Optik 125 (2014) 3524–3529

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Laplacian regularized kernel minimum squared error and its application to face recognition

Haitao Gan *

School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China

ARTICLE INFO

Article history:
Received 12 July 2013
Accepted 4 January 2014

Keywords:
Kernel minimum squared error
Semi-supervised learning
Laplacian regularization
Face recognition

ABSTRACT

Kernel minimum squared error (KMSE) has been receiving much attention in data mining and pattern recognition in recent years. Generally speaking, training a KMSE classifier, which is a kind of supervised learning, needs sufficient labeled examples. However, labeled examples are usually insufficient and unlabeled examples are abundant in real-world applications. In this paper, we introduce a semi-supervised KMSE algorithm, called *Laplacian regularized KMSE* (LapKMSE), which explicitly exploits the manifold structure. We construct a p nearest neighbor graph to model the manifold structure of labeled and unlabeled examples. Then, LapKMSE incorporates the structure information of labeled and unlabeled examples in the objective function of KMSE by adding a Laplacian regularization term. As a result, the labels of labeled and unlabeled examples vary smoothly along the geodesics on the manifold. Experimental results on several synthetic and real-world datasets illustrate the effectiveness of our algorithm. Finally our algorithm is applied to face recognition and achieves the comparable results compared to the other supervised and semi-supervised methods.

1. Introduction

In the last decades, kernel method has been receiving more and more attention in nonlinear classification and regression. A training example can be mapped into a high-dimensional feature space by using kernel trick satisfying the Mercer condition [1,2] and then a classifier can be trained in the new feature space. In most case, the kernel trick can achieve good generalization performance. Hence, many researchers have been studying the idea and various methods have been proposed, such as kernel minimum squared error (KMSE) [3], support vector machine (SVM) [4], least squares SVM (LS-SVM) [5], kernel principal component analysis (KPCA) [6], kernel Fisher discriminant analysis (KFDA) [7]. Among the above methods, KMSE has received many attention due to its higher computational efficiency in the training phase. However, the solution of KMSE is not stable and affects the generalization ability [8]. Ref. [9] has presented a novel solution method, which yields the unique solution, by maximizing the between-class geometric margin. And experimental results show the feasibility of the method. Ref. [10] proposed two versions of KMSE using different regularization terms and proved their relation to KFDA and LS-SVM.

Nevertheless, the performance of KMSE, which is a kind of supervised learning, relies on sufficient labeled examples to train a good classifier (*sufficient* usually means that the labeled examples can roughly represent the underlying structure of the entire feature space). In fact, labeled examples are usually insufficient while unlabeled data are *often* abundant in real world. Consequently, semi-supervised learning, which uses both labeled and unlabeled examples to train a classifier, has become an attractive researched topic. In semi-supervised learning, how to learn from unlabeled examples is still an open problem. One of the most used ways is manifold regularization. Ref. [11] proposed Laplacian regularized least squares (LapRLS) and Laplacian support vector machines (LapSVM) which both employ Laplacian regularization to learn from labeled and unlabeled examples. Ref. [12] introduced a semi-supervised discriminant analysis (SDA) where unlabeled examples are used to exploit the intrinsic manifold structure through a graph regularization. We refer the readers to some excellent surveys [13,14] for more details.

In this paper, we propose a semi-supervised KMSE algorithm, called Laplacian regularized KMSE (LapKMSE), which explicitly reveals the manifold structure of the labeled and unlabeled examples. Our basic intuition is that two examples are likely to be drawn from the same class if they are close on the manifold. In fact, the manifold is usually unknown. Hence, we construct a p nearest neighbor graph to model the manifold and employ

* Fax: +86 2787543594.
E-mail address: ght1102@gmail.com

http://dx.doi.org/10.1016/j.ijleo.2014.01.058
0030-4026/

graph Laplacian to incorporate the Laplacian regularized term in the objective function of KMSE. Based on this, the information of labeled and unlabeled examples are exploited by Laplacian regularization which smooths the labels of labeled and unlabeled examples along the geodesics on the manifold.

The rest of the paper is organized as follows: In Section 2, we briefly review the naïve KMSE. In Section 3, we describe our algorithm in detail. Section 4 presents the experimental results on several datasets and we will apply our algorithm to face recognition. Finally, we conclude the paper and discuss some future directions in Section 5.

2. 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 $x_i \in \omega_1$ or $y_i=1$ if $x_i \in \omega_2$. By using a nonlinear mapping function Φ, a training example is transformed into a new feature space $\Phi(x_i)$ from the original feature space. The task of KMSE is to build a linear model in the new feature. The outputs of the training examples 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 [4,7], one can note that w can be expressed as

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

By substituting Eq. (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 the matrix K is kernel matrix whose entry $k(x_i, x_j) = (\Phi(x_i) \cdot \Phi(x_j))$.

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

$$\alpha^* = (K^T K)^{-1} K^T Y \tag{5}$$

From Eq. (5), we can find that the dimension of α^* is $l+1$ and $Rank(K^TK) \leq l$. In other words, K^TK is always singular. Consequently, the solution α^* is not unique. In the last decades, the regularization approach [10] is 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 coefficient of the regularization term.

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

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

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 presented 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 $f(x) > 0$.

3. Method

In this section, we will discuss how to learn from labeled and unlabeled examples in KMSE.

3.1. Manifold regularization

Recall the standard learning framework. There is a probability distribution P on $X \times \mathbb{R}$ according to which examples are generated for function learning. Labeled examples are (x, y) pairs drawn according to P. Unlabeled examples are $x \in X$ generated according to the marginal distribution $\mathcal{P}_X$ of P. In many applications, the marginal distribution $\mathcal{P}_X$ is unknown. Related works show that there may be a relationship between the marginal and conditional distribution [11]. It is assumed that if two examples $x_1, x_2 \in X$ are *similar* in the intrinsic geometry of $\mathcal{P}_X$, then the conditional distribution $\mathcal{P}(y|x_1)$ and $\mathcal{P}(y|x_2)$ are *similar*. This is referred to as *manifold assumption* which is *often* used in semi-supervised learning [14].

Given a data set $X=\{(x_1, y_1), \ldots, (x_l, y_l), x_{l+1}, \ldots, x_n\}$ with l labeled examples and $u = n - l$ unlabeled examples. In order to exploit the manifold structure, Ref. [11] introduced a Laplacian regularization term by using graph Laplacian. The Laplacian regularization is defined as

$$\mathcal{R} = f^T L f \tag{9}$$

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

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

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

3.2. Laplacian regularized KMSE (LapKMSE)

In this section, we introduce Laplacian regularized KMSE (LapKMSE) which is extended from KMSE by incorporating Laplacian regularizer into the objective function of KMSE.

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

$$\mathcal{J}_r(\alpha) = (Y - GK\alpha)^T (Y - GK\alpha) + \gamma_A \alpha^T \alpha + \gamma_I \mathcal{R} \tag{10}$$

where

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

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