

Research Article

Kernel Fisher Discriminant Analysis Based on a Regularized Method for Multiclassification and Application in Lithological Identification

Dejiang Luo¹ and Aijiang Liu²

¹ College of Management Science, Chengdu University of Technology, Chengdu 610059, China ² College of Geophysics, Chengdu University of Technology, Chengdu 610059, China

Correspondence should be addressed to Aijiang Liu; 75747141@qq.com

Received 19 July 2014; Revised 8 October 2014; Accepted 10 October 2014

Academic Editor: Jun Cheng

Copyright © 2015 D. Luo and A. Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study aimed to construct a kernel Fisher discriminant analysis (KFDA) method from well logs for lithology identification purposes. KFDA, via the use of a kernel trick, greatly improves the multiclassification accuracy compared with Fisher discriminant analysis (FDA). The optimal kernel Fisher projection of KFDA can be expressed as a generalized characteristic equation. However, it is difficult to solve the characteristic equation; therefore, a regularized method is used for it. In the absence of a method to determine the value of the regularized parameter, it is often determined based on expert human experience or is specified by tests. In this paper, it is proposed to use an improved KFDA (IKFDA) to obtain the optimal regularized parameter by means of a numerical method. The approach exploits the optimal regularized parameter selection ability of KFDA to obtain improved classification results. The method is simple and not computationally complex. The IKFDA was applied to the *Iris* data sets for training and testing purposes and subsequently to lithology data sets. The experimental results illustrated that it is possible to successfully separate data that is nonlinearly separable, thereby confirming that the method is effective.

1. Introduction

China's tight clastic rock reservoir is considerably wide, containing sediments that were deposited during the Carboniferous, Permian, Triassic, and Jurassic periods. The reservoirs in Western Sichuan and Erdos are more representative. The West Sichuan depression is located in the Western Sichuan Basin, which belongs to the western depression belt that is located in Yangtze Platform or Longmenshan Fault Zone. A tight gas reservoir was discovered in the Xujiahe and Shaximiao formations that occur in this area. Tight clastic reservoirs are noted for their low porosity, because of their dense, multilayered stacking and strong heterogeneous characteristics that are caused by their complexity and particularity. These characteristics complicate the identification of the lithology, which would enable the prediction of the properties of a reservoir. Previous research [1, 2] identified the reservoir lithology consisting of mudstone, sandstone, and siltstone in the AC area of western Sichuan. The cross plot

of sandstone and siltstone shows that these rock types overlap and are mixed together, which is a linearly nonseparable case. The cross plot and mathematical models have been applied extensively in lithology identification in previous studies. For example, Hsieh et al. constructed a fuzzy lithology system from well logs to identify the formation lithology [3], while Shao et al. applied an improved BP neural network algorithm, based on a momentum factor, to lithology recognition [4]. Zhang et al. used Fisher discrimination to identify volcanic lithology using regular logging data [5]. However, it is very difficult to identify the lithology of tight clastic rock reservoirs with the above method. Thus, this paper proposes the Kernel Fisher discriminant analysis (KFDA) for tight clastic rock lithology identification.

The KFDA has its roots in Fisher discriminant analysis (FDA) and is the nonlinear scheme for two-class and multiclass problems [6]. KFDA functions by mapping the lowdimensional sample space into a high-dimensional feature space, in which the FDA is subsequently conducted. The KFDA study focuses on applied and theoretical research. Billings et al. replaced the kernel matrix with its submatrix in order to simplify the computation [7, 8]; Liu et al. proposed a new criterion for KFDA to maximize the uniformity of class-pair separabilities that was evaluated by the entropy of the normalized class-pair separabilities [9]; Wang et al. considered discriminant vectors to be linear combinations of "nodes" that are part of the training samples and therefore proposed a fast kernel Fisher discriminant analysis technique [10, 11]. Wang et al. proposed the nodes to be the most representative training samples [12]. Optimal kernel selection is one of the areas of theoretical research that has been attracting considerable attention. Fung et al. have developed an iterative method based on a quadratic programming formulation of FDA [13]. Khemchandani et al. considered the problem of finding the data-dependent "optimal" kernel function via second-order cone programming [14]. The use of KFDA in combination with strong nonlinear feature extraction ability is becoming a powerful tool for solving identification or classification problems. Hence, it has been applied widely and successfully in many areas. Examples of the application of KFDA are face recognition, fault diagnosis, classification, and the prediction of the existence of hydrocarbon reservoirs [15-20].

The principle that underlies KFDA is that input data are mapped into a high-dimensional feature space by using a nonlinear function, after which FDA is used for recognition or classification in feature space. KFDA requires factorization of the Gram matrix into the kernel within-class scatter matrix K_w and the between-class scatter matrix K_b . KFDA can finally be attributed to the solution of a generalized eigenvalue problem $K_b \alpha = \lambda K_w \alpha$. As the matrix K_w is often singular, a regularized method is often used to solve the problem, which is transformed into a general eigenvalue problem by choosing a smaller positive number μ , in which case K_w is replaced with $K_w + \mu I$. Previous studies have shown the classification ability of KFDA to depend on the value of μ ; therefore, the appropriate values for KFDA are very important. In many practical applications, the parameter μ is specified according to experience or experimental results.

This paper proposes a new approach for the selection of the regularized parameter μ to gain the best classification results, and KFDA is improved for both *Iris* data sets and lithology data sets. The paper is organized as follows. Section 2 summarizes kernel Fisher discriminant analysis. In Section 3, a numerical method for finding an optimal parameter μ is proposed by introducing a regularized method for KFDA. The experimental results are given in Sections 4 and 5, while Section 6 presents the concluding remarks.

2. Kernel Fisher Discriminant Analysis

Let $S = \{x_1, x_2, ..., x_N\}$ be the data set that contains K classes in the d-dimensional real space \mathbb{R}^d . Let Ni samples belong to the *j*th class, $(N = N_1 + N_2 + \cdots + N_k)$. FDA is used for lithology identification by searching the optimal projection vectors w, and then a different class of lithology samples has minimum within-class scatter. FDA is given by the vector $w \in \mathbb{R}^d$ that maximizes the Fisher discriminant function as

$$J(W) = \frac{w^T S_b w}{w^T S_w w},\tag{1}$$

where S_w is the within-class scatter matrix and S_b is the between-class scatter matrix. FDA is essentially a linear method, which makes it very difficult to separate the nonlinear separable sample.

KFDA significantly improves the classification ability for the nonlinear separable sample of FDA via the use of a kernel trick. To adapt to nonlinear cases, $\varphi(x)$ is mapped from the lower dimensional sample space into a high-dimensional feature space. Note that $\phi(x_i^j)$ ($i = 1, 2, ..., N_j$, j = 1, 2, ..., K) represents the *i*th projection value in the class ω_j . Let m^{ϕ} be the mean vector of the population, and let m_j^{ϕ} be the mean vector of class ω_j . In the feature space *F*, the total scatter matrix S_t , the within-class scatter matrix S_w , and the between-class scatter matrix S_b can be defined as

$$S_{t}^{\phi} = \frac{1}{N} \sum_{i=1}^{N} \left(\phi(x_{i}) - m^{\phi} \right) \left(\phi(x_{i}) - m^{\phi} \right)^{T},$$

$$S_{w}^{\phi} = \frac{1}{N} \sum_{j=1}^{K} \sum_{i=1}^{N_{j}} \left(\phi(x_{i}^{j}) - m_{j}^{\phi} \right) \left(\phi(x_{i}^{j}) - m_{j}^{\phi} \right)^{T}, \quad (2)$$

$$S_{b}^{\phi} = \sum_{j=1}^{K} \frac{N_{j}}{N} \left(m_{j}^{\phi} - m^{\phi} \right) \left(m_{j}^{\phi} - m^{\phi} \right)^{T}.$$

Lithology identification by KFDA can be attributed to the optimization of kernel Fisher criterion function as follows:

$$J(v) = \frac{v^T S_b^{\phi} v}{v^T S_w^{\phi} v},\tag{3}$$

where v represents the different optimal projection vector. The high dimension of feature space *F* and the infinite dimension make it impossible to directly calculate the optimal discriminant vector *v*. A solution for this problem is to use the kernel trick as follows:

$$K\left(x_{i}, x_{j}\right) = \left(\phi\left(x_{i}\right), \phi\left(x_{j}\right)\right). \tag{4}$$

According to the theory of reproducing a kernel [6], any solution v must lie in the feature space F, which spans $\phi(x_1), \phi(x), \dots, \phi(x_N)$ as follows:

$$\nu = \sum_{i=1}^{N} \alpha_i \phi(x_i).$$
 (5)



FIGURE 1: Scatter gram plots of Iris data: red (C1, setosa); green (C2, virginica); blue (C3, versicolor).

In F, any test samples can be projected into w to give the following equation:

$$v^{T}\phi(x) = \sum_{i=1}^{N} \alpha_{i} (\phi(x), \phi(x_{i}))$$

= $\alpha^{T} ((\phi(x)\phi(x_{1})), (\phi(x), \phi(x_{2})), ..., (6))$
 $(\phi(x), \phi(x_{N})))$
= $\alpha^{T} (K(x, x_{1}), K(x, x_{2}), ..., K(x, x_{N})).$

In F, the kernel within-class scatter matrix K_w and the between-class scatter matrix K_b can be defined as

$$\mu_{0} = \left(\frac{1}{N}\sum_{i=1}^{N} K(x_{1}, x_{i}), \frac{1}{N}\sum_{i=1}^{N} K(x_{2}, x_{i}), \dots, \frac{1}{N}\sum_{i=1}^{N} K(x_{N}, x_{i})\right)^{T},$$
(7)

$$\mu_{i} = \left(\frac{1}{N_{i}}\sum_{j=1}^{N_{i}} K\left(x_{1}, x_{j}^{i}\right)\right), \dots, \frac{1}{N_{i}}\sum_{j=1}^{N_{i}} K\left(x_{N}, x_{j}^{i}\right), \quad (8)$$

$$K_{b} = \sum_{i=1}^{K} \frac{N_{i}}{N} \left(\mu_{i} - \mu_{0}\right) \left(\mu_{i} - \mu_{0}\right)^{T}, \qquad (9)$$

$$K_{w} = \frac{1}{N} \sum_{i=1}^{K} \sum_{j=1}^{N_{i}} \left(\xi_{x_{j}} - \mu_{i} \right) \left(\xi_{x_{j}} - \mu_{i} \right)^{T}, \quad (10)$$

$$J(\nu) = \frac{\nu^T S_b^{\phi} \nu}{\nu^T S_w^{\phi} \nu} = \frac{\alpha^T K_b \alpha}{\alpha^T K_w \alpha} = J(\alpha).$$
(11)

According to the properties of the generalized Rayleigh quotient, the optimal solution vector w is obtained by maximizing the criterion function in (11) by setting it equivalent

to the solution of the generalized characteristic equation as follows:

$$K_b \alpha = \lambda K_w \alpha. \tag{12}$$

3. Choosing the Regularized Parameter

If K_w is a nonsingular matrix, then optimal vectors α , obtained by maximizing (11), are equivalent to the feature vectors corresponding to the top *m* largest eigenvalues [12, 21]. Equation (12) can be described as

$$\left(K_{w}\right)^{-1}K_{b}\alpha = \lambda\alpha. \tag{13}$$

The solution of practical problems requires the use of n training samples to estimate the variance of n-dimensional structure; therefore, K_w is a singular matrix. This means that it is often not possible to use (13). However, it is possible to promote the stability of the numerical method by using regularized method as follows:

$$(K_w)_\mu = K_w + \mu I, \tag{14}$$

where μ is a small, positive number, and *I* is the identity matrix. Then, (12) can be expressed as

$$\left(K_{w}+\mu I\right)^{-1}K_{b}\alpha=\lambda\alpha.$$
(15)

When KFDA is used to solve problems of an applied nature, parameter μ is determined according to the experience or the result of the experiment. This paper uses a numerical analysis method to solve the parameter; hence, the determinant of the value of $|K_w + \mu I|$ can be regarded as a function of μ :

$$f(\mu) = |K_w + \mu I|. \tag{16}$$

When function f is stable and the value of the function tends to zero (17), the parameter μ is the best classification parameters

$$\lim_{\mu \to 0} f(\mu) \longrightarrow 0.$$
(17)

4. Experiments

4.1. Experiments Settings. The Iris data set is often used to test the discriminant analysis algorithm [22–24]. This data set is divided into three classes, which represent three different varieties of the Iris flower: C1, Iris setosa, C2, Iris versicolor, and C3, Iris virginica, which takes petal length, petal width, sepal length, and sepal width as four-dimensional variables. The three classes represent three different varieties of Iris flowers. There are 150 samples in this data set, and there are 50 samples in each class. The results are plotted as scatter grams (Figure 1) and show that classes C1 and C2 and classes C1 and C3 are linearly separable and classes C2 and C3 are nonlinearly separable. The aim was to address this problem by using the KFDA with different values of μ for classification purposes [25].

The following experiments involve a comparative study of the different selection schemes of μ . In this study,



FIGURE 2: Value of the determinant versus the value of regularized μ .

the *Iris* data set is used to conduct algorithm training. Thus, the three sets of samples (based on petal length, petal width, sepal length, and sepal width) were chosen with each of these sets comprising 30 samples. The kernel function employs in KFDA the Gauss kernel function (18), for which the kernel parameter σ is set to be the norm of the covariance matrix of the training samples

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma}\right).$$
 (18)

4.2. Experimental Results and Discussions. The experiment was processed within a MATLAB 7.0 environment running on a PC powered by a Pentium 3.3 GHz CPU. The experimental results are shown in Figures 2 and 3. The optimal value of regularized μ is identified in Figure 1, which shows that the function $f(\mu) = |K_w + \mu I|$ has an obvious inflexion point and that the value of the function $f(\mu)$ approaches zero when the parameter is equal to 0.09.

This is further illustrated in Figure 3, which shows the classification performance when the parameter μ has different values. A simulation was conducted by selecting 30 data points from each of the three classes and the results were constructed in the form of scatter plots that show the distribution of the data by using KFDA. The optimal regularized parameter ($\mu = 0.09$) guaranteed separation of classes C2 and C3. However, when the value of the regularized parameter ($\mu = 0.18$) increased, it impacted negatively the classification effect.

5. Application to Lithology Data Sets

5.1. The Geological Settings of the AC Region. The AC region is located in the central segment of the West Sichuan depression and was formed during Upper Triassic and Jurassic periods. The Triassic-Jurassic stratum is the part of the thick entity of the Western Sichuan foreland basin, with a total thickness



FIGURE 3: Classification results with different value of regularized μ .

of 6500 m. The AC region is located in a large uplift belt of the West Sichuan depression, which shows NEE trend [1, 2]. This paper uses KFDA for logging parameters identification of lithology, and the purpose of the study horizon is Xujiahe formation. The Xujiahe formation in Chengdu, Deyang, and JiangYou is a typical tight rock formation characterized by low porosity, because of its dense, multilayered stacking and strong heterogeneity. The formation has a typical thickness of 400–700 m, and the thickness of the formation in Anxian county in front of Longmenshan thickness is up to 1000 m.

The Xujiahe formation consists of alternating layers of sandstone, siltstone, mudstone, shale, and coal series, in

which the rock types are more complex. According to logging data obtained from wells and the extraction of the physical parameters, the rock in the Xujiahe formation can be divided into mudstone, siltstone, and sandstone based on the physical intersection diagram and histogram analysis of rocks.

5.2. Logging Parameters Lithological Identification. Logging parameters provide a comprehensive reflection of lithology, and their sensitivity is different for lithology identification. The sensitivity of these parameters for lithology identification was studied using the method of correlation analysis and finally determined acoustic (AC), natural gamma ray (GR),



FIGURE 4: Cross-sectional plots of the logging parameters: green O (sandstone); red (siltstone); blue (mudstone).

density (DEN), and compensated neutron logging (CNL) as the characteristic variables. Training and testing sets in the standard layer, each of which contained 50 samples, were obtained. The cross-sectional plot is displayed in Figure 4. A comparison of Figures 1 and 4 reveals the characteristics that are similar to both, namely, in which the sandstone and siltstone sample data cannot be separated.

In the following experiments, FDA and KFDA are compared on logging attribute data sets. The FDA method is used to extract the optimal and suboptimal discriminant vector of the training sets, and then the testing sets of sample are projected onto vectors (Figure 5). As can be seen in Figure 5, the mudstone can be separated from the sandstone and siltstone. The latter two rock types were still mixed together with respect to the siltstone and sandstone data, although the separation degree is higher than indicated by cross plot.

Experiments were conducted on logging attribute sets using IKFDA. The kernel function employed in IKFDA is the Gaussian kernel function, and a numerical method was used to obtain the optimal value of parameter $\mu = 0.06$. IKFDA

was used to obtain the first and second kernel feature vectors, following which the cross plot of the test sets was obtained (Figure 6). As can be seen in Figure 6, it is possible to obtain a good separation between the three types of samples, with each sample forming its own cluster center. The experimental results show that the performance of IKFDA is superior to that of FDA for lithology identification purposes.

6. Conclusion

The optimal kernel Fisher projection of KFDA can be expressed as a generalized characteristic equation by using a regularized method. For multiclass problems, the value of the regularized parameter is a key factor in the application effect, which is largely influenced by human experience. This paper proposes a novel method to optimize the regularized parameter using a numerical method. Thus, by selecting an optimal value for the regularized parameter, it becomes possible to solve the generalized characteristic equation, thereby eliminating the human factor. The effectiveness of



FIGURE 5: FDA identification result.



FIGURE 6: IKFDA identification result.

the IKFDA was demonstrated by applying it to the nonlinearly separable *Iris* and lithology data sets. The experimental results indicated that successful separation was achieved.

In this paper, the selection of the regularized parameter depended on the numerical method. We analyzed the applied validity of the improved Kernel Fisher discriminant analysis without performing deep theoretical analysis. This needs to be addressed by conducting further research in the future.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors are most grateful for the valuable advice on the revision of the paper from the reviewers. The work was supported by the Foundation of Geomathematics Key Laboratory of Sichuan Province.

References

- A. J. Liu, *The Rock Physics Research of Typical Tight Clastic Reservoir in Western Sichuan*, Chengdu University of Technology, ChengDu, China, 2010.
- [2] S. Q. Wang, Comprehensive Reservoir Evaluation of Xujiahe Formation of the West Sichuan by Log Welling in XC Structure, Chengdu University of Technology, Chengdu, China, 2009.
- [3] B.-Z. Hsieh, C. Lewis, and Z.-S. Lin, "Lithology identification of aquifers from geophysical well logs and fuzzy logic analysis: Shui-Lin Area, Taiwan," *Computers & Geosciences*, vol. 31, no. 3, pp. 263–275, 2005.
- [4] Y. X. Shao, Q. Chen, and D. M. Zhang, "The application of improved BP neural network algorithm in lithology recognition," in Advances in Computation and Intelligence: Proceedings of the 3rd International Symposium, ISICA 2008 Wuhan, China, December 19–21, 2008, L. Kang, Z. Cai, X. Yan, and Y. Liu, Eds., vol. 5370 of Lecture Notes in Computer Science, pp. 342–349, Springer, Berlin, Germany, 2008.
- [5] J. Z. Zhang, Q. S. Guan, J. Q. Tan et al., "Application of Fisher discrimination to volcanic lithology identification," *Xinjiang Petroleum Geology*, vol. 29, no. 6, pp. 761–764, 2008.
- [6] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Muller, "Fisher discriminant analysis with kernels," in *Proceedings of the* 9th IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing (NNSP '99), pp. 41–48, Madison, Wis, USA, August 1999.
- [7] S. A. Billings and K. L. Lee, "Nonlinear Fisher discriminant analysis using a minimum squared error cost function and the orthogonal least squares algorithm," *Neural Networks*, vol. 15, no. 1, pp. 263–270, 2002.
- [8] A. J. Smola and B. Schökopf, "Sparse greedy matrix approximation for machine learning," in *Proceedings of the 17th International Conference on Machine Learning*, pp. 911–918, San Francisco, Calif, USA, 2000.
- [9] J. Liu, F. Zhao, and Y. Liu, "Learning kernel parameters for kernel Fisher discriminant analysis," *Pattern Recognition Letters*, vol. 34, no. 9, pp. 1026–1031, 2013.
- [10] Y. Xu, J. Y. Yang, J. F. Lu, and D.-J. Yu, "An efficient renovation on kernel Fisher discriminant analysis and face recognition experiments," *Pattern Recognition*, vol. 37, no. 10, pp. 2091–2094, 2004.
- [11] Q. Zhu, "Reformative nonlinear feature extraction using kernel MSE," *Neurocomputing*, vol. 73, no. 16–18, pp. 3334–3337, 2010.
- [12] J. Wang, Q. Li, J. You, and Q. Zhao, "Fast kernel Fisher discriminant analysis via approximating the kernel principal component analysis," *Neurocomputing*, vol. 74, no. 17, pp. 3313– 3322, 2011.
- [13] G. Fung, M. Dundar, J. Bi, and B. Rao, "A fast iterative algorithm for Fisher Discriminant using heterogeneous kernels," in *Proceedings of the 21st International Conference on Machine Learning (ICML '04)*, pp. 264–272, ACM, July 2004.
- [14] R. Khemchandani, Jayadeva, and S. Chandra, "Learning the optimal kernel for Fisher discriminant analysis via second order

cone programming," *European Journal of Operational Research*, vol. 203, no. 3, pp. 692–697, 2010.

- [15] G. Q. Wang and G. Q. Ding, *Face Recognition Using KFDA-LLE*, Springer, Berlin, Germany, 2011.
- [16] J. H. Li and P. L. Cui, "Improved kernel fisher discriminant analysis for fault diagnosis," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1423–1432, 2009.
- [17] H.-W. Cho, "An orthogonally filtered tree classifier based on nonlinear kernel-based optimal representation of data," *Expert Systems with Applications*, vol. 34, no. 2, pp. 1028–1037, 2008.
- [18] Q. Zhang, J. W. Li, and Z. P. Zhang, *Efficient Semantic Kernel-Based Text Classification Using Matching Pursuit KFDA*, Springer, Berlin, Germany, 2011.
- [19] J. H. Xu, X. G. Zhang, and Y. D. Li, "Application of kernel Fisher discriminanting technique to prediction of hydrocarbon reservoir," *Oil Geophysical Prospecting*, vol. 37, no. 2, pp. 170–174, 2002.
- [20] N. Louw and S. J. Steel, "Variable selection in kernel Fisher discriminant analysis by means of recursive feature elimination," *Computational Statistics & Data Analysis*, vol. 51, no. 3, pp. 2043–2055, 2006.
- [21] M. Volpi, G. P. Petropoulos, and M. Kanevski, "Flooding extent cartography with Landsat TM imagery and regularized kernel Fisher's discriminant analysis," *Computers & Geosciences*, vol. 57, pp. 24–31, 2013.
- [22] K.-L. Wu and M.-S. Yang, "A cluster validity index for fuzzy clustering," *Pattern Recognition Letters*, vol. 26, no. 9, pp. 1275– 1291, 2005.
- [23] Z. Volkovich, Z. Barzily, and L. Morozensky, "A statistical model of cluster stability," *Pattern Recognition*, vol. 41, no. 7, pp. 2174– 2188, 2008.
- [24] S. S. Khan and A. Ahmad, "Cluster center initialization algorithm for *K*-means clustering," *Pattern Recognition Letters*, vol. 25, no. 11, pp. 1293–1302, 2004.
- [25] X. Zhang, Performance Monitoring, Fault Diagnosis, and Quality Prediction Based on Statistical Theory, Shanghai Jiao Tong University, Shanghai, China, 2008.



The Scientific World Journal





Decision Sciences







Journal of Probability and Statistics



Hindawi Submit your manuscripts at





International Journal of Differential Equations





International Journal of Combinatorics





Mathematical Problems in Engineering



Abstract and Applied Analysis



Discrete Dynamics in Nature and Society







Journal of Function Spaces



International Journal of Stochastic Analysis



Journal of Optimization