Journal of Petroleum Science and Engineering 143 (2016) 95–102

Contents lists available at ScienceDirect

Journal of Petroleum Science and Engineering

journal homepage: www.elsevier.com/locate/petrol

Lithology identification using kernel Fisher discriminant analysis with well logs ☆

Shaoqun Dong [a,b], Zhizhang Wang [a,b,*], Lianbo Zeng [a,b,*]

[a] College of Geoscience, China University of Petroleum, Beijing 102249, China
[b] State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum, Beijing 102249, China

ARTICLE INFO

Article history:
Received 15 August 2015
Received in revised form
28 January 2016
Accepted 22 February 2016
Available online 23 February 2016

Keywords:
Lithology identification
Kernel Fisher discriminant analysis
Linear discriminant analysis
Kernel trick

ABSTRACT

Lithology identification is a key step in reservoir characterization. Linear discriminant analysis (LDA) is a widely used method for lithology identification. However, LDA suffers from the disadvantage that it can only extract linear features, whereas nonlinear features in the lithological feature space often play a role in lithology identification. In this paper, we introduce kernel Fisher discriminant analysis (KFD), an improved LDA with kernel trick, to overcome the shortcoming of LDA for lithology identification. It includes two processes: raising dimensions to get nonlinear information and reducing dimensions to get classification features. By these processes, it can obtain nonlinear classification features efficiently. To examine the effect of KFD for lithology identification, experiments are implemented on a field data set by KFD and auxiliary methods, namely LDA and traditional nonlinear discriminant analysis (quadratic discriminant analysis, QDA). By comparisons from different aspects, the results show that KFD outperforms LDA and QDA and it is a practicable method for lithology identification.

1. Introduction

Lithology identification is extremely important in reservoir characterization. However, the lithology cannot be measured by well logs directly, and evaluating the lithology from core data is generally of limited use. In fact, it is often predicted based on well log data (Ma, 2011; Rogers et al., 1992; Ali et al., 2006; Grana et al., 2012).

Since the introduction of well logs, many mathematical methods have been used for predicting lithology based on well log data (Delfiner et al., 1987). Linear discriminant analysis (LDA) was introduced early on for lithology identification and it has been proved practical (Busch et al., 1987). However, with the complicating practical reservoir, nonlinear features begin to play a significant role, and LDA underperformed nonlinear algorithm, such as artificial neural network (ANN) (Dubois et al., 2007) and support vector machines (SVM) (Al-Anazi and Gates, 2010a,b). The characteristic that LDA can only extract linear features (Kim and Kittler, 2005) restricts its performance. Nevertheless, LDA has an advantage that both ANN and SVM lack. LDA is a feature extraction method and a dimensionality reduction method, so the process of lithology classification can be easily visualized (Wang and Paliwal, 2003; Tominaga, 1999), while artificial intelligence algorithm (ANN, SVM, etc.) cannot, because they are black boxes to some extent (Tu, 1996; Gardner and Dorling, 1998). Geologists can see and control the lithology identification better by LDA, because classifier is visible when the extracted features are shown in a plane or a space.

To alleviate the shortcomings of LDA mentioned above and remain its advantages, kernel Fisher discriminant analysis (KFD) (Scholkopft and Mullert, 1999; Baudat and Anouar, 2000) is introduced for lithology identification. It is an improved LDA by kernel trick (Shawe-Taylor and Cristianini, 2004). The implementation of KFD includes two steps: (a) map the data into an implicit feature space (usually nonlinear space) by kernel trick; (b) apply LDA in the feature space (Yang et al., 2004; Xu et al., 2004).

In this paper, a field data set is to examine the potential of KFD to identify lithology. The data set from Junggar Basin in China is chosen to carry out comparative experiments by KFD and auxiliary methods. Auxiliary methods include LDA and quadratic discriminant analysis (QDA), which is an extension of LDA with a quadratic decision surface, while LDA has a linear decision surface (Friedman, 1989; Wu et al., 1996).

☆This study is financially supported by the National Natural Science Foundation of China (Grant no. 41272164).

* Corresponding authors at: College of Geoscience, China University of Petroleum, Beijing 102249, China.
E-mail addresses: wang_zhizhang@126.com (Z. Wang), lbzeng@sina.com (L. Zeng).

http://dx.doi.org/10.1016/j.petrol.2016.02.017
0920-4105/

2. Method

In KFD, kernel trick is used to map the data into an implicit feature space. In this way, the nonlinear relations with different forms in the input data can be converted into linear ones indirectly (Liu et al., 2004). Then, LDA can be applied in the feature space, thereby yielding nonlinear decision surfaces in the input space. In this process, the algorithm can actually be implemented in the feature space by virtue of kernel trick, and therefore an explicit mapping process is not required. To better present KFD from simple to complex, we will introduce LDA first and then KFD.

2.1. Linear discriminant analysis (LDA)

LDA is a classical multivariate technique for both dimension reduction and classification (Duda et al., 2001; Subasi and Gursoy, 2010). The data vectors will be transformed into a low dimensional subspace so that the class centroids are spread out as much as possible. In this subspace, LDA works as a simple prototype classifier with linear decision boundaries (Martínez and Kak, 2001).

Given a set of N centered samples (zero mean and unit variance) $\mathbf{x}_j^{(i)} \in \mathbb{R}^p$, where $\mathbf{x}_j^{(i)}$ is the j-th sample in the i-th class, and there are p elements in vector $\mathbf{x}_j^{(i)}$. Unless a special declaration, vectors in this paper are all column vectors. n_i is the number of samples in the i-th class and it satisfies $\sum_{i=1}^{c} n_i = N$, where c is the number of classes. The transform process can be translated to find several projection vectors and map the original data upon the vectors. The mapped data meets the requirements of both maximizing the distance between classes and minimizing the distance within classes. The requirements can be formulated as maximizing

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}} \tag{1}$$

where $\mathbf{S}_w$ is the within-class scatter matrix, and $\mathbf{S}_b$ is the between-class scatter matrix:

$$\mathbf{S}_w = \sum_{i=1}^{c} \sum_{j=1}^{n_i} \frac{1}{N} (\mathbf{x}_j^{(i)} - \mathbf{m}_i)(\mathbf{x}_j^{(i)} - \mathbf{m}_i)^T \tag{2}$$

$$\mathbf{S}_b = \sum_{i=1}^{c-1} \sum_{k=i+1}^{c} \frac{N_i}{N} \frac{N_k}{N} (\mathbf{m}_i - \mathbf{m}_k)(\mathbf{m}_i - \mathbf{m}_k)^T \tag{3}$$

Here, $\mathbf{m}_i$ is the centroid of the i-th class, $\mathbf{m}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_j^{(i)}$

It has been proven that solving the optimization problem is equivalent to finding the largest eigenvectors of $\mathbf{S}_b\mathbf{u} = \lambda \mathbf{S}_w\mathbf{u}$, and $\mathbf{w} = \mathbf{u}/\sqrt{\mathbf{u}^T\mathbf{S}_w\mathbf{u}}$. The classification ability of each projection vector can be measured by the corresponding eigenvalue. When the eigenvalue is larger than the others, the contribution, which the corresponding vector makes to spread out class centroids, is more than the contributions of the other vectors. After obtaining the projected new vectors, they can be employed to identify lithology instead of the original ones by linear classifier.

Take well logs for example. Four logs are available, which are gamma ray (GR), compensated neutron log (CNL), density (DEN) and acoustic log (AC). There are five types of lithology and then $c=5$. For each sample, $\mathbf{x}_j^{(i)} = (GR, CNL, DEN, AC)^T$, $i=1, 2, \ldots, 5$, $j = 1, 2, \ldots, n_i$ and then $p=4$.

Suppose that the number of eigenvectors is four in this example, although it may be 3 or 2, or even 1. $\mathbf{w}_1 = (w_{11}, w_{12}, w_{13}, w_{14})^T$, $\mathbf{w}_2 = (w_{21}, w_{22}, w_{23}, w_{24})^T$, $\mathbf{w}_3 = (w_{31}, w_{32}, w_{33}, w_{34})^T$ and $\mathbf{w}_4 = (w_{41}, w_{42}, w_{43}, w_{44})^T$ are the eigenvectors corresponding to the eigenvalues $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ ($\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$) respectively. In fact, the number of eigenvectors is usually not more than both $c-1$ and p. Then the mapped image becomes $\mathbf{z}_j^{(i)} = (\mathbf{w}_1^T\mathbf{x}_j^{(i)}, \mathbf{w}_2^T\mathbf{x}_j^{(i)}, \mathbf{w}_3^T\mathbf{x}_j^{(i)}, \mathbf{w}_4^T\mathbf{x}_j^{(i)})^T$. However, the first several elements can contain most of the classification features. To determine the number of elements used, the contribution is often used. For instance, $\lambda_k/(\lambda_1 + \lambda_2 + \lambda_3)$ is the contribution of the k-th eigenvector. If $\lambda_1/(\lambda_1 + \lambda_2 + \lambda_3) < threshold$ and $(\lambda_1 + \lambda_2)/(\lambda_1 + \lambda_2 + \lambda_3) \geq threshold$, then $\mathbf{z}_j^{(i)LDA} = (z_1, z_2)^T$. In practice, *threshold* is often more than 0.8. After the extracted features are determined, the centroid of the i-th class can be obtained, $\mathbf{m}_i^{LDA} = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{z}_j^{(i)}$. So far, the LDA model has been built.

For a sample with unknown lithology label, $\mathbf{x} = (GR, CNL, DEN, AC)^T$.

Firstly, get the projected new vector $\mathbf{z} = (z_1, z_2)^T$. Here $z_1 = GR \times w_{11} + CNL \times w_{12} + DEN \times w_{13} + AC \times w_{14}$, $z_2 = GR \times w_{21} + CNL \times w_{22} + DEN \times w_{23} + AC \times w_{24}$.

Secondly, decide the lithology label according to the distance to the class centroid:

$$dist_i = \left| \mathbf{z}^{LDA} - \mathbf{m}_i^{LDA} \right| \tag{4}$$

where $dist_i$ is the distance between $\mathbf{z}^{LDA}$ and the centroid of the i-th class.

When $dist_k$ is the smallest, it will be identified as the k-th lithology.

2.2. Kernel Fisher discriminant analysis (KFD)

KFD was first proposed by Mika et al. for two-class classification (Scholkopft and Mullert, 1999). Then Baudat and Anouar (2000) improved it for multi-class classification and called it general discriminant analysis (GDA). Afterwards KFD in many papers is mostly for multi-class classification (Kodipaka et al., 2007; Cai et al., 2011; Chu et al., 2011; Li et al., 2003). Some papers also call it KFDA (Wei and Wu, 2008; Jin et al., 2012). In this paper, the term KFD is used. It enables solving both two-class and multi-class classification problems.

2.2.1. Mathematical principle

In the pattern recognition theory, linearly non-separable patterns in low-dimensional space can become linearly separable if they are mapped into a high-dimensional space (Cover, 1965). In particular, the original data in KFD are mapped into a feature space by kernel trick. In this section, a detailed introduction to the mathematical background of KFD will be provided.

For the data set $\mathbf{x}_j^{(i)} \in \mathbb{R}^p$ in Section 2.1 and a given nonlinear mapping function ϕ, the input data space $\mathbb{R}^p$ can be mapped into a feature space F, and then the images of the original samples under the map ϕ are denoted as $\phi(\mathbf{x}_j^{(i)})$. It is a column vector with t elements. In fact both t and ϕ are difficult to determine and usually unknown. For this problem, we will give a detailed explanation how kernel trick solves it in the following text.

Based on that illustrated in Section 2.1, the features in the feature space extracted by LDA are $\mathbf{w}_k^\phi \cdot \phi(\mathbf{x}_j^{(i)})$, $k = 1, 2, \ldots, m$, and $\{\mathbf{w}_k^\phi | k = 1, 2, \ldots, m\}$ is a set of generalized eigenvectors corresponding to the first m largest generalized eigenvalues $\{\lambda_i | i = 1, 2, \ldots, m\}$ obtained by maximizing $(\mathbf{S}_w^\phi)^{-1}\mathbf{S}_b^\phi$, namely

$$J^\phi(\mathbf{w}^\phi) = \frac{(\mathbf{w}^\phi)^T \mathbf{S}_b^\phi \mathbf{w}^\phi}{(\mathbf{w}^\phi)^T \mathbf{S}_w^\phi \mathbf{w}^\phi} \tag{5}$$

Here, $\mathbf{S}_w^\phi$ and $\mathbf{S}_b^\phi$ are respectively the within-class and between-class scatter matrices defined in the feature space F.

From the theory of reproducing kernels (Aronszajn, 1950), any solution $\mathbf{w}^\phi \in F$ must lie in the span of all training samples in F, i.e.

$$\mathbf{w}^\phi = \sum_{i=1}^{c} \sum_{j=1}^{n_i} \alpha_{ij} \phi(\mathbf{x}_j^{(i)}) = \mathbf{Q}\boldsymbol{\alpha} \tag{6}$$

where $\mathbf{Q} = [\phi(\mathbf{x}_1^{(1)}), \phi(\mathbf{x}_2^{(1)}), \ldots, \phi(\mathbf{x}_{n_1}^{(1)}), \ldots, \phi(\mathbf{x}_{n_c}^{(c)})]$ and $\boldsymbol{\alpha} = (\alpha_{11}, \alpha_{12}, \ldots, \alpha_{1n_1}, \ldots, \alpha_{cn_c})^T$.