ARTICLE IN PRESS

ISA Transactions ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

ISA Transactions

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

Discriminative sparse subspace learning and its application to unsupervised feature selection

Nan Zhou [a], Hong Cheng [a,*], Witold Pedrycz [b,d,e], Yong Zhang [a], Huaping Liu [c]

[a] Center for Robotics, School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China
[b] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada
[c] Department of Computer Science and Technology, Tsinghua University, Beijing, China
[d] Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[e] Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

ARTICLE INFO

Article history:
Received 30 August 2015
Received in revised form
3 December 2015
Accepted 18 December 2015

Keywords:
Machine learning
Feature selection
Subspace learning
Unsupervised learning
Kernel learning

ABSTRACT

In order to efficiently use the intrinsic data information, in this study a Discriminative Sparse Subspace Learning (DSSL) model has been investigated for unsupervised feature selection. First, the feature selection problem is formulated as a subspace learning problem. In order to efficiently learn the discriminative subspace, we investigate the discriminative information in the subspace learning process. Second, a two-step TDSSL algorithm and a joint modeling JDSSL algorithm are developed to incorporate the clusters' assignment as the discriminative information. Then, a convergence analysis of these two algorithms is provided. A kernelized discriminative sparse subspace learning (KDSSL) method is proposed to handle the nonlinear subspace learning problem. Finally, extensive experiments are conducted on real-world datasets to show the superiority of the proposed approaches over several state-of-the-art approaches.

1. Introduction

With the improvement of technology, the dimensionality of the data increases fast in many fields, such as data processing, machine learning and pattern recognition. The high-dimensional data not only increase requirements for storage space and processing time, but also introduce much noise and redundancy which can negatively influence the performance of algorithms of pattern recognition. Therefore, dimensionality reduction is badly required to overcome this problem [1,2]. There are two widely used approaches to dimensionality reduction, namely: subspace learning and feature selection. Subspace learning aims to find the projection which can transform the data to lower-dimensional subspace [3,4]. Feature selection aims to select a smaller subset of features [5,6].

Depending on whether the process is supplied with class labels, both the subspace learning and the feature selection methods can be categorized into supervised and unsupervised ones. The classic unsupervised subspace learning method is Principal Component Analysis (PCA) [4,7]. Some other graphic embedding subspace learning methods include Locally Linear Embedding (LLE) [8], Linear Discriminant Analysis (LDA) [9] and Locality Preserving Projection (LPP) [10].

The supervised feature selection methods can often result in better performance than unsupervised ones, because the labels contain essential discriminative information which can be used to guide the feature selection process to select more useful features. There are some commonly used supervised feature selection methods, such as, Fisher Criterion (FC) [11], Mutual Information (MI) [5], and Pearson Correlation Coefficient (PCC) [12]. However, unlabeled data are much easier to obtain than the labeled ones, which makes the unsupervised feature selection indispensable. Without label information, unsupervised feature selection methods have to explore the intrinsic information of the data, such as, samples' similarity [6], clustering label [13], global reconstruction information [14] and data structure [15,16]. However, there are a few studies on subspace learning considering both global reconstruction and clusters' assignment for unsupervised feature selection. Inspired by this, we formulate the unsupervised feature selection problem in discriminative subspace learning form and then develop the efficient iterative algorithms to solve the problem.

This paper offers the following contributions:

* Corresponding author.
E-mail addresses: nzhouuestc@gmail.com (N. Zhou), hcheng@uestc.edu.cn (H. Cheng), wpedrycz@ualberta.ca (W. Pedrycz), mathzy@uestc.edu.cn (Y. Zhang), hpliu@tsinghua.edu.cn (H. Liu).

http://dx.doi.org/10.1016/j.isatra.2015.12.011
0019-0578/

Please cite this article as: Zhou N, et al. Discriminative sparse subspace learning and its application to unsupervised feature selection. ISA Transactions (2016), http://dx.doi.org/10.1016/j.isatra.2015.12.011

1. A novel unsupervised subspace learning framework is proposed, which incorporates data's discriminative information into subspace learning process. Because of the group sparsity regularization term which can control the row sparsity of the transformation matrix and intuitively point whether the feature is selected or not, this model is suitable for feature selection.
2. Class label as the most efficient discriminative information always be used in supervised learning scenario. Because of unsupervised learning scenario does not exhibit class label information, the clusters' assignment is used as the pseudo class label. Combining with the data's reconstruction information, two methods for discriminative sparse subspace learning are proposed. One is a two-step technique which includes two subproblems: (1) nonnegative Laplacian embedding, which can form the pseudo class labels; (2) discriminative sparse subspace learning which incorporates the pseudo class labels discriminative information into the subspace learning. Then, to improve the flexibility and performance of the model, another joint modeling method is developed, which combines both steps of the two-step method. The convergence analysis of the proposed methods is carried out to present the efficiency of the algorithms.
3. To better capture the intrinsic information of the data, a kernelized discriminative sparse subspace learning method is proposed to tackle the features in nonlinear manifold situation.
4. The effectiveness of the proposed methods are presented by extensive experimental studies. The proposed methods are tested on six real-world datasets coming from different areas and compared with eight state-of-the-art unsupervised feature selection algorithms. The results demonstrate the superiority of TDSSL and JDSSL methods over all other compared methods. The kernelized method is even better than the proposed methods which verifies the effectiveness of the kernel learning in the unsupervised feature selection scenario. In addition, the sensitivity of the joint modeling method's parameter is studied. It is observed that the JDSSL method can well perform in a stable manner within a large range of the value of the parameter.

The paper is organized as follows. Section 2 briefly reviews the related studies on subspace learning and feature selection. Section 3 presents the discriminative sparse subspace learning model and reviews the Laplacian embedding, multi-way ratio cut and the relation between them. Then, a brief review of the nonnegative Laplacian embedding is given. Section 4 presents two algorithms about discriminative sparse subspace learning and demonstrates their convergence analysis. Section 5 presents kernelized discriminative sparse subspace learning method. Section 6 discusses how the improvements are achieved in comparison with other related methods. Experimental results are reported in Section 7. Finally, the conclusion is offered in Section 8. The overall scheme of the paper is shown in Fig. 1.

To facilitate the presentation of the material, Table 1 contains the notation used in this study.

2. Related studies

Data reconstruction, which can capture the discriminative information, plays an important role in pattern recognition [17]. The most commonly used subspace learning method using the reconstruction information is PCA [4], which maximizes the data's global information in the principal space hence it is optimal for data reconstruction. Because of its sound flexibility and interpretation, graphic embedding method is widely used in subspace learning [4,10,18]. The graphic embedding can be used to preserve the data's similarity structures or network structures which can be seen as the local information of the data.

Table 1
Notations used in the study.

Notation	Description
n	The number of instances
d	The number of features
$A_{i.}$	The ith row of the matrix A
K	The number of selected features
m	The number of nearest neighbors
$\Vert W \Vert_{2,1}$	$\sum_i \Vert W_{i.} \Vert_2$, the sum of the ℓ_2-norm of rows in W
$\Vert \mathbf{x} \Vert_0$	$\sharp\{x_i \neq 0\}$, the number of nonzero elements in vector $\mathbf{x}$
$\vert \mathcal{I} \vert$	Cardinality of set $\mathcal{I}$
$\mathrm{Tr}(A)$	The trace of matrix A, i.e., $\mathrm{Tr}(A) = \sum_{i=1}^{n} A_{ii}$
$\Vert A \Vert_F$	The Frobenius norm of A, i.e., $\Vert A \Vert_F = \sum_{i=1}^{n} \sum_{j=1}^{m} A_{ij}^2$

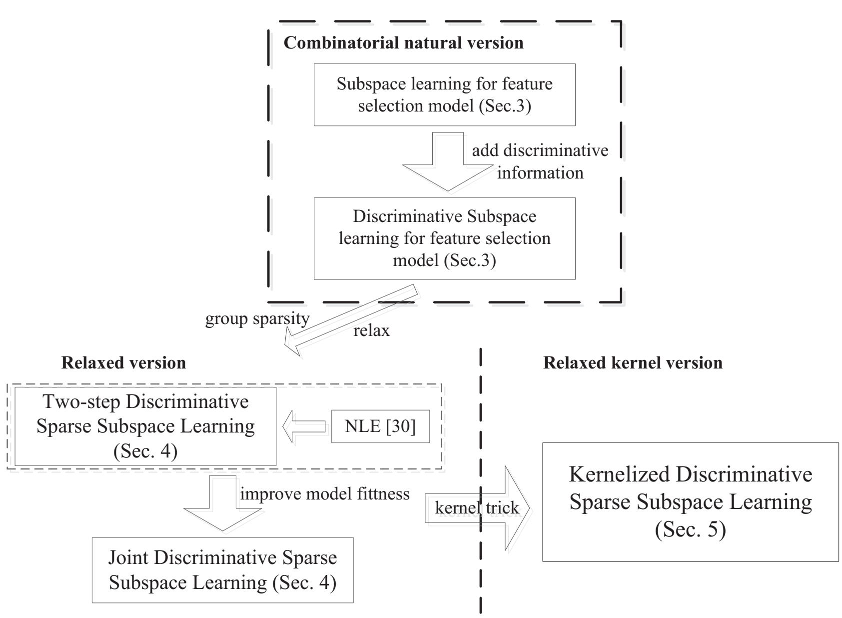

Fig. 1. The overall scheme of the study.

Please cite this article as: Zhou N, et al. Discriminative sparse subspace learning and its application to unsupervised feature selection. ISA Transactions (2016), http://dx.doi.org/10.1016/j.isatra.2015.12.011