ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 4, Issue 3, March 2017

Classification of Time Series Trajectories Based on Shape Features using SVM

[1] Gajanan S. Gawde [2] Jyoti D. Pawar

[1]Computer Engineering Department, Goa College of Engineering, Farmagudi – Ponda , Goa – India

[2] Department of Computer Science & Technology, Goa University, Taleigao Plateau– Panaji , Goa - India.

Abstract— Classification of time series trajectories is a non trivial problem and few researchers have contributed to this field. Time series trajectories such as Character trajectories, ECG trajectories, Stock Exchange Trajectories need to be classified accurately in order to improve the performance of system. The classification of time\ series trajectories is carried out using different techniques such as Neural networks, Support vector machines, Fourier Transforms and Bayesian Network. Most of the existing classification techniques work on simple features such as co-ordinates, distance, and velocity. These techniques are not able to consider shape features of trajectories and classify the test trajectories. Shape of the time series trajectory is an important feature and can be used in the classification method. A support vector machine is better classification technique compared to the existing techniques. In this paper, we have proposed our novel classification method to classify time series trajectories using SVM with shape as the feature vector for classification purpose. Thorough experimental study was carried out on proposed technique with different datasets. Experimental results shows that SVM technique is better compared to Neural Network, Fourier Transform, Bayesian Network. SVM with shape features is efficient compared to SVM without shape features. We have tested SVM with different kernel functions. Our experimental results show that RBF SVM with shape features method is efficient compared with Linear and Non Linear kernels.

***Keywords*– RBF SVM, Shape Based SVM, Time Series Trajectories Classification**

I. INTRODUCTION

Moving objects are responsible for generating huge amount of time series trajectories. The co-ordinates of the moving objects changes as the time changes. The moving objects are tracked and their co-ordinates are stored in the log file along with time interval. The log file contains the information of time series trajectories and it is acting as a database repository. The time series trajectories have some kind of shape and this shape is important feature to classify the time series trajectories. Initially, we extracted the feature vector of time series trajectory and then shape based feature vector is passed to the SVM classifier for classification. The block diagram of our proposed model is shown in the figure 1.

There are many application of the classification of time series trajectory such as Hand Written Character Recognition, ECG classification, Stock Exchange Patterns etc. In hand written character recognition characters are compared with standard characters and converted into digital form. ECG signal of the patient are recorded and are compared with the normal ECG signal. If the signal is matched with the normal signal, then patient is said to be normal else patient is abnormal or not well. Stock exchange is having some interesting hidden patterns and these hidden patterns are useful to the share market investors. The hidden patterns can be extracted using classification technique from stock exchange datasets. The time series trajectories can be classified as good, average and poor patterns for investment purpose.

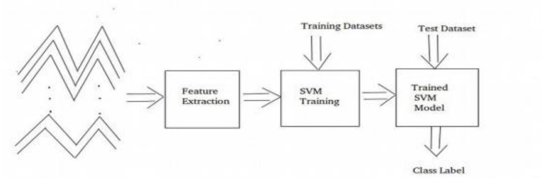

Figure 1: SVM model

The rest of paper is organised as follows. In section 2 we present related work. In section 3, we propose PruneMatrix and PruneHash algorithms to prune redundant nodes of tree. Section 4, provides the experimental results and finally, section 5 concludes the paper.

IFERP
connecting engineers... developing research

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 3, March 2017

Symbols	Meaning
NN	Neural Network
BN	Bayesian Network
FF	Fourier Transform
SVM	Support Vector Machine

Table 1: Meaning Of Symbols Used

II. RELATED WORK

Classification of time series sequence has gained lot of importance and few researchers have been contributed to this field. In paper [1], authors have used constraints to accurately classify the time series. DTW measure is expensive in term of processing and hence DTW measure is improved by pruning in order to improve overall performance. In paper [2], Genetic algorithm technique is used to extract feature from time series and subsequently these features were used in SVM model to classify time series. Authors of [3] have focused on interval based feature for the purpose of classification. The features extracted from time series trajectories were used in SVM to classify data. Statistical method is used to extract the features from time series and neural network was used to classify time series in [4]. In paper [5], features are extracted with the help of grammar and these features vector are used by SVM to classify time series. Cepstral coefficients are fetched as a feature vector and same is classified using Hidden Markov Model in paper [6]. In paper [9], Authors have used entropy to extract features and subsequently classify the time series. Dynamic Hidden Markov model is used in [12] to classify time series. In paper [11], Bayesian network is used to classify time series. In paper [14] , shapelet is used as feature for the classification and same is classified accurately.

In paper [16] , Gaussian Mixture Model is used to classify time series sequence. Classification of time series is done using maximum likelihood of the time series. In paper [17], SVM classification model is proposed to classify imbalance time series. In paper [19] latent is used as feature vector and subsequently classified using Hidden Markov Model. Authors of [20] have proposed technique to improve the performance of SVM to classify the time series. In paper [21], recurrent neural network is used to classify time series sequences. Multivariate time series classification is proposed in [23] with temporal abstraction. In paper [23], heart beat classification is proposed using SVM model. The authors of [24] have proposed a survey of classification technique of time series sequences.

III. SUPPORT VECTOR MACHINES BASED CLASSIFICATION OF TIME SERIES TRAJECTORIES

Problem Definition

Let 'N' be number of time series trajectories of moving objects. Training datasets records have class label defined. There is need to classify given input test samples as one in one of the class defined. The classification should be able to reduce misclassification and also improve the accuracy of classification.

Shape Based Feature Vector

Generally time series trajectories have some kind of shape and this shape can be used as feature vector to classify data. The shape features of time series trajectories are extracted and mapped to feature vector. There is need to mapped feature vector to higher dimension so that data points can be easily differentiable. SVM kernel is used to map feature vector to higher dimension. SVM along with kernel function, maps the feature vector to higher dimension where data is separable. The accuracy of SVM is depends on the quality of shape feature vector and the kind of kernel function used.

The feature extraction process of time series trajectory is shown in the figure 2. The time series trajectory is processed in such a way that, polygons are identified and then using this polygons feature vector is constructed. The time series trajectory T1 is divided into three polygons such as P1 , P2 and P3. The feature vector is represented in two dimensional spaces. Row wise trajectories are defined and column wise polygons and its turning function are defined.

IFERP
connecting engineers... developing research
ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 3, March 2017

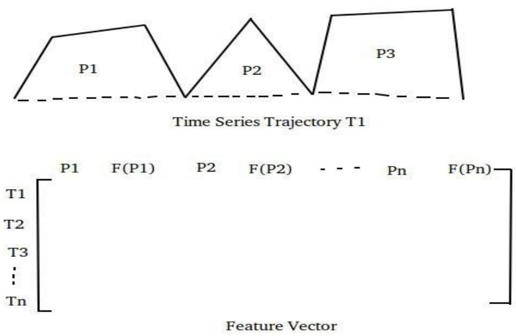

Figure 2: Feature Vector of Time Series Trajectories

Support Vector Machines for Classification

Support Vector Machines is a supervised learning classification method. There are two types of SVM model, first one is two class and second one is multi class SVM models. In two class SVM model, there are only two classes whereas in multi class SVM model, there are more than two class labels. The accuracy of SVM model depends on the quality of shape based feature vector and the kernel function used in SVM model. The kernel function maps the feature vector to higher dimension where datasets can be easily separable.

First, shape feature vector is extracted from raw input time series trajectories. Extracted feature vector is used to configure the SVM model and subsequently SVM is trained with training datasets. During training of SVM, training datasets are passed to SVM model, such that support vectors are identified. These support vectors are the boundary marks which separate the two class samples from each other. Once SVM model is trained with the training datasets, SVM mode is ready to classify test time series trajectories. During classification of test time series trajectory, the support vectors are compared with the test trajectories and accordingly the sample is classified as one of the output class.

Support Vector Machines Classifier: Mathematical Formulation

A key concept required for defining a linear classifier is the dot product between two vectors, also referred to as an inner product or scalar product, defined

$$W^T X = \sum_i WiXi \; .$$

A linear classifier is based on a linear discriminant function of the form

$$y(x) = W^T X + b$$

A pattern x is given a class y=+/- 1 by first transforming the pattern into a feature vector y(x). The hyperplane y(x)=0 defines a decision boundary in the feature space. The parameter W and b are determined by running a learning procedure on a training set (x1,y1),…(xn,yn).
Minimised P(w,b) = ½ w2
Subject to for all i, yi (WT X +b) >= 1

SVM Classification algorithm:-

Input: - Time Series Trajectories
Output: - Class Label
Step1 : Model the SVM with a kernel.
Step2 : Trained the SVM model with given training datasets. Step 3: Classify given test sample using the trained SVM model.
Step 4: SVM model classify the given sample as one of the class.

IV. EXPERIMENTAL STUDY

Configuration

We have used python and C++ languages to write programs and Ubuntu 12.04 operating system. The hardware configuration used for experimental study was Pentium V processor, 4 GB ram and 500 GB harddisk space. We have written C++ and python programs and were tested for correctness with different datasets and found to be correct.

Datasets

GPS Trajectories (GPS): - The dataset has been feed by Android app called Go Track. The dataset is composed by two tables. The first table go_track_tracks presents general attributes and each instance has one trajectory that is represented by the table go_track_trackspoints using latitude, longitude and altitude parameters.

IFERP
connecting engineers... developing research

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 3, March 2017

ECG datasets (ECG): - Concerning the study of H. Altay Guvenir: "The aim is to distinguish between the presence and absence of cardiac arrhythmia and to classify it in either one of the group. Class 01 refers to 'normal' ECG class and Class 02 refers to "abnormal ECG" class.

Istanbul Stock Exchange datasets (Stock) : - Data sets includes returns of Istanbul Stock Exchange with seven other international index; SP, DAX, FTSE, NIKKEI, BOVESPA, MSCE_EU, MSCI_EM from Jun 5, 2009 to Feb 22, 2011.

V. RESULTS

Neural Network, Bayesian Network, Fourier Transform and SVM techniques were compared for the performance. Table 1 shows the results of different classification techiques with their accuracy. The accuracy of Neural Network, Fourier Transform and Bayesian Network is at lower side as compared to Simple SVM model without shape features as can be easily seen from the table 1. Thus SVM technique is efficient compared to other methods. Therefore, we have chosen SVM technique and applied shape features.

	Accuracy			
Dataset	NN	BN	FT	SVM
GPS	78.24	67.23	75.24	80.13
ECG	76.45	72.03	77.45	81.24
Stock	75.12	74.19	74.38	79.25

Table 1: Accuracy of different classification Techniques

The SVM Classification shape based method was tested with existing SVM without shape features method. The experimental results are shown in the Table 2. The SVM without shape features method is showing low accuracy as compared to SVM with shape features. Shape features are able to consider the shape of trajectories and classify the test samples accordingly. The average accuracy of SVM with shape features is 70.27 whereas SVM without shape features is 39.71. Thus, SVM with shape features method is efficient compare to SVM without shape features.

	Accuracy	
Dataset	SVM without shape feature	SVM with shape feature
GPS	80.13	88.56
ECG	81.24	87.23
Stock	79.25	89.34

Table 2: Classification Accuracy of SVM Model with Shape Features

The classification accuracy of different SVM kernels was recorded of different datasets. The table 3 shows the results of experimental study carried out with different datasets such as GPS Trajectories, ECG datasets and Istanbul Stock Exchange. The experimental results show that RBF kernel performance is better compared to linear and non linear kernels.

	Accuracy		
Kernel	GPS	ECG	Stock
Linear	87.56	89.23	88.34
Non-Linear	86.34	88.59	87.98
RBF	91.03	90.49	91.89

Table 3: Classification Accuracy with different SVM kernel

VI. CONCLUSION

Time Series classification model was successfully proposed using Support Vector Machines with shape features. SVM technique was compared with Neural Network, Bayesian Network and Fourier Transform experimentally. Experimental results show that SVM method was efficient compared to Neural Network, Bayesian Network and Fourier Transform. Further, SVM with shape features method was efficient compared to SVM without shape features. SVM method was tested with different kernel functions. Our experimental study show that SVM with RBF was efficient compared to Linear and Non Linear functions. Therefore, RBF SVM with shape features method was efficient compared with the rest of the techniques.

REFERENCES

[1] Chotirat Ann Ratanamahatana , Eamonn Keogh , 2004,Making Time-series Classification More Accurate Using Learned Constraints. SDM.

[2] Damian Eads Daniel Hill, Sean Davis,2002, Genetic Algorithms and Support Vector Machines for Time Series Classification. Application and Science of Neural Networks.

[3] Juan Jose Rodriguez , 2005, Support Vector Machines of Interval-based Features for Time Series Classification. Research and Development in Intelligent Systems.

[4] Alex nanopoulos , Rob Alcock, Yannis Manolopouslos,1999, Feature based classification of time series data. ICDE.

[5] Damian Eads, Karen Glocer, Simon Perkins, James Theiler, 2005, Grammar-guided Feature Extraction for Time Series Classification. NIPS.

[6] Mark D. Skowronski , John G. Harris, 2006, Minimum mean squared error time series classification using an echo state network prediction model. IEEE International Symposium on Circuits and Systems.

[7] Sofiane Brahim-Belhouari , Amine Bermak , 2004, Gaussian process for non stationary time series prediction. Computational statistics and Data Analysis.

[8] Richard P. Lippmann , 2004, Pattern Classification Using Neural Networks. Pattern Recognition.

[9] Houtao Deng , George Runger, Eugene TuvMartyanov Vladimir , Eugene TuvMartyanov Vladimir , 2013,A Time Series Forest for Classification and Feature Extraction. Journal of Information Science.

[10] Stefan Ruping , 2006, SVM Kernels for Time Series Analysis. Artificial Neural Network.

[11] Krisztian Buza , Lars Schmidt-Thieme, 2009, Motif-based Classification of Time Series with Bayesian Networks and SVMs.Data Analysis , Data Handling and Business Intelligent.

[12] Vladimir Pavlovi´cy, Brendan J. Freyz, Thomas S. Huang, 1999, Time-Series Classification Using Mixed-State Dynamic Bayesian Networks. Computer Vision and Pattern Recognition.

[13] Mehran Kafai, Bir Bhanu, 2011, Dynamic Bayesian Networks for Vehicle Classification in Video. IEEE Transaction on Industrial Informatics.

[14] Jon Hills, Jason Lines , Edgaras Baranauskas , James Mapp , Anthony Bagnall,2014, Classification of time series by shapelet transformation. Data Mining and Knowledge Discovery.

[15] Xiaopeng Xi , Eamonn Keogh , Christian Shelton Li Wei , 2006,Fast Time Series Classification Using Numerosity Reduction. ICML.

[16] Richard J. Povinelli, Michael T. Johnson, Andrew C. Lindgren, Jinjin Ye, 2004, Time Series Classification Using Gaussian Mixture Models of Reconstructed Phase Spaces. IEEE Transaction on Knowledge and Data Engineering.

[17] Suzan Köknar-Tezel , Longin Jan Latecki,2011, Improving SVM classification on imbalanced time series data sets with ghost points. Knowledge and Information System. [18] Jason Lines , Luke M. Davis , Jon Hills , Anthony Bagnall, 2009, A shapelet transform for time series classification. SIGKDD.

[19] Peter Sykacek , Stephen Roberts , Bayesian time series classification.

[20] Edgar Osuna , Robert Freund , 1997, An improved training algorithm for support vector machines. Neural Network for signal Processing.

[21] Michael Husken , Peter Stagge , 1993, Recurrent Neural Networks for Time Series Classification. International conference of Neural Network.

[22] Iyad Batal, Lucia Sacchi and Riccardo Bellazzi , Milos Hauskrecht , 2009, Multivariate Time Series Classification with Temporal Abstractions. FLAIRS.

[23] Argyro Kampouraki, George Manis, and Christophoros Nikou, 2008, Heartbeat Time Series Classification With Support Vector Machines. IEEE Transaction on Information Technology in Bioinformatics.

[24] Zhengzheng Xing , Jian Pei, Eamonn Keogh , 2006, A Brief Survey on Sequence Classification. ICML.