ARTICLE IN PRESS

Knowledge-Based Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Multi-view learning based on nonparallel support vector machine

Jingjing Tang[a,b], Dewei Li[a,b], Yingjie Tian[*,b,c,d], Dalian Liu[*,e]

[a] School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[b] Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing 100190, China
[c] Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China
[d] School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China
[e] Department of Basic Course Teaching, Beijing Union University, Beijing 100101, China

ARTICLE INFO

Keywords:
Multi-view learning
Nonparallel support vector machine
Alternating direction method of multipliers

ABSTRACT

Multi-view learning (MVL) focuses on the problem of learning from the data represented by multiple distinct feature sets. Various successful SVM-based multi-view learning models have been proposed to improve existing learning tasks performance. Since nonparallel support vector machine (NPSVM) is proposed with several incomparable advantages over the state-of-the-art classifiers, it is potentially beneficial to perform the multi-view classification task using NPSVM. In this paper, we build a new multi-view learning model based on nonparallel support vector machine, termed as MVNPSVM. By combining the large margin mechanism and the consensus principle, MVNPSVM not only inherits the advantages of both NPSVM and multi-view learning, but also brings a new insight of extending NPSVM to the multi-view learning field. To solve MVNPSVM efficiently, we adopt the alternating direction method of multipliers (ADMM) as the solution. We theoretically analyze the performance of MVNPSVM from the viewpoints of the consensus analysis and the comparisons with the other two similar methods SVM-2K and multi-view twin support vector machines. Experimental results on 95 binary data sets confirm the effectiveness of the proposed method.

1. Introduction

Multi-view learning (MVL) focuses on data sets that can be represented by multiple distinct feature sets. Multiple views offer complementary strengths, and in many cases, using multiple relational views together can yield better performance in the machine learning tasks [1,2]. Conventional algorithms for handling multi-view data are either to concatenate multi-view into one single view with comprehensive description (as shown in Fig. 1(a)), or to build a learning function for each separate view and then jointly optimize the learning function by exploiting redundant views (as shown in Fig. 1(b)). The concatenating strategy ignores the statistical property of each view and leads to the *curse of dimensionality* problem. While the separation strategy considers each view independently. In fact, views are inherently related since they describe the same set of objects through different feature spaces. A number of methods [3–6] have shown that learning with multiple views jointly is better than the naive approach of using one concatenated view or learning from each view separately.

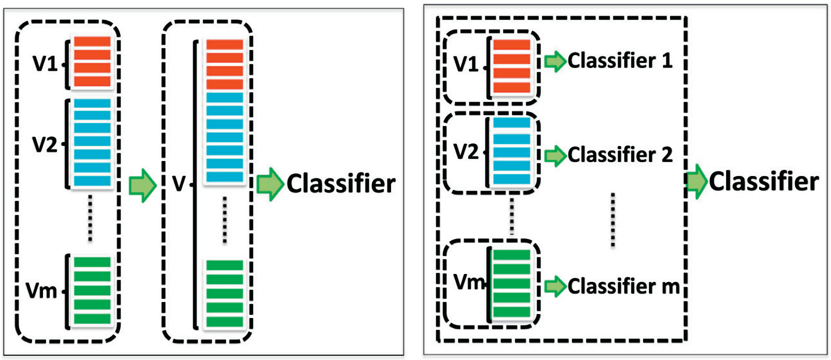
Considering the relationship between multiple views and the combination way for them, the success of multi-view learning algorithms are guaranteed by two significant principles: consensus and complementarity principles [7]. The consensus principle aims at maximizing the agreement on multiple distinct views to achieve an accurate classifier on each view. The complementarity principle emphasizes the complementary information shared by views so that the data can be described comprehensively. In multi-view learning, two principles are crucial to guide model construction.

Under the consensus or complementarity principles, numerous multi-view learning algorithms have been proposed that can be divided into two categories: co-training style algorithms [4,8,9] and co-regularization style algorithms [1,10–12]. Co-training style algorithms yield the classifiers on each view by iterative maximizing the mutual agreement of multi-view and exchanging complementary information from one to another. Co-regularization style algorithms take the disagreement among different views as a new regularization term in the learning objective function that needs to be minimized.

In both categories, many SVM-based algorithms have been proposed. SVM-based co-training style algorithms [4,8] train separate yet correlated classifiers based on three assumptions, i.e., sufficiency, compatibility and conditional independence. The independence assumption is often too rigorous to realize. On the other hand, SVM-based co-regularization style algorithms [10,12,13] directly solve the learning

* Corresponding authors.
E-mail addresses: tangjingjing13@mails.ucas.ac.cn (J. Tang), lidewei15@mails.ucas.ac.cn (D. Li), tyj@ucas.ac.cn (Y. Tian), ldtdalian@buu.edu.cn (D. Liu).

https://doi.org/10.1016/j.knosys.2018.05.036
Received 10 October 2017; Received in revised form 8 April 2018; Accepted 27 May 2018



(a) Concatenating Strategy (b) Separation Strategy

Fig. 1. Two strategies for handling multi-view data: (a) the concatenating strategy: converting multi-view data into single-view data by concatenating heterogeneous feature spaces into one homogeneous feature space; and (b) the separation strategy: building a learning function for each feature view separately and then optimizes the learning function by exploiting redundant views of the same input data.

problem without any assumptions. Since the SVM-based multi-view classifiers ignore the prior information within the classes, the effectiveness of these methods is limited. For this reason, we have concentrated on learning the multi-view classifiers considering the potential information within each class.

As a successful extension of SVM, nonparallel support vector machine (NPSVM) [14] pursues two nonparallel hyperplanes such that each hyperplane is closer to one of the two classes and is at least one distance from the other. It takes the information of each class into account and possesses several incomparable advantages over the state-of-the-art classifiers: (a) under the structural risk minimization (SRM) principle, a pair of primal problems of NPSVM are constructed instead of a large one of standard SVM; (b) it has inherited the sparseness of standard SVM; (c) comparing with existing nonparallel classifiers such as twin support vector machines (TWSVMs) [15], NPSVM can directly apply the kernel trick in the elegant formulations of dual problem for the nonlinear case; (d) existing TWSVMs are only the special cases of the NPSVM when the parameters of which are appropriately chosen. Thus, it is potentially beneficial to fulfill the multi-view classification task using NPSVM.

For the above considerations, this paper aims to integrate NPSVM into the multi-view classification framework for performance improvement. To achieve a new learning model, we need to address the following challenges:

- Most existing multi-view classification models are established based on SVM. How to build a new multi-view learning model based on NPSVM has not been considered.
- Because of the increased number of variables and parameters for multi-view analysis, how to develop effective algorithms is the second challenge.
- Assuming a new NPSVM-based multi-view learning model has been built, how to theoretically guarantee its performance is the third challenge.

In this paper, we propose a new multi-view learning model based on nonparallel support vector machine (MVNPSVM). It combines the large margin mechanism and the consensus principle into a joint framework for multi-view classification. Under the large margin mechanism, it pursues two nonparallel hyperplanes by fully digging the within-class information as well as the between-class separation information. The consensus principle can be guaranteed through forcing the consistency constraints of two-view. Therefore, MVNPSVM not only inherits the advantages of both NPSVM and MVL, but also brings a new insight of extending NPSVM to the MVL field.

To address the second challenge, we adopt ADMM as the solution. To address the third challenge, we theoretically analyze the performance of MVNPSVM from the viewpoint of the consensus analysis. Due to the great relevance to other two-view classification methods, i.e., SVM-2K and multi-view twin support vector machines (MvTSVMs), we compare MVNPSVM with them and conclude that MVNPSVM can be taken as an improved version of them. A great many numerical experiments confirm the effectiveness of the proposed method.

Our contributions are summarized as follows:

(1) We propose a new multi-view learning model called MVNPSVM under the large margin mechanism and the consensus principle. MVNPSVM not only preferably unleashes the power of multiple views, but also fully digs the within-class information as well as the between-class separation information for better classification performance.

(2) In order to improve the training efficiency of our model, we develop ADMM to obtain the solution.

(3) We theoretically analyze the performance of MVNPSVM from two aspects, i.e., the consensus principle and the comparisons with SVM-2K and MvTSVMs. The consensus principle in MVNPSVM is guaranteed by KCCA-style analysis using Rademacher complexity. The comparisons with SVM-2K and MvTSVMs reveal that MVNPSVM serves as an improved and extended version of SVM-2K and MvTSVMs, and inevitably inherits the advantages of them.

(4) We conduct experiments on 95 binary data sets. The results confirm the effectiveness of the proposed method.

The remainder of this paper is organized as follows. Section 2 briefly reviews the related works. Section 3 introduces the MVNPSVM model. Section 4 provides the corresponding algorithm. Section 5 theoretically analyzes the performance of the proposed method. Experimental results and discussions follow in Section 6. Finally, this paper is concluded in Section 7.

2. Related work

In this section, a brief overview is provided from both multi-view learning and nonparallel support vector machine.

2.1. Multi-view learning

Existing multi-view learning models can be divided into two categories [16]: co-training style algorithms [4,8,9] and co-regularization style algorithms [1,10–12,17]. Representative co-training style algorithms include tri-training semi-supervised learning algorithm [18], the graph-based and disagreement-based semi-supervised co-training method [4] and the co-training cross-view based graph random walk approach [9], etc. In contrast, typical co-regularization style algorithms include sparse multi-view SVMs [11], multi-view vector-valued manifold regularization (MV^3MR) [17], multi-view maximum entropy discrimination (MVMED) [19], multi-view twin support vector machine (MvTSVMs) [20], to name a few.

For both categories, lots of SVM-based multi-view classification methods have appeared in the literature. For example, Xu et al. extended the information bottleneck (IB) theory to multi-view learning