Method of Time Series Similarity Measurement Based on Dynamic Time Warping

Lianggui Liu[1, *], Wei Li[1] and Huiling Jia[1]

Abstract: With the rapid development of mobile communication all over the world, the similarity of mobile phone communication data has received widely attention due to its advantage for the construction of smart cities. Mobile phone communication data can be regarded as a type of time series and dynamic time warping (DTW) and derivative dynamic time warping (DDTW) are usually used to analyze the similarity of these data. However, many traditional methods only calculate the distance between time series while neglecting the shape characteristics of time series. In this paper, a novel hybrid method based on the combination of dynamic time warping and derivative dynamic time warping is proposed. The new method considers not only the distance between time series, but also the shape characteristics of time series. We demonstrated that our method can outperform DTW and DDTW through extensive experiments with respect to cophenetic correlation.

1 Introduction

With the rapid development of the Internet, cloud computing, information technology and communication technologies, the rapid growth of data has caused many companies to face a variety of severe challenges and valuable opportunities. Nowadays, data is everywhere, our living and working habits have a fundamental change with the continuous emergence of data. Therefore, mankind has entered the era of big data from the age of information. Among the numerous data, time series data is the most widely used by people and has been applied in many fields including biology, economic, multi-media, image analysis and so on. Researchers and participants of data mining are also applying their technique and algorithms to this type of data. Among this data, mobile phone communication data is widely used in urban functional area division, population flow analysis, urban anomaly detection and so on. In the procedure of processing these time series data, the most common method is to apply a clustering algorithm after decomposing these data. Clustering is unsupervised learning. It can classify data according to certain rules, and clustering is also an important step in data mining.

---

[1] School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou, 310018, China.

[*] Corresponding Author: Lianggui Liu. Email: lgliu@zstu.edu.cn.

Due to the large amount of time series data, it is difficult to obtain satisfactory results by directly performing similarity queries, classification and clustering, and pattern mining on the original time series. Therefore, it is very necessary to carry out dimensionality reduction of time series data. The most intuitive benefit of data dimensionality reduction is that it is easy to calculate and visualize data when the dimension is reduced. And the deeper meaning lies in the extraction of effective information and the abandon of useless information. There are many methods for data dimensionality reduction, which can be mainly divided into linear dimensionality reduction and nonlinear dimensionality reduction. Linear dimensionality reduction means that the low-dimensional data obtained through dimensionality reduction can maintain the linear relationship between high-dimensional data points. The representative methods of linear dimensionality reduction include principal component analysis (PCA) and linear discriminant analysis (LDA). Nonlinear dimensionality reduction is based on kernels, representative methods have nucleation methods, and manifold learning (ISOMap, LLE, LPP). Among so many dimensionality reduction methods, Principal Component Analysis (PCA) is a basic mathematical analysis method. And its practical application is very extensive. It is also a common method for analyzing multivariate problems. Through PCA dimensionality reduction, a high-dimensional time series with a large amount of data can be compressed into a low-dimensional space, so that the time series can perform other related operations simply and quickly.

In the process of clustering operation, the most important step is to calculate the similarity between time series. Nowadays, there are many classic algorithms that are worth learning. Euclidean distance is one of the most commonly used methods for calculating the similarity of time series. Although Euclidean distance can quickly calculate the similarity, the disadvantage of Euclidean distance is also obvious. First, it can only measure the time series of equal length. The second is that it is sensitive to noise, because some individual points that deviate from the average can have a great impact on the results. And because the dimensions of various data in time series are not the same, it is difficult to determine the similarity measure method between time series. Therefore, compared with the traditional similarity measure method, we need to find a more appropriate method. Dynamic Time Warping [Muller (2007)] (DTW) is one of the most robust methods for similarity measurement. It was first applied in speech recognition, and now it has been widely applied in the study of the similarity measure of time series. DTW can not only calculate the similarity of time series with different length, but also the quality of calculation can reflect the difference between data. However, the DTW cannot reflect the overall morphological characteristics of the time series. We all know that the derivative of the data can be used to describe the characteristics of these sequences. Therefore, Derivative Dynamic Time Warping (DDTW) [Łuczak (2016)] is introduced to achieve the above purpose. Above all, our method can consider the morphological characteristics when calculating the similarity between two time series.

This paper introduces some different methods for measuring the similarity of mobile communication time series. And the main contributions of this paper are two parts.
- DTW and DDTW are applied to measure the similarity of mobile communication time series.
- Combining the DTW and DDTW algorithm with a weight, a more accurate

similarity measure can be obtained by changing the value of the weight.

The rest of this paper is as follows. Section II introduces some methods used in mobile communication time series analysis. Section III mainly introduces the principle of DTW and DDTW algorithm. Section IV is the experimental verification and analysis. And the summary of the full paper is placed in Section V.

2 Related work

2.1 Dynamic time warping

Nowadays, almost everyone uses mobile phones for communication. And mobile communication data has become an important time series for research. Meanwhile, Dynamic Time Warping is widely used in time series research due to its unique advantages. Dynamic Time Warping [Keogh and Ratanamahatana (2005)] has been widely applied to many fields such as time series data mining [Muscillo, Conforto, Schmid et al. (2007)], speech recognition [Itakura (1975)] and so on. However, DTW's quadratic time and space complexity limits its performance. So far, plenty of methods are used to speed up the calculation process of DTW [Zhou and Wong (2005); Keogh and Pazzani (2000)]. The Sakoe-Chuba band and the Itakura parallelogram are the most commonly applied methods. They can limit the number of elements in the cost matrix R, and reduce the path range for searching for sub-optional warping paths. However, the performance of DTW using the two methods always relies on a constant factor. Moreover, it cannot retrieve the optimal warping path which is often exceeds the path range. FTW [Sakurai, Yoshikawa and Faloutsos (2005)] is often applied to the similarity search of time series and can get results faster than the original DTW method.

Dynamic Time Warping is a method that can analyze time series with different lengths. And DTW can align time series with each other by expanding and compressing the dimension of these time series. For the warping function to be truly reflected the progress of the process, some constraints need to be imposed on the dynamic time warping algorithm to avoid an alignment which is over-aggressive or involves pathological warping. Spooner et al. [Spooner, Kold and Kulahci (2017)] studied the use of local constraints in dynamic time warping and defined the standard to evaluate the degree of time distortion and get the synchronization of variables. When calculating the similarity of time series, in order to get the minimum cumulative distance, the DTW distance may map multiple points in one time series to one point in another. This can make the time series overstretched and compressed, resulting in the loss of important feature information and affecting the classification accuracy. To solve this problem, Wan et al. [Wan, Chen and Shi (2017)] proposed an adaptive cost dynamic time warping distance (AC-DTW) method that can adjust the number of points on one time series mapped to the points on another. At the same time, AC-DTW records the trajectories of all points and then adaptively allocates the cost rate to each point by calculating cost function at the next step. Compared with other existing methods in experiment, AC-DTW has higher accuracy. In the study of time series, Górecki at al. [Górecki and Łuczak (2013)] proposed a novel distance function based on the derivative of time series. Compared with traditional metrics, the new method considers the general shape of the time series, not just the point-to-point function comparison. The new distance is applied to classification

using the nearest neighbor rule. And experiments have shown that this method can provide high quality classification for most inspection data sets. After this, Górecki et al. [Górecki and Łuczak (2014)] improved the previous algorithm by adding second derivatives and achieved good results. In addition, they [Górecki and Łuczak (2015)] proposed a novel method for multivariate time series classification using parametric derivative dynamic time warping distance. This method combines the DTW distance of multivariate time series and the DTW distance of the derivative of multivariate time series. And this method is used in classification with the nearest neighbor rule. Through experiments, this method has a good classification effect for 18 data sets. Jeong et al. [Jeong, Jeong and Omitaomu (2011)] proposed a new distance metric, called Weighted DTW (WDTW). This method penalizes points which have higher phase difference between a reference point and a testing point to prevent the minimum distance distortion caused by the abnormal value. At the same time, a new weighting function called the modified logical weight function is also proposed. And this function can systematically assign weights based on the phase difference between a reference point and a testing point. By applying different weights to adjacent points, the algorithm can enhance the detection of similarity between two time series. And through experiments, the proposed method can improve the accuracy of time series classification and clustering problems. Hsu et al. [Hsu, Huang, Yang et al. (2015)] proposed a flexible dynamic time warping method for measuring the similarity of two time series. This method uses the additional score as a reward for consecutive long one-to-one segments. Compared with separate methods such as DTW and DDTW, the average error rate of this method is lower.

2.2 Other methods

Mobile communication time series is used in many fields, and lots of researches have focused on understanding aggregated urban mobility patterns based on mobile phone datasets. And Dynamic Time Warping is used to measure the similarity of different mobile phone communication data. Yuan et al. [Yuan and Raubal (2012)] use mobile phone dataset to extract and represent the dynamic mobility patterns in different urban areas. They can get the detection of outlier urban patterns by classifying the time series based on hierarchical clustering. In addition to DTW, time series clustering algorithms use many kinds of distance or similarity measures. And each distance may reflect a different kind of similarity between time series. Distances reflecting model-based similarities include the Hidden Markov Models (HMM) distance [Smyth (1997)], and Auto-Regressive Moving Average (ARMA) distance [Kalpakis, Gada and Puttagunta (2001)]. Shape-based measures, reflecting similarities in the time and shape domain of the time series include Longest Common Sub-Sequence (LCSS) distance [Vlachos, Kollios and Gunopulos (2002)] and Minimal Variance Matching (MVM) distance [Latecki, Megalooikonomou, Wang et al. (2005)]. However, most of these methods may not measure the similarity between time series which has different length. Meanwhile, they may not consider the shape characteristics when calculate the distance. Our method can solve these problems and may have good results.

3 Method

3.1 Dynamic time warping

Euclidean Distance and Dynamic Time Warping are two methods that are often used in calculating the similarity of time series. Euclidean Distance is very sensitive to slight changes on the time axis, and DTW distance can effectively eliminate this defect. DTW is an algorithm that searches for the best contrast between two time series. It returns a distance measure that calculates the similarity between them. The sequence allows local nonlinear distortion in the time dimension, and the DTW deals with local deformation to some extent. The principle of DTW is shown in the figure below:
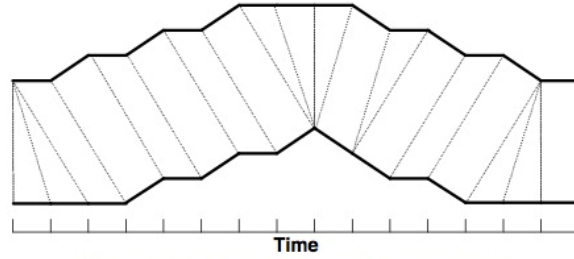

Figure 1: Schematic diagram of dynamic time warping

Suppose there are two time series Q and C, their lengths are n and m respectively, and

$$Q = q_1, q_2, ..., q_i, ..., q_n \tag{1}$$

$$C = c_1, c_2, ..., c_j, ..., c_m \tag{2}$$

In order to align two sequences by using DTW, we construct an n-by-m matrix, where the element (i^{th}, j^{th}) in the matrix is the distance $d(q_i, c_j)$ between two points q_i and c_j. Generally, the distance between two points is represented by Euclidean Distance, which is $d(q_i, c_j) = (q_i - c_j)^2$. Each element (i, j) in the matrix corresponds to the alignment distance between two points q_i and c_j. The warping path W is a group of contiguous sets of matrix elements, reflecting the mapping between sequences Q and C. And the k-th element in W is defined as $w_k = (i, j)_k$, then

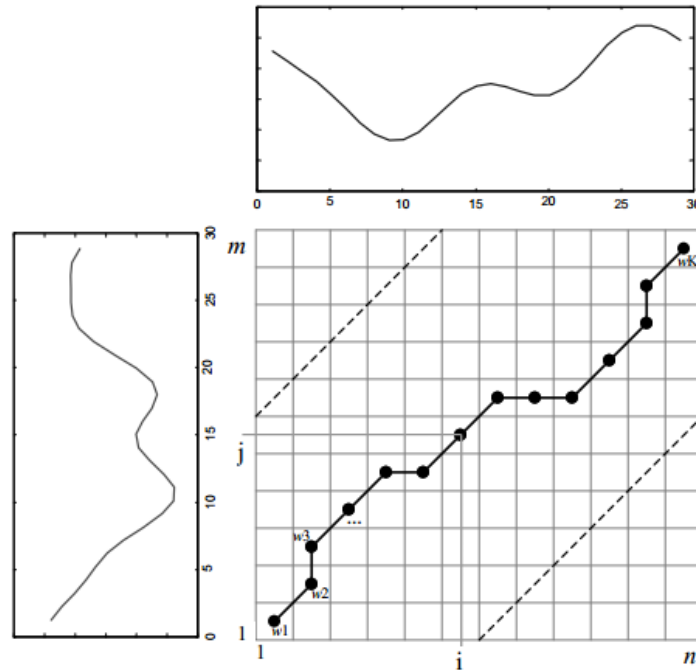$$W = w_1, w_2, ..., w_k, ..., w_K \qquad \max(m, n) \le K < m + n - 1 \tag{3}$$

Figure 2: The warping path of dynamic time warping

In general, the warping path is subject to the following constraints:

1) Boundary conditions: Generally, $w_1 = (1,1)$ and $w_k = (m,n)$, this requires that the warping path must start from the lower left corner and end in the upper right corner.

2) Continuity: If $w_{k-1} = (a',b')$, the next point $w_k = (a,b)$ in the path needs to satisfy $a - a' \leq 1$ and $b - b' \leq 1$. That means that it is impossible to match across a certain point, and it can only be aligned with its neighboring points. This ensures that every coordinate in Q and C appears in W.

3) Monotonicity: If $w_{k-1} = (a',b')$, the next point $w_k = (a,b)$ in the path needs to satisfy $a - a' \geq 0$ and $b - b' \geq 0$. This requires that the points above W must be monotonous over time, ensuring that the dashed lines in the graph do not intersect.

There are many paths that meet these constraints, and then we get the minimum cost path through the following formula:

$$DTW(Q,C) = min \left\{ \sqrt{\sum_{k=1}^{K} w_k} \Big/ K \right\} \tag{4}$$

The K in the denominator is mainly used to compensate for different length of warping paths. The idea of DTW is to extend and shorten the time series to get the shortest similarity distance between two time series, which is the most similar one. The shortest distance is the final distance measure of two time series. To do this, we define a cumulative distance, starting from point $(0,0)$, which matches the two sequences Q and C. At each point, the previously calculated distances for all points are accumulated. After

reaching the end point (n, m), this cumulative distance is the final total distance, which is the similarity between Q and C. The cumulative distance $\gamma(i, j)$ is the sum of the current grid distance $d(i, j)$ and the minimum distance of adjacent elements that can reach this point:

$$\gamma(i, j) = d(q_i, c_j) + min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\} \tag{5}$$

The best path is the one that makes the cumulative distance along the path reach a minimum value. And this path can be obtained through dynamic programming algorithm.

3.2 Derivative dynamic time warping

Since the DTW attempts to interpret the variability of time series value by distorting the time axis, it easily leads to non-intuitive alignment. That is, a single time-series point maps to the majority of points in another time series, and this point is called a "singular point". Furthermore, if q_i is a point on the uptrend in a time series and c_j is a point on the downtrend in another time series, due to this obvious characteristic, we will not map the two-time series directly. In order to solve such problems, DDTW, which comes from modifying DTW, considers more about the shape characteristic of time series.

In the DTW algorithm, the distance between q_i and c_j is expressed as $d(i, j) = (q_i - c_j)^2$. However, in the DDTW algorithm, we first compute the difference between the derivatives of q_i and c_j, then compute the square of the difference to replace $d(i, j)$. It uses

$$D_i(q) = q_i - q_{i-1} \tag{6}$$

to evaluate the value of q_i. Similarly, the derivative of points in the time series C can also be obtained by the above formula, and can be expressed as $D_j(c)$. Then the new distance formula is:

$$D'(i, j) = (D_i(q) - D_j(c))^2 \tag{7}$$

Then, the original distance D is replaced by the new distance D', and the remaining steps of the DDTW are the same with the DTW's. From the figure below, we can also see that DDTW can reflect the trend of the sequence, but DTW does not achieve this effect.
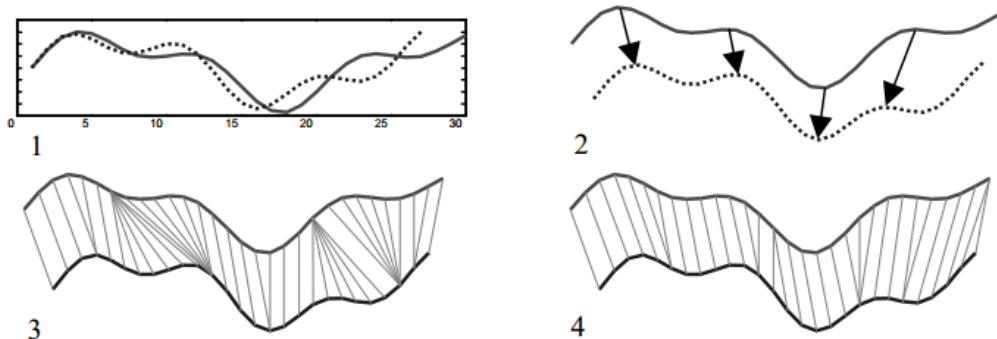

Figure 3: 1) Two artificial time series. 2) The intuitive feature to feature warping alignment. 3) The alignment produced by DTW. 4) The alignment produced by DDTW

In this paper, in order to better calculate the similarity between time series, we combine DTW and DDTW to get the similarity distance between sequences. This method not only calculates the similarity, but also considers the comparison of trends between sequences, which is:

$$Dist = (\alpha)\,DTW + (1-\alpha)\,DDTW \tag{8}$$

And *Dist* in the above formula is a measure of similarity between time series that meets our requirements.

4 Experiments and analysis

The data set used in this paper is a seasonal time series data decomposed from mobile phone communication data in Cici et al. [Cici, Gjoka, Markopoulou et al. (2015)]. There are a total of 2726 time series, and the length of each time series is 4032. Because the length of time series used in this paper is too long, we use the PCA to extract features of the time series. The length of new time series after dimensional reduction is 8, which greatly reduces the length of the time series and improves computational efficiency. Next, we use DTW, DDTW and hybrid method respectively to calculate the similarity between time series, and apply it to hierarchical clustering algorithm. In the experiment, α is set in the range of 0 to 1 and in steps of 0.01. Finally, we calculate cophenetic correlation coefficient. The cophenetic correlation for a cluster tree is defined as the linear correlation coefficient between the cophenetic distances obtained from the tree, and the original distances used to construct the tree. Thus, it is a measure of how faithfully the tree represents the dissimilarities among observations. Finally, we get the figure as shown below:
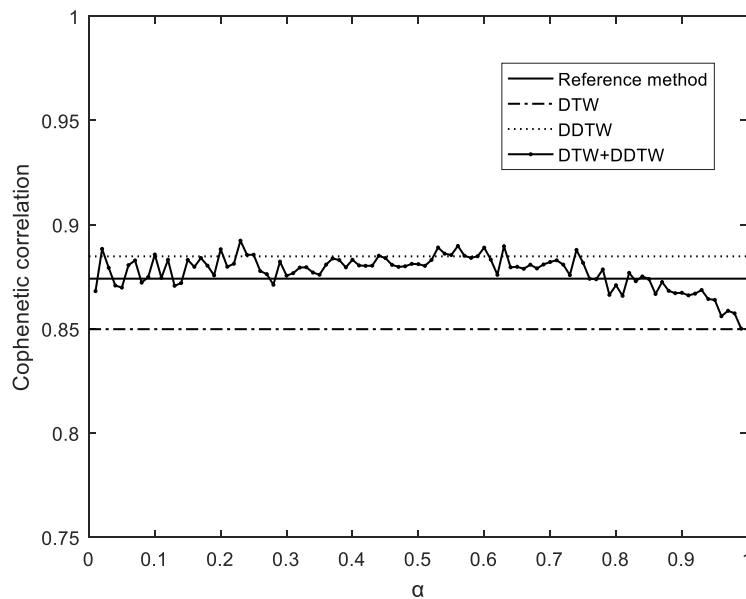

Figure 4: The results with different methods

It can be observed in the figure that the method only using the DTW distance is slightly less effective. It may be because the two time series with large differences in characteristics are warping to the optimal path, resulting in a low correlation coefficient. On contrary, DDTW considers the morphological characteristics of the time series in the calculation process, so the result obtained is better than DTW. When using mixed metric method, different weights have different effects, and when the weight is 0.23, a 2% performance improvement is obtained.

5 Conclusion

There are lots of methods to measure the similarity of time series, but most of them ignore the comparison of morphological characteristics between sequences. DTW can directly calculate the distance between time series and has good robustness. DDTW can compare the differences in morphological characteristics between time series, and also prevent over-warping of the DTW. This paper proposes a hybrid method based on the combination of DTW and DDTW to calculate the similarity between time series. By combining the two methods, the similarity measure of time series can be made more accurate, and the effectiveness of the method is verified by experiments.

Reference

Cici, B.; Gjoka, M.; Markopoulou, A.; Butts, C. T. (2015): On the decomposition of cell phone activity patterns and their connection with urban ecology. *ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 317-326.

Górecki, T.; Łuczak, M. (2013): Using derivatives in time series classification. *Data Mining and Knowledge Discovery*, vol. 26, no. 2, pp. 310-331.

Górecki, T.; Łuczak, M. (2014): First and second derivatives in time series classification using DTW. *Communications in Statistics-Simulation and Computation*, vol. 43, no. 9, pp. 2081-2092.

Górecki, T.; Łuczak, M. (2015): Multivariate time series classification with parametric derivative dynamic time warping. *Expert Systems with Applications,* vol. 42, no. 5, pp. 2305-2312.

Hsu, C. J.; Huang, K. S.; Yang, C. B.; Guo, Y. P. (2015): Flexible dynamic time warping for time series classification. *Procedia Computer Science*, vol. 51, pp. 2838-2842.

Itakura, F. (1975): Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 23, no. 1, pp. 67-72.

Jeong, Y. S.; Jeong, M. K.; Omitaomu, O. A. (2011): Weighted dynamic time warping for time series classification. *Pattern Recognition*, vol. 44, no. 9, pp. 2231-2240.

Kalpakis, K.; Gada, D.; Puttagunta, V. (2001): Distance measures for effective clustering of ARIMA time-series. *IEEE International Conference on Data Mining*, pp. 273-280.

Keogh, E. J.; Pazzani, M. J. (2000): Scaling up dynamic time warping for datamining applications. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 285-289.

Keogh, E.; Ratanamahatana, C. A. (2005): Exact indexing of dynamic time warping. *Knowledge and Information System*, vol. 7, no. 3, pp. 358-386.

Latecki, L. J.; Megalooikonomou, V.; Wang, Q.; Lakaemper, R.; Ratanamahatana, C. A. et al. (2005): *Elastic Partial Matching of Time Series*. Springer Berlin Heidelberg, pp. 577-584.

Łuczak, M. (2016): Hierarchical clustering of time series data with parametric derivative dynamic time warping. *Expert Systems with Applications*, vol. 62, pp. 116-130.

Muller, M. (2007): Dynamic time warping. *Information Retrieval for Music and Motion.* Springer Berlin Heidelberg, pp. 69-84.

Muscillo, R.; Conforto, S.; Schmid, M.; Caselli, P.; Alessio, T. D. (2007): Classification of motor activities through derivative dynamic time warping applied on accelerometer data. *International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4930-4933.

Sakurai, Y.; Yoshikawa, M.; Faloutsos, C. (2005): FTW: Fast similarity search under the time warping distance. *ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pp. 326-337.

Smyth, P. (1997): Clustering sequences with hidden markov models. *Nips*, vol. 9, pp. 648-654.

Spooner, M.; Kold, D.; Kulahci, M. (2017): Selecting local constraint for alignment of batch process data with dynamic time warping. *Chemometrics and Intelligent Laboratory Systems*, vol. 167, pp. 161-170.

Vlachos, M.; Kollios, G.; Gunopulos, D. (2002): Discovering similar multidimensional trajectories. *International Conference on Data Engineering*, pp. 673-684.

Wan, Y.; Chen, X. L.; Shi, Y. (2017): Adaptive cost dynamic time warping distance in time series analysis for classification. *Journal of Computational and Applied Mathematics*, vol. 319, pp. 514-520.

Yuan, Y.; Raubal, M. (2012): *Extracting Dynamic Urban Mobility Patterns from Mobile Phone Data*. Springer Berlin Heidelberg, pp. 354-367.

Zhou, M.; Wong, M. H. (2005): A segment-wise time warping method for time scaling searching. *Information Sciences*, vol. 173, no. 1-3, pp. 227-254.