Neurocomputing 177 (2016) 564–571

Contents lists available at ScienceDirect

Neurocomputing

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

Adaptive semi-supervised dimensionality reduction with sparse representation using pairwise constraints

Jia Wei [a], Meng Meng [a], Jiabing Wang [a], Qianli Ma [a], Xuan Wang [b]

[a] School of Computer Science and Engineering, South China University of Technology, Guangzhou, China
[b] Computer Application Research Center, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, China

ARTICLE INFO

Article history:
Received 20 August 2014
Received in revised form
29 September 2015
Accepted 19 November 2015
Communicated by Feiping Nie
Available online 2 December 2015

Keywords:
Semi-supervised learning
Dimensionality reduction
Pairwise constraints
Sparse representation

ABSTRACT

With the rapid accumulation of high dimensional data, dimensionality reduction plays a more and more important role in practical data processing and learning tasks. This paper studies semi-supervised dimensionality reduction using pairwise constraints. In this setting, domain knowledge is given in the form of pairwise constraints, which specifies whether a pair of instances belong to the same class (must-link constraint) or different classes (cannot-link constraint). In this paper, a novel semi-supervised dimensionality reduction method called Adaptive Semi-Supervised Dimensionality Reduction with Sparse Representation (ASSDR-SR) is proposed, which can get the optimized low dimensional representation of the original data by adaptively adjusting the weights of the pairwise constraints and simultaneously optimizing the graph construction using the ℓ_1 graph of sparse representation. Experiments on clustering and classification tasks show that ASSDR-SR is superior to some existing dimensionality reduction methods.

1. Introduction

The goal of dimensionality reduction is to reduce the complexity of the input data while some desired intrinsic information of the data is preserved. Two of the most popular methods for dimensionality reduction are Principal Component Analysis (PCA) [1] and Linear Discriminant Analysis (LDA) [2], which are unsupervised and supervised respectively.

In many real world applications such as image segmentation, web page classification and gene-expression clustering, a labeling process is costly and time-consuming; in contrast, unlabeled examples can be easily obtained. Therefore, in such situations, it may be beneficial to incorporate the information which is contained in unlabeled examples into a learning problem, i.e., Semi-Supervised Learning (SSL) [3] should be applied instead of supervised learning. Meanwhile, dimensionality reduction in semi-supervised situation has also attracted more and more attention [4,5].

However, in many cases, people cannot tell which category an instance belongs to, that is we do not know the exact label of an instance, and what we know is the constraint information of whether a pair of instances belong to the same class (must-link constraint) or different classes (cannot-link constraint) [6]. The above pairwise constraint information is called “Side Information” [7]. It can be seen that constraint information is more general than label information, because we can get constraint information from label information but it cannot work contrariwise [8].

Some related works have been proposed to make use of the pairwise constraints to extract low dimensional structure in high dimensional data. Bar-Hillel et al. proposed Relevant Component Analysis (RCA) which can make use of the must-link constraints for semi-supervised dimensionality reduction [9]. Xing et al. [7], Tang et al. [10], Yeung et al. [11] and An et al. [12] proposed some constraints based semi-supervised dimensionality reduction methods, which can make use of both the must-link constraints and cannot-link constraints. Zhang et al. proposed Semi-Supervised Dimensionality Reduction (SSDR) [13] and Chen et al. used SSDR in hyperspectral image classification [14]. SSDR can use the pairwise constraints as well as preserve the global covariance structure of the unlabeled data in the projected low dimensional subspace. Cevikalp et al. proposed Constrained Locality Preserving Projections (CLPP) [15] which is the semi-supervised version of LPP [16]. The method can make use of the information provided by the pairwise constraints and can also use the unlabelled data by preserving the local structure used in LPP. Wei et al. proposed Neighborhood Preserving based Semi-Supervised Dimensionality Reduction (NPSSDR) [17] by using the pairwise constraints and preserving the neighborhood structure used in LLE [18]. Baghshah et al. used the idea of NPSSDR in metric learning and used a heuristic search algorithm to solve the proposed constrained trace ratio problem [19]. Davidson proposed a graph driven constrained

E-mail address: csjwei@scut.edu.cn (J. Wei).

http://dx.doi.org/10.1016/j.neucom.2015.11.048
0925-2312/

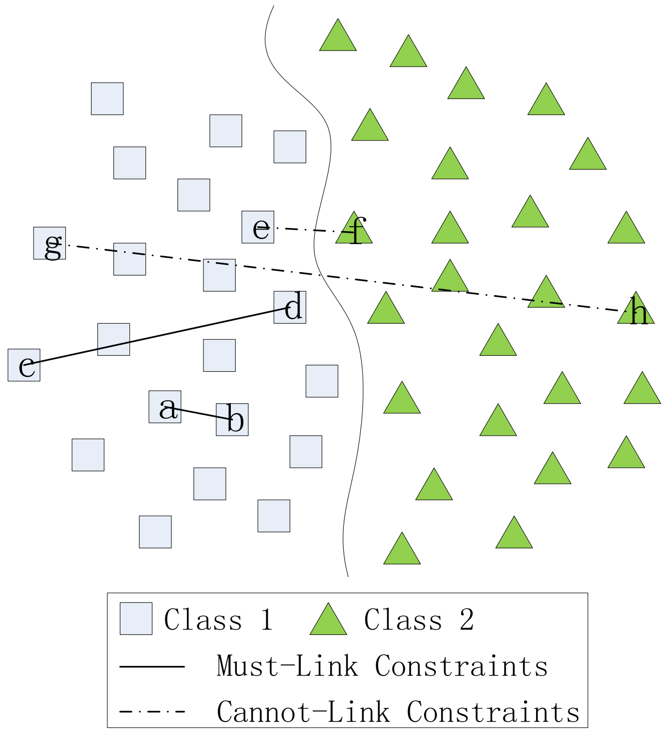

Fig. 1. Illustration of pairwise constraints with different amounts of information.

dimensionality reduction approach GCDR-LP for clustering [20]. In this approach, a constraint graph is firstly created by propagate the constraints due to transitivity and entailment in the graph, and then the dimensionality reduction can be conducted by the constraint graph. Related works also include Dual Subspace Projections (DSP) [21], integrating Local and Global topological Structure for SSDR (LGS3DR) [22] et al.

However, a common problem of the aforementioned methods is that the pairwise constraints are equally treated in the algorithms which ignore the fact of unequally informativeness owned by different pairwise constraints. For example, consider a binary-class case in Fig. 1, (a, b) and (c, d) are two must-link constraints of class 1, (e, f) and (g, h) are two cannot-link constraints between class 1 and class 2. It is sound to say that the must-link constraint (c, d) has more information than (a, b), because the distance of (c, d) is larger than that of (a, b), which indicates c and d are more likely to be located on the margin of class 1. On the contrary, the cannot-link constraint (e, f) has more information than (g, h), because the distance of (e, f) is smaller than that of (g, h), which indicates e and f are more likely to be located on the margin between class 1 and class 2. So, it is sensible to handle different pairwise constraints with different importances.

On the other hand, in order to utilize unlabeled data, most graph-based semi-supervised dimensionality reduction methods (e.g., CLPP and NPSSDR) generally construct a neighborhood graph from the available data. Conventional graph construction processes are generally divided into two independent steps, i.e., adjacency searching and weight computing. However, such graph tends to work poorly due to the high dimensions of the original space because it is hard to search the appropriate adjacencies of a given datum and compute the weights between them in the high dimensional space. To solve the problem, one should integrate graph construction with specific semi-supervised dimensionality reduction process into a unified framework, which results in an optimized graph rather than a predefined one. Recently, Nie et al. [23] proposed a clustering model to learn the data similarity matrix and clustering structure simultaneously which is similar to our idea. Related works also include [24,25].

In this paper, a novel semi-supervised dimensionality reduction method called Adaptive Semi-Supervised Dimensionality Reduction with Sparse Representation (ASSDR-SR) is proposed. ASSDR-SR first initializes all the pairwise constraints with equal weights and construct the adjacency weight matrix using the ℓ_1 graph of the sparse representation [26–28], and then the following procedure is repeated until the stop condition is satisfied: (1) reducing the dimensionality of the original space with the current weighted pairwise constraints and the current adjacency weight matrix; (2) optimizing the ℓ_1 graph of the sparse representation as the adjacency weight matrix in the reduced subspace with the current weighted pairwise constraints; (3) clustering in the reduced subspace and updating the weights of the pairwise constraints according to the clustering result. As a result, we can get the optimized weights for the pairwise constraints and the optimized adjacency weight matrix for the neighborhood graph, as well as the projection matrix.

2. Objective function of ASSDR-SR

2.1. The problem

Here we define the weighted pairwise constraints based semi-supervised dimensionality reduction problem as follows: Suppose we have a set of D-dimensional data samples $\{cx_i\}_{i=1}^n$, where $x_i \in R^D$, together with some pairwise must-link constraints (M) and cannot-link constraints (C) as domain knowledge: $(x_i, x_j) \in M$, if x_i and x_j are in the same class; $(x_i, x_j) \in C$, if x_i and x_j are in the different classes. In addition, we suppose each pairwise constraint (x_i, x_j) has a pending weight S_{ij} ($S_{ij} < 0$ for must-link constraints; $S_{ij} > 0$ for cannot-link constraints. For consistency, we set $S_{ij} = 0$ for unconstrained pairs) to indicate the informativeness owned by itself, which means one should be paid more attention to the pairwise constraint (x_i, x_j) if $|S_{ij}|$ is large. In this case, what we want to do is to find a set of linear projection vectors $W = [w_1, w_2, \ldots, w_d] \in R^{D \times d}$, where $d \ll D$, such that the transformed low dimensional projections $\{y_i\}_{i=1}^n \subset R^d$, where $y_i = W^T x_i$, can preserve some properties of the original dataset as well as the pairwise constraints in M and C.

2.2. Making use of pairwise constraints

To make use of the pairwise constraints to control the similarity of pairwise instances, those in M should end up close to each other while the pairwise points in C should end up far from each other. This means the instances belong to the same class in the original space should be close to each other in the reduced subspace, and the instances belong to different classes in the original space should be far from each other in the reduced subspace. In addition, if S_{ij} is negative and $|S_{ij}|$ is large, it means $(x_i, x_j) \in M$ and they should be closer to each other than with smaller $|S_{ij}|$ in the reduced subspace; if S_{ij} is positive and $|S_{ij}|$ is large, it means $(x_i, x_j) \in C$ and they should be farther from each other than with smaller $|S_{ij}|$ in the reduced subspace.

As for the weighted must-link constraints M, the intraclass compactness is characterized by the term as follows:

$$Q^M(W, S) = \sum_{(x_i, x_j) \in M \text{ or } (x_j, x_i) \in M} \| W^T x_i - W^T x_j \|_2^2 |S_{ij}| \tag{1}$$

$Q^M(W, S)$ should be as small as possible which means the weighted distance sum in the transformed low dimensional subspace between instances involved in the must-link constraints M should be small.