Zhikai ZHAO, Jiansheng QIAN and Jian CHENG
China University of Mining and Technology

Marginal Discriminant Projection for Coal Mine Safety Data Dimensionality Reduction

***Abstract**. Marginal Fisher Analysis (MFA) is a novel dimensionality reduction algorithm. However, the two nearest neighborhood parameters are difficult to select when constructing graphs. In this paper, we propose a nonparametric method called Marginal Discriminant Projection (MDP) which can solves the problem of parameters selection in MFA. Experiment on several benchmark datasets demonstrated the effectiveness of our proposed method, and appreciate performance was achieved when applying on coal mine safety data dimensionality reduction.*

***Streszczenie.** W artykule zaproponowano nieparametryczna metodę nazwaną MDOP (marginal discriminant projection) która pomaga rozwiązać problem selekcji danych w algorytmie MFA (marginal Fisher analysis). Metodę zastosowano do redukcji danych w systemach bezpieczeństwa klopalni węglowych. (**Metoda Marginal Discriminant Projection w zastosowaniu do redukcji wymiaru danych w systemach bezpieczeństwa kopalni węglowych**)*

Keywords: coal mine safety, marginal fisher analysis, dimensionality reduction, marginal discriminant projection.
Słowa kluczowe:.bezpieczeństwo kopalni węglowych, metoda MFA

Introduction

Safety is a key problem we concern during the production process of coal mine. With the increasing informational degree of coal mining enterprises, we can get a large number of sensor information every minute. Many researches have been carried out to use the information for coal mine safety warning. Most of the researches on coal mine gas safety information process focus on the prediction of gas emission or coal and gas outburst forecast [1,2,3]. Cheng et al. have proved the chaotic character of gas signals, and established a chaotic time series based gas concentration forecasting model [4,5,6]. It is a new way for coal mine safety warning to perform workface pattern classification on multi-sensor time series data. In real applications, the same workface involves multiple sensors so that the learning algorithms have to be carried in a high dimensional space. However, previous studies have demonstrated that learning performance can be improved in lower dimensional subspaces. Therefore, before using classification algorithms on coal mine safety data, it is necessary to carry dimensionality reduction firstly. In the field of machine learning and pattern recognition, researches on dimensionality reduction have attracted much attention during the last few years. The aim of dimensionality reduction is to extract few important features from the high dimensions and keep the intrinsic property of samples in the subspace, so it is also called subspace learning. A linear dimensionality reduction algorithm aims to achieve a linear mapping from high to low dimensions space, and discrimination structures and geometry information are preserved during the mapping. There are already a large number of dimensionality reduction algorithms which have different forms but the same goal of finding a low embedding to make the classification or recognition more efficient.

During the large amount of dimensionality algorithms, the linear ones are popular for their direct mapping which can be used to the unseen samples. The classical linear dimensionality reduction methods include Principle Component Analysis (PCA) [7], Locality Preserving Projections (LPP) [8], Linear Discriminant Analysis (LDA) [7] and Marginal Fisher Analysis (MFA) [9,10,11]. PCA and LDA both consider the mean and covariance of sample data so they are methods based on statistics, however, worse performance will be gotten when the samples are not sufficient. Besides this kind of method, there is another kind of algorithms based on graph, which include LPP and MFA. LPP is a linear unsupervised dimensionality reduction method. It seeks for an embedding transformation such that the neighborhood relation of data pairs can be preserved in the embedding space. MFA constructs an intrinsic graph that characterizes the intraclass compactness and another penalty graph which characterizes the interclass separability. Though MFA is a novel method and could achieve desiring performance some times, the two nearest neighborhood parameters is difficult to select when constructing the graphs. Several researches have considered this problem. Hu [12] proposed subspace learning method called Locality Discriminating Projection (LDP) which provides a new scheme for discriminant analysis by considering both the manifold structure and the prior class information. Zeng [13] presented a Synthesized Discriminant Projection (SDP) which maximizes the distance between marginal points and the distance between different class centers. Kokiopoulou [14] introduced the repulsion graph, and proposed a methodology based on repulsion graphs.

In the paper, we propose a new linear dimensionality reduction method called Marginal Discriminant Projection (MDP). A new definition method for the marginal points is presented. Specifically, for each point, two distances were computed, respectively the distance between it and its nearest point from different class and the distance between it and its farthest point from same class. According to the value of these two distances, the points will be classified to marginal points and non-marginal points. The parameter selection problem of MFA was solved by this kind of definition method. Based on the graph embedding framework, we also design two graphs using the marginal points and the non-marginal points. A new discrimination criterion was designed on the graphs. A directly mapping can be achieved by solving a generalized eigenvalue decomposition problem.

The rest of the paper is organized as follows: We give a brief description of the graph based dimensionality reduction method. In next section we introduced our proposed Marginal Discriminant Projection Dimensionality Reduction method. Then the algorithm is tested in several experiments. Finally, we provide some summary and suggestions for future work.

Graph based Dimensionality Reduction

Most of the existing methods can be unified under a graph embedding framework [9], the proposed algorithm is also based on this framework, so in this section we introduce the graph based dimensionality reduction framework and the process to compute a linear projection.

The basic idea of graph embedding is to consider the sample point in high dimension space as a vertex to construct a graph and the weight of sides between the vertexes represent the similarity of their corresponding samples points. The similarity is measured using a graph similarity matrix which represents the statistics and geometry property of the data set. Graph embedding aims to find a low dimension vector to represent the vertex of graph and keep the similarity in the low dimension space. Linear dimensionality reduction aims to find an explicit mapping from original feature space to the embedding space. In this situation, we can get the embedding of new test point through this mapping. The definition of dimensionality reduction show followed and also the basic process of graph based dimensionality reduction.

For a pattern recognition problem, there is a sample set $X=[x_1, \cdots, x_n]$, $x_i \in R^m$, n is the sample number and m is the feature dimension. In a supervised task, label of sample x_i is $l_i \in \{1, \cdots, C\}$. The aim to application dimensionality reduction is to extract few dimension features and preserve most intrinsic information of samples. This will bring convenient for the follows work such as recognition, and avoid the problem of curse of dimensionality. Linear algorithm aims to get a mapping through which all samples can get a embedding. Suppose the mapping is $F{:}x \to y$, where $x \in R^m$ is the sample in the original space, and is the corresponding embedding.

In the view of graph embedding, the samples X form an undirected weighted graph $G=\{X, \mathbf{W}\}$. $\mathbf{W} \in R^{N\times N}$ is the weight matrix of the graph, in which the elements represent the similarity of two vertexes; the mostly using measure method is the Euclidean distance between the vertexes. For the graph G, its diagonal matrix $\mathbf{D}$ and Laplacian matrix $\mathbf{L}$ are defined as:

(1) $$\mathbf{L} = \mathbf{D} - \mathbf{W}, \mathbf{D}_{ii} = \sum_{j \neq i} \mathbf{W}_{ij}, \forall i.$$

The nature of graph embedding is to find the low embedding of all vertexes in G, and at the same time maintain maximum similarity of point to point in G. The points which are more similarity in the original feature space should keep the same in the low embedding space. Suppose y_i is the corresponding low embedding of sample x_i, during the embedding process the rule should be followed:

(2) $$y^{(opt)} = \arg\min \sum_{i \neq j} \| y_i - y_j \|^2 \mathbf{W}_{ij} = \arg\min y^{\mathrm{T}} \mathbf{L} y$$

The optimal embedding of the algorithm is to solving the minimize problem above. To avoid the trivial solution which all points are embedding to the same point, another graph $G^p=\{X, \mathbf{W}^p\}$ usually is constructed, where $\mathbf{W}^p$ describe the similarity should be avoid during the embedding. That is the lower similarity between two points in original space should maintain low in the embedding space, which can be formulated as:

(3) $$y^{p(opt)} = \arg\max \sum_{i \neq j} \| y_i - y_j \| \mathbf{W}_{ij}^p = \arg\max y^{\mathrm{T}} \mathbf{L}^p y$$

Similar to mentioned earlier, the diagonal matrix $\mathbf{D}^p$ and Laplacian matrix $\mathbf{L}^p$ of graph G^p are defined as:

(4) $$\mathbf{L}^p = \mathbf{D}^p - \mathbf{W}^p, \mathbf{D}_{ii}^p = \sum_{j \neq i} \mathbf{W}_{ij}^p, \forall i.$$

As a conclusion of the above optimal problem, we should optimal the following formulation:

(5) $$y^* = \arg\min \frac{y^{\mathrm{T}} \mathbf{L} y}{y^{\mathrm{T}} \mathbf{L}^p y}$$

The above dimensionality reduction method can only get the embedding of the trained samples, and can not be used on new coming points. So it is necessary to get an explicit mapping from the high to low dimensions space which also can be used for the "out of sample" points. Suppose there is a linear mapping ω from the high feature space to the low embedding space, for any sample point x_i in the high feature space, we can get its corresponding low embedding through:

(6) $$y_i = x_i^{\mathrm{T}} \omega$$

Using this formulation, the problem above can be reformed:

(7) $$y^* = \arg\min \frac{\omega^{\mathrm{T}} \mathbf{X} \mathbf{L} \mathbf{X}^{\mathrm{T}} \omega}{\omega^{\mathrm{T}} \mathbf{X} \mathbf{L}^p \mathbf{X}^{\mathrm{T}} \omega}$$

Finally, we can get an explicit linear mapping through solving above formulation.

Marginal Discriminant Projection

In this section, we introduced a new way to define marginal points, and then constructed an intrinsic graph on each non-marginal point and its farthest point from same class and another penalty graph on each marginal pint and its nearest point from different class.

Fig. 1 shows the basic idea to define marginal point. For each point x_1 from class 1, find the farthest point x_2 from class 1 and the nearest data x_3 from class 2. We use S_m and S_n as the set of marginal points and non-marginal points respectively.

We define point x_1 as:

(8) $$x_1 \in \begin{cases} S_m, \text{ if } \|x_1 - x_3\| \leq \|x_1 - x_2\| \\ S_n, \text{ otherwise.} \end{cases}$$

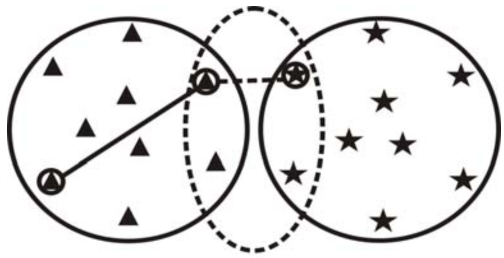

Fig.1. The basic idea to define marginal points in MDP

From the point of graph embedding, we construct a graph to represent the within class compactness called within class graph G. Link the non-marginal point with its farthest point from same class and set the weight of the side to 1, the others to 0, and then we can get a weight matrix:

(9) $$\mathbf{W}_{ij} = \begin{cases} 1, \text{ if } x_i \in S_n \text{ and } x_j = S_n(x_i) \\ 0, \text{otherwise.} \end{cases}$$

where $x_j=S_n(x_i)$ represents that x_j is the farthest point of x_i within same class.

The locality neighborhood relation should be kept during the projection. The projection can be got through optimize the follow formulation:

$$\arg\min \sum_{i=1}^{N}\sum_{j=1}^{N} \| y_i - y_j \|^2 \ \mathbf{W}_{ij}$$

(10) $$= \arg\min \sum_{i=1}^{N}\sum_{j=1}^{N} \| \omega^{\mathrm{T}} x_i - \omega^{\mathrm{T}} x_j \|^2 \ \mathbf{W}_{ij}$$

$$= \arg\min 2\omega^{\mathrm{T}} \mathbf{X}(\mathbf{D}-\mathbf{W})\mathbf{X}^{\mathrm{T}} \omega$$

$$= \arg\min 2\omega^{\mathrm{T}} \mathbf{X}\mathbf{L}\mathbf{X}^{\mathrm{T}} \omega$$

$\mathbf{D}$ is a $n \times n$ diagonal matrix, and

(11) $$\mathbf{D}_{ii} = \sum_{j=1}^{n} \mathbf{W}_{ij}$$

For convenient, we set $\mathbf{L}$:

(12) $$\mathbf{L} = \mathbf{D} - \mathbf{W}$$

The aim of dimensionality reduction is to get better performance during the following task, such as classification or recognition. So when finding the embedding of the sample points, the points from different classes should be separated. As in MFA, we also construct a penalty graph G^p with the marginal points defined by our proposed method. The weight matrix of the graph is:

(13) $$\mathbf{W}_{ij}^{p} = \begin{cases} 1, \text{if } \ x_i \in S_m \text{ and } x_j = S_m(x_i) \\ 0, \text{otherwise.} \end{cases}$$

where $x_j = S_n(x_i)$ represents that x_j is the nearest point of x_i from different class.

During the embedding, we should enhance the separability of the marginal point from different classes, and the projection can be got through optimize the following formulation:

$$\arg\min \sum_{i=1}^{N}\sum_{j=1}^{N} \| y_i - y_j \|^2 \ \mathbf{W}_{ij}^{p}$$

(14) $$= \arg\min \sum_{i=1}^{N}\sum_{j=1}^{N} \| \omega^{\mathrm{T}} x_i - \omega^{\mathrm{T}} x_j \|^2 \ \mathbf{W}_{ij}^{p}$$

$$= \arg\min 2\omega^{\mathrm{T}} \mathbf{X}(\mathbf{D}^{p}-\mathbf{W}^{p})\mathbf{X}^{\mathrm{T}} \omega$$

$$= \arg\min 2\omega^{\mathrm{T}} \mathbf{X}\mathbf{L}^{p}\mathbf{X}^{\mathrm{T}} \omega$$

G^p is a $n \times n$ diagonal matrix, and

(15) $$\mathbf{D}_{ii}^{p} = \sum_{j=1}^{n} \mathbf{W}_{ij}^{p}$$

For convenient, we set $\mathbf{L}^p$:

(16) $$\mathbf{L}^{p} = \mathbf{D}^{p} - \mathbf{W}^{p}$$

We combine the two criterions and the final optimal projection should be the following optimization problem:

(17) $$\omega^{*} = \arg\max_{\omega} \frac{\omega^{\mathrm{T}} \mathbf{X}\mathbf{L}_{b}\mathbf{X}^{\mathrm{T}} \omega}{\omega^{\mathrm{T}} \mathbf{X}\mathbf{L}_{w}\mathbf{X}^{\mathrm{T}} \omega}$$

This optimal solution is got by solving the generalized eigenvalue decomposition problem:

(18) $$\lambda \mathbf{X}\mathbf{L}_{w}\mathbf{X}^{\mathrm{T}} \omega = \mathbf{X}\mathbf{L}_{b}\mathbf{X}^{\mathrm{T}} \omega$$

A solution is d eigenvector correspond to the first d largest eigenvalues.

The process of the algorithm is:

1. Marginal point partition. Divide the samples to marginal point set and non-marginal point set according the definition of marginal point.

2. Construct the intrinsic graph. In the intrinsic graph, put a side between a non-marginal point and its farthest neighbor within the same class, and set the weight follow equation (9).

3. Construct the penalty graph. In the penalty graph, for each marginal point, put a side between it and its nearest neighbor from different classes, and set the weight follow equation (13).

4. Find the marginal discriminant project follow the optimize problem:

(19) $$\omega^{*} = \arg\max_{\omega} \frac{\omega^{\mathrm{T}} \mathbf{X}\mathbf{L}_{b}\mathbf{X}^{\mathrm{T}} \omega}{\omega^{\mathrm{T}} \mathbf{X}\mathbf{L}_{w}\mathbf{X}^{\mathrm{T}} \omega}$$

5. Final projection:

(20) $$\omega = \omega^{*}$$

Experimental Analysis

To evaluate the proposed Marginal Discriminant Projection algorithm, we take experiments on four real world data sets and compare it to three dimensionality reduction algorithms. Firstly, we take a visualization experiment on the Coil20 data set and compare the embedding performance under each algorithm. Then the recognize accuracy was showed using the proposed algorithm and the other three algorithms.

We compare the two dimensions embedding using our proposed algorithm and the other three algorithms on the Coil20 data set. Coil20 data set is gray pictures of 20 objects, in this visualization experiments we choose 4 objects to analysis the embedding performance using our proposed method and other three methods.

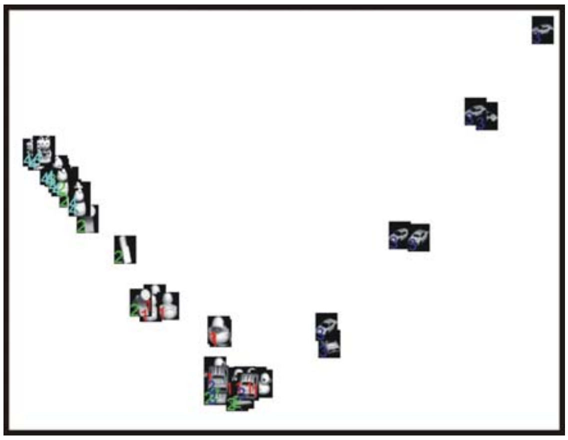

Fig.2. The embedding performance under LPP

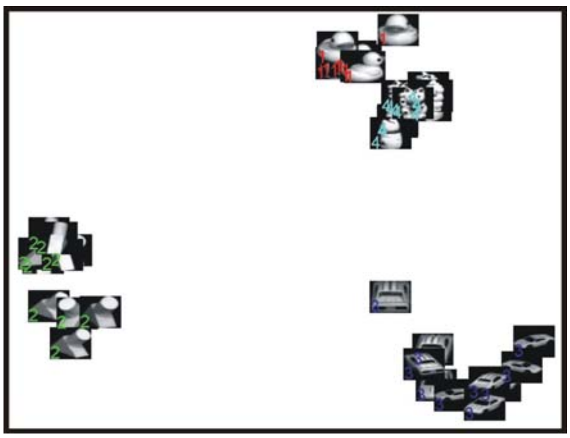

Fig.3. The embedding performance under LDA

Fig.4. The embedding performance under MFA

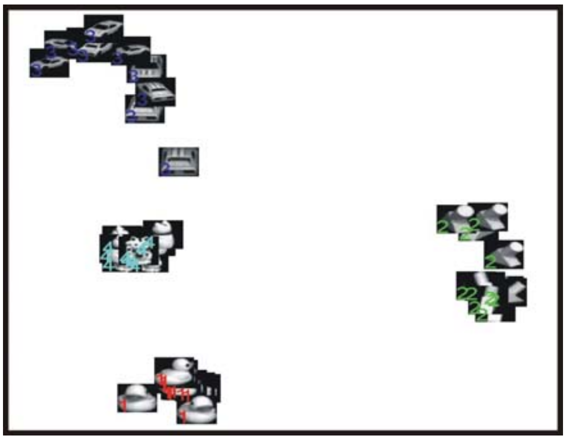

Fig.5. The embedding performance under MDP

Fig. 2 to Fig. 5 show the embedding performance of four algorithms (LPP, LDA, MFA and MDP) respectively. It is clear that the supervised algorithms (LDA, MFA and MDP) have better performance than the unsupervised algorithm (LPP), and at the same time our proposed algorithm MDP can separate the different class better than the other three algorithms.

Experiments are taken on four benchmark datasets, ORL, Yale, MU PIE and Coil20 to evaluate the separability of the lower dimensional representation derived from MDP in comparison to LPP, LDA and MFA. The ORL database contains 400 images of 40 individuals. The images were captured at different times and with different variations including expression and facial details, the size of each image is 92×112, in our experiment, we resize each image to 32×32 with the eyes in the same place. The Yale database contains 165 images of 15 individuals. We also resize each image to 32×32 with the eyes in the same place. The CMU PIE database contains 41,368 facial images of 68 people. The images were acquired across different poses, under variable illumination conditions, and with different facial expressions. In our experiment, a sub database is chosen for evaluation of our proposed algorithm. The sub database referred to as PIE-10 includes five near frontal poses (C05, C07, C09, C27 and C29) under all illumination and each person has 170 images with size of 32×32.

The purpose of our experiments is to compare the recognition accuracy under different algorithms. The classification algorithm is 1NN. Each dataset is splits to train set and test set randomly 50 times. In ORL database 5 images each person are chosen to compose the train set, the rest 5 images are test ones. It is respectively 6 and 10 images each person in the Yale and CMU PIE database.

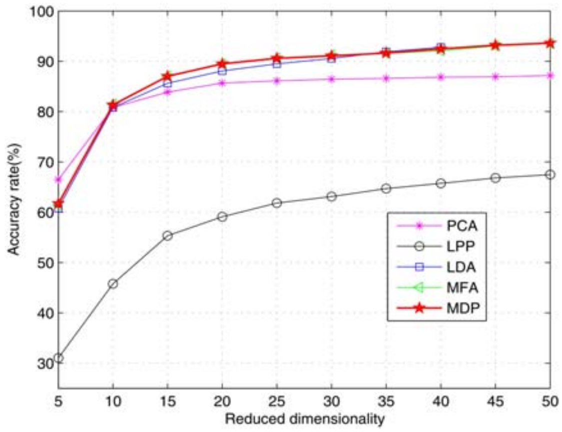

Fig.6. The recognition rate on ORL using different algorithms

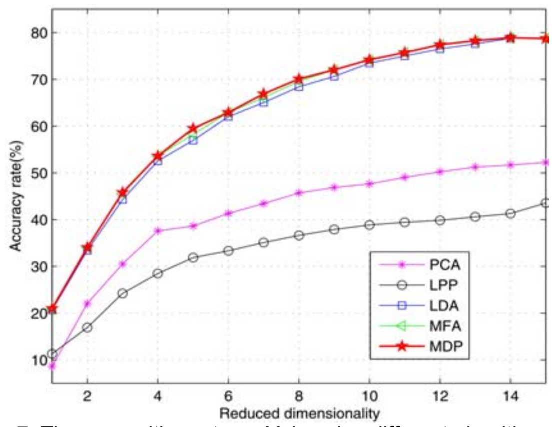

Fig.7. The recognition rate on Yale using different algorithms

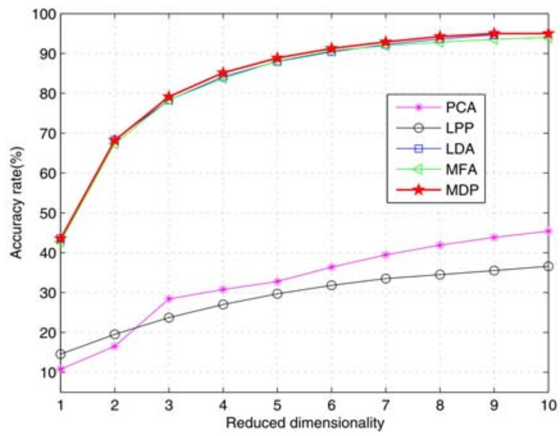

Fig.8. The recognition rate on PIE using different algorithms

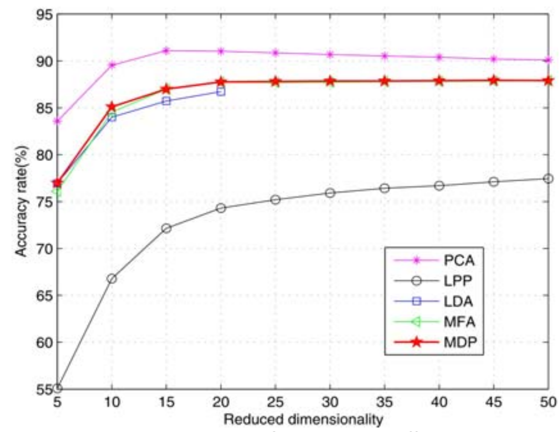

Fig.9. The recognition rate on Coil20 using different algorithms

From Fig. 6 to Fig. 9 we can conclude that compared with other algorithms, MDP could get better performance in most situations except on the Coil20 dataset.

In this subsection experiments ware taken on coal mine gas time series to compare our proposed method MDP with three other algorithms. The data set was sampled from the KJ2006 safety monitoring system of Luling Coal Mine in Anhui Province. 16 sensors were chosen from 882 workface, including □882-1#yan<T1>, □882-1# yan <T2>, □882-1# yan <T3>, □882-1#shimen<T1>, □882-1#shimen<T2>, □882-5#yan Concentration, □882-5#yan Flow, □882-5#yan Temperature, □882-4#yan Pressure, □882-4# yan Flow, □882-4# yan Concentration, □882-3# yan Temperature, □882-2# yan Flow, □882-2# yan Temperature, □882-2# yan Pressure and □882-2# yan Concentration. The data were samples from January 10, 2007 to January 29, 2007. We labelled the samples according to the values of 5 sensors, which were □882-1#yan<T1>, □882-1# yan <T2>, □882-1# yan <T3>, □882-1#shimen<T1> and □882-1#shimen<T2>. Two classes were generated to express the safety situation of the workface. If the values of 5 sensors were bigger than 0.4 at the same time, the situation is labelled to 1, otherwise labelled to -1. We choose 600 samples in our experiment, which include 300 positive samples and 300 negative ones.

The aim of this experiment is to test the classification accuracy using the low dimensional data which is achieved by four different dimensionality reduction methods. The classification algorithm is KNN, and the parameter choice was using grid search. We spit the coal mine sensor data set randomly to two parts, which are the train set and the test set. Training set includes 60 samples and test set includes 260 samples.

Table 1 shows the classification accuracy using KNN on the low dimensionality data under four different algorithms. The result shows dimensions of 1 to 10 of each dimensionality reduction algorithm. It is clear that the proposed algorithm MDP achieved better result than the other three methods.

Table 1. The classification accuracy by KNN using different dimensionality reduction algorithm (%)

	1	2	3	4	5	6	7	8	9	10
PCA	91.15	96.54	95.77	96.92	96.54	94.23	94.23	94.23	94.23	94.23
LDA	38.85	38.85	38.85	38.85	38.85	38.85	37.69	65.00	20.00	23.08
MFA	91.54	92.31	93.08	93.08	93.08	93.46	93.46	93.46	93.46	93.46
MDP	91.15	99.23	98.46	99.23	99.23	94.23	94.23	94.23	94.23	94.23

Conclusions

A new dimensionality reduction algorithm called Marginal Discriminant Projection is proposed. Based on the graph embedding framework, we present a new method to define the marginal points, and then the marginal points are using to increase the separability of samples from different classes, at the same time keep the intraclass compactness of samples from same class. Though we avoid the parameter selection problem, the new definition method of marginal point could not always get the most optimal performance. So constructing a more optimal nonparametric method is what we should research in the future.

This work was supported by National Natural Science Foundation of China under Grant 60835002, the Fundamental Research Funds for the Central Universities under Grant 2011QNB24.

REFERENCES

[1] Y. Zeng, C. Wu. Research on fuzzy fractal neural network for prediction of mine gas emission. Coal Science and Technology. vol.32, no.2. pp.62-65,2004.
[2] C. Cai, G. Jing. Fuzzy Synthetic Evaluation of Prediction of Coal and Gas Outburst Scale. Journal of Safety and Environment. Vol.4, no.2, pp.54-56,2004.
[3] N. Li, D. Guo, M. Fan. Ash related analysis method applied to analysis of coal and gas outburst controlled factors. Coal Science and Technology. vol.32, no.2. pp.67-69,2004.
[4] J. Cheng, J. Bai, J. Qian and S. Li. Short- Term Forecasting Method of Coalmine Gas Concertration Based on Chaotic Time Series. Journal of China University of Mining & Technology. vol.37, no.2. pp.231-235, 2008.
[5] J. Zhang, J. Cheng, J. Qian, Forecasting Coalmine Gas Concentration Based on Adaptive Neuro-Fuzzy Inference System. Journal of China University of Mining & Technology. vol.36, no.4. pp.494-498, 2007.
[6] Y. Hou, J. Cheng, S. Li. Forecasting coalmine gas concentration based on RBF neural network. Proceedings of the International Conference on Information Acquisition. pp.192-194, 2007.
[7] A. M. Artinez and A. C. Kak, "PCA Versus LDA". IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp.228-233, 2001.
[8] X. He and P. Niyogi. "Locality preserving projection". In Advances in Neural Information Processing System 16. MIT Press, Massachusetts, 2003.
[9] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang and S. Lin. "Graph Embedding and Extensions: A General Framework for Dimensionality Reduction. Pattern Analysis and Machine Intelligence, IEEE Transactions on", vol. 29, no. 1, pp. 40-51, 2007.
[10] J. Xu, J. Yang. "Nonparametric Marginal Fisher Analysis for Feature Extraction". Lecture Notes in Computer Science, Vol. 6216, pp. 221-228, 2010.
[11] D. Xu , S. Yan , D. Tao, S. Lin and H. Zhang. "Marginal Fisher Analysis and Its Variants for Human Gait Recognition and Content- Based Image Retrieval". IEEE transactions on image processing, vol. 16, no. 11, pp.2811-2821, 2007.
[12] J. Hu, W. Deng, J. Guo and W. Xu. "Learning a locality discriminating projection for classification. Knowledge-Based System, vol.22, no.8, pp.562-568, 2009.
[13] Z. Zeng, C.Liu, L. Huang. "Analysis on two Fisher Methods and a Synthesized Discriminant Projection". Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. No. (8-11), pp.1-4, 2008.
[14] E. Kokiopoulou, Y. Saad. "Enhanced graph-based dimensionality reduction with repulsion Laplaceans". Pattern Recognition. vol. 42, no. 11, pp. 2392-2402, 2009.

***Authors**: Zhikai ZHAO. School of Computer Science and Technology, IoT Perception Mine Research Center, China University of Mining and Technology, E-mail: zhikaizh@gmail.com; Jiansheng QIAN. School of Information and Electrical Engineering, China University of Mining and Technology, E-mail: qianjsh@cumt.edu.cn; Jian CHENG. School of Information and Electrical Engineering, IoT Perception Mine Research Center, China University of Mining and Technology, E-mail: fantastcj@126.com.*

The correspondence address is:

e-mail:
zhikaizh@gmail.com