We SP3 05

Integrating Markov Chains for Bayesian Estimation of Vertical Facies Sequences through Linear Discriminant Analysis

W.J.M. Al-Mudhafer* (Louisiana State University)

SUMMARY

The paper introduces new procedure for vertical facies distribution with respect to well logs and core data in a well from a Sandstone formation by improving Linear Discriminant Analysis (LDA) considering cross-validation and Bayes' Theorem. The independent variables are gamma rays, formation density, water saturation, shale volume, log porosity, core porosity, and core permeability. LDA has been chosen to estimate the maximum likelihood and minimize the standard error for the nonlinear relationships between facies & core and log data. LDA seeks linear transformation (discriminate function) of both the independent and dependent variables to produce a new set of transformed values that provides a more accurate discrimination with dimensionality reduction. The counts of facies have been formulated through the transition probability matrix of first-order Markov Chains to be prior knowledge into the Bayesian construction. The resulted predicted probability (posterior) has been estimated from LDA based on Bayes' theorem that represents the relationship between posterior with the conditional probability and the prior knowledge. For assessing LDA model, the Cross-validation was considered to check how well the estimation procedure can be expected to perform. The cross-validation results in decrease the squared difference between the estimated and observed facies leading to decrease the uncertainty.

Introduction

Improving reservoir characterization should be carried out considering precisely prediction of vertical facies distribution among all the wells.

Well logs and core data such as gamma rays, formation density, water saturation, shale volume, log porosity, core porosity, and core permeability have been used to predict the geological facies using Bayesian Linear Discriminant Analysis (LDA) with Jackknifed Prediction.

Discriminant Analysis (DA) is a multivariate statistical method that is used to seek linear transformation (discriminate function) of both the independent (well logs) and dependent variables (facies) in order to produce a new set of features that best separates observations into different classes with dimensionality reduction (Pires and Branco, 2010).

LDA assumes that the data in each class has normal distribution with unique covariance matrix for each class (Albert, 2007).

Methodology

LDA has been chosen to estimate the maximum likelihood and minimize the standard error for the nonlinear relationships between facies, core and well log data. The new set of linear combination separates the multinomial outcomes in order to reduce the dimensionality.

LDA has been done in two different ways. Either separate the most of observation for training and the rest for validating Or consider Cross-Validation which no need to separate the data. Both of the methods have been done in this paper using the statistical computing packages "R".

In order to use LDA, I need to first split the data into a part used to train the classifier, and another part to test the classifier. For this problem, I considered a proportion 80:20 split, approximately. The above has told R to train the classifier on the data available in the independent variables of the data set "strain" using the knowledge that this data is grouped to the categorical variable (Facies) of "strain". The counts of facies have been formulated through the transition probability matrix of first-order Markov Chains to be prior knowledge and the resulted predicted probability (posterior) has been estimated based on Bayes' theorem that represents the relationship between predicted probability (posterior) with the conditional probability (maximum likelihood) and the prior knowledge.

$$P(F_i | L_1, L_2, L_3, \dots, L_{13}) = \frac{P(L_1, L_2, L_3, \dots, L_{13} | F_i)}{P(L_1, L_2, L_3, \dots, L_{13})} \times P(F_i)$$

Where:

$L_1, L_2, L_3, \dots, L_{13}$: the 13 well log data.

Prior: $P(F_i)$

Posterior: $P(F_i | L_1, L_2, L_3, \dots, L_{13})$

Maximum Likelihood: $P(L_1, L_2, L_3, \dots, L_{13} | F_i)$

For assessing the LDA model, the Cross-validation has been considered to check how well the estimation procedure can be expected to perform. It is accomplished by comparing estimates with true values, that is, re-estimate each known sample value for surrounding information to analyze the error distribution by plotting the residual on maps to check any persistent overestimation or underestimation. The cross-validation results in decrease the squared difference between the estimated and observed categorical variables (facies) leading to decrease the degree of uncertainty.

Figure 1 Linear Discriminant Predicted Probability.

Figure 2 Bayesian Linear Discriminant Predicted Probability that represents the distribution of all facies with depths.

Conclusions

LDA has been efficiently applied to estimate the maximum likelihood and minimize the standard error for the nonlinear relationships between facies, core and well log data. Bayes' theorem has improved the predicted probability by including the beta distribution as prior knowledge. The cross-validation results in decrease the squared difference between the estimated and observed categorical variables (facies) leading to decrease the degree of uncertainty.

References

- Pires, A.M. and Branco, J.A. [2010] Projection-pursuit approach to robust linear discriminant analysis. *Journal of Multivariate Analysis*, **101**, 2464-2485.
- Albert, J. [2007] *Bayesian computation with R. Use R!* New York, Springer.
- Al-Mudhafer, W.J. [2014] Multinomial Logistic Regression for Bayesian Estimation of Vertical Facies Modeling in Heterogeneous Sandstone Reservoirs. OTC 24732, paper accepted for presentation at *Offshore Technology Conference Asia*, Kuala Lumpur, Malaysia, 25-28 March 2014