RECOGNIZING CHARACTERISTIC PATTERNS IN DISTORTED DATA COLLECTIONS

Tomáš Kocyan [(a)], Jan Martinovič[(b)], Pavla Dráždilová[(c)], Kateřina Slaninová[(d)]

[(a,b)]VŠB - Technical University of Ostrava,
IT4Innovations,
17. listopadu 15/2172, 708 33 Ostrava, Czech Republic
[(c,d)]VŠB - Technical University of Ostrava,
Department of Computer Science,
17. listopadu 15/2172, 708 33 Ostrava, Czech Republic

[(a)]tomas.kocyan@vsb.cz, [(b)] jan.martinovic@vsb.cz, [(c)]pavla.drazdilova@vsb.cz, [(d)]katerina.slaninova@vsb.cz

ABSTRACT
Many models and artificial intelligence methods work with the inputs in the form of time series. Generally, success of many of them strongly depends on ability to successfully manage input data, which often contains repeating similar episodes (patterns). If these patterns are recognized, they can be used for instance for indexing, prediction or compression. These operations can also be very useful for improving the already existing model performance and accuracy. Our effort is to provide a robust mechanism for retrieving these characteristic patterns from the collections that are subject of various distortions. The whole process of our pattern recognition consists of receiving the episodes, their clustering into the groups of similar episodes and deriving the representatives of each cluster. These representatives will be used for further indexing collections. This paper is focused on the last step of this process – receiving the representatives of concrete clusters using Dynamic Time Warping method.

Keywords: dynamic time warping, clustering, pattern recognition, time series

1. INTRODUCTION

Processing and analyzing time series data is very important task in many domains, especially in modeling and simulations. In this domain, time series data is often used as one of simulation inputs, or can be produced as one of the simulation outputs. For this purpose, it is appropriate to be able to manage this type of data, e.g. describe the data nature, search in data in reasonable time, or to recognize characteristic patterns in collection. If such patterns are recognized, then they may be used for instance in data compression, for prediction or for indexing large collections. Time series analysis covers the 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. However, the problem arises for data collections that are a subject to different types of distortions, because the patterns can differ in time, shape or amplitude. In these cases, the classic methods for pattern recognition can fail.

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 (Keogh, Chu, Hart, and Pazzani 2004) 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 (Kocyan, Martinovic, Podhorányi, and Vondrak 2012) can be mentioned. These obtained episodes exactly belong to a previously mentioned group of distorted collections, because there are no strictly defined rules for generating the data collection (time series). Our effort is to provide a robust mechanism for retrieving characteristic patterns just from such distorted time series. In our case, the obtained patterns will be used for further creation of an index file, which will allow much faster and more accurate searching for similar episodes in large data collections. This will be used for better and faster prediction in our Case-Based Reasoning (Aamodt and Plazza, 1994) system (Kocyan, Martinovic, Unucka, and Vondrak 2009). The structure of suggested index file is shown in Figure 1, where each of the found patterns will contain its own group of similar episodes in original data collection.

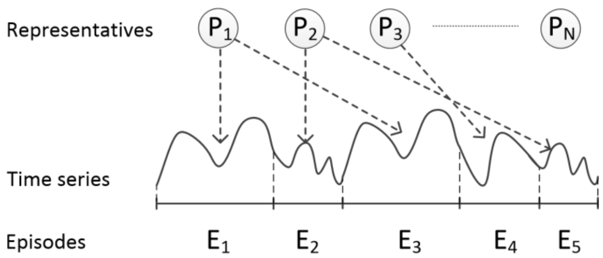

Figure 1: Collection of Representatives Pointing to Locations in Time Series

Then, once the most similar episodes in data collection will need to be found, a suitable pattern, which corresponds with the input sequence, will be searched first. Thereafter, it is possible to search in depth in a group of the selected pattern or a set of patterns, which are similar to a found episode from the input. By this way, the process of searching similar episodes will be speed up. However, there is question how to receive the patterns from distorted data collection and make the index file. 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 (Koutroumbas and Theodoridis 2008; Hand, Smyth, and Mannila 2001).

Our suggested approach is done in following manner. First of all, it is necessary to receive the particular episodes from data collection (i.e. cut the collection into the episodes). For instance, this can be done by the Voting Experts algorithm (Cohen, Adams, and Heeringa 2007) or by our unsupervised algorithm for retrieving characteristic patterns from time-warped data collections (Kocyan, Martinovic, Podhorányi, and Vondrak 2012). Once these episodes are obtained, they should be processed by a suitable clustering algorithm and divided into the clusters (Guojun, Chaoqun, and Jianhong 2007). 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. Finding the representative of a cluster is defined as finding such set of representative patterns P, which describe episodes E inside these clusters by the most appropriate way. There are two basic generally known ways for finding representatives. 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 new representative episode using the combination of episodes in the cluster.

While searching the representative, it is important to define a mechanism for comparing two episodes. In common, the Euclidean distance and other common methods for measuring the similarity between the episodes can be used. However, it is possible only while working with the undistorted episodes of the identical length. In cases where we have distorted 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 a new approach which combines the both ways for finding representatives using Dynamic time warping method (DTW) is presented in Section 2.

The organization of the paper is following: 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 optimization 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 (ECG, 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 2. 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).

DTW is a technique for finding the optimal matching of two warped episodes using pre-defined rules (Muller 2007). 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 2 can be seen in Figure 3. This approach was used for example for comparison of two voice patterns during an automatic recognition of voice commands (Rabiner 1993). The main goal of DTW method is a comparison of two time dependent episodes X and Y, where $X = (x_1, x_2, \dots, x_n)$ of the length $n \in \mathbb{N}$ and $Y = (y_1, y_2, \dots, y_m)$ of the 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 (Muller 2007).

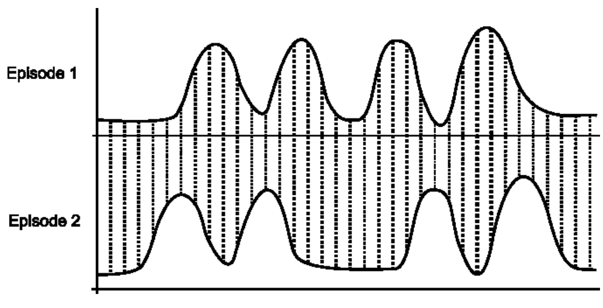

Figure 2: Standard Metrics Comparison

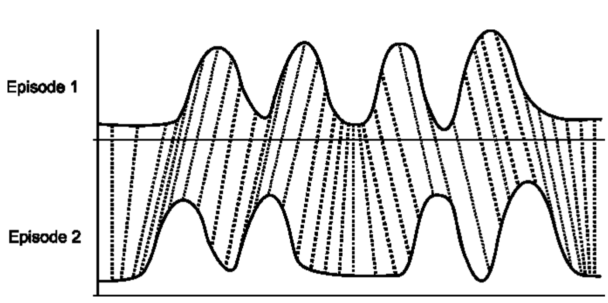

Figure 3: DTW Comparison

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 episode R is valid, that:

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

However, this approach is not sufficient in cases, where we have data with distortion. Examples of such episodes are presented in Figure 4a and 4b. If only we used simple average presented in Equation 1, we would achieve an episode showed in Figure 4c. As we can see, this episode absolutely is not a representative and all the information about the episode course is loosed.

As we can see from Figure 4, 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.

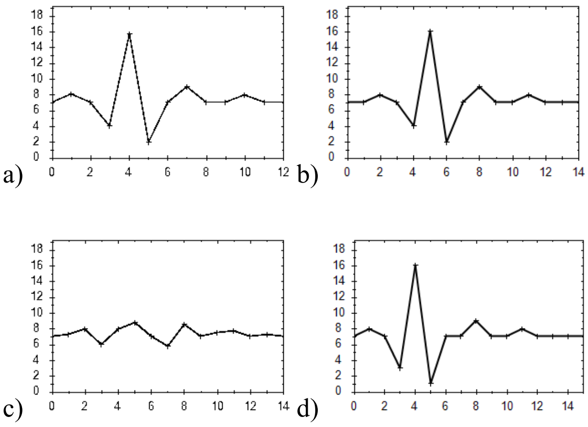

Figure 4: Sample Figure Caption

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 warping path $p^* = (p_1, p_2, \dots, p_L)$, which allows to find a representative. The approach for finding such representative is described in Algorithm 1. The output of presented algorithm applied on episodes in Figure 4 is presented in Figure 4d.

Algorithm 1: Searching Representative of a Pair
Input: Episodes X and Y
Output: Representative episode R
Steps:

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^*.
3. Value in the first position in R is determined as average of values in the first positions of episodes X and Y, ie. $r_1 = \frac{x_1 + y_1}{2}$.
4. ***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 $r_q = \frac{x_{n_l} + y_{m_l}}{2}$ is inserted into R;

   else if $(p_l - p_{l-1}) = (0,1)$ or $(1,0)$ ***then***
   - no item is inserted into R;

   end if
   - $l = l + 1$
   - Repeat Step 3.
5. ***end if***
6. Output of the algorithm is representative episode R of length q.

Algorithm 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:

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

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 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 a representative is 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 2.

Algorithm 2: Searching Representative of a Set
Input: Collection C of N episodes
Output: Representative episode R
Steps:

1. Initialization: N is count of input episodes, u=1 is level of collection; $C^1 = C$ is the first level of collection; $M = N - u + 1$ is count of processed episodes in level u.
2. Create from collection C^U, which consists of episodes $\{e_1, e_2, \dots, e_N\}$ 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.
   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 $\sum_{j=1}^{M} d_{min,j}^u = min_{\forall i=1,\dots,M}(\sum_{j=1}^{M} d_{ij}^u)$. 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.
3. 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

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 (Mean Absolute Error, Mean Percentage Error, Root Mean Square Error etc.). In such cases, it is necessary to adapt steps 2 and 4 of Algorithm 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 practical demonstrations of previously introduced methods are presented. First of all, the step by step example for better understanding of proposed algorithm will be demonstrated. Then, several outputs of the algorithm will be showed.

It must be noted that meaning and usage of DTW method is closer to a human judgment and perception of similarity than a machine definition of physical distance. For this reason, it is hard or almost impossible to perform a numerical evaluation for the following outputs (Berndt and Clifford, 1994), so the results will be presented only visually.

4.1. Step by Step Example

Consider we have two episodes $X = (1,\ 1, 4, 1, 10, 1)$, $Y = (1,\ 6, 1, 1, 10, 10, 1)$ and we want to find their mutual representative. First of all, a distance matrix (see Table 1), accumulated distance matrix (see Table 2) and warping path $p^* = \{(1,1),\ (2,1),\ (3,2),\ (4,3),\ (4,4),\ (5,5),\ (5,6),\ (6,7)\}$ have to be found according the steps listed in (Muller 2007). Visualization of found mutual episodes' mapping can be seen in Figure 5.

1	0	0	9	0	81	0
10	81	81	36	81	0	81
10	81	81	36	81	0	81
1	0	0	9	0	81	0
1	0	0	9	0	81	0
6	25	25	4	25	16	25
1	0	0	9	0	81	0
X/Y	**1**	**1**	**4**	**1**	**10**	**1**

Table 1: Distance Matrix

1	187	146.5	75.5	66.5	83	2
10	187	146.5	66.5	71	2	42.5
10	106	65.5	30.5	57.5	2	83
1	25	12.5	17	2	42.5	17
1	25	12.5	11	2	62	17
6	25	12.5	2	21.5	17	42
1	0	0	9	9	90	90
X/Y	**1**	**1**	**4**	**1**	**10**	**1**

Table 2: Accumulated Distance Matrix

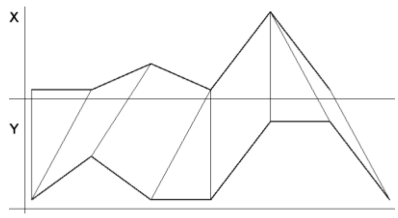

Figure 5: Found Mapping of Episodes

Now, the process of searching a representative can start. The first element of the representative is determined as $r_1 = \frac{x_1 + y_1}{2} = \frac{1+1}{2} = 1$. Then, we move to the next pair in the warping path. Since $(p_2 - p_1) = (2,1) - (1,1) = (1,0)$, no new element is added to the representative episode. However, the next step $(p_3 - p_2) = (3,2) - (2,1) = (1,1)$ causes an addition of a new element r_2, where $r_2 = \frac{x_3 + y_2}{2} = \frac{4+6}{2} = 5$. In the same way, the rest of the representative R is constructed:

$(p_4 - p_3) = (4,3) - (3,2) = (1,1) \rightarrow r_3 = \frac{1+1}{2} = 1$
$(p_5 - p_4) = (4,4) - (4,3) = (0,1) \rightarrow nothing\ to\ add$
$(p_5 - p_4) = (5,5) - (4,4) = (1,1) \rightarrow r_4 = \frac{10+10}{2} = 10$
$(p_6 - p_5) = (5,6) - (5,5) = (0,1) \rightarrow nothing\ to\ add$
$(p_7 - p_6) = (6,7) - (5,6) = (1,1) \rightarrow r_5 = \frac{1+1}{2} = 1$

The final found representative is then specified as $R = (1, 5, 1, 10, 1)$.

4.2. Demonstrations of Algorithm Outputs

As it was mentioned earlier in Section 4, it is almost impossible to numerically evaluate the success of the algorithm. For this reason, the following samples of output will be demonstrated only graphically and each of the result will be visualized in the following manner: the first row contains episodes, which were used as the input to the algorithm, whereas the second row consists of outputs for the different approaches. The first output is always the average of input episodes (defined in Equation 1), the second output is obtained from the proposed approach described in Section 3.2, and in the third case the Euclidean distance instead of DTW is used.

The first input dataset was a set of similar signals (see Figure 6), which shapes resemble ECG records. The signal is ended with tiny swings. As we can see from the second row of the episodes in Figure 6, the average of values from the both episodes absolutely degraded the signal information; the shift of signal peaks and drops was smoothed nearly to one level. Also usage of Euclidean distance did not provide sufficient results, which did not differ from averaged 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.

The next set of episodes contains signals with the three peaks mutually shifted in time, while each of them had a variable duration (see Figure 7). It was supposed that the representative would have a curve with the three evident peaks. It is obvious from the results, that even though the Euclidean distance worked much better, the loss of information was still noticeable.

The last input dataset represents the situation, in which the signal consists of two waves - one in a positive and one in a negative part (see Figure 8). 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.

The most important advantage of the proposed solution is the fact that the Algorithm 1 in combination with DTW is able to process even episodes with different lengths. This is very difficult while using other methods. 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 9 and 10, there are presented outputs from proposed algorithm applied on episodes with different lengths.

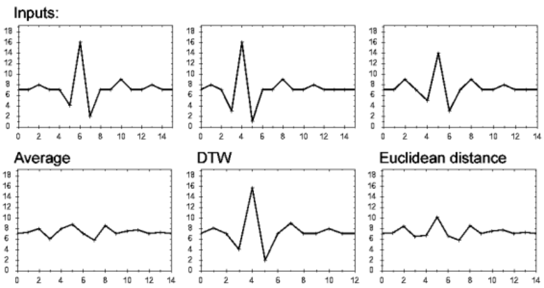

Figure 6: First set of inputs and outputs

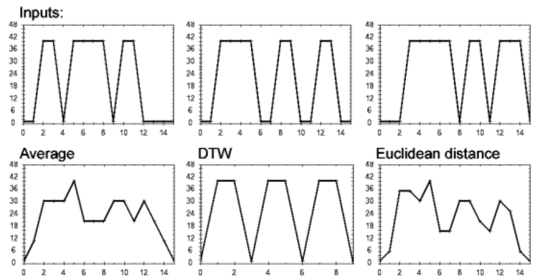

Figure 7: Second set of inputs and outputs

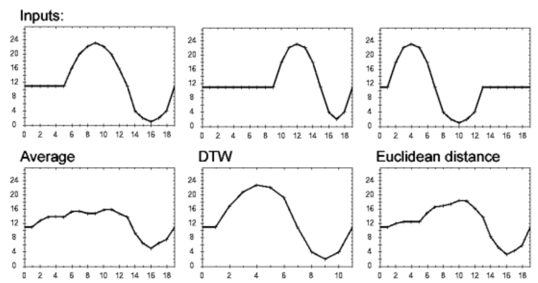

Figure 8: Third set of inputs and corresponding outputs

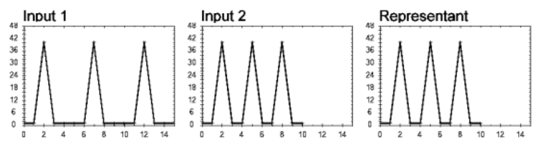

Figure 9: First set of inputs with variable lengths

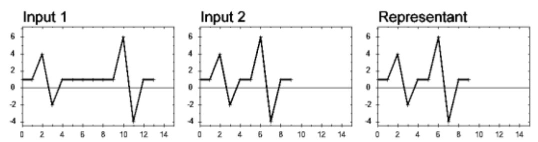

Figure 10: Second set of inputs with variable lengths

Observing the value of the weight parameter and its influence onto the resulting representative is also very interesting. If the weight parameter is not set (respectively if the weight is set to 1:1 – it means that the importance of the current level representative R^U and other episodes is equal), searching representative looks like as in the Figure 10. If the weight is set to 10:1

(the R^U is strongly preferred), the episodes in the next level will be strongly influenced by R^U as it is shown in Figure 11. On the other hand, if the weight is set to 1:10, the influence of R^U is minimal (in substance the R^U almost ignored) and the episodes in the next level are almost unchanged (see Figure 12).

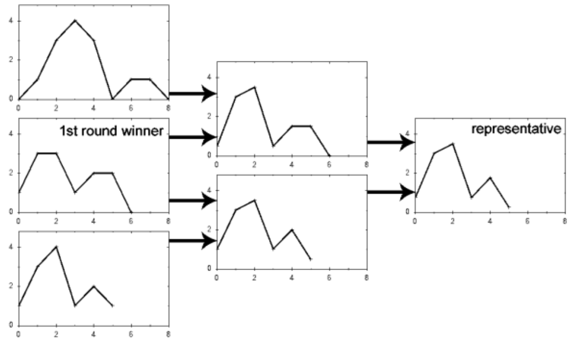

Figure 10: Weight parameter set to 1:1

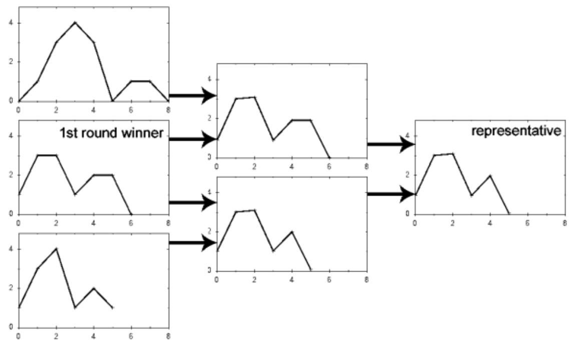

Figure 11: Weight parameter set to 10:1

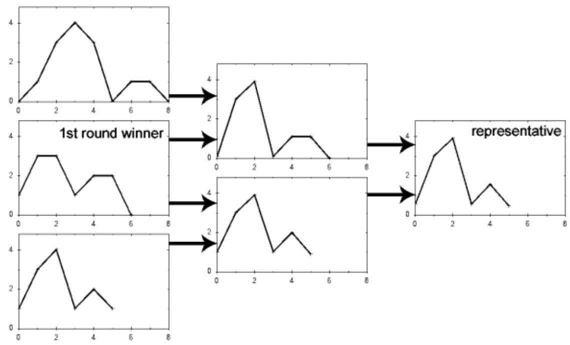

Figure 12: Weight parameter set to 1:10

5. CONCLUSION AND FUTURE WORK

The real application of proposed algorithm 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 which were distorted by time shifting. The proposed solution was compared with conventional methods, in which much worse success was obvious.

Further work will be focused 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.

ACKNOWLEDGMENTS

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 grant project of the Technology Agency of the Czech Republic under number TA01021374, titled New Technologies for Environmental Protection against Negative Effects of Natural Mass Transfer.

REFERENCES

Aamodt, A., Plaza, E., 1994. *CBR: Foundational issues, methodological variations, and system approaches,* AI COMMUNICATIONS, 39-59.

Berndt, D., Clifford, J., 1994. *Using Dynamic Time Warping to Find Patterns in Time Series*, In KDD Workshop (1994), pp. 359-370.

Cohen, P. R., Adams, N., and Heeringa, B, 2007. Voting Experts: An Unsupervised Algorithm for Segmenting Sequences. *In Journal of Intelligent Data Analysis*.

Evans, W.A., 1994. Approaches to intelligent information retrieval. *Information Processing and Management*, 7 (2), 147–168.

Guojun, G., Chaoqun, M., Jianhong, W, 2007. Data Clustering: Theory, Algorithms, and Applications. *ASA-SIAM Series on Statistics and Applied Probability*.

Hand, J., Smyth, P., Mannila, H, 2001. *Principles of Data Mining,* MIT Press, ISBN: 0-262-08290-X.

Keogh, E., Chu, S., Hart, D., Pazzani, M., 2004. *Segmenting Time Series: A Survey and Novel Approach,* Data mining in Time Series Databases, World Scientific, 1-21.

Kocyan, T., Martinovic, J., Podhorányi, M., Vondrak., I., 2012. Unsupervised algorithm for retrieving characteristic patterns from time-warped data collections. *Proceedings of The 11th International Conference on Modeling and Applied Simulation*.

Kocyan, T., Martinovic, J., Unucka, J., Vondrak., I., 2009. FLOREON+: using Case-Based Reasoning in a system for flood prediction. *Proceedings of ZNALOSTI 2009*.

Koutroumbas, K., Theodoridis S., 2008. *Pattern Recognition, 4th Edition.* Academic Press, ISBN:9781597492720 .

Liao, T. W., 2005. Clustering of time series data—a survey. *Journal of Pattern Recognition*, Volume 46, Issue 9, 2391-2612.

Muller, M., 2007. Dynamic Time Warping. *Information Retrieval for Music and Motion*, Springer, ISBN 978-3-540-74047-6, 69-84.

Rabiner, L., Biing-Hwang Juang, 1993. *Fundamentals of Speech Recognition,* Prentice Hall, ISBN: 0130151572.