Journal of Biodiversity and Environmental Sciences (JBES)
ISSN: 2220-6663 (Print) 2222-3045 (Online)
Vol. 6, No. 3, p. 489-495, 2015
http://www.innspub.net

INNSPUB
RESEARCH PAPER
OPEN ACCESS

Determination of lithology boundary of jahrom formation in hendijan field in persian gulf using fuzzy clustering

Seyede Tayebe Khalili*, Arash Vakili

Earthquake Research Center, Ferdowsi University of Mashhad (FUM), Iran

Article published on March 28, 2015

Key words: Fuzzy clustering, Jahrom information, Lithology, Log.

Abstract

Clustering is one of the main tools in introduction of similar patterns recognition which makes the analysis of existing data more accurate and comfortable. Clustering is used in different branches. In the hydrocarbon exploration activities, determination of boundary between formations is an important factor of hydrocarbon fields. Therefore, in this research, the logs collected from the area under study, were clustered using fuzzy clustering methods and the results were compared with the results from determination of actual boundary. The dolomite formation of Jahrom which Pabde shale formation located at its lower boundary has been studied. The goal of the study is to recognition of the boundary between the two formations using fuzzy clustering. A thickness with 300 meters length has been studied. Input data are logs data including DT, RHOB, PE, FDC, CGR, SGR, GR, CNL, NPHI and PEF which are classified in six separate groups with 3 members and one group with 4 members. To determine the degree of success of clustering, the ratio of within cluster distance to between cluster distances has been used. Because used logs in this study are able to recognize lithology. Fuzzy clustering with 3 member was partly successful to recognition the number of lithology. In group of 4 members, clustering was able to recognition lithology with great successful. Since this study has done on the one formation, obviously the logs group presented is valid for similar lithology. But it proves fuzzy clustering is useful and efficient for lithology determination in hydrocarbon field.

***Corresponding Author:** Seyede Tayebe Khalili ✉ st.khalili_geo@ymail.com

Introduction

Clustering is one of the main tools in introduction of similar sets and patterns recognition which makes the analysis of existing data more accurate and comfortable. Clustering is used in different branches. In the hydrocarbon exploration activities, determination of boundary between formations is also considered to be an important factor of hydrocarbon fields. The management of a hydrocarbon reservoir is possible while having proper knowledge and accurate picture of properties of the reservoir. It is time-consuming to compute some of the properties of the reservoir such as boundary determination. Precise understanding of boundaries between lithologies makes future decisions more reliable. In recent years, intelligent systems as a powerful tool for modeling and estimation of different parameters in different branches have been used. Data and pattern are two significant parameters in the world of information. Clustering (Witold and Kaoru, 2008) and its ability to enter data space and detect their structures have made it one of the most ideal mechanisms to work with huge amount of data. Recent advances in clustering have led it to be applied in different activities and issues (Yang, 1993). Clustering is the action of finding structures in a data set which lack any classification (Zhong *et al.*, 2008). In the other words, it can be said that clustering is organizing data into groups according to their similarities. Therefore, members of one group are similar to each other and different from members of other groups. (Zhong *et al.*, 2008, Feng *et al.*, 2007, Dong *et al.*, 2006). Here distance is the similarity criterion (Lee and Pedrycz, 2010) meaning that objects which are close together are organized into the same group (de Carvalho *et al.*, 2006). In clustering, determination of distance is very important. Distance which refers to dissimilarity helps us to move in data space and form clusters. By calculating the distance between two data, it can be found that how much they are close to each other and based on this, they can be put in the same group. There are different mathematical functions for calculating distance such as Euclidean, Hamming, etc which are used in different clustering algorithms including fuzzy c-means, k- means and hierarchical clustering. Among all clustering algorithms, fuzzy algorithm is well-known. What makes this algorithm different from other algorithm is that here one data can belong to two or more clusters simultaneously. The belonging degree of a data to a cluster is called membership degree (Soto *et al.,* 2008). It should be noted clustering is unsupervised data classification and it uses in various field. But it has not used for lithology determination yet. In this case, it is possible evaluated performance of fuzzy clustering. The geologists can lithology by used logs in this study. Therefore they have used as data for clustering. Formation selected that it has heterogeneous lithology, and logs are data that it is necessary to be classified as regular and logical. If there is reasonable conformity among clustering answer and geologist answer (actual column), it shows that clustering can succeed in this area. Because of the high number of effective logs should be classified in different groups and their effect will be investigated.

The aim of clustering is detecting similar clusters of objects among input samples. (Zhu *et al.,* 2011) and this the input samplesare logs. One of the important problems in clustering is the number of clusters. In some algorithms, it is a prior value, while in others the algorithm determines how many clusters to be made from data. In logs clustering number of clusters Itis determined by a number of different lithologies.

Materials and methods

The geology of the region

In the most parts of the formation, the boundary between the two lithologies is unknown due to the interference of the lithologies. Jahrom information is dolomiteic information which Pabde shale formation located at its lower boundary. This formation has thickness of about 300 m and starts at the depth of about 2600 meters. It is composed of two types of lithologies including dolomite at upper 120 meters, and limestone-dolomite at Lower 170 meters. The aim of this research is to recognition of the boundary between the two lithologies through the use of the fuzzy clustering. Moreover, it is attempted the proper logs for separating the two lithologies to be presented.

Clustering

At first it should be noted clustering is unsupervised data classification and it uses in various field. But it has not used for lithology determination yet. Therefore one formation selected that it has heterogeneous lithology. In this case, it is possible evaluated performance of fuzzy clustering. The geologists can lithology by used logs in this study. Therefore they have used as data for clustering. In this study, different groups are organized into six separate groups and form a clustering, the results from clustering each of which are analyzed for determining boundaries. The following shows groups that are used for clustering:

1. RHOB, GR, DT
2. RHOB, PHIE, GR
3. RHOB, CALL, DT
4. RHOB, NPHI, GR
5. RHOB, PHIE, DT
6. SGR, CGR, PEF

Since clustering with the presence of an improper log may shows incorrect responses. Loges have been combined into different groups and clustering. Then best group of loges and clustering validity has been discussed. The number of clusters is selected to be equal to 3, since a high number of clusters may decrease the efficiency of the results and cause to insufficient details to be considered in the final result. In addition, two clusters may lead to error in the case of gradual boundary. In all samples depth column, compeer with logs, are clustered in order to better organizing the clusters. A histogram has been presented to illustrate between distance to within-cluster distance and the accuracy of the clustering. Finally, results have been exploited from the clustering and the best results which are compatible with the reality have introduced.

Results and discussions

What is fuzzy clustering? To gain a better understanding of fuzzy clustering and its different algorithms first let us to introduce fuzzy sets and their differences with classic sets. In classic sets, a member of the reference set has two statuses. If the member belongs to set A, it corresponds to 1, otherwise 0 refers to its status (Salski, 2007), whereas a member in fuzzy sets can have values between 0 and 1(Salski, 2007, Pedrycz and Hirota, 2008, Lucieer and Lucieer, 2009). In classic clustering any input sample belongs to one and only one cluster. In the other word, clusters do not overlap each other (Salski, 2007).

Now consider a case in which a sample is similar to two or more clusters. In this case, since a sample in classic clusters must be a member of only one set, user should select the final set. It is the main difference between classic and fuzzy clustering. In fact, it reflects the concept of multi-value rather than single-value.

There are many phenomenons that are satisfied by a continuous range between 0 and 1 more properly. Fuzzy logic is defined based on fuzzy sets. Contrary to ordinary sets which have certain boundaries, a fuzzy set is a data set with uncertain boundaries. In practice, we encounter cases in which any sample has the same possibilities for belonging to both societies. The possibility is reflected by a quantity named membership degree which varies from 0 to 1. The set of membership degrees of members of a fuzzy set, A, is called the membership function of A. membership function of a fuzzy set is a mapping of members of A in [0, 1] so that:

$A: x \rightarrow [0, 1]$

In general, any function doing this mapping can be considered as a membership function of a fuzzy set.

C-means clustering algorithm

Similar to the classic c-means algorithm, in this algorithm the number of clusters is a priori. The predefined goal function of this algorithm is as following (Eq.1):

$$J = \sum_{j=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} d_{ik}^{2} = \sum_{j=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} \left\| x_k - v_i \right\|^2 \qquad (1)$$

where m is a real number larger than 1 which is usually considered to be equal to 2. Xk is k-th sample and V refers to the center of i-th cluster. U shows the belonging amount of i-th sample to k-th cluster. The sign ||*|| indicates how similar the sample is to the center of the cluster which can be any function indicating the similarity between the sample and cluster. A matrix, U, with n rows and c column can be defined using matrix Uik whose components can range from 0 to 1. If all elements of U are considered to be 0 or 1, algorithm will be like classic c-mean algorithm. The sum of elements of any column in U is considered to be equal 1 (Eq.2) that:

$$\sum_{j=1}^{c} u_{ik} = 1, \forall k = 1,\ldots,n \qquad (2)$$

It means that the sum of the belonging of any sample to cluster c should be equal to 1. Using above criterion and minimizing the goal function we have

$$u_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{2/(m-1)}} \qquad (3)$$

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^{m} x_k}{\sum_{k=1}^{n} u_{ik}^{m}} \qquad (4)$$

The algorithm stages

1- First cluster matrices with n rows and c columns must be initialized using values between 0 and 1.

2- The center of clusters is calculated.

3- The belonging matrix is calculated using the matrix from stage 2.

4- If $\| U_{i+1} - U_i \| \leq \varepsilon$, the algorithme will terminate otherwise it goes to stage 2.

The groups have been clustered and the steplike chart (Fig.1) of all groups has been illustrated. The chart in the horizontal axis shows the number of clusters which is considered to be 3 in the current study. Indeed drop step place and close to straight line shows the best number of clusters.

The more clusters makes more straight line. But that is not interesting, because inefficient detail and processing time increase. As shown in Fig1 that is logical selecting 3 clusters. The vertical axis explains between distances to within-cluster distances. The more between distance cause increase number of cluster (each data as a cluster) and the less within-cluster distance cause decrease number of cluster (all data as a cluster).

According to the least within-cluster distance and the most out-of-cluster distance, it is clear that, the smaller number the diagram shows, the more accurate responses are obtained.

Since boundary determination is done using fuzzy clustering, in order to be assured of the accuracy of the results and the ability of fuzzy clustering in separating lithologies, it is necessary to compare them with those from real boundary.

The figure 2 compares the fuzzy clustering of all six groups. Grey and white parts refer to dolomite and lime respectively. The ratio of the between cluster distance to within-cluster distance has been shown. In the groups DT, GR and RHOB the ratio is 0.17, the boundaries between grey and white layers are not clear and they are different from those of real columns. Compared to the previous group, the results from clustering for GR, PHIE and RHOB have improved; however, they are not so much exact that can rely on their response. Although this group with b/w is approximately equal 0.132, has the best success rate of clustering, but Clustering has been efficient at bottom parts than top parts. The set of RHOB, DT and CALL has reasonable results for dolomite layers and depicts upper parts with higher quality than other groups. Nevertheless, this group was also unable to recognize boundaries. In addition, the ratio b/w is approximately equal 0.17 is another sign of inability of this group in comparison to other groups. In the group including GR, RHOB and NPHI, despite effective logs have been used fuzzy clustering has not been able to present reasonable responses.

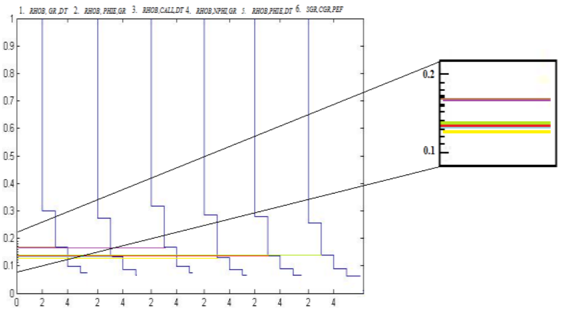

Fig. 1. Steplike chart, The horizontal axis shows the number of clusters and The vertical axis explains between distances to within-cluster distances.

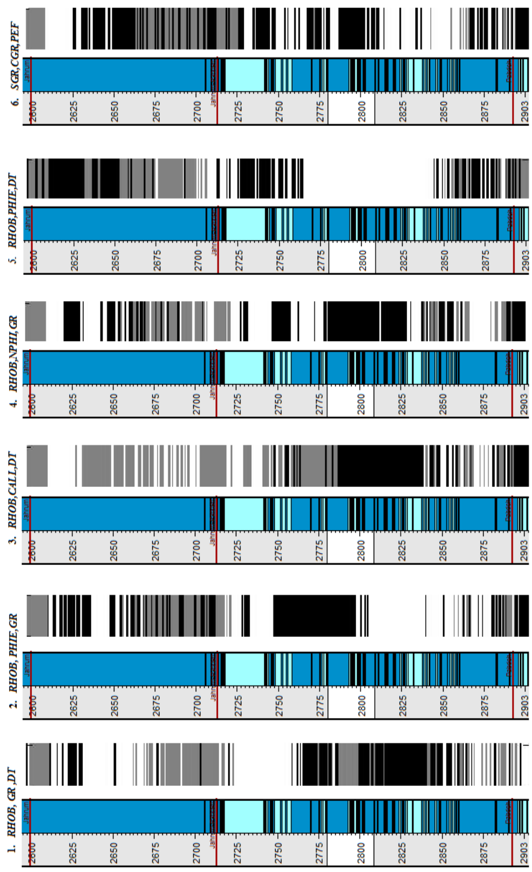

Fig. 2. Fuzzy clustering of all six groups, Grey and white parts refer to dolomite and lime respectively.

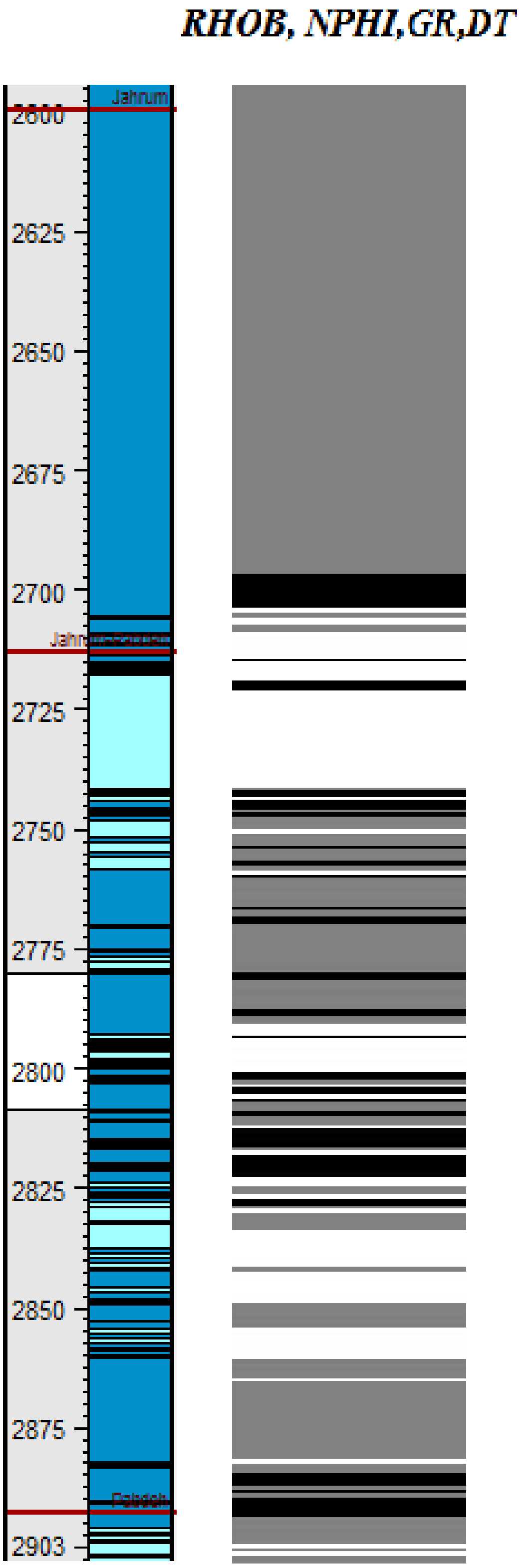

Fig. 3. Lithology is close responses to reality and it is too clear the boundary. There is much compatible between clustering interpretative column and geology column.

Clustering just has been clear responses in top parts. Because with b/w = 0.13, the boundary between two lithologies are opposing and relying on their results will lead to incorrect responses. The two last groups, which have been studied, are RHOB, PHIE and DT with b/w=0.135, and RHOB, NPHI and GR with b/w = 0.14. Similar to other groups, the clustering result for these groups were unsuccessful in determining the boundaries between the two lithologies. But Group 6 has shown better response than group 5. The boundary recognition differs from the real boundary. Although success rates of clustering in all groups are not low, the visual results imply that the formed clusters are not consistent with reality and should not be relied for recognizing boundaries. Partial lithology determination by clustering shows clustering ability at this field. Therefore loge puts on groups of 4 members and finally one group shows the best responses, Figure 3. The maim logs in this are RHOB, NPHI, GR and DT. As shown in Figure 4 lithology is close responses to reality and it is too clear the boundary. There is much compatible between clustering interpretative column and geology column. In this group W/B is about 1.4 that it has been successful than groups of 3 members.

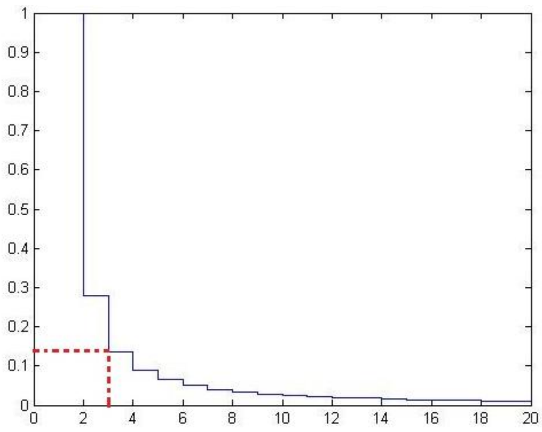

Fig. 4. In RHOB, NPHI,GR And DT group W/B is about 1.4 that it has been successful than groups of 3 members.

Conclusion

Fuzzy clustering and classification of existing data without presence of human has been accepted in different brunches of sciences for far many years. In the current study, the ability of the fuzzy clustering in determining the boundary between two lithologies (dolomite and lime) and the accuracy of the responses has been examined. Regardless log groups examined during this study (Because used logs in this study are able to recognize lithology). Fuzzy clustering with 3 members was unable to present reliable responses. Although success rates were not low, the results from the clustering and real columns were extremely differing. It seems that the use of this method still needs to be examined. It should be noted fuzzy clustering was successful to recognition the number of lithology and it determined 3 layers as optimized layers. However it may shows sufficient result for classifying other lithologies. Inefficiency of the produced clusters caused any idea of the most effective log group to be unreliable. In group of 4 members, clustering was able to recognition lithology with great successful. Since this study has done on the one lime and dolomiteic formation, obviously the logs group presented is valid for similar lithology. But it proves fuzzy clustering is useful and efficient. It is necessary to determine other appropriate logs for other lithologies. As well as according to W/B recommended to study other clustering method such as Hierarchical clustering.

Acknowledgement

Authors would like to express their appreciation to Dr. Seyed Reza Mousavi Harmi, for his help in paper, The authors also would like to thank the Iranian Offshore Oil Company (IOOC), for providing adequate data, and the detailed geological background of the reservoir and Mr. Ali Chehrazi and Mr. Behzad Nazari from for useful peer review of the paper.

References

De Carvalho FAT, Teorio CP, Cavalcanti Junior NL. 2006. Partitional fuzzy clustering methods based on adaptive quadratic distances. Fuzzy Sets and Systems **157**, 2833 – 2857.

Dong Y, Zhuang Y, Chen K, Tai X. 2006. A hierarchical clustering algorithm based on fuzzy graph connectedness. Fuzzy Sets and Systems **157**, 1760 – 1774.

Feng Z, Zhou B, Shen J. 2007. A parallel hierarchical clustering algorithm for PCs cluster system. Neurocomputing **70**, 809–818.

Lee M, Pedrycz W. 2010. Adaptivelearningo fordinal–numericalmappingsthrou ghfuzzy clustering fortheobjectsofmixedfeatures. Fuzzy Sets and Systems **161**, 564–577.

Lucieer V, Lucieer A. 2009. Fuzzy clustering for seafloor classification. Marine Geology **264,** 230–241.

Pedrycz W, Hirota K. 2008. A consensus-driven fuzzy clustering. Pattern Recognition Letters **29**, 1333–1343.

Salski A. 2007. Fuzzy clustering of fuzzy ecological data. Ecological informatics **2**, 262 – 269.

Soto J, Flores-Sintas A, Palarea-Albaladejo J. 2008. Improving probabilities in a fuzzy clustering partition. Fuzzy Sets and Systems **159**, 406 – 421.

Witold P, Kaoru H. 2008. A consensus-driven fuzzy clustering. Pattern Recognition Letters **29**, 1333–1343.

Yang MS. 1993. A Survey of Fuzzy Clustering. Mathl. Comput. Modelling **18,11**, 1-16.

Zhong C, Miao D, Wanga R, Zhou X. 2008. DIVFRP: An automatic divisive hierarchical clustering method based on the furthest reference points. Pattern Recognition Letters **29**, 2067–2077.

Zhu W, Jiang J, Song C, Bao L. 2011. Clustering Algorithm Based on Fuzzy C-means and Artificial Fish Swarm. Procedia Engineering **29,** 3307–3311.