

Searching Time Series Based On Pattern Extraction Using Dynamic Time Warping

Tomáš Kocyan¹, Jan Martinovič¹, Pavla Dráždilová², and Kateřina Slaninová²

¹ VŠB - Technical University of Ostrava,
IT4Innovations,

17. listopadu 15/2172, 708 33 Ostrava, Czech Republic
{tomas.kocyan,jan.martinovic}@vsb.cz

² VŠB - Technical University of Ostrava,
Department of Computer Science,

17. listopadu 15/2172, 708 33 Ostrava, Czech Republic
{pavla.drazdilova,katerina.slaninova}@vsb.cz

Abstract. Many types of data collections processed by time series analysis often contain repeating similar episodes (patterns). If these patterns are recognized, then they may be used for instance in data compression, for prediction or for indexing large collections. Extraction of these patterns from data collections with components generated in equidistant time and in finite number of levels is now a trivial task. The problem arises for data collections that are a subject to different types of distortions in all axes. In this type of collections, the found similar episodes do not have to be exactly the same; they can differ in time, shape or amplitude. In these cases, it is necessary to pick the suitable one from each group of similar episodes and to declare it as a representative member of the whole group. This paper discusses the possibilities of using the Dynamic Time Warping (DTW) method for deriving the representative member of a group of similar episodes that are subjects to the previously mentioned distortions. The paper is also focused on providing a suitable mechanism for more effective searching of distorted time series.

Keywords: Dynamic Time Warping, Time Series, Pattern Mining

1 Introduction

Time series analysis covers methods for analysis of time series data with a focus on extraction of various types of information like statistics and other characteristics of the data. During time series processing, it is common that a time series is divided into a large amount of smaller parts named episodes, which are interconnected or partially overlapped [6] and which are important for further processing. For example, interconnected outputs of hydrological models, data collections from traffic monitoring of selected stretches, or long time series divided by segmentation algorithm like Voting experts [7] can be mentioned. Obtained episodes may be processed by a suitable clustering algorithm and divided into the clusters [3, 5].

Various approaches in spheres like recommended systems, decision support systems or tasks based on Case Base Reasoning (CBR) are focused on finding similar sequences (time series episodes) to a sequence entered on the input. In such cases, a suitable cluster of similar sequences is found, which represents the input sequence. Much faster searching is allowed due to finding in set of the cluster representatives which were selected in indexing phase. Thereafter, it is possible to search in depth in a selected cluster or a set of clusters, which are similar to a found episode from the input.

Since each obtained cluster contains a concrete amount of similar episodes, it is suitable to select an appropriate representative, which would describe the whole cluster. Given selected representative is named pattern. Research area aimed to finding patterns, pattern mining, has been studied in several fields. Pattern mining, or pattern recognition, is a scientific discipline focused on object classification into categories or classes [10, 4].

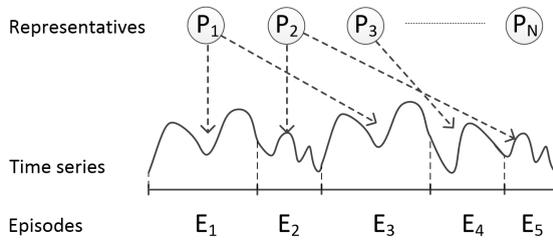


Fig. 1. Collection of Representatives Pointing to Locations in Time Series

Finding the representative of a cluster is defined as finding such set of representative patterns P , which describes episodes E inside these clusters by the most appropriate way. Obtained representatives may be used for the creation of an index file, in which each representative contains a set of pointers to episodes from the base collection (see Figure 1).

Two basic ways for finding representatives are generally known. The first approach is based on selecting one episode, which is the most accurate for a given cluster. The second approach is based on the creation of a representative using the combination of episodes in the cluster. Euclidean distance and other common methods for measuring the similarity between the episodes can be used only while working with the episodes of the identical length. In cases where we have episodes of different lengths, we need a specific algorithm which respects this requirement or an algorithm which is immune to sequence distortions. In the paper, it is described the comparison of the both approaches, and the introduction of an approach which combines the both ways for finding representatives using DTW method is presented (for more details, see Section 2).

The organisation of the paper is following: Dynamic time warping method (DTW) and the utilization of DTW for finding cluster representatives is described in Section 2 and in Section 3. Afterwards, in Section 4, a practical

demonstration of proposed approach is presented. The paper is concluded by Section 5, in which obtained results of suggested approach are discussed and the future work is outlined.

2 Dynamic Time Warping

Recently, finding a signal similar to a signal generated by computers, which consists of accurate time cycles and which achieves a determined finite number of value levels, is a trivial problem. A main attention is focused more likely on the optimisation of searching speed. A non-trivial task occurs while comparing or searching the signals, which are not strictly defined and which have various distortions in time and amplitude. As a typical example, we can mention measurement of functionality of human body (EKG, EEG) or the elements (precipitation, flow rates in riverbeds), in which does not exist an accurate timing for signal generation. Therefore, comparison of such episodes is significantly difficult, and almost excluded while using standard functions for similarity (distance) computation. Examples of such signals are presented in Figure 2a.

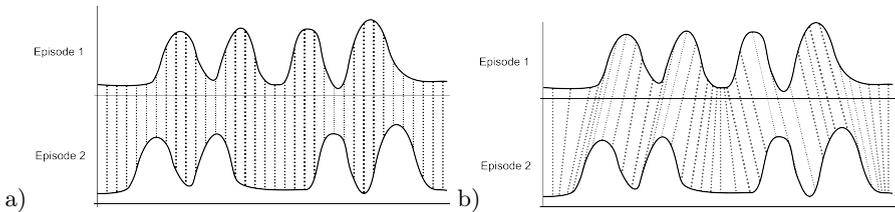


Fig. 2. Standard and DTW Mapping of Episodes

A problem of standard functions for similarity (distance) computation consists in sequential comparison of opposite elements in both episodes (comparison of elements with the identical indexes). Dynamic time warping (DTW) is a technique for finding the optimal matching of two warped episodes using pre-defined rules [1, 9]. Essentially, it is a non-linear mapping of particular elements to match them in the most appropriate way.

The output of such DTW mapping of episodes from Figure 2a can be seen in Figure 2b. This approach was used for example for comparison of two voice patterns during an automatic recognition of voice commands [8].

The main goal of DTW method is a comparison of two time dependent episodes X and Y , where $X = (x_1, x_2, \dots, x_N)$ is of length $N \in \mathbb{N}$ and $Y = (y_1, y_2, \dots, y_M)$ is of length $M \in \mathbb{N}$, and to find an optimal mapping of their elements. A detailed description of DTW including particular steps of the algorithm is presented in [1].

3 Using DTW for Finding Cluster Representative

In cases, where it is necessary to gain the most suitable representative of the set of similar episodes, we need to find an algorithm appropriate to a given domain. Sometimes it is possible to use simple average of episodes X and Y , which means that for a representative R is valid, that:

$$R_i = \frac{X_i + Y_i}{2}, \forall i = 1, \dots, P, \text{ where } P = |X| = |Y|. \quad (1)$$

However, this approach is not sufficient in cases, where we have data with distortion. Examples of such episodes are presented in Figure 3a and 3b. If only we used simple average presented in Equation 1, we would achieve an episode showed in Figure 3c. As we can see, this episode absolutely is not a representative and all the information about the episode course is loosed.

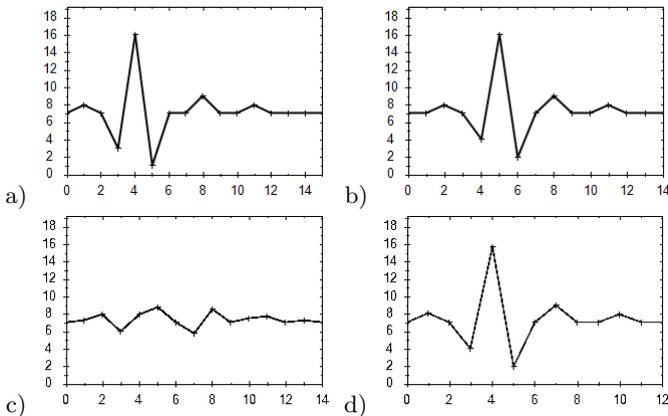


Fig. 3. Similar Episodes X and Y , their Average and Representative Found by DTW

As we can see from Figure 3, it is necessary to find a more appropriate algorithm for domains which yield to distortion. The algorithm should be immune to such distortions. This paper is focused on using DTW for finding a representative of set of similar, but distorted episodes.

3.1 Finding Representative for Episode Couples

The approach for finding a representative of two episodes X and Y by finding the optimal mapping of two episodes using DTW was described in Section 2. In this method, the most important is obtained warped path $p^* = (p_1, \dots, p_L)$, which allows to find a representative. The approach for finding such representative is described in Algorithm 3.1. The output of presented algorithm applied on episodes in Figure 3 is presented in Figure 3d.

Algorithm 3.1 Searching for Representative from Pair of Episodes

Input: Episodes X and Y

Output: Representative episode R

1. Compute $DTW(X, Y)$ for episodes X and Y ; obtain warping path p^* .
2. **Initialization:**
 - R is a representative episode for episodes X and Y .
 - $q = 1$ gives a position in R , $l = 2$ gives a position in warping path p^* .
 - Value in the first position in R is determined as average of values in the first positions of episodes X and Y , e.g. $r_1 = \frac{x_1 + y_1}{2}$.
3. **if** $l \leq L$ **then** for couple of the subsequent points of warping path p_l and p_{l-1} perform:
 - if** $(p_l - p_{l-1}) = (1, 1)$ **then**
 - $q = q + 1$;
 - A new item $r_q = \frac{x_{n_l} + y_{m_l}}{2}$ is inserted into episode R ;
 - else if** $(p_l - p_{l-1}) = (0, 1)$ or $(p_l - p_{l-1}) = (1, 0)$ **then**
 - No item is inserted into representative episodes R ;
 - end if**
 - $l = l + 1$
 - Repeat Step 3.
- end if**
4. Output of the algorithm is representative episode R of length q .

Algorithm 3.1 finds a representative common for two episodes, where both episodes have the same importance. It finds such episode, which is the most similar to the both two episodes. If it is necessary, a one of the episodes may be preferred by adding a weight $w \in (0; \infty)$ and by adjusting a computation of element r_1 and r_q by Equation 2:

$$r_1 = \frac{(x_1 * w) + y_1 * (w - 1)}{w + 1} \text{ and } r_q = \frac{(x_{n_l} * w) + y_{m_l} * (w - 1)}{w + 1}. \quad (2)$$

The impact of adding a weight on achieved representative R for episodes X and Y is following:

- $w = 1$: episodes are equal
- $w \in (1, \infty)$: episode X is preferred
- $w \in (0, 1)$: episode Y is preferred

3.2 Finding Representative for Set of Episodes

Algorithm 3.1 can be applied only on two episodes. However, this is often insufficient in common practice; we need to find a representative for the whole set of episodes in most cases. Given a collection C with generally N episodes,

$C = \{e_1, e_2, \dots, e_N\}$. The question is, how the presented approach applies on generally N episodes.

A first solution is based on an approach, in which is a representative found step by step by finding particular representatives for episode couples. More precisely, the first step consists of finding representative R_{1-2} for the first two episodes e_1 and e_2 . Then, representative R_{1-2-3} is found for a new obtained episode R_{1-2} and for episode e_3 . Then, such approach is used for the rest of episodes in the cluster.

However, our experiments showed that this approach is not as much suitable as it could be. It is strongly dependent on the order of particular episodes in collection. The solution is to find an approach that would be immune to the order of elements in an episode. Our proposed approach which solves this problem is presented in Algorithm 3.2.

The presented approach is not restricted only to using DTW as a method for the expression of episode similarity. Of course, DTW could be replaced by any other indicator, for example Euclidean distance or statistical indicators for time series (MAE, MPE, RMSE, etc.). In such cases, it is necessary to adapt steps 2 and 4 of Algorithm 3.2, where instead of finding a representative for the episodes couple by DTW is necessary to use (weighted) average of two compounded episodes. Section 4 describes both two approaches with a visual comparison of the impact to a found representative.

4 Experiments

In this section, a method for determination of similarity between two episodes is presented. Furthermore, the proposed method is compared with other methods. The achieved outputs are visualized with the following structure. The first row of the Figures 4 - 8 consists of episodes, which were used as the input to the algorithm, the second row consists of outputs for the different approaches.

The first output was average of episodes, defined in Equation 1. The second output was from the proposed approach described in Section 3.2. Both outputs are followed by the results using Mean Absolute Error (MAE), and finally as reference, Euclidean distance.

Meaning and usage of DTW method is closer to a human judgement and perception of similarity than a machine definition of physical distance. It is impossible to use a numerical evaluation for the following outputs. The experiments presented in this section were focused on finding such representative, which would describe the characteristics and the important parts of particular episodes.

The first input dataset was a set of similar signals (see Figure 4), which shapes resembled ECG records (described for example in [2]). The signal ended with tiny swings. As we can see from the second row of the episodes in Figure 4, average of values from both episodes absolutely degraded signal information; the shift of signal peaks and drops was smoothed nearly to one level. Also usage of MAE method and Euclidean distance did not provide sufficient results, which

Algorithm 3.2 Searching for Representative from Set of Episodes

Input: Collection C of N episodes

Output: Representative episode R

1. **Initialization:**

- N is count of input episodes.
- u is level of collection; $u = 1$.
- C^1 is the first level of collection; $C^1 = C$.
- M is count of processed episodes in level u ; $M = N - u + 1$.

2. Create from collection C^u , which consists of episodes $\{e_1^u, e_2^u, \dots, e_M^u\}$, distance matrix $D^u \in \mathbb{R}^{(M \times M)}$, where particular matrix elements are defined as $d_{ij}^u = DTW(e_i^u, e_j^u)$, i.e. matrix elements are created by values of reciprocal mapping of particular episodes.
3. Calculate sum for each row r_i^u in matrix D^u and select a row with the lowest sum value. Find row r_{min}^u , where

$$\sum_{j=1}^M d_{min,j}^u = \min_{i=1, \dots, M} (\sum_{j=1}^M d_{ij}^u)$$

The found row refers to the episode, which is selected as the most similar to the others in the current collection, and which could be declared as representative R^u of the collection for u -th level.

4. Remove representative R^u from the current collection and create $(N - u)$ new episodes by application of method for searching representative from couple (R^u, e_i^u) , described in Section 3.1. This algorithm can be modified by adding weight (preference) to one of the episodes, which can prefer (or discriminate) the importance of the representative R^u .

if $M > 2$ **then**

$u = u + 1$;

$M = M - 1$;

Repeat from Step 2 for remaining $(N - u)$ episodes;

else if $M = 2$ **then**

Select a representative from the two episodes as a representative of the whole original set of episodes C ;

end if

did not differ from average outputs much. On the other way, usage of DTW method for finding representative fully depicted a character of the signal and brought the most accurate results.

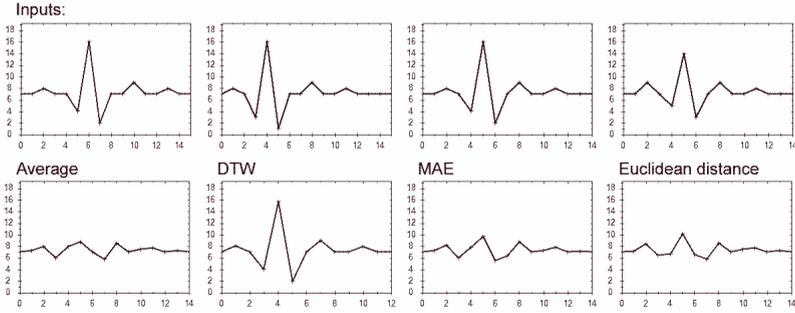


Fig. 4. Experiment with Simplified ECG Signals

The next episode quartet contained signals with the three peaks mutually shifted in time, while each of them had a variable duration (see Figure 5). It was supposed that the representative would have a curve with the three evident peaks. It is obvious from the results, that even though MAE and Euclidean distance worked much better, the loss of information was still noticeable.

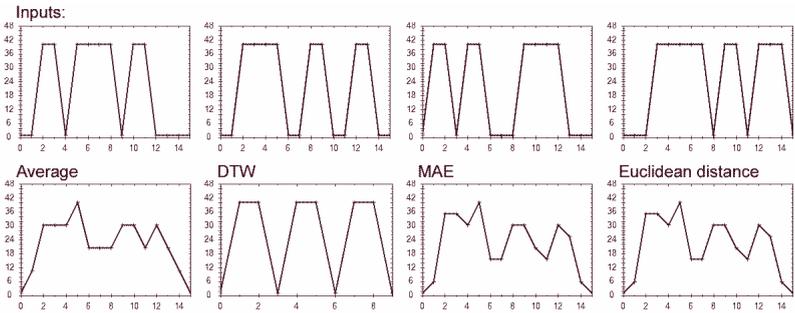


Fig. 5. Experiment with Three Distorted Peaks

The last input dataset represented the situation, in which the signal consisted of two waves - one in a positive and one in a negative part (see Figure 6). These waves were deformed in time, while they were spread or shrunk in X axis. Although the other methods achieved seemingly the best results, the distortion was evident again. The output representative did not contained as high amplitudes as the input waves, did not have smoothed waves and did not detect the constant segments, which were distorted.

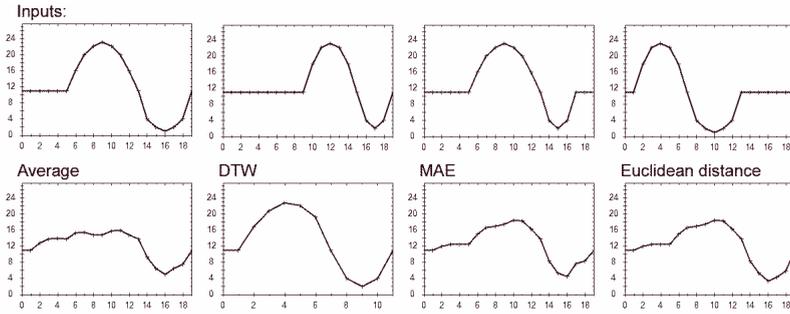


Fig. 6. Experiment with Waves

The most important advantage of the proposed solution is the fact that the Algorithm 3.2 in combination with DTW is able to process even episodes with different lengths. This is very difficult while using other methods, and in some cases even impossible. In these cases it is necessary to shrink the episodes into the identical length, which of course cause the loss of information. Using DTW, we are able to process such episodes with different lengths without any loss of information. In Figures 7 and 8 are presented outputs from proposed algorithm applied on episodes with different lengths.

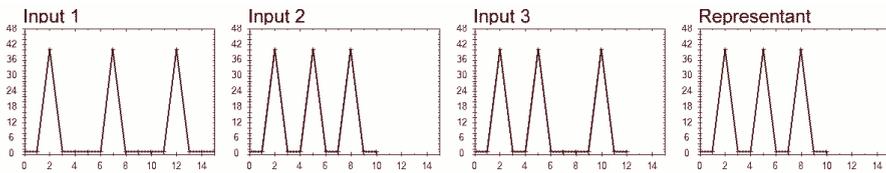


Fig. 7. Set 1 of Episodes with Variable Length

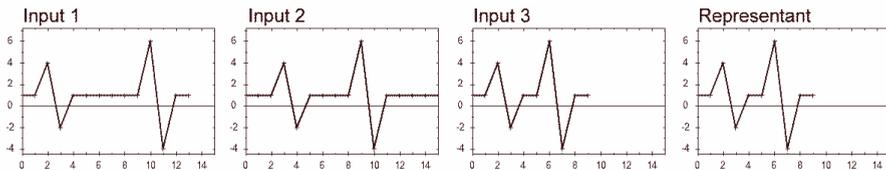


Fig. 8. Set 2 of Episodes with Variable Length

5 Conclusion and Future Work

The real application of proposed algorithm “Searching for Representative from Set of Episodes” described in Section 3.2 showed that it is able to find a representative not only from the set of typical episodes, but also from their distorted variants. The tested input datasets consisted of signals with changed amplitudes and were distorted by time shifting. The proposed solution was compared with conventional methods, in which much worse success was obvious.

Further work will be concentrated on creation of index file, which structure was defined in Section 1, and which visual representation was presented in Figure 1. The aim is to create a sufficiently robust mechanism, which will be able to find all the similar episodes to the selected pattern in data collection during the shortest time. Furthermore, these found episodes will be used for a prediction using the Case-Based Reasoning method. This method requires a suitable mechanism that is able to extract the most similar patterns from the input.

Acknowledgement

This work was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and by the SGS, VŠB – Technical University of Ostrava, Czech Republic, under the grant No. SP2013/167 Analysis of Users’ Behaviour in Complex Networks.

References

1. Dynamic time warping. In *Information Retrieval for Music and Motion*, pages 69–84. Springer Berlin Heidelberg, Jan. 2007.
2. G. D. Clifford, F. Azuaje, P. McSharry, et al. *Advanced methods and tools for ECG data analysis*. Artech House, 2006.
3. G. Gan, C. Ma, and J. Wu. *Data Clustering: Theory, Algorithms, and Applications*. ASA-SIAM Series on Statistics and Applied Probability. SIAM, MAY 2007.
4. D. J. Hand, P. Smyth, and H. Mannila. *Principles of Data Mining*. MIT Press, Cambridge, MA, USA, 2001.
5. A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Computing Surveys*, 31(3):264–323, 1999.
6. E. Keogh, S. Chu, D. Hart, and M. Pazzani. Segmenting time series: A survey and novel approach. *Work*, 57:121, 2004.
7. T. Kocyan, J. Martinovic, M. Podhoranyi, and I. Vondrak. Unsupervised algorithm for retrieving characteristic patterns from time-warped data collections. In *Proceedings of the MAS 2012, The 11th International Conference on Modeling and Applied Simulation*, 2012.
8. L. R. Rabiner and B. B.-H. Juang. *Fundamentals of Speech Recognition*. Prentice Hall, 1993.
9. P. Senin. Dynamic time warping algorithm review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, pages 1–23, 2008.
10. S. Theodoridis. *Pattern Recognition*. Elsevier, 3 edition, 2006.