Hierarchical Agglomerative Clustering of Time-Warped Series

Marian Kotas[1(✉)], Jacek Leski[1], Tomasz Moroń[1], and Jader Giraldo Guzmán[2]

[1] Institute of Electronics, Silesian University of Technology, Gliwice, Poland
{Marian.Kotas,Jacek.Leski,Tomasz.Moron}@polsl.pl
[2] Department of Electrical and Electronic Engineering, Univ. Tecnológica de Bolívar, Parque Industrial y Tecnologico Carlos Velez Pombo Km 1 via Turbaco, Cartagena, Colombia
JGiraldo@utbvirtual.edu.co

Abstract. We have developed a procedure for hierarchical agglomerative clustering of time series data. To measure the dissimilarity between these data, we use classically the Euclidean distance or we apply the costs of the series nonlinear alignment (time warping). In the latter approach, we use the classical costs or the modified ones. The modification consists in matching short signal segments instead of single signal samples. The procedure is applied to a few datasets from the internet archive of time series. In this archive, the series of the same classes possess visual similarity but their time evolution is often different (the characteristic waves have different location within the individual signals). Therefore the use of the Euclidean distance as the dissimilarity measure gives poor results. After time warping, the nonlinearly aligned signals match each other better, and therefore the total cost of the alignment appears to be a much more effective measure. It results in higher values of the Purity index used to evaluate the results of clustering. In most cases, the proposed modification of the alignment costs definition leads to still higher values of the index.

Keywords: Hierarchical clustering · Single/complete linkage · DTW

1 Introduction

Clustering is the operation of unsupervised data classification, employing different measures of their similarity or dissimilarity. Usually the analyzed data are described by vectors of a fixed length whose entries correspond to the specific features of these data. Calculation of the Euclidean distance between them is one of the most often applied approaches to the assessment of the data dissimilarity. However, when clustering of time series is considered, we often have to compare vectors of different length, corresponding to signals of different duration. Calculation of the Euclidean distance between them is not always possible without any auxiliary operations. One of such operations is the linear time-alignment of the signals compared. This operation consists in shifting a selected segment

A. Gruca et al. (eds.), *Man-Machine Interactions 5*,
Advances in Intelligent Systems and Computing 659, DOI 10.1007/978-3-319-67792-7_21

of one signal along with the second signal until the best matching of the both signals is achieved. After this operation, we are again able to use the Euclidean distance between the aligned signal segments as the measure of their dissimilarity. With the development of the technique of nonlinear alignment, called as dynamic time-warping (DTW), new possibilities emerged. Not only comparison of different length signals appeared possible but also compensation of their different time evolution. And indeed, in the 70s of the previous century the technique was applied to time normalization in spoken words recognition applications [17]. In spite of its high computational cost, since its introduction in the early 70s [16] DTW has often been applied in numerous research fields. Among others, it was used for biomedical signals processing [5,8–10,12], and its applications to time-series clustering are described e.g. in [6,7,13–15].

The goal of this paper is to apply the classical and the modified measure of dissimilarity based on DTW to hierarchical agglomerative clustering of time-series and to study the factors that influence the clustering results. The rest of the paper is organized as follows. In Sect. 2, an outline of the algorithm for hierarchical agglomerative clustering is provided. In Sect. 3, the similarity measures based on DTW are defined. Numerical experiments are reported in Sect. 4 and concluded in Sect. 5.

2 Hierarchical Agglomerative Clustering

We can distinguish two approaches to hierarchical clustering: divisive or agglomerative. In the former one, we start with a large cluster containing all data, and we divide it, and then the ensuing smaller clusters, until all data points are separated. In the agglomerative approach an opposite strategy is applied. We start with all data points belonging to different clusters and subsequently we merge the closest of them, one after another until all data are conjunct. Results of both approaches can be presented using a hierarchical tree structure called dendrogram [4]. A dendrogram consists of nodes corresponding to the created clusters. Whereas the root node represents the whole datasets, the leaf nodes are associated with the individual data points, and the intermediate nodes correspond to the respective clusters formed. Two vertical dendrograms are presented in the experimental section in Fig. 3. The leaf nodes start at the bottom of the dendrograms. The nodes are connected by horizontal lines called edges. Their heights correspond to the distances between the merged nodes. The structure of a dendrogram depends on the applied linking strategy. Such strategy depends on the measure of a distance between a pair of clusters. In this study, we have applied two very basic linking strategies: the single and the complete linkage methods (the definition of a distance between clusters is called as linkage itself). In the single linkage algorithm a distance between two clusters Ω_i and Ω_j is defined as the distance between two nearest points representing both clusters [4]

$$d_{\min}(\Omega_i, \Omega_j) = \min_{\mathbf{x}\in\Omega_i,\, \mathbf{x}'\in\Omega_j} d(\mathbf{x}, \mathbf{x}'). \tag{1}$$

In the complete linkage algorithm a distance between two clusters is defined as the distance between two farthest points representing both clusters

$$d_{\max}(\Omega_i, \Omega_j) = \max_{\mathbf{x}\in\Omega_i,\ \mathbf{x}'\in\Omega_j} d(\mathbf{x}, \mathbf{x}'). \tag{2}$$

Although there are also other linking strategies [2,11], in this paper we confined our experiments to the above two. In the further part of the study, when Euclidean distance between data points is considered the single linkage will be denoted as d^{euc}_{min} and the complete linkage as d^{euc}_{max}.

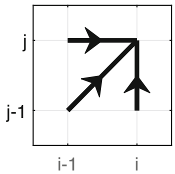

Fig. 1. Allowable step directions of a warping path

3 Dynamic Time Warping Based Linkages

Dynamic Time Warping (DTW) is a technique developed to solve the problem of matching two signals of possibly different length. DTW uses the technique of dynamic programming [1,17] to determine the best alignment (optimal warping path) between the signals considered. Having two signals $x(n), n = 1, 2, ..., N_x$ and $y(n), n = 1, 2, ..., N_y$ (where N_x can be different from N_y), we appropriately squeeze or stretch their temporal axes to obtain the minimal cost of their alignment. These costs can be exploited by clustering algorithms, trying to assign the signals to the proper clusters. The warping path consists of the ordered pairs of time indices: $\{(i_k, j_k)|k = 1, 2, \cdots, K\}$, which indicate the elements $y(i_k)$ and $x(j_k)$ of the time-warped series that are assigned with each other. Classically the cost of matching $y(i)$ and $x(j)$ is defined as

$$d_{i,j} = (y(i) - x(j))^2 \tag{3}$$

and the warping path minimizes the total cost of the alignment:

$$Q = \sum_{k=1}^{K} d_{i_k, j_k} \tag{4}$$

preserving the specified constraints, assuring that

- the alignment function aligns all points

$$i_1 = j_1 = 1 \text{ and } i_K = N_x, j_K = N_y \tag{5}$$

- the function is increasing monotonically

$$i_k \geq i_{k-1} \text{ and } j_k \geq j_{k-1} \tag{6}$$

- no point is omitted

$$i_{k+1} - i_k \leq 1, j_{k+1} - j_k \leq 1. \tag{7}$$

These constraints correspond to the so-called step directions of the warping path presented in Fig. 1.

Using the square root of the classical definition of the total alignment costs $\sqrt{Q}$, we obtain the definition of the distance between two time series d^{cdtw} which is equal to the Euclidean distance between the time-warped series. The corresponding single and complete linkages will be denoted as d^{cdtw}_{min} and d^{cdtw}_{max}, respectively (*cdtw* stands for classical DTW). In [10] the DTW technique was applied to ECG cycles or EEG evoked potentials alignment prior to their time averaging. The aim of this operation was to suppress noise and enhance the desired signals. It appeared that the operation was much more effective when the classical definition of alignment costs was replaced be the following one

$$d'_{i,j} = \left\| \mathbf{x}^{(i)} - \mathbf{y}^{(j)} \right\| \tag{8}$$

where $\|\cdot\|$ denotes the Euclidean norm, and vectors $\mathbf{x}^{(n)}$ and $\mathbf{y}^{(n)}$ are defined as

$$\mathbf{x}^{(n)} = [x(n-v), x(n-v+1), \ldots, x(n), \ldots, x(n+v)]^T \tag{9}$$

which means that they are composed of $2v+1$ successive signal samples (with the nth sample being the central one). With this modified definition of the alignment costs, using the ensuing total alignment costs as the definition of the distance between two time series, we obtain the single and the complete linkages, denoted as $d^{mdtw}_{min}$and d^{mdtw}_{max}, respectively (*mdtw* stands for modified DTW). **Remark** For the modified definition of the alignment costs (8), the following difficulty emerges. For $v \geq i$ or $v \geq j$ we cannot construct vectors defined by (9). Therefore we find $l = min(i-1, j-1)$ and we construct shorter vectors

$$\mathbf{x}^{(n)} = [x(n-l), x(n-l+1), \ldots, x(n), \ldots, x(n+v)]^T \tag{10}$$

Similarly we proceed on the right side ends of the time-warped series.

4 Numerical Experiments

We have applied the developed algorithms to cluster signals contained in five datasets from the UCR (University of California Riverside) time series classification archive [3]. We used the following datasets: Trace, Mallat, BeetleFly, OSULeaf and Plane. The first (presented in Fig. 2) and the second one are synthetic; the other three are one dimensional time series created on the basis of

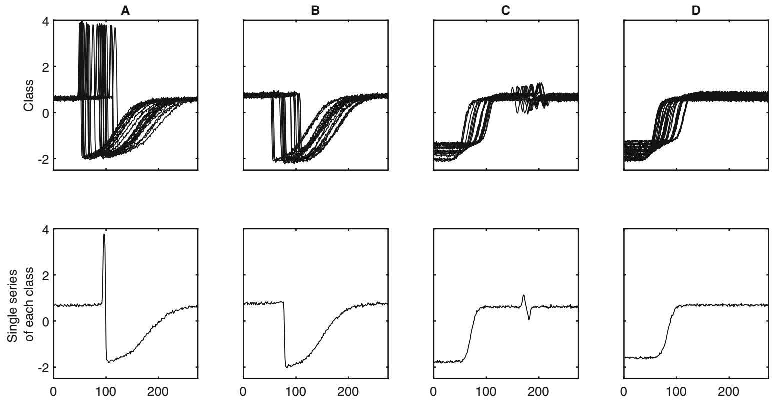

Fig. 2. The Trace dataset from the UCR archive. In upper subplots, we can see groups of signals of the existing four classes, contained in the learning part of the dataset. Below individual examples are presented.

two-dimensional pictures of beetles or flies, leaves, and planes, respectively. Each database contains both learning and testing part, prepared for experiments on time series classification. In our experiments, we used the signals contained in the learning parts of these datasets (each signal was normalized to be of zero mean and unit variance). Parameter v was established separately for each of the above datasets, so as the length $2v + 1$ of vectors (9) used to calculate d^{mdtw}_{min} and d^{mdtw}_{max} was equal to approximately one 10th of the signals length. After the signals clustering, the information on the classes they belong to was used to evaluate the results of clustering. To this end, the Purity index was calculated, showing if the results of clustering were similar to the true assignment of the signals to the classes specified. If each of the created clusters contains the data of the same class, the index reaches the value of 1, otherwise it is smaller (but always greater than 0).

First, to visualize the operation of hierarchical clustering, we chose 8 time series from the Trace dataset: two examples from each of the existing data classes. The selected signals were clustered using the d^{euc}_{min} and d^{cdtw}_{min} linkages. The obtained dendrograms are presented in Fig. 3. By cutting the dendrogram obtained using d^{cdtw}_{min} at the level of about 3, a correct data clustering is achieved. We can see that the height of the edges connecting the leaf nodes corresponding to the same class signals (e.g. number 1 and 2) are relatively low what shows that the costs of nonlinear alignment were small. The edges connecting larger clusters (each consisting of two elements) are much higher, what corresponds to minimal distances between signals of different classes (see definition (1) of the applied d^{cdtw}_{min} linkage). Using the d^{euc}_{min} linkage for the series of the same class, in most cases we obtained much higher edges (their relative height is considered). As a result, no correct clustering appeared possible for this, Euclidean distance based linkage. To visualize the reasons of such a failure, in Fig. 4 we have

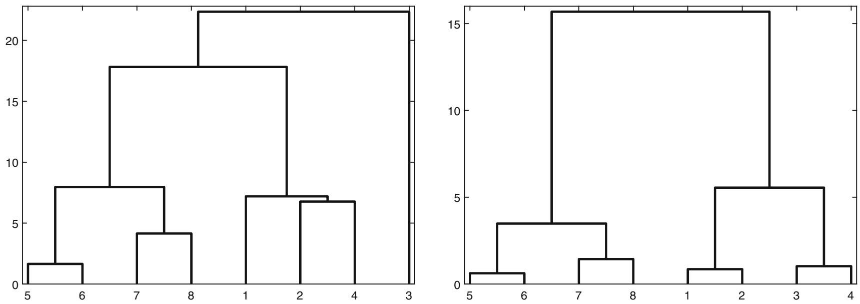

Fig. 3. Dendrograms obtained for the selected 8 signals from the Trace dataset using the Euclidean distance based linkage: d^{euc}_{min} (on the left) and the classical DTW costs based one: d^{cdtw}_{min} (on the right). The numbers of the signals of the selected subset, corresponding to the leaf nodes, are presented at the bottom.

presented a few signals corresponding to the leaf nodes of the dendrograms. Dealing with signals of limited length, we can assume that they are the selected segments of longer traces. Therefore the operation performed to determine the Euclidean distance d^{euc} can be regarded as a kind of linear time alignment. Results of such an alignment are presented in subplots A and B, whereas the nonlinearly aligned signals are shown in C and D. The differences between the linearly aligned signals can be watched in E and F, and the differences between the time-warped ones, in G and H. The pair of signals presented in the left subplots belongs to the same class, whereas the pair on the right to different classes. We can notice that even for signals that are visually very similar (A) the difference between them can be relatively large (E) because of the different location of the characteristic waves occurring within these signals. After the nonlinear alignment, the signals match to each other very precisely (G). As a result, the difference between them is much reduced, and the calculated d^{cdtw} distance small. Even when the signals are of different classes, time warping often leads to the distance reduction (if compared to the Euclidean one); however, for the same class signals the operation is usually much more effective, and for such signals smaller distances are more likely. In Fig. 5, we have presented the dendrograms horizontally, showing additionally the time series selected when using the d^{euc}_{min} and d^{cdtw}_{min} linkages to calculate distances among clusters. In the upper subplot of Fig. 5, where the results of correct clustering are presented, we can see that to compose clusters containing two elements, the algorithm was able to select the series of the same classes. In the uppermost group, however, the two series (number 3 and 4), although visually similar, are distant, as far as the Euclidean distance is regarded, which results from different positions of the characteristic negative deflections within the two series considered. Therefore in the dendrogram created using the d^{euc}_{min} linkage, they are assigned to completely different clusters.

Results of the experiment performed on the whole learning subsets of the datasets from the UCR archive are gathered in Table 1. This table contains

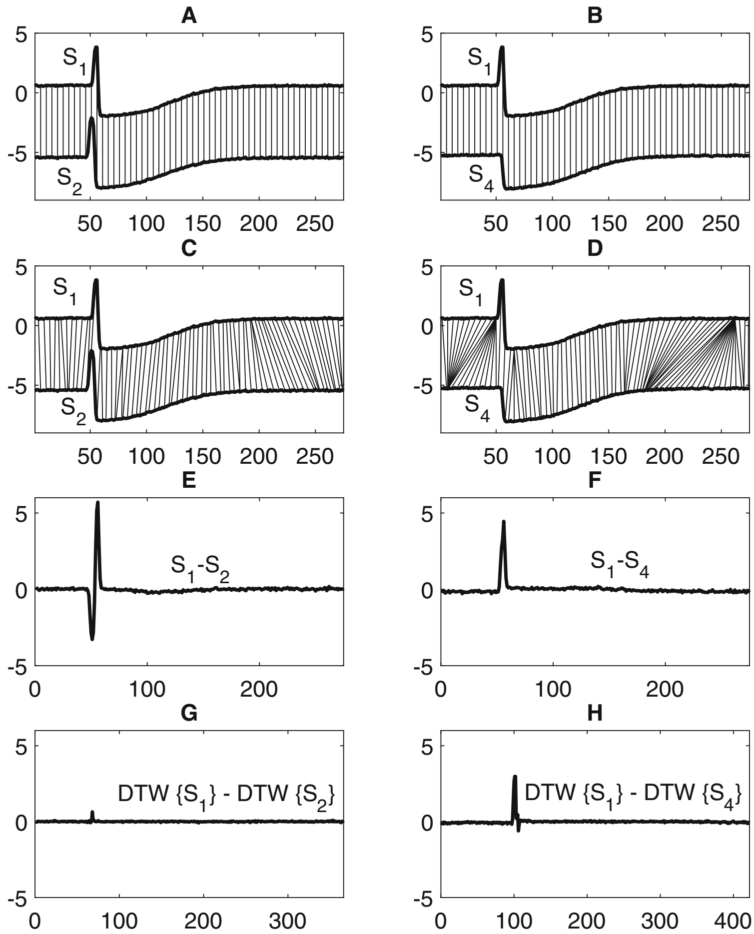

Fig. 4. Results of linear (A,B) and nonlinear (C,D) alignment of the signals of the same (on the left) and different (on the right) classes. Differences E,F and G,H correspond to the calculations of the Euclidean distance and the DTW based one, respectively. The numbers of the signals: 1, 2 and 4 correspond to the numbers of the leaf nodes in Fig. 3. In A, B, C and D the lower signals are vertically shifted for better presentation.

the Purity index obtained for algorithms using the DTW based distances or the Euclidean one, and the single or the complete linkage. As we could have expected, for the Trace dataset the algorithms using the DTW based distances are much more effective than those using the Euclidean one. This observation is valid for all datasets investigated. However, we can also notice that it is justified to use the complete linkage (d_{max}) instead of the single one (d_{min}). In three cases, using the modified definition of the alignment costs and the complete linkage, we have obtained significant improvement of the Purity index. For the synthetic datasets, however, no improvement was observed; on the contrary, for the Mallat dataset even deterioration of the results was caused.

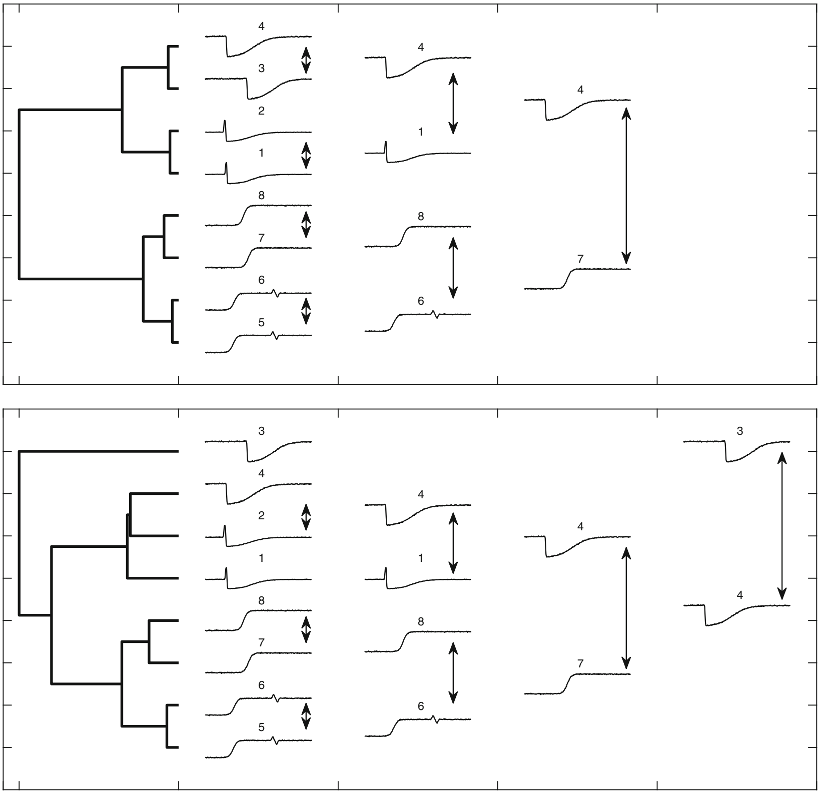

Fig. 5. Dendrograms from Fig. 3 presented horizontally, with the time series on the basis of which the between clusters distances d_{min}^{cdtw} (up) and d_{min}^{euc} (down) were computed. The signals numbers correspond to the leaf nodes in Fig. 3.

Table 1. The purity index obtained for the selected datasets from the UCR archive while using different clustering algorithms. For each dataset, the best results achieved are bolded.

	Linkage					
Dataset NAME	d_{min}^{euc}	d_{max}^{euc}	d_{min}^{cdtw}	d_{max}^{cdtw}	d_{min}^{mdtw}	d_{max}^{mdtw}
Trace	0.57	0.57	**0.78**	**0.78**	**0.78**	0.57
Mallat	0.67	0.75	0.64	**0.95**	0.67	0.76
OSULeaf	0.29	0.40	0.29	0.41	0.31	**0.46**
BeetleFly	0.55	0.60	0.55	0.80	0.6	**0.8**
Plane	0.70	0.70	0.88	0.98	0.88	**1**

5 Conclusions

We have developed a procedure for time series clustering, using different definitions of distances between pairs of signals and between pairs of clusters. To measure the dissimilarity between signals, we have applied the Euclidean distance and two definitions based on the costs of the series nonlinear alignment. To measure the distance between data clusters, we have used the single linkage and the complete one. For the clustered datasets, an easy conclusion can be inferred that it is favorable to use dynamic time warping to assess dissimilarity between signals, instead of the Euclidean distance. It has also appeared advantageous to apply the complete linkage instead of the single one. However, the major contribution of this study is the application of the new definition of the alignment costs. With this definition, we have achieved the highest values of the Purity index for most datasets studied; however, for one dataset the results were poorer. Therefore additional studies are necessary to find the reasons of such antithetical results of the modification proposed.

Acknowledgements. This work was partially supported by the Ministry of Science and Higher Education funding for statutory activities (BK-220/RAu-3/2016) and the Ministry of Science and Higher Education funding for statutory activities of young researchers (BKM-508/RAu-3/2016). The work was performed using the infrastructure supported by POIG.02.03.01-24-099/13 grant: GeCONiI—Upper Silesian Center for Computational Science and Engineering.

References

1. Bellman, R.E., Dreyfus, S.E.: Applied Dynamic Programming. Princeton University Press, Princeton (2015)
2. Bien, J., Tibshirani, R.: Hierarchical clustering with prototypes via minimax linkage. J. Am. Stat. Assoc. **106**(495), 1075–1084 (2011)
3. Chen, Y., Keogh, E., Hu, B., Begum, N., Bagnall, A., Mueen, A., Batista, G.: The UCR Time Series Classification Archive (2015). http://www.cs.ucr.edu/~eamonn/time_series_data/
4. Everitt, B.S., Landau, S., Leese, M., Stahl, D.: Hierarchical Clustering, pp. 71–110. Wiley, Hoboken (2011)
5. Gupta, L., Molfese, D.L., Tammana, R., Simos, P.G.: Nonlinear alignment and averaging for estimating the evoked potential. IEEE Trans. Biomed. Eng. **43**(4), 348–356 (1996)
6. Keogh, E.: Exact indexing of dynamic time warping. In: VLDB 2002, Hong Kong, China, pp. 406–417 (2002)
7. Keogh, E.J., Pazzani, M.J.: Scaling up dynamic time warping for datamining applications. In: KDD 2000, Boston, MA, US, pp. 285–289 (2000)
8. Kotas, M.: Projective filtering of time warped ECG beats. Comput. Biol. Med. **38**(1), 127–137 (2008)
9. Kotas, M.: Robust projective filtering of time-warped ECG beats. Comput. Methods Programs Biomed. **92**(2), 161–172 (2008)
10. Kotas, M., Pander, T., Leski, J.M.: Averaging of nonlinearly aligned signal cycles for noise suppression. Biomed. Sig. Process. Control **21**, 157–168 (2015)

11. Leski, J.M., Kotas, M.: Hierarchical clustering with planar segments as prototypes. Pattern Recogn. Lett. **54**, 1–10 (2015)
12. Moroń, T.: Averaging of time-warped ECG signals for QT interval measurement. In: Piętka, E., Badura, P., Kawa, J., Wieclawek, W. (eds.) Information Technologies in Medicine. AISC, vol. 471, pp. 291–302. Springer International Publishing, Switzerland (2016)
13. Niennattrakul, V., Ratanamahatana, C.A.: On clustering multimedia time series data using k-means and dynamic time warping. In: MUE 2007, Seoul, South Korea, pp. 733–738 (2007)
14. Petitjean, F., Ketterlin, A., Gançarski, P.: A global averaging method for dynamic time warping, with applications to clustering. Pattern Recogn. **44**(3), 678–693 (2011)
15. Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J., Keogh, E.: Searching and mining trillions of time series subsequences under dynamic time warping. In: KDD 2012, Beijing, China, pp. 262–270 (2012)
16. Sakoe, H., Chiba, S.: A similarity evaluation of speech patterns by dynamic programming. In: Nat. Meeting of Institute of Electronic Communications Engineers of Japan, p. 136 (1970)
17. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acoust. Speech Sig. Process. **26**(1), 43–49 (1978)