Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Engineering 29 (2012) 1028 – 1033
Procedia Engineering
www.elsevier.com/locate/procedia

2012 International Workshop on Information and Electronics Engineering (IWIEE)

Spare Projections with Pairwise Constraints

Xiaodong Chen[a]*, Jiangfeng Yu[a]

[a]School of Information and Engineering, Zhejiang Radio & TV University, Hangzhou, 310030, China

Abstract

In this paper, we propose a new semi-supervised DR method called sparse projections with pairwise constraints (SPPC). Unlike many existing techniques such as locality preserving projection (LPP) and semi-supervised DR (SSDR), where local or global information is preserved during the DR procedure, SPPC constructs a graph embedding model, which encodes the global and local geometrical structures in the data as well as the pairwise constraints. After obtaining the embedding results, sparse projections can be acquired by minimizing a L1 regularization-related objective function. Experiments on real-world data sets show that SPPC is superior to many established dimensionality reduction methods.

Keywords: Sparse projections; Pairwise constraints; Dimensionality reduction; global and local structures.

1. Introduction

In general, existing dimensionality reduction methods can be roughly categorized into supervised ones, semi-supervised ones and unsupervised ones. In supervised DR scenarios where data samples are accompanied with class labels, Fisher Linear Discriminant (FLD) [1] is the most popular one. It can find the optimal discriminant projection in terms of class labels obtained in advance. FLD has two famous variants, called Marginal Fisher Analysis (MFA) [2] and Local Fisher Discriminant Analysis (LFDA) [3]. Unsupervised DR methods project input data onto a low-dimensional manifold without any constraint or class label information. Principle Component Analysis (PCA) [4] is the most popular method among this category, which obtains linear subspace spanned by the leading eigenvectors of the data's covariance matrix. Locally Linear Embedding (LLE) [5] is another popular unsupervised algorithm which computes

* Corresponding author. Tel.: +86-571-89983172; fax: +86-571-89983165.
E-mail address: cxd@zjtvu.edu.cn.

doi:10.1016/j.proeng.2012.01.084

low dimensional neighborhood preserving embeddings of high dimensional data. Other unsupervised ones include Neighborhood Preserving Embedding (NPE) [6], Locality Preserving Projections (LPP) [7], Laplacian Eigenmap [8], ISOMAP [9], etc. Semi-supervised DR (SSDR) [10] is essentially semi-supervised PCA, which uses pairwise constraints to find the projective direction together with unlabeled instances. Semi-supervised Discriminant Analysis (SDA) [11] expands FDA to semi-supervised scenarios in terms of labeled instances. In addition, other semi-supervised DR methods based on pairwise constraints are related to semi-supervised clustering [12-18].

In this paper, motivated by the recent development of sparse learning [19-21], we propose a novel semi-supervised DR algorithm called Sparse Projection with Pairwise Constraints (SPPC). Specifically, we present a graph Laplacian formulation to integrate global and local structures of the data. After getting the embedding results, lasso regression can be naturally applied to obtain sparse basis functions for finding the optimal projective direction. Moreover, a tuning parameter is employed to control the weights between supervised and unsupervised terms of the objective function.

2. Spare Projections with Pairwise Constraints

2.1. Model formulation

We may apply a k-nearest neighbor graph to formulate the relationship between nearby data instances. More specifically, an edge is put between nodes i and j if x_i and x_j are close, i.e. x_i and x_j are among k nearest neighbors. The corresponding weight matrix P is defined as follows:

$$P_{ij} = \begin{cases} e^{\frac{-\|x_i - x_j\|^2}{2\sigma^2}} & \text{if } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $N_k(x_i)$ denotes the set of k nearest neighbors of x_i. Thus, we minimize the following loss function [7] to learn an appropriate representation with the local geometrical structure.

$$J_L(a) = \sum_{i,j} (a^T x_i - a^T x_j)^2 P_{ij} \,. \tag{2}$$

The goal of Eqn. (2) is to make near instances in the original space as close as possible in the embedding space. Thus, the objective function in semi-supervised DR algorithm is minimized as follows:

$$J_{MC}(a) = \frac{1}{2n_M} \sum_{(x_i,x_j)\in M} (a^T x_i - a^T x_j)^2 - \frac{1}{2n_C} \sum_{(x_i,x_j)\in C} (a^T x_i - a^T x_j)^2 \tag{3}$$
$$\text{subject to} \quad a^T a = 1$$

where n_M and n_C are the numbers of must-link constraints and cannot-link constraints, respectively.

Concretely, we minimize the following objective:

$$J(a) = \frac{1}{2n} \sum_{i,j} (a^T x_i - a^T x_j)^2 P_{ij} - \frac{1}{2n} \sum_{i,j} (a^T x_i - a^T x_j)^2 \tag{4}$$
$$+ \eta\left(\frac{1}{2n_M} \sum_{(x_i,x_j)\in M} (a^T x_i - a^T x_j)^2 - \frac{1}{2n_C} \sum_{(x_i,x_j)\in C} (a^T x_i - a^T x_j)^2\right)$$

To understand the proposed algorithm easily, we reduce Eqn. (4) as follows:

$$J(a) = \frac{1}{2} \sum_{i,j} (a^T x_i - a^T x_j)^2 S_{ij} \tag{5}$$

where

$$S_{ij} = \begin{cases} \frac{P_{ij}-1}{n} + \frac{\eta}{n_M} & if\,(x_i, x_j) \in M \\ \frac{P_{ij}-1}{n} - \frac{\eta}{n_C} & if\,(x_i, x_j) \in C \\ \frac{P_{ij}-1}{n} & otherwise \end{cases} \tag{6}$$

Thus, we can simplify Eqn. (5) or Eqn. (6) as minimizing J(a) w.r.t. aT a = 1, where

$$\min\; J(a) = a^T XLX^T a \qquad subject\ to\quad a^T a = 1 \tag{7}$$

Clearly, the projective vector a that minimizes Eqn. (7) is given by the minimum eigenvalue solution to the generalized eigenvalue problem:

$$XLX^T a = \lambda a \tag{8}$$

2.2. A general model for sparse projection

After obtaining weight matrix S from Eqn. (6), we aim to minimize the following objective function to obtain a sparse projection:

$$\min\quad J(a) = \frac{1}{2}\sum_{i,j}(a^T x_i - a^T x_j)^2 S_{ij} \qquad subject\ to\quad a^T a = 1,\quad card(a) \le q \tag{9}$$

Directly solving Eqn. (9) is NP-hard. In the following we present an efficient method to overcome this issue, which is the same as [19]. Firstly, the projective vector a is obtained by optimizing Eqn. (7). Secondly, we may get the embedding results:

$$y_i = a^T x_i \tag{10}$$

Finally, lasso regression is used to get a sparse projective vector. Let $\tilde{a}$ be the sparse approximation of a. Minimizing Eqn. (9) is transformed to minimizing the following objective function:

$$\min_{\tilde{a}}\quad (\sum_{i=1}^{n}(a^T x_i - \tilde{a}^T x_i)^2 + \lambda\sum_{j=1}^{D}|\tilde{a}_j|) \tag{11}$$

Further reducing Eqn. (11), we have

$$\min_{\tilde{a}}\quad (\sum_{i=1}^{n}(y_i - \tilde{a}^T x_i)^2 + \lambda\sum_{j=1}^{D}|\tilde{a}_j|) \tag{12}$$

3. Experimental Results

3.1. Experimental setup

It is worthwhile to point out that we apply K-means to cluster the instances in the embedding space. We also compare the proposed algorithm with representative DR algorithms including the following five methods: SSDR [10], LMDM [16], LPP [7], SPG [19], and SDRS [23].

We compared all these algorithms on three benchmark data sets, including TDT2, Reuters[a] and USPS[b]. The statistics of all data sets are described in Tables 1-2. We ran different algorithms for 40 times on each data set and the comparison was based on the average performance.

---

[a] http://www.zjucadcg.cn/dengcai/Data/TextData.html
[b] http://www.zjucadcg.cn/dengcai/Data/MLData.html

The clustering result is evaluated by comparing the obtained cluster label of each instance with that provided by each data set. Normalized mutual information metric (NMI) is applied to evaluate these algorithms, which is defined as follows [15,17,19]:

$$NMI = \frac{I(A,B)}{\sqrt{H(A)H(B)}} \tag{13}$$

Table 1. Summary of TDT2 and USPS used in our experiment

Data set	Dimension	Instance	Class
TDT2	1500	8761	9
TDT2	1500	9030	10
TDT2	1500	9594	11
TDT2	1500	9968	12
USPS	256	6010	7
USPS	256	7000	8
USPS	256	8126	9
USPS	256	9298	10

Table 2. Summary of Reuters used in our experiment

Data set	Dimension	Instance	Class
Reuters	1000	3979	3
Reuters	1000	5030	4
Reuters	1000	4594	5
Reuters	1000	5168	6
Reuters	1000	6766	7
Reuters	1000	7980	8
Reuters	1000	8538	9
Reuters	1000	8137	10

3.2. Experimental results

We selected the same number of must-link constraints and cannot-link constraints from the data sets randomly according to the ground truth in our experiments. The same constraints are used for SSDR, LMDM and SDRS. The corresponding results are shown in Figs. 1 and Table 3.

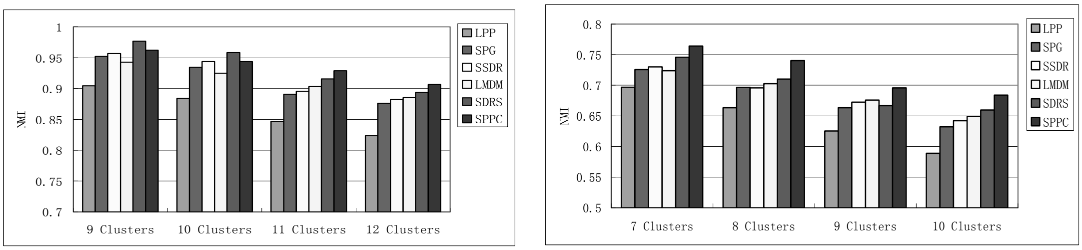

Fig. 1. (a)Clustering performance on TDT2 with 600 constraints ; (b)Clustering performance on USPS with 600 constraints

Table 3. Clustering performance on Reuters with 600 constraints (%)

c	SSDR	LMDM	LPP	SPG	SDRS	SPPC
3	56.66	55.36	47.90	49.25	60.22	**66.31**
4	58.87	52.47	48.29	54.24	61.09	**68.52**
5	57.40	51.67	49.02	53.57	60.87	**67.98**
6	56.31	50.34	50.11	49.87	59.31	**65.30**
7	59.42	56.97	49.37	54.83	61.24	**69.67**
8	53.04	51.02	42.41	46.54	57.36	**62.55**
9	51.00	48.55	39.33	44.90	55.39	**60.37**
10	53.29	49.64	43.20	48.49	57.04	**61.28**

4. Conclusions

In this paper, we propose a simple but efficient sparse projection with pairwise constraints algorithm called SPPC. We construct a new affinity graph to encode the global and local geometrical structures in the data as well as the pairwise constraints. The sparse projections can be obtained by solving an optimization problem which using techniques from lasso regression. Experimental results on real-world data sets show that SPPC leads to considerable improvements in embedding and clustering over conventional dimensionality reduction methods.

Acknowledgment

The research reported in this paper has been partially supported by Natural Science Foundation of Zhejiang Province under Grant No. Y1100349 , Foundation for the Open University of China under Grant No. GFQ1601 and Foundation for University Young Teacher by Zhejiang Province (2010).

References

[1]K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, Maryland, MO, 1990.

[2]S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, S. Lin. Graph Embedding and Extensions: A General Framework for Dimensionality Reduction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(1): 40-50.

[3]M. Sugiyama. Local Fisher discriminant analysis for supervised dimensionality reduction. the 23th International Conference on Machine Learning, Pittsburgh, Pennsylvania, 2006, 905-912.

[4]I. Jolliffe. Principal Component Analysis. Springer, 2nd edition, 2002.

[5]L. K. Saul and S. T. Roweis. Think globally, fit locally. Unsupervised learning of low dimensional mani-folds. Journal of Machine Learning Research, 2003, 4:119-155.

[6]X.F. He, D. Cai, S.C. Yan, H.J. Zhang, Neighborhood preserving embedding, in: Proceedings of IEEE International Conference on Computer Vision (ICCV), 2005.

[7]X.F. He, P. Niyogi, Locality preserving projections, in: Proceedings of Neural Information Processing Systems (NIPS), 2003.

[8]M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Computation 15 (6) (2003) 1373-1396.

[9]J. Tenenbaum, V. de Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. Science, 290(5500):2319–2323, 2000.

[10]D. Zhang, Z. Zhou, S. Chen, Semi-supervised dimensionality reduction, in: The 7th SIAM International Conference on Data Mining, 2007, pp. 11-393.

[11]D. Cai, X. He, J. Han, Semi-supervised discriminant analysis, in: Proceedings of the International Conference on Computer Vision, 2007, pp. 1-7.

[12]Y. Song, F. Nie, C. Zhang, S. Xiang, A Unified framework for semi-supervised dimensionality reduction, Pattern Recognition 41, (2008) 2789–2799.

[13]H. Li, T. Jiang, K. Zhang, Efficient and robust feature extraction by maximum margin criterion, IEEE Trans. Neural Networks, 2006, 17 (1): 157–165.

[14]R. Chatpatanasiri, B. Kijsirikul. A unified semi-supervised dimensionality reduction framework for manifold learning. Neurocomputing, 2010, 73(3): 1631-1640.

[15]X. Yin, E. Hu. Distance Metric Learning Guided Adaptive Subspace Semi-supervised Clustering. Frontiers of computer science in china, 2011, 5(1): 100-108.

[16]W. Tang, H. Xiong, S. Zhong, J. Wu, Enhancing semi-supervised clustering: a feature projection perspective, in: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, California, USA, 2007, pp. 707–716.

[17]S. Xiang , F. Nie and C. Zhang. Learning a Mahalanobis distance metric for data clustering and classification. Pattern Recognition (PR), Volume 41, Issue 12, Pages 3600 - 3612, 2008.

[18]X. Yin, S. Chen, E. Hu, D. Zhang. Semi-supervised clustering with metric learning: An adaptive kernel method, Pattern Recognition, 2010, 43(4), pp. 1320-1333.

[19]D. Cai, X. He, J. Han, Sparse Projections over Graph, In Proc. of the 23rd national conf. on Artificial intelligence, Chicago. Illinois. USA, 2008, 2, pp.601-615.

[20]J. Wright, A. Yang, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 2009, 31 (2): 210–227.

[21]J. Yang, H. Tang, Y. Ma, and T. Huang. Robust Sparse Coding for Face Recognition. 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs, USA, 2011.

[22]Efron, B.; Hastie, T.; Johnstone, I.; Tibshirani, R. Least angle regression. Annals of Statistics, 2004, 32(2):407–499.

[23]S. Chen, D. Zhang. Semi-supervised Dimensionality Reduction with Pairwise Constraints for Hyperspectral Image Classification. IEEE Geoscience & Remote Sensing Letters, 2011, 8(2): 369-373.