

¿Requiere la clasificación de series temporales métodos específicos?

Amaia Abanda^{1,2}

¹ Basque Center for Applied Mathematics (BCAM)

² Intelligent Systems Group (ISG)

Department of Computer Science and Artificial Intelligence,
University of the Basque Country UPV/EHU

Bilbao, Spain

aabanda@bcamath.org

Usue Mori^{2,3}

² Intelligent Systems Group (ISG)

Department of Computer Science and Artificial Intelligence

³ Department of Applied Mathematics
Statistics and Operational Research

University of the Basque Country UPV/EHU

Bilbao, Spain

usue.mori@ehu.es

Jose A. Lozano^{1,2}

¹ Basque Center for Applied Mathematics (BCAM)

² Intelligent Systems Group (ISG)

Department of Computer Science and Artificial Intelligence,

University of the Basque Country UPV/EHU

Bilbao, Spain

ja.lozano@ehu.es

Resumen—La clasificación de series temporales tiene la peculiaridad, respecto a otros problemas de clasificación supervisada, de que las observaciones de las series o variables predictoras tienen un orden específico. La mayoría de las soluciones propuestas en la literatura consideran que este orden es discriminatorio para la clasificación y, por tanto, emplean métodos específicos que tienen en cuenta el orden. El objetivo de esta investigación es explorar, de una manera preliminar, si realmente es siempre necesario el uso de métodos específicos para series temporales o si hay algunos casos en los que los métodos de clasificación no específicos, habitualmente utilizados para datos de tipo convencional, obtienen mejores resultados que los específicos. La experimentación llevada a cabo en 40 bases de datos del repositorio UCR muestra como en los casos en los que el orden de las observaciones no es relevante para la clasificación, los clasificadores no específicos consiguen mejorar la precisión, mientras que en los casos en los que el orden es un factor clave no lo consiguen.

Index Terms—Clasificación, series temporales, orden, vector, información temporal, métodos específicos.

I. INTRODUCTION

Una serie temporal es una secuencia de datos u observaciones que viene ordenada respecto al tiempo -la mayoría de las veces-, o respecto a otros aspectos como el espacio, por ejemplo. Es un tipo particular de datos precisamente porque tiene una naturaleza ordinal que la mayoría de datos no tienen. Las series temporales aparecen naturalmente en diferentes áreas como la bio-informática o economía [1], y están adquiriendo una gran relevancia en el ámbito del aprendizaje automático debido al gran reto que supone trabajar con este tipo especial de datos. De esta manera, continuamente están surgiendo nuevos métodos para representar, indexar, agrupar y clasificar

series temporales [2]. El presente trabajo se sitúa dentro del área de la clasificación de series temporales -Time Series Classification (TSC)-, donde cada serie del conjunto de entrenamiento tiene asociada una clase, y el objetivo es encontrar una función tal que, dada una serie nueva, sea capaz de predecir cuál es su clase. La diferencia principal con el problema clásico de clasificación es que, mientras en éstos las instancias vienen descritas por atributos sin orden específico, en la clasificación de series temporales las instancias vienen definidas por las propias series temporales completas [3][4]. Además de que las variables predictoras están ordenadas en este caso, las series pueden tener un número muy alto de observaciones, las longitudes pueden ser variables dentro de una misma base de datos y, dependiendo del contexto, pueden ser observaciones con mucho ruido [5].

La peculiaridad de las series temporales, su posible interpretabilidad y su significado semántico han llevado a la comunidad científica a asumir que éstas tienen que ser tratadas con métodos específicamente diseñados para series [4][6][7][8]. Estos métodos tienen en cuenta la información temporal que intrínsecamente contienen las series para extraer atributos, construir modelos, medidas de similitud, etc. útiles para la clasificación. El orden de las observaciones ha sido hasta ahora una de las características más comúnmente asumidas como información temporal discriminatoria; en los métodos basados en distancias elásticas [3][9][10], en los métodos basados en modelos autorregresivos [11][12] en los modelos ocultos de Markov [13], o en los métodos extraen atributos temporales de las series [5].

Esta asunción implica que los métodos no específicos, es decir, los métodos basados en clasificadores tradicionales



que no tienen en cuenta la información temporal, no deberían ser competitivos con los específicos en cuanto a precisión se refiere. En este trabajo se pretende explorar hasta qué punto esta afirmación es cierta y, sobre todo, en qué casos se cumple o no se cumple.

Este trabajo parte de la clasificación de series temporales basada en distancias [14], donde se utiliza la distancia/similitud entre series como criterio para la clasificación; dos series que sean similares pertenecerán a la misma clase. En particular, el método que hasta ahora más se ha utilizado para en la clasificación de series temporales basada en distancias es el 1-Nearest Neighbour (1NN), clasificador que aunque haya demostrado buenos resultados en el ámbito [2][9][15][16], no es muy robusto. En cuanto a las distancias utilizadas, en el ámbito de la clasificación de series temporales, podemos distinguir entre distancias que tienen en cuenta el orden de las observaciones y distancias que no lo tienen, pero muchas de las distancias situadas en el segundo grupo son muy costosas computacionalmente. En este contexto, sería interesante saber en qué casos no es necesario el uso de métodos específicos -casos en los que se podrían utilizar clasificadores estándares de una manera eficiente y con mejores resultados-.

El trabajo se organiza de la siguiente forma: en la Sección II se introducen algunos conceptos básicos para contextualizar la clasificación de series temporales basada en distancias, en la Sección III se presenta la propuesta y la experimentación realizada y, finalmente, en la Sección IV presentamos las principales conclusiones obtenidas.

II. CONCEPTOS BÁSICOS Y MOTIVACIÓN

En esta sección se presenta un breve resumen de los conceptos básicos necesarios para abordar la clasificación de series temporales basada en distancias y se introducen algunas de las distancias más conocidas para series.

II-A. Clasificación de series temporales

Los métodos de clasificación de series temporales se pueden dividir entre tres principales categorías [15]: métodos basados en atributos, basados en modelos y basados en distancias. Dentro de la primera línea, los investigadores tratan de extraer características fundamentales de las series para obtener una nueva representación no-ordenada y de menor dimensión que, además, contenga la información más relevante de las series. Esta transformación evita tener que aprender directamente con las series originales y las traslada a un nuevo espacio donde las características discriminatorias pueden ser más detectables [17]. Algunos ejemplos de estas representaciones incluyen Discrete Fourier Transformation (DFT) [5], Discrete Wavelet Transformation (DWT) [18] o Piecewise Aggregate Approximation (PAA) [19]. Por otro lado, en los métodos basados en modelos, se asume que todas las series de una clase han sido generadas por un modelo subyacente y a una serie nueva se le asigna la clase del modelo al que mejor se ajuste. Algunos de los modelos más utilizados son los

modelos auto-regresivos [11] [12] o los modelos ocultos de Markov [13]. Por último, en los métodos basados en distancias los investigadores tratan de definir una similitud, o disimilitud, entre series que tenga en cuenta diferentes características semánticas o temporales. La definición de una distancia adecuada es una cuestión crucial, ya que cada distancia refleja diferentes características de las series. Una vez elegida la distancia adecuada, la clasificación se lleva a cabo empleando métodos basados en distancias, en la mayoría de los casos clasificadores basados en el vecino más cercano (1NN). Entre las distancias para series temporales más conocidas están la distancia Euclídea, Dynamic Time Warping (DTW) [20] o Edit Distance with RealPenalty (ERP) [21].

II-B. Distancias de series temporales

Las distancias entre series temporales suelen categorizarse habitualmente en dos grupos principales mostradas en la Figura 1 [22]; las *medidas rígidas*, que se refieren a aquellas medidas que comparan el punto i -ésimo de una serie con el punto i -ésimo de la otra, y las *medidas elásticas*, que tratan de crear un mapeo no lineal entre las series para alinearlas, permitiendo una comparación de uno-a-varios puntos. La mayor diferencia entre estos dos tipos de medidas es que las medidas rígidas, al no considerar los puntos de alrededor, tratan la serie como si fuera un vector y por tanto, no consideran la ordenación de las mediciones. De esta forma, aún desordenando la serie, obtendríamos los mismos resultados de clasificación. Las medidas elásticas, por el contrario, consideran que el orden de las observaciones es determinante y hacen uso de la misma en su cálculo. Por tanto, otra ordenación de las observaciones podría cambiar por completo los resultados de la clasificación.

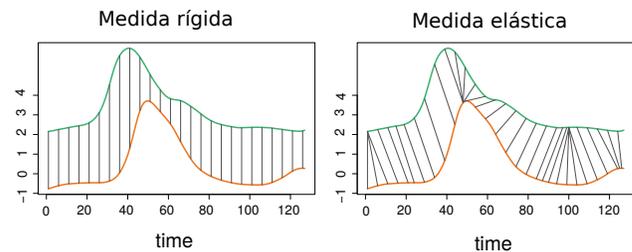


Figura 1: Alineamientos de medidas rígidas y elásticas

La distancia rígida más habitual es la denominada distancia Euclídea. Se define de la siguiente forma: dadas dos series temporales $T=(t_1, \dots, t_n)$ y $S=(s_1, \dots, s_n)$,

$$ED(T, S) = \sqrt{\sum_{i=1}^n (t_i - s_i)^2}$$

Esta distancia es ampliamente conocida debido que cumple buenas propiedades, como ser métrica, así como su bajo coste computacional y simplicidad, razones por las

que es habitualmente utilizada en numerosos problemas de aprendizaje automático. Sin embargo, tiene también algunos inconvenientes como que sólo acepta dos series que tengan la misma longitud, o como se ha comentado anteriormente, que no considera el orden de las observaciones.

En vista de las carencias de la distancia Euclídea, en los últimos años se han presentado numerosas distancias elásticas específicamente diseñadas para medir la disimilitud entre series temporales. La más habitual entre ellas es la Dynamic Time Warping (DTW). Esta distancia alinea las dos series con tal de minimizar la distancia entre ellas y, por lo tanto, es robusta frente a desfases y distorsiones en el tiempo. Además, a diferencia de la distancia Euclídea, permite calcular la distancia entre series de diferentes longitudes. Dadas dos series temporales T y S , el objetivo de la DTW es buscar la alineación óptima que minimice la distancia entre ellas. Para ello, el primer paso es construir una matriz de distancias de $n \times m$ donde cada posición (i, j) contiene la distancia $(t_i - s_j)^2$, que representa el coste de alinear la observación i -ésima de T con la observación j -ésima de S . Así, un alineamiento entre dos series se define como un camino π en la matriz de distancias, que tiene que cumplir ciertas restricciones (continua, monótona creciente y puntos de inicio y final fijos) (ver Figura 2). El alineamiento óptimo entre dos series es aquel que minimiza la distancia acumulativa. Finalmente, la distancia $DTW(T, S)$ se define como la distancia acumulativa del camino óptimo entre T y S . En la práctica, este camino óptimo se encuentra mediante métodos de programación dinámica [23] y uno de las mayores desventajas de ésta distancia es que este proceso tiene un coste computacional de $\mathcal{O}(N^2)$, siendo N el número de series.

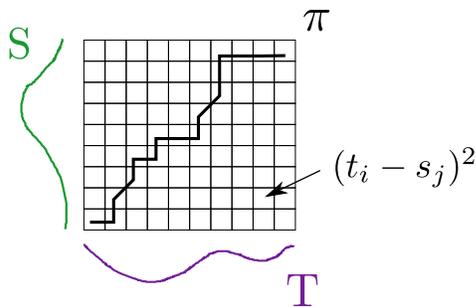


Figura 2: Matriz de distancias y camino óptimo de DTW

La mayoría de los métodos de clasificación de series temporales asumen que el orden de las observaciones es fundamental para discriminar entre clases y emplean medidas elásticas como la DTW [24][16][10]. El objetivo de este trabajo es realizar un estudio preliminar para explorar si realmente es, o en qué casos, la información temporal -en particular, el orden de las observaciones- fundamental para la clasificación. Esta cuestión proporcionaría información importante sobre la utilidad de clasificadores específicos

para series temporales, ya que si el orden de las observaciones no es relevante, no habría necesidad de usar métodos específicos -clasificadores basados en la temporalidad-, sino que las series podrían entenderse como simples vectores y podrían utilizarse también otros clasificadores clásicos, obteniendo quizás resultados mejores.

III. PROPUESTA Y EXPERIMENTACIÓN

Como hemos expuesto anteriormente, el objetivo de este trabajo es explorar si realmente la información temporal y, en particular, el orden de las observaciones, juega siempre un papel fundamental en la clasificación de series temporales. Además, la hipótesis principal es que en aquellos casos donde el orden no es discriminatorio, los clasificadores no específicos podrían funcionar mejor, debido a que son más complejos y robustos que el 1-NN. Es decir, en los casos donde el orden no es relevante, las series pueden ser interpretadas como vectores (*feature vectors*) y, por tanto, las observaciones puede entenderse como dimensiones que no tienen un orden pre-definido. En los casos donde el orden es un aspecto discriminatorio para la clasificación, el hecho de que cada observación esté en una posición concreta es relevante y por tanto no se pueden interpretar las series como vectores. Los clasificadores no específicos, o tradicionales, esperan como entrada un vector sin orden, por lo que en el primer caso podrían funcionar bien mientras que en el segundo, al no ser capaces de tener en cuenta el orden, deberían obtener peores resultados.

El estudio parte de la clasificación basada en distancias y, más particularmente, tomamos el clasificador 1-NN como referencia. Como distancia que no considera el orden de las observaciones se utiliza la distancia Euclídea, mientras que como distancia que sí considera el orden consideraremos la DTW.

Asumiremos que en los casos en los que la distancia Euclídea obtiene mayor precisión que la DTW, el orden de las observaciones no es relevante para la discriminación, mientras que en las que la distancia DTW obtiene mayor precisión sí lo es. El objetivo es explorar si en los casos en los que 1-NN-EUC (método no específico) tiene mayor precisión que el 1-NN-DTW (método específico), los clasificadores estándares (métodos no específicos), como el Support Vector Machine o naive Bayes, mejoran más (veces) los resultados que en los casos en los que el 1-NN-DTW tiene mayor precisión que el 1-NN-EUC. En particular, se quiere comprobar si, en el caso de que el 1-NN-EUC obtiene mejores resultados que el 1-NN-DTW, los clasificadores estándares pueden llegar a mejorar los resultados obtenidos con el 1-NN.

Con el objetivo de poder contrastar las hipótesis planteadas, la clasificación se ha llevado a cabo utilizando, por un lado, 1-NN-Euclídea y 1-NN-DTW y, por otro, 3 clasificadores estándares (no específicos): Support Vector Machine (SVM), naive Bayes (NB) y Random Forest (RF). No se han ajustado los parámetros de ninguno para poder establecer una base de referencia. Los experimentos se han



realizado con 40 bases de datos del repositorio de series temporales UCR [25], un repositorio de bases de datos de series temporales que es habitualmente utilizada como referencia para evaluar nuevos métodos de clasificación de series temporales.

En el Cuadro I se muestran los resultados de la clasificación de las bases de datos donde el 1-NN-EUC obtiene mayor precisión que el 1-NN-DTW (15/40), mientras que en el Cuadro II la de aquellas en las que el 1-NN-DTW obtiene mayor precisión que 1-NN-EUC (25/40). Estas proporciones eran de esperar debido a que el 1-NN-DTW se considera uno de los métodos que mejores resultados obtienen en la clasificación de series temporales [2][9][16][15]. Se han separado las bases de datos en estas dos tablas para poder visualizar mejor si hay una tendencia diferente en cada uno de los casos.

En el primer cuadro, donde el método no específico obtiene mejores resultados que el específico, se puede observar que los clasificadores estándares superan en algunos casos (12/15) al 1-NN. En el segundo cuadro, en cambio, donde el método específico funciona mejor que el no específico, los clasificadores estándares no superan al 1-NN en casi ningún caso (6/25).

BBDD	EUC	DTW	SVM	NB	RF
Adiac	0.61	0.52	0.40	0.57	0.64
Beef	0.53	0.50	0.50	0.47	0.57
Chlothe	0.65	0.62	0.58	0.35	0.71
CinC	0.90	0.69	0.73	0.88	0.76
ECG	0.88	0.80	0.87	0.77	0.83
ECGF5	0.80	0.75	0.84	0.78	0.80
GunPoint	0.91	0.89	0.81	0.78	0.92
Haptics	0.37	0.35	0.45	0.44	0.46
SonyAI	0.70	0.66	0.78	0.94	0.68
SonyAI II	0.86	0.82	0.82	0.79	0.80
Swedish	0.79	0.75	0.88	0.86	0.88
WaveX	0.74	0.73	0.76	0.66	0.76
WaveY	0.66	0.64	0.67	0.56	0.69
Wafer	0.99	0.98	0.99	0.70	0.99
FetalECG2	0.88	0.83	0.90	0.83	0.91

Cuadro I: Resultados de clasificación de las bases de datos donde 1-NN-Euc obtiene mayor precisión que 1-NN-DTW.

Con fin de entender mejor los resultados, el Cuadro III sintetiza y compara los dos cuadros anteriores. En él se refleja para cada una de las tablas (representadas por las filas EUC y DTW) el número de veces en el que al menos uno -y al menos dos- de los clasificadores no-específicos empleados supera la precisión del 1-NN. Se puede apreciar que en los casos del Cuadro I (fila EUC), los métodos no-específicos tienden a superarlo al menos una vez en el 80% de los casos, mientras que los casos del Cuadro II (fila DTW), este porcentaje baja hasta el 24%. Lo mismo sucede si contamos las veces en las que los métodos no-específicos superan al menos 2 veces al 1NN: el valor es 53% para las bases de datos donde la Euclídea funciona mejor y 4% para las del DTW.

Por lo tanto, se puede apreciar una tendencia a que los

BBDD	EUC	DTW	SVM	NB	RF
50Words	0.63	0.71	0.62	0.58	0.66
CBF	0.85	0.97	0.87	0.90	0.91
Coffee	0.75	0.79	0.71	0.68	0.68
Cricket_X	0.57	0.73	0.54	0.44	0.62
Cricket_Y	0.64	0.75	0.63	0.53	0.74
Cricket_Z	0.62	0.73	0.58	0.42	0.66
Diatom	0.93	0.96	0.90	0.87	0.91
Face_all	0.71	0.75	0.74	0.69	0.81
Face_four	0.78	0.84	0.64	0.90	0.77
FacesUCR	0.77	0.92	0.73	0.74	0.79
Fish	0.78	0.85	0.78	0.66	0.79
InlineSkate	0.34	0.38	0.25	0.22	0.35
Lightning2	0.75	0.80	0.70	0.67	0.75
Lightning7	0.58	0.77	0.63	0.71	0.75
MALLAT	0.91	0.92	0.77	0.87	0.93
MedicalImg	0.68	0.73	0.57	0.44	0.73
MoteStrain	0.88	0.90	0.85	0.84	0.89
OSU Leaf	0.52	0.63	0.52	0.38	0.51
Symbols	0.90	0.95	0.76	0.64	0.87
Synthetic	0.88	0.95	0.98	0.96	0.96
Trace	0.76	0.99	0.71	0.80	0.86
Two Lead	0.75	0.93	0.72	0.69	0.73
Two Patterns	0.91	0.99	0.89	0.46	0.85
WaveY	0.65	0.65	0.70	0.56	0.71
WordSyn	0.62	0.67	0.53	0.48	0.57

Cuadro II: Resultados de clasificación de las bases de datos donde 1-NN-Dtw obtiene mayor precisión que 1-NN-Euc.

métodos no-específicos funcionen mejor en los casos en los que la distancia Euclídea funciona mejor que la DTW con el 1-NN. Es decir, en las bases de datos en las que el orden de las observaciones no es relevante para clasificar y los métodos no-específicos -que no tienen en cuenta el orden- funcionan mejor que en las bases de datos que en las que el orden si es relevante.

	≥ 1 vez	≥ 2 veces
EUC	12/15 = 80%	8/15 = 53%
DTW	6/25 = 24%	1/25 = 4%

Cuadro III: Comparativa de las veces que los clasificadores estándares superan el 1-NN en el método no específico (EUC) y específico (DTW).

Para evaluar estadísticamente los Cuadros I y II, se han realizado múltiples test estadísticos emparejados de Wilcoxon utilizando el paquete 'scmamp' [26] para R. En el primer caso, se ha realizado 3 test estadísticos para comprobar si hay diferencias significativas entre los métodos del Cuadro I. En particular, se han comparado las siguientes columnas: 1-NN-EUC con SVM, 1-NN-EUC con NB y 1-NN-EUC con RF. Después de corregir los p-valores obtenidos de cada uno de ellos (0.34, 0.09 y 0.28, correspondientemente), no podemos rechazar la hipótesis nula y, por tanto, no podemos afirmar que haya diferencias estadísticas de que los métodos no-específicos y el 1-NN tengan un comportamiento diferente. En el segundo caso, se han comparado los resultados de los métodos del Cuadro II; en particular, comparando el 1-NN-DTW con SVM,

1-NN-DTW con NB y 1-NN-DTW con RF. Después de corregir los p-valores obtenidos (0.00004, 0.00004 y 0.0008) podemos rechazar la hipótesis nula y afirmar que hay diferencias significativas entre los métodos. En particular, podemos afirmar que los métodos no-específicos tienen peores resultados de clasificación que el 1-NN-DTW. Resumiendo, en los casos donde la distancia Euclídea funciona mejor que la DTW (el orden no tiene importancia), los métodos no-específicos se comportan parecido al 1-NN, mientras que en los casos donde la DTW funciona mejor (el orden sí tiene importancia), funcionan peor que el 1-NN.

Para terminar, se ha realizado otros 2 test estadísticos para comparar si hay diferencias significativas entre los resultados del 1-NN y el mejor de los 3 métodos no-específicos en cada caso (Cuadro IV). En ambos casos, los test afirman que hay diferencias significativas con p-valores 0.01 y 0.002. En los casos donde el 1-NN-EUC obtiene mejores resultados que el 1-NN-DTW, el test indica que al menos uno de los 3 métodos no-específicos obtiene mejores resultados que el 1-NN. En los casos donde el 1-NN-DTW obtiene mejores resultados que el 1-NN-EUC, por el contrario, el test muestra que los 3 métodos específicos obtienen peores resultados que el 1-NN.

BBDD	EUC	Mejor	BBDD	DTW	Mejor
Adiac	0.61	0.64	50Words	0.71	0.66
Beef	0.53	0.57	CBF	0.97	0.91
Chlothe	0.65	0.71	Coffee	0.79	0.71
CinC	0.90	0.88	Cricket_X	0.73	0.62
ECG	0.88	0.87	Cricket_Y	0.75	0.74
ECGF5	0.80	0.84	Cricket_Z	0.73	0.66
GunPoint	0.91	0.92	Diatom	0.96	0.91
Haptics	0.37	0.46	Face_all	0.75	0.81
SonyAI	0.70	0.94	Face_four	0.84	0.90
SonyAI II	0.86	0.82	FacesUCR	0.92	0.79
Swedish	0.79	0.88	Fish	0.85	0.79
WaveX	0.74	0.76	InlineSkate	0.38	0.35
WaveY	0.66	0.69	Lightning2	0.80	0.75
Wafer	0.99	0.99	Lightning7	0.77	0.75
FetalECG2	0.88	0.91	MALLAT	0.92	0.93
			MedicalImg	0.73	0.73
			MoteStrain	0.90	0.89
			OSU Leaf	0.63	0.52
			Symbols	0.95	0.87
			Synthetic	0.95	0.98
			Trace	0.99	0.86
			Two Lead	0.93	0.73
			Two Patterns	0.99	0.89
			WaveY	0.65	0.71
			WordSyn	0.67	0.57

Cuadro IV: Resultados de clasificación del 1-NN y el mejor de los 3 clasificadores estándares en cada base de datos.

IV. CONCLUSIONES

En este trabajo el objetivo era explorar de una manera preliminar si realmente es necesario siempre utilizar los métodos específicos que tienen en cuenta la información temporal, en particular, el orden de las observaciones, para la clasificación de series temporales. La mayoría de los

métodos habitualmente utilizados en clasificación de series temporales, especialmente aquellos basados en distancias, asumen, por defecto, que el orden es un factor discriminatorio para la clasificación y que, por tanto, interpretar las series como vectores sin orden no aportaría beneficios. Esta cuestión limita el uso de clasificadores estándares que no están basados en la temporalidad (por ejemplo SVM, naive Bayes o Random Forest), ya que si el orden de las series es crucial, utilizar estos clasificadores que no lo tienen en cuenta no tendría sentido. Sin embargo, la hipótesis de este estudio es que no hay que asumir siempre que el orden es discriminatorio y hay casos en los que los clasificadores estándares pueden funcionar igual o mejor que el tan valorado 1-NN.

Para evaluar esta hipótesis se han tomado como métodos de referencia el 1-NN-EUC (que no tiene en cuenta el orden de las observaciones) y el 1-NN-DTW (que sí tiene en cuenta el orden de las observaciones). De esta manera, en la experimentación se observa que en los casos donde el 1-NN-EUC obtiene mejores resultados que el 1-NN-DTW, los clasificadores estándares pueden mejorar la precisión de clasificación (en el 80% de los casos), mientras que en los otros casos los clasificadores estándares no consiguen mejorarlo (solo en el 24% de los casos lo hacen).

Este resultado era esperado ya que si en una base de datos el 1-NN-EUC obtiene mayor precisión que el 1-NN-DTW se puede suponer que el orden de las observaciones no es un factor determinante para la clasificación. En ese caso, los clasificadores estándares, que son más complejos y robustos que el 1-NN, y no tienen en cuenta el orden, deberían funcionar bien. Por otro lado, si en una base de datos el 1-NN-DTW obtiene mayor precisión que el 1-NN-EUC se puede interpretar que el orden de las observaciones sí es un factor determinante para la clasificación, y tiene sentido que los clasificadores estándares, que no tienen en cuenta el orden de las observaciones, rindan peor. De hecho, los test estadísticos realizados a los resultados confirman las hipótesis. En otras palabras, cuando el orden de las observaciones no es un factor discriminatorio, los clasificadores estándares mejoran el resultado del 1-NN, mientras que cuando el orden es sí es relevante no lo consiguen.

En conclusión, no es que los clasificadores estándares, y métodos no específicos en general, sean inadecuados para la clasificación de series temporales, si no que dependiendo de las características discriminatorias de cada base de datos (en particular, la relevancia del orden de las observaciones para la clasificación), pueden llegar a ser beneficiosos. Por ello, sería interesante entender más en profundidad cuáles son las características temporales discriminatorias de las series, seguir estudiando la importancia del orden para la clasificación y buscar otras maneras de evaluar la influencia de este factor. Además, sería también enriquecedor seguir explorando el uso de clasificadores estándares, o métodos no específicos, para series temporales, ya que son mucho más robustos que el



1-NN y dependiendo de la distancia empleada en el 1-NN, mucho más rápidos.

AGRADECIMIENTOS

Esta investigación esta subvencionada por el Gobierno Vasco a través del programa BERC 2018-2021, así como del Ministerio de Economía y Competitividad (MINECO) mediante la acreditación de excelencia BCAM Severo Ochoa SEV-2013-0323. También del proyecto TIN2017-82626-R, financiado por AEI/FEDER (UE) y acrónimo “GECECPAST”. Además, gracias al Programa de Grupos de Investigación 2013-2018 (IT-609-13) (Gobierno Vasco) y TIN2016-78365-R (MINECO). A. Abanda está subvencionada con la beca BES-2016-076890.

REFERENCIAS

- [1] E. Keogh and S. Kasetty, “On the need for time series data mining benchmarks,” *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 102, 2002. [Online]. Available: <http://dl.acm.org/citation.cfm?id=775047.775062>
- [2] P. Esling and C. Agon, “Time-series data mining,” *ACM Computing Surveys*, vol. 45, no. 1, pp. 1–34, 2012. [Online]. Available: <http://dl.acm.org/citation.cfm?doi=2379776.2379788>
- [3] A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh, “The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances,” *Data Mining and Knowledge Discovery*, vol. 31, no. 3, pp. 606–660, 2017.
- [4] T. C. Fu, “A review on time series data mining,” *Engineering Applications of Artificial Intelligence*, vol. 24, no. 1, pp. 164–181, 2011. [Online]. Available: <http://dx.doi.org/10.1016/j.engappai.2010.09.007>
- [5] C. Faloutsos, M. Ranganathan, and Y. Manolopoulos, “Fast subsequence matching in time-series databases,” *ACM SIGMOD International Conference on Management of Data*, pp. 419–429, 1994.
- [6] M. Cuturi and J. Vert, “A kernel for time series based on global alignments,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 1, pp. 413–416, 2007. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4217433
- [7] R. Povinelli, “Time series classification using Gaussian mixture models of reconstructed phase spaces,” *Knowledge and Data ...*, vol. 16, no. 6, pp. 779–783, 2004. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1294898
- [8] D. Eads, D. Hill, S. David, S. Perkins, J. Ma, R. Porter, and J. Theiler, “Genetic Algorithms and Support Vector Machines for Time Series Classification,” *Applications and Science of Neural Networks, Fuzzy Systems, and Evolutionary Computation*, vol. 4787, 2002.
- [9] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and E. Keogh, “Experimental comparison of representation methods and distance measures for time series data,” *Data Mining and Knowledge Discovery*, vol. 26, no. 2, pp. 275–309, 2013.
- [10] J. Lines and A. Bagnall, “Time series classification with ensembles of elastic distance measures,” *Data Mining and Knowledge Discovery*, vol. 29, no. 3, pp. 565–592, 2015.
- [11] A. Bagnall and Gareth Janacek, “A run length transformation for discriminating between auto regressive time series,” *Journal of Classification*, vol. 31, no. October, pp. 274–295, 2014.
- [12] M. Corduas and D. Piccolo, “Time series clustering and classification by the autoregressive metric,” *Computational Statistics and Data Analysis*, vol. 52, no. 4, pp. 1860–1872, 2008.
- [13] P. Smyth, “Clustering sequences with hidden Markov models,” *Advances in Neural Information Processing Systems*, vol. 9, pp. 648–654, 1997. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.44.3648>
- [14] A. Abanda, U. Mori, and J. A. Lozano, “A review on distance based time series classification,” <https://arxiv.org/abs/1806.04509>, pp. 1–28, 2018. [Online]. Available: <http://arxiv.org/abs/1806.04509>
- [15] Z. Xing, J. Pei, and E. Keogh, “A brief survey on sequence classification,” *ACM SIGKDD Explorations Newsletter*, vol. 12, no. 1, p. 40, 2010. [Online]. Available: <http://portal.acm.org/citation.cfm?doi=1882471.1882478>
- [16] Y. Chen, B. Hu, E. Keogh, and G. E. Batista, “DTW-D: Time Series Semi-Supervised Learning from a Single Example,” *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 383, 2013. [Online]. Available: <http://dl.acm.org/citation.cfm?doi=2487575.2487633>
- [17] A. Bagnall, J. Lines, J. Hills, and A. Bostrom, “Time-series classification with COTE: The collective of transformation-based ensembles,” *Proceedings of 32nd ICDE International Conference on Data Engineering*, vol. 27, no. 9, pp. 1548–1549, 2016.
- [18] I. Popivanov and R. J. Miller, “Similarity Search Over Time-Series Data Using Wavelets,” *Proceedings 18th International Conference on Data Engineering (ICDE)*, pp. 212–221, 2002.
- [19] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, “Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases,” *Knowledge and Information Systems*, vol. 3, no. 3, pp. 263–286, 2001. [Online]. Available: <http://link.springer.com/10.1007/PL00011669>
- [20] D. Berndt and J. Clifford, “Using dynamic time warping to find patterns in time series,” *Workshop on Knowledge Knowledge Discovery in Databases*, vol. 398, pp. 359–370, 1994. [Online]. Available: <http://www.aaai.org/Papers/Workshops/1994/WS-94-03/WS94-03-031.pdf>
- [21] L. Chen and R. Ng, “On The Marriage of Lp-norms and Edit Distance,” in *International conference on Very large data bases*, 2004, pp. 792–803.
- [22] H. Kaya and . Gündüz-Ö üdücü, “A distance based time series classification framework,” *Information Systems*, vol. 51, pp. 27–42, 2015.
- [23] H. Sakoe and S. Chiba, “Dynamic Programming Algorithm Optimization for Spoken Word Recognition,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.
- [24] X. Xi, E. Keogh, C. Shelton, L. Wei, and C. A. Ratanamahatana, “Fast time series classification using numerosity reduction,” *Proceedings of the 23rd ICML International Conference on Machine Learning*, pp. 1033–1040, 2006. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1143974>
- [25] Y. Chen, E. Keogh, B. Hu, N. Begum, A. Bagnall, A. Mueen, and G. E. Batista, “The UCR Time Series Classification Archive,” 2015. [Online]. Available: www.timeseriesclassification.com
- [26] B. Calvo and G. Santafe, “scmamp: Statistical comparison of multiple algorithms in multiple problems,” *The R Journal*, vol. Accepted for publication, 2015.