Joint Sparse Neighborhood Preserving Embedding

Haibiao Liu¹, Zhihui Lai²,∗ and Yudong Chen³
Shenzhen University, Guangdong, China

*Corresponding author e-mail: laizhihui@szu.edu.cn

Abstract. Feature extraction and feature selection are the two most commonly used methods for dimensionality reduction in the field of machine learning. A new method can do feature extraction and selection simultaneously was presented in this paper, we named the method as Joint Sparse Neighborhood Preserving Embedding (JSNPE). Our model can retain the local relationship when the data were projected in the new subspace, because we construct an adjacency graph matrix based on the neighborhood relation of data points. What’s more, our model can obtain jointly sparse projection by introducing the $L_{2,1}$-norm regularization on the projection matrix. By using Lagrange multiplier method, we iteratively solve the objection function. We get the experimental results on two public face dataset to validate JSNPE is a better method than some well-known methods that used to feature extraction and selection.

1. Introduction
The dimensionality of a facial image can easily be very high in practice. The redundant features of the image data result in large waste of computing and storage resources[1].

Feature extraction and feature selection can obtain dimensional reduction from two different perspectives. At present, all feature extraction methods can be roughly fall into two types according to different purposes when we projected the data to the new subspace, the first kind is to try to retain the global information of data, including the classical principal component analysis (PCA)[2], linear discriminant analysis (LDA)[3] and multidimensional scaling problems (MDS)[4]. The second kind is to preserve the local manifold structure information of original data, including locally linear embedding (LLE)[5], isometric feature mapping (ISOMAP)[6], Laplacian Eigenmap (LE)[7], Locality preserving projections (LPP)[8] and so on.

PCA, LDA and MDS only aiming to preserve the data global structure, which are not completely satisfied the need of reality when the data is distributed on a manifold. It is more likely that the face images lie on a manifold[9]. Therefore, many manifold learning methods were designed to find the manifold structure of data in dimensionality reduction. Such as ISOMAP, LE, LLE and so on. Although these methods all can preserve the neighborhood relationship, these methods are computationally expensive because they are nonlinear. Moreover, these methods get the projections only on the training data points, but they are hard to extended for the testing data. Therefore, a linear method named Neighborhood Preserving Embedding (NPE) was proposed[10]. And furthermore, they introduce the $L_{2,1}$-norm on the learned projection and proposed the sparse version of NPE[10], the function of simultaneous feature selection and extraction can be achieved.

Therefore, we presented a new method that has both advantages, it named Joint Sparse Neighborhood Preserving Embedding (JSNPE). By introducing the $L_{2,1}$-norm regularization of the
learned projection matrix, it can get jointly row sparse projection matrix and can act on feature selection. There is the main work of this paper:

(1) The new method named Joint Sparse Neighborhood Preserving Embedding (JSNPE) that can get jointly sparse projection was presented.

(2) Two experimental result verify that JSNPE is better than the classical sparse feature extraction and selection methods used to dimensionality reduction in the application of image recognition.

2. Objective Function and its Solution

The objective function of JSNPE includes regression item and penalty item, the regression item is the $L_2$-norm loss error regression, the penalty item is the $L_{2,1}$-norm projection matrix, and the objection function have matrix constraints:

$$\min J(A,P,W) = \sum_i \|x_i - \sum_j A^{\top}P_j x_j W_{ij}\|^2_2 + \lambda \|P\|_{2,1} \ s.t. \ A^{\top}A = I, \ \sum_j W_{ij} = 1$$

where $x_i$ and $x_j$ is a column vector denoted a sample of the data, $i$ and $j$ is their serial number, the matrix $A$ is a basic matic for reconstructing the sample in a new subspace, the matrix $P$ is the projection matrix that transforming the sample from the high-dimensional representation space to the low-dimensional representation space. The matrix $W$ is the weight matrix to construct the neighboring samples. The parameter $\lambda$ is a regular parameter for the regularization terms.

By the mathematics operations, to minimize the matrix can transform to the problem of minimizing the trace optimization of matrix. Then we get the formula as follows:

$$\min \text{tr} \left( X^{\top}X - AP^T X(W + W^T)X^T + P^T X^T W^T W^T X + \lambda \text{tr}(P^TGP) \right) \ s.t. \ A^{\top}A = I, \ \sum_j W_{ij} = 1$$

where the diagonal matrix $G_{ij} = 1$ and $P^i$ denotes the $i$-th row of matrix $P$. We use Lagrange multiplier method to solve the optimal solution. The algorithm mainly consists of three steps. Firstly, fix $A$ to compute $P$, then update $P$ and fix $P$ to compute $A$, after getting the optimal solution $P$ and $A$ in one iteration, we need to update the $W$; then iterate these three steps until convergence.

When get matrix $A$, taking the partial deviation of (2), we can solve the matrix $P$

$$-AX(W + W^T)X^T + 2XW^T W^T X P + 2\lambda GP = 0$$

This gives

$$P = 2(XW^T W^T X + \lambda G)^{-1} AX(W + W^T)X^T$$

(3)

When local solution of $P$ is obtained, then use (3) to update the $A$. Hence the minimum problem in (2) is translated into the maximum problem:

$$\max \text{tr} \left( A^T P X(W + W^T)X^T \right) \ s.t. \ A^T A = I$$

(4)

Then use the SVD of $P^T X(W + W^T)X^T$ to obtain the optimal solution in a iteration. Let

$$P^T X(W + W^T)X^T = \hat{U} \hat{D} \hat{V}^T$$

Then the optimal solution of (4) is

$$A = \hat{U} \hat{V}^T$$

(6)

In one iteration, when we obtain the optimal $A$ and $P$, we need to update the weight matrix $W$ using the local linear reconstruction relationship as the on in LLE algorithm.

3. JSNPE Algorithm

We give the pseudocode of JSNPE algorithm in this section:

BEGIN

Input: the training dataset matrix $X$, the weight graph matrix $W$, the dimensionality $d$ of the sample, the desired dimensionality $k$ of matrix $A$ and $P$, and the regularization parameter $\lambda$.

Program:
1: Initialize the basic matrix $A$, the projection matrix $P$.
2: For $i = 1: T$
   - Compute the weight matrix $W$ using local linear reconstruction as in LLE.
   - Fix $A$, Compute $P$ using (3)
   - Fix $P$, Compute $A$ using (6).
End
3: output the final matrix $P$ for feature extraction.
4: reducing dimension by multiplying the projection matrix $P$: $[y_1, y_2, ..., y_n] = P^T[x_1, x_2, ..., x_n]$.

END

4. Experiment
In this section, in order to verify the effectiveness of our algorithm JSNPE for feature extraction and selection and recognition, we chose some classical and recently new methods to compared with JSNPE on two public face database, they are NIR and YALE face database. We compared it with PCA, LPP, NPE, SPCA, FSSL[11], L21R21, UDFS[12], SAIR[13], FOLPP[14], JELSR[15]. When we get the new representation of the sample in the subspace, we all use KNN classifier to do classification and calculate recognition rate.

In all the experiment, we divided the data into two parts: training set and testing set, use $l (l = 4, 5, 6)$ to denote that $l$ images used to train, and regularization parameter $\gamma$ belongs to the set $\{10^{-1}, 10^{-2}, 10^{-3}, 10^0, 10^1\}$.

4.1 Performance comparisons on the NIR database
A total of 3920 images were collected from 196 individuals in the NIR database. The pixels of all images are $64 \times 64$.

We get the average experimental result on 10 time running. Table 1 shows the average recognition rates of the all methods when the training samples is different, Fig. 1 shows average recognition rates of the all methods when the dimensions of the projection changes.

| samples | PCA | LPP | NPE | SPCA | FSSL | L21R21 | UDFS | SAIR | FOLPP | JELSR | JSNPE |
|---------|-----|-----|-----|------|------|--------|------|------|-------|-------|-------|
| 4       | 86.1| 83.39| 86.86| 84.78| 78.72| 81.43| 87.02| 87.02| 68.57| 84.51| 87.32 |
|         | ±1.96| ±2.41| ±2.11| ±2.19| ±1.30| ±2.69| ±2.16| ±2.11| ±3.29| ±2.35| ±2.05 |
| 5       | 87.14| 87.73| 89.02| 88.52| 83.87| 84.33| 89.58| 89.68| 70.53| 88.22| 89.78 |
|         | ±2.69| ±2.40| ±1.98| ±1.45| ±2.07| ±2.97| ±2.16| ±2.13| ±4.00| ±2.08| ±2.11 |
| 6       | 90.62| 90.79| 90.63| 90.64| 87.35| 86.08| 91.12| 91.12| 74.04| 90.41| 91.32 |
|         | ±2.34| ±2.39| ±1.95| ±1.25| ±2.38| ±3.15| ±2.26| ±2.36| ±3.24| ±2.06| ±2.35 |
|         | 60 | 60 | 50 | 90 | 165 | 155 | 150 | 150 | 50 | 150 | 130 |
Fig.1. When the training number in a class is 5, the recognition rates (%) variation curve of all methods

4.2 Performance comparisons on the YALE Database

There are 165 images that were collected from 15 people on the Yale face database. All images are normalized to 50×40 pixels.

The table 2 listed the average performance of all methods on the 10 times experimental results, and when the dimension of projection dimension increases gradually, Fig.3 show the changing regularity of the performance of every method.

| samples | PCA    | LPP    | NPE    | SPCA   | FSSL   | UDFS   | FOLPP  | JELSR  | JSNPE  |
|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 4       | 91.14  | 84.19  | 90.86  | 91.43  | 91.14  | 90.38  | 86.57  | 89.62  | 93.33  |
|         | ±3.65  | ±5.41  | ±3.46  | ±3.46  | ±3.30  | ±4.10  | ±3.86  | ±4.13  | ±2.98  |
|         | 20     | 40     | 20     | 20     | 10     | 40     | 50     | 150    | 25     |
| 5       | 92.78  | 88.89  | 92.78  | 93.33  | 91.89  | 92.00  | 91.00  | 90.56  | 94.44  |
|         | ±3.81  | ±2.87  | ±4.56  | ±3.77  | ±4.26  | ±4.25  | ±5.67  | ±4.22  | ±3.68  |
|         | 25     | 35     | 30     | 25     | 10     | 40     | 50     | 150    | 35     |
| 6       | 92.53  | 92.53  | 93.47  | 92.67  | 93.33  | 91.47  | 94.40  | 90.67  | 94     |
|         | ±3.69  | ±3.33  | ±3.49  | ±3.93  | ±3.82  | ±4.22  | ±4.53  | ±4.46  | ±3.60  |
|         | 25     | 20     | 35     | 25     | 10     | 40     | 50     | 125    | 35     |
5. Conclusion

We proposed a new discriminant feature extraction and selection method named Joint Sparse Neighborhood Preserving Embedding (JSNPE) for jointly sparse projection learning and feature selection. The $L_{2,1}$ norm of projection matrix $P$ was introduced, and iterative solve the optimization solution by Lagrange multiplier method. The critical step of solving the objective function is to solve the three matrices: the reconstruction matrix, orthogonal matrix and sparse projection matrix. Through the analysis of the experimental result on two public face datasets, compared with the traditional linear global feature extraction methods and manifold-based feature learning methods, JSNPE can obtain higher recognition rate, because $L_{2,1}$ norm based method can provide more powerful properties of feature selection compared with the $L_1$ and $L_2$ norm based feature selection methods.

References

1. Kamishima T, Toshihiro K. A Survey of Recent Clustering Methods for Data Mining (Part 2); Challenges to Conquer Giga Data Sets and the Curse of Dimensionality[J]. Journal of Japanese Society for Artificial Intelligence, 2003, 18:170-176.
2. K. Pearson, “On lines and planes of closest fit to systems of points in space,” Phil. Mag., vol. 6, no. 6, pp. 559–572, 1901
3. K. Fukunaga, Introduction to Statistical Pattern Recognition, 2nd ed. San Diego, CA, USA: Academic, 1990.
4. P. Paatero, “Least squares formulation of robust non-negative factor analysis,” Chemometr. Intell. Lab. Syst., vol. 37, pp. 23–35, 1997.
5. S.T. Roweis and L.K. Saul, “Nonlinear Dimensionality Reduction by Locally Linear Embedding,” Science, vol. 290, Dec. 2000.
6. J.B. Tenenbaum, V. de Silva, and J.C. Langford.: A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science, vol. 290, Dec. 2000.
7. Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering[J]. Advances in Neural Information Processing Systems, 2001, 14(6):585-591.
8. He, X., & Niyogi, P. (2004).: Locality preserving projections. Neural Information Processing Systems, 16, 153.
9. Seung H S, Lee D D. Cognition. The manifold ways of perception. [J]. Science, 2000, 290(5500):2268-2269.
10. Xiaofei He, Deng Cai, Shuicheng Yan, and HongJiang Zhang. Neighborhood preserving embedding. In ICCV, pages 1208–1213, 2005.
11. Q. Gu, Z. Li, and J. Han, “Joint feature selection and subspace learning,” in Proc. 22nd Int. Joint Conf. Artif. Intell. (IJCAI), Barcelona, Spain, 2011, pp. 1294–1299.
12. Yang Y, Shen H T, Ma Z, et al. l 2,1 -norm regularized discriminative feature selection for unsupervised learning[C]// International Joint Conference on Artificial Intelligence. AAAI Press, 2011:1589-1594.
13. Z. Ma, Y. Yang, N. Sebe, K. Zheng, and A. G. Hauptmann, “Multimedia event detection using a classifier-specific intermediate representation,” IEEE Trans. Multimedia, vol. 15, no. 7, pp. 1628–1637, Nov. 2013.
14. Wang R, Nie F, Hong R, et al. Fast and Orthogonal Locality Preserving Projections for Dimensionality Reduction. [J]. IEEE Transactions on Image Processing, 2017, PP(99):1-1.
15. Hou C, Nie F, Li X, et al. Joint embedding learning and sparse regression: a framework for unsupervised feature selection[J]. IEEE Trans Cybern, 2017, 44(6):793-804.