Expert Systems With Applications 45 (2016) 1–7

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Towards designing risk-based safe Laplacian Regularized Least Squares

Haitao Gan [a,*], Zhizeng Luo [a], Yao Sun [a], Xugang Xi [a], Nong Sang [b], Rui Huang [b]

[a] School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China
[b] School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China

ARTICLE INFO

Keywords:
Semi-supervised learning
Laplacian Regularized Least Squares
Safe mechanism
Risk degree

ABSTRACT

Recently, Safe Semi-Supervised Learning (S3L) has become an active topic in the Semi-Supervised Learning (SSL) field. In S3L, unlabeled data that may affect the performance of SSL both positively and negatively are exploited more safely through different risk-based strategies, and such S3L methods are expected to perform at least the same as the corresponding Supervised Learning (SL) methods. While the previously proposed S3L methods considered the risk of unlabeled data, they did not explicitly model the different risk degrees of unlabeled data on the learning procedure. Hence, we propose risk-based safe Laplacian Regularized Least Squares (RsLapRLS) by analyzing the different risk degrees of unlabeled data in this paper. Our motivation is that unlabeled data may be risky in SSL and the risk degrees are different. We assign different risk degrees to unlabeled data according to the different characteristics in supervised and semi-supervised learning. Then a risk-based tradeoff term between supervised and semi-supervised learning is integrated into the objective function of SSL. The role of risk degrees is to determine the way of exploiting the unlabeled data. Unlabeled data with large risk degrees should be exploited by SL and others by SSL. In particular, we employ Regularized Least Squares (RLS) and Laplacian RLS (LapRLS) for SL and SSL, respectively. Experimental results on several UCI and benchmark datasets show that the performance of our algorithm is never significantly inferior to RLS and LapRLS. In this way, our algorithm improves the practicability of SSL.

1. Introduction

Past decades have witnessed the success of Semi-Supervised Learning (SSL) (Zhu, 2005; Zhu & Goldberg, 2009) in the machine learning field and various tasks, such as object detection and tracking (Chen, Li, Su, Cao, & Ji, 2014; Grabner Helmut & Horst, 2008; Qi, Xu, Wang, & Song, 2011; Tan, Zhang, & Wang, 2011), image classification (Cao, He, & Huang, 2011; Gan, Sang, & Huang, 2014; Lu & Wang, 2015; Richarz, Vajda, Grzeszick, & Fink, 2014; Van Vaerenbergh, Santamaria, & Barbano, 2011), speech recognition (Tur, Hakkani-Tur, & Schapire, 2005; Varadarajan, Yu, Deng, & Acero, 2009), etc. SSL aims at exploiting the information of both labeled and unlabeled data and achieving better performance than Supervised Learning (SL). How to utilize the unlabeled data is the core problem. Generally speaking, SSL utilizes the following assumptions on the data space: (1) smoothness; (2) cluster; (3) manifold; and (4) disagreement. Many algorithms (Adankon and Cheriet, 2011; Belkin, Niyogi, and Sindhwani, 2006; Blum and Mitchell, 1998; Reddy, Shevade, and Murty, 2011; Zhou and Li, 2005) have been proposed and achieved the encouraging performance using one or more assumptions in many tasks. Among these assumptions, manifold regularization (Belkin et al., 2006; Gan, Sang, & Chen, 2013) based methods have received much attention which exploit the intrinsic manifold structure of both labeled and unlabeled data. Belkin et al. (2006) proposed Laplacian Regularized Least Squares (LapRLS) and Support Vector Machines (LapSVM) which both employed a Laplacian regularization term to learn from labeled and unlabeled data. Experimental results show that the manifold regularization technique can effectively exploit the information of unlabeled data.

Among these SSL methods, a common assumption is that all the unlabeled data are safe to be exploited. However, some literatures (Li & Zhou, 2011b; Wang & Chen, 2013; Yang & Priebe, 2011) show that the information of unlabeled data has the dual characteristics: (1) helpfulness; and (2) harmfulness. For a given SSL method, If the unlabeled data can improve the performance, they can be considered helpful. If the unlabeled data degenerate the performance, they can be considered harmful. Since different SSL methods utilized the different assumptions as mentioned above, the unlabeled data may have different characteristics in different SSL methods. When the employed assumption is not consistent with data distribution disclosed by the whole dataset, the unlabeled data may be harmful for learning in SSL. Some previous studies (Cohen, Cozman, Sebe, Cirelo, & Huang, 2004; Singh, Nowak, & Zhu, 2009; Yang & Priebe, 2011) have

* Corresponding author. Tel.: +86 571 86919130.
E-mail addresses: htgan@hdu.edu.cn (H. Gan), luo@hdu.edu.cn (Z. Luo), sunyao@hdu.edu.cn (Y. Sun), xixugang@hdu.edu.cn (X. Xi), nsang@hust.edu.cn (N. Sang), ruihuang@hust.edu.cn (R. Huang).

http://dx.doi.org/10.1016/j.eswa.2015.09.017
0957-4174/

discussed the impact of model assumption and unlabeled data on SSL in theory. Meanwhile other works (Chawla & Karakoulas, 2005; Gan et al., 2013) have investigated the effect of unlabeled data in empirical analysis. If the unlabeled data cannot be safely used, it will limit the scope of practical applications of SSL to some extent. Consequently, it is necessary to design a Safe Semi-Supervised Learning (S3L) method that never performs worse than the corresponding SL methods using only the labeled data.

To the best of our knowledge, there are only a few studies about S3L to this day. Li and Zhou (2011a) presented an S3VM_*us* method where a hierarchical clustering method is used to select the helpful unlabeled data. The selected helpful unlabeled data are then trained by transductive SVM (TSVM) (Joachims, 1999) and the remaining are trained by SVM. Hence, the probability of performance degeneration is much smaller than that of TSVM. Meanwhile, Li and Zhou (2011b) also proposed safe semi-supervised SVMs (S4VMs). Unlike S3VM which only tried to find one optimal low-density separator, S4VMs exploited the candidate low-density separators simultaneously to reduce the risk of identifying a poor separator with unlabeled data. The performance of S4VMs is never significantly inferior to that of SVM. In order to extend S4VMs for multi-class problems, Covoes, Barros, da Silva, Hruschka, and de Carvalho (2013) proposed a hierarchical bottom-up S4VMs tree scheme to take advantage of S4VMs. After that, Wang and Chen (2013) developed a safety-aware SSCCM (SA-SSCCM) which is extended from the semi-supervised classification method based on class membership (SSCCM). The safe mechanism is implemented through a tradeoff between least-square SVM (LS-SVM) and SSCCM. The performance of SA-SSCCM is never significantly inferior to that of LS-SVM and seldom significantly inferior to that of SSCCM. Recently, Kawakita and Takeuchi (2014) proposed a S3L method based on weighted likelihood which is expected to be safe in any situation. Experimental results on the regression and classification problems illustrated the effectiveness over SL although they did not compare the method with SSL.

Though the above-mentioned S3L methods considered the risks using unlabeled data, they did not explicitly take into account the different risk degrees of unlabeled data. In other words, it is an reasonable assumption that different unlabeled data should have different risk degrees. Hence, we propose a novel safe mechanism to design a risk-based S3L method. Our basic idea is to assign a risk degree to each unlabeled data. The risk degree is obtained through analyzing the behavior of unlabeled data in supervised and semi-supervised learning. When the unlabeled data may be helpful to train a semi-supervised classifier, the risk degree of the unlabeled data would be small. Otherwise, the risk degree would be large and we prefer to utilize SL to predict the labels of unlabeled data. Hence, the prediction of our algorithm is a tradeoff between those of supervised and semi-supervised learning. In particular, we employ Regularized Least Squares (RLS) and LapRLS for SL and SSL, respectively.

To sum up, the main contributions of our work are:

(1) The risk degree of unlabeled data is explicitly defined according to their characteristics in SL and SSL.
(2) The proposed safe mechanism can easily be applied to the other objective function-based SSL methods.
(3) The performance of RsLapRLS is never significantly inferior to that of RLS and LapRLS. And it is relatively stable with respect to the tradeoff parameter λ.

The rest of the paper is organized as follows: In Section 2, we firstly review the related work, including RLS and LapRLS. In Section 3, we will give the details of our algorithm. Section 4 includes a series of experiments on several UCI and benchmark datasets and analysis of the results. Finally, we will conclude the paper and present some future work in Section 5.

2. Related work

In this section, we mainly review the related work, including RLS and LapRLS.

2.1. Regularized Least Squares (RLS)

Given a dataset $X = \{(x_1, y_1), \ldots, (x_l, y_l)\}$ with size l, $x_i \in \mathbb{R}^D$ and $y_i \in \mathbb{R}$. For a two-class classification problem, $y_i = -1$ if $x_i \in \omega_1$ or $y_i = 1$ if $x_i \in \omega_2$. RLS aims to learn a decision function $f(x)$ by minimizing the objective function as follows:

$$\mathcal{J}(f) = \frac{1}{l}\sum_{i=1}^{l}(f(x_i) - y_i)^2 + \gamma\|f\|_K^2 \tag{1}$$

here, $\|\cdot\|_K$ is the norm defined in $\mathcal{H}_K$ which is a Reproducing Kernel Hilbert Space (RKHS) associated with a Mercer kernel $K : X \times X \to \mathbb{R}$.

Then the Gram matrix K can be calculated using a Mercer kernel:

$$\mathbf{K} = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_l) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_l, \mathbf{x}_1) & \cdots & k(\mathbf{x}_l, \mathbf{x}_l) \end{bmatrix} \tag{2}$$

According to the Representer Theorem, the solution can be given as (Belkin et al., 2006)

$$f(x) = \sum_{i=1}^{l}\alpha_i^* k(x_i, x) \tag{3}$$

where α^* denotes the optimal value of α.

By Substituting Eq. (3) into Eq. (1), the objective function can be rewritten as:

$$\mathcal{J}(\alpha) = \frac{1}{l}(K\alpha - Y)^T(K\alpha - Y) + \gamma\alpha^T K\alpha \tag{4}$$

where $\alpha = [\alpha_1, \ldots, \alpha_l]^T$, $Y = [y_1, \ldots, y_l]^T$, and K is the Gram matrix with size $l \times l$ whose entry $K_{ij} = K(x_i, x_j)$.

The derivative of Eq. (4) with respect to α is:

$$\frac{\partial \mathcal{J}}{\partial \alpha} = \frac{2}{l}(K\alpha - Y)^T K + 2\gamma K\alpha \tag{5}$$

By setting the derivative to zero, we can obtain the optimal solution:

$$\alpha^* = (K + \gamma lI)^{-1}Y \tag{6}$$

where I is an identity matrix with size $l \times l$.

2.2. Laplacian Regularized Least Squares (LapRLS)

Suppose we have a dataset $X = \{(x_1, y_1), \ldots, (x_l, y_l), x_{l+1}, \ldots, x_n\}$ with l labeled data and $u = n - l$ unlabeled data, LapRLS (Belkin et al., 2006) incorporated the geometry manifold structure of both labeled and unlabeled data into the objective function of RLS. To exploit the intrinsic geometrical structure, LapRLS introduced a regularization term defined on a graph Laplacian as:

$$\mathcal{R} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(f(x_i) - f(x_j))^2 W_{ij} = f^T L f \tag{7}$$

where $f = [f(x_1), \ldots, f(x_n)]^T$, and L is the graph Laplacian defined as $L = D - W$. Here D is a diagonal matrix whose entry $D_{ii} = \sum_{j=1}^{n} W_{ij}$ and W is an edge weight matrix. The edge weight matrix can be computed as follows:

$$W_{ij} = \begin{cases} \exp\left\{-\frac{\|x_i - x_j\|_2^2}{2\sigma^2}\right\} & \text{if } x_i \in N_p(x_j) \text{ or } x_j \in N_p(x_i) \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

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