Neurocomputing 171 (2016) 149–155

Contents lists available at ScienceDirect

Neurocomputing

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

Towards a probabilistic semi-supervised Kernel Minimum Squared Error algorithm

Haitao Gan [a], Rui Huang [b,*], Zhizeng Luo [a], Yingle Fan [a], Farong Gao [a]

[a] School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China
[b] NEC Laboratories China, Beijing 100084, China

ARTICLE INFO

Article history:
Received 12 March 2015
Received in revised form
25 May 2015
Accepted 17 June 2015
Communicated by Liang Lin
Available online 27 June 2015

Keywords:
Semi-supervised learning
Kernel Minimum Squared Error
Laplacian regularization

ABSTRACT

Recently, semi-supervised learning has received much attention in data mining and machine learning, and a number of algorithms are proposed to discuss how to make good use of the unlabeled data. Some algorithms deal with the unlabeled data in an exact way, in which each unlabeled sample is assigned to one single class and then treated as a labeled sample. Other algorithms use the unlabeled data to regularize the objective function but do not explicitly model the influence of the unlabeled data towards different classes. In many applications, however, the unlabeled data may be ambiguous and belong to multiple classes with different probabilities. Based on this assumption, this paper presents a Probabilistic Laplacian-regularized Kernel Minimum Squared Error algorithm (named PrLapKMSE), in which the probabilities of the unlabeled data belonging to different classes are adaptively generated. "Adaptively" means that these probabilities are recalculated iteratively along with the reformulated objective function so that the unlabeled data may have increasingly accurate effects on the semi-supervised learning procedure. Experimental results on several simulated and real-world datasets illustrate the effectiveness of our algorithm.

1. Introduction

Machine learning has been widely used in various computer vision and pattern recognition applications, such as object detection and tracking [1–3], face recognition [4–6], speech recognition [7], handwritten digit recognition [8], and scene categorization [9,10]. Recently, semi-supervised learning [11,12] has received more and more attention, which investigates how to exploit the information of both labeled and unlabeled data to achieve better performance than supervised learning. How to use the information of unlabeled data is the core problem. Among the various semi-supervised learning methods, Self-training [13] and Co-training [14] are the widely used approaches where the unlabeled data with the highest confidence are labeled and added to the training set at each iteration. In these approaches, an unlabeled data sample is usually assigned to exactly one single class. However, when some unlabeled data lie in the overlapping region of two or more classes, it may be beneficial not to give the unlabeled data hard assignments during the learning process, especially in the early stage.

On the other hand, manifold regularization based methods have become popular, which exploit the intrinsic manifold structure of the labeled and unlabeled data without explicitly labeling the unlabeled data. Belkin et al. [15] proposed Laplacian Regularized Least Squares (LapRLS) and Laplacian Support Vector Machines (LapSVM), both of which employ Laplacian regularization to the labeled and unlabeled data. Due to the high computational efficiency of Kernel Minimum Squared Error (KMSE) in the training phase, Gan et al. [16] proposed a semi-supervised KMSE (LapKMSE) which incorporated the manifold structure of the labeled and unlabeled data in the objective function of KMSE. Cai et al. introduced a Semi-supervised Discriminant Analysis (SDA) [17] and Graph Regularized Nonnegative Matrix Factorization (GNMF) [18] where the intrinsic manifold structure is modeled through a p nearest neighbor graph. However, these methods only use the unlabeled data in the regularization terms but not the fidelity terms, that is, they do not explicitly model the influence of the unlabeled data towards different classes.

Based on the above observations, we propose a Probabilistic Laplacian-regularized KMSE (PrLapKMSE) algorithm in this paper. In particular, we assign class probabilities to each unlabeled sample and incorporate the probabilities into the objective function of the original LapKMSE. The class probabilities represent the degrees of each unlabeled sample belonging to the different classes. Note that here we interpret this ambiguity property of the samples with the Bayesian interpretation of probability. Another way is to use the fuzzy membership function in fuzzy logic. The distinction between the two interpretations is not critical in our derivation. The probabilities are adaptively computed to control the contributions of the unlabeled data to the

* Corresponding author.
E-mail address: huang-rui@nec.cn (R. Huang).

http://dx.doi.org/10.1016/j.neucom.2015.06.031
0925-2312/

learning of the classifier. Furthermore, we show in our derivation that the optimal probability function for the unlabeled data is essentially a probabilistic variant of the optimal decision function. We conduct extensive experiments on several simulated and real-world datasets to demonstrate the robustness and effectiveness of our method.

The rest of the paper is organized as follows: In Section 2, we briefly review the related work and necessary background. Section 3 explains our algorithm in detail. An illustrative example on simulated data is shown in Section 4 to highlight the advantage of our method, and extensive experiments are carried out in Section 5 on publicly available datasets with detailed analysis. Finally, we conclude the paper and discuss some future directions in Section 6.

2. Background

2.1. Naïve KMSE

Let $X=\{(x_1,y_1),\ldots,(x_l,y_l)\}$ be a training set of size l, where $x_i\in\mathbb{R}^D$ and $y_i\in\mathbb{R}$. For the binary classification problem, $y_i=-1$ if $x_i\in\omega_1$ and $y_i=1$ if $x_i\in\omega_2$. By using a nonlinear mapping function Φ, a training data sample x_i is transformed into a new feature space $\Phi(x_i)$ from the original data space. The task of KMSE is to build a linear model on the new features so that the outputs of the training data obtained by the linear model are equal to the labels

$$\Phi W=Y \tag{1}$$

where

$$\Phi=\begin{bmatrix}1 & \Phi(x_1)^T\\ \vdots & \vdots\\ 1 & \Phi(x_l)^T\end{bmatrix},\quad W=\begin{bmatrix}\alpha_0\\ w\end{bmatrix}\quad\text{and}\quad Y=[y_1,\ldots,y_l]^T$$

According to the reproducing Kernel theory [19,20], w can be expressed as

$$w=\sum_{i=1}^{l}\alpha_i\Phi(x_i) \tag{2}$$

By substituting (2) into Eq. (1), we can get

$$K\alpha=Y \tag{3}$$

where

$$K=\begin{bmatrix}1 & k(x_1,x_1) & \cdots & k(x_1,x_l)\\ \vdots & \vdots & \ddots & \vdots\\ 1 & k(x_l,x_1) & \cdots & k(x_l,x_l)\end{bmatrix}\quad\text{and}\quad\alpha=\begin{bmatrix}\alpha_0\\ \vdots\\ \alpha_l\end{bmatrix}$$

here K is the Gram matrix whose entries are $k(x_i,x_j)=(\Phi(x_i)\cdot\Phi(x_j))$.

The goal of KMSE is to find the optimal vector α by minimizing the objective function as follows:

$$\mathcal{J}_0(\alpha)=(Y-K\alpha)^T(Y-K\alpha) \tag{4}$$

By setting the derivative of $\mathcal{J}_0(\alpha)$ with respect to α to zero, we can obtain the solution:

$$\alpha^*=(K^TK)^{-1}K^TY \tag{5}$$

From Eq. (5), we can find that the dimension of α^* is $l+1$ and $Rank(K^TK)\le l$. In other words, K^TK is always singular. Consequently, the solution α^* is not unique. Regularization approaches, e.g., [21], are often used to deal with the ill-posed problem. The corresponding regularized objective function can be described as

$$\mathcal{J}_1(\alpha)=(Y-K\alpha)^T(Y-K\alpha)+\mu\alpha^T\alpha \tag{6}$$

where μ is the weight of the regularization term.

By minimizing the above objective function (6), we can obtain

$$\alpha^*=(K^TK+\mu I)^{-1}K^TY \tag{7}$$

where I is an identity matrix of size $(l+1)\times(l+1)$.

When the optimal weight coefficients α^* is obtained, the linear model of KMSE can be written as

$$f(x)=\sum_{i=1}^{l}\alpha_i^*k(x_i,x)+\alpha_0^* \tag{8}$$

In the testing phase, $x\in\omega_1$ if $f(x)<0$ and $x\in\omega_2$ if $f(x)>0$.

2.2. Laplacian-regularized KMSE

Given a data set $X=\{(x_1,y_1),\ldots,(x_l,y_l),x_{l+1},\ldots,x_n\}$ with l labeled data and $u=n-l$ unlabeled data, LapKMSE [16] incorporated the manifold structure of all the data into the objective function of KMSE. In order to exploit the manifold structure, Gan et al. introduced a Laplacian regularization term by using graph Laplacian. The Laplacian regularization term is defined as

$$\mathcal{R}=f^TLf \tag{9}$$

where L is the graph Laplacian, and $f=[f(x_1),\ldots,f(x_n)]^T$ is the output of the decision function on the labeled and unlabeled data. The graph Laplacian is defined as $L=D-S$, where D is a diagonal matrix whose entry $D_{ii}=\sum_j S_{ij}$ and the edge weight matrix $S=[S_{ij}]_{n\times n}$ can be defined as follows:

$$S_{ij}=\begin{cases}1 & \text{if } x_i\in N_p(x_j) \text{ or } x_j\in N_p(x_i)\\ 0 & \text{otherwise}\end{cases} \tag{10}$$

where $N_p(x_i)$ denotes the data sets of p nearest neighbors of x_i.

By integrating the regularization term Eq. (9) into the objective function (6), the objective function of LapKMSE can be given as

$$\mathcal{J}_r(\alpha)=(Y-GK\alpha)^T(Y-GK\alpha)+\gamma_A\alpha^T\alpha+\gamma_I\mathcal{R} \tag{11}$$

where

$$G=\begin{bmatrix}I_{l\times l} & \mathbf{0}_{l\times u}\\ \mathbf{0}_{u\times l} & \mathbf{0}_{u\times u}\end{bmatrix},\ Y=[y_1,\ldots,y_l,0,\ldots,0]^T,$$

$$K=\begin{bmatrix}1 & k(x_1,x_1) & \cdots & k(x_1,x_n)\\ \vdots & \vdots & \ddots & \vdots\\ 1 & k(x_n,x_1) & \cdots & k(x_n,x_n)\end{bmatrix}$$

According to the Representer Theorem [15], the solution can be written as

$$f(x)=\sum_{i=1}^{n}\alpha_i^*k(x_i,x)+\alpha_0^* \tag{12}$$

Substituting Eq. (12) into Eq. (11), the modified objective function becomes

$$\mathcal{J}_r(\alpha)=(Y-GK\alpha)^T(Y-GK\alpha)+\gamma_A\alpha^T\alpha+\gamma_I\alpha^TK^TLK\alpha \tag{13}$$

Setting the derivative of Eq. (13) with respect to α to zero,

$$(-GK)^T(Y-GK\alpha)+\gamma_A\alpha+\gamma_IK^TLK\alpha=0 \tag{14}$$

we can get

$$\alpha^*=((GK)^TGK+\gamma_AI+\gamma_IK^TLK)^{-1}(GK)^TY \tag{15}$$

3. Probabilistic Laplacian-regularized KMSE

In the last section, we reviewed LapKMSE, and in this section, we present our probabilistic extension.