Signal Processing: Image Communication 47 (2016) 207–217

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Semi-supervised classification via discriminative sparse manifold regularization

Zhuang Zhao, Wei Qi, Jing Han, Yi Zhang, Lian-fa Bai *

Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sense, Nanjing University of Science and Technology, Nanjing 210094, China

ARTICLE INFO

Article history:
Received 1 February 2016
Received in revised form
9 May 2016
Accepted 18 June 2016
Available online 23 June 2016

Keywords:
Manifold regularization
Semi-supervised learning
Classification
Sparse representation

ABSTRACT

In this paper, a newly semi-supervised manifold learning algorithm named Discriminative Sparse Manifold Regularization (DSMR) is proposed. In DSMR, the whole unlabeled sample set is used to reconstruct the mean vector of each class, then obtains the sparse coefficient. For each sample of labeled samples, the new dictionary is composed of samples from the same class and the samples from the unlabeled sample set according to the corresponding rows of the sparse coefficient. For each unlabeled sample, the new dictionary is composed of samples from the whole unlabeled samples and the samples from the labeled class according to the corresponding columns of the sparse coefficient. Additionally, a discriminative term is added to stabilize performance of the algorithm. Extensive experiments on the several UCI datasets and face datasets demonstrate the effectiveness of the proposed DSMR.

1. Introduction

Classification is one of the fundamental problems in many scientific fields [1]. The goal of classification is to separate different classes as far as possible. During the past decades, many algorithms for classification have been proposed [8–10]. Classification functions of these algorithms are obtained by minimizing the empirical prediction loss functions.

In the real world, data tend to have very high dimension, but many researches have demonstrated that the performance of classifier will decline with the dimension increase. Thus, in order to improve the performance of classifier and reduce the computational load, many feature extraction algorithms have been proposed. Among these algorithms, global Euclidean structure based algorithms are proposed, such as Principal Component Analysis (PCA) [2,3] and manifold structure based algorithms, such as Isometric Feature Mapping (ISOMAP) [4], Local Linear Embedding (LLE) [5], Laplacian Eigenmap (LE) [6] and Locality Preserving Projections (LPP) [7] are proposed. After feature extraction, classification algorithms, such as Linear Discriminant Analysis (LDA) [8], Constrained Maximum Variance Mapping (CMVM) [9] and Multi-Manifold Discriminant Analysis (MMDA) [10], can be implemented to obtain the classification results.

The classification algorithms mentioned above are supervised algorithms they will perform well when there are enough labeled samples, but in the real world, it is difficult to get enough labeled samples. In this way, semi-supervised learning algorithms have attracted many researchers' interest [11,12]. Among these semi-supervised learning algorithms, graph based semi-supervised learning algorithms also known as semi-supervised manifold learning algorithms are widely used since their good performance.

In semi-supervised manifold learning algorithms, a graph is constructed to character the distribution of all samples. The traditional methods including K nearest neighbor (KNN) based methods and ε-ball based methods [13]. In general, these algorithms construct the graph in two main steps, firstly, choosing the samples needed to connected, secondly, determining the edges weights. However, these semi-supervised manifold learning algorithms have a common drawback: how many samples need to be connected. However, there are no specific solutions to solve it.

The semi-supervised manifold learning algorithms have two common assumptions: cluster assumption and manifold assumption. The cluster assumption indicates that if two samples are located in the same cluster, they have high probability to belong to the same class. The manifold assumption indicates that the high-dimensional data reside on a low-dimensional sub-manifold. However, many semi-supervised manifold learning algorithms may connect samples from different classes in the cluster boundaries because they do not use the discriminative information or the labeled samples information when constructing the graph.

In order to overcome the shortcomings of traditional semi-supervised manifold learning algorithms, a newly semi-supervised learning algorithm named Discriminative Sparse Manifold Regularization (DSMR) is proposed. The graph in DSMR is constructed through sparse representation (SR) and with SR, following

* Corresponding author.
E-mail addresses: 476336296@qq.com (Z. Zhao), blf@njust.edu.cn (L.-f. Bai).

http://dx.doi.org/10.1016/j.image.2016.06.008
0923-5965/© 2016 Published by Elsevier B.V.

advantages are achieved:

(1) The graph constructed in DSMR relies on sparse representation [14], which is generally superior to KNN or ε-ball, especially for high-dimensional data.
(2) Parameter free. The neighborhood size K or the radius ε and the edge weight for each sample, which are hard to determine in traditional semi-supervised manifold learning algorithms, are automatically established by SR. In fact, it is unreasonable to believe that all samples of different classes have the same parameter for KNN or ε-ball based method. However, through SR, different samples will have different neighborhood size K or the radius ε and the edge weight, which is more adaptive for unknown high-dimensional data distribution.
(3) Robust to noise. The graph constructed by KNN or ε-ball based method is sensitive to data noise, especially for high-dimensional data. However, SR has shown its robustness in [14] and experimental results in this paper have verified this point.
(4) Naturally discriminative. Discriminative information is important for classification. However, many semi-supervised manifold learning algorithms only focus on the smoothness of manifold and ignore the discriminative information. SR has natural discriminative power and can work well for high-dimensional data [15] and experimental results in this paper have verified this point.

The organization of this paper is organized as follows. In Section 2, the semi-supervised manifold learning algorithms and SR are introduced. In Section 3, the motivation and implement of DSMR are introduced in detail. In Section 4, experiments are conducted on several UCI datasets and face image datasets and we analyze the relationship of the proposed DSMR and other algorithms. Conclusion of this paper is made in Section 5.

2. Related works

Our work is closely related to the semi-supervised manifold learning algorithms and SR. Now we introduce them in detail.

2.1. Semi-supervised manifold learning algorithms

Semi-supervised manifold learning algorithms can make full use of all samples to explore the underlying geometric structure and get better results. It can be used in many areas, such as image annotation [16], medical diagnosis [17] and classification [20,35,13]. Cai et al. use the labeled samples to maximize the discriminating power and use the unlabeled samples to estimate the intrinsic geometric structure of the samples in SSDA [18]. However, in SSDA the label information is ignored when constructing the graph. Belkin et al. introduce the underlying samples distribution information of manifold structures into the traditional regularization and get Manifold Regularization (MR) [19]. MR has two regularization terms, one regularization term controls the complexity of the classifier, and the other one controls the complexity, which is measured by the manifold geometry of the samples distribution. On the basis of MR, many semi-supervised manifold regularization algorithms have been proposed, such as Wu et al. combine the MR and a discriminative term get Semi-supervised Discriminant Regularization (SSDR) [20], Zhao et al. combine the MR and local and global regression get Learning from Local and Global Discriminative Information (LLGDI) [21] and Local and Global Regression (LGR), Fan et al. combine the MR and sparse representation get Sparse Regularized Least Square Classification (S-RLSC) [23], Gu et al. also combine the MR and sparse representation get Discriminant Sparsity Preserving Projection (DSPP) [36].

The traditional semi-supervised manifold learning algorithms only focus on the smoothness of the manifold and ignore discriminative information. However, the discriminative information or label information is helpful to remove the edges link samples from different classes. By exploring the discriminative information, the proposed DSMR has the ability to characterize the whole manifold structure of all samples and gets more accurate results.

2.2. Sparse representation

Recently, sparse representation (SR) has attracted a great deal of attentions due to its success in improving the performances of various machine learning algorithms, such as classification [24], image super-resolution [25,26], image de-noising [27], feature extraction [28,29], signal reconstruction [30].

Given a set of samples $X=[x_1, x_2, \cdots, x_n]^T \in R^{m*n}$, where n is the number of samples and m is dimensional of each sample. The aim of SR is to reconstruct each sample x_i using as few samples in X as possible. The objective function of SR can be described as:

$$\min_{s_i} \|s_i\|_0,\ s.\ t.\quad x_i = Xs_i \tag{1}$$

or

$$\min_{s_i} \|s_i\|_0,\ s.\ t.\quad \|x_i - Xs_i\| \le \varepsilon \tag{2}$$

where ε is a small constants parameter, $s_i = [s_{i,1}, \cdots, s_{i,i-1}, 0, s_{i,i+1}, \cdots, s_{i,n}]^T$ is a n-dimensional column vector with the ith element equal zero, which states that the x_i is removed from X, other elements in s_i denote the contribution of x_j for reconstructing x_i. Unfortunately, the solution of Eqs. (1) or (2) is a NP-hard problem. If we use l_1 instead l_0, the solution of Eqs. (1) or (2) can be solved by LASSO [31] or LARS [32]. In this paper, SPAMS (Sparse Modeling Software) [33,34] is used to solve the problem. After obtaining all optimal reconstruction coefficient s_i for each x_i, then the sparse matrix S is constructed by

$$S = \left[s_1, s_2, \cdots, s_n\right] \tag{3}$$

Then, the new constructed graph is $G=\{X, S\}$, where X is the training samples set and S is the sparse matrix.

3. DSMR

In this section, we elaborate on the motivation and the detail of DSMR.

3.1. Motivation of DSMR

MR provides a framework for semi-supervised manifold learning algorithms, but MR ignores the discriminative information and its graph is hard to construct. SR is a good solution to solve it. In Sparse Neighborhood Preserving Embedding (SNPE) [35], SR is used to preserve the local structure of manifold. The objective function of SNPE is shown in Eq. (4).

$$\min \sum_{i=1}^{n} \left\|w^T x_i - w^T X s_i\right\|^2 \tag{4}$$

where w is the projection matrix. The objective function can be simplified to Eq. (5) with simple algebraic formulation.