2015 14th International Symposium on Distributed Computing and Applications for Business Engineering and Science

Improvement of Dynamic Time Warping (DTW) Algorithm

Yuansheng Lou
College of Computer and Information
Ho Hai University
Nanjing,China
wise.lou@163.com

Huanhuan Ao*, Yuchao Dong
College of Computer and Information
Ho Hai University
Nanjing,China
*Corresponding author: aohuanhuanhhu@163.com

Abstract: **Dynamic time warping algorithm is widely used in similar search of time series. However, large scales of route search in existing algorithms resulting in low operational efficiency. In order to improve the efficiency of dynamic time warping algorithm, this paper put forward an improved algorithm, which plans out a three rectangular area in the search area of the existing algorithms, search path won't arrive the points outside the rectangular area, thus further reduced the search range of regular path. To some extent, this algorithm reduces computation of original algorithm, thus improving the operational efficiency. The improvement is more pronounced when two time series are longer.**

Keywords: dynamic time warping; time series; similar search; warping path; Euclidean distance

I. INTRODUCTION

In the early 1970s, Japanese scholars Itakura put forward dynamic time warping algorithm (DTW), introducing the concept of dynamic programming into the hard problem of recognition of the talk with isolated words in an uneven speed, thus significantly improving the efficiency of the recognition of isolated words [1]. With the spring up of time series data mining, the concept of DTW has been introduced in similar search field of time series by scholars at home and abroad [2]. After large amount of experiments, great achievements are made. The idea of DTW algorithm is a procedure using dynamic programming techniques to solve the optimization problem, during which a complex global optimization problem is broken into several local optimization problems, then decisions are made step by step [3], finally an optimal solution of the global problem is gotten. The procedure of time series similarity matching makes use of local optimization to find a path, along which he cumulative bending distance between two time series is minimum. But during the measurement of distance between two time series with DTW, it is vulnerable to be interfered by "noise" and "outlier" [4] of time series, since DTW algorithm matches point by point. And when the time series are too long, the calculation is very large. Therefore a suitable constraints and an appropriate matching range is the key to improve DTW matching accuracy and to shorten time of matching.

Based on analyzing existing DTW algorithm [5], this paper plans out three rectangles area outside the parallelogram path search range. The lattice outside the rectangular does not appear in the regular path; therefore there is no need to make a judgment that whether the lattice is within the area of parallelogram, which in turn decrease the calculation and increase the operation efficiency in some degree.

II. DTW ALGORITHM

A. Dynamic Time Warping Distance

Definition 1: Assume there are two certain length of time series R and T, of m and n. And $R = (r_1, r_2, r_3, \dots, r_m)$, $T = (t_1, t_2, t_3, \dots, t_n)$. In order to align these two sequences in time, firstly rectangle D with n row and m column need to be created [6]. Each matrix element $D_{ij} = d(r_i, t_j)$ represents the distance between the point r_i of series R and the point t_j of series T. This distance is usually Euclidean distance $d(r_i, t_j) = (r_i - t_j)^2; (1 \le i \le m, 1 \le j \le n)$. The shorter distance indicates the more similar between points, on the contrary, the longer distance indicates the less similarity [7].

$$D = \begin{bmatrix} d(r1, tn) & \cdots & d(rm, tn) \\ d(r1, tn-1) & \cdots & d(rm, tn-1) \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ d(r1, t1) & \cdots & d(rm, t1) \end{bmatrix}. \quad (1)$$

In the distance matrix D, solve a successive collection of matrix element with dynamic programming algorithm. The connection of each element is called regular path W, $W = w_1, w_2, w_3, \dots, w_k, \dots, w_K$. The k-th element of W $w_k = (i, j)_k$ is the alignment of i-th point of series R and j-th point of series T. The distance of regular path W is the warping distance of dynamic time. Regular path is not chosen at random, it must meet the following constraints [8-9]:

a) ***Boundedness****: The length of regular path W should within this range* $max(m, n) \le K \le m + n - 1$;

b) ***Boundary conditions****: The starting point of regular path W is* $w_1 = (1, 1)$, *and the end point is* $w_K = (m, n)$;

c) ***Continuity:*** *Suppose the previous point* $w_{k-1} = (i', j')$, *next point* $w_k = (i, j)$ *in regular path, then there must be* $(i - i') \le 1$ *and* $(j - j') \le 1$;

978-1-4673-6593-2/15 $31.00 © 2015 IEEE
DOI 10.1109/DCABES.2015.103

d) ***Monotonic:*** *Suppose the previous point* $w_{k-1} = (i', j')$*, next point* $w_k = (i, j)$ *in regular path, then there must be* $0 \le (i - i')$ *and* $0 \le (j - j')$*;*

Combining the continuity and monotonic of these constraints, if regular path W passes through the point (i, j), then the previous point must be one of point $(i-1, j)$, $(i-1, j-1)$ and $(i, j-1)$. The number of regular path which meets the above constraints can be exponential, but the one we need is the one with least cost to regular following formula:

$$\mathrm{DTW}(R,T) = \min\left\{\sqrt{\sum_{k=1}^{K} w_k}/K\right\} \quad (2)$$

B. Classic DTW Algorithm

In order to avoid unnecessary path search during DTW algorithm, the actual search path is limited at a parallelogram of 1/2 to 2 slopes in literature [10]. As shown below, m and n are two time series in parallelogram OABC. As $k_{OA} = 2$, $k_{OC} = 1/2$ are known, the functional relation of four sides and the coordinates of four points can be determined as follows: $y_{OA} = 2x$, $y_{OC} = 0.5x$, $y_{AB} = 0.5x + m - 0.5n$, $y_{CB} = 2x + m - 2n$, point $O(0,0)$, point $A(\frac{2m-n}{3}, \frac{2(2m-n)}{3})$, point $B(n,m)$ and point $C\left(\frac{2(2n-m)}{3}, \frac{(2n-m)}{3}\right)$. x_a ($x_a = \frac{2m-n}{3}$) and x_c ($x_c = \frac{2(2n-m)}{3}$) are two nearest integer point, therefore the constraints for the length of m and n are $2m - n \ge 3$ and $2n - m \ge 2$. If these two constraints cannot be satisfied, then these two series cannot be dynamic warped due to the big difference.

The computation of distance matrix D is very large. When the search path is limited at parallelogram OABC, there is no need to calculate the matching distance of the lattice points outside OABC. Therefore the computation is cut down largely [11-12].

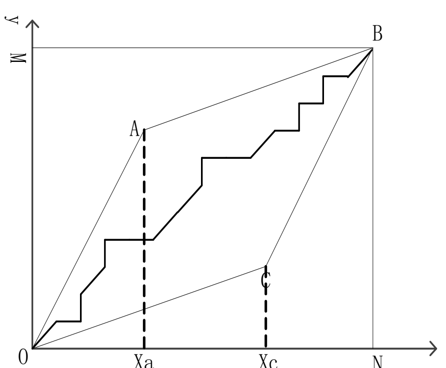

Figure1: Search Route Constraints Of Classic DTW Algorithm

III. Algorithm Improvement

A. Problems of Current Algorithms

In current DTW algorithm, distance matrix between two time series and accumulated distance matrix are need to be calculated, so the calculation is very large originally. Even if the search path is limited in a parallelogram, upper and lower boundary calculation is needed at each link of path search. The calculation is still very large, especially when two time series is very long, then this kind of repeated calculation is even more and the calculating efficiency is significantly reduced.

B. Principle of Algorithm Improvement

In order to avoid n*m times calculation of points' coordinates, firstly it needs to quickly determine which points are within OABC. Plan three rectangular areas Ω_1, Ω_2, and Ω_3 in figure 2. In the analysis, Ω_1 is determined by point $O(0,0)$ and point $A'\left(\left\lceil\frac{2m-n}{3}\right\rceil, \left\lceil\frac{2(2m-n)}{3}\right\rceil\right)$ which is the nearest integer point at top of A. Vertical line A'E intersects OC at point E' .Point $E\left(\left\lceil\frac{2m-n}{3}\right\rceil, \left\lfloor\frac{2m-n}{6}\right\rfloor\right)$ is the nearest integer point below point E'. Point $C'\left(\left\lceil\frac{2(2n-m)}{3}\right\rceil, \left\lfloor\frac{2n-m}{3}\right\rfloor\right)$ is the nearest integer at right of point C. The vertical line through point C' intersects AB at point D'. Point $D\left(\left\lceil\frac{2(2n-m)}{3}\right\rceil, \left\lceil\frac{4m+n}{6}\right\rceil\right)$ is the nearest integer point above point D'. Point D and point E can determine Ω_2. Point $C'(\left\lceil\frac{2(2n-m)}{3}\right\rceil, \left\lfloor\frac{2n-m}{3}\right\rfloor$ and point $B(n,m)$ can determine Ω_3. Firstly, the set of points in parallelogram OABC the set of points within $\Omega_1+\Omega_2+\Omega_3$ the set of points within rectangular OMBN, so the points outside $\Omega_1+\Omega_2+\Omega_3$ are not involved in the calculation of slope range. Most of points which are not on regular path are filtered. In other words, to locate points in OABC accurately means to find the set of points in $\Omega_1+\Omega_2+\Omega_3$ which meet the constraints as follows:

$$\begin{cases} y - 2x \le 0 \\ y - 0.5x \ge 0 \\ y - 0.5x - m + 0.5n \le 0 \\ y - 2x - m + 2n \ge 0 \end{cases} \quad (3)$$

Since not all the set of points in OMBN are involved in the calculation of slope range， the calculation is reduced largely that in turn improve the calculation efficiency. Especially when time series are longer, the improvement of calculating speed is even more evident.

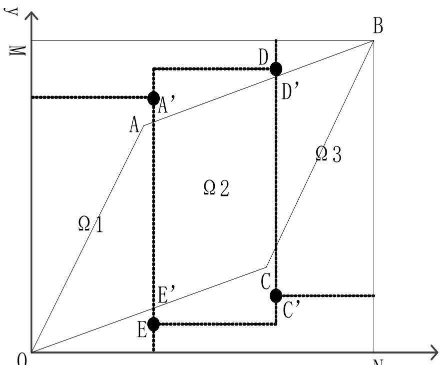

Figure2: Constraints of Matching Route

C. The realization of improved algorithm

Take time series $R = (r_1, r_2, r_3, \dots, r_m)$ and $T = (t_1, t_2, t_3, \dots, t_n)$ as input, improved DTW algorithm is realized as follows table I:

TABLE I. DTW(R, T) ALGORITHM

```
DTW(R, T) algorithm:
m = R.length; n = T.length
d, D = new[n × m]
warpingDistance = 0.0
for i = 0; i < ⌈(2m − n)/3⌉; i++
        for j = 0; j < ⌈2(2m − n)/3)⌉; j++
            if (i, j) in OABC
                d[i][j] = (R[i] – T[j])^2
for i = ⌈(2m − n)/3⌉; i < ⌈2(2n − m)/3)⌉; i++
        for j = ⌊(2m − n)/6⌋; j < ⌈(4m − n)/6⌉; j++
            if (i, j) in OABC
                d[i][j] = (R[i] – T[j])^2
for i = ⌈2(2n − m)/3)⌉; i < n; i++
        for j = ⌊(2n − m)/3⌋; j < m; j++
            if (i, j) in OABC
                d[i][j] = (R[i] – T[j])^2
for i = 0; i < n; i++
        for j = 0; j < m; j++
            if d[i][j] == 0.0
                continue;
            if i == 0 && j == 0
                D[i][j] = d[i][j];
            elseif i == 0 && j != 0
                D[i][0] = d[i][0] + D[i - 1][0];
            elseif j == 0 && i != 0
                D[0][j] = d[0][j] + D[0][j - 1];
            else
                warpingDistance = min(D[i - 1][j] == 0.0 ? : +∞, D[i - 1][j - 1]) == 0.0 ? : +∞, D[i][j - 1]) == 0.0 ? : +∞)
                warpingDistance += d[i][j];
        D[i][j] = warpingDistance;
return warpingDistance
```

In the improved algorithm above, d[n][m], D[n][m] are respectively distance matrix and cumulative distance matrix. The lattice point outside the three rectangular in the figure 2 are not be calculated, and algorithm eventually outputs warping distance which is the distance of regular path, that is the shortest warping distance required by DTW algorithm.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental environment: Eclipse + java, win7;

System parameters: P320 AMD Athlon、2G memory;

Experimental data: In this experiment, the data is provided by Professor Eamonn Keogh of University of California. The website address: http: //www.cs.ucr.edu/~eamonn/time_series_data/. Select 15.dat data from the folder mariohWords as the experimental data.

Experimental procedure: There are 1861 pairs of data in sample data set. In order to compare efficiency of arithmetic of the conventional DTW algorithm with the improved algorithm with different lengths of time series, select samples as follows, sample 1: m = 100, n = 120; sample 2: m = 500, n = 540; sample 3: m = 1000, n = 1500; sample 4: m = 1500, n = 1800.Every two time series were tested five times and calculate average test results in both arithmetic. Results are shown in table II:

TABLE II. COMPARISON OF OPERATIONAL TIME IN TWO ALGORITHMS

Length of time series	Traditional DTW (ms)	Improved DTW (ms)	Improvement rate (%)
100×150	15.6	15.2	2.56
500×550	31.2	29.8	4.49
1000×1500	143.2	135.3	5.38
1500×1800	188.4	177.6	5.73

By table II: When the length of time series is 100 × 150, the improved algorithm saves 2.56% time than the traditional method. When the length is 500 × 550, the improved algorithm saves 4.49%. When the length is 1000 × 1500, the percentage is 5.38%. When the length is 1500 × 1800, the improvement rate is 5.73%. Results prove that improved algorithm proposed in this paper, to a certain extent, reduces the amount of computation and improve the operational efficiency and the improvement is even more obvious when the time series are long.

ACKNOWLEDGMENT

The research of this paper is partially supported by the National Key Technology Research and Development Program of the Ministry of Science and Technology of China under Grant No. 2013BAB06B04, No. 2013BAB05B00 and No. 2013BAB05B01.

REFERENCES

[1] Yuxin Z, Miyanaga Y. An improved dynamic time warping algorithm employing nonlinear median filtering[C]//Communications and Information Technologies (ISCIT), 2011 11th International Symposium on. IEEE, 2011: 439-442.

[2] Berndt D J, Clifford J. Using Dynamic Time Warping to Find Patterns in Time Series[C]//KDD workshop. 1994, 10(16): 359-370.

[3] Xu L, Ke M. Research on isolated word recognition with DTW-based[C]//Computer Science & Education (ICCSE), 2012 7th International Conference on. IEEE, 2012: 139-141.

[4] Wang Y, Lei P, Zhou H, et al. Using DTW to measure trajectory distance in grid space[C]//Information Science and Technology (ICIST), 2014 4th IEEE International Conference on. IEEE, 2014: 152-155.

[5] Islam S, Hannan M A, Basri H, et al. Solid waste bin detection and classification using Dynamic Time Warping and MLP classifier[J]. Waste Management, 2014, 34(2):págs. 281-290.

[6] Rakthanmanon T, Campana B, Mueen A, et al. Addressing Big Data Time Series: Mining Trillions of Time Series Subsequences Under Dynamic Time Warping[J]. Acm Transactions on Knowledge Discovery from Data, 2013, 7(3):965-991.

[7] Cleuziou G, Moreno J G. Kernel methods for point symmetry-based clustering[J]. Pattern Recognition, 2015, 48:2812–2830.

[8] Keogh E J, Pazzani M J. Derivative Dynamic Time Warping[C]//SDM. 2001, 1: 5-7.

[9] Tan L N, Alwan A, Kossan G, et al. Dynamic time warping and sparse representation classification for birdsong phrase classification using limited training dataa)[J]. The Journal of the Acoustical Society of America, 2015, 137(3): 1069-1080.

[10] Ratanamahatana C A, Keogh E. Making time-series classification more accurate using learned constraints[C]. SDM, 2004.

[11] Li C, Ma Z, Yao L, et al. Improvements on EMG-based handwriting recognition with DTW algorithm.[C]// Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference. 2013:2144-2147.

[12] Adwan S, Arof H. On improving Dynamic Time Warping for pattern matching[J]. Measurement, 2012, 45(6):1609–1620.