ARTICLE IN PRESS

Information Sciences xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Information Sciences

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

Graph-based semi-supervised learning by mixed label propagation with a soft constraint

Xiaolan Liu [a,*], Shaohua Pan [a], Zhifeng Hao [b], Zhiyong Lin [c]

[a] School of Science, South China University of Technology, Guangzhou 510640, China
[b] Faculty of Computer, Guangdong University of Technology, Guangzhou 510006, China
[c] School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou 510665, China

ARTICLE INFO

Article history:
Received 3 September 2009
Received in revised form 16 June 2010
Accepted 11 February 2014
Available online xxxx

Keywords:
Semi-supervised learning
Graph
Dissimilarity
Fractional quadratic program
Collaborative filtering

ABSTRACT

In recent years, various graph-based algorithms have been proposed for semi-supervised learning, where labeled and unlabeled examples are regarded as vertices in a weighted graph, and similarity between examples is encoded by the weight of edges. However, most of these methods cannot be used to deal with dissimilarity or negative similarity. In this paper we propose a mixed label propagation model with a single soft constraint which can effectively handle positive similarity and negative similarity simultaneously, as well as allow the labeled data to be relabeled. Specifically, the soft mixed label propagation model is a fractional quadratic programming problem with a single quadratic constraint, and we apply the global optimal algorithm [1] for solving it, yielding an ϵ-global optimal solution in a computational effort of $O(n^3 \log \epsilon^{-1})$. Numerical comparisons with several existing methods for common test datasets and a class of collaborative filtering problems verify the effectiveness of the method.

1. Introduction

It is well known that supervised learning often requires a great amount of labeled examples in order to establish a good classifier. In many practical applications, it is difficult to collect enough labeled examples due to time consuming and expensive human labor; by contrast, unlabeled examples are rather easier to access; for example, in web search one may obtain a lot of unlabeled webs by crawling the web. Thus, semi-supervised learning, which may exploit a large quantity of unlabeled data but few labeled examples to construct a classifier, becomes an important way in machine learning.

Recently, graph-based semi-supervised learning has gained active research; see, e.g., [2–4,6,11–13,16]. Its basic idea is to construct a graph $G = (v, \varepsilon)$ with all data points and nonnegative weights S_{ij} on edge $e_{ij} \in \varepsilon$ to characterize the similarity between examples x_i and x_j, and then propagate labels from labeled data to unlabeled data through the pairwise similarity. However, as will be shown in the next section, most of existing graph-based semi-supervised learning methods have difficulty in dealing with dissimilarity, or negative similarity. Note that dissimilarity, or negative similarity frequently appears in many practical applications. For example, in collaborative filtering problems, two users who have different interests will have opposite ratings to the same thing, and in this case negative correlation coefficient (negative similarity) is often adopted to measure the similarity between the two users.

* Corresponding author.
E-mail address: liuxl@scut.edu.cn (X. Liu).

http://dx.doi.org/10.1016/j.ins.2014.02.067
0020-0255/

To handle dissimilarity in graph-based semi-supervised learning, Goldberg et al. [5] first proposed a semi-supervised classification algorithm by a graph-based encoding for dissimilarity. Later, Tong and Rong [10] proposed a mixed label propagation model which can effectively deal with similarity and dissimilarity simultaneously. We observe that in Tong et al. 's model, the labels of those points to be labeled are mandatorily required to be same as those of the labeled data. When the labeled data is contaminated or the available labels are with noise, this will inevitably bring negative effects on the correct classification. Taking into account that the labeled data from practice is often contaminated, in this paper we relax the hard requirements for the labels, and establish a mixed label propagation model with a soft constraint. More specifically, this model is a fractional quadratic program with a single quadratic constraint. In addition, in [10] Tong et al. 's model is relaxed to a semidefinite programming problem with $(nc)*(uc)+1$ variables, and then the interior point method software SeduMi [9] is applied for solving it, where n is the size of dataset, u is the number of unlabeled data and c is the total number of classes. Since a system of linear equations is required to solve at each iteration of interior point methods, and the reformulated semidefinite programming problem has a much larger scale than the original model, the method in [10] is not suitable for those large-scale problems such as collaborative filtering problems.

In this paper, we apply the global optimal algorithm [1] for solving our model. This method may yield an ϵ-global optimal solution within a computational effort of $O(n^3 \log \epsilon^{-1})$, though our model is a fractional quadratic programming with a quadratic constraint, instead of a convex programming model. Numerical comparisons with the mixed label propagation method in [10] for five test datasets indicate that our method does not only have better performance in accuracy, but also requires much less running time. In addition, numerical comparisons with the label propagation approach [15], the local and global consistency method [14] and Pearson correlation coefficient method [7] for a class of collaborative filtering problems show that in most cases our method yields the lowest mean absolute error (MAE). This means that our method has the best performance in terms of accuracy, although it requires more running time.

The paper is organized as follows. In Section 2, we propose a mixed label propagation model with a soft constraint for binary classification problems (BCPs). In Section 3, we extend the binary classification model to a multiclass classification model which is a fractional quadratic program with a single quadratic constraint. Section 4 describes the specific iteration steps of global optimization algorithm [1] for solving a fractional quadratic program of Section 3. Section 5 reports numerical experiment results. Finally, we conclude this paper with two future research direction.

Throughout this paper, unless otherwise stated, all vectors are column ones and $^{\mathrm{T}}$ denotes the transpose of a vector or a matrix. The notations $\mathbb{R}^n$ represents the n-dimensional Euclidean space, I_m and $\mathbf{0}_{m\times m}$ means an $m\times m$ identity matrix and zero matrix, respectively. Given $c_1,\ldots,c_n\in\mathbb{R}$, we denote $\mathrm{diag}(c_1,\ldots,c_n)$ by an $n\times n$ diagonal matrix with $c_1,\ldots,c_n$ as diagonal entries. Similarly, given a group of square matrices $C_1,\ldots,C_n\in\mathbb{R}^{l\times l}$, $\mathrm{diag}(C_1,\ldots,C_n)$ means a block diagonal matrix with C_i being the ith block.

2. Mixed label propagation with a soft constraint for BCPs

Assume that the given data set X consists of n points $x_1,\ldots,x_n\in\mathbb{R}^d$, and the first l points $x_1,\ldots,x_l$ are labeled as $y_1,\ldots,y_l\in\{-1,1\}$, and the rest $x_{l+1},\ldots,x_n$ are unlabeled. Our goal is to predict the labels of $x_{l+1},\ldots,x_n$.

As mentioned in the introduction, graph-based semi-supervised classification methods construct a graph $G=(v,\varepsilon)$ with all data points $x_1,\ldots,x_n$ and nonnegative weights S_{ij} on edge $e_{ij}\in\varepsilon$ to reflect the similarity between x_i and x_j. If x_i and x_j has better similarity, then the weight S_{ij} should have a larger value so as to reflect that they tend to have similar labels. In mathematics, this can be achieved by introducing the following penalty term

$$E(S,f) := \frac{1}{2}\sum_{i,j=1}^{n} S_{ij}\big(f(x_i)-f(x_j)\big)^2 = f^T L f \tag{1}$$

into the discriminant function $f: X\mapsto\mathbb{R}$, where $L=D-S$ is the Laplacian matrix with $D=\mathrm{diag}(D_1,\ldots,D_n)$ and $D_i=\sum_{j=1}^{n}S_{ij}$. As pointed out in [5], if S_{ij} is very large, minimizing $E(S,f)$ indeed may lead to $f(x_i)\approx f(x_j)$; but when S_{ij} is very small or tends to zero which reflects that x_i and x_j tend to have different labels, minimizing $E(S,f)$ cannot guarantee $f(x_i)\neq f(x_j)$. In other words, when using $E(S,f)$, a zero or small edge weight does not yield preference. A negative weight $S_{ij}<0$ can embody the difference between x_i and x_j, but it will cause that $E(S,f)$ becomes unbounded below, and thereby the corresponding optimization problem has no optimal solution.

To overcome the shortcoming of penalty term $E(S,f)$, Goldberg et al. [5] incorporated both similarity and dissimilarity by introducing an auxiliary matrix W, where $W_{ij}=-1$ if there is a dissimilarity edge between x_i and x_j, and $W_{ij}=1$ if there is a similarity edge. Thus, the resulting penalty term is given by

$$E(S,W,f) := \frac{1}{2}\sum_{i,j=1}^{n} S_{ij}\big(f(x_i)-W_{ij}f(x_j)\big)^2. \tag{2}$$

The major advantage of this model is a convex programming problem. However, it probably specifies the relation between the discriminant function f at dissimilarity samples x_i and x_j more than necessary. For example, the minimization of $E(S,W,f)$ will give $f(x_i)=-f(x_j)$, instead of the opposite sign of $f(x_i)$ and $f(x_j)$ which is precisely needed by the users.

Please cite this article in press as: X. Liu et al., Graph-based semi-supervised learning by mixed label propagation with a soft constraint, Inform. Sci. (2014), http://dx.doi.org/10.1016/j.ins.2014.02.067