2017 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2017, Cancun, Mexico, December 4-7, 2017

Time series Classification with New Similarity Measure : An Application for Automatic Detection of Driver's Distraction

Basabi Chakraborty† and Sho yoshida‡

†Faculty of Software and Information Science, Iwate Prefectural University
152-52 Sugo, Takizawa, Iwate, Japan
‡Graduate School of Software and Information Science
152-52 Sugo, Takizawa, Iwate, Japan
Email: basabi@iwate-pu.ac.jp, g231n033@s.iwate-pu.ac.jp

Abstract—Classification or grouping of time series data is now increasingly needed to solve various real life problems. As time series data is huge, a proper representation method and an efficient similarity measure are important factors for the success of any clustering or classification method involving time series data. Though a lot of research has already been done in this line, dynamic time warping (DTW) seems to be the most common method used for measuring similarity of two time series data. Though classification accuracy of time series classification with DTW is quite satisfactory, computational cost is also very high. In this work, newly proposed measures by the authors have been used for time series classification problem. Publicly available benchmark data sets as well as time series data from a real life problem of detecting driver's distraction with cognitive load are used for classification with the proposed measures. The comparative effectiveness of the proposed measures over DTW has been examined by the experimental results.

1. Introduction

Time series data is defined as the data which varies with time and is collected over a period of time. With the increasing use of different sensors to monitor and analyse the dynamic behavior of a system, huge time series data is generated and their efficient analysis is becoming necessary for meaningful use of the data [1]. Various approaches of time series classification have been developed so far ranging from neural and bayesian networks to genetic algorithms, support vector machines and characteristic pattern extraction[2]. Traditional classification techniques like bayesian classifier or decision tree are modified for time series data and temporal naive bayesian model (T-NB) and temporal decision tree (T-DT) are developed [3]. In [4] time series data is transformed to a lower dimensional compact representation by extracting characteristic features to facilitate the use of classical machine learning algorithms for classification.

Now for any classification task, computation of pair-wise similarity measure for any two time series is essential. For static data, several distance metric are available which cannot be used directly for temporal data. The most popular method to compare two time series is warping the time axis in order to achieve an alignment between the data points of the series. The Dynamic Time Warping (DTW) algorithm, first being used in speech recognition, has been shown to be an effective solution for measuring the distance between time series [5]. Unlike euclidean distance which is easier to compute, DTW allows a time series to be stretched or compressed to provide a better match with another time series and can handle time series with local time shifting and different lengths. Despite the effectiveness of DTW algorithm, it has a computational cost of $O(N^2)$ which makes it computationally difficult to use for longer time series. Several measures have been introduced to speed up DTW computations as well as to better control the possible routes of the warping path [6], [7].

A new similarity measure Cross Translation Error (CTE) based on multidimensional delay vector representation of time series has been proposed by author in [8]. This measure, though computationally efficient, was found to have poor recognition accuracy compared to dynamic time warping. In [9], Dynamic Translation Error (DTE), a modification of CTE by integration with DTW has been proposed. The classification accuracy of time series benchmark data set with this measure is found to be higher compared to DTW, but the computational cost is also higher than CTE. To reduce the computational cost of DTW, another new measure DTW-GA has been proposed in this work.

In this work the efficiency of the proposed measures over DTW and its other variants are examined with benchmark data sets and are also applied to a real life problem of detecting driver's distraction with cognitive load. The next section presents a brief description of popular DTW measure and proposed similarity measures for time series classification. The following section introduces the classification problem for automatic detection of driver's distraction from the time series data of driving behaviour collected from driving simulator. The next section describes the classification experiments and results. The last section contains conclusion.

2. Similarity measures for time series classification

In this section DTW , its variant fastDTW and proposed similarity measures are presented in brief.

2.1. Dynamic Time Warping (DTW) distance measure

DTW belongs to the group of elastic measures and works by optimally aligning the time series in temporal domain so that the accumulated cost of the alignment is minimal.The accumulated cost can be calculated by dynamic programming, recursively applying

$$D_{i,j} = f(x_i, y_j) + \min(D_{i,j-1}, D_{i-1,j}, D_{i-1,j-1}) \quad (1)$$

for $i = 1 \ldots M$ and $j = 1 \ldots N$ where M and N are the length of the time series x and y respectively and $f(x_i, y_j) = \sqrt{(x_i - y_j)^2}$.

FastDTW, introduced in [7], is an approximation of DTW that has a linear time and space complexity. FastDTW uses a multilevel approach which recursively produces a solution from a coarse one and refines it.

2.2. Proposed Similarity Measures

Two proposed similarity measures by authors involving DTW are presented here.

2.2.1. Dynamic Translation Error DTE

This similarity measure is based on time series representation by delay coordinate embedding like earlier proposed measure CTE [8]. According to Taken's theorem [10] , a deterministic time series signal $\{s_n(t)\}_{t=1}^{T_n}(n = 1, 2, \ldots, N)$ can be embedded as a sequence of time delay co-ordinate vector $v_{s_n}(t)$ known as experimental attractor, with an appropriate choice of embedding dimension m and delay time τ for reconstruction of the original dynamical system as follows:

$$v_{s_n}(t) \equiv \{s_n(t), s_n(t+\tau), \ldots, s_n(t+(m-1)\tau)\}, \quad (2)$$

Now for correct reconstruction of the attractor, a fine estimation of embedding parameters (m and τ) is needed. There are variety of heuristic techniques for estimating those parameters [11]. In CTE, distance between two time series is calculated as the average distance of two nearest vectors from two time series respectively. Dynamic Translational Error is a combination of CTE and DTW. Here distance calculation of DTW is done according to the strategy of calculation of CTE. The measure considers the two time series represented by multidimensional delay vectors and aligns them in the phase space considering the nearest vectors so that the accumulated cost is minimum.

The algorithm in brief is as follows:

1. The time series is to be converted to multidimensional delay vector form

$$v_{s_n}(t) \equiv \{s_n(t), s_n(t+\tau), \ldots, s_n(t+(m-1)\tau)\}, \quad (3)$$

2. Calculate the similarity matrix as in DTW, but here $f(x_i, y_j)$ of Equation 1 is CTE

   $D_{i,j} = f(x_i, y_j) + \min(D_{i,j-1}, D_{i-1,j}, D_{i-1,j-1})$

$$f(x_i, y_j) = \frac{1}{2}\left(\frac{|v_x(i) - \bar{v}|}{|\bar{v}|} + \frac{|v_y(j) - \bar{v}|}{|\bar{v}|}\right) \quad (4)$$

   where $\bar{v}$ denotes average vector. $D_{0,0} = 0$

3. $D_{M,N}$ is the distance between time series x and y

   $D_{dte}(x, y) = D_{M,N}$

   where M and N are length of time series x and y.

2.2.2. GA masked Dynamic Time Warping DTW-GA

Another new similarity measure is proposed in this work with the objective of decreasing computational cost of DTW without much decrease in classification accuracy. DTW-GA is a combination of dynamic time warping (DTW) with genetic algorithm (GA). DTW-GA use masking of time series with the optimum gene of GA as the representation method of time series and DTW as the comparison method of two time series. The main concept behind this new measure is that GA searches the best(optimum) gene for masking so that the most important points of the time series is used for comparison instead of the whole series to reduce the computational cost.

The time series is sliced into portions by the length of the optimum gene with the slicing window. The time series points corresponding to 1 is used for comparison while the points corresponding to zero are discarded. The distance between two time series are computed and average is taken as the distance between two time series. As the most important points of the time series (which can be called feature pattern of the time series) is different for different classes, the optimum gene code is different for different classes. So at the first step GA is used to find out the optimum gene code of each class from the training samples of the class. Minimum intra class distance is used as the fitness function of the genetic algorithm. So the gene code is considered optimum for the specific class for which the average of the pairwise distance of the training samples is minimum. In the second step 1NN classifier is used to find out the class of the unknown test time series using class specific gene codes of each class for comparison of the unknown time series and training time series of the class.

The details of the process is as follows:

1. Generation of initial solution group

   - Generate N searching genes having cells stored 0 or 1 randomly

- Generate C comparison genes stored 0 where C is number of class
- Initialize evaluation value of comparison genes to infinity

2. Fitness evaluation

```
for each Searching genes n
   for each Comparison genes c
     for M (parameter for stabilization)
  get two c class time series from
    training data randomly
   compare two time series using
       gene n
    if distance is less than d_c, store
         distance in d_c
  if d_c is lower than evaluation value of
   class c,comparison gene and evaluation
    value of class c are overwritten by gene
      n and d_c
store minimum value d_1 to d_ c in
   evaluation value of gene n
```

3. Genetic operation (selection , crossover and mutation as in ordinary GA)
4. Repeat step 2 and 3 until stopping criterion is fulfilled

3. Detection of driving distraction from driving data

Distracted driving is the cause of most of the road accidents and research on automatic detection of distracted driving from sensor data attached to car is gaining increasing attention. Among various distractions, two main types are visual and cognitive distraction. Visual distraction also known as eye-off the road can be detected by tracking driver's eye movement. Detecting cognitive distraction, also known as mind-off-the road, is more complex. We restricted our study to the area of cognitive distraction. In our study [12] the driving behaviour is assessed from driving simulator output which contains time series data (steering angle, steering torque, accelerator stroke, brake stroke, car speed, car angle, engine speed etc). The experimental study in brief is as follows:

1. For 4 subjects, driving data for three situations have been collected for a) normal driving with attention b) driving while continuing conversation with co passenger c) driving while doing mental arithmetic at the elementary school level, such as simple addition, subtraction and multiplication.
2. For each situation, different driving scenarios are used for example, simple route, route having curves and sharp bending and routes with multiple diversions.

Table 1: Classification accuracy with different measures

Data Name	Classification accuracy with				
	Euclid	DTW	Fast DTW	DTE	DTW GA
50Words	0.668	**0.709**	0.614	0.681	0.681
Coffee	0.75	0.82	0.963	**0.964**	0.786
FaceAll	0.721	0.772	0.772	**0.747**	0.722
Haptics	0.36	0.36	0.367	0.37	**0.399**
InlineSkate	0.353	0.369	**0.376**	0.320	0.365
ItalyPD	0.962	0.947	0.943	0.945	**0.967**
Mallat	**0.924**	0.915	0.917	0.91	0.91
StarLC	0.853	0.886	0.850	0.890	**0.915**
TwoPatterns	0.961	**1.00**	0.947	0.836	0.933
uWGL-y	0.666	0.645	0.665	0.589	**0.689**
wafer	0.995	0.983	0.953	**0.999**	0.995

Table 2: Average accuracy with different measures

Euclid	DTW	FastDTW	DTE	DTW-GA
0.76	0.79	0.73	0.79	0.77

3. Each subject is asked to drive following a car speeding 60km per hour with a more or less constant separation in the designated routes (from simple to complex) consecutively and repeat driving for 5 times with driving duration of 3min.
4. The time series output data from the driving simulator for steering wheel angle, steering torque, accelerator torque, brake stroke, car speed and engine speed have been recorded.

We selected 150 driving samples for each person, normal driving 60 samples, driving with conversation 45 samples and driving with mental arithmatic 45 samples. For each sample, 6 time series (steering wheel angle SA, steering torque ST, accelerator torque AT, brake stroke BS, car speed CS and engine speed ES) for 3 minutes are obtained. The data is first preprocessed by using moving avarage filter and then normalized. We have used several methodologies to find out the best feature subset for individual driver in order to classify normal driving data from driving data with cognitive loads and results are reported [12], [13]. Here in this work, we use the time series data for classification into two classses (normal and with cognitive load) using newly proposed similarity measures.

4. Simulation Experiments

Simulation experiments for judging the efficiency of the two proposed measures over popular measures have been done with bench mark data sets and the collected data from driving simulator.

Table 3: Average accuracy for driving data with different measures

Euclid	DTW	FastDTW	DTE	DTW-GA
0.69	0.69	0.67	0.73	0.70

4.1. Bench mark data set

The benchmark data sets consisting of 43 different time series data from University of California, Riverside (UCR) time series repository [14] are used for the simulation experiments. The training data is used as labeled data for 1NN classifier and classification accuracy is calculated on the test set. The average classification accuracy for twenty trials on different partitions of training and test data are noted for all the data sets.

4.1.1. Results for benchmark data

Table 1 represents the classification results for 11 data sets with several similarity measures, simple Euclid, popular DTW, its variant fast DTW, priviously proposed DTE and the newly proposed DTW-GA. Table 2 represents the average classification results for 43 data sets. It is seen from the table that no measure is the best for all the data sets. However DTE and DTW-GA seems to be fairly well compared to DTW in producing the best classification accuracy. The computational cost of DTW-GA is on the average 7 times lower than DTW, the effect is considerable for longer time series.

4.1.2. Results for driving simulator data

Table 3 represents the average classification results of 1NN classifier for classifying normal driving and driving with cognitive loads using simulator data for six time series individually. 5 similarity measures are used for classification. It is found by statistical testing that there is no significant difference in the classification result for any two measures. However , on average, DTE produces the best accuracy but computational cost is high. Computational costs of DTW-GA and FastDTW are of same order and much less than DTW while DTW-GA produces better classification accuracy.

5. Conclusion

In this work, a new similarity measure for time series classification has been proposed and its efficiency in terms of recognition accuracy and computational cost compared to other similarity measures has been evaluated with bench mark data sets. Finally the proposed measure is also used to classify collected data from driving simulator to automatically classify normal driving with distracted driving. The proposed measure seems to be efficient in time series classification problem.

References

[1] H. Kantz, T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, UK, 2004.

[2] K. Buza, A. Nanopoulos, L. Schmidt-Thieme, Time series classification based on individual error prediction. *IEEE International Conference on Computational Science and Engineering*, HongKong, Dec,2010.

[3] B. Chakraborty.: Feature selection and classification techniques for multivariate time series. *Proc. of ICICIC 2007*. September 2007.

[4] Zhang, X., Wu, J., et.al.: A novel pattern extraction method for time series classification. *Optimiz.Engin.* 10. No.2, pp. 253–271, 2009.

[5] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series",*Proc. AAAI Workshop on Knowledge Discovery in Databases*, pp. 359–370, 1994.

[6] S. Chu, E. Keogh, D. Hart and M. Pazzani, "Iterative deepening dynamic time warping for Time series", *Proc. 2nd SIAM International Conference on Data Mining*, 2002.

[7] S. Salvador and P. Chan, " FastDTW: Toward accurate dynamic time warping in linear time and space", *Intelligent Data Analysis.*, Vol.11(5), pp. 561–580, 2007.

[8] Chakraborty, B., Manabe, Y.: An Efficient Approach for Person Authentication using Online Signature Verification. *Proceedings of SKIMA 2008*. pp.12–17,2008

[9] Basabi Chakraborty and Sho Yoshida, "Proposal of a new similarity measure for time series classification", *Proc. ITISE 2016*, Spain, 2016.

[10] K. Alligood, et.al, *Chaos : An Introduction to Dynamical Systems*. Springer. New York,1997.

[11] H. D. I. Aberbanel, *Analysis of Observed Chaotic Data*. Springer. New York,1996.

[12] B. Chakraborty, Y. Manabe, S. Yoshida and k. Nakano, "Effect of Cognitive Distraction on Driving Behaviour, *Proc. of FITAT 2016*, Zuhai, China, March/April 2016.

[13] B. Chakraborty and K. Nakano, Automatic detection of drivers awareness with cognitive task from driving behavior, *Proc. of IEEE SMC 2016*, Budapest, Hungary, October 2016.

[14] Keogh, E., Zhu, Q., Hu, B., Hao. Y., Xi, X., Wei, L. & Ratanamahatana, C. A. (2011). The UCR Time Series Classification/Clustering Homepage: www.cs.ucr.edu/ eamonn/time_series_data/