Journal of Petroleum Science and Engineering 133 (2015) 633–645

Contents lists available at ScienceDirect

Journal of Petroleum Science and Engineering

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

Application of dimensionality reduction technique to improve geophysical log data classification performance in crystalline rocks

Ahmed Amara Konaté [a,d,*], Heping Pan [a,*], Huolin Ma [a,*], Xiurong Cao [b], Yao Yevenyo Ziggah [c], Margaret Oloo [b], Nashir Khan [a]

[a] Institute of Geophysics and Geomatics, China University of Geosciences (Wuhan), Lumo Road 388, Postal code: 430074, Wuhan, Hubei, China
[b] Faculty of Earth Resource, China University of Geosciences (Wuhan), Lumo Road 388, Postal code: 430074, Wuhan, Hubei, China
[c] Faculty of Information Engineering, China University of Geosciences (Wuhan), Lumo Road 388, Postal code: 430074, Wuhan, Hubei, China
[d] Institut Supérieur des Mines et Géologie de Boké / Ministère de l'Enseignement Supérieur et de la Recherche Scientifique, République de Guinée

ARTICLE INFO

Article history:
Received 11 March 2015
Received in revised form
2 June 2015
Accepted 30 June 2015
Available online 8 July 2015

Keywords:
Well log interpretation
Dimensionality reduction
Crystalline rocks
Principal component analysis
Linear discriminant analysis
Machine learning

ABSTRACT

In this study, dimensionality reduction technique to improve geophysical log data classification performance in crystalline rocks is presented. In fact, in complex geological situations such as the study area in context, more complex nonlinear functional behaviors exist for well log classification purpose; thus posing challenges in accurate identification of log curves for this purpose. Dimensionality reduction (DR) using Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were used here to reduce the dimensionality of the original log set of Chinese Continental Scientific Drilling Main Hole to a convenient size, and then feed the reduced-log set into the classifiers. Three classifiers were addressed, namely, Support vector Machines, Feed forward Back Propagation and Radial Basis Function Neural Networks in the classification of metamorphic rocks. The strategy of combining dimensionality reduction methods and classifiers was demonstrated and discussed. The results showed that the reduced log sets found from DR can separate the metamorphic rocks types better or almost as well as the original log set. Therefore LDA and PCA can be suitable to be performed before geophysical well log data classification in the context of crystalline rocks.

1. Introduction

Crystalline rocks cover a wide spectrum of igneous, metamorphic and some sedimentary rocks in which recrystallization process has been important to their formation. These occur in a range of continental and oceanic settings, where a number of boreholes have been and continued to be drilled in order to provide information of the Earth's crust by means of geophysical surveys and geological mapping (Harvey et al., 2005). Several studies have been reported by geophysicists based on crystalline rocks using well logging data (see for example Pratsone et al., 1992; Pechnig et al.,1997, 2005; Bartetzko et al., 2005; Maiti and Tiwari, 2009; Luo and Pan, 2010; Pan et al., 2010). The general understanding collected from these studies show that, the well logging data from crystalline rocks are not simple to analyze because of their complicated geological characteristics and the difficulty in understanding and using the intensive information content in these data. Moreover, up to now, there is no systematic formulated interpretation/classification methods available for crystalline rocks (Pechnig et al., 2005) in geophysics; thus posing challenges in accurate identification of log curve for this purpose.

By some criteria of relevance, the log specialist/interpreter always considerably reduces the dimensionality of the log set to be made. Such reduction may be based on (1) exclusion of logs by specific judgment (e.g. visual inspection); (2) statistical techniques. The former approach tend to a reduction of logs in that the number of log to be observed is reduced. However, this method has proven to be time consuming and error prone (Perez-Munoz et al., 2013 and reference therein). Based on this, there has been a move towards the use of computer aided statistical techniques for extracting/selecting meaningful logs from the original log set. This latter approach, in general, can reduce logs by showing that a subset of the logs is “important” for discriminating rocks type. This study is concerned with the application of this latter approach. It is worth stating that there has been a concerted effort to apply statistical cross plotting method on well logs analyses in crystalline rocks. For example, Pechnig et al. (2001, 2005); Bartetzko et al. (2005); including CCSD-MH (Niu et al. 2004; Pan et al., 2005; Salim et al., 2008; Pan et al., 2010; Luo and Pan, 2010) demonstrated that cross plotting graphic representation can be applied

* Corresponding authors.
E-mail addresses: konate77@yahoo.fr (A.A. Konaté), panpinge@163.com (H. Pan), mhl70@163.com (H. Ma).

http://dx.doi.org/10.1016/j.petrol.2015.06.035
0920-4105/

successfully in crystalline rocks. Although cross-plotting can be utilized to explore relationships between two logs, nevertheless it is unable to clearly reveal with understanding the relationships that may be in the whole data (Kassenaar, 1991). Another problem in this area is that the cross-plots method relies on the experience of the log specialist. It can therefore can generate multiple interpretations (Saggaf and Nebrija, 2000; Amirgaliev et al., 2014). In looking for a reduced set of logs, statistical methods such as principal component analysis (PCA), and linear discriminant analysis (LDA) have been largely concerned with Dimensionality Reduction (DR), which consist of reducing the dimensionality by projecting the original p-dimensional space on a k-dimensional subspace (where $k < p$) through a transformation (Khosla, 2004). DR is important for the following reasons (Cunningham and Carney, 2000): (i) To build better classifiers: better quality classifiers can be built by eliminating irrelevant features; (ii) Economy of representation: allow problems to be represented as compactly as possible; (iii) Knowledge discovery: discover what features are and are not influential dataset.

PCA and LDA have been successfully applied in geophysics (see example Posadas et al., 1993; Rezaie and Smorgrav, 2012) including well logging problems (see exemple Kassenaar, 1991; Urbancic and Bailey, 1988; Xu et al., 2006; Gelfort, 2006). However, so far they have not yet been widely used in the context of crystalline rocks. Especially within the study area of this research. Therefore, this study constitutes a good foundation for quality control of PCA and LDA dimensionality reduction methods from future deep boreholes drilled in similar geological contexts.

With reference to the above, the purpose of this study was focuses on reducing the dimensionality of the original log set of Chinese Continental Scientific Drilling Main Hole (CCSD-MH) to a convenient size, holding as much of the original information as

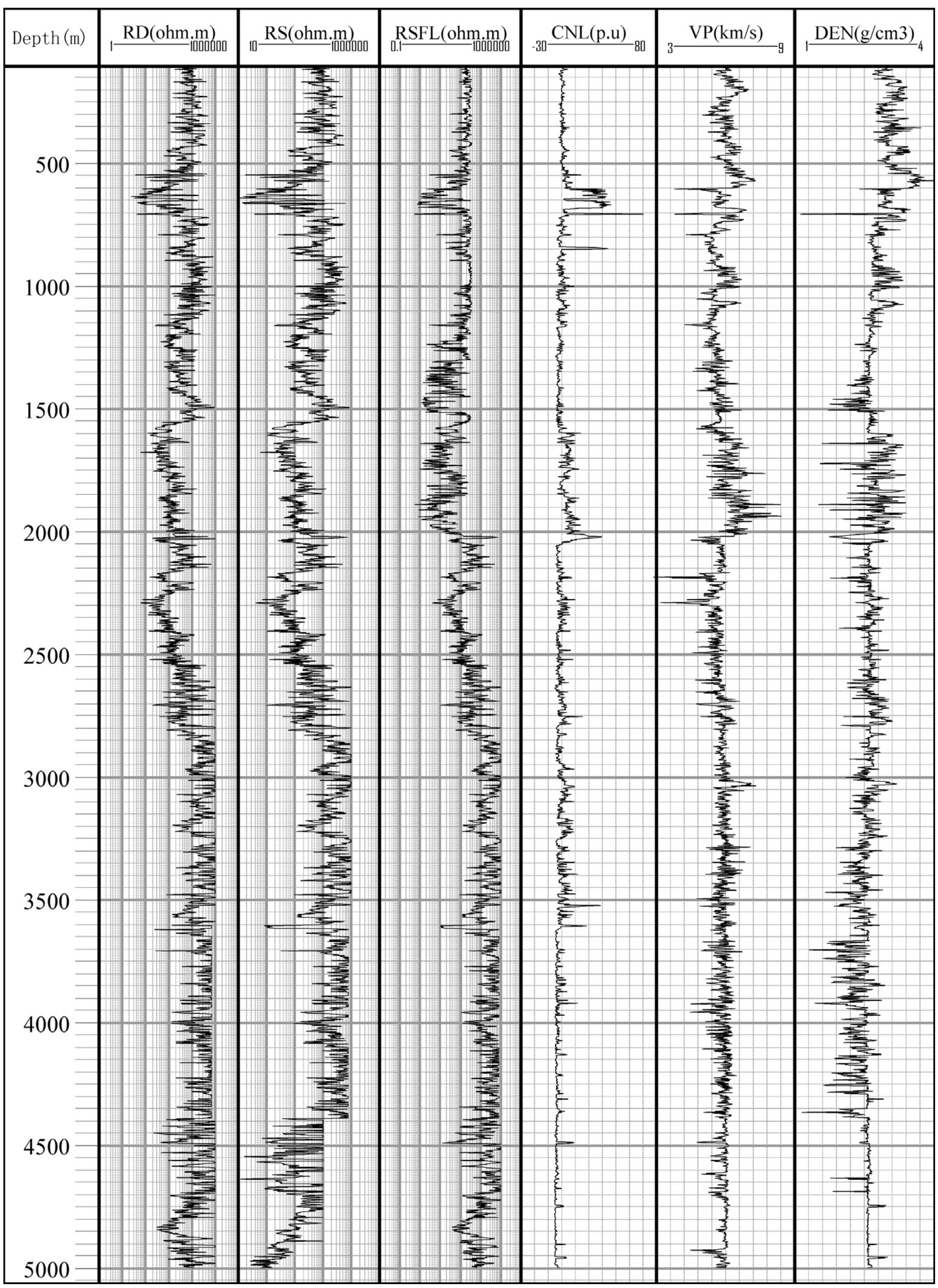

Fig. 1. Profile of the logs used in this study.,