

# Constraint Sparse Neighborhood Preserving Embedding

Shuhua Xu\*

Department of Maths, Shaoxing University, Shaoxing, China

**Abstract:** Neighborhood Preserving Embedding (NPE) and extensions of NPE are hot research topics of data mining at present. An algorithm called Constraint Sparse Neighborhood Preserving Embedding (CSNPE) for dimensionality reduction is proposed in the paper. The algorithm firstly creates the local sparse reconstructive relation information of samples; then, extracts the pairwise constrain information of samples. Finally, projections are obtained by infusing the two kinds of information with linear weighted way. In contrast to existing semi-supervised dimensionality reduction algorithms on NPE, CSNPE is available with the following characteristics: 1) Sparse reconstruction of local neighborhood of samples cost little because the number of them is limited. 2) CSNPE inherits the great robustness from sparse learning. 3) CSNPE infuses pairwise constrain information with weighted, preserving more discriminant information and local neighborhood sparse reconstruction information. Experiments conducted on real word facial databases demonstrate the effectiveness of the proposed algorithm.

**Keywords:** Information infuse, neighborhood preserving embedding, pairwise constrain, semi-supervised dimensionality reduction, sparse learning.

## 1. INTRODUCTION

Dimensionality reduction is an important step of disposing high-dimensional data in practical applications of data mining. Usually, dimensionality reduction methods are divided into unsupervised dimensionality reduction methods and supervised dimensionality reduction methods. Principal Component Analysis (PCA) [1, 2] and Linear Discriminant Analysis (LDA) [3] are respectively representational dimensionality reduction methods. PCA and LDA are linear, which are not successfully applied in data with nonlinear structures. Although kernel versions of them [4] can be fit for nonlinear data, the problem of kernel functions and parameters remains. Therefore, manifold learning-based dimensionality reduction algorithms are proposed, discovering data lying on nonlinear manifolds. Locally Linear Embedding (LLE) [5] assumes the points within a neighborhood have a linear reconstruction relation and the reconstructing coefficients, which represent the local manifold structure, are preserved in the low dimensional embedding space. ISOMAP [6] maintain the global geodesic distances; and Laplacian eigenmaps (LE) [7] is designed to preserve the relative distances. Neighborhood Preserving Embedding (NPE) [8] is the linear approximations of LLE and aims to find a low-dimensional embedding that optimally preserves the local neighborhood structure on the original data manifold. NPE shares some similar properties with Locality Preserving Projections (LPP) [9, 10] algorithm. However, NPE is a linear approximation to LLE, which leads to some problems in practical face recognition.

Firstly, usually the number of face image samples is smaller than the dimension of face image data space. It is difficult to directly apply NPE to high dimensional matrices because of computational complexity. Moreover, in such case, NPE often suffers from the singularity problem of eigenmatrix, which makes the direct implementation of the NPE algorithm almost impossible. Aiming at overcoming the out-of-sample problem, Kokiopoulou *et al.* proposed Orthogonal Neighborhood Preserving Projections (ONPP) [11]. ONPP can be viewed as a synthesis of PCA and LLE, preserving both the intrinsic neighborhood geometry of the data samples and the global geometry. But ONPP only focuses on the intraclass geometrical information while ignores the interaction of samples from different classes. Zhang *et al.* proposed Discriminative orthogonal neighborhood preserving projections (DONPP) [12] and semi-supervised DONPP (SDONPP) for classification. DONPP takes into account both intraclass and interclass geometries, following the orthogonality property of ONPP. Different from orthogonal linear algorithms, Wang *et al.* proposed a novel approach named Complete Neighborhood Preserving Embedding (CNPE) [13]. CNPE transforms the singular generalized eigensystem computation into two eigenvalue decomposition problems.

However, like LLE, NPE and above algorithms are based on the assumption of the local linearity and constitutes coordinates of one sample from its  $k$  neighbors, which cause that the performance of NPE is not ideal on nonlinear data with such external disturbs as illuminations and occlusions. More precisely, the robustness of NPE is poor. Recently, sparse representation (SR) has been proposed. Recent researches [14] showed that classifier based on SR has natural discriminating power and achieves the best recognition rate on facial databases with varying expression, illumi-

nation, occlusion and disguise, which, thanks to the sparse reconstructive weight reflects some intrinsic geometric properties of the data. Therefore, Cheng *et al.* proposed sparse neighborhood preserving embedding (SNPE) [15]. However, SNPE suffers from a limitation that it does not encode discriminant information. Gui *et al.* proposed a discriminant sparse neighborhood preserving embedding (DSNPE) [16] algorithm by combining SNPE and maximum margin criterion (MMC) [17] methods, which can be viewed as a new algorithm integrating Fisher criterion and sparsity criterion. But MMC is based on the linear covariance matrix of data and could be slow when the dataset is large in practice.

Owing to great convenience in obtaining them and more supervised discriminant information in them, Pairwise Constraint (PC) [18, 19] has attained more attentions. As side-information, pairwise constraint doesn't require people to tell which category an instance belongs to, that is we do not know the exact label of an instance, and only need to judge whether a pair of instances belong to the same class (must-link constraints) or different classes (cannot-link constraints). It can be seen that side information is more general than label information, because we can get side information from label information but it cannot work contrariwise. Moreover pairwise constraint encodes more discriminant information, especially on image data. Nowadays pairwise constraint has been successfully applied in dimensionality reduction [20-26].

Motivated by above analysis, an algorithm called Constraint Sparse Neighborhood Preserving Embedding (CSNPE) for dimensionality reduction is proposed in this paper. CSNPE firstly achieves sparse reconstruction information of samples of within the same class; then exacts pairwise constraint information of samples; finally fuses these two feature information with linear weighted way and projections are obtained. Experimental results on real world facial datasets demonstrate the effectiveness of the proposed algorithm.

CSNPE is available with following merits:

(1) Sparse reconstructions of A considerable amount of samples is significantly expensive. Sparse neighborhood reconstructions in the same class cost little time.

(2) As a semi-supervised dimensionality reduction algorithm, CSNPE not only preserve local sparse reconstructions information but also preserve pairwise constrain information, retaining the sparsity characteristics of sparse learning and side information.

The remainder of this paper is organized as follows: In Section 2 we will introduce related works. A theoretical analysis of CSNPE is given in Section 3. The experimental results on real word datasets will be presented in Section 4, followed by the conclusions in Section 5.

## 2. RELATED WORKS

### 2.1. Sparse Representation and Sparse Reconstruction

Given a set of training samples  $X = \{x_1, x_2, x_3, \dots, x_n\} \in R^{d \times n}$ , sparse representation aims to reconstruct each sample  $x_i$  with else sample, using as few

samples as possible, namely, seek a sparse reconstructive weight vector  $s_i$  for each  $x_i$  through the following minimization problem:

$$\begin{aligned} \min_{s_i} \|s_i\|_0 \\ s.t. x_i = Xs_i \end{aligned} \tag{1}$$

where  $S_{ij}$  denotes the contribution of each  $x_j$  to reconstructing  $x_i$ .  $\|s_i\|_0$  is the pseudo- $\ell_0$  norm which is equal to the number of non-zero components in  $S$ . However, Eq.(5) is NP-hard. The solution of  $\ell_0$  minimization problem is equal to the solution of  $\ell_1$  minimization problem. Therefore, this difficulty can be bypassed by transforming the problem and solving as follows:

$$\begin{aligned} \min_{s_i} \|s_i\|_1 \\ s.t. x_i = Xs_i \end{aligned} \tag{2}$$

Sparse reconstruction seeks a sparse reconstructive weight vector  $x_j$  for each  $x_i$  through the following modified  $\ell_1$  minimization problem:

$$\begin{aligned} \min_{s_i} \|s_i\|_1 \\ s.t. x_i = Xs_i \\ 1 = 1^T s_i \end{aligned} \tag{3}$$

where,  $\|s_i\|_1$  denotes the  $\ell_1$  normal of  $s_i$ ,  $s_i = [s_{i1}, \dots, s_{ii-1}, 0, s_{ii+1}, \dots, s_{in}]^T \in R^n$  is a vector in which  $s_{ij}$  denotes the contribution of each  $x_j$  to reconstructing  $x_i$ , and  $1 \in R^n$  is a vector of all ones.

$$x_i = s_{i1}x_1 + \dots + s_{ii-1}x_{i-1} + s_{ii+1}x_{i+1} + \dots + s_{in}x_n \tag{4}$$

The sparse reconstruction matrix  $S = [s_1, s_2, \dots, s_n]^T$  is attained through calculating  $s_i$ .

### 2.2. Neighborhood Preserving Embedding (NPE)

NPE employs a nearest neighborhood search in the low-dimensional space to yield results to that in the high-dimensional image space as similar as possible. Given samples  $X = \{x_1, x_2, x_3, \dots, x_n\} \in R^{d \times n}$ , NPE attempts to seek an optimal transformation matrix  $T$  to map high-dimensional data  $X$  into low-dimensional data  $Y = T^T X$ , in which the local neighborhood structure of  $X$  can be preserved. There are some basic steps of NPE as follows:

(1) Construct neighborhood adjacent graphic G. The adjacent graphic G is composed of N nodes. Node i corresponds to sample  $x_i$ . If sample  $x_j$  is the neighborhood of sample  $x_i$ , there is a line between  $x_j$  and  $x_i$ . Common methods for constructing neighborhood adjacent graphic G are  $k$ -neighborhood and  $\epsilon$ -neighborhood.

(2) Calculate the sparse reconstructive weights. According to G, Each sample in training samples is reconstructed through the linear combination of neighborhood nodes of the sample as follows:

$$\begin{aligned} \min_T \sum_i \left\| x_i - \sum_{j=1}^k w_{ij} x_j \right\|^2 \\ \text{s.t.} \sum_{j=1}^k w_{ij} = 1 \end{aligned} \quad (5)$$

(3) Projected low-dimensional data Y satisfy:

$$\begin{aligned} \min_T \sum_i \left\| y_i - \sum_{j=1}^k w_{ij} y_j \right\|^2 \\ = \min_T \|Y(I-W)\|^2 \\ = \min_T \left( Y(I-W)(I-W)^T Y^T \right) \\ = \min_T \left( TX(I-W)(I-W)^T XT^T \right) \\ = \min_T (TXMX^T) \end{aligned} \quad (6)$$

where,  $M = (I-W)(I-W)^T$ ,  $I$  represents an identity matrix. In order to make the optimization problem well posed, constrain conditions are introduced as follows:

$$\begin{aligned} \sum_{i=1}^N y_i = 0 \\ \frac{1}{N-1} \sum_{i=1}^N y_i^T y_i = I \end{aligned} \quad (7)$$

According to  $Y = T^T X$ , we replace  $Y$  with  $T^T X$  in Eq. (10) and Eq. (11) and get the objective function

$$\begin{aligned} \min_T T^T XMX^T T \\ \text{s.t.} T^T XX^T T = I \end{aligned} \quad (8)$$

### 2.3. PCFP

Given training samples  $X = \{x_1, x_2, x_3, \dots, x_n\}$ , must-link set  $ML = \{(x_i, x_j) | x_i \text{ and } x_j \text{ are in the same class}\}$  and cannot-link set  $CL = \{(x_i, x_j) | x_i \text{ and } x_j \text{ are not in the same class}\}$ . PCFP aims to find an optimal projection matrix T that maximizes the following function:

$$\begin{aligned} \max_T \left[ \sum_{(x_i, x_j) \in CL} \|(T^T x_i - T^T x_j)\|^2 - \sum_{(x_i, x_j) \in ML} \|(T^T x_i - T^T x_j)\|^2 \right] \\ \text{s.t.} T^T T = I \end{aligned} \quad (9)$$

where  $(\beta X S_\alpha X^T + (1-\beta) P_\alpha) \varphi = \lambda (\beta X X^T + (1-\beta) I) \varphi$  denotes the  $T^{d \times r}$  ( $r < d$ ) norm.

Eq.(9) may be understood in such a sentence that two samples of must-link set in high-dimensional data space should be more near in low-dimensional data space and two samples of cannot-link set in high-dimensional data space should be more further in low-dimensional data space.

## 3. CONSTRAINT SPARSE NEIGHBORHOOD PRESERVING EMBEDDING (CSNPE)

### 3.1. Basic Idea

In the term of information infusion, semi-supervised dimensionality reduction is the process of information infusion, which not only preserves supervised feature information but also preserves unsupervised feature information. According to the information level, information fusions are divided into the data level, the feature level and the decision-making level. Information fusion based on the feature level may make sure relations are maintained among different feature information. Fusion of supervised feature information and unsupervised feature information based on the linear weighted way has proved to be an efficient fusion way of semi-supervised dimensionality reduction [15, 16].

### 3.2. Objective Function

Given samples  $X = \{x_i | x_i \in R^{n_1 \times n_2}, 1 \leq i \leq n\}$  and classes  $C = \{1, 2, 3, \dots, m\}$ .

(1) Firstly, the sparse construction of neighbors is defined as follows:

$$\begin{aligned} \min_{S_i} \|S_i\|_1 \\ \text{s.t.} x_i = \sum_{j=1}^h S_{ij} x_j^{(k)} \\ 1 = 1^T S_i \end{aligned} \quad (10)$$

where  $x_j^{(k)}$  denotes  $x_j$  among the k nearest neighbors of  $x_i$ ,  $S_{ij}$  denotes the contribution of each  $x_j$  to reconstructing  $x_i$ . Eq.(14) gives us description of sparse reconstructive weights of  $x_i$  among the k nearest neighbors of  $x_j$ , preserving local intrinsic geometric property. The weight matrix  $W^S$  is calculated with the sparse weight matrix  $S$ .

(2) We can get  $M^{(s)} = (I-W^S)(I-W^S)^T$  replacing  $S$  with  $W$  in Eq.(10) and get the following objective function:

$$\min_T \sum_i \left\| y_i - \sum_{j=1}^k S_{ij} y_j \right\|^2 = \min_T (TXM^{(s)}XT^T) \quad (11)$$

(3) According to Eq.(9) and Eq.(11), we can draw the following objective function:

$$\begin{aligned} \max_T \left[ \beta \left( \sum_{(x_i, x_j) \in CL} \|(T^T x_i - T^T x_j)\|^2 - \sum_{(x_i, x_j) \in ML} \|(T^T x_i - T^T x_j)\|^2 \right) \right. \\ \left. - (1-\beta) TXM^{(s)}XT^T \right] \\ \text{s.t.} T^T T = I \end{aligned} \quad (12)$$

where  $M^S = (I-W^S)(I-W^S)^T$ .

where  $\beta$  denote the trade-off parameter. Eq.(12) reduce to Eq.(9) when  $\beta = 1$  and Eq.(12) reduce to Eq.(11) when  $\beta = 0$ .



Fig. (1). A group of face images on AR.



Fig. (2). A group of face images on YaleB.

### 3.3. Algorithm Steps

**Input:** training samples  $X = \{x_1, x_2, x_3, \dots, x_n\} \in R^{d \times n}$ , must-link set  $ML$ , cannot-link set  $CL$  and the trade-off parameter  $\beta$ .

**Output:** the projection matrix  $T = [t_1, t_2, \dots, t_d] \in R^{r \times d}$  ( $r < d$ ).

Steps:

- (1) According to Eq.(10), calculate sparse reconstructive weight  $S$  of samples in each class.
- (2) The weight matrix  $P_\alpha$  is obtained with  $S$ .
- (3) Transform Eq.(12) to the generalized eigenvalue problem and calculate the projection matrix  $T^{r \times d}$  ( $r < d$ ).

### 3.4. Computational Complexity Analyses

Given samples  $X = \{x_1, x_2, x_3, \dots, x_n\} \in R^{d \times n}$ , main part cost considerable time in CSNPE as follows :

- (1) The computational complexity of selecting the  $k$  nearest neighbors of each sample cost with  $K$  nearest neighbors (KNN) is  $O(n^2 \times d^2)$ .
- (2) The computational complexity of sparse learning is nearly that of solution of  $l_1$  norm minimization problems which is  $O(d^3)$ [17]. Therefore the computational complexity of solving  $S$  is  $O(n \times d^3)$ .
- (3) The computational complexity of the scatter difference of  $CL$  and  $ML$  is  $O(n \times d^2)$ .
- (4) The eigen-problem on a symmetric matrix can be efficiently computed by the singular value decomposition (SVD) which is  $O(d^3)$ .

To sum up, the computational complexity of CSNPE is  $O(n^2 \times d^2 + 2n \times d^2 + d^3)$ .

## 4. EXPERIMENTS AND ANALYSES

In this section, in order to evaluate the performance of CSNPE, we compare it with the NPE, SNPE, DSNPE and PCFP on AR and YaleB face databases which are disturbed by external environment.

### 4.1. Face Datasets

(1) AR: the facial database consists of over 4000 facial images of 126 individuals. For each individual, 26 pictures were taken in two sessions that were separated by two weeks and each session contains 13 images, which includes front view of faces with different expressions, illuminations and occlusions. Fig. (1) shows a group of face images on AR. For computational convenience, we firstly resize these face images to  $30 \times 30$  pixels.

(2) YaleB: the face database contains 2414 front-view facial images of 38 individuals. For each individual about 64 pictures were taken under various laboratory-controlled lighting conditions. A group face images of YaleB are shown in Fig. (2). In our experiment, these images are resized to  $32 \times 32$  pixels

### 4.2. Experimental Settings

SNPE, PCFP, SDONPP and DSNPE are compared with the proposed OTSNPE for analyses on performances of CSNPE. Apart from the trade-off parameter  $\beta$ , the detail settings of other parameters on algorithms are shown in Table 1.

Table 1. The detail settings of parameters on algorithms.

Algorithms Name	Parameters Settings
SNPE	$\kappa = 7$
PCFP	no
SDONPP	$\lambda = 0.5$
DSNPE	$\gamma = 0.5$
CSNPE	$\kappa = 7, \beta = 0.5$

Where, the parameter  $k$  denotes the neighborhood size.

The Nearest Neighbor Classifier (NNC) is applied for calculating the recognition rate. For computational convenience, we firstly resize face image to  $30 \times 30$  pixels.

According to Eq. (12), CSNPE is sensitive to the number of training samples and the scale of pairwise sets. We randomly select  $T$  images from each group of faces for training samples and remainder for testing. In order to evaluate the performance of our algorithm, we select respectively re-

tained feature with the increment  $D$  and calculate corresponding recognition accuracy rates. Moreover, pairwise sets with the certain number CS are created randomly.

### 4.3. Experimental Results and Analyses

#### 4.3.1. Analyses on the Performance of Dimensionality Reduction

Concrete experimental results under different  $T$  and CS are shown in Figs. (3 and 4).

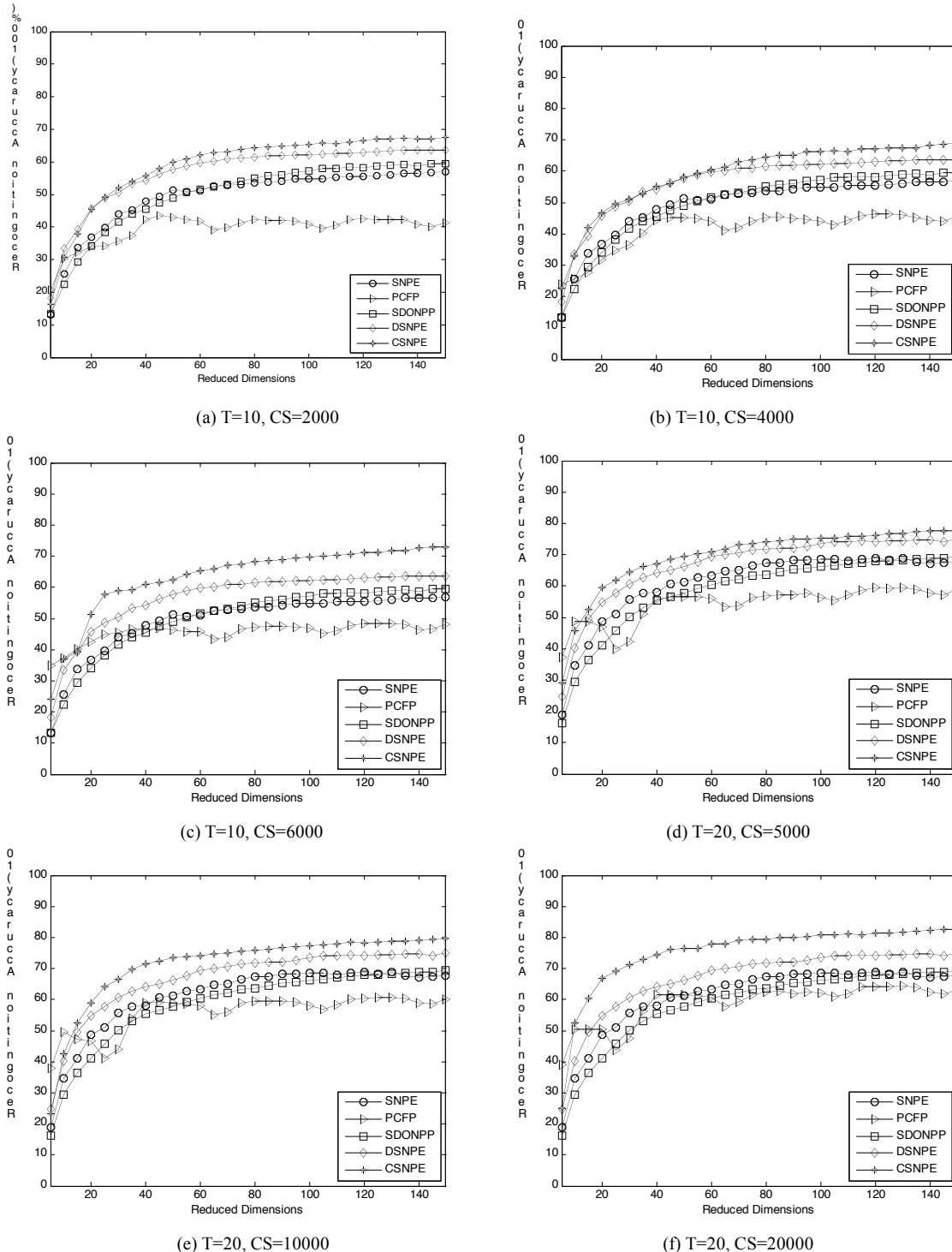


Fig. (3). The recognition rate VS. dimensions on AR under  $T$  and CS.

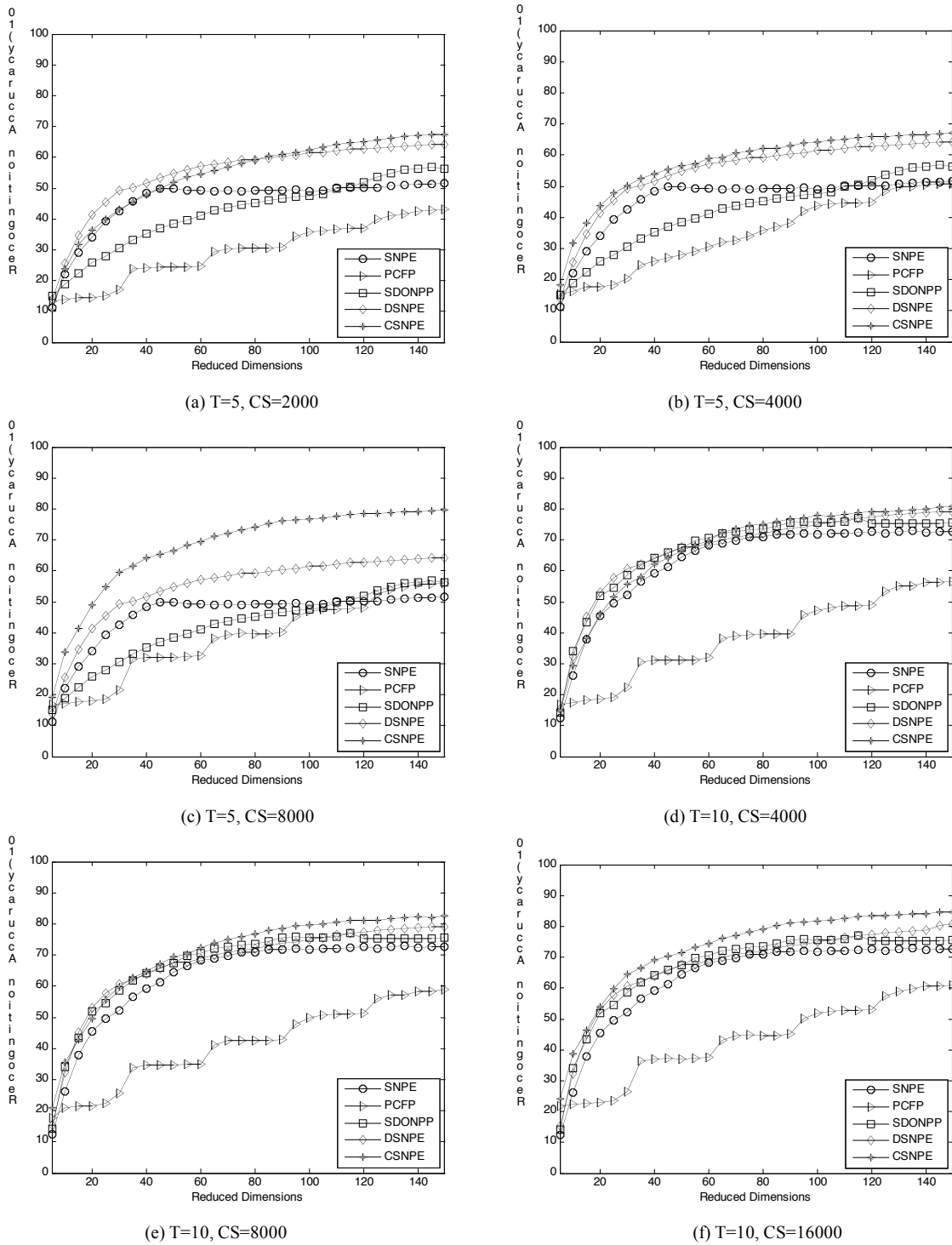


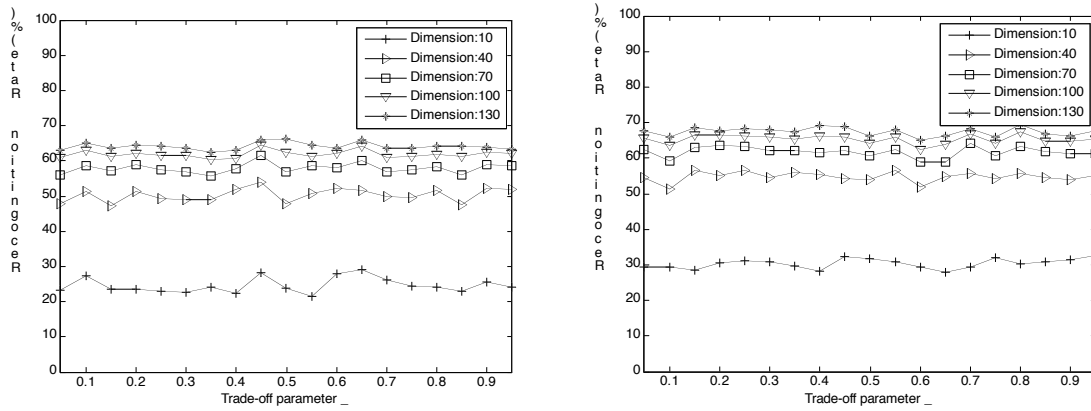
Fig. (4). Recognition rate VS. dimensions on YaleB under T and C.

From above Figs. (3 and 4), we can draw conclusions as follows:

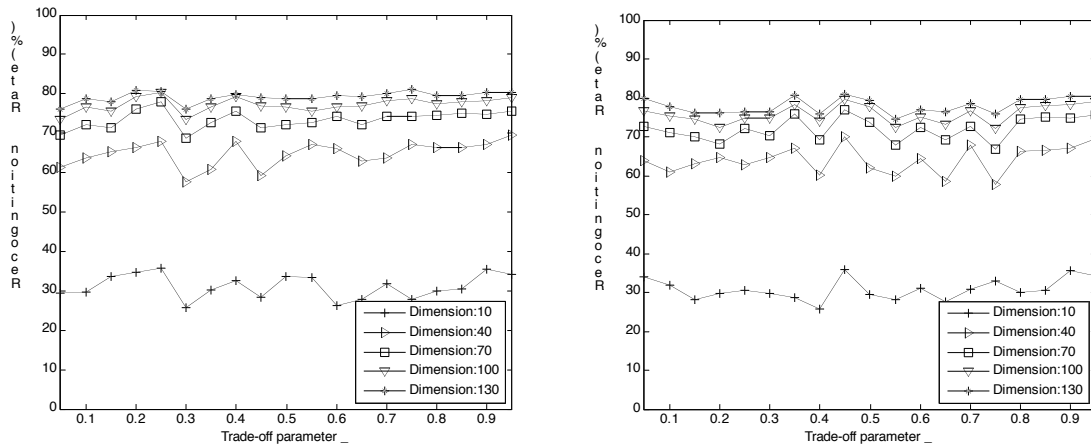
- (1) As a semi-supervised dimensionality reduction algorithm, Our CSNPE algorithm outperforms them obviously, which is attributed to CSNPE inheriting their characteristics and infusing them efficiently.
- (2) SDONPP generalize DONPP by introducing new part optimizations based on unlabeled samples and then incorporating them into the whole alignment stage, with less

exact discriminant information of samples including unlabeled samples and labeled samples than CSNPE.

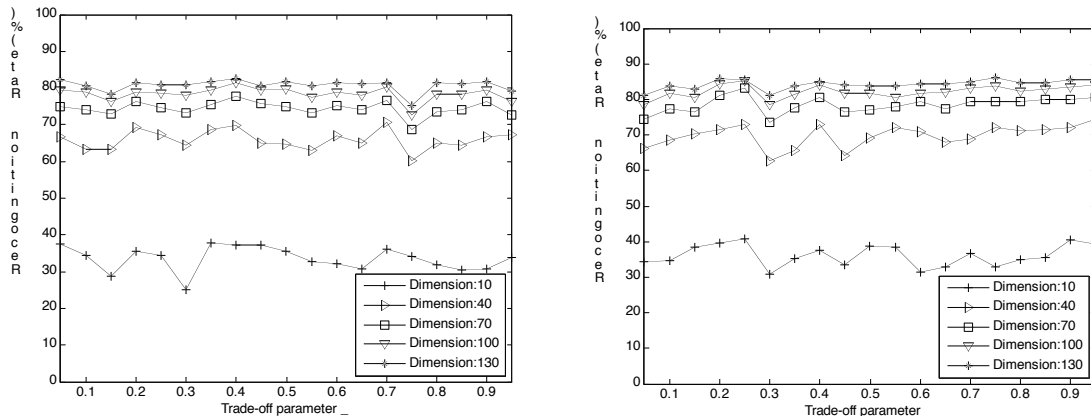
- (3) Similarly, DSNPE combines sparsity criterion and maximum margin criterion (MMC) together to project the input high-dimensional image into a low-dimensional feature vector, sharing sparsity learning with CSNPE. CSNPE is superior to DSNPE, which is illustrated by that the constrains information of samples is available for power discriminant information than MMC.



(a) T=5, CS=2000 (b) T=5, CS=4000



(c) T=5, CS=8000 (d) T=10, CS=4000



(e) T=10, CS=8000 (f) T=10, CS=16000

Fig. (5). Recognition rate VS. Trade-off parameter  $\beta$  on AR with different dimension under T and CS.

(4) The size of constrain sets play an impact on the performance of CSNPE. More constrain sets contain more supervised information and help CSNPE to get more performance, which is illustrated by comparisons under the same training samples on AR and YaleB.

4.3.2. Analyses on the Effect of the Trade-off Parameter  $\beta$

In order to further verify the effect of the trade-off parameter  $\beta$  in Eq.(12), we set  $\beta$  from 0.1 to 0.95 with the

increment 0.05 and calculate corresponding recognition rates.

From Figs. (5 and 6), following conclusions are drawn:

(1) The recognition rate fluctuates greatly when reduced dimensions are low and it is smooth when reduced dimensions are high, which illustrate that the trade-off parameter  $\beta$  is sensitive to reduced dimension.

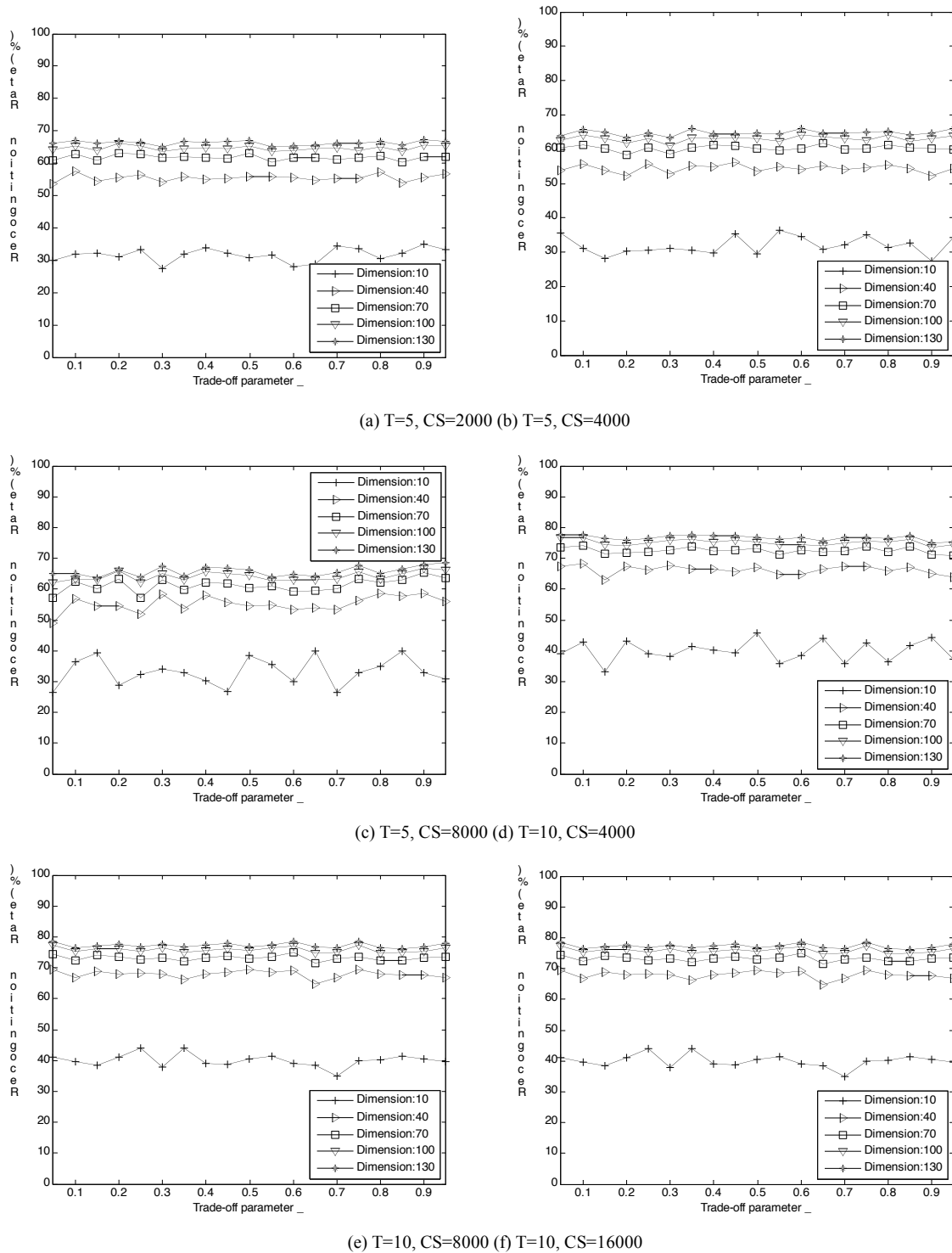


Fig. (6). The recognition rate VS. The trade-off parameter  $\beta$  on YaleB with different dimension under T and CS.

(2) When training samples of each individual increase from 10 to 20 on AR, there is no change on the curve, describing the effect of the trade-off parameter  $\beta$  on the recognition rate when reduced dimensions are high, which also happen with YaleB.

To sum up, the trade-off parameter  $\beta$  is set to 0.5 in our CSNPE, attaining nearly most of the performance without spending plenty of time in calculating the optimal value.

### 5. CONCLUSION

In the paper, in order to attain power discriminant supervised information, Constraint Sparse Neighborhood Preserving Embedding (CSNPE) is proposed for semi-supervised dimensionality reduction. CSNPE considers explicitly the local neighborhood sparse reconstruction by the objective function (10), which captures local nonlinear structure based on sparse representation. Meanwhile, the power supervised



constrain feature information is infused in the objective function (12), ensuring CSNPE to attain power discriminant information than other supervised feature. Experimental results on AR and Yale demonstrate the effectiveness of our algorithm.

However, as a vector-based reduction algorithm, CSNPE fails to take into account the spatial relation of image pixels. How to extend it to tensor version to deal with tensor data is among the future work.

## CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

## ACKNOWLEDGEMENTS

Declared none.

## REFERENCES

- [1] M. Turk and A.P. Pentland, "Face recognition using eigenfaces". In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 586-591, 1991.
- [2] M. Turk and A.P. Pentland, "Eigenfaces for recognition," *J. Cognit. Neurosci.*, vol. 3, no. 1, pp. 71-86, 1991.
- [3] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 7, pp. 711-720, 1997.
- [4] W. Tang and S. Zhong, "Pairwise constraints-guided dimensionality reduction", In: Proc. of SDM Work-shop on Feature Selection for Data Mining, pp. 345-355, 2006.
- [5] C. Cortes, M. Mohri, and A. Rostamizadeh, "Two-stage learning kernel algorithms", In: Proc. of the 27th International Conference on Machine Learning (ICML 2010), pp. 249-256, 2010.
- [6] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear bedding". *Science*, vol. 290, no. 5500, pp. 2323-2326, 2000.
- [7] J.B. Tenenbaum, V.D. Silva, and J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction", *Science*, vol. 290, no. 5500, pp.2319-2323, 2000.
- [8] X. F. He, S. C. Yan, Y. X. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using Laplacianfaces", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328-340, 2005.
- [9] M. Belkin and P. Niyogi, "Laplacian Eigenmaps spectral techniques for embedding and clustering", In: Proc. of the Conference on Advances in Neural Information Processing Systems, pp. 585-591, 2001.
- [10] E. Kokiopoulou and Y. Saad, "Orthogonal neighborhood preserving projections: A projection-based dimensionality reduction technique," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2143-2156, 2007.
- [11] J.Wright, A.Yang, S.Sastry, and Y.Ma, "Robust face recognition via sparse representation", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp.210-227, 2009.
- [12] D. Zhang, Z. Zhou, and S. Chen, "Semi-Supervised dimensionality reduction," In: Proc. of the SDM, pp. 629-634, 2007.
- [13] H. Cevikalp, J. Verbeek, F. Jurie, and A. Kläser, "Semi-Supervised dimensionality reduction using pairwise equivalence constraints", In: Proc. of the VISAPP, pp. 489-496, 2008.
- [14] B. A. Hillel T. Hertz, N. Sental, and D. Weinshall, "Learning a mahalnobis metric from equivalence Constraints", *J. Mach. Learning Res.*, vol. 6, no. 6, pp. 937- 965, 2006.
- [15] J. Wei and H. Peng, "Neighbourhood preserving based semi-supervised dimensionality reduction", *Electro. Lett.*, vol. 44, no. 20, pp.1190-1191, 2008.
- [16] G. Yu, H. Peng, J. Wei , and Q. Ma, "Robust Locality Preserving Projections with Pairwise Constraints", *J. Comput. Info. Syst.*, vol. 6, no. 5, pp.1631-1636, 2010.
- [17] B. Cheng, J. C. Yang, S. C. Yan, Y. Fu, and T. S. Huang, "Learning with l(1)-graph for image analysis," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 858-866, 2010.
- [18] X. F. He, D. Cai, and S. C. Yan, H. J. Zhang, "Neighborhood Preserving Embedding", In: Proc. of the 10th IEEE International Conference on Computer Vision. Washington: IEEE Computer Society Press, pp. 1208-1213, 2005.
- [19] X. He and P. Niyogi, "Locality Preserving Projections", In: Proc. of Conference on Advances in Neural Information Processing Systems, pp. 153-160, 2003.
- [20] T. Zhang, K. Huang, X. Li, J. Yang, and D. Tao, "Discriminative orthogonal neighborhood-preserving projections for classification", *IEEE Trans. Syst. Man. Cybern. B Cybern.*, vol. 40, no. 1, pp. 253-263, 2010.
- [21] Y. Wang and Y. Wu, "Complete neighborhood preserving embedding for face recognition", *Pattern Recognit.*, vol. 43, no. 1, pp.1008-1015, 2010.
- [22] J. Gui, Z. Sun, W. Jia, R. Hu, Y. Lei, and S. Ji, "Discriminant sparse neighborhood preserving embedding for face recognition", *Pattern Recognit.*, vol. 44, no. 8, pp. 2884-2893, 2012.
- [23] G.F. Lu, Z. Lin, and Z. Jin. Face recognition using discriminant locality preserving projections based on maximum margin criterion", *Pattern Recognit.*, vol. 43, no. 10, pp. 3572-3579, 2010.
- [24] E. Xing, A. Ng, M. Jordan, and S. Russell, "Distance metric learning with application to clustering with side information", *Advances in Neural Information Processing Systems*. MIT Press, pp. 505-512, Cambridge, MA.
- [25] R. Yan, J. Zhang, J. Yang, and A. Hauptmann, "A discriminative learning framework with pairwise constraints for video object classification", In: Proc. of the IEEE computer society conference on computer vision and pattern recognition, pp. 284-291, 2004.
- [26] Y. Peng and D. Q. Zhang, "Semi-Supervised canonical correlation analysis algorithm", *J. Softw.*, vol. 19, no. 11, pp. 2822-2832, 2008.

Received: September 22, 2014

Revised: November 30, 2014

Accepted: December 02, 2014

© Shuhua Xu; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.