2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010)

A Similarity Measure of Jumping Dynamic Time Warping

Lin Feng, Xiaoyan Zhao, Yiwei Liu, Yuan Yao, Bo Jin*
School of Innovation Experiment
Dalian University of Technology
Dalian, P.R. China

Abstract—The problem of similarity measure for time series has attracted considerable research interest. Most of the recently used algorithms utilize the Dynamic Time Warping (DTW) distance for measuring the similarity of time series, in various areas such as science, medicine, industry, and finance. DTW is a considerably more robust distance measure for time series, which allows similar shapes to match even if they are of different lengths. Unfortunately however, several serious problems are associated with the use of DTW, such as high complexity and "one to many" problems. The present study is aimed at introducing a novel technique for improving the DTW algorithm, known as Jumping Dynamic Time Warping (JDTW). It is proven that this approach improves the efficiency with lower omission factor and reduces the noise impact of query sequence.

Keywords-time series; dynamic time warping; similarity measure

I. INTRODUCTION

A time series is a sequence of data points, measured typically at successive times, and spaced at (often uniform) time intervals such as that of stock prices, exchange rates, and weather data [1]. One of the key issues of time series analysis is similar sequence matching. Similar sequence matching in time series databases involves a problem of finding data sequences similar to the given query sequence in the database [2]. There are two aspects of similar sequence matching: whole matching and part matching.

Most algorithms used for matching time series utilize the Euclidean distance [3] or Dynamic Time Warping (DTW) distance [4]. However, it has been clearly shown that the Euclidean technique is a very brittle distance measure. On the other hand, DTW is a flexible distance measure and hence it has been widely used in the areas of science, medicine, industry, and finance. Unfortunately however, there may be a limitation of DTW for performance on very large databases [5]. In order to overcome this limitation, some lower bounding distance measures of DTW were proposed for improving the efficiency, which, however, reduced the precision. It was found that Fast search method for dynamic Time Warping (FTW) [6] improved the efficiency with the segmented DTW. Nevertheless, it faced the problem of "one to many", which would lead to multi-point aggregation.

In the present study, a novel technique is introduced for similar sequence matching of time series known as Jumping Dynamic Time Warping (JDTW). In the JDTW technique, each point of the query sequence can match in a certain range. Thus, after calculating the distances among the points in the range, the shortest distance is selected and the longer ones are jumped over. There are two main characteristics of JDTW, one is the certain range for improving the efficiency, and the other is jumping over the further points for reducing the noise impact.

II. RELATED WORKS

Similar sequence matching in time series databases is a problem of finding data sequences similar to the given query sequence from the database. It has been used in several data mining applications such as rule discovery [7], clustering [8], and classification [9,11].

A. DTW Distance

DTW is a flexible distance measure, which allows warping of the time axes in order to better align the shapes of the two times series.

Suppose there are two time series: $Q(q_1,q_2,...,q_n)$ and $C(c_1,c_2,...,c_m)$. Then, an n-by-m matrix should be constructed, as shown in Fig. 1. The element (i^{th}, j^{th}) of matrix is the distance between two points q_i and c_j. A warping path is a continuous set of matrix elements that defines a mapping between Q and C. The warping path is typically subject to several constraints: boundary conditions, continuity, and monotonicity. The shortest warping path is defined as DTW distance.

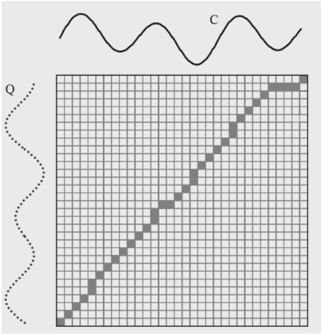


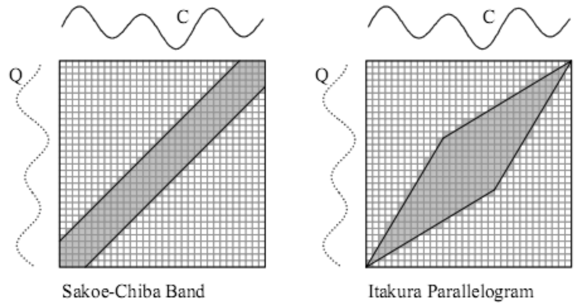
Figure 1. The n-by-m matrix for DTW distance

---

* Corresponding Author. E-mail: jinbo@dlut.edu.cn

B. *Lower Bounding of DTW*

Search similarity under DTW is a heavily CPU bound. Pruning sequences which can not possibly be a best match will save a considerable amount of time. Thus the lower bounding of DTW is introduced. There are two typical constraints used for the lower bounding of DTW, as shown in Fig. 2. With a lower bounding technique, the time complexity of DTW can be reduced from $O(n^2)$ to $O(n)$ [5].



Sakoe-Chiba Band

Itakura Parallelogram

Figure 2. Constraints for lower bounding of DTW

C. *Path Constraints of DTW*

Reference [10] presents a commonly used local path constraint to set the fan-in of only 27°-45°-63°, as shown in Fig. 3. It supports the jumping process. If there are noise points, the mapping path could ignore the noise points. Since the optimum-value function is based on the total distance, the mapping path will sometimes take the local paths of 27° or 63° instead of 45° in an attempt to minimize the total distance.

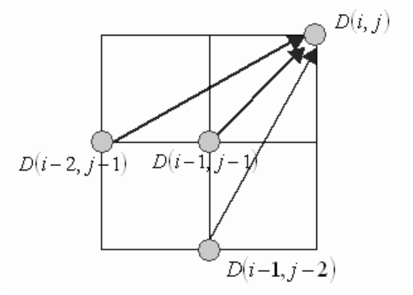


Figure 3. Local path constraints

With the “anchored beginning and anchored end”, the global path restriction can be got, as shown in Fig. 4.

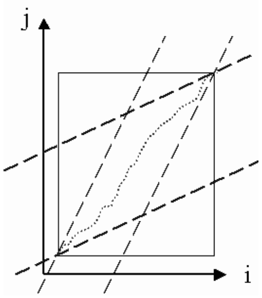


Figure 4. Global path constraints

The global path constraints shown in Fig.4 are deduced from the local path constraints of 27°-45°-63° shown in Fig. 3. The global path constraints serve two purposes: (1) reduce the computation load; and (2) limit the mapping path to be a reasonable one.

III. JUMPING DYNAMIC TIME WARPING

The present study presents a similarity measure of jumping dynamic time warping (JDTW). In JDTW, each point of a sequence is allowed to match with several points of another sequence. Subsequently, the nearest matched distance is chosen, and the further distances are jumped over. Thus, JDTW has the characteristic of reducing noise impact.

A. *Local Path Constraints of JDTW*

In reference [10], the mapping path will jump only one point each time. Subsequently, the length ratio of the two time series should be limited from 0.5 to 2, otherwise, the method will invalidate. For this purpose, in JDTW a point matching range *r is presented*. Suppose there are two time series $X(x_1,x_2,\ldots,x_n)$ and $Y(y_1,y_2,\ldots,y_m)$, then, r can be formulized as:

$$r=\left\{\begin{array}{ll}\lfloor n/m\rfloor & n\ge m\\ \lfloor m/n\rfloor & n<m\end{array}\right\} \qquad (1)$$

With the point matching range, the mapping path of JDTW can jump over more than one point. Each point can be matched with several points, shown in Fig. 5.

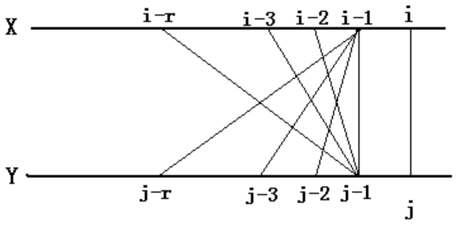


Figure 5. The range limitation of point matching

The local path constraints shown in Fig. 6 can be obtained with the point matching range r. Subsequently, the fan-in can be set from *arctan(1/r)* to *arctan(r)* for the local path constraints. The nearest matched distance is chosen, and the further distances are jumped over.

Similar to DTW, JDTW presents an n-by-m matrix for calculating the distance between two time series, as shown in Fig. 7.

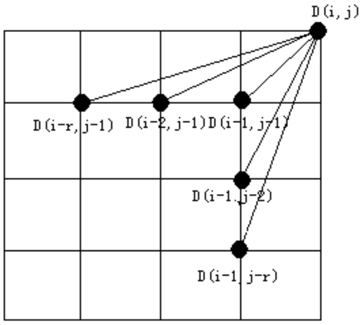


Figure 6. Local path constraints of JDTW

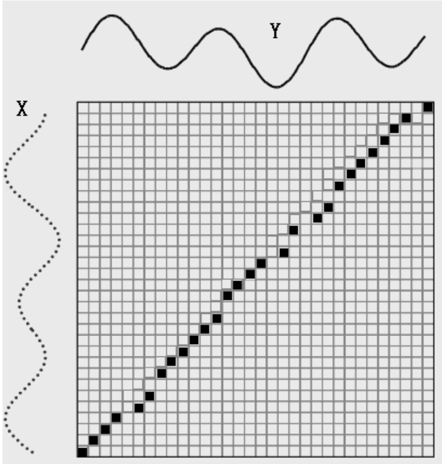


Figure 7. The mapping path supporting jumping

A mapping path *W* is a correspondingly contiguous set of matrix elements that defines a mapping between *Q* and *C*. The k-th element of *W* is $w_k=(i_k,j_k)$, then *W* can be formulized as:

$$W=(w_1,w_2,\ldots,w_k,\ldots,w_K) \quad \max(n,m)/r \le K \le \max(n,m) \tag{2}$$

Where, the mapping path should satisfy the following conditions:

(1) Bounding condition: $w_1=(1,1)$ and $w_K=(n,m)$.

(2) Continuity: If $w_k=(a,b)$, and $w_{k-1}=(a',b')$, then $(a-a'\le r)$ $\&\&(b-b'=1)$ or $(b-b'\le r)\&\&\ (a-a'=1)$.

(3) Monotonicity: If $w_k=(a,b)$, and $w_{k-1}=(a',b')$, then $a-a'\ge 1$ and $b-b'\ge 1$.

The shortest mapping path is defined as JDTW distance which formulized as:

$$D_{jdtw}(X,Y)=\min(\sum_{k=1}^{K} w_k) \tag{3}$$

The shortest mapping path can be calculated by dynamic programming by defining the cumulative distance $D_{dist}(i,j)$ as the distance *d(i,j)* found in the current cell and the minimum of the cumulative distances of the adjacent elements.

$$\begin{aligned} D_{dist}(i,j) &= d(x_i-y_j)+\min\{D_{dist}(i-1,j-1),D_{dist}(i-1,j-2), \\ &\ldots,D_{dist}(i-1,j-r),D_{dist}(i-2,j-1),\ldots,D_{dist}(i-r,j-1)\} \end{aligned} \tag{4}$$

where, $d(x_i,y_j)$ is the distance between the i-th point of X and the j-th point of Y.

The 27-45-63 local paths in reference [10] can be taken as a special case of JDTW when r=2.

B. Global Path Constraints of JDTW

Since the required mapping path is "anchored beginning, anchored end", the local path constraints can induce the global path constraints, as shown in Fig. 8. The global path constraints are addressed from *arctan(1/r)* to *arctan(r),* and thus the mapping path can be limited to be a reasonable one.

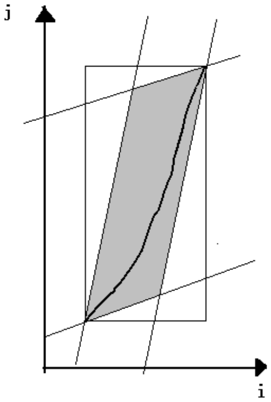


Figure 8. Global path constraints of JDTW

The feasible region for the mapping path is a trapezoid, as shown in Fig. 8. The computation time can be saved by the use of global path constraints. It is not required to calculate the white area. The non-calculated area can be formulized as:

$$\begin{cases} w(i,1)\cdots w(i,\left\lceil \dfrac{i-1}{r} \right\rceil) & (i=2\cdots n) \\ w(i,m-\dfrac{n-i+1}{r}+1)\cdots w(i,m) & (i=2\cdots n) \\ w(1,j)\cdots w(\left\lceil \dfrac{j-1}{r} \right\rceil,j) & (j=2\cdots m) \\ w(n-\dfrac{m-j+1}{r}+1,j)\cdots w(n,j) & (j=2\cdots m) \end{cases} \tag{5}$$

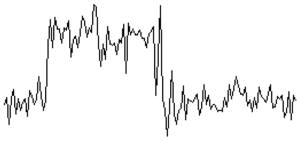
In summary, there are two aspects of global path constraints:

(1) To reduce the computation load.

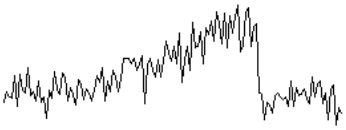
(2) To limit the mapping path to be a reasonable one.

IV. EXPERIMENTAL EVALUATION

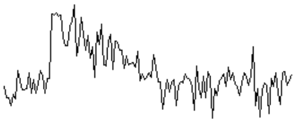
Thirty thousand CBF (Cylinder-Bell-Funnel) data were selected as the test data. Subsequently, noise data were injected into the original CBF data to create the test database. The samples of the original CBF data are shown in Fig. 9, and the samples of processed CBF data are shown in Fig. 10.



a. Cylinder

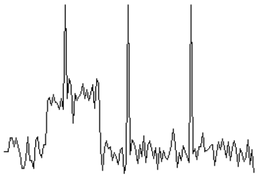


b. Bell

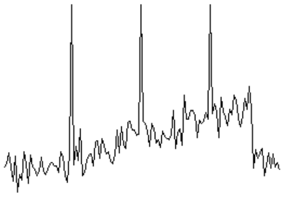


c. Funnel

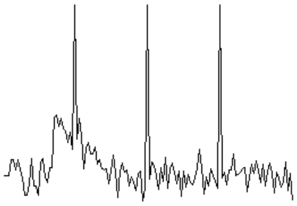
Figure 9. Samples of original CBF (Cylinder-Bell-Funnel) dataset



a. Cylinder



b. Bell



c. Funnel

Figure 10. Samples of processed CBF (Cylinder-Bell-Funnel) dataset

In the present study, comparison experiments were performed with the k-nearest neighbor (k-NN) algorithm (k=5000). Two similarity measures (JDTW and DTW) were compared in the same k-NN algorithm, and the measure quality was evaluated by precision. Precision is the proportion of the returned sequences that are correctly matched. In JDTW, the computation time can be saved by skipping the elements outside the parallelogram, whereas in DTW, the whole matrix must be computed. Thus, the improvement in efficiency is obvious, and therefore, the cost time is not included in the comparison experiment.

In reference [10], the length ratio of the two time series should be limited from 0.5 to 2; otherwise, the method will invalidate. It should be noted that a comparison experiment between JDTW and the method in reference [10] has not been designed because it can be seen as a special case of JDTW.

A. Comparison Experiment with Original CBF Data

In this experiment, the original CBF data is used, and the lengths of data sequence and query sequence are both equal to 128. There are three aspects of this experiment:

- Data sequences use the original CBF data, and query sequences use random sequence data.
- Data sequences use the processed CBF data, and query sequences use random non-noise sequence data.
- Data sequences use the processed CBF data, and query sequences use random noise sequence data.

Each aspect of the experiment includes ten query sequences. The experimental results are shown in Table 2. The JDTW achieves nearly the same results when non-noise sequence is used. When the noise sequence is used, it is found that the precision of JDTW is 9 percent higher than that of DTW.

TABLE I. PRECISION COMPARISON BETWEEN DTW AND JDTW

Query Sequence	**DTW**	**JDTW**
Random sequence from original CBF database	0.9969	0.9859
Random non-nosie sequence from processed CBF database	0.9980	0.9988
Random nosie sequence from processed CBF database	0.8035	0.8926

B. Comparison Experiment with Processed CBF Data

In this experiment, we used the processed CBF data. There are two aspects of this experiment:

- Data sequences use the processed CBF data with length of 128, and query sequences use random non-noise sequence data with length of 32, 64, 96, and 128.
- Data sequences use the processed CBF data with length of 128, and query sequences use random noise sequence data with length of 32, 64, 96, and 128.

Each aspect of the experiment includes ten query sequences. The precisions of DTW and JDTW under different query sequence lengths are shown in Figs. 11 and 12. Different length ratios require different constraints; precision comparisons among all kinds of different constraints are shown in Figs. 11 and 12.

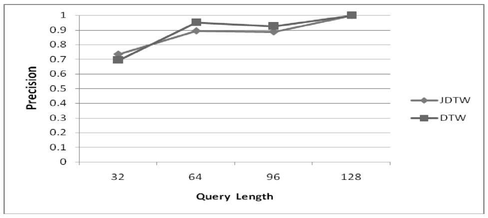


Figure 11. Precision comparision with non-noise query sequence

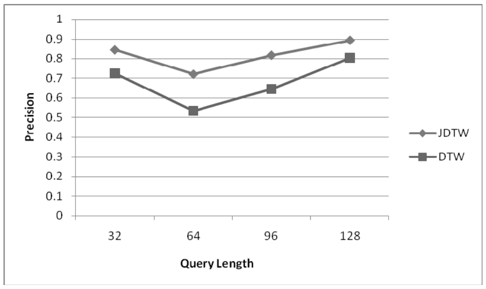


Figure 12. Precision comparision with noise query sequence

As seen in Fig. 11, JDTW and DTW achieve similar precisions with non-noise query sequence. However, it can be clearly seen that JDTW has an advantage over DTW in the noise condition, as shown in Fig. 12, since JDTW can reduce the noise impact. In summary, the novel approach improved the efficiency with lower omission factor and reduced the noise impact of query sequence.

V. CONCLUSIONS

This study was aimed at presenting a similarity measure of jumping dynamic time warping. The advantage of the new measure, of being able to jump over noise points, was proven. It is thus evident from the results of the experiments that the measure is effective in the noise condition. However, to make the technique more effective for the real world data, the constraints optimization algorithm must be further developed.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of P.R. China (Grant No. 60773213) and Natural Science Foundation of Liaoning Province, P.R. China (Grant No. 20071092).

REFERENCES

[1] G. J. Mao, L. J. Duan, S. Wang, Data Mining and Algrithm. Tsinghua University Press, 2007.

[2] Y. S. Moon, K. Y. Whang, and W. K. Loh, “Efficient time-series subsequence matching using duality in constructing windows,” In Information Systems , vol. 26, No. 4, 2001, pp. 279–293.

[3] E. Keogh, K. Chakrabarti, M. Pazzani and Mehrotra, “Locally adaptive dimensionality reduction for indexing large time series databases,” In Proc of ACM SIGMOD Conference on Management of Data, pp. 151-162, May, 2001.

[4] D. Berndt and J. Clifford, “Using dynamic time warping to find patterns in time series,” In AAAI-94 Workshop on Knowledge Discovery in Databases, pp. 229-248, 1994.

[5] E. Keogh. “Exact Indexing of Dynamic Time Warping,” In Proceedings of VLDB, pp.406-417, August 2002.[Proceedings of the 28th VLDB Conference].

[6] Y. Sakurai, M. Yoshikawa and C. Faloutsos, “FTW: Fast similarity search under the time warping distance,” In F. Afrati, pp. 326-337, 2005. [Proceedings of the 24th ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (PODS'05)].

[7] G. Das, K. Lin, H. Mannila, G. Renganathan and P. Smyth, “Rule discovery form time series,” AAAI Press, pp. 16–22, 1998.[Proc. of the 4th International Conference of Knowledge Discovery and Data Mining.]

[8] A. Debregeas and G. Hebrail, “Interactive interpretation of Kohonen maps applied to curves,” In Proc. of the 4th International Conference of Knowledge Discovery and Data Mining, pp. 179-183, 1998.

[9] J. J. R. Diez and C. A. Gonzalez, “Applying boosting to similarity literals for time series Classification,” Multiple Classifier Systems, pp. 210-219, 2000.[1st Inter’ Workshop]

[10] R. Jang, “Audio Signal Processing and Recognition”, http://www.cs.nthu.edu.tw/~jang.

[11] M. W. Kadous, “Learning comprehensible descriptions of multivariate time series,” In Proc. of the 16th International Machine Learning ,1999.