, respectivamente.

Para el cálculo de la distancia  $d(q_i, p_j)$  se utiliza típicamente la distancia euclidiana (3.).

$$d(q_i, p_j) = (q_i - p_j)^2 \quad (3.1)$$

El valor obtenido al calcular la distancia local  $d(q_i, p_j)$  expresa el efecto de alinear los puntos  $q_i$  y  $p_j$  de las series. Un camino distorsionado  $W$  es un conjunto de elementos de la matriz que definen una correspondencia entre  $P$  y  $Q$ .

El  $k$ -ésimo elemento contenido en  $W$  se define como se muestra en (3.) y tanto  $i_k$  como  $j_k$  son los índices de puntos contenidos en  $P$  y  $Q$ , respectivamente.

$$w_k = (i_k, j_k) \quad (3.2)$$

$$W = w_1, w_2, \dots, w_k, \dots, w_K$$

Respecto a la longitud del camino  $W$  encontrado por el algoritmo tenemos que se cumple que (3.3):

$$\max(m, n) \leq K \leq m + n - 1 \quad (3.3)$$

El camino distorsionado está comúnmente sujeto a varias restricciones que garantizan una adecuada alineación entre las series. Estas restricciones se conocen como las condiciones de frontera, continuidad y monotonía.

Las restricciones de frontera se definen como  $w_1 = (1, 1)$  y  $w_K = (m, n)$ . Esto hace que el camino distorsionado comience y termine diagonalmente. Por lo tanto se garantiza que el punto inicial como el final de la serie  $P$  se alineará con el inicial y el final de la serie  $Q$ , respectivamente.



Otra restricción a cumplir es respecto a la continuidad. Dado  $w_k = (a, b)$ , entonces  $w_{k-1} = (a', b')$ , donde  $a - a' \leq 1$  y  $b - b' \leq 1$ . Estas desigualdades restringen los pasos posibles en el camino a las celdas adyacentes y controlan que no existan saltos en el camino  $W$ . Además, las restricciones  $a - a' \leq 1$  y  $b - b' \leq 1$  fuerzan a los puntos en  $W$  a ser monótonamente espaciados en el tiempo.

La Figura 3.1 A. Dos series temporales  $P$  y  $Q$  con formas similares pero desfasadas en el tiempo. B) El emparejamiento entre los puntos de cada serie usando DTW permite detectar desfases en el tiempo., muestra cómo dos series  $P$  y  $Q$  son emparejadas entre sí a lo largo del tiempo. Cada punto de la serie  $Q$  es conectado con el punto más similar posible en la serie  $P$  mediante una línea recta que los une. Si existiera el caso que ambas series en la figura fueran idénticas, todas las líneas entre ellas serían verticales pues no se necesitaría de un emparejamiento diferente para alinearlas entre sí de forma óptima.

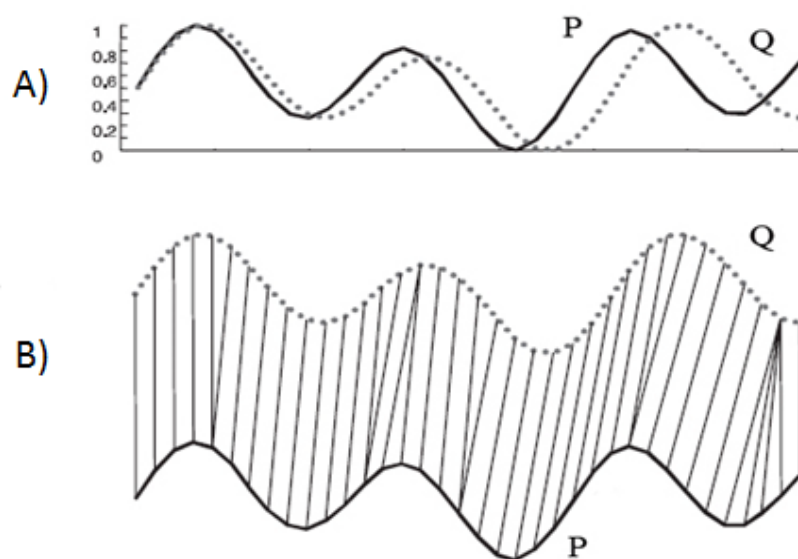


Figura 3.1 A. Dos series temporales  $P$  y  $Q$  con formas similares pero desfasadas en el tiempo. B) El emparejamiento entre los puntos de cada serie usando DTW permite detectar desfases en el tiempo.

La métrica DTW es una medida de la diferencia entre dos series temporales después de que ambas hayan sido alineadas de forma óptima. Dicha medida se corresponde con la suma de las distancias entre cada par de puntos  $w_k$  conectados en el camino  $W$ . Existe un número bastante grande de caminos que satisfacen las condiciones antes mencionadas. Sin embargo, solo estamos interesados en el camino que minimice el costo del alineamiento.

### 3.2 Cálculo del alineamiento óptimo

Este camino (Figura 3.2) puede ser eficientemente hallado usando programación dinámica pues la solución de este problema se realiza a partir de la evaluación de la ecuación de recurrencia (3.). Dicha ecuación de recurrencia define la distancia acumulada  $\gamma(t, f)$  como la distancia  $d(t, f)$  encontrada en la celda actual y el mínimo de las distancias acumuladas de los elementos adyacentes.

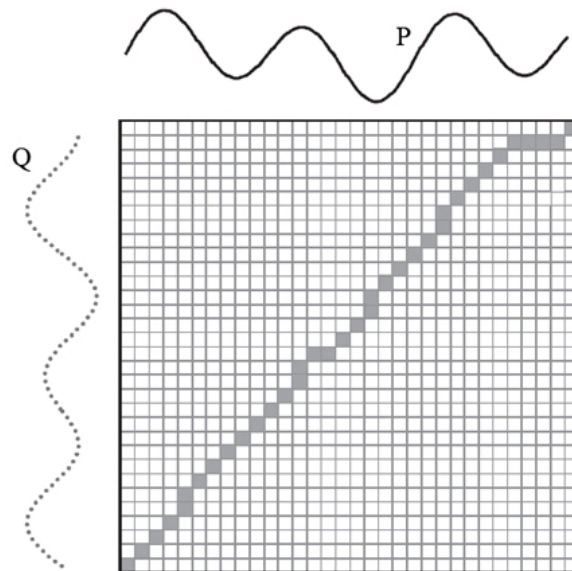


Figura 3.2. Matriz  $M$  donde se forma el camino  $W$  mínimo para los alineamientos entre las series  $P$  y  $Q$

$$\gamma(t, f) = d(q_t, p_f) + \min\{\gamma(t-1, f-1), \gamma(t-1, f), \gamma(t, f-1)\} \quad (3.4)$$

Esta forma de calcular la distancia acumulada no es la única que se reporta en la literatura pues la forma de calcularla varía en dependencia de la alineación que se desea obtener. Por ejemplo la ecuación 3.4 favorece los caminos diagonales y puede que este comportamiento no sea el deseado. Para más información sobre los distintos patrones existentes para el cálculo de DTW, consultar Giorgino (2009).

Un camino de alineamiento  $W$ , tal que la distancia global entre los puntos de datos sea mínima, puede calcularse mediante la ecuación (3.).

$$DTW(Q, P) = \min \left\{ \sum_{k=1}^K d(w_k) \right\} \quad (3.5)$$

Donde  $d(w_k)$  puede definirse como muestra la ecuación (3.). Detalles sobre este tratamiento pueden verse en Kruskal, (1983).

$$d(w_k) = d(q_{i_k} - p_{j_k})^2 \quad (3.6)$$

El cálculo de DTW tiene una complejidad temporal y espacial de orden  $O(n, m)$  donde  $n$  y  $m$  son las longitudes de las series  $P$  y  $Q$ , respectivamente. En caso que las series a comparar tengan igual longitud entonces la complejidad se puede expresar como una función cuadrática  $O(n^2)$ .

### 3.3 Reducción del costo computacional al calcular DTW

Aunque aparentemente la implementación computacional no es muy costosa esta situación se agrava cuando los puntos de datos de las series superan los miles, situación que se repite con frecuencia en el dominio de las series temporales. Diferentes métodos se han propuesto

para acelerar la velocidad del proceso del cálculo de las distancias, estos pueden ser divididos en las categorías siguientes:

- restricciones globales: Limitar el número de celdas que son evaluadas en la matriz de costo
- abstracción de los datos: Ejecutar el algoritmo DTW en una representación reducida de los datos
- indexado: Usar funciones de acotación para reducir el número de veces que DTW deberá ejecutarse durante la clasificación o el agrupamiento

### 3.4 Restricciones globales

El establecimiento de restricciones globales que controlan el subconjunto de la matriz que el algoritmo es capaz de visitar durante la construcción del camino de costo mínimo, Figura 3.3, es uno de los métodos más usados en el mejoramiento de la eficiencia de DTW (Ratanamahatana and Keogh, 2004b). El subconjunto de la matriz que el camino de alineamiento es capaz de visitar es también conocido como ventana.

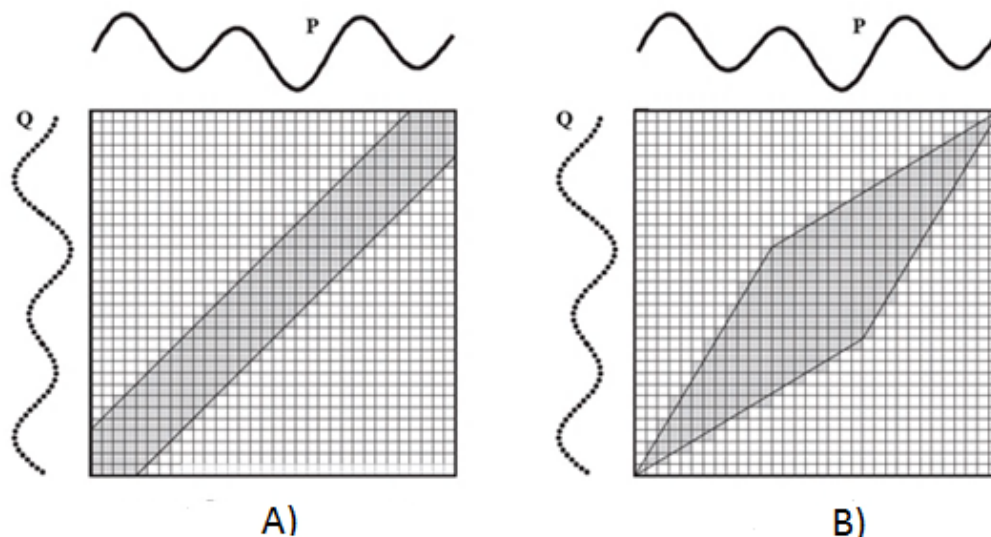
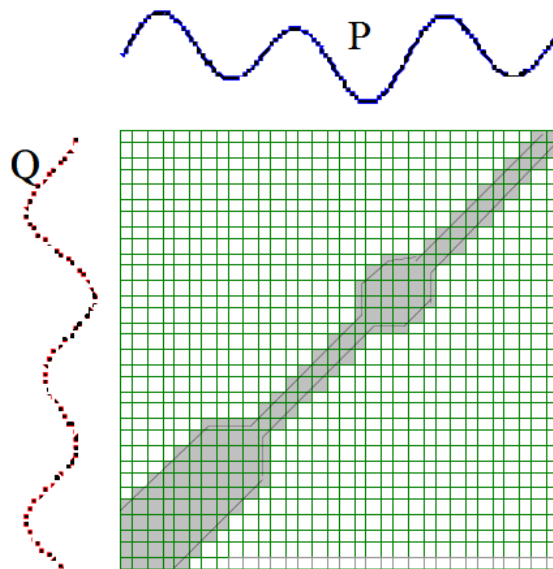


Figura 3.3. Dos de las restricciones globales más usadas. A) Banda de Sakoe-Shiba. B) Paralelogramo de Itakura

De esta forma, se restringen los índices del camino  $w_k = (i_k, j_k)$  tal que  $j-r \leq i \leq j+r$ , donde  $r$  es el término que regula la elasticidad permitida para un punto dado de la serie (normalmente el valor de  $r$  es considerado como un parámetro de la función que implementa la métrica DTW). En el caso de la banda Sakoe-Chiba (Sakoe and Chiba, 1978),  $r$  es independiente de  $i$  y su valor representa el ancho que alcanza la ventana. En el caso del paralelogramo de Itakura (Itakura, 1975)  $r$  se calcula en función de  $i$  pues como se observa en la figura el ancho de la ventana es variable, permitiendo una mayor elasticidad para los puntos de datos localizados en la parte central de la serie.

Los trabajos anteriores han inspirado investigaciones (Ratanamahatana and Keogh, 2004b) donde las restricciones globales se ajustan al problema que se está tratando. El objetivo primario de este ajuste es incrementar la eficacia de la clasificación y como consecuencia de la reducción de la ventana se mejora además la eficiencia del cálculo de DTW. La Figura 3.4 muestra las restricciones globales resultantes al aplicar este método.



**Figura 3.4. Matriz que muestra las restricciones globales aprendidas de forma automática a partir de los datos de un problema**

Evaluaciones empíricas (Keogh and Ratanamahatana, 2005) en numerosos conjuntos de datos han mostrado que la reducción de la ventana no solo aumenta la eficiencia del cálculo DTW sino que además mejora el desempeño de algoritmos como el  $k$ -NN. Esto se debe a que en ocasiones la elasticidad de la métrica provoca la obtención de alineamientos absurdos en el contexto del problema que se está resolviendo. Por tanto, limitar la elasticidad en este caso reduce este tipo de alineamientos erróneos. En muchos casos los mejores resultados se obtienen cuando el valor de  $r$  no supera el 10% de la longitud de la serie.

### 3.4.1 Cotas inferiores

Los métodos analizados anteriormente constituyen variantes para el cálculo de DTW y las modificaciones tienen lugar dentro de la propia función. En este epígrafe se analiza el uso de cotas inferiores en el contexto de las búsquedas por similitud que utilizan la función DTW, por ejemplo el algoritmo  $k$ -NN. El objetivo de usar una cota inferior es aproximar el resultado del cálculo de la función DTW original pero con un costo computacional menor. El uso de la cota inferior permite evitar algunos de los cálculos que se realizan de DTW que es una función relativamente costosa como se analizó con anterioridad.

En la Figura 3.5 se muestra el algoritmo de búsquedas por similitud que saca provecho de la utilización de una cota inferior donde el objetivo es encontrar, dado un conjunto  $C$  de series temporales, la serie que más se asemeja a la serie objetivo  $Q$ . En cada iteración donde resulte falsa la sentencia condicional número cuatro, se ahorra un cálculo de la función DTW.

---

**Algoritmo** Lower\_Bounding\_Sequential\_Scan( $Q, C$ )

---

```
1:  $best\_so\_far = Inf$ 
2: for all sequences in database  $C$  do
3:    $LB\_dist = lower\_bound\_distance(C_i, Q)$  // Cálculo menos costoso
4:   if  $LB\_dist < best\_so\_far$  then
5:      $true\_dist = DTW(C_i, Q)$  // Cálculo costoso
6:     if  $true\_dist < best\_so\_far$  then
7:        $best\_so\_far = true\_dist$ 
8:        $index\_of\_best\_match = i$ 
```

---

Figura 3.5. Algoritmo de búsqueda por similitud que utiliza una cota inferior para el cálculo de DTW con el objetivo de podar cálculos innecesarios.

Existe una multitud de trabajos que presentan cotas inferiores para DTW (Yi *et al.*, 1998; Kim *et al.*; 2001, Sakurai *et al.*, 2005; Niennattrakul *et al.*, 2010). Entre los más destacados se encuentra la cota inferior LB Keogh (Keogh and Ratanamahatana, 2005) e incluso se afirma en (Wang *et al.*, 2013) que no tiene sentido intentar obtener una cota mejor pues el ahorro de cálculos de DTW no sería significativo.

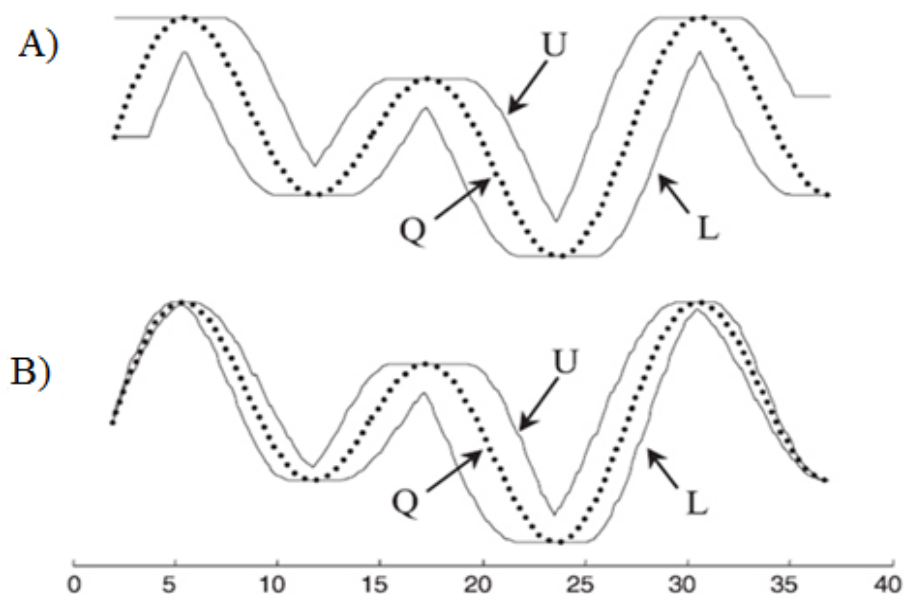
Para realizar el cálculo de la cota inferior LB Keogh primeramente se definen, a partir de una serie objetivo  $Q = (q_1, \dots, q_i, \dots, q_n)$ , dos nuevas series temporales en base a un parámetro  $r$ . Este es el mismo que representa el ancho de la ventana en la restricción global de Sakoe-Shiba (Ecuación 3.7).

$$U_i = \max(q_{i-r}, q_{i+r}) \quad (3.1)$$

$$L_i = \min(q_{i-r}, q_{i+r})$$

Las series  $U$  y  $L$  representan la envolvente superior e inferior, respectivamente, de una serie  $Q$  dada. Como se puede apreciar en la Figura 3.6. Una ilustración de las series  $U$  y  $L$  creadas para la serie  $Q$ : A) usando la banda de Sakoe-Chiba y B) usando el paralelogramo

de Itakura A, ambas series conforman una banda que envuelve totalmente la serie  $Q$ . Es posible notar que, aunque la banda de Sakoe-Chiba tiene un ancho constante en la matriz, la banda correspondiente que envuelve la serie  $Q$  generalmente no tiene un espesor uniforme. En particular, dicha envolvente se hace más ancha en los puntos en los que la serie experimenta cambios repentinos y se achica en sus mesetas. Es posible utilizar distintos valores para el parámetro  $r$  ya que el paralelogramo de Itakura también es aplicable en el cálculo de las envolventes como lo muestra la Figura 3.6 B).



**Figura 3.6.** Una ilustración de las series  $U$  y  $L$  creadas para la serie  $Q$ : A) usando la banda de Sakoe-Chiba y B) usando el paralelogramo de Itakura

Una propiedad obvia pero importante de las series  $U$  y  $L$  en la ecuación (3.2), donde se garantiza que cada punto de datos  $q_i$  se encuentra en el interior de las envolventes o a lo sumo toma el valor de alguna de ellas.

$$\forall_i: U_i \geq q_i \geq L_i \quad (3.2)$$



Habiendo definido las series  $U$  y  $L$ , se define la medida de acotación inferior para DTW mediante la ecuación (3.3).

$$LB\_Keogh(C, Q) = \sqrt{\sum_{i=1}^n \begin{cases} (c_i - U_i)^2 & \text{si } c_i > U_i \\ (c_i - L_i)^2 & \text{si } c_i < L_i \\ 0 & \text{en otro caso} \end{cases}} \quad (3.3)$$

Esta función puede ser interpretada, para el caso de aquellos puntos  $c_i$  que no pertenecen al intervalo entre las envolventes, como la distancia euclidiana entre los puntos de la serie candidata y la envolvente más cercana a dicha serie. En la Figura 3.1 se muestra una interpretación visual de la cota inferior LB Keogh.

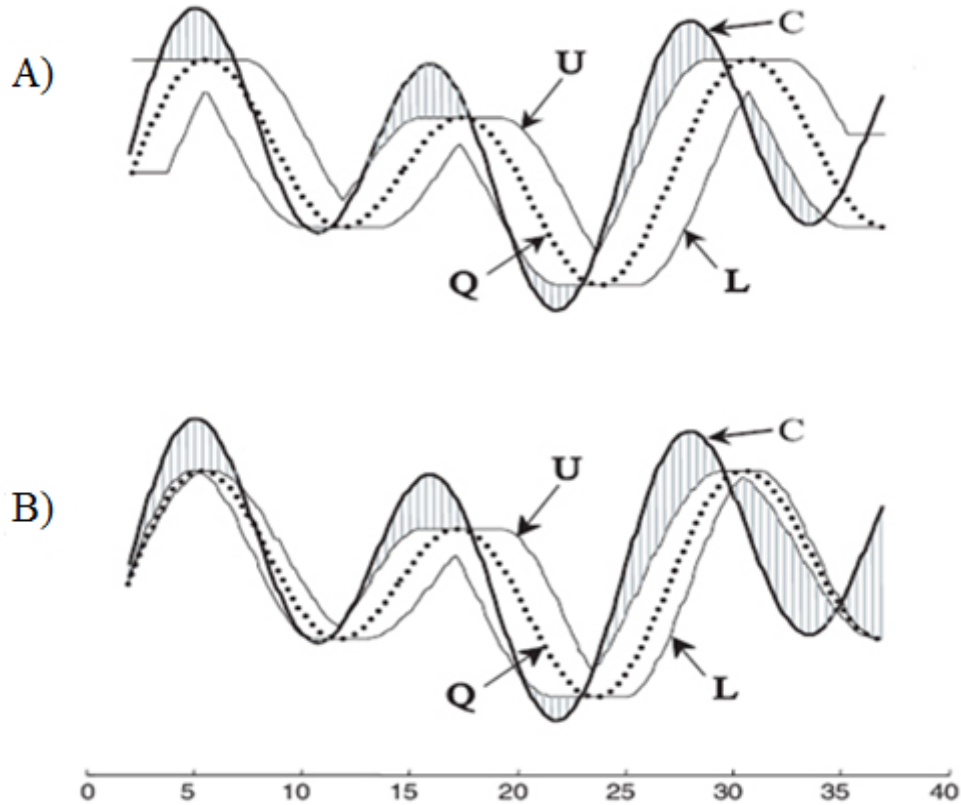


Figura 3.1. LB\_Keogh calcula el cuadrado de la suma de la distancia euclidiana entre los puntos de la serie  $C$  fuera de la envolvente y el borde ortogonal de la envolvente más cercana. A) Banda de Sakoe-Chiba, B) Paralelogramo de Itakura

De aquí se hace posible probar la desigualdad planteada en la ecuación (3.4) (Keogh *et al.*, 2002b). Toda cota inferior para que cumpla su función correctamente nunca debe sobrestimar la función que está aproximando.

$$\forall_i: U_i \geq q_i \geq L_i \quad LB\_Keogh(P, Q) \leq DTW(P, Q) \quad (3.4)$$

Cabe señalar que el costo computacional del cálculo de esta cota inferior es  $O(n)$ , o sea, un costo lineal mucho menor que el costo  $O(n^2)$  que conlleva el cálculo de DTW.

### 3.4.2 Reducción de la numerosidad

Muchos algoritmos han sido propuestos en el campo de la clasificación de series temporales. Como ya se ha dicho, aquellos basados en los  $k$  vecinos más cercanos que implementan DTW como función de distancia, son los que mejores resultados ofrecen y por consiguiente son difíciles de mejorar. Sin embargo existe una dificultad significativa en cuanto al tiempo de obtención de los resultados con este tipo de algoritmos; en gran medida producto de la numerosidad de los datos que generalmente conforman las series temporales, su alta dimensionalidad y la necesidad de su constante actualización.

Todo esto, sumado a la demanda de resultados inmediatos, por aplicaciones en tiempo real que los requieren, trae consigo la necesidad de ejecutar la clasificación lo más rápidamente posible. En este sentido, se ha trabajado mucho para acelerar la velocidad en los cálculos de DTW. Numerosos avances se han dado, no obstante existe un claro límite en cuanto a estas mejoras; incluso se ha sugerido la existencia de un límite asintótico en cuanto a qué tanto se podría mejorar la eficiencia de DTW (Ratanamahatana and Keogh, 2005).

La idea de reducir la numerosidad de los datos ofrece ventajas adicionales (Pełalska *et al.*, 2006). Es bien conocido que, si se escogen con cuidado las instancias que va a descartar un clasificador, esto puede reducir significativamente el tiempo de ejecución de la clasificación, a la vez que se mantiene la efectividad del clasificador (en muchos casos incluso los resultados obtenidos son mejores luego de efectuada la reducción).

Investigaciones recientes han mostrado que la utilización de DTW con restricciones óptimas, unido al uso de algoritmos que reducen convenientemente la numerosidad de los

datos, dan como resultado conjuntos de datos muy compactos y con poca o ninguna pérdida de precisión en la clasificación de series temporales (Xi *et al.*, 2006).

La idea del método de reducción basado en jerarquías expuesto en (Xi *et al.*, 2006) consiste en la eliminación de aquellas series que menos aporten a la clasificación, teniendo siempre presente la interdependencia entre el atributo clase y los valores de los demás atributos de la serie. La idea intuitiva en que se inspira el algoritmo es simple: si la eliminación de una instancia  $p$ , en un conjunto de datos  $S$ , no produce que otras instancias en  $S$  sean mal clasificadas entonces  $p$  puede ser extraída de  $S$  sin que por ello se afecte la clasificación.

El método que funciona en forma de filtro supervisado tiene dos etapas principales, las cuales se denominan “jerarquización” y “desempate”. Primeramente se asigna una jerarquía a cada una de las series temporales (instancias), que van a ser clasificadas con 1NN-DTW, según su aporte a la clasificación de todo el conjunto de datos. Una vez definidas las clases que van a ser reducidas y el porcentaje que se desea reducir de las mismas, se aplican ambas etapas del algoritmo. Finalmente se obtiene un conjunto de datos reducido el cual contiene aquellas series con mayor valor de jerarquía durante cada etapa.

Durante la “jerarquización”, se comienza con la eliminación de todas aquellas series duplicadas (si existen), pues estas no le aportan información nueva al clasificador. Luego se aplica 1NN-DTW en todo el conjunto de datos, asignándole una menor jerarquía a aquellas series que ofrecen información “ruidosa” para la clasificación. Esto es debido a que estas series generalmente afectan la clasificación de otras series cercanas. Para cada serie se le asigna un valor de jerarquía según la ecuación (3.5):

$$jerarquía(x) = \sum_j \begin{cases} 1 & \text{si } clase(x) = clase(x_j) \\ -2 & \text{en otro caso} \end{cases} \quad (3.5)$$

Donde  $x_j$  es la serie que tiene a  $x$  como su vecino más cercano. Por consiguiente, se le asigna mayor jerarquía a aquellas series que más aportan en la clasificación del resto, y aquellas que peor lo hacen obtienen valores negativos.

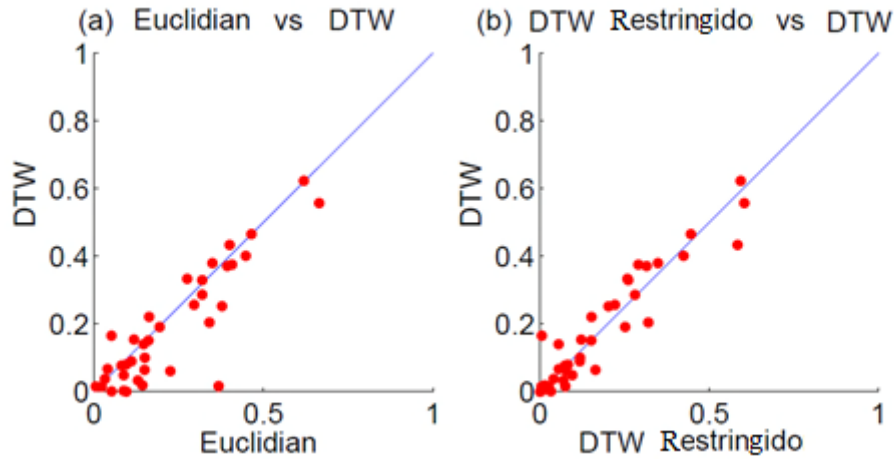
Si dos series tienen la misma jerarquía, entonces el empate se rompe asignándoles diferentes prioridades; esta es la etapa denominada “desempate”. La prioridad de una serie  $x$  es calculada según la ecuación siguiente (3.6):

$$prioridad(x) = \sum_j \left( \frac{1}{(DTW(x, x_j))^2} \right) \quad (3.6)$$

Donde  $x_j$  es la serie que tiene a  $x$  como su vecino más cercano y  $DTW(x, x_j)$  representa el cálculo de DTW entre  $x$  y  $x_j$ . El supuesto en este caso es que si una serie está demasiado alejada de su vecino más cercano, entonces este ejerce una menor influencia en la clasificación de la serie vecina. Luego, si dos series tienen la misma jerarquía, aquella con la menor prioridad será descartada primero. Notar que, gracias a que en la primera etapa se han eliminado las instancias duplicadas, se puede asegurar que el denominador de la fracción será distinto de cero. Teniendo en cuenta que el algoritmo resulta costoso, al tener varios cálculos de DTW, se recomienda la utilización de una cota inferior para acelerar los cálculos.

### 3.5 Eficacia de la métrica DTW en comparación con la distancia euclidiana

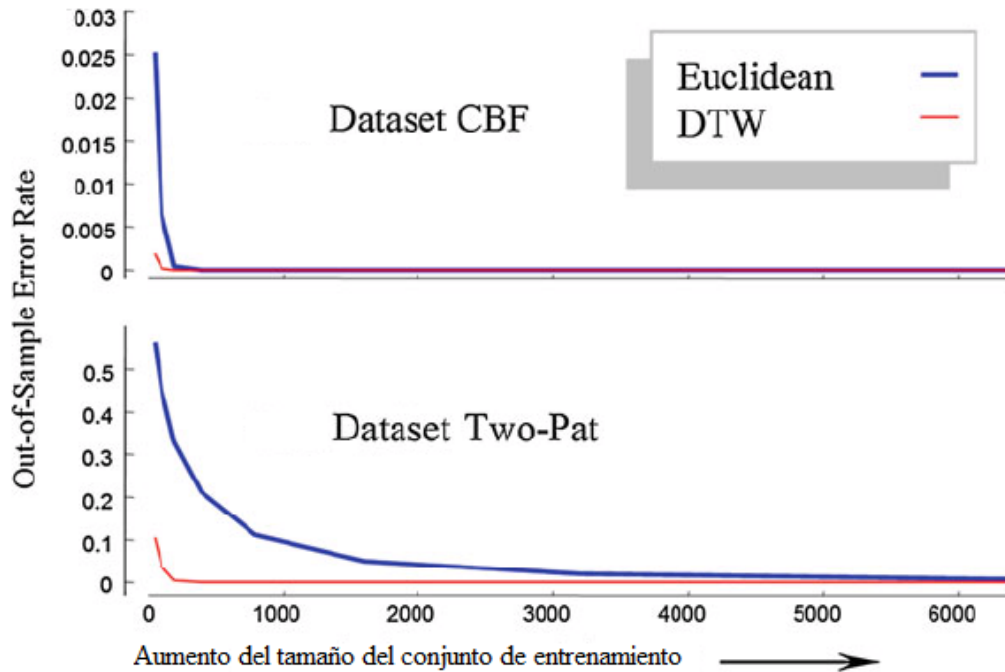
En Wang *et al.* (2013) se lleva a cabo una evaluación general de DTW comparándola con otras métricas de distancia conocidas. Solamente haremos alusión a las comparaciones de nuestro interés. La Figura 3.2 muestra los resultados obtenidos para una colección de 38 conjuntos de datos de series temporales, en la cual cada punto rojo representa un conjunto de datos, y los ejes de coordenadas los índices de error de cada una de las métricas. Las tasas de error de cada conjunto de datos son utilizadas como coordenadas del punto que lo representa. La comparación por pares “A contra B”, un punto rojo sobre la línea indica que “A” es superior a “B”:



**Figura 3.2. Comparación entre las medidas de similitud A) Euclidiana contra DTW sin restricciones globales B) DTW sin restricciones globales contra DTW con un ancho de ventana igual al 10 % del conjunto de datos**

Como se muestra en la figura anterior, se reafirma que la distancia DTW sin restricciones globales es claramente superior a la euclidiana. Además se aprecia que DTW con un ancho de ventana igual al 10 % del conjunto de datos (el cual no tiene por qué ser el tamaño óptimo) es casi igual (e incluso ligeramente superior) a DTW sin restricciones globales. Por lo que se valida el uso de restricciones globales en lugar de realizar el cálculo completo de DTW, reduciendo de esta forma el costo computacional empleado en su cómputo.

Por otro lado, se ha comprobado empíricamente que tanto los porcentajes de clasificación correcta como la velocidad de los cálculos amortizados dependen en gran medida de la numerosidad del conjunto de datos a utilizar. Como una forma de ilustrar la afirmación anterior, en Wang *et al.* (2013) se realizaron experimentos individuales en los conjuntos de datos Two Patterns y CBF (por ser estos los más utilizados en la literatura para comprobar la superioridad de una u otra medida de similitud). Debido a que ambos son conjuntos de datos sintéticos, es posible generar tantas instancias de ellos como se deseen, por lo que existen versiones con múltiples tamaños. Se midieron los índices de errores durante la clasificación usando 1-NN, obtenidos con la distancia euclidiana y con DTW, como se muestra en la Figura 3.3. Como se puede apreciar, la tasa de error relativo de la distancia DTW es significativamente menor a la euclidiana, sobre todo cuando los conjuntos de datos son menos numerosos.



**Figura 3.3. Tasa de error relativo para  $I$ -NN usando la distancia euclidiana y DTW, al incrementar la numerosidad de las series en dos conjuntos de datos tradicionales**

Para el conjunto de datos CBF, cuando el número de series temporales supera las 400, no hay diferencias estadísticas significativas entre una y otra métrica de distancia. En el caso de Two Patterns, la distancia euclidiana necesita de un aumento mucho mayor en la numerosidad del conjunto de datos para converger a la precisión que exhibe la métrica DTW.

Esto pudiera parecer desalentador, teniendo en cuenta que la distancia euclidiana tiene una complejidad temporal de  $O(n)$  y que un cálculo simple de DTW tiene una complejidad de  $O(n, r)$ , donde  $r$  es al ancho de la ventana luego de aplicarse la restricción global de Sakoe-Shiba. No obstante la complejidad amortizada de DTW durante la clasificación es realmente de  $O((P \cdot n) + (1 - P) \cdot n \cdot r)$ , donde  $P$  es la fracción de cálculos de DTW que se omiten con el uso del algoritmo LB\_Keogh para la búsqueda por similitud en el algoritmo  $I$ -NN.

### **3.6 Conclusiones del capítulo**

La distancia DTW constituye uno de los aportes más significativos en la minería de datos para series temporales. Aunque surge en la década del setenta se ha popularizado a partir de esa fecha y actualmente es una de las más utilizadas en las tareas de la minería de datos para series temporales. Son incontables los dominios donde se ha aplicado exitosamente tales como análisis de electrocardiogramas, bioinformática, control de procesos, por solo mencionar algunos. Debido a su flexibilidad se han propuesto innumerables variantes, entre ellas las restricciones globales y las distintas formas de estimar la distancia acumulada. La búsqueda de cotas inferiores, para agilizar los cálculos durante los algoritmos de búsqueda por similitud, también ha sido otra temática abordada en las investigaciones. Debido a su probada fiabilidad es considerada actualmente una de las métricas de referencia en el área de la minería de datos para series temporales.

## Referencias bibliográficas

- AGRAWAL, R., FALOUTSOS, C. & SWAMI, A. 1993. *Efficient similarity search in sequence databases*, Springer.
- AREF, W. G., ELFEKY, M. G. & ELMAGARMID, A. K. 2004. Incremental, online, and merge mining of partial periodic patterns in time-series databases. *Knowledge and Data Engineering, IEEE Transactions on*, 16, 332-342.
- ÅSTRÖM, K. J. 1969. On the choice of sampling rates in parametric identification of time series. *information Sciences*, 1, 273-278.
- BOX, G. E. P. & JENKINS, G. M. 1976. Time series analysis: Forecasting and control (rev. ed.) Holden-Day. *San Francisco*, 575.
- BROCKWELL, P. J. & DAVIS, R. A. 2002. *Introduction to time series and forecasting*, Taylor & Francis US.
- CAMONES, F. C. 2002. Control de calidad en los procesos estadísticos. Una aproximación basada en los modelos ARIMA con Análisis de Intervención. Lima: INEI.
- CARAÇA-VALENTE, J. P. & LÓPEZ-CHAVARRÍAS, I. Discovering similar patterns in time series. 2000. ACM, 497-505 %@ 1581132336.
- CHAN, P. K. & MAHONEY, M. V. Modeling multiple time series for anomaly detection. 2005. IEEE, 8 pp. %@ 0769522785.
- CHATFIELD, C. 2003. *The analysis of time series: an introduction*, CRC press.
- CHUNG, F.-L., FU, T.-C., NG, V. & LUK, R. W. P. 2004. An evolutionary approach to pattern-based time series segmentation. *Evolutionary Computation, IEEE Transactions on*, 8, 471-489.
- COWPERTWAIT, P. S. & METCALFE, A. V. 2009. *Introductory time series with R*, Springer.
- DAS, G., LIN, K.-I., MANNILA, H., RENGANATHAN, G. & SMYTH, P. 1998. Rule discovery from time series. *Knowledge Discovery and Data Mining*, 16-22.
- DOUZAL-CHOUAKRIA, A. & AMBLARD, C. 2012. Classification trees for time series. *Pattern Recognition*, 45, 1076-1091.
- FALOUTSOS, C., RANGANATHAN, M. & MANOLOPOULOS, Y. 1994. *Fast subsequence matching in time-series databases*, ACM.
- FU, T.-C. 2011. A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 24, 164-181.
- GEURTS, P. 2001. Pattern extraction for time series classification. *Principles of Data Mining and Knowledge Discovery*. Springer.
- GIORGINO, T. 2009. Computing and visualizing dynamic time warping alignments in R: the dtw package. *Journal of Statistical Software*, 31, 1-24.
- GONZALEZ-ABRIL, L., VELASCO, F., ORTEGA, J. A. & CUBEROS, F. J. 2009. A new approach to qualitative learning in time series. *Expert Systems with Applications*, 36, 9924-9927.



- HAN, J. & KAMBER, M. 2001. Data mining: Concepts and techniques. *China Machine Press*, 8, 3-6.
- HAN, W.-S., LEE, J., MOON, Y.-S. & JIANG, H. Ranked subsequence matching in time-series databases. 2007. VLDB Endowment, 423-434 %@ 1595936491.
- HOCHHEISER, H. & SHNEIDERMAN, B. 2004. Dynamic query tools for time series data sets: timebox widgets for interactive exploration. *Information Visualization*, 3, 1-18
- ITAKURA, F. 1975. Minimum prediction residual principle applied to speech recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 23, 67-72.
- KEOGH, E. Fast similarity search in the presence of longitudinal scaling in time series databases. 1997. IEEE, 578-584
- KEOGH, E., CHAKRABARTI, K., PAZZANI, M. & MEHROTRA, S. Locally adaptive dimensionality reduction for indexing large time series databases. 2001. ACM, 151-162.
- KEOGH, E., HOCHHEISER, H. & SHNEIDERMAN, B. 2002a. An augmented visual query mechanism for finding patterns in time series data. *Flexible Query Answering Systems*. Springer.
- KEOGH, E., LONARDI, S. & CHIU, B. Y.-C. Finding surprising patterns in a time series database in linear time and space. 2002b. ACM, 550-556 %@ 158113567X.
- KEOGH, E. & RATANAMAHATANA, C. A. 2005. Exact indexing of dynamic time warping. *Knowledge and information systems*, 7, 358-386.
- KEOGH, E., ZHU, Q., HU, B., HAO, Y., XI, X., WEI, L. & RATANAMAHATANA, C. A. 2011. *The UCR Time Series Classification/Clustering Homepage* [Online]. Available: [www.cs.ucr.edu/~eamonn/time\\_series\\_data/](http://www.cs.ucr.edu/~eamonn/time_series_data/)
- KEOGH, E. J. & PAZZANI, M. J. Derivative dynamic time warping. 2001.
- KIM, S.-W., PARK, S. & CHU, W. W. An index-based approach for similarity search supporting time warping in large sequence databases. Data Engineering, 2001. Proceedings. 17th International Conference on, 2001. IEEE, 607-614.
- KRUSKALL, J. B., LIBERMAN 1983. *The symmetric time warping algorithm: from continuous to discrete, TimeWarps, String Edits and Macromolecules*.
- LAGUS, K., HONKELA, T., KASKI, S. & KOHONEN, T. Self-organizing maps of document collections: A new approach to interactive exploration. 1996. Menlo Park, CA: AAAI, 238-243.
- LAROSE, D. T. 2005. *Discovering knowledge in data: an introduction to data mining*, Wiley. com.
- LERNER, A., SHASHA, D., WANG, Z., ZHAO, X. & ZHU, Y. Fast algorithms for time series with applications to finance, physics, music, biology, and other suspects. 2004. ACM, 965-968 %@ 1581138598.
- LIN, J., KEOGH, E. & LONARDI, S. 2005. Visualizing and discovering non-trivial patterns in large time series databases. *Information Visualization*, 4, 61-82
- MA, J. & PERKINS, S. Online novelty detection on temporal sequences. 2003. ACM, 613-618 %@ 1581137370.

- MEESRIKAMOLKUL, W., NIENNATTRAKUL, V. & RATANAMAHATANA, C. A. 2012. Shape-Based clustering for time series data. *Advances in knowledge discovery and data mining*. Springer.
- MOON, Y.-S., WHANG, K.-Y. & HAN, W.-S. General match: a subsequence matching method in time-series databases based on generalized windows. 2002. ACM, 382-393 %@ 1581134975.
- MOON, Y.-S., WHANG, K.-Y. & LOH, W.-K. Duality-based subsequence matching in time-series databases. 2001. IEEE, 263-272
- MUEEN, A., KEOGH, E., ZHU, Q., CASH, S. S., WESTOVER, M. B. & BIGDELY-SHAMLO, N. 2011. A disk-aware algorithm for time series motif discovery. *Data Mining and Knowledge Discovery*, 22, 73-105.
- NIENNATTRAKUL, V., RUENGRONGHIRUNYA, P. & RATANAMAHATANA, C. A. 2010. Exact indexing for massive time series databases under time warping distance. *Data Mining and Knowledge Discovery*, 21, 509-541.
- OATES, T. Identifying distinctive subsequences in multivariate time series by clustering. 1999. ACM, 322-326 %@ 1581131437.
- ORSENIGO, C. & VERCELLIS, C. 2010. Combining discrete SVM and fixed cardinality warping distances for multivariate time series classification. *Pattern Recognition*, 43, 3787-3794.
- PANUCCIO, A., BICEGO, M. & MURINO, V. 2002. A Hidden Markov Model-based approach to sequential data clustering. *Structural, Syntactic, and Statistical Pattern Recognition*. Springer.
- PEKALSKA, E., DUIN, R. P. W. & PACLIK, P. 2006. Prototype selection for dissimilarity-based classifiers. *Pattern Recognition*, 39, 189-208.
- POVINELLI, R. J. 1999. *Time series data mining: identifying temporal patterns for characterization and prediction of time series events*. Faculty of the Graduate School, Marquette University.
- POVINELLI, R. J., JOHNSON, M. T., LINDGREN, A. C. & YE, J. 2004. Time series classification using Gaussian mixture models of reconstructed phase spaces. *Knowledge and Data Engineering, IEEE Transactions on*, 16, 779-783.
- RATANAMAHATANA, C. A. & KEOGH, E. Everything you know about dynamic time warping is wrong. 2004a. 22-25.
- RATANAMAHATANA, C. A. & KEOGH, E. Making time-series classification more accurate using learned constraints. 2004b. Lake Buena Vista, Florida, 11-22.
- RATANAMAHATANA, C. A. & KEOGH, E. Three myths about dynamic time warping data mining. 2005. 506-510.
- RATH, T. M. & MANMATHA, R. Word image matching using dynamic time warping. 2003. IEEE, II-521-II-527 vol. 2 %@ 0769519008.
- RODPONGPUN, S., NIENNATTRAKUL, V. & RATANAMAHATANA, C. A. 2012. Selective subsequence time series clustering. *Knowledge-Based Systems*, 35, 361-368.

- RODRÍGUEZ, J. J. & ALONSO, C. J. Interval and dynamic time warping-based decision trees. 2004. ACM, 548-552 %@ 1581138121.
- RODRÍGUEZ, J. J., ALONSO, C. J. & BOSTRÖM, H. Learning first order logic time series classifiers. Proceedings of the 10th International Workshop on Inductive Logic Programming, 2000. Citeseer, 260-275.
- SAKOE, H. & CHIBA, S. 1978. Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 26, 43-49.
- SAKURAI, Y., YOSHIKAWA, M. & FALOUTSOS, C. FTW: fast similarity search under the time warping distance. Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, 2005. ACM, 326-337.
- SHUMWAY, R. H. & STOFFER, D. S. 2011. *Time series analysis and its applications: with R examples*, Springer.
- SMYTH, P. & KEOGH, E. Clustering and mode classification of engineering time series data. 1997. Citeseer, 24-30.
- VAN WIJK, J. J. & VAN SELOW, E. R. Cluster and calendar based visualization of time series data. 1999. IEEE, 4-9, 140 %@ 0769504310.
- WANG, H., WANG, W., YANG, J. & YU, P. S. Clustering by pattern similarity in large data sets. 2002. ACM, 394-405 %@ 1581134975.
- WANG, X., MUEEN, A., DING, H., TRAJCEVSKI, G., SCHEUERMANN, P. & KEOGH, E. 2013. Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery*, 26, 275-309.
- WEBER, M., ALEXA, M. & MÜLLER, W. Visualizing time-series on spirals. 2001. 7.
- WEI, L. & KEOGH, E. Semi-supervised time series classification. Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 2006. ACM, 748-753.
- XI, X., KEOGH, E., SHELTON, C., WEI, L. & RATANAMAHATANA, C. A. Fast time series classification using numerosity reduction. 2006. ACM, 1033-1040 %@ 1595933832.
- YI, B.-K., JAGADISH, H. & FALOUTSOS, C. Efficient retrieval of similar time sequences under time warping. Data Engineering, 1998. Proceedings., 14th International Conference on, 1998. IEEE, 201-208.
- YI, B. K., & Faloutsos, C.(2000). Fast time sequence indexing for arbitrary Lp norms.
- ZHANG, H., HO, T. B. & LIN, M. S. 2004. A non-parametric wavelet feature extractor for time series classification. *Advances in knowledge discovery and data mining*. Springer.