Knowledge-Based Systems 81 (2015) 131–147

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A comparison on multi-class classification methods based on least squares twin support vector machine

Divya Tomar *, Sonali Agarwal

Indian Institute of Information Technology, Allahabad, India

ARTICLE INFO

Article history:
Received 9 September 2014
Received in revised form 25 January 2015
Accepted 9 February 2015
Available online 18 February 2015

Keywords:
Least squares twin support vector machine
Twin support vector machine
Multi-class classification
Support vector machine
Multi-class least squares twin support vector machine

ABSTRACT

Least Squares Twin Support Vector Machine (LSTSVM) is a binary classifier and the extension of it to multiclass is still an ongoing research issue. In this paper, we extended the formulation of binary LSTSVM classifier to multi-class by using the concepts such as "One-versus-All", "One-versus-One", "All-versus-One" and Directed Acyclic Graph (DAG). This paper performs a comparative analysis of these multi-classifiers in terms of their advantages, disadvantages and computational complexity. The performance of all the four proposed classifiers has been validated on twelve benchmark datasets by using predictive accuracy and training–testing time. All the proposed multi-classifiers have shown better performance as compared to the typical multi-classifiers based on 'Support Vector Machine' and 'Twin Support Vector Machine'. Friedman's statistic and Nemenyi post hoc tests are also used to test significance of predictive accuracy differences between classifiers.

1. Introduction

Support Vector Machine (SVM), proposed by Vapnik et al., is a supervised machine learning approach. SVM has advantage over other existing classification approaches as it provides global solution for the data classification [1,2]. It generates a unique global hyper-plane to separate the data points of different classes rather than local boundaries as compared to the other existing data classification approaches. Since SVM follows the Structural Risk Minimization (SRM) principle, so it reduces the occurrence of risk during the training phase [1–4]. SVM is used for both classification and regression tasks [5–11]. Initially, SVM was developed for binary classification; later researchers successfully extended the formulation of binary SVM to multi-class problem scenario [12–22]. Due to its better performance, SVM is one of the most widely used classification techniques of data mining that has its application in many fields, for example, in disease detection, text categorization, software defect prediction, speech recognition, face identification, bankruptcy prediction, intrusion detection, time series forecasting, music emotion detection and etc. [23–48]. But one of the main issues with the conventional SVM is to obtain the solution of a complex Quadratic Programming Problem (QPP). SVM solves a complex QPP with inequality constraints and constructs an optimal separating hyper-plane that maximizes the margin of two classes. The dual formulation of SVM depends on the size of training dataset and all data points give constraints to QPP. If the size of training dataset is l, then the computational complexity of SVM is $O(l^3)$ which is very expensive.

Recently, Mangasarian et al. proposed a binary classifier named as Generalized Eigen-value Proximal Support Vector Machine (GEPSVM) which classifies the data points of two classes by generating two non-parallel hyper-planes [49]. In order to reduce the computational complexity of SVM, Jayadeva et al. proposed Twin Support Vector Machine (TWSVM), which is a binary classifier [50]. TWSVM is inspired by the concept of SVM and GEPSVM and classifies the data points of two classes by generating two non-parallel hyper-planes. For this purpose, it solves two smaller size QPPs rather than a complex QPP as in conventional SVM which makes the learning of TWSVM classifier four times faster than the standard SVM. As opposed to SVM, the data points of one of the two classes give constraints to each QPP in TWSVM. Some improvements to the TWSVM have been proposed by the researchers to obtain higher predictive accuracy with lower computational time such as Twin Bounded Support Vector Machine (TBSVM), Twin Parametric Margin Support Vector Machine (TPMSVM), Structural TPMSVM, Sparse TWSVM, Least Squares Twin Support Vector Machine (LSTSVM) etc. [51–66]. TWSVM is also extended to multi-class scenario and appears as a better alternative to Multi-class SVM approaches due to its better computational speed and comparable predictive accuracy [67–70].

* Corresponding author. Tel.: +91 9026021639.
E-mail addresses: divyatomar26@gmail.com (D. Tomar), sonali@iiita.ac.in (S. Agarwal).

http://dx.doi.org/10.1016/j.knosys.2015.02.009
0950-7051/

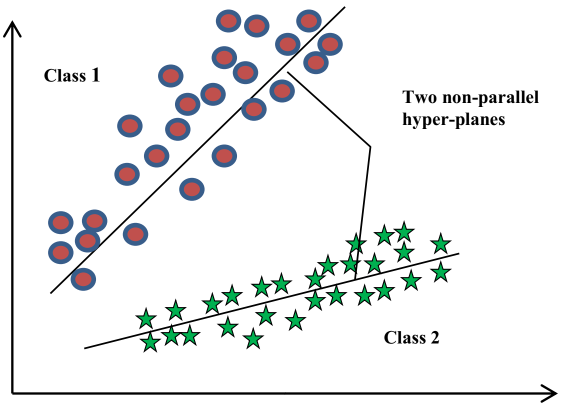

Fig. 1. Twin Support Vector Machine.

Although TWSVM is faster as compared to the SVM, yet it involves solving of two QPPs which is a complex process. Therefore, Kumar et al. proposed a binary classifier LSTSVM, which is the least squares variant of TWSVM [56]. LSTSVM determines two non-parallel hyper-planes by solving two linear equations rather than two QPPs as in TWSVM. It shows better generalization performance and is faster than the TWSVM. Inspired by LSTSVM, Least Squares Recursive Projection TWSVM (LSPTSVM) has been proposed by Shao et al. which is the least squares version of Projection TWSVM (PTSVM) [71,72]. LSTSVM and LSPTSVM generate their separating hyper-plane by solving linear equations and have fast training speed.

LSTSVM is suitable for the binary classification problem. However, most of the real life applications are related to multi-class classifications such as disease detection, activity recognition, speaker identification, digit recognition, text categorization etc. Similar to binary TWSVM, the QPPs of multi-class TWSVM classifiers have inequality constraints. Hence, the problem of binary TWSVM classifier also exists in multi-TWSVM classifiers which involve the solution of complex QPPs. Therefore, in order to utilize the advantages of binary LSTSVM in time complexity, simplicity and comparable predictive accuracy, we extend the formulation of it to multi-class problem domains and propose four novel multi-classifiers. These classifiers are based on One-versus-All (OVA), One-versus-One (OVO), All-versus-One (AVO) and Direct Acyclic Graph (DAG) concepts. The first multi-classifier is based on OVA strategy, in which the data points of a class are trained with the data points of rest of the other classes. For K-class classification problem, it solves K-linear equations and determines K non-parallel hyper-planes, one for each class. For a test data point, its distance is calculated from each hyper-plane and a class is assigned to the given data point from which it lies nearest. On the other hand, OVO Multiclass Least Squares Twin Support Vector Machine (MLSTSVM) classifier solves K (K-1) linear equations and constructs K (K-1) binary LSTSVM classifiers where each classifier is trained with the data points of two classes. The class is assigned to a given data point on the basis of max-win voting strategy i.e., the class with maximum vote is assigned to the data point. The vote is given to the class on the basis of its distance from the point. If a point lies nearer to a class as compared to another class, then the vote is given to it. In the construction of the third classifier, we adopt "All-versus-One" concept. AVO MLSTSVM classifier generates K-binary LSTSVM classifiers and K non-parallel hyper-planes, one for each class. The data points of *i*th class provide constraints to *i*th LSTSVM classifier i.e., it considers the data points of other classes with positive class labels and data points of *i*th class with negative class labels. The concept of AVO is different from the concept of OVA in which the data points of other classes provide constraint to the *i*th classifier. In AVO MLSTSVM classifier, the given data point is assigned to a class which lies farthest from it. All these classifiers suffer from the unclassifiable region problem. In order to handle this problem, we propose fourth classifier named as DAG MLSTSVM. This classifier not only solves the problem of unclassifiable region but also shows better generalization ability and takes lesser time in testing. The training phase of DAG MLSTSVM is similar to OVO MLSTSVM classifier. In testing phase, DAG MLSTSVM constructs a binary rooted DAG which includes K (K-1)/2 internal nodes and each node corresponds to a binary LSTSVM classifier of *i*th and *j*th classes. In this research work, we analyze and compare the computational complexity and predictive accuracy of each classifier with the other existing classifiers such as Multi SVM, Multiple Birth Twin Support Vector Machine (MBSVM) and Twin KSVC.

This study follows the recommendations given by Demsar in order to make the statistical inferences from the observed difference in predictive accuracy. Demsar provided recommendation regarding statistical comparison of classifiers over multiple datasets [73]. Therefore, in this research work, the performance of each classifier is compared using Friedman's average rank test and Nemenyi post hoc test is employed to test the significance of differences in the rank of individual classifiers. Modified version of Demsar significance diagram is also plotted to display the results.

The rest of the paper is organized as follows. The brief introduction of fundamental approaches such as TWSVM, LSTSVM and several existing multi-classifier approaches based on TWSVM and SVM is given in Section 2. In Section 3, we propose the formulations of MLSTSVM classifiers for both linear and non-linear cases. In Section 4, we analyze the computational complexity of each multi-classifier and also discuss their advantages and disadvantages. Section 5 presents and discusses the results of experiment. Finally, concluding remarks and recommendations for further research work are given in Section 6.

2. Background

This section includes the brief description of traditional TWSVM, LSTSVM, MBSVM, Twin KSVC and Multi-class SVMs. For binary classification, the training dataset is represented as:

$$T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)\} \tag{1}$$

where $x_i \in R^n, i = 1, 2, \ldots, l$, indicates input data points in n-dimensional real space R and $y_i \in \{+1, -1\}$ corresponds to the class label. Let class +1 contains l_1 data points and class -1 contains l_2 data points and $l = l_1 + l_2$.

2.1. Twin Support Vector Machine

Consider the matrices $X_1 \in R^{l_1 \times n}$ and $X_2 \in R^{l_2 \times n}$ comprise the data points of class +1 and class -1 respectively. TWSVM classifier seeks following two non-parallel hyper-planes

$$x^T w_1 + b_1 = 0 \quad \text{and} \quad x^T w_2 + b_2 = 0 \tag{2}$$

by solving two QPPs:

$$\begin{aligned} \min(w_1, b_1, \xi)\ & \frac{1}{2}\|X_1 w_1 + e_1 b_1\|^2 + c_1 e_2^T \xi \\ & \text{s.t.}\ -(X_2 w_1 + e_2 b_1) + \xi \geqslant e_2,\ \ \xi \geqslant 0 \end{aligned} \tag{3}$$

$$\begin{aligned} \min(w_2, b_2, \eta)\ & \frac{1}{2}\|X_2 w_2 + e_2 b_2\|^2 + c_2 e_1^T \eta \\ & \text{s.t.}\ (X_1 w_2 + e_1 b_2) + \eta \geqslant e_1,\ \ \eta \geqslant 0 \end{aligned} \tag{4}$$

where $e_1 \in R^{l_1}$ *and* $e_2 \in R^{l_2}$ are the vectors of 1's, c_1 and c_2 are non-negative penalty parameters and $\xi \in R^{l_2}$ and $\eta \in R^{l_1}$ are slack