Cluster analysis of 3D seismic data for oil and gas exploration

D. R. S. Moraes, R. P. Espíndola, A. G. Evsukoff
& N. F. F. Ebecken
COPPE/Federal University of Rio de Janeiro, Brazil

Abstract

Seismic attributes are information extracted from seismic data and constitute important tools to estimate the geological structure of a place, helping the understanding of the subsurface and reducing the uncertainties on interpretations. This comprehension is crucial to tasks such as lithology prediction and reservoir characterization. Seismic attributes are generated by transforming data from a seismic line (two dimensional data) or a seismic volume (three dimensional data). This work presents a study of clustering algorithms to these attributes and the techniques employed follow two distinct approaches: a self organizing map to perform crisp clustering and fuzzy c-means to perform partial clustering. The evaluations of the partitions are performed with the PBM index which indicates the best number of groups. Data from a Brazilian oil field is used to test the algorithms.
Keywords: clustering, neural networks, fuzzy sets, cluster validity.

1 Introduction

In oil industry, the decision making of exploration and exploitation processes involves dealing with geological, geophysical and geochemical data. The evolution of computer hardware technology and the development of data base management systems (DBMS) allow the data to processed and stored in very huge databases. This fact represents a great opportunity for a company to record all data from its operations. However, it makes the analysis and understanding of such data difficult. Thus, Data Mining techniques may help the experts on their task by automatically discovering useful information from these huge volumes of data. Examples of activities that can be improved by the use of these techniques


WIT Transactions on Information and Communication Technologies, Vol 37, © 2006 WIT Press
www.witpress.com, ISSN 1743-3517 (on-line)
doi:10.2495/DATA060071

are reservoir characterization, prediction and classification of rock porosity, optimal placement of wells and detection of oil spills [1].

In this work, two well known clustering techniques are employed on 3D seismic data. It was expected that the clusters may help geologists to identify different lithologies on the region studied. The quality of the groups formed is assessed using a cluster validation index called PBM [2]. The clustering algorithms are fuzzy c-means [3] and a neural network approach known as self-organizing map [4]. The dataset was obtained from *Namorado* oil field in Brazil.

This paper is organized as follows: the next section presents the clustering methods used in the application. Section 3 describes the PBM index, used to validate the results generated by the clustering methods and section 4 presents details about the data base and how it was treated. Section 5 shows the results obtained and, in last section, some concluding remarks and suggestions of future research are done.

2 Clustering methods

Clustering is one of the most usual tasks in the process of data mining, helping the discovery and identification of distributions and patterns of interest. It aims to organize a data set into groups of similar elements by detecting implicit relationships between attributes recorded on data [5].

The algorithms studied on this work perform partitioning clustering. This kind of clustering splits a data set into a pre-defined amount of groups or partitions. From an initial partition, such techniques search for the best possible splitting until some stopping criteria is reached. A good partitioning must yields groups very different among themselves as well as formed by similar objects.

Following, the algorithms here studied are presented.

2.1 Fuzzy c-means

On classic partitioning, a data set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ with n objects $\mathbf{x}_i \in R^p$ must be allocated into K groups such that each object belongs to only one group. In other words, given $G_1, G_2, ..., G_K$ groups, it is verified that $G_i \neq \varnothing$ and $G_i \cap G_j = \varnothing$ for every $i, j = 1, ..., K$, $i \neq j$. The partition matrix U(**X**) presents the result, such that each element has $u_{ij} = 1$ if $\mathbf{x}_i \in G_j$ and $u_{ij} = 0$ otherwise.

In fuzzy clustering algorithms, an object can be assigned to different groups with partial membership grades $u_{ij} \in [0,1]$. Fuzzy c-means splits a data set into K groups and assigns a membership grades for all objects by optimizing the following objective function:

$$F = \sum_{i=1}^{n} \sum_{j=1}^{K} u_{ij}^{m} \left\| \mathbf{x}_i - \mathbf{w}_j \right\|^2 \quad (1)$$

in which the fuzzifier m is a real value greater than 1, u_{ij} is the membership grade of object $\mathbf{x}_i$ to group G_j, and $\mathbf{w}_j$ is the center of G_j.

As the optimization process continues, the membership grades and the centers of groups are updated by the following expressions:

$$u_{ij} = \sum_{k=1}^{K} \left(\frac{\left\| \mathbf{x}_i - \mathbf{w}_j \right\|}{\left\| \mathbf{x}_i - \mathbf{w}_k \right\|} \right)^{\frac{2}{1-m}} \tag{2}$$

$$\mathbf{w}_j = \frac{\sum_{i=1}^{n} u_{ij}^m \cdot \mathbf{x}_i}{\sum_{i=1}^{n} u_{ij}^m} \tag{3}$$

When no more changes occur on partition matrix, the algorithm is stopped.

2.2 SOM neural networks

Artificial neural networks are computing models characterized by systems that remind (in some level) the human brain structure. Self-organizing maps (SOM) is a network model which main characteristic is the neighborhood concept; that is, the network learns to recognize neighbor sections in the training data as well as the topology of the training set [4].

Neurons in the SOM's competition layer are originally arranged in fixed positions according to a topology function. It is possible to arrange the neurons topology into rectangular or hexagonal shapes. This network employs competitive learning and the closest neuron to the input data is the output unit. All the neurons from a certain neighborhood of this winner neuron are updated by Kohonen's rule and the value of radius determines the neighborhood space.

The neural network determines the winner neuron by

$$\left\| \mathbf{X} - \mathbf{G}_w \right\| = \min_j \left\{ \left\| \mathbf{X} - \mathbf{G}_j \right\| \right\}, \quad \text{or} \quad \mathbf{w} = \arg\min_j \left\{ \left\| \mathbf{X} - \mathbf{G}_j \right\| \right\}. \tag{4}$$

in which X is the input data vector, $\mathbf{G}_j$ is the synapses vector and w is the index of the winner neuron. The synapses of the winner neuron are adjusted by means of a linear combination between the preceding and the current weight:

$$\mathbf{G}_j(t+1) = \mathbf{G}_j(t) + h_{wj}(t)[\mathbf{X}(t) - \mathbf{G}_j(t)] \tag{5}$$

in which t is the discrete-time coordinate and $h_{wj}(t)$ is the neighborhood kernel. Usually this kernel is calculated as

$$h_{wj}(t) = h(\left\| r_w - r_j \right\|, t) \tag{6}$$

where r_w and r_j are the radius of the neuron w and j, respectively.

The last expression reflects the general definition of a neighborhood kernel.

In this work, it was used the concept of bubble neighborhood, defined by

$$h_{wj} = \begin{cases} \alpha(t) & \text{, if neuron j belongs to neuron w neighborhood} \\ 0 & \text{, otherwise} \end{cases} \quad (7)$$

in which $\alpha(t)$ is a monotonically decreasing function of time $(0 < \alpha(t) < 1)$.

In SOM networks, the number of groups is determined by the network size. In this study, to permit the comparison of the results with the fuzzy c-means clustering, the first SOM layer was formed by only one neuron and the second SOM layer has the number of groups it is applied to.

3 Cluster validation index

One of the greatest difficulties of applying a partitioning algorithm is to set the amount of groups and frequently no one knows the correct value with antecedence. Thus, the clustering algorithm is executed with diverse values of groups and the best one can be identified by cluster validation indexes. As mentioned before, a good clustering result exhibits compact and separated groups and these indexes evaluate the groupings based on these features. The measure used here is PBM index:

$$PBM(K) = \left(\frac{1}{k} \times \frac{E_1}{E_K} \times D_K \right)^2, \quad (8)$$

in which K is the number of clusters, E_K is the sum of intra-cluster distances, E_1 is the sum of distances of all points to the center of data, and D_K is the maximum of within-cluster distances. They are given by the following expressions:

$$E_K = \sum_{j=1}^{k} \sum_{i=1}^{n} u_{ij} d(\mathbf{x}_i, \mathbf{w}_j), \quad (9)$$

$$D_K = \max_{i,j=1}^{k} d(\mathbf{w}_i, \mathbf{w}_j). \quad (10)$$

On these formulas, u_{ij} is degree of membership of object $\mathbf{x}_i$ to group G_j with center $\mathbf{w}_j$ and d is the distance function. The greater is the index value the best is the partition.

4 Seismic data and experiments performed

In order to recognize and classify reservoir characteristics, geoscientists analyze the variability in seismic reflections [6]. This activity demands a professional able to identify subtle waveform characteristics and tools to measuring these characteristics are given by seismic attributes. So these attributes are important in reservoir characterization [7] and the correct use of them depends on the experience of a geoscientist, that is, his understanding and interpretation of their behavior.

The 3D seismic data used in this work are from *Namorado* oil field at *Campos* basin in Brazil. The seismic attributes used on this research were created using OpendTect software (www.opendtect.org), which is able to generate and visualize them. The attributes used are amplitude, cosine phase and Hilbert transformation. A dataset with over 223000 records was created using these ones.

Both clustering algorithms were run from 2 to 10 partitions in order to identify the best number of groups. As they are randomly initiated (centers coordinates and memberships grades), each one was performed 10 times to obtain the mean results. Table 1 shows the algorithms settings.

Therefore the data set was approached by one setting of fuzzy c-means and two of SOM network, with rectangular or hexagonal topology.

Table 1: The settings of algorithms.

Parameter	Value
Fuzzy c-means	
fuzzifier	1.5
error tolerance	0.1
SOM	
neurons on first layer	1
neurons on second layer	from 2 to 10
epochs	10^6
learning rate	0.1
radius	0
neighborhood function	Bubble
topology	rectangular or hexagonal

Table 2: Centers coordinates of groups.

Algorithm	Coordinates								
	Amplitude			Cosine phase			Hilbert transform.		
	G1	G2	G3	G1	G2	G3	G1	G2	G3
Fuzzy c-means	.476	.476	.476	.895	.497	.103	.171	.165	.167
SOM rectangular	.476	.472	.476	.890	.502	.109	.177	.173	.171
SOM hexagonal	.476	.472	.476	.890	.502	.109	.177	.173	.171

5 Results

Figures 1-3 shows the PBM index variations with the number of groups for the three experiments. As it can be noticed, both clustering algorithms found three groups as the best value for partition the data, which suggests that it is the correct one. Moreover, almost all the values of PBM index were very similar, indicating that the groups formed on this data by fuzzy c-means and SOM networks have comparable quality. Table 2 shows the mean values (after normalization) of the centers coordinates of the three groups for each experiment and it can be seen that they are almost the same.

Figure 4 shows a clustering result visualized in OpendTect software and it can be seen three groups (black, white and gray).

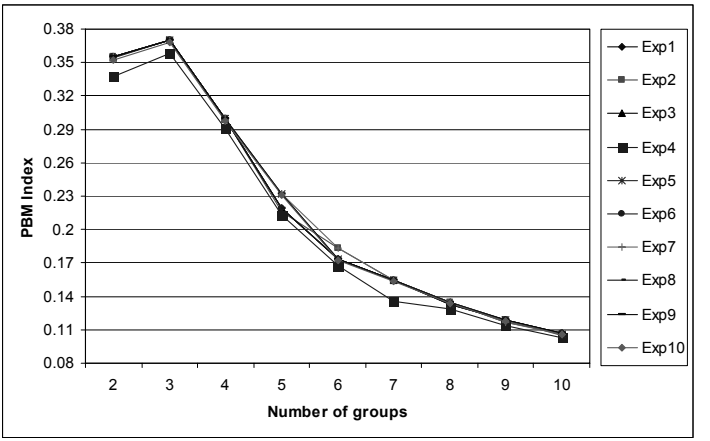

Figure 1: Results from fuzzy c-means.

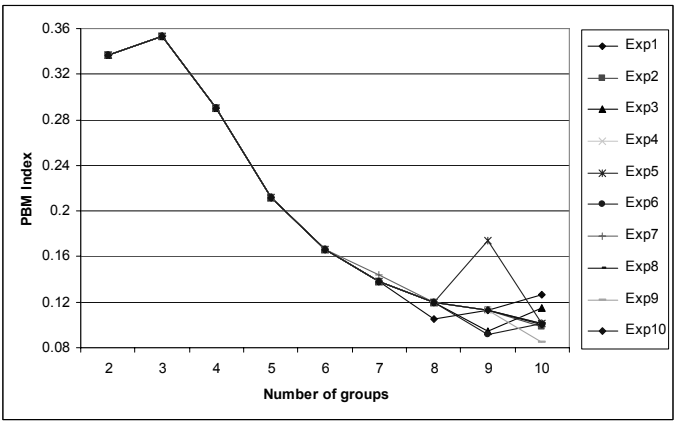

Figure 2: Results from SOM with rectangular topology.

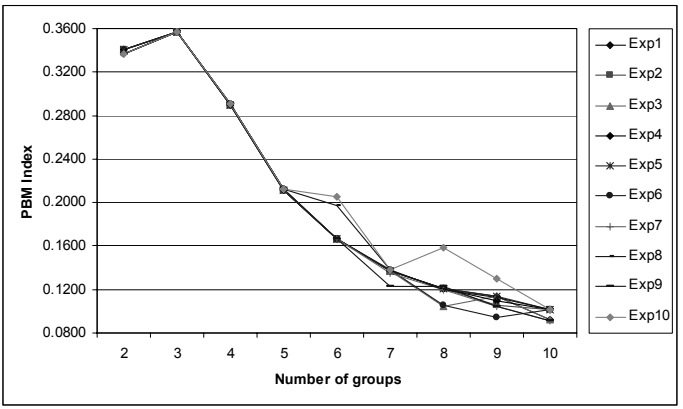

Figure 3: Results from SOM with hexagonal topology.

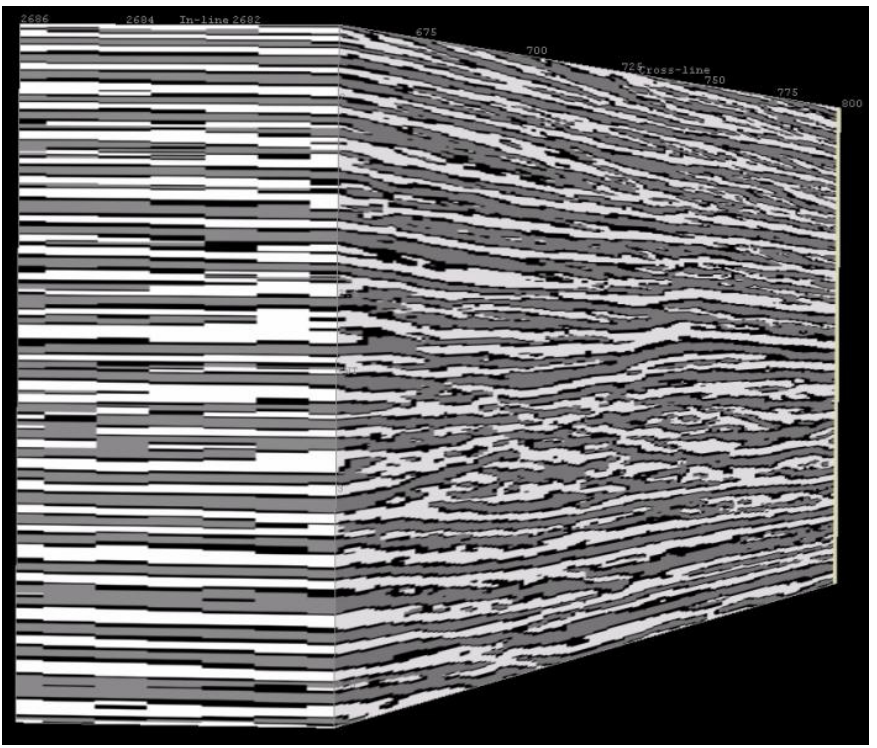

Figure 4: Visualization of a clustering result.

6 Conclusions

This study aimed to show that clustering seismic data may help geologists understand the lithologies of the subsurface. It was generated a dataset with

thousands records and three relevant seismic attributes and was possible to see that both clustering algorithms, self-organizing maps and fuzzy c-means, achieved very similar partitioning – the same number of groups with centers almost coincident. In order to verify the quality of a partitioning, it was employed the PBM cluster validation index.

Future works should consider the generation and use of other seismic attributes and the presentation of the results to a specialist to qualify them. Other clustering techniques specific to signal processing will be used too.

Acknowledgments

This work was partially supported by HP Brazil R&D and the Brazilian research institutions FINEP and CNPq. The authors are also grateful to the Brazilian Petroleum Agency (ANP) who provides the data for this study.

References

[1] Aminzadeh, F., Applications of AI and soft computing for challenging problems in the oil industry. *Journal of Petroleum Science and Engineering*, **47**, pp. 5-14, 2005.

[2] Pakhiraa, M.K., Bandyopadhyayb, S., Maulikc, U., Validity index for crisp and fuzzy clusters, *Pattern Recognition*, **37**, pp. 487-501, 2004.

[3] Bezdek, J.C., *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press: New York, 1981.

[4] Kohonen, T., *Self-Organizing Maps*, Springer: New York, 2001.

[5] Jain, A.K., Murty, M.N., Flynn, P.J., Data Clustering: A Review, *ACM Computing Surveys*, **31(3)**, pp. 264-323, 1999.

[6] Bodine, J.H., Waveform analysis with seismic attributes. *Oil & Gas Journal*, **84(24)**, pp. 59-63, 1984.

[7] Taner, M.T., Koehler, F., Sheriff, R.E., Complex seismic trace analysis, *Geophysics*, **44**, pp. 1041-1063, 1979.