ARTICLE IN PRESS

JID: NEUCOM [m5G;September 7, 2017;23:50]

Neurocomputing 000 (2017) 1–13

Contents lists available at ScienceDirect

Neurocomputing

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

Scalable transfer support vector machine with group probabilities

Tongguang Ni[a], Xiaoqing Gu[a,*], Jun Wang[b,c], Yuhui Zheng[d], Hongyuan Wang[a]

[a] School of Information Science and Engineering, Changzhou University, Changzhou, Jiangsu 213164, China
[b] School of Digital Media, Jiangnan University, Wuxi, Jiangsu 214122, China
[c] Department of Radiology and BRIC, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[d] School of Computer and software, Nanjing University of Information Science & Technology, Nanjing, Jiangsu 210044,China

ARTICLE INFO

Article history:
Received 26 July 2016
Revised 22 August 2017
Accepted 29 August 2017
Available online xxx

Communicated by Dr Zhang Zhaoxiang

Keywords:
Large datasets
Classification
Support vector machine
Transfer learning
Group probability

ABSTRACT

A novel transfer support vector machine called TSVM-GP with group probabilities is proposed for the scenarios where plenty of labeled data in the source domain and the group probabilities of unlabeled data in the target domain are available. TSVM-GP integrates a transfer term and group probabilities into the support vector machine (SVM) to improve the classification accuracy. In order to reduce the high computational complexity of TSVM-GP, the scalable version of TSVM-GP called scalable transfer support vector machine with group probabilities (STSVM-GP) is further developed by selecting the representative set of the training samples as the training data in the source domain. Experimental results on synthetic datasets as well as several real-world datasets show the effectiveness of the proposed classifiers, and especially STSVM-GP is very feasible for large scale transfer datasets.

1. Introduction

Learning from group probabilities [1–6] is attractive in the scenarios where the samples are provided as groups and only the label proportion of the samples is available. One of the most natural applications comes in analyzing the outcomes of political elections, where the population of all voters in an electoral district is known, but only the total number of votes per party in each district is revealed. However, from an analysis of this data, e.g. the dependence of votes on variables such as income or household types, can show up interesting connections, and may be used to uncover election fraud when outliers from this model are uncovered. This case is quite different from the traditional supervised, unsupervised and semi-supervised learning problems. Fig. 1 shows the difference among several traditional learning algorithms, including: supervised learning, unsupervised learning, semi-supervised learning and learning from group probabilities. Learning from group probabilities can be regarded as an algorithm lying somewhere between supervised learning algorithms and semi-supervised learning algorithms.

Various algorithms have been developed by utilizing the group probabilities. For example, Quadrianto et al. [2] applied consistent estimators which could reconstruct the correct labels with high probability into a uniform convergence sense. Later, Rüping [3] proposed a parametric classifier, called IC-SVM, which integrated inverse calibration technology into support vector regression (SVR). Besides, Stolpe and Morik [4] developed a clustering based algorithm to learn from label proportions. Recently, Qi et al. [6] introduced an effective model called LLPs via nonparallel support vector machine (LLP-NPSVM). LLP-NPSVM determined the label of samples according to two nonparallel hyper-planes under the supervision of label proportion information. The current study on learning from group probabilities often assumes that the training set is large enough to train a robust classifier [3]. However, this assumption may not always hold. In practice, the data features or data distributions may be different, which leads to the lack of group probabilities since there are not enough groups in corresponding data. As a result, it may not be able to directly apply the classifiers learned on the learning tasks with group probabilities. It would be helpful if the samples of relative source domains can be transferred into the target domain. For the application of political election discussed above, the vote data in previous years will be helpful for the learning tasks of this year. In such cases, transfer learning between task domains would be desirable [7–11]. Transfer learning is motivated by the fact that people can apply previously learned knowledge to solve new problems. Transfer learning builds a model for the target domain by leveraging the label information from another related domain (source domain), such as the vote data collected in other time frames or with other data collection setups, thus it avoids the costly process of

* Corresponding author.
E-mail addresses: hbxtntg-12@163.com (T. Ni), czxqgu@163.com (X. Gu), zhengyh@vip.126.com (Y. Zheng), hywang@cczu.edu.cn (H. Wang).

http://dx.doi.org/10.1016/j.neucom.2017.08.049
0925-2312/© 2017 Published by Elsevier B.V.

Please cite this article as: T. Ni et al., Scalable transfer support vector machine with group probabilities, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.08.049

ARTICLE IN PRESS
JID: NEUCOM [m5G;September 7, 2017;23:50]

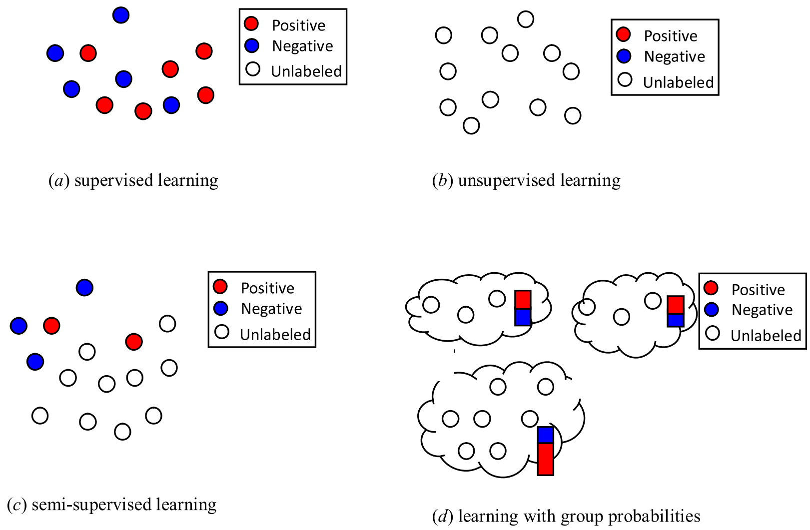

Fig. 1. Difference between (among) the proposed classifiers and the often-used learning algorithms (colors encoding class labels). (a): supervised learning: labeled data are given; (b): unsupervised learning: unlabeled data are given; (c): semi-supervised learning: labeled and unlabeled data are given; (d): learning with group probabilities: class label proportion of unlabeled data are given.

generating labels for the samples in the target domain. Different from conventional machine learning algorithms which assume that the training data should have the same distribution as that of the test data, transfer learning is able to utilize knowledge from data which follows a different distribution. Up to now, many transfer learning algorithms have been successfully applied in many areas, such as speech recognition, computer vision, information retrieval and natural language processing [12–15].

In this work, we construct a novel transfer learning framework with group probabilities using labeled samples in the source domain and the group probabilities in the target domain. Furthermore, inspired by Inverse Calibration [3], with the ε-insensitive loss function, a transfer support vector machine with group probabilities (called TSVM-GP) is proposed. TSVM-GP can be solved using the classical quadratic programming (QP) solver [16].

However, the number of samples in the source domain is usually very large and computing the corresponding kernel matrix using the naive QP solver is $O(n^3)$ (n is the number of training samples) computationally complex. All these factors heavily limit the applicability of TSVM-GP on large scale transfer datasets. Many endeavors have been made to develop various techniques for scaling up the QP solver. Typical techniques involve chunking and some other complicated decomposition methods such as sequential minimal optimization (SMO) algorithm [17], minimum enclosing ball (MEB) [18], core vector machines (CVM) [19], fastKDE [20] and so on. Recently, AESVM [21] is proposed as a fast SVM training algorithm, whose implementation is found to be much faster than many state-of-the-art kernel methods, whereas its classification accuracy is comparable to these existing algorithms. Based on the definition of the convex hull in the kernel space, AESVM effectively finds the convex hull vertices of the training data (called representative set) in the kernel space, then uses the representative set as the training data to build a SVM classifier. Obviously, the size of the training data is effectively decreased and the training time is much shorter.

In this paper, in order to scale TSVM-GP on large scale transfer datasets, we introduce the AESVM algorithm into TSVM-GP and develop a scalable transfer support vector machine with group probabilities (STSVM-GP). First, the convex hull vertices of samples (representative set) in the source domain and their corresponding weights are computed based on the idea of AESVM algorithm. Then, these selected samples and their weights as well as the group probabilities of the target domain are feed into TSVM-GP to build a classifier. Therefore, the scalable version of TSVM-GP, i.e., STSVM-GP, is proposed to implement the fast training algorithm on large scale transfer datasets.

The contribution of this work can be summarized as the following aspects.

(1) A novel transfer learning classifier TSVM-GP is proposed. TSVM-GP simultaneously utilizes both the labeled samples in the source domain and the group probabilities of the unlabeled samples in the target domain in a transfer learning framework. The training procedure of TSVM-GP can be equivalently transformed as a classical QP problem and it guarantees a global optimal solution.
(2) By using both the representative set selected from the source domain and group probabilities of the target domain as the training set, the scalable classifier STSVM-GP of TSVM-GP is developed for fast training on large scale transfer datasets. It guarantees that the selected samples can retain the greatest amount of information of data in the source domain. The performance of STSVM-GP is much closed to that of TSVM-GP but STSVM-GP consumes much less training time.
(3) Extensive experiments on synthetic and real-world datasets are conducted and the experimental results demonstrate that the proposed classifiers are at least comparable to several state-of-the-art algorithms in terms of classification accuracy; moreover, STVM-GP is very feasibility for large scale transfer datasets in terms of training time.

The rest of this paper is organized as follows. The related concepts of classic learning of group probabilities: Inverse Calibration and AESVM are reviewed in Section 2. In Section 3, the proposed classifier TSVM-GP is proposed. In Section 4, the proposed classifier STSVM-GP is proposed. The experimental results on synthetic and real-world datasets are reported in Section 5. Finally, conclusions and the potential of the proposed classifiers are given in the last section.

2. Related works

2.1. Inverse Calibration (IC)

Inverse Calibration [3] learns a classifier from group probabilities based on the idea of support vector regression (SVR) and