A Proposal for Classification of Multisensor Time Series Data based on Time Delay Embedding

Basabi Chakraborty
Faculty of Software and Information Science
Iwate Prefectural University
152-52 Sugo, Takizawamura, Iwate 020-0193, Japan
basabi@iwate-pu.ac.jp

Abstract—**Multisensor time series data is common in many applications of process industry, medical and health care, biometrics etc.Analysis of multisensor time series data requires analysis of multidimensional time series(MTS) which is challenging as they constitute a huge volume of data of dynamic nature. Traditional machine learning algorithms for classification and clustering developed for static data can not be applied directly to MTS data. Various techniques have been developed to represent MTS data in a suitable manner for analysis by popular machine learning algorithms. Though a plethora of different approaches have been developed so far, 1NN classifier based on dynamic time warping (DTW) has been found to be the most popular due to its simplicity. In this work, an approach for time series classification is proposed based on multidimensional delay vector representation of time series. Multivariate time series is considered here as a group of single time series and each time series is processed separately to be represented by a multidimensional delay vector (MDV). A simple simulation experiment with online handwritten signature data has been done with a similarity measure based on the MDV representation and classification performance is compared with DTW based classifier. The simulation results show that classification accuracy of the proposed approach is satisfactory while computational cost is lower than DTW method.**

Index Terms—**Time series data, multisensor data, delay embedding,cross translational error, classification**

I. Introduction

In many areas such as process industry, medical and health care, behavioural biometrics, varieties of sensor data are collected over a time period. These multisensor time series data need to be analysed for some decision or action. The multivariate time series classification technologies are needed for the mining of multisensor timeseries data. Though handling of multivariate time series (MTS) data is relatively new in the area of data mining, the study of multivariable time series has a long history in statistics, signal processing or control theory. An MTS data item is generally represented by a $m \times n$ matrix, where m is the number of observations in the time scale and n represents the number of features or the number of dimension of the observed variables. Due to the huge volume and the dynamic nature of MTS data, its analysis is a challenging task. Classical machine learning algorithms developed for static data are quite unsuitable for mining and classification of MTS data.

Due to increased interest and need for classification of MTS data, various approaches have been developed ranging from Neural and Bayesian networks to genetic algorithms, support vector machines and characteristic pattern extraction[1].Traditional classification techniques like Bayesian classifier or decision tree are modified for MTS data and temporal naive Bayesian model (T-NB) and temporal decision tree (T-DT) are developed [2]. In [3] MTS data is transformed to a lower dimensional compact representation by extracting characteristic features to facilitate the use of classical machine learning algorithms for classification. A brief review of the related research works on time series classification is presented in the next section. Now for any classification task, similarity measure for grouping the data or identifying the class is the most important.Euclidean distance is widely used as the simple similarity (dissimilarity) measure. Dynamic time Warping (DTW) and its various variants are considered the most successful similarity measure. Among other measures, Longest Common Subsequence (LCSS) and edit distance are quite popular [4].

In this work, an algorithm for multivariate time series classification is proposed which is based on multidimensional delay vector representation of time series. A similarity measure for measuring similarity of two time series proposed in [5] is extended and used for proposed classification algorithm. Section 3 represents the proposed algorithm followed by simulation experiments and results in the following section. The final section represents discussion and conclusion.

II. Related Research Work on MTS Classification

Existing approaches for time series classification can be broadly classified into three categories [6]:

- *Feature based classification* in which a multidimensional time series is transformed into a feature vector and then classified by traditional classification algorithms.The choice of appropriate features plays an important part in this approach. A number of techniques has been proposed for feature selection or feature subset selection by using compact representation of high dimensional MTS into one row to facilitate application of traditional feature selection algorithms like recursive feature elimination (RFE), zero norm optimization etc. [7] [2]. Time series shapelets, characteristic subsequences of the original series, is recently

978-1-4799-0652-9/13/$31.00 ©2013 IEEE

proposed as the features for time series classification [8]. Another group of techniques extract features from the original time series by using various transformation techniques like Fourier, Wavelet etc. In [9], a family of techniques have been introduced to perform unsupervised feature selection on MTS data based on common principal component analysis (CPCA), a generalization of PCA for multivariate data items where all the data items have the same number of dimensions.

- *Model based classification* in which a model is constructed from the data and the new data is classified according to the model that best fits it. Models used in time series classification are mainly statistical such as Gaussian, Poisson, Markov and Hidden Markov Model (HMM) or based on neural networks. Naive Bayes is the simplest model and is used in text classification [11]. Hidden Markov models (HMM) are successfully used for biological sequence classifications. Some neural network models such as recurrent neural network (RNN) are suitable for temporal data classification.
- *Distance based classification* in which a distance function which measures the similarity between two time series is used for classification. Similarity or dissimilarity measures are the most important component of this approach. Eucledean distance is the most widely used measure with 1NN classifier for time series classification. Though computationally simple, it requires two series to be of equal length and is sensitive to time distortion. Elastic similarity measures such as Dynamic Time Warping (DTW) and its variants overcome the above problems and seem to be the most successful similarity measure for time series classification in spite of high computational cost. However recent results strongly suggest that for large data sets, accuracy of DTW converges with Euclidean distance. For multivariate time series, correlation between pairs of time series also has to be taken into account in classification algorithm. In [10] a new distance measure has been developed to deal with MTS data.

III. A NEW APPROACH FOR MTS CLASSIFICATION

In the proposed approach, time series is represented by a multidimensional delay vector (MDV) by delay coordinate embedding which is a standard approach for analysis and modelling of nonlinear time series [12]. The similarity between two time series is measured by the proposed similarity measure *Cross Translation Error*(CTE) [13] based on MDV representation of time series. In the next two subsections a brief introduction on delay coordinate embedding and Cross Translation Error is represented.

A. Multidimensional Delay Vector

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 coordinate 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:

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

Figure 1 shows the concept of multidimensional delay vector with various m and τ. 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 [14]. The author proposed an approach for fine estimation of optimal embedding parameters in [15].

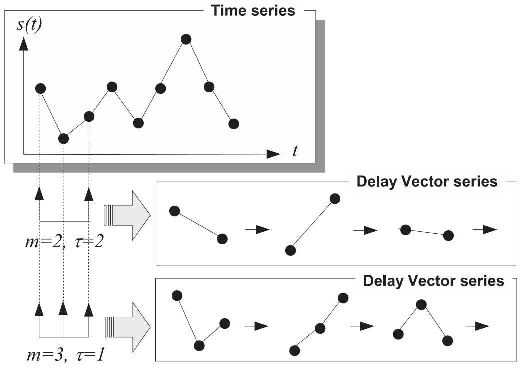

Fig. 1. Time Series and Multidimensional Delay Vector

B. Cross Translation Error

Cross Translation Error (CTE), based on Translation Error, originally proposed by [16] for detecting determinism in a time series, has been proposed in [13] for calculating similarity between two time series. The algorithm is described below, the details can be found in [13].

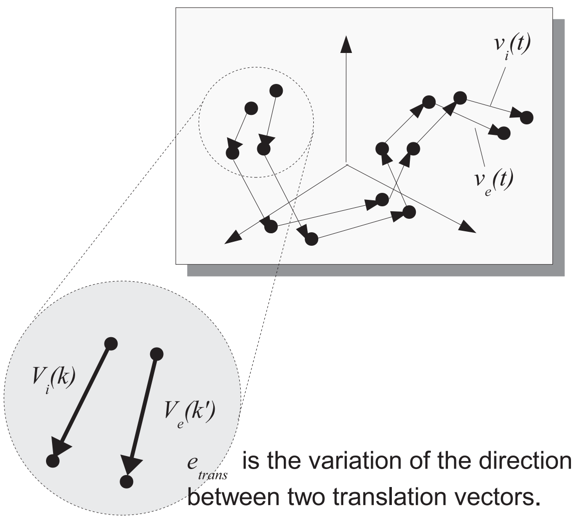

Fig. 2. Concept of Cross Translation Error

1) Multi-dimensional delay vector $v_{s_n}(t)$ can be generated from time series $\{s_n(t)\}_{t=1}^{T_n}(n = 1, 2, \ldots, N)$ based on Eq. (1). $(m+1)$ dimensional vector $v_{s_n}(t)'$ including the normalized time index t/T_n is defined as follows;

$$v_{s_n}(t)' \equiv \{s_n(t), s_n(t+\tau), \ldots, s_n(t+(m-1)\tau), t/T_n\}. \quad (2)$$

2) Let $v_{s_i}(t)$ and $v_{s_e}(t)$ denote m-dimensional delay vectors generated from time series $s_i(t)$ and time series $s_e(t)$ respectively. $v_{s_i}(t)'$ and $v_{s_e}(t)'$ denote the corresponding $(m + 1)$-dimensional vector including the normalized time index t/T_n.
3) A random vector $v_{s_i}(k)$ is picked up from $v_{s_i}(t)$. Let the nearest vector of $v_{s_i}(k)$ from $v_{s_e}(t)$ be $v_{s_e}(k')$. The index k' for the nearest vector is defined as follows;

$$k' \equiv \arg\min_t ||v_{s_i}(k)' - v_{s_e}(t)'|| \quad (3)$$

4) For the vectors $v_{s_i}(k)$ and $v_{s_e}(k')$, the transition in the each orbit after one step are calculated as follows;

$$V_{s_i}(k) = v_{s_i}(k+1) - v_{s_i}(k), \quad (4)$$

$$V_{s_e}(k') = v_{s_e}(k'+1) - v_{s_e}(k'). \quad (5)$$

5) Cross Translation Error (CTE) e_{cte} is calculated from $V_{s_i}(k)$ and $V_{s_e}(k')$ as

$$e_{cte} = \frac{1}{2}(\frac{|V_{s_i}(k) - \bar{V}|}{|\bar{V}|} + \frac{|V_{s_e}(k') - \bar{V}|}{|\bar{V}|}), \quad (6)$$

where $\bar{V}$ denotes average vector between $V_{s_i}(k)$ and $V_{s_e}(k')$. Figure 2 shows the concept of e_{trans}.
6) e_{cte} is calculated for L times for different selection of random vector $v_{s_i}(k)$ and the median of e^i_{cte} $(i = 1, 2, \ldots, L)$ is calculated as

$$M(e_{cte}) = Median(e^1_{cte}, \ldots, e^L_{cte}). \quad (7)$$

The final cross translation error E_{cte} is calculated by taking the average, repeating the procedure Q times to suppress the statistical error generated by random sampling in the step (3).

$$E_{cte} = \frac{1}{Q}\sum_{i=1}^{Q} M_i(e_{cte}). \quad (8)$$

Cross translation error is a distance metric, so lower value of E_{cte} represents higher similarity. For multivariable time series, each dimension is considered separately as a single time series and represented by a multidimensional vector. So MTS data can be represented as a set of multidimensional vectors, each element corresponding to a single variable time series.

C. Proposed Algorithm for Multivariate Time Series Classification using CTE

The similarity measure between two time series, calculated in the phase space described above, can be used in conjunction with popular machine learning classification algorithm for MTS data classification. The proposed measure E_{cte} in Eq (8), integrated with DTW, has been successfully applied in biometric authentication with online handwritten signature [17]. In this work, algorithms for supervised as well as unsupervised classification of MTS data has been proposed.

The algorithms are represented as follows:

- For *supervised classification* problems, where training data set is available, MTS data with class label is used for pairwise intraclass and inter class similarity calculation of two time series. As an example, for two class problems, similarity measure of all possible pairs of time series of class A and class B training samples are to be calculated separately. All possible pairs of time series (one from class A and the other from class B) are also to be calculated for inter class similarity of classes A and B
- The average intraclass similarity values of class A, CTE_A and class B, CTE_B are calculated from all possible pairwise similarity values of class A and B respectively. The average interclass similarity value for class pairs A and B, CTE_{AB}, is calculated from all possible interclass similarity values.
- The similarity values of the unlabelled time series and all time series from class A and B are to be calculated respectively and the averages are denoted as $CTE_{(Un,A)}$ and $CTE_{(Un,B)}$ respectively. The unlabelled sample will be labeled as class A if the average similarity value of the sample from all the class A samples matches CTE_A better than CTE_B and the average similarity value of the sample from all class B samples matches CTE_{AB} within a certain limit. In short,
  Case 1: For unlabelled sample to be in class A,
  $CTE_{(Un,A)} \leq CTE_A$ and $CTE_{(Un,B)} \approx CTE_{AB}$ within a certain limit ϵ.
  Case 2: For unlabelled sample to be in class B,
  $CTE_{(Un,B)} \leq CTE_B$ and $CTE_{(Un,A)} \approx CTE_{AB}$ within a certain limit ϵ.
  Some samples might be undecided if they do not satisfy either one of the above conditions.
- Simple K-NN classifier can also be used for classification of the unlabelled sample after calculating all the pairwise similarity values of the unlabelled samples with the labelled training samples.
- For *unsupervised classification* or *clustering* of MTS data, all the pairwise similarity values between two time series using proposed measure are to be calculated in the first stage and a similarity matrix can be formed.
- Now as the similarity values are real numbers, they can be clustered according to euclidean distance and a dendogram is generated. The appropriate number of clusters can be decided by examining the dendogram.
- Now other simple traditional clustering algorithms like k-means or self organizing map (SOM) can also be used to cluster the similarity values and by examining the cluster contents, the original time series classification can be achieved.
- The algorithm is explained above with two class examples, but is is very simple to extend to multiclass data.

IV. Simulation Experiments and Results

Simulation experiments have been done using benchmark data set of signature verification contest 2004, SVC 2004. The detail of data description is presented below:

A. Data Set

SVC 2004 data set contains two data sets; task1 and task2. Task1 data set includes mainly pen position (x and y) time series data. Task 2 data set includes pen position ($x(t)$ and $y(t)$), pen pressure($p(t)$), pen inclination (Azimuth and Altitude) time series data. We have used task2 data set for our experiment. So in our experiment we used 5 dimensional time series data set. Data acquisition is achieved by the digitizing pen tablet system (WACOM intuos) and data is sampled while pen touches the tablet (pen down condition). Total number of data in is 1600 (40 writers $\times$ 40 signatures/writer). 40 signatures are divided into two classes; genuine signature set (S1 to S20) and skilled forgery signature set (S21 to S40). The data has been preprocessed to eliminate the sampled point $s(t) = (x(t), y(t))$ as a redundant point when $s(t) = s(t+1)$ and to scale the data from 0.0 to 1.0. This data set is also used in authentication problem in our earlier paper [17].

B. Simulation Methods

In this work, the data set is used for classification purpose. The two classes are genuine and forged signature for a particular writer. The classification experiment has been done with the proposed similarity measure and DTW for comparison. The procedure in brief is described below.

1) For a particular writer, 10 genuine and 10 forged samples are randomly chosen from the 20 genuine and 20 forged samples respectively for training set and the rest are used for test set.
2) The classification algorithms described above with proposed similarity measure, cross translation error, has been performed and the recognition accuracy has been noted.
3) The experiment has been repeated 10 times for a particular writer with different random selection of training and test samples and the average recognition accuracy has been calculated.
4) The above procedure is repeated for 40 writers individually to classify the genuine and forged samples of individual writers.
5) In another experiment, K- means clustering procedure is used to group the sample signatures of individual writers in two groups genuine and forged.
6) The clustering experiment is repeated for all writers.
7) Finally the classification experiment is done by Dynamic Time Warping (DTW) algorithm for comparison of efficiency of the proposed measure.

C. Simulation Results

Table I shows the average recognition rate of Genuine class and Forgery class for all the 40 writers individually for supervised classification algorithm. From the results it seems that the classification accuracy for '"Genuine class" is better for CTE measure for most of the writers. Only for writer 35, 24, 9, DTW performs better. In the case of some writers like 33,32,29,25,22,21 and 19, the performance are more or less similar for both the measures. On average over all the writers, performance of proposed algorithm based on CTE seems to be better than DTW based classification. For "Forgery class" the classification rate is lower than the classification rate of "Genuine class" as "Forgery class" has higher variance. For the "Forgery class" also the same result as " Genuine class" is found. On average, CTE based classifier performs better than DTW based classifier

For representing the time series with MDV, the value of the embedding parameters "m" and "tau" are used from earlier work [17].The embedding parameters are different for different time series. Now it seems that the performance of CTE depends on proper values of embedding parameters but DTW

TABLE I
Classification Results

Writer	Recognition rate (%)			
	Class Genuine		Class Forgery	
	CTE	DTW	CTE	DTW
1	92.1	83.5	86.5	82.5
2	91.8	84.7	88.2	83.9
3	95.3	90.2	90.2	90.1
4	93.8	92.3	90.2	90.3
5	94.3	92.8	85.3	84.4
6	93.3	94.3	86.8	85.3
7	97.5	95.8	91.3	90.8
8	96.3	92.3	88.9	87.8
9	90.2	95.3	85.6	85.4
10	92.3	91.8	83.5	85.3
11	93.8	92.5	80.5	81.2
12	92.9	90.2	85.3	86.2
13	93.8	91.3	86.9	86.7
14	94.5	93.2	87.5	86.3
15	94.7	90.8	88.3	85.7
16	95.2	93.2	90.1	90.1
17	95.5	92.5	90.3	90.7
18	93.2	91.3	88.5	83.4
19	92.5	92.5	88.3	86.7
20	93.2	92.9	85.3	86.7
21	91.5	91.6	82.3	80.7
22	93.8	93.7	88.9	88.7
23	97.5	92.3	90.2	90.3
24	90.5	94.6	80.3	86.5
25	95.6	95.6	90.1	90.3
26	95.3	90.4	90.2	88.5
27	96.7	92.3	89.3	88.8
28	97.8	95.8	98.9	90.2
29	96.8	96.8	90.5	90.2
30	95.8	94.2	90.2	90.3
31	96.2	91.3	90.8	90.5
32	90.5	90.5	91.3	89.8
33	91.3	91.5	90.2	89.3
34	94.5	93.2	90.2	89.3
35	91.3	96.7	91.2	90.2
36	92.3	90.8	91.2	91.3
37	92.8	91.3	98.5	92.5
38	92.9	90.8	98.5	90.5
39	93.5	91.9	92.5	90.5
40	94.5	92.3	89.8	87.6

is not dependent on embedding parameters.By proper tuning of embedding parameters for different writers (we used here same values for all writers, different values for different variables of time series), the results can be further improved, though it will cost some higher computational time. At this stage computational cost CTE based measure is lower than DTW based measure as DTW grows exponentially with the length of the time series but CTE has a linear growth..The clustering result by Kmeans with CTE based similarity measure also indicates that the overall recognition accuracy is comparable with DTW based measure while computational cost is lower.

The time series data used in this simulation experiment is basically collected for authentication problem by signature verification. For classification problem, the number of data in each class is small and the two classes are not balanced. The simple experiment with this data shows that the performance of CTE is comparable. For multivariable data, the correlation of the time series is also not taken into account here.

V. Conclusion and Discussion

The analysis of multivariable time series data is challenging due to its dynamic nature. Efficient technique of representation or transformation of MTS data is needed for their analysis by traditional machine learning algorithms.In this work, classification algorithm based on a proposed similarity measure with the representation of the time series by time delay embedding approach has been proposed. The simulation experiment with online signature data, an example of multivariable time series, shows that the CTE based classification algorithms are promising. Though the data used here is not actually suitable for classification experiment and other bench mark time series data must be analysed by the proposed measure to justify its efficiency, the results are quite positive. The measure should also need to be modified to include correlation of the time series of different variable and should be extended properly to deal multivariable data. This approach of time series classification is expected to be suitable for fault analysis in process industry from multi sensor data. Currently further experiments are being carried out to further examine the usefulness of the proposed approach for time series classification and clustering problems with other benchmark time series data sets.

References

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

[2] B. Chakraborty, "Feature selection and classification techniques for multivariate time series",in *Proc. of ICICIC 2007*, September 2007.

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

[4] K. Buza, A. Nanopoulos, L. Schmidt-Thieme, "Fusion of similarity measure for time series classification",*LNCS* vol.6679, pp.253-261, 2011.

[5] B, Chakraborty and Y. Manabe,'An Efficient Approach for Person Authentication using Online Signature Verification', *Proceedings of SKIMA 2008*, pp.12–17.2008.

[6] Z.xing,J.Pei and E. Keogh, "A brief survey on sequence classification",*ACM SIGKDD Explorations Newsletter*, Vol 12, No.1, pp.40-48, 2010.

[7] Lal, T. N. et al.,'Support Vector Channel Selection in BCI', *IEEE Trans. on Biomedical Engineering,* Vol. 51, No. 6, pp.1003-1010, 2004.

[8] L. Ye and E. Keogh," Time series shapelets: A new primitive for data mining", in *Proc.ACM SIGKDD International Conference of Knowledge Discovery and Data Mining*, 2009.

[9] Yoon, H., Yang, K. and Sahabi,C., 'Feature subset selection and feature ranking for multivariate time series', *IEEE Trans. on Knowledge and Data engneering,* Vol 17, No. 9, pp. 1186–1198, 2005.

[10] H. Yoon, and C. Shahabi, 'Feature Subset Selection on Multivariate Time Series with Extremely Large Spatial Features', *Proceedings of Sixth International Conference on Data Mining (ICDMW'06)*, 2006..

[11] S. B. Kim, K. S. Han et al. " Some effective techniques for naive bayes text classification", *IEEE Trans. on Knowledge and Data Engineering*, Vol 18, No. 11,pp. 1457–1466, 2006.

[12] Alligood, K, Sauer, T, Yorke, J. A. *Chaos : An Introduction to Dynamical Systems,* Springer, New York, 1997.

[13] Manabe, Y. and Chakraborty, B., 'Identity Detection from On-line Handwriting Time Series', *Proceedings of SMCia/08,* pp. 365–370.2007.

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

[15] Manabe, Y. and Chakraborty,B., 'A novel approach for estimation of optimal embedding parameters of nonlinear time series by structural learning of neural network', *Neurocomputing,* Vol 70, No. 3, pp.1360–1371, 2007.

[16] Wayland,R., Bromley, D., Pickett, D. and Passamante, A., "Recognizing Determinism in a Time Series', *Physical Review Letters,* Vol 70, No. 5, pp.580-582.1993.

[17] Y. Manabe and B. Chakraborty, " Efficient On-line Handwritten Signature Verification by Integrating DTW and Proposed Local Dynamic Similarity Measure",*Journal of SOFT*, vol 24, no.6 pp 1165–1176, 2012.