Signal Processing 142 (2018) 188–193

Contents lists available at ScienceDirect

Signal Processing

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

Semi-supervised graph-based retargeted least squares regression

Haoliang Yuan[a,*], Junjie Zheng[a], Loi Lei Lai[a], Yuan Yan Tang[b]

[a] School of Automation, Guangdong University of Technology, Guangzhou 510006, China
[b] Department of Computer and Information Science, University of Macau, Macau 999078, China

ARTICLE INFO

Article history:
Received 17 April 2017
Revised 20 July 2017
Accepted 22 July 2017
Available online 24 July 2017

Keywords:
Graph learning
Retargeted least squares regression (ReLSR)
Multicategory classification

ABSTRACT

In this paper, we propose a semi-supervised graph-based retargeted least squares regression model (SSGReLSR) for multicategory classification. The main motivation behind SSGReLSR is to utilize a graph regularization to restrict the regression labels of ReLSR, such that similar samples should have similar regression labels. However, in SSGReLSR, constructing the graph structure and learning the regression matrix are two independent processes, which can't guarantee an overall optimum. To overcome this shortage of SSGReLSR, we also propose a semi-supervised graph learning retargeted least squares regression model (SSGLReLSR), where linear squares regression and graph construction are unified into a same framework to achieve an overall optimum. To optimize our proposed SSGLReLSR, an efficient iteration algorithm is proposed. Extensive experiments results confirm the effectiveness of our proposed methods.

1. Introduction

Multicategory classification, where the aim is to assign a test sample into one of several classes, is an important task in the fields of computer vision and pattern recognition [1–8].

Least squares regressions (LSR) is a typical and fundamental technique for multicategory classification. First of all, given a labeled sample matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^T \in \mathbb{R}^{n\times d}$, where $\mathbf{x}_i \in \mathbb{R}^{d\times 1}$ is a labeled sample, letting $y_i \in \{1, 2, \cdots, c\}$ be the class label of $\mathbf{x}_i$ and c is the total number of classes. The aim of LSR is to learn a regression matrix $\mathbf{W} \in \mathbb{R}^{d\times c}$ to represent the label matrix $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n]^T \in \mathbb{R}^{n\times c}$, where $\mathbf{y}_i$ is a label vector with 0 or 1 for the $\mathbf{x}_i$. If $\mathbf{x}_i$ belongs to the class j, its label is $\mathbf{y}_i = [0, \ldots, 0, 1, 0, \ldots, 0]$ with the jth element being equal to 1. As the following, the objective function of LSR is formulated as:

$$\min_{\mathbf{W}} \|\mathbf{XW} - \mathbf{Y}\|_F^2 + \lambda\|\mathbf{W}\|_F^2 \tag{1}$$

where λ is a regularization parameter. The closed-form solution of LSR is represented as $\mathbf{W} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{Y}$, where $\mathbf{I} \in \mathbb{R}^{d\times d}$ is the identity matrix. Since LSR has simple yet effective formulation and is easy to obtain the closed-form solution, many variations have been proposed to improve the classification performance of LSR, such as kernel ridge regression [9], partial LSR [10], weighted LSR [11], LASSO regression [12], and nonnegative least squares [13]. To further utilize the class information, some discriminative LSR methods have been proposed to enhance the learning performance of LSR. For example, Xiang et al. [14] propose a discriminative least squares regression (DLSR) model, which introduces a ϵ-dragging technique to yield the soft labels to enlarge the distance of the regression labels of different classes. Zhang et al. [15] present a framework of retargeted least squares regression (ReLSR), which directly learns the regression targets from data other than using the traditional zero-one matrix as regression targets. The objective function of ReLSR is formulated as:

$$\begin{aligned} &\min_{\mathbf{W},\mathbf{T}} \|\mathbf{XW} - \mathbf{T}\|_F^2 + \lambda\|\mathbf{W}\|_F^2 \\ &\text{s.t. } T_{i,y_i} - \max_{j\neq y_i} T_{i,j} \geq 1, \quad i = 1, 2, \ldots, n \end{aligned} \tag{2}$$

where the constraint of target matrix $\mathbf{T}$ is to guarantee that each labeled sample is correctly classified with the large margin, i.e., the margin between correct and incorrect classes should be larger than 1.

Although these discriminative LSR methods improve the robustness and effectiveness of LSR, they only utilize the labeled samples. In real applications, there exist lots of unlabeled samples to help enhancing the classification performance of the existing regression methods. Therefore, in this paper, we propose a semi-supervised graph-based retargeted least squares regression model (SSGReLSR), which utilizes the graph structure of labeled and unlabeled samples to restrict the regression labels of ReLSR for classification. However, in SSGReLSR, constructing the graph structure and learning the regression matrix are two independent processes, which can't guarantee an overall optimum. To overcome this shortage of SSGReLSR, we also propose a semi-supervised graph learning retargeted least squares regression method (SSGLReLSR), where

* Corresponding author.
E-mail address: hunteryuan@126.com (H. Yuan).

http://dx.doi.org/10.1016/j.sigpro.2017.07.027
0165-1684/

the processes of linear squares regression and graph construction are unified into a same framework so as to achieve an overall optimum. SSGLRELSR can be solved efficiently by an alternating optimization algorithm with guaranteed convergence.

The rest of the paper is organized as follows. In Section 2, we first introduce the mathematical model of SSGReLSR and SSGLReLSR, and then design an efficient iteration algorithm to solve the SSGLReLSR model. Experimental results are conducted and discussed in Section 3. Finally, the conclusion of this paper is drawn in Section 4.

2. Proposed methods

2.1. SSGReLSR

Aforementioned ReLSR model only utilizes the information of labeled samples to learn the regression matrix, which belongs to the supervised learning framework. However, in the real applications, they also exist lots of unlabeled samples, which can help to improve the learning performance of ReLSR. Letting $\mathbf{Z}=[\mathbf{X},\mathbf{U}]=[\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_m]\in\mathbb{R}^{m\times d}$ be a sample matrix, where $\mathbf{X}=[\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_n]\in\mathbb{R}^{n\times d}$ is the labeled sample matrix and $\mathbf{U}=[\mathbf{z}_{n+1},\mathbf{z}_{n+2},\ldots,\mathbf{z}_m]\in\mathbb{R}^{(m-n)\times d}$ is the unlabeled sample matrix. To fully utilize these labeled and unlabeled samples, a reasonable assumption here is that similar samples $\mathbf{z}_i$ and $\mathbf{z}_j$ should have similar regression label vectors of these two samples, i.e., $\mathbf{W}^T\mathbf{z}_i$ and $\mathbf{W}^T\mathbf{z}_j$. Based on the idea from manifold learning [16–18], the class compactness graph is proposed to address this problem. The core idea is that the similar samples should be kept close together in the regression label space. This assumption is usually referred to as manifold assumption.

Given a set of d-dimensional samples $\mathbf{z}_1,\ldots,\mathbf{z}_m$, we can construct a graph $\mathcal{G}$ with m vertices, where each vertex represents a sample. Let $\mathbf{S}\in\mathbb{R}^{m\times m}$ be the weight matrix of $\mathcal{G}$. In this paper, we adopt two common methods, i.e., k-nearest neighbors method and sparse representation method, to yield $\mathbf{S}$ as:

- *k-nearest neighbors*: if $\mathbf{z}_i$ is within the k nearest neighbors of $\mathbf{z}_j$ or $\mathbf{z}_j$ is within the k nearest neighbors of $\mathbf{z}_i$, $S_{i,j}=e^{-\frac{\|\mathbf{z}_i-\mathbf{z}_j\|_2}{2\sigma^2}}$, otherwise, $S_{i,j}=0$.
- *Sparse representation*: For sample matrix $\mathbf{Z}$, its sparse matrix $\mathbf{A}$ is achieved by solving a nonnegative ℓ_1-norm optimization problem:

  $$\min_{\mathbf{S}}\|\mathbf{Z}^T-\mathbf{Z}^T\mathbf{A}\|_F^2+\alpha\|\mathbf{A}\|_1$$
  $$\text{s.t. } \mathbf{A}\geq 0,\quad A_{i,i}=0,\quad i=1,2,\ldots,m.$$

  where α is a regularization parameter and $A_{i,i}=0$ is used to avoid the trivial solution. After obtaining the sparse matrix $\mathbf{A}$, the weight matrix of $\mathbf{Z}$ is defined as $\mathbf{S}=\frac{\mathbf{A}+\mathbf{A}^T}{2}$.

Consider the problem of mapping the weighted graph $\mathcal{G}$ to the regression label matrix $\mathbf{ZW}$, the objective function of graph regularization is formulated as:

$$\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\|\mathbf{W}^T\mathbf{z}_i-\mathbf{W}^T\mathbf{z}_j\|_2^2S_{i,j}=\mathrm{Tr}(\mathbf{W}^T\mathbf{Z}^T\mathbf{LZW}) \tag{3}$$

where $\mathbf{L}=\mathbf{D}-\mathbf{S}\in\mathbb{R}^{m\times m}$ is the graph Laplacian and $\mathbf{D}$ is a diagonal matrix and its diagonal entries are defined as $D_{i,i}=\sum_{j=1}^{m}S_{i,j}$, $i=1,2,\ldots,m$.

By incorporating the graph constraint (3) into ReLSR, the objective function of SSGReLSR is formulated as:

$$\min_{\mathbf{W},\mathbf{T}}\|\mathbf{XW}-\mathbf{T}\|_F^2+\lambda\|\mathbf{W}\|_F^2+\lambda_1\mathrm{Tr}(\mathbf{W}^T\mathbf{Z}^T\mathbf{LZW})$$
$$\text{s.t. } T_{i,y_i}-\max_{j\neq y_i}T_{i,j}\geq 1,\quad i=1,2,\ldots,n \tag{4}$$

where λ_1 is a regularization parameter.

2.2. SSGLReLSR

For SSGReLSR, it predefines the graph structure and then uses it as a regularization term to learn the regression matrix $\mathbf{W}$, which ignores the relationships between these two processes and can't guarantee to obtain an overall optimum. To overcome this problem, we propose a semi-supervised graph learning ReLSR (SSGLReLSR) model, where the processes of graph construction and linear regression are unified into a same framework. The objective function of SSGLReLSR is formulated as:

$$\min_{\mathbf{W},\mathbf{T},\mathbf{S}}\|\mathbf{XW}-\mathbf{T}\|_F^2+\lambda\|\mathbf{W}\|_F^2+\lambda_1\sum_{i=1}^{m}\sum_{j=1}^{m}\|\mathbf{W}^T\mathbf{z}_i-\mathbf{W}^T\mathbf{z}_j\|_2^2S_{i,j}$$
$$+\|\mathbf{Z}^T-\mathbf{Z}^T\mathbf{S}\|_F^2$$
$$\text{s.t. } \mathbf{S}\geq 0,\quad S_{i,i}=0,\quad T_{i,y_i}-\max_{j\neq y_i}T_{i,j}\geq 1,\quad i=1,2,\ldots,n, \tag{5}$$

The first and second terms in (5) denote the ReLSR model. The third term $\|\mathbf{W}^T\mathbf{z}_i-\mathbf{W}^T\mathbf{z}_j\|_2^2S_{i,j}$ builds a relationship between $\mathbf{W}$ and $\mathbf{S}$, which allows the similar samples to have nearly similar regression label vectors. The fourth term in (5) is to proceed the linear reconstruction. The contribution of one sample to reconstruct another sample is a good indicator of similarity between these two samples [19]. Hence the reconstruction matrix $\mathbf{S}$ can be used to represent the weight matrix. Comparing with SSGReLSR, the regression matrix $\mathbf{W}$ and the weight matrix $\mathbf{S}$ in SSGLReLSR are optimized in the learning process simultaneously, which leads to an overall optimum. As the following, we design an effective iteration algorithm to solve the optimization problem of SSGLReLSR.

2.2.1. Computing $\boldsymbol{W}$

To compute $\mathbf{W}$, we need to fix $\mathbf{T}$ and $\mathbf{S}$. Then, the optimization problem of Eq. (5) is rewritten as:

$$\min_{\mathbf{W}}\|\mathbf{XW}-\mathbf{T}\|_F^2+\lambda\|\mathbf{W}\|_F^2+\lambda_1\sum_{i=1}^{m}\sum_{j=1}^{m}\|\mathbf{W}^T\mathbf{z}_i-\mathbf{W}^T\mathbf{z}_j\|_2^2S_{i,j} \tag{6}$$

Then, the objective function of model (6) can be rewritten as:

$$\min_{\mathbf{W}}\|\mathbf{XW}-\mathbf{T}\|_F^2+\lambda\|\mathbf{W}\|_F^2+\lambda_1\mathrm{Tr}(\mathbf{W}^T\mathbf{Z}^T\mathbf{LZW}) \tag{7}$$

where $\mathbf{L}=\mathbf{D}-(\mathbf{S}+\mathbf{S}^T)$ and $\mathbf{D}$ is a diagonal matrix and its diagonal entries are defined as $D_{i,i}=\sum_{j=1}^{m}(S_{i,j}+S_{j,i})$, $i=1,2,\ldots,m$. If the derivative of model (7) with respect to $\mathbf{W}$ is set equal to zero, we can obtain:

$$\mathbf{W}=(\mathbf{X}^T\mathbf{X}+\lambda\mathbf{I}+\lambda_1\mathbf{Z}^T\mathbf{LZ})^{-1}\mathbf{X}^T\mathbf{T} \tag{8}$$

where $\mathbf{I}\in\mathbb{R}^{d\times d}$ is an identity matrix.

2.2.2. Computing $\boldsymbol{T}$

To compute $\mathbf{T}$, we need to fix $\mathbf{W}$. Then, the optimization problem of SSGLReLSR is rewritten as:

$$\min_{\mathbf{T}}\|\mathbf{XW}-\mathbf{T}\|_F^2$$
$$\text{s.t. } T_{i,y_i}-\max_{j\neq y_i}T_{i,j}\geq 1,\quad i=1,2,\ldots,n \tag{9}$$

which is the ReLSR model. The optimal target matrix $\mathbf{T}$ in (9) is solved by Algorithm 1 (i.e., Algorithm 2 in [15]).

2.2.3. Computing $\boldsymbol{S}$

To compute $\mathbf{S}$, we need to fix $\mathbf{W}$. Then, the optimization problem of SSGLReLSR is rewritten as:

$$\min_{\mathbf{s}_i}\|\mathbf{Z}-\mathbf{ZS}\|_F^2+\lambda_1\sum_{i=1}^{m}\sum_{j=1}^{m}\|\mathbf{W}^T\mathbf{z}_i-\mathbf{W}^T\mathbf{z}_j\|_2^2S_{i,j}$$
$$\text{s.t. } \mathbf{S}\geq 0,\quad S_{i,i}=0,\quad i=1,2,\ldots,m \tag{10}$$