Pattern Recognition Letters 80 (2016) 58–63

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Manifold regularized matrix completion for multilabel classification☆

Bin Liu[a,b], Zenglin Xu[a,b,*], Shuang Wu[c], Fei Wang[d]

[a] School of Computer Science & Engineering, University of Electronic Science and Technology of China, 2006 Xiyuan Avenue, West Hi-tech Zone, Chengdu 611731, China
[b] Big Data Research Center, University of Electronic Science and Technology of China, 2006 Xiyuan Avenue, West Hi-tech Zone, Chengdu 611731, China
[c] Zhejiang University, 388 Yuhangtang Road, Hangzhou 310058, China
[d] University of Connecticut, 2131 Hillside Road, Unit 3088 Storrs, CT 06269-3088, United States

ARTICLE INFO

Article history:
Received 24 November 2015
Available online 19 May 2016

Keywords:
Multi-label learning
Matrix completion
Manifold regularization

ABSTRACT

Multilabel learning is an important research problem arising in a number of practical applications from diverse fields. Recent studies on multilabel learning have suggested the approach of matrix completion as a novel and promising approach to transductive multilabel learning. Here the missing labels of test data are regarded as missing values from the construction matrix composed of feature-by-item and label-by-item matrices. With the assumption of the low rank of the construction matrix, by minimizing its rank under the constraints of observed data and labels, we can recover all the missing labels. Despite its success, however, naive matrix completion methods ignore the smoothness assumption of the large amount of unlabel data, i.e., similar data should share similar labels, which may under exploit the intrinsic structure of data. To this end, we propose to solve the multi-label learning problem as an enhanced matrix completion problem with manifold regularization, where the graph Laplacian is used to ensuring the label smoothness over the label space. The resulting nuclear norm minimization problem is solved with a modified fixed-point continuation method that is guaranteed to find the global optimum. Experiments on both synthetic and real-world data have shown the promising results of the proposed approach.

1. Introduction

Multilabel learning is an important research problem in a number of real world applications, such as automatic image annotation and text categorization, where each example of data can be assigned to a set of labels, simultaneously. For example, in natural language processing, a wikipedia term “Albert Einstein” covers 54 categories (such as “Nobel laureates in Physics”, “Jewish physicists”, “American physicists” and so on); in bioinformatics, most genes are associated with more than one functional classes (e.g., metabolism, transcription and protein synthesis); in automatic image annotation, each image can also be assigned with a number of tags (e.g., sunset, sea, water, and hill).

During the past decades, many methods have been proposed to learn from multi-label data. According to a recent survey by Zhang and Zhou [21], most of existing methods fall into two categories – *problem transformation methods* and *algorithm adaptation methods*. Problem transformation methods try to transform multilabel learning into an existing and well-established learning architecture, such as binary classification or label ranking problems. Typical examples of this kind of methods include the works of Boutell et al. [2], Read et al. [18] and Fürnkranz et al. [11]. However, this kind of approaches usually ignores the underlying correlations among the multiple different labels, which could be an important hint for deciding the class memberships. While algorithm adaptation methods try to adapt popular learning techniques to deal with multi-label data, such as ML-*k*NN [20] and Rank-SVM [10], which may be reliable when the amount of training instances is sufficient and the number of classes is relatively small.

In contrast, real-world applications of multi-label learning usually contain a large number of classes and a relatively small size of training data. As a result, the amount of training data related to each class is usually small and insufficient for learning a reliable classifier. In addition, part of the labels associated with some examples may be unobserved and some instances may even be completely unlabeled. Take automatic image annotation as an example, digital images uploaded by users to the Internet usually have multiple semantic meanings, however most of them are only labeled with part of semantic meaning or unlabeled.

☆ This paper has been recommended for acceptance by Eckart Michaelsen.

* Corresponding author at: School of Computer Science & Engineering, University of Electronic Science and Technology of China, 2006 Xiyuan Avenue, West Hi-tech Zone, Chengdu 611731, China. Tel.: +862861830885.
E-mail address: zlxu@uestc.edu.cn (Z. Xu).

http://dx.doi.org/10.1016/j.patrec.2016.04.017
0167-8655/

To address this problem, we present a novel framework for multi-label learning based on regularized low rank minimization of the joint matrix of the feature-by-item matrix and label-by-item matrix. It can: (1) explicitly explore the correlation among different labels via low rank minimization, (2) leverage the requirement of large amount of training data via manifold regularization, and (3) automatically complete the missing labels via matrix completion. Different from traditional approaches for multi-label learning that also explore the class correlation, the proposed framework provides a natural way for exploring the intrinsic structure of unlabeled data, the label correlation, and the missing of labels, simultaneously, thus effective for the learning scenarios with a large number of missing labels and a small size of training data.

Our paper extends traditional multilabel learning approaches based on matrix completion, proposed by Goldberg et al. [12], Cabral et al. [4] and Liu et al. [14], with a more reasonable modeling of the intrinsic structures of data. In more detail, in order to leverage the scarcity of labeled examples, we adopt the manifold assumption, i.e., instances lie in a small local neighborhood region should also share similar set of labels. To this end, the graph Laplacian is calculated to constrain the process of matrix completion, where we call the proposed model Matrix Completion with Graph Laplacian (MCLA). The resulted nuclear norm minimization problem can be solved with a modified fixed-point continuation method which is guaranteed to converge to the global optimum. Experiments on both synthetic and real-world data have shown that our proposed method can achieve promising results.

2. Related work

2.1. Semi-supervised learning

Traditional supervised learning methods use only labeled data to train classifiers. However, labeled data are often expensive and difficult to obtain. Meanwhile plenty of unlabeled data may be relatively easy to collect. Semi-supervised learning addresses this problem by using both of labeled data and the large amount of unlabeled data to make better classification. Semi-supervised learning usually requires less human effort and gives higher accuracy [23].

However, there are basic assumptions behind the above conclusion, which states that classification performance can be improved when we utilizing unlabeled data in training. In other words, one could say that the knowledge on $p(x)$ that one gains through the unlabeled data has to carry information that is useful in the inference of $p(y|x)$ [8]. One condition that unlabeled data could help semi-supervised learning is that it complies with the clustering assumption. The clustering assumption considers that if two instances lie on a same cluster, they share same labels with a very high probability. Another basic assumption is manifold assumption [17] or smoothness assumption, which refers to instances lie in a small local neighborhood region should also share similar set of labels. Semi-supervised learning has also been studied to address the problem of the small number of labeled examples in multi-label classification [9,15,19] under various assumptions of semi-supervised learning. However, most of them are with less consideration on the large number of (probably partially observed) classes in real world multilabel learning applications.

2.2. Multi-label learning with matrix completion

Considering an $m \times n$ low rank or approximately low-rank matrix A with only partially observed entries A_{ij}, we use Ω to denote the set of indices of observed entries. The goal of matrix completion is to recover the rest of matrix A by a same-size substitute matrix X with $X_{ij} = A_{ij} \forall (i, j) \in \Omega$. Actually, in very general settings, the work of [5] shows that the goal of perfectly recovering all of the unobserved entries from a sufficiently large random subset can be achieved by using a minimizer obtained with the Nuclear Norm. And many works have been done to perform this optimization efficiently [13,16]. For semi-supervised multi-label learning, Goldberg et al. in [12] assumed that the feature-by-item and label-by-item matrices are jointly low rank, then abstracted this task as a matrix completion problem. In [4], matrix completion was applied in multilabel image classification and achieved a remarkable performance. Alternatively, Liu et al. [14] have modeled the multilabel classification problem as a non-negative matrix factorization problem. Despite the successes of these works, they may under exploit the intrinsic structure conveyed in the large amount of unlabeled data, and thus suboptimal for classification. In contrast, the proposed algorithm in our paper can take the advantages of the data geometry conveyed in data to improve multilabel classification.

3. Matrix completion with manifold regularization

We consider the settings of transductive multilabel learning, where only a limited number of (partially) labeled examples are given and most of examples are not labeled. We denote the number of labels by t. Let $X_L = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{l}$ be the labeled set where l denotes the number of labeled examples, and $X_U = \{\mathbf{x}_i\}_{i=l+1}^{n}$ be the unlabeled set, where n denotes the whole amount of observed examples and each example $\mathbf{x}_i \in \mathbb{R}^d$. We denote the label vector by $\mathbf{y}_i \in \{-1, +1\}^t$. Such that $\mathbf{y}_i(k) = +1$ if $\mathbf{y}_i$ belongs to the k-th class ($1 \le k \le t$), and -1 otherwise. We further denote the observed label set as Y_L, and the partially observed and unobserved label set as Y_U. Given the whole set of examples and partial label information, we hope to complete the label set Y_U.

In the following, we will first model the multilabel learning problem as a matrix completion problem, then enhance it with manifold regularization for better modeling intrinsic manifold structure of data, followed by the details of the optimization method.

3.1. Matrix completion for multilabel learning

We first introduce an intermediate instance matrix (X_L^0, X_U^0), such that the observed instance matrix (X_L, X_U) is sampled from (X_L^0, X_U^0) with i.i.d Gaussian noises: $(X_L, X_U) = (X_L^0 + \boldsymbol{\epsilon}, X_U^0 + \boldsymbol{\epsilon})$, where $\epsilon_{ij} \sim N(0, \sigma^2)$. Then the intermediate label matrix (Y_L^0, Y_U^0) can be represented as a linear combination of instance matrix under a weight matrix $W \in \mathbb{R}^t \times \mathbb{R}^d$: $(Y_L^0, Y_U^0) = W(X_L^0, X_U^0)$. Given the linear projection W, the ranks of these two matrices (X_L^0, X_U^0) and $(Y_L^0, Y_U^0; X_L^0, X_U^0)$ satisfies:

$$\operatorname{rank}\left(\begin{pmatrix} Y_L^0 & Y_U^0 \\ X_L^0 & X_U^0 \end{pmatrix}\right) \le \operatorname{rank}((X_L^0, X_U^0)) + 1.$$

We further constructed the augmented matrix Z as follows:

$$Z = \begin{pmatrix} Y_L^0 & Y_U^0 \\ X_L^0 & X_U^0 \\ 1^\top & 1^\top \end{pmatrix}.$$

Due to the correlation among the label set $[Y_L^0 Y_U^0]$ and the connection between data and labels, the rank of Z should be small. However, it is very hard to directly solve a rank minimization problem. Fortunately, rank minimization can be relaxed as minimization of the nuclear norm of Z, that is $\|Z\|^* = \sum_k \sigma_k(Z)$ where $\sigma_k(Z)$ is the k-th singular value of Z [6,7]. Now with the consideration of the noises on (X_L^0, X_U^0) and (Y_L^0), the optimization problem can be