Knowledge-Based Systems 75 (2015) 184–191

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Support vector-based algorithms with weighted dynamic time warping kernel function for time series classification

Young-Seon Jeong [a,*], Raja Jayaraman [b]

[a] Department of Industrial Engineering, Chonnam National University, Gwangju, Republic of Korea
[b] Department of Industrial & Systems Engineering, Khalifa University, Abu Dhabi, United Arab Emirates

ARTICLE INFO

Article history:
Received 17 July 2014
Received in revised form 30 November 2014
Accepted 1 December 2014
Available online 6 December 2014

Keywords:
Time series classification
Weighted dynamic time warping kernel function
Multiclass support vector data description
Multiclass support vector machines
Dynamic time warping

ABSTRACT

In this paper, we propose support vector-based supervised learning algorithms, called multiclass support vector data description with weighted dynamic time warping kernel function (MSVDD-WDTWK) and multiclass support vector machines with weighted dynamic time warping kernel function (MSVM-WDTWK), which provides a flexible and robust kernel function for time series classification between non-aligned time series data resulting in improved accuracy. The proposed WDTW kernel function provides an optimal match between two time series data by not only allowing a non-linear mapping between two data sequences, but also considering relative significance depending on the phase difference between points on time series data. We validate the proposed approaches using extensive numerical experiments on a number of multiclass UCR time series data mining archive, and demonstrate that our proposed methods provide lower classification error rates compared with existing techniques.

1. Introduction

Time series data, which represents ordered sequence of observations are quite common in several domains such as aviation, healthcare, financial, seismology, signal processing, computer vision, to name just a few [10,18,20,22,23,37,42,51]. In the last few decades, the explosive growth of data combined with superior computing power has drawn the attention of several researchers to classify, model and predict time series based data. Exploring time series data presents enormous challenges due to volume of data, high dimensionality, and measurement noise, varying statistical properties which require sophisticated methods to index, and perform analysis. We refer the readers to excellent survey articles and background material by Längkvist et al. [31], Kotsiantis [29], Wei and Keogh [54], Xing et al. [55], Yuan and Liu [58], Górecki [14].

Several classification methods have been used for modeling time series data such as Artificial Neural Networks (ANN), Bayesian Networks, Hidden Markov Models (HMM), Genetic Algorithms (GA), and support vector machine (SVM) [6,9,34,49,60]. Among the classification methods, SVM is most commonly employed to classify high dimensional representations of data by constructing hyperplanes in a multidimensional space that separates different class labels. The main idea of SVM is to determine an optimum hyperplane for linearly separable data sets, extended to patterns that are not linearly separable by transformations of original data into a new space using kernel functions. In addition, support vector data description (SVDD) method, which is one class classification method based on SVM, aims to find a set of support vectors describing a sphere or domain with minimum volume containing the target data by mapping into a higher dimensional space [50]. SVDD has been used in a number of real-world applications such as face recognition, image processing, pattern detection, quality control, and more typically employed for outlier detection and classification [30,38,44,53,56,57]. Time series classification techniques using HMM are popular to model human activity recognition (HAR) in predicting trajectories based on surveillance videos. Nascimento et al. [39] developed a framework for modeling and recognition of human trajectories shared by several classes incorporating probabilistic switching that resembles an HMM-based classifier. Conditional Random Fields (CRFs) have been used for labeling sequence data for speech, gesture recognition in human activities. Extending the work of Nascimento et al. [39], Gao and Sun [13] used hidden conditional random fields for HAR. Gao and Sun [11] employed a beta process to discover a set of hidden motion common to multiple trajectories. Model parameters were learned using Markov Chain Monte Carlo (MCMC) algorithm to obtain a classifier for HAR by maximizing the log-likelihood of a test trajectory. Gao and Sun [12] proposed a method for HAR modeling based on the hierarchical Dirichlet process hidden Markov

* Corresponding author. Tel.: +82 62 530 1790; fax: +82 62 530 1789.
E-mail address: young.jeong@jnu.ac.kr (Y.-S. Jeong).

http://dx.doi.org/10.1016/j.knosys.2014.12.003
0950-7051/

models (HDP-HMM) employing Gibbs sampler to obtain parameter estimates for the model. Recent extensions to time series classification methods include ensemble methods commonly referred as time series forest that combines multiple or weighted classifiers to produce a more accurate prediction than individual classifiers [7]. Rodriguez and Kuncheva [43] benchmarked the performance of three types of classifiers nearest neighbor (using DTW), SVM (linear and perceptron kernel) and decision forests (boosting, random forest, rotation forests) to conclude that there is a lot of room for improvement in devising a more reliable error estimation that is flexible enough to accommodate very small labeled data sets and make use of unlabeled data.

In time series classification, the distance between two time series data is critical to identify structural similarities for improved prediction accuracy. Several distance measures have been used such as Euclidean distance (ED) and dynamic time wrapping (DTW). Gudmundsson et al. [15] investigated the effectiveness of time series classification when pairwise similarities are based on DTW distance measure. Although Euclidean distance metric is popular, requires both input series to be of same length, while DTW provides a more elastic measure to classify time series data [2]. DTW aligns two sequences by wrapping the time axis iteratively till an optimal match is found. Regardless of the phase difference between a reference point and a testing point, however, DTW assigns equal weights to each point, leading to misclassification for the shape similarity between two sequences. To overcome this drawback, Jeong et al. [24] proposed the weighted DTW (WDTW) measure, which considers relative significance depending on the phase difference between points.

The main objective of this paper is to propose effective support vector-based algorithms for time series classification. The proposed approaches are based on a weighted DTW (WDTW) kernel function combining multiclass supervised learning algorithms, called multiclass support vector data description with weighted dynamic time warping kernel function (MSVDD-WDTWK) and multiclass support vector machines with weighted dynamic time warping kernel function (MSVM-WDTWK), which provide a flexible and robust matching algorithm for time series classification. The WDTW kernel function contains a generalized distance measure that penalizes the points according to the phase difference between a test point and a reference point to prevent minimum distance distortion by outliers. We experimentally validate the performance of the proposed approaches using several real life multiclass time series datasets from "UCR Time Series Data Mining Archive".

The rest of the paper is organized as follows. We introduce the basic concepts of the related methodology of supervised learning algorithms and discuss DTW and weighted DTW technique in time series classification problems in Section 2. In Section 3, we discuss our proposed approaches, namely MSVDD-WDTWK and MSVM-WDTWK. The experimental results are presented in Section 4. We present conclusions and some future research directions in Section 5.

2. Related methodology

2.1. Supervised learning algorithms

Supervised learning is analogous to instructed learning system that associates class labels based on the training data. The primary goal of supervised learning algorithms is to use annotated classes from training data to model and classify the output of unobserved time series. Supervised learning is a two-step process, the first step is to learn the model using training data and followed by testing the model using unseen data to assess the model accuracy. Variety of supervised learning algorithms including fisher linear discriminant, K-nearest neighbors (KNN), decision trees, neural networks based, Bayesian networks, support vector machines have been developed and used in various applications. The nature of these algorithms varies based on robustness, speed, scalability, model compactness, prediction accuracy, bias and transparency.

In addition, support vector machine (SVM) is an important class of supervised learning algorithms introduced by Vapnik [52] in 1990s relies on the notion of "margin", which represents either side of the hyperplane that separate two data classes. Maximizing the margin enables the largest distance between separating hyperplane to reduce the upper bound on the expected generalization error. Fig. 1 gives a representation of the maximum margin hyperplane, where distance between support vectors to the hyperplane is maximized. When the training data is linearly separable it is possible to show an optimum separating hyperplane can be found by minimizing the squared norm of the separating hyperplane. The minimization procedure can be modeled as a convex quadratic optimization problem and solved [17]. We refer the readers to excellent references written by Cristianini and Shawe-Taylor [5], Vapnik [52], and Burges [3] for more details on SVM. Shawe-Taylor and Sun [47] review several optimization techniques used in SVM classification algorithms and characterization of effective kernels. Standard SVM involves solving time consuming quadratic programming problem and has a computational complexity of $O(n^3)$, where n is total size of training data. To improve the computation speed and accuracy, several variants of SVM have been developed and applied including SVM light, least squares SVM, library for SVM, twin support vector machines, weighted SVM, sampling SVM, etc. [46]. Some of the recent progresses in multiclass SVM include learning using privileged information, Ji et al. [26] propose a multitask multiclass privileged information SVM as constrained optimization problem with quadratic objective function. Ji and Sun [25] modeled a multitask multiclass SVM as a constraint optimization problem with quadratic objective function.

As a variant of SVM, the support vector data description (SVDD) proposed by Tax and Duin [50], aims to find a spherically shaped description for target data. The fundamental idea of SVDD is to map training data nonlinearly into a higher feature space and minimize the volume of the hypersphere containing the most mapped training data. The target data may contain more than one class of objects and each class of objects need to be described and distinguished simultaneously [21]. SVDD is modeled as a constrained convex optimization problem and is quite popular in outlier analysis, anomaly detection applied to quality control, manufacturing, speech recognition, etc. Several recent improvements to SVDD have been proposed including, Liu et al. [36] used a weighting factor for each data point based on their spatial position distribution in train-

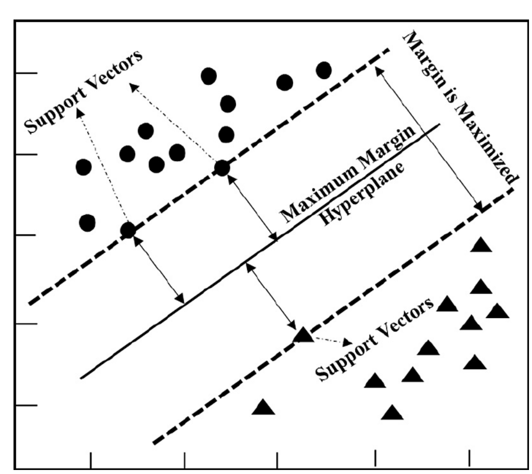

Fig. 1. Optimal hyperplane maximizing the distance to support vectors of SVM.