Combining Locality Preserving Projection with Global Information for Efficient Recognition

Heesung Lee

Department of Railroad Electrical and Electronics Engineering, Korea National University of Transportation,
 Uiwang-si, Korea

Ijfis

Abstract

This paper proposes a new feature extraction scheme, combining global and local features. The proposed method uses principal component analysis (PCA) and linear discriminant analysis (LDA) for global property and locality preserving projections (LPP) for local property of the pattern. PCA and LDA are known for extracting the most descriptive ones after projection while LPP is known for preserving the neighborhood structure of the data set. The proposed combining method integrates global and local descriptive information and finds an efficient set of alternatives beyond PCA, LDA and LPP in the parametric space. Further, In order to find optimal parameters, the genetic algorithm (GA) is employed. Experiments are performed with four data sets selected in UCI machine learning repository to show the performance of the proposed algorithm.

Keywords: PCA, LDA, LPP, GA, Feature space, UCI machine learning repository

1. Introduction

For general pattern recognition systems, the patterns can be represented as certain features [1]. It is also well-known fact that the irrelevant or redundant features degrade the classification accuracy, so constructing an appropriate feature subset is essential to build a good classification algorithm. Therefore, developing recognition scheme of real world objects or concepts using computational methods, the selection of an appropriate set of features is of considerable importance [2]. In this paper, we propose a new feature extraction scheme combining locality preserving projection with global information. The proposed method uses principal component analysis (PCA) and linear discriminant analysis (LDA) for global property and locality preserving projections (LPP) for local property of the pattern to improve the classification ability. PCA and LDA are well known for preserving the most descriptive ones after projection while LPP is known for preserving the neighborhood structure of the data set [3]. Our combining method can integrate global and local descriptive information and find a richer set of alternatives beyond PCA, LDA and LPP. In order to find the hybrid features adaptively and find optimal parameters, the genetic algorithm (GA) is employed. GAs have been used for many scientific and engineering optimization and search problems [4]. GAs provide an adaptive and robust computational procedure modeled on the mechanics of natural genetic systems [5]. GAs can be considered as a good solution for finding a richer set of alternatives beyond PCA, LDA and LPP in search space, because they are complicated problem in a large

Received: May 8, 2018

Revised : Jun. 1, 2018

Accepted: Jun. 7, 2018

Correspondence to: Heesung Lee

(hslee0717@ut.ac.kr)

©The Korean Institute of Intelligent Systems

©This is an Open Access article distributed 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 original work is properly cited.

search space [6].

We introduce some preliminaries including PCA, LDA, LPP and GA briefly in Section 2. The proposed global and local features combining method based on GA is explained in Section 3. In Section 4, we assess the experimental performance of our proposed system. Finally, the conclusion is drawn in Section 5.

2. Preliminaries

2.1 Principal Component Analysis

PCA can choose a dimensionality reducing linear projection that maximizes the scatter of all projected samples. Let us assume that an object is represented by the n -dimensional features

$$X = [x_1, x_2, \dots, x_n]^T, \quad (1)$$

and we have a set of objects

$$Z = [X_1, X_2, \dots, X_N], \quad (2)$$

where each object belongs to one of C classes. If the total scatter matrix S_T is defined as

$$S_T = \sum_{j=1}^N (X_j - \mu)(X_j - \mu)^T, \quad (3)$$

where $\mu \in R^n$ is the mean vector of all objects and computed by

$$\mu = \frac{1}{N} \sum_{j=1}^N X_j. \quad (4)$$

The objective of PCA is maximizing the determinant of the total scatter matrix of the projected samples as follows:

$$\arg_W \max S_T. \quad (5)$$

The new feature vectors $Y_j \in R^m$ for $j = 1, 2, \dots, N$ are defined by the following linear transformation

$$Y_j = W^T X_j \text{ for } j = 1, 2, \dots, N, \quad (6)$$

where $W \in R^{n \times m}$ is a matrix with orthonormal columns. W_{opt} that maximizes the objective function is given by the set of n -dimensional eigenvectors of S_T corresponding to the m -largest eigenvalues. A drawback of PCA is that the scatter being maximized is due not only to the between-class scatter that is useful for classification, but also to the within-class scatter that,

for classification purposes, is unwanted information [7].

2.2 Linear Discriminant Analysis

The LDA is a linear projection method that maximizes the ratio of determinant of between-class covariance matrix to the determinant of within-class covariance matrix [8]. The within-class scatter matrix S_W and the between class scatter matrix S_b are defined by

$$S_W = \sum_{c=1}^C \sum_{j=1}^{N_c} (X_j^c - \mu_c) \cdot (X_j^c - \mu_c)^T, \quad (7)$$

and

$$S_b = \sum_{c=1}^C (\mu_c - \mu) \cdot (\mu_c - \mu)^T, \quad (8)$$

where X_j^c is the j th object of class c , N_c is the number of objects in class c , and $\mu_c \in R^n$ is the mean of class c .

S_W and S_b represent the scatter of features around the mean of each class and scatter of features around the overall mean for all classes, respectively [9]. The objective of LDA is to maximize S_b while minimizing S_W as follows:

$$\arg_W \max \frac{S_b}{S_W}. \quad (9)$$

It has been proven that that this ratio is maximized when the column vectors of the projection matrix W_{opt} are the eigenvectors of $S_W^{-1} S_b$.

2.3 Locality Preserving Projections

Though the information of features is important, the local structure is also significant in many real-world applications [3, 10]. LPP is proposed for learning a locality preserving subspace and seeks to preserve the intrinsic geometry of the samples and local structure. The objective function of LPP is as follows:

$$\min \sum_{ij} (Y_i - Y_j)^2 S_{ij}, \quad (10)$$

where S is a similarity matrix and computed by

$$S_{ij} = \begin{cases} e^{(-\|X_i - X_j\|^2/t)}, & \|X_i - X_j\|^2 < \varepsilon, \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

In (11), ε is a radius of the local neighborhood and defines the locality. Following some simple algebraic steps, we see that

$$\begin{aligned}
 & \frac{1}{2} \sum_{ij} (Y_i - Y_j)^2 S_{ij} \\
 &= \frac{1}{2} \sum_{ij} (W^T X_i - W^T X_j)^2 S_{ij} \\
 &= \sum_{ij} W^T X_i S_{ij} X_i^T W - \sum_{ij} W^T X_i S_{ij} X_j^T W \\
 &= W^T Z D Z^T W - W^T Z S Z^T W \\
 &= W^T Z (D - S) K^T W \\
 &= W^T Z L Z^T W,
 \end{aligned} \tag{12}$$

where D is a diagonal matrix and its entries are column sums of S and $L = D - S$ is the Laplacian matrix. Because the bigger value D is, the more important is Y_i , following constraint can be imposed

$$Y^T D Y = 1 \rightarrow W^T Z D Z^T W = 1. \tag{13}$$

Finally, we can obtain reduced optimal problem as follows:

$$\begin{aligned}
 & \arg_W \min W^T Z L Z^T W, \\
 & W^T Z D Z^T W = 1.
 \end{aligned} \tag{14}$$

The W_{opt} that minimizes the objective function is given by the m -smallest eigenvalues to the generalized eigenvalue problem [11]

$$Z L Z^T W = \lambda Z D Z^T W. \tag{15}$$

2.4 Genetic Algorithm

The GA typically maintains a fixed size population of individuals that represent the set of solution candidates for the optimization problem to be solved. The goodness of each candidate solution is evaluated using a fitness function. The population of the GA evolves with a set of genetic operators being applied. The basic genetic operators are selection, crossover and mutation. In the selection process, some individuals are selected to be copied into a tentative next population. An individual is selected for the next population with the number of copies proportional to the fitness value. The copied individuals are altered to search for a global optimal solution in the crossover and mutation processes. The GA is simple yet provides adaptive and robust optimization methodology. Given below is the basic configuration of the GA [6].

procedure Genetic Algorithm

```

begin
  initialize  $P(t)$ ;
  while termination-condition not satisfied do
    begin
      evaluate  $P(t)$ ;
      select  $P(t+1)$  from  $P(t)$ ;
      crossover  $P(t+1)$ ;
      mutation  $P(t+1)$ ;
       $t = t + 1$ ;
    end
  end

```

3. Global and Local Information Combine

We propose a method which uses PCA and LDA for global property and LPP for local property of the pattern to improve the classification ability. It can integrate global and local descriptive information and find a richer set of alternatives beyond PCA, LDA and LPP. Based on them, we can obtain new features which preserve global and local structure, simultaneously as follows:

$$H_i = \omega_1 Y_i^1 + \omega_2 Y_i^2 + \omega_3 Y_i^3 \text{ for } i = 1, 2, \dots, m, \tag{16}$$

where ω_1 , ω_2 , and ω_3 are three weight parameters and Y_i^1 , Y_i^2 and Y_i^3 are the features obtained by

$$Y_i^j = W_{opt}^j X_i \text{ for } i = 1, 2, \dots, m, \quad j = 1, 2, 3, \tag{17}$$

where W_{opt}^j for $j = 1, 2, 3$ is the optimal linear transformation matrixes of PCA, LDA and LPP, respectively.

Because with different ω_1 , ω_2 , and ω_3 values, (16) provides a rich set of alternatives to PCA, LDA and LPP, we have to find the best parameter setting. For these reason, we employ the GA to select weights parameters. GA can be considered as a good solution for finding a richer set of alternatives beyond PCA, LDA and LPP in a parametric space, because they are complicated combinational problem. Further, we select locality parameter ε using GA to find local features efficiently. The chromosome used for optimization is shown as Figure 1.

4. Experimental Results

For experiments, four databases of the UCI repository [12] are used to test the proposed method. The databases used in the experiments are summarized in Table 1.

Figure 1. Structure of the chromosome.

Table 1. Databases used for experiments

Database	Number of instances	Number of classes	Number of features
Sonar	208	3	60
Spambase	4601	2	57
SPECTF	267	2	44
Heart	270	2	13

Table 2. GA parameters

Parameter	Value
Crossover rate	0.6
Mutation rate	0.05
Population size	100
Generation	200

Table 3. Average, the best, and the worst accuracy using PCA

Database	Average	Best	Worst
Sonar	82.35 (0.053)	86.27	74.51
Spambase	81.84 (0.004)	82.35	81.39
SPECTF	69.70 (0.081)	74.24	57.58
Heart	62.31 (0.022)	64.17	59.70

In this experiment, we use the genetic parameters given in Table 2. We divide each database into the four subsets. We use three subsets as training and the remaining one subset as test set. We made independent four runs to show the general performance of the algorithms. The training set is used to train classifier and find optimal parameter and the test set is used to evaluate the performance of the proposed method. The fitness function of a chromosome is computed by training accuracy using leave-one-out cross-validation of the training set with nearest neighbor (NN) classifier [13]. We denote the results of previous three methods using only PCA, LDA and LPP in Tables 3, 4, and 5, respectively. We also denote the accuracy of the proposed method in Table 6. In Tables 3-6, the numbers in the parenthesis denote the standard deviation of four runs.

Further, the result of the proposed method is compared with

Table 4. Average, the best, and the worst accuracy using LDA

Database	Average	Best	Worst
Sonar	74.02 (0.079)	80.40	62.75
Spambase	78.37 (0.022)	81.65	77.13
SPECTF	73.11 (0.054)	75.76	65.15
Heart	53.36 (0.056)	59.70	46.27

Table 5. Average, the best, and the worst accuracy using LPP

Database	Average	Best	Worst
Sonar	84.80 (0.040)	90.20	80.39
Spambase	82.04 (0.002)	82.26	81.82
SPECTF	73.86 (0.050)	80.30	68.19
Heart	55.97 (0.019)	58.21	53.73

Table 6. Average, the best, and the worst accuracy of the proposed method

Database	Average	Best	Worst
Sonar	86.27 (0.679)	92.16	76.47
Spambase	83.13 (0.009)	83.57	82.26
SPECTF	75.37 (0.084)	83.33	63.64
Heart	63.43 (0.026)	65.67	59.70

Figure 2. Correct classification ratio for four databases of the UCI repository.

those of the conventional feature selection methods in Figure 2. It can be seen that, in all four databases, the proposed methods show better recognition accuracy than the previous methods. The reason for the better performance of the proposed method might be that proposed method considered global and local structures, simultaneously while conventional methods considered only one aspect structure. Further, optimal parameters can be obtained using GA to improve classification ability.

5. Conclusion

In this paper, we propose the improved the recognition techniques of real-world objects or concepts using computational methods. In classifier, the selection of an appropriate set of features is of considerable importance. Though PCA, LDA and LPP aim to preserve the global and local structure, respectively, it is more important that local and global structures are considered, simultaneously. Therefore, we have proposed a new feature extraction method which can preserve both extrinsic and intrinsic geometry of structures. To find optimal parameters, GA is used. For these reason, we can obtain feature extraction scheme that can have the more classification ability and use features which have global and local properties. The proposed method was applied to four databases selected in UCI machine learning repository and its effectiveness was demonstrated.

Conflict of Interest

No potential conflict of Interest relevant to this article was reported.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP; Ministry of Science, ICT & Future Planning) (No. NRF-2017R1C1B5018408).

References

- [1] Y. Liu and F. Dellaert, "A classification based similarity metric for 3D image retrieval," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA, 1998, pp. 800-805. <https://doi.org/10.1109/CVPR.1998.698695>
- [2] H. Lee, S. Hong, I. F. Nizami, and E. Kim, "An efficient design of nearest neighbor classifier for various-scale problems," *Pattern Recognition Letters*, vol. 31, no. 9, pp. 1020-1027, 2010. <https://doi.org/10.1016/j.patrec.2010.01.001>
- [3] H. Lee, S. Hong, S. An, and E. Kim, "Efficient classification scheme based on hybrid global and local properties of feature," in *Proceedings of International Conference on Control, Automation and Systems*, Seoul, Korea, 2008, pp. 2126-2129. <https://doi.org/10.1109/ICCAS.2008.4694447>
- [4] L. Davis, *Handbook of Genetic Algorithms*. New York, NY: Van Nostrand Reinhold, 1991.
- [5] H. Lee and E. Kim, "Genetic outlier detection for a robust support vector machine," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 15, no. 2, pp. 96-101, 2015. <https://doi.org/10.5391/IJFIS.2015.15.2.96>
- [6] H. Lee, E. Kim, and M. Park, "A genetic feature weighting scheme for pattern recognition," *Integrated Computer-Aided Engineering*, vol. 14, no. 2, pp. 161-171, 2007.
- [7] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997. <https://doi.org/10.1109/34.598228>
- [8] K. Kim and S. Yeom Investigation on the growth of green bean sprouts with linear discriminant analysis," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 17, no. 4, pp. 315-322, 2017. <https://doi.org/10.5391/IJFIS.2017.17.4.315>
- [9] H. Cho and S. Moon, "Comparison of PCA and LDA based face recognition algorithms under illumination variations," in *Proceedings of ICROS-SICE International Joint Conference*, Fukuoka, Japan, 2009, pp. 4025-4030.
- [10] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328-340, 2005. <https://doi.org/10.1109/TPAMI.2005.55>
- [11] Z. Zheng, F. Tang, W. Tan J. Jia, and J. Yang, "Gabor feature-based face recognition using supervised locality preserving projection," *Signal Processing*, vol. 87, no. 10, pp. 2473-2483, 2007. <https://doi.org/10.1016/j.sigpro.2007.03.006>
- [12] P. M. Murphy and D. W. Aha, "UCI repository for machine learning databases," *Department of Information and Computer Science*, University of California, Irvine, CA, 1994.
- [13] S. Ho, C. Liu, and S. Liu, "Design of an optimal nearest neighbor classifier using an intelligent genetic algorithm," *Pattern Recognition Letters*, vol. 23, no. 13, pp. 1495-1503, 2002. [https://doi.org/10.1016/S0167-8655\(02\)00109-5](https://doi.org/10.1016/S0167-8655(02)00109-5)

Heesung Lee received the B.S., M.S., and Ph.D. degrees in Electrical and Electronic Engineering from Yonsei University, Seoul, Korea, in 2003, 2005, and 2010, respectively. From 2011 to 2014, he was a managing researcher with the S1 Corporation, Seoul, Korea. Since 2015, he

has been with the railroad electrical and electronics engineering at Korea National University of Transportation, Uiwang-si, Gyeonggi-do, Korea, where he is currently an assistant professor. His current research interests include computational intelligence, biometrics, and intelligent railroad system.

E-mail: hslee0717@ut.ac.kr