Neurocomputing 147 (2015) 299–306

Contents lists available at ScienceDirect

Neurocomputing

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

Semi-supervised classification learning by discrimination-aware manifold regularization

Yunyun Wang [a,b], Songcan Chen [b,*], Hui Xue [c], Zhenyong Fu [a]

[a] Department of Computer Science and Engineering, Nanjing University of Posts & Telecommunications, Nanjing 210046, PR China
[b] Department of Computer Science and Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, PR China
[c] School of Computer Science and Engineering, Southeast University, Nanjing 210096, PR China

ARTICLE INFO

Article history:
Received 13 December 2013
Received in revised form
20 April 2014
Accepted 23 June 2014
Communicated by Feiping Nie
Available online 30 June 2014

Keywords:
Semi-supervised classification
Manifold regularization
Discrimination
Unsupervised clustering

ABSTRACT

Manifold regularization (MR) provides a powerful framework for semi-supervised classification (SSC) using both the labeled and unlabeled data. It first constructs a single Laplacian graph over the whole dataset for representing the manifold structure, and then enforces the smoothness constraint over such graph by a Laplacian regularizer in learning. However, the smoothness over such a single Laplacian graph may take the risk of ignoring the discrimination among boundary instances, which are very likely from different classes though highly close to each other on the manifold. To compensate for such deficiency, researches have already been devoted by taking into account the discrimination together with the smoothness in learning. However, those works are only confined to the discrimination of the labeled instances, thus rather limited in boosting the semi-supervised learning. To mitigate such an unfavorable situation, we attempt to discover the possible discrimination in the available instances first by performing some unsupervised clustering over the whole dataset, and then incorporate it into MR to develop a novel discrimination-aware manifold regularization (DAMR) framework. In DAMR, instances with high similarity on the manifold will be restricted to share the same class label if belonging to the same cluster, or to have different class labels, otherwise. Our empirical results show the competitiveness of DAMR compared to MR and its variants likewise incorporating the discrimination in learning.

1. Introduction

In many real applications, the unlabeled data can be easily and cheaply collected, while the acquisition of labeled data is usually quite expensive and time-consuming, especially involving manual effort. For instance, in web page recommendation, huge amounts of web pages are available, but few users are willing to spend time marking which web pages they are interested in. In spam email detection, a large number of emails can be automatically collected, yet few of them have been labeled spam or not by users. Consequently, semi-supervised learning, which exploits a large amount of unlabeled data jointly with the limited labeled data for learning, has attracted intensive attention during the past decades. In this paper, we focus on semi-supervised classification, and so far, lots of semi-supervised classification methods have been developed [1–4].

Generally, semi-supervised classification methods attempt to exploit the intrinsic data distribution information disclosed by the unlabeled data in learning, and the information is usually considered to be helpful for learning. To exploit the unlabeled data, some assumption should be adopted for learning. Two common assumptions in semi-supervised classification are the cluster assumption and the manifold assumption [3–5]. The former assumes that similar instances are likely to share the same class label, thus guides the classification boundary passing through the low density region between clusters. The latter assumes that the data are resided on some low dimensional manifold represented by a Laplacian graph, and similar instances should share similar classification outputs according to the graph. Almost all off-the-shelf semi-supervised classification methods adopt one or both of those assumptions explicitly or implicitly [1,4]. For instance, the large margin semi-supervised classification methods, such as transductive Support Vector Machine (TSVM) [6], semi-supervised SVM (S3VM) [7] and their variants [8,9], adopt the cluster assumption. The graph-based semi-supervised classification methods, such as label propagation [10,11], graph cuts [12] and manifold regularization (MR) [13], adopt the manifold assumption. Furthermore, there are also methods combining both assumptions for better performances, such as RegBoost [14] and SemiBoost [15], etc.

In this paper, we concentrate on the MR framework [13], which provides an effective way for semi-supervised classification [16], and has been applied in diverse applications such as image

* Corresponding author. Tel.: +86 25 84892956; fax: +86 25 84892811.
E-mail address: s.chen@nuaa.edu.cn (S. Chen).

http://dx.doi.org/10.1016/j.neucom.2014.06.059
0925-2312/

retrieval [17] and web spam identification [18], etc. At the same time, the manifold learning concept has also successfully applied in many other learning tasks including clustering [19], dimensionality reduction [20], and non-negative matrix factorization [21,22], etc.

MR for semi-supervised classification represents the manifold structure for the whole dataset by a single Laplacian graph, which is different from MR for supervised classification constructing the respective Laplacian graphs for individual classes, and then imposes the smoothness constraint over such a representation by a Laplacian regularizer in learning. However, the smoothness constraint imposed over a single Laplacian graph may take the risk of ignoring the discrimination among the boundary instances, which are very likely to belong to different classes though close over the manifold, consequently, MR may misclassify the boundary instances between clusters [16].

In fact, many researches have already been devoted to compensating for this deficiency by utilizing the dissimilarity or discrimination in the learning of MR. In [23], Andrew et al. considered both the label similarity and dissimilarity in learning, and developed a new dissimilarity encoded MR framework based on mixed graph. However, the dissimilarity should be given beforehand. In [24], Wang and Zhang constructed an unsupervised discriminative kernel based on discriminant analysis, and then used it to derive specific algorithms, including semi-supervised discriminative regularization (SSDR) and semi-parametric discriminative semi-supervised classification (SDSC). However, the derived methods do not fall into the methodology of manifold regularization. Recently in [16], Wu et al. incorporated linear discriminant analysis (LDA) and MR into a coherent framework and developed a semi-supervised discriminative regularization (SSDR). Specifically, the intra-class and inter-class graphs are constructed first in SSDR based on the labeled data, and then the corresponding intra-class compactness and inter-class separation are optimized simultaneously in the learning of MR. However, SSDR in [16] only utilize the discrimination of the labeled data, while the label information is usually rather limited in semi-supervised learning, consequently, its improvement over MR is not so distinct in the experiments.

In this paper, we attempt to incorporate the discrimination of both the labeled and unlabeled data into MR so as to develop a discrimination-aware MR framework for semi-supervised classification. In fact, due to the lack of label information in semi-supervised learning and thus the difficulty for formulating the discrimination of the whole data, SSDR in [16] only uses the discrimination of the labeled data, while the label instances are usually scarce in semi-supervised classification. For discovering the discrimination of all given data, we adopt the strategy of a pre-performed unsupervised clustering method as an example. Specifically, by performing some unsupervised clustering method such as FCM beforehand, we can get the within/between-cluster information of all instance pairs, which is much analog to the must/cannot-link information in semi-supervised clustering. Then we incorporate such information into MR such that for instances with high similarity over the manifold structure, they are restricted to share the same class label if belonging to the same cluster, or to have different class labels, otherwise. In this way, DAMR actually utilizes both the cluster and manifold assumptions in learning. It has been demonstrated by previous work that methods working on multiple data distribution assumptions can achieve better classification than those working on a single one [14,15], thus DAMR is able to be expected to perform better than MR.

The rest of this paper is organized as follows. Section 2 introduces the related works, Section 3 presents the proposed discrimination-aware manifold regularization framework, Section 4 presents a specific algorithm DA_LapRLSC through adopting the square loss function, Section 5 gives the empirical results, and some conclusions are drawn in Section 6.

2. Related works

2.1. Manifold regularization

Manifold assumption is one of the most commonly-used data distribution assumptions in semi-supervised learning [2,4]. Generally, the manifold structure is captured by an undirected graph according to some similarity measure, in which the vertices represent the instances and the edge-weights represent the similarities between instance pairs, and the manifold assumption assumes that similar instances over the manifold structure should share similar classification outputs. Lots of semi-supervised classification methods have been proposed based on the manifold assumption, mainly including the graph-based methods such as label propagation, graph cuts and manifold regularization, etc. Most graph-based methods, including label propagation and graph cuts, aim to learn only the class labels for the available unlabeled instances, thus learn in the transductive learning style [4]. However, many real applications actually need inductive methods for predicting unseen instances [15], and manifold regularization (MR) is exactly an inductive learning framework for semi-supervised classification based on the manifold assumption, which has been applied in diverse applications during the recent years.

Given labeled data $X_l=\{x_i\}_{i=1}^{l}$ with the corresponding labels $Y=\{y_i\}_{i=1}^{l}$, and unlabeled data $X_u=\{x_j\}_{j=l+1}^{n}$, where each $x_i \in R^d$ and $u=n-l$. $G=\{W_{ij}\}_{i,j=1}^{n}$ is a Laplacian graph over the whole dataset, where each weight W_{ij} represents the similarity between instances x_i and x_j. The Laplacian graph can be defined by many strategies such as the 0–1 weighting, i.e., $W_{ij}=1$ if and only if x_i and x_j are connected by an edge over the graph, the heat kernel weighting with $W_{ij}=e^{-\|x_i-x_j\|^2/\sigma}$ if x_i and x_j are connected, or the dot-product weighting with $W_{ij}=x_i^T x_j$ if x_i and x_j are connected.

Then with a decision function $f(x)$, the framework of MR can be formulated as

$$\min_f \frac{1}{l}\sum_{i=1}^{l} V(x_i,y_i,f)+\gamma_A\|f\|_K^2+\frac{\gamma_I}{2(l+u)^2}\sum_{i,j=1}^{l+u} W_{ij}(f(x_i)-f(x_j))^2 \tag{1}$$

where $V(x_i,y_i,f)$ is some loss function, such as the hinge loss max $[0,1-y_if(x_i)]$ for support vector machine (SVM) or the square loss $(y_i-f(x_i))^2$ for regularized least square classifier (RLSC), in this way, the MR framework naturally embodies the specific algorithms LapSVM and LapRLSC [13]. $\|f\|_K^2$ is a regularization term for smoothness in the Reproducing Kernel Hilbert Space (RKHS). The third term guarantees the prediction smoothness over the graph, which can be further written as

$$\frac{1}{2}\sum_{i,j=1}^{l+u} W_{ij}(f(x_i)-f(x_j))^2=\mathbf{f}^T\mathbf{L}\mathbf{f} \tag{2}$$

where $\mathbf{f}=[f(x_1),\ldots,f(x_{l+u})]^T$, and $\mathbf{L}$ is the graph Laplacian given by $\mathbf{L}=\mathbf{D}-\mathbf{W}$, $\mathbf{W}$ is the weight matrix of graph G and $\mathbf{D}$ is a diagonal matrix with the diagonal component given by $\mathbf{D}_{ii}=\sum_{j=1}^{n}\mathbf{W}_{ij}$. According to the Representer theorem [13], the minimizer of problem (1) has the form

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

where K: $X\times X\rightarrow R$ is a Mercer kernel (the bias of the decision function can be omitted by augmenting each instance with an 1-valued element).

It is clear that in MR, if instances x_i and x_j are similar in terms of W_{ij}, then it is restricted that their class labels are similar as well. *Such a smoothness restriction is also imposed on the boundary instance pairs, however, instance pairs in the boundary area are very*