International Journal of Engineering
Journal Homepage: www.ije.ir
RESEARCH NOTE

Oil Reservoirs Classification Using Fuzzy Clustering

S. Askari*

Mechanical Engineering Department, Amirkabir University of Technology, Tehran, Iran

PAPER INFO

Paper history:
Received 13 February 2017
Received in revised form 18 March 2017
Accepted 07 July 2017

Keywords:
Enhanced Oil Recovery
Oil Reservoirs
Fuzzy C-Means
Fuzzy Clustering
Outlier
Possibilistic C-Means

ABSTRACT

Enhanced Oil Recovery (EOR) is a well-known method to increase oil production from oil reservoirs. Applying EOR to a new reservoir is a costly and time consuming process. Incorporating available knowledge of oil reservoirs in the EOR process eliminates these costs and saves operational time and work. This work presents a universal method to apply EOR to reservoirs based on the available data by clustering the data into compact and well-separated groups. A label is then assigned to each cluster which is in fact class of the data points belonging to that cluster. When EOR is intended to be applied to a new reservoir, class of the reservoir is determined and then EOR method used for the reservoirs of that class is applied to this one with no further field studies and operations. In contrast to classification, clustering is unsupervised and number of clusters within the data is not known a priori. Some well-known methods for determining number of clusters are tried but they failed. A novel method is presented in this work for number of clusters based on difference of membership grades of the data points in the clusters. It is applied to both synthetic and real life data including reservoirs data and it is shown that this method finds number of clusters accurately. It is also shown the raw data could be easily represented as fuzzy rule-base for better understanding and interpretation of the data.

***doi**: 10.5829/idosi.ije.2017.30.09c.12*

1. INTRODUCTION

Enhanced Oil Recovery (EOR) is a technique for augmenting oil production from reservoirs. Three methods are used for EOR namely thermal injection, gas injection, and chemical injection. EOR allows extraction of about 60% of oil of the reservoir compared to the 40% which is usually extracted. Therefore, it is possible to increase oil extraction by 20% using EOR. Gas injection or miscible flooding is the most popular method in EOR by injecting miscible gases into the reservoir [1, 2]. Gas injection retrieves reservoir internal pressure and increases oil displacement by diminution of the tension between water and oil. CO_2, nitrogen, and natural gas are commonly used for this type of EOR. However, CO_2 is the most proper gas for this purpose since it decreases oil viscosity and facilitates its flow through the reservoir. In thermal injection, crude oil is heated to decrease its viscosity and surface tension which increases its permeability and eases its motion through the pores of the reservoir [3]. Chemical injection (i.e. alkaline or surfactants [4-7] like sulfonates, rhamnolipids [8], etc) dilutes the crude oil and increases its mobility by reducing surface tension [9-14].

One of the most important characteristics of carbonate oil reservoirs [15-17] is their natural fracture networks. Oil is mainly stored in these fractured carbonate reservoirs rather than sandstones. Depending on their fracture intensity, reservoirs are divided into three groups of high, medium, and low fracture intensity.

It is possible to increase oil and gas production from old undeveloped fields and matured fields using EOR which allows enhancement of gas and oil production from low productive reservoirs. Increasing either natural gas [18, 19] or oil production is economically very important which is extreme goal of any EOR process. To apply a suitable EOR method including gas injection (either miscible or immiscible), chemical injection, and thermal injection, one should have the full understanding and description of the reservoir rock and fluid. To achieve this goal, clustering techniques have become quite interesting to researchers. Rock and fluid

*Corresponding Author's Email: s_askari@aut.ac.ir (S. Askari)

properties characterize the reservoirs and are important in order to assign an appropriate EOR method to the reservoir. These properties are viscosity, gravity, oil saturation, pressure, temperature, reservoir depth, thickness, porosity, and permeability.

Different methods are used for clustering as a main tool for data mining. Hard clustering methods i.e. K-Means algorithm are based on crisp logics which leads to strict clustering of the data. A data vector is just in one cluster in hard clustering methods. A paradigm shift happened by the presentation of fuzzy sets by which binary logics is replaced with multi-valued logics. New clustering methods are introduced with the advent of fuzzy sets. Fuzzy clustering methods i.e. Fuzzy C-Means algorithm are soft in which each data vector belongs to all clusters to some degree. Extreme points of fuzzy sets are traditional crisp sets. This study employes fuzzy clustering methods to apply EOR to oil reservoirs.

Fuzzy C-Means (FCM) is the basic fuzzy clustering algorithm which is widely used in the literature and is developed for different purposes [20, 21]. The following objective function is used in FCM. Minimizing this function yields partition matrix and cluster centers [20].

$$J_{FCM} = \sum_{j=1}^{n}\sum_{i=1}^{c} u_{ij}^{m}\left\|\vec{x}_j - \vec{v}_i\right\|_A^2, \sum_{i=1}^{c} u_{ij} = 1 \quad (1)$$

where, n is number of observations, c is number of clusters, $\vec{v}_i$ is i^{th} cluster center, r is number of variables, $\vec{x}_j$ is j^{th} observation, u_{ij} is membership grade of j^{th} observation in i^{th} cluster, $\left\|\vec{x}_j - \vec{v}_i\right\|_A^2 = (\vec{x}_j - \vec{v}_i)^T A(\vec{x}_j - \vec{v}_i)$ is distance, $m > 1$ is degree of fuzziness, and $A_{r\times r}$ is covariance norm matrix defined as:

$$A = \left(\frac{1}{n}\sum_{j=1}^{n}(\vec{x}_j - \bar{\vec{v}})(\vec{x}_j - \bar{\vec{v}})^T\right)^{-1}, \bar{\vec{v}} = \frac{1}{n}\sum_{j=1}^{n}\vec{x}_j \quad (2)$$

The following cluster centers matrix $V_{r\times c}$ and partition matrix $U_{c\times n}$ minimize Equation (1) [20].

$$u_{ij} = \left[\sum_{k=1}^{c}\left(\frac{\left\|\vec{x}_j - \vec{v}_i\right\|_A^2}{\left\|\vec{x}_j - \vec{v}_k\right\|_A^2}\right)^{\frac{1}{m-1}}\right]^{-1}, \vec{v}_i = \frac{\sum_{j=1}^{n} u_{ij}^{m}\vec{x}_j}{\sum_{j=1}^{n} u_{ij}^{m}} \quad (3)$$

The main drawback with FCM is sensitivity of the cluster centers to noise and outliers. It is well-known that Possibilistic C-Means (PCM) is capable of handling data with outliers [20], but PCM itself has two main problems, coincident clusters and sensitivity to initialization. A combination of FCM and PCM namely Possibilistic Fuzzy C-Means (PFCM) is presented which has none of the above shortcomings and efficiently clusters datasets with outliers. Objective function of PFCM algorithm is as follows [20].

$$J_{PFCM} = \sum_{j=1}^{n}\sum_{i=1}^{c}\left(u_{ij}^{m} + t_{ij}^{\eta}\right)\left\|\vec{x}_j - \vec{v}_i\right\|_A^2 + \sum_{i=1}^{c}\gamma_i\sum_{j=1}^{n}(1 - t_{ij})^{\eta}, \sum_{i=1}^{c} u_{ij} = 1 \quad (4)$$

where, t_{ij} is typicality and $\eta > 1$. The following $\vec{v}_i$, u_{ij}, and t_{ij} minimize J_{PFCM} [20].

Since there are some outliers in oil reservoir dataset, PFCM algorithm is preferred in this work to calculated cluster centers insensitive to the outliers.

$$\vec{v}_i = \frac{\sum_{j=1}^{n}\left(u_{ij}^{m} + t_{ij}^{\eta}\right)\vec{x}_j}{\sum_{j=1}^{n}\left(u_{ij}^{m} + t_{ij}^{\eta}\right)}, u_{ij} = \left[\sum_{k=1}^{c}\left(\frac{\left\|\vec{x}_j - \vec{v}_i\right\|_A^2}{\left\|\vec{x}_j - \vec{v}_k\right\|_A^2}\right)^{\frac{1}{m-1}}\right]^{-1}$$
$$t_{ij} = \left(1 + \left(\frac{\left\|\vec{x}_j - \vec{v}_i\right\|_A^2}{\gamma_i}\right)^{\frac{1}{\eta-1}}\right)^{-1}, \gamma_i = \frac{\sum_{j=1}^{n} u_{ij}^{m}\left\|\vec{x}_j - \vec{v}_i\right\|_A^2}{\sum_{j=1}^{n} u_{ij}^{m}} \quad (5)$$

Fuzzy clustering is widely used for different problems such as fuzzy time series [22-24], structure identification of fuzzy systems [25], etc. Increasing application of fuzzy clustering in different fields proves its superiority over its crisp counterparts (hard clustering i.e. K-Means). Objective of the present work is to apply fuzzy clustering for knowledge extraction from oil reservoirs raw data to assign proper Oil Enhanced Recovery method to increase oil production from the oil fields which significantly reduces costs and operational time. There are two main problems with these data including outliers and unknown number of clusters. It is shown that fuzzy clustering is able to efficiently handle outliers. Moreover, a new method is proposed to determine number of clusters in a given dataset and it is then applied to reservoir data and two other datasets to show its accuracy. Finally, a universal method is presented to handle any given dataset with outliers and unknown number of clusters as well as the reservoir data.

2. DATASET DESCRIPTION

There are 151 different reservoirs with nine variables including Depth, Thickness, Permeability, Pressure, Temperature, Saturation, Viscosity, Gravity, and Porosity. These variables characterize the reservoirs and are measured through wide-range field studies. Type of

EOR method to be applied to a reservoir depends on the values of these variables. The reservoirs are grouped into similar clusters based on these variables. For a given reservoir, it is determined that to which cluster it belongs. The EOR method applied to at least one representative reservoir of that cluster is then used for this reservoir and there is no need for further field studies or operations which significantly reduces EOR projects cost and time.

3. DATA CLUSTERING

There are some outliers in the data which influence the cluster centers as shown in Figure 1. The ability of PFCM algorithm to handle noisy data is a common knowledge and is repeatedly circulated in the literature with numerous applications. We use PFCM algorithm for clustering the data where possibilistic term is supposed to damp impacts of the outliers on the cluster centers.

As discussed earlier, one of the main drawbacks with PCM algorithm is coincident cluster centers. PFCM algorithm as a combination of FCM and PCM algorithms suffers from the same problem. When the data are clustered into three clusters using PFCM, coincident clusters result as shown in Figure 2 where cluster centers are indicated by *. PFCM calculates the following cluster centers V_{PFCM} . It is because of possibilistic terms of PFCM algorithm inherited from PCM algorithm which tend to coincident clusters. Distinct clusters could be found by dropping these possibilistic terms of PFCM which yields FCM algorithm. Results of clustering the data using FCM algorithm is shown in Figure 3 where three distinct cluster centers are observed but cluster centers are displaced because of the outliers. The following cluster centers are computed by FCM algorithm.

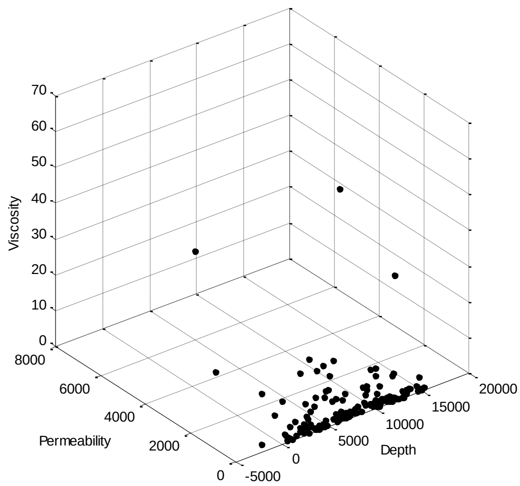

Figure 1. Reservoir data

$$V_{PFCM} = \begin{bmatrix} 7316.60 & 7319.88 & 7323.84 \\ 995.86 & 995.45 & 995.06 \\ 132.53 & 132.53 & 132.54 \\ 4101.92 & 4103.04 & 4104.61 \\ 197.41 & 197.44 & 197.48 \\ 69.81 & 69.81 & 69.81 \\ 2.56 & 2.56 & 2.56 \\ 31.24 & 31.24 & 31.24 \\ 11.85 & 11.85 & 11.85 \end{bmatrix}$$

$$V_{FCM} = \begin{bmatrix} 2439.66 & 7188.91 & 11888.88 \\ 1311.30 & 890.45 & 787.32 \\ 177.09 & 272.17 & 47.481 \\ 2475.46 & 3588.144 & 6357.37 \\ 150.91 & 194.02 & 241.87 \\ 67.09 & 69.55 & 70.97 \\ 2.72 & 3.95 & 4.04 \\ 34.57 & 29.45 & 31.39 \\ 9.57 & 14.44 & 10.98 \end{bmatrix}$$

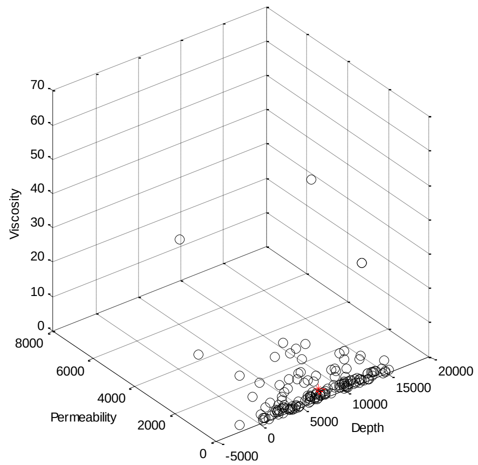

Figure 2. Cluster centers of the data computed by PFCM algorithm

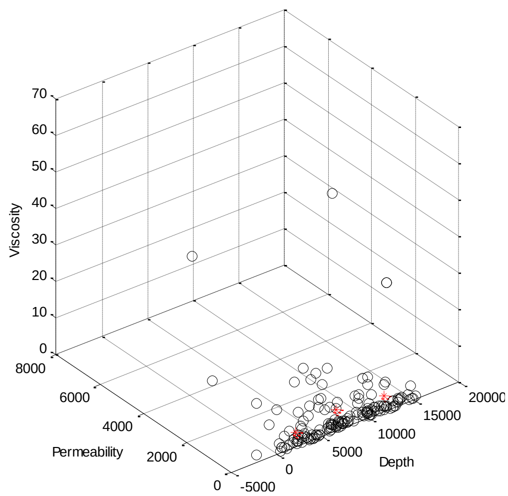

Figure 3. Cluster centers of the data computed by FCM algorithm

Therefore none of the FCM and PFCM algorithms is capable of damping outliers' impacts on the cluster centers and cluster centers are still displaced towards the outliers.

Recently, a clustering algorithm called Generalized Entropy based Possibilistic Fuzzy C-Means (GEPFCM) is presented for noisy data [26, 27]. This algorithm initializes by FCM algorithm.

Since FCM does not produce coincident clusters, it is expected that this algorithm is capable of handling these data by computing distinct clusters insensitive to outliers. Index of GEPFCM is defined as [27]:

$$J_{GEPFCM} = \sum_{j=1}^{n}\sum_{i=1}^{c}\Big(u_{ij}^{m} f_{i,FCM}\left(\|\vec{x}_j - \vec{v}_i\|_A^2\right) + t_{ij}^{\eta} f_{i,PCM}\left(\|\vec{x}_j - \vec{v}_i\|_A^2\right) + c_E s_{ij} f_{i,E}\left(\|\vec{x}_j - \vec{v}_i\|_A^2\right)\Big) + \sum_{j=1}^{n}\sum_{i=1}^{c} s_{ij}\ln(s_{ij}) + \sum_{i=1}^{c}\gamma_i \sum_{j=1}^{n}(1-t_{ij})^{\eta}, \sum_{i=1}^{c} u_{ij} = 1 \quad (6)$$

where, $f_{i,FCM}$, $f_{i,PCM}$, and $f_{i,E}$ are some functions related to fuzzy, possibilistic, and entropy terms, respectively and s_{ij} indicates entropy. c_E is a weighting coefficient associated to entropy and c_{FCM} and c_{PCM} are constants.

We drop possibilistic and entropy terms of the index and just use the fuzzy term which yields Generalize Fuzzy C-Means (GFCM). We also use f_i instead of $f_{i,FCM}$ for simplicity. In fact, we set $c_{FCM} = 1, c_{PCM} = 0, c_E = 0$ which yields a special case of GEPFCM algorithm called GFCM and is capable of handling data with outliers as well as the GEPFCM algorithm itself. Therefore objective function of GFCM is:

$$J_{GFCM} = \sum_{j=1}^{n}\sum_{i=1}^{c} u_{ij}^{m} f_i\left(\|\vec{x}_j - \vec{v}_i\|_A^2\right), \sum_{i=1}^{c} u_{ij} = 1 \quad (7)$$

Using Lagrange Multipliers method for this constrained optimization problem yields

$$J = \sum_{j=1}^{n}\sum_{i=1}^{c} u_{ij}^{m} f_i\left(\|\vec{x}_j - \vec{v}_i\|_A^2\right) + \sum_{j=1}^{n}\lambda_j\left(\sum_{i=1}^{c} u_{ij} - 1\right)$$

Zeroing derivatives of J with respect to $\vec{v}_i$ and u_{ij} gives clusters centers of the data and partition matrix.

$$\begin{cases} \dfrac{\partial J}{\partial \vec{v}_i} = \sum_{j=1}^{n} u_{ij}^{m} f_i'\left(\|\vec{x}_j - \vec{v}_i\|_A^2\right)\left(A + A^T\right)\left(\vec{x}_j - \vec{v}_i\right) = 0 \\ \dfrac{\partial J}{\partial u_{ij}} = m u_{ij}^{m-1} f_i\left(\|\vec{x}_j - \vec{v}_i\|_A^2\right) + \lambda_j = 0, \sum_{k=1}^{c} u_{kj} = 1 \end{cases} \Rightarrow$$

$$u_{ij} = \left[\sum_{k=1}^{c}\left(\frac{f_i\left(\|\vec{x}_j - \vec{v}_i\|_A^2\right)}{f_k\left(\|\vec{x}_j - \vec{v}_k\|_A^2\right)}\right)^{\frac{1}{m-1}}\right]^{-1}$$

$$\vec{v}_i^{k+1} = \frac{\sum_{j=1}^{n} u_{ij}^{m} f_i'\left(\|\vec{x}_j - \vec{v}_i^{k}\|_A^2\right)\vec{x}_j}{\sum_{j=1}^{n} u_{ij}^{m} f_i'\left(\|\vec{x}_j - \vec{v}_i^{k}\|_A^2\right)} \quad (8)$$

where, f_i is defined as:

$$f_i\left(\|\vec{x}_j - \vec{v}_i\|_A^2\right) = 1 - \exp\left(-\frac{\rho\|\vec{x}_j - \vec{v}_i\|_A^2}{R_i^2}\right)$$

$$R_i^2 = \frac{\sum_{j=1}^{n} u_{ij}^{m}\left(\vec{x}_j - \vec{v}_i\right)^T\left(\vec{x}_j - \vec{v}_i\right)}{\sum_{j=1}^{n} u_{ij}^{m}}$$

where, ρ is the key parameter of GFCM which is calculated by minimizing the following index.

$$J_\rho = \sum_{j=1}^{n}\sum_{i=1}^{c} s_{ij}\|\vec{x}_j - \vec{v}_i\|_A^2$$
$$s_{ij} = \exp\left(-1 - \|\vec{x}_j - \vec{v}_i\|_A^2\right) \quad (9)$$

One simple way of finding optimal ρ is to compute J_ρ for different values of ρ and then choose the ρ corresponding to the minimum of J_ρ. This method is used in this work which gives $\rho = 49.25$.

Results of clustering the data using GFCM algorithm are shown in Figure 4. Cluster centers computed by GFCM algorithm are:

$$V_{GFCM} = \begin{bmatrix} 2378.07 & 7071.57 & 11736.28 \\ 1103.56 & 943.87 & 1328.59 \\ 1.46 & 24.60 & 5.58 \\ 2475.15 & 3663.50 & 6294.02 \\ 150.44 & 189.91 & 252.75 \\ 74.66 & 54.32 & 57.55 \\ 1.30 & 1.23 & 0.50 \\ 32.01 & 28.25 & 32.51 \\ 14.65 & 11.05 & 13.97 \end{bmatrix}$$

4. NUMBER OF CLUSTERS

Clustering is an unsupervised procedure and number of clusters is not known. Cluster Validity Index (CVI) [27] is usually used to determine number of clusters.

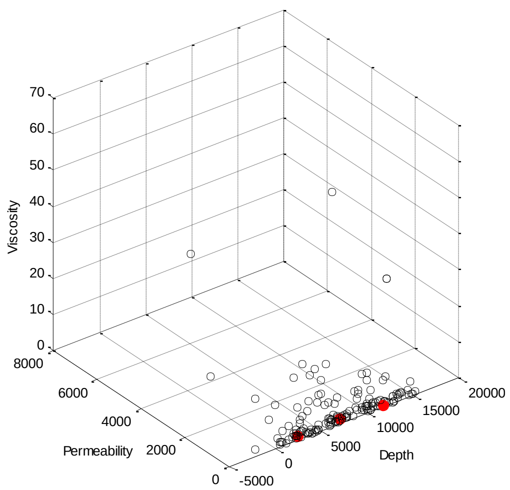

Figure 4. Clustering the data using GFCM algorithm

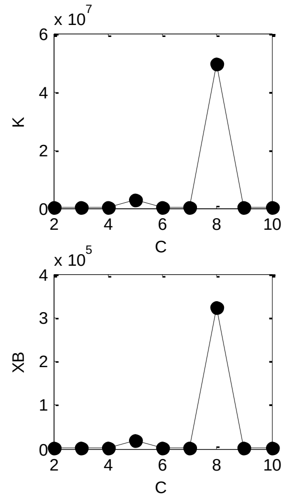

Figure 5. Xie-Beni and Kwon indices for oil reservoir data

The most popular CVIs are Xie-Beni [27] and Kwon [27] CVIs which are widely used in the literature. These indices are ratios of compactness to separation. Minimizing these indices maximizes compactness and separation. These CVIs are as follows where $\bar{\vec{v}}$ is mean of the cluster centers. Optimal number of clusters is the one that minimizes either of these indices. We cluster the reservoir data for different numbers of clusters and compute Xie-Beni and Kwon indices to find number of clusters corresponding to the minimum of these indices. Results of these computations are shown in Figure 5. It is observed that graphs are generally similar to each other and definite number of clusters can be deduced from none of them since the indices are minimum in several numbers of clusters.

$$XB = \frac{\sum_{j=1}^{n}\sum_{i=1}^{c} u_{ij}^{m}\left\|\vec{x}_j - \vec{v}_i\right\|_A^2}{n \min\limits_{\substack{i,j=1\\ i\neq j}}^{c}\left\{\left\|\vec{v}_i - \vec{v}_j\right\|_A^2\right\}}$$
$$K = \frac{\sum_{j=1}^{n}\sum_{i=1}^{c} u_{ij}^{m}\left\|\vec{x}_j - \vec{v}_i\right\|_A^2 + \frac{1}{c}\sum_{i=1}^{c}\left\|\vec{v}_i - \bar{\vec{v}}\right\|_A^2}{\min\limits_{\substack{i,j=1\\ i\neq j}}^{c}\left\{\left\|\vec{v}_i - \vec{v}_j\right\|_A^2\right\}} \tag{10}$$

A novel index is proposed here to determine optimal number of clusters in a given dataset using the difference between membership functions of the clusters.

Assume there are only two clusters with membership functions $u_{1j}, u_{2j}\ j \in [1,n]$ related to cluster centers $\vec{v}_1$ and $\vec{v}_2$ as depicted in Figure 6. If cluster centers are well-separated, the difference between their corresponding membership functions will be significant but if they are close to each other or coincident, this difference will be insignificant.

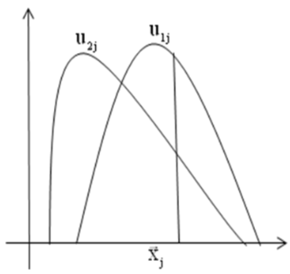

Figure 6. Separation of two clusters via membership functions

Therefore, the average difference between these membership functions is a measure of their corresponding clusters separation which is:

$$S = \frac{1}{2}\sum_{j=1}^{n}\left(u_{1j} - u_{2j}\right)^2 \tag{11}$$

Higher value of S indicates higher separation of the clusters. If there are c clusters, the difference between membership functions of each pair of the clusters should be considered as the separation measure. It is represented as follows which is maximized for optimal number of clusters.

$$S = \frac{1}{c}\sum_{k=1}^{c}\sum_{i=1}^{c}\sum_{j=1}^{n}\left(u_{kj} - u_{ij}\right)^2 \tag{12}$$

The proposed measure of separation is shown in Figure 7 for different numbers of clusters which suggests three clusters. Therefore, the oil reservoir data are grouped into three clusters according to this measure. These three clusters are visually detectable in Figure 6 which confirms accuracy of the proposed index. It is surprising to note that the three clusters identified by the above index are confirmed by the common knowledge of experts of oil industry mentioned in the introduction as reservoirs high, medium, and low fracture intensity.

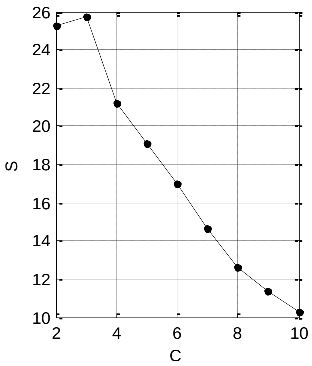

Figure 7. Clusters separation measure versus number of clusters for reservoir data

A question arises here that any process could be divided into three clusters with linguistic labels high, medium, and low as the reservoir data. Why three clusters are chosen? In fact, the data can be divided into many clusters. For example

1. Two clusters: low and high.
2. Three clusters: low, medium, and high.
3. Four clusters: low, medium, high, and very high.
4. Five clusters: very low, low, medium, high, and very high.
5. etc

So, the data could be clustered into any of the above groups and one does not know which of them is true. The role of the proposed index in Equation (12) is to determine what number of clusters is true. As shown in Figure 7, these data are grouped into three clusters. The proposed measure outperforms Xie-Beni and Kwon indices which are the most popular indices in the literature and unable to determine number of clusters in these data. This measure is applied to two datasets illustrated in Figure 8. Figure 8 (a) shows a synthetic dataset with six distinct clusters. Figure 8 (b) shows IRIS data which are well-known and widely used in the literature. These data contain some information about three different types of flowers with four variables. As shown in the figure, two of the clusters overlap but the third one is distant from the two. The separation measure Equation (12) is computed for these datasets and shown in Figure 9. It is observed that number of clusters is identified correctly for both datasets.

5. A UNIVERSAL METHOD

The method applied to the data to find optimal number of clusters and then group the reservoirs into similar clusters could be presented in a universal form as shown in Figure 10 where, $X_{r\times n}$ is the data, r is number of variables, n is number of observations, c_m is maximum number of clusters used for computing optimal number of clusters, and ε is convergence criterion (in this work $\varepsilon = 0.00001$). The algorithm computes r and n from the X matrix. This flowchart first decides on the number of clusters c_{opt} by maximizing the index S given in Equation (12). Having number of clusters, then it clusters the data into similar groups and computes cluster center matrix V and partition matrix U and then terminates. If the data are noisy, GFCM algorithm is used and if they are not noisy, FCM algorithm is employed. For FCM algorithm, $f_i\left(\left\|\vec{x}_j-\vec{v}_i\right\|_A^2\right)=\left\|\vec{x}_j-\vec{v}_i\right\|_A^2, f_i'\left(\left\|\vec{x}_j-\vec{v}_i\right\|_A^2\right)=1$.

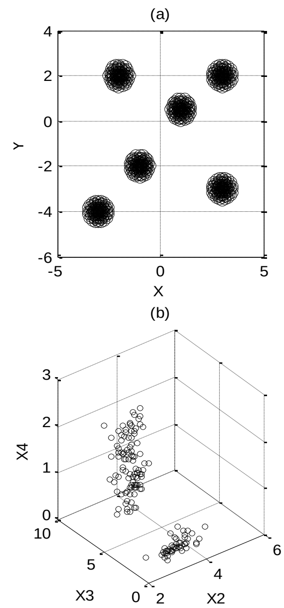

Figure 8. (a) Synthetic dataset with six clusters and (b) IRIS data with three clusters

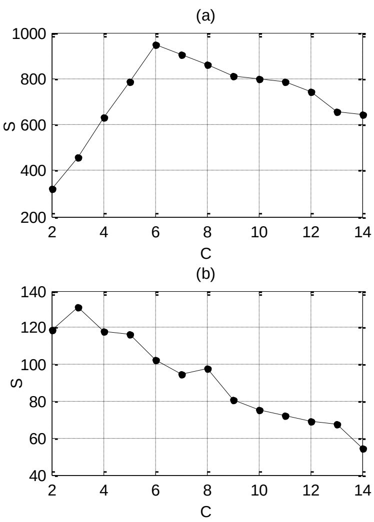

Figure 9. Separation measure for (a) synthetic dataset and (b) IRIS data

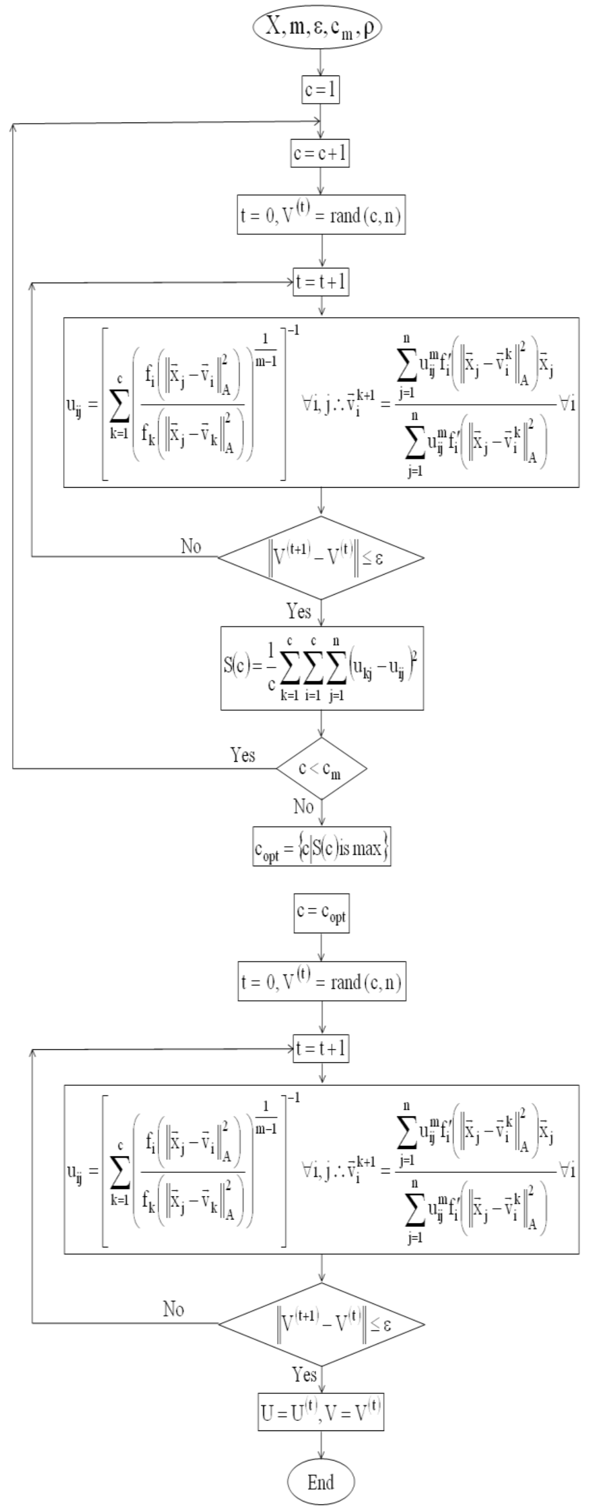

Figure 10. Flowchart of the universal method

6. DATA ANALYSIS

As discussed earlier, the reservoir data contains three well-separated and compact clusters as confirmed by the index S given in Equation (12) and shown in Figures 4 and 7. Values of each of the nine variables determining nature of reservoirs in these clusters are given in Table 1. Each color in this table represents a linguistic (fuzzy) concept. Blue indicates Low, Green indicates Medium, and Red indicates High. $\vec{v}_1$, $\vec{v}_2$, and $\vec{v}_3$ are centers of the fuzzy clusters. Therefore, characteristics of the reservoirs are interpretable as a fuzzy IF-THEN rule-base with three rules as follows:

Rule 1: IF Depth is Low and Thickness is Medium and Permeability is Low and Pressure is Low and Temperature is Low and Saturation is High and Viscosity is High and Gravity is Medium and Porosity is High THEN the Reservoir is **Cluster 1**.

Rule 2: IF Depth is Medium and Thickness is Low and Permeability is High and Pressure is Medium and Temperature is Medium and Saturation is Low and Viscosity is Medium and Gravity is Low and Porosity is Low THEN the Reservoir is **Cluster 2**.

Rule 3: IF Depth is High and Thickness is High and Permeability is Medium and Pressure is High and Temperature is High and Saturation is Medium and Viscosity is Low and Gravity is High and Porosity is Medium THEN the Reservoir is **Cluster 3**.

Therefore, if these nine parameters are known for a new reservoir, it is determined that the reservoir matches what cluster. Enhanced Oil Recovery process is then applied to this reservoir as it is applied to any of the reservoirs belongs to that cluster. So, the existing knowledge of the reservoirs is used for the new reservoir which makes furthers field studies unnecessary and results considerable financial and time savings. The proposed method is not limited to the present reservoirs and is easily applied to the reservoirs of any region as discussed in preceding section and shown in Figure 10. The knowledge extracted from the raw data of the reservoirs and recapitulated as three fuzzy IF-THEN rules can be transferred to standard fuzzy rule-base in terms of membership functions as shown in Figure 11 where output of each rule is designated by a relevant cluster center. The following membership functions are used for the reservoir dataset.

$$\mu_{is}\left(x_{sj}\right) = exp\left(-\left(\frac{x_{sj}-v_{si}}{\sigma_{si}}\right)^2\right) \quad (13)$$
$$i \in [1,c], s \in [1,r], j \in [1,n]$$

where, μ_{is} is membership function of s^{th} variable in the i^{th} cluster, x_{sj} is the entry of s^{th} row and j^{th} column of X matrix, v_{si} is entry of s^{th} row and i^{th} column of cluster centers matrix V. σ_{si} s are computed from the partition matrix U and cluster centers matrix V as follows.

$$\sigma_{si}^{2}=\frac{\sum_{j=1}^{n}u_{ij}^{m}\left(x_{sj}-v_{si}\right)^{2}}{\sum_{j=1}^{n}u_{ij}^{m}},i\in\left[1,c\right],s\in\left[1,r\right] \tag{14}$$

TABLE 1. Values of variables in each cluster and their linguistic labels. The colors Blue, Green, and Red indicate Low, Medium, and High, respectively

Variable	Cluster 1	Cluster 2	Cluster 3
Depth	2378.07	7071.57	11736.28
Thickness	1103.56	943.87	1328.59
Permeability	1.46	24.60	5.58
Pressure	2475.15	3663.50	6294.02
Temperature	150.44	189.91	252.75
Saturation	74.66	54.32	57.55
Viscosity	1.30	1.23	0.50
Gravity	32.01	28.25	32.51
Porosity	14.65	11.05	13.97
Cluster Center	$\vec{v}_1$	$\vec{v}_2$	$\vec{v}_3$

Each reservoir is assigned to a cluster using either the above rule-base or the partition matrix U . Consider the partition matrix U given in Equation (3). The j^{th} data vector $\vec{x}_j$ belongs to the cluster in which it attains the maximum membership grade. So, this data vector is assigned to a cluster as follows.

$$C(j)=\{l|U(l,j)\geq U(i,j)\,\forall i\in[1,c]\} \tag{15}$$

This observation is then fuzzified to $[C(j)]^{th}$ cluster and its characteristics (variables) are most similar to those of $V_{C(j)}$. Each of the 151 data vectors are fuzzified using this approach. Numbers of data vectors assigned to each cluster using Equation (15) are as Cluster 1: 51; Cluster 2: 41; Cluster 3: 59.

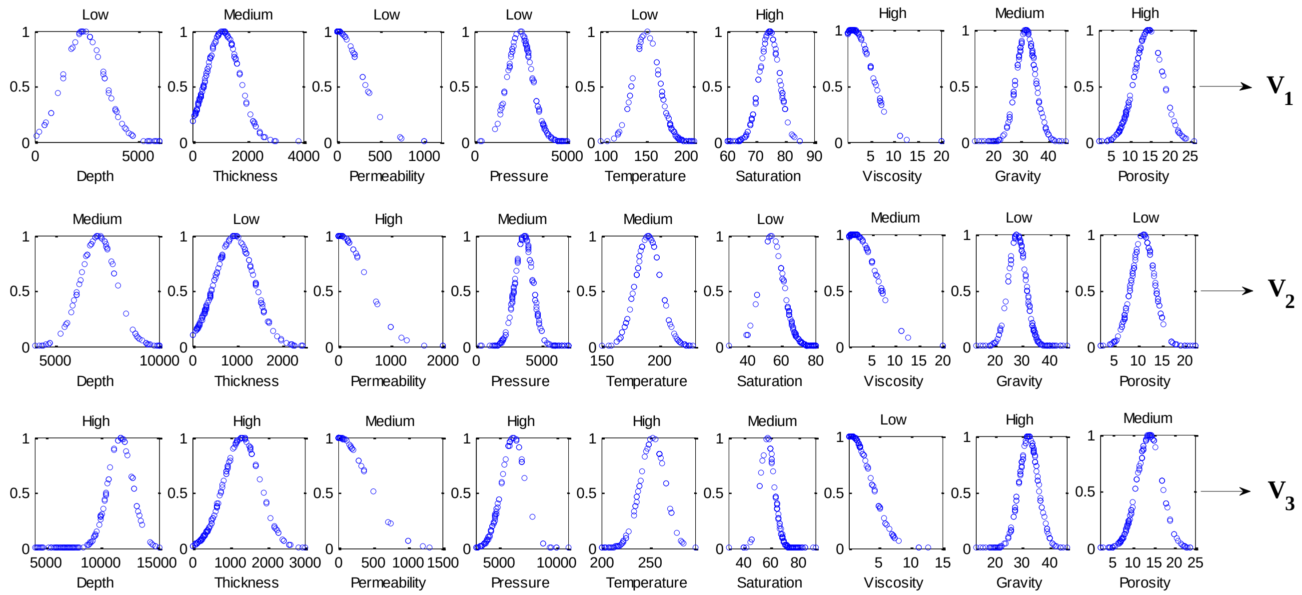

Figure 11. Rule-base of the fuzzy system developed for reservoir dataset

7. CONCLUSION

This work employs fuzzy clustering for knowledge extraction from oil reservoirs raw data with outliers and unknown number of clusters and then generalizes the method to any given dataset by presenting a universal method. Possibilistic Fuzzy C-Means (PFCM) algorithm is used to cluster the data because there are some outliers in the data and possibilistic terms of PFCM algorithm are supposed to cancel outliers impacts on the cluster centers. However, PFCM yields three coincident clusters because of these possibilistic terms. The data are then clustered using Fuzzy C-Means (FCM) algorithm which yields three distinct clusters. However, these cluster centers are displaced towards the outliers that causes the clusters to mismatch the actual nature of the data. However, these cluster centers are displaced towards the outliers that causes the clusters to mismatch the actual nature of the data. Finally, the recently developed Generalized Entropy based Possibilistic

Fuzzy C-Means (GEPFCM) algorithm is applied to the data to cancel outliers' contributions in determination of cluster centers and it is observed that this algorithm clusters the data satisfactorily. The other problem is that clustering is an unsupervised method and number of clusters is not known a priori. The well-known cluster validity indices including those of Xie-Beni and Kwon are applied to the data to determine number of clusters within the data but both of them fail. A new method is then presented for this purpose and it is shown that it works for both synthetic and real life data. This method is applied to the oil reservoirs data and three clusters are identified which are exactly the same as the number of clusters suggested by common knowledge of the reservoirs experts. A universal method is presented to extract knowledge from the raw data with outliers and unknown number of clusters for any given dataset and the clustering results are translated into a fuzzy rule-base for better interpretability and understandins.

8. REFERENCES

1. Mohsenzadeh, A., Escrochi, M., Afraz, M., Karimi, G., Al-Wahaibi, Y. and Ayatollahi, S., "Non-hydrocarbon gas injection followed by steam–gas co-injection for heavy oil recovery enhancement from fractured carbonate reservoirs", ***Journal of Petroleum Science and Engineering***, Vol. 144, (2016), 121-130.
2. Mokhtari, R., Ashoori, S. and Seyyedattar, M., "Optimizing gas injection in reservoirs with compositional grading: A case study", ***Journal of Petroleum Science and Engineering***, Vol. 120, (2014), 225-238.
3. Kovscek, A., "Emerging challenges and potential futures for thermally enhanced oil recovery", ***Journal of Petroleum Science and Engineering***, Vol. 98, (2012), 130-143.
4. Pakzad, H., Alemzadeh, I. and Kazemii, A., "Encapsulation of peppermint oil with arabic gum-gelatin by complex coacervation method", ***International Journal of Engineering-Transactions B: Applications***, Vol. 26, No. 8, (2013), 807-815.
5. Bakhtiari, L., Rezaie, H., Javadpour, J., Erfan, M. and Shokrgozar, M., "Effect of micellization pH on properties of sphere-like mesoporous hydroxyappatite", ***International Journal of Engineering-Transactions A: Basics***, Vol. 28, No. 7, (2015), 1068.
6. Tanzifi, M., Kolaei, Z.T. and Roushani, M., "Characterization of polypyrrole-hydroxyethylcellulose/tio2 nanocomposite: Thermal properties and afm analysis", ***International Journal of Engineering-Transactions B: Applications***, Vol. 28, No. 5, (2014), 654-662.
7. Morenov, V. and Leusheva, E., "Development of drilling mud solution for drilling in hard rocks", ***International Journal of Engineering, Transactions A: Basics***, Vol. 30, No. 620-626, (2017).
8. Amani, H., "Study of enhanced oil recovery by rhamnolipids in a homogeneous 2d micromodel", ***Journal of Petroleum Science and Engineering***, Vol. 128, (2015), 212-219.
9. Armstrong, R.T. and Wildenschild, D., "Investigating the pore-scale mechanisms of microbial enhanced oil recovery", ***Journal of Petroleum Science and Engineering***, Vol. 94, (2012), 155-164.
10. Baldygin, A., Nobes, D.S. and Mitra, S.K., "Water-alternate-emulsion (wae): A new technique for enhanced oil recovery", ***Journal of Petroleum Science and Engineering***, Vol. 121, (2014), 167-173.
11. Santanna, V., Curbelo, F., Dantas, T.C., Neto, A.D., Albuquerque, H. and Garnica, A., "Microemulsion flooding for enhanced oil recovery", ***Journal of Petroleum Science and Engineering***, Vol. 66, No. 3, (2009), 117-120.
12. Demikhova, I.I., Likhanova, N.V., Perez, J.R.H., Falcon, D.A.L., Olivares-Xometl, O., Berthier, A.E.M. and Lijanova, I.V., "Emulsion flooding for enhanced oil recovery: Filtration model and numerical simulation", ***Journal of Petroleum Science and Engineering***, Vol. 143, (2016), 235-244.
13. Hua, S., Liu, Y., Di, Q., Chen, Y. and Ye, F., "Experimental study of air foam flow in sand pack core for enhanced oil recovery", ***Journal of Petroleum Science and Engineering***, Vol. 135, (2015), 141-145.
14. Cao, R., Yang, H., Sun, W. and Ma, Y.Z., "A new laboratory study on alternate injection of high strength foam and ultra-low interfacial tension foam to enhance oil recovery", ***Journal of Petroleum Science and Engineering***, Vol. 125, (2015), 75-89.
15. Liu, D., Zhong, X., Shi, X., Qi, Y. and Qi, Y., "Enhanced oil recovery from fractured carbonate reservoir using membrane technology", ***Journal of Petroleum Science and Engineering***, Vol. 135, (2015), 10-15.
16. Katterbauer, K., Arango, S., Sun, S. and Hoteit, I., "Enhanced heavy oil recovery for carbonate reservoirs integrating cross-well seismic–a synthetic wafra case study", ***Journal of Petroleum Science and Engineering***, Vol. 134, (2015), 1-13.
17. Jeong, C., Kallivokas, L., Kucukcoban, S., Deng, W. and Fathi, A., "Maximization of wave motion within a hydrocarbon reservoir for wave-based enhanced oil recovery", ***Journal of Petroleum Science and Engineering***, Vol. 129, (2015), 205-220.
18. Askari, S., Montazerin, N. and Zarandi, M.F., "High-frequency modeling of natural gas networks from low-frequency nodal meter readings using time-series disaggregation", ***IEEE Transactions on Industrial Informatics***, Vol. 12, No. 1, (2016), 136-147.
19. Askari, S., Montazerin, N. and Zarandi, M.F., "Gas networks simulation from disaggregation of low frequency nodal gas consumption", ***Energy***, Vol. 112, (2016), 1286-1298.
20. Pal, N.R., Pal, K., Keller, J.M. and Bezdek, J.C., "A possibilistic fuzzy c-means clustering algorithm", ***IEEE transactions on fuzzy systems***, Vol. 13, No. 4, (2005), 517-530.
21. Askari, S., "A novel and fast mimo fuzzy inference system based on a class of fuzzy clustering algorithms with interpretability and complexity analysis", ***Expert Systems with Applications***, Vol. 84, (2017), 301-322.
22. Askari, S. and Montazerin, N., "A high-order multi-variable fuzzy time series forecasting algorithm based on fuzzy clustering", ***Expert Systems with Applications***, Vol. 42, No. 4, (2015), 2121-2135.
23. Chen, S.-M. and Tanuwijaya, K., "Multivariate fuzzy forecasting based on fuzzy time series and automatic clustering techniques", ***Expert Systems with Applications***, Vol. 38, No. 8, (2011), 10594-10605.
24. Askari, S., Montazerin, N. and Zarandi, M.F., "A clustering based forecasting algorithm for multivariable fuzzy time series using linear combinations of independent variables", ***Applied Soft Computing***, Vol. 35, (2015), 151-160.
25. Askari, S., Montazerin, N. and Zarandi, M.F., "Forecasting semi-dynamic response of natural gas networks to nodal gas consumptions using genetic fuzzy systems", ***Energy***, Vol. 83, (2015), 252-266.

26. Askari, S., Montazerin, N., Zarandi, M.F. and Hakimi, E., "Generalized entropy based possibilistic fuzzy c-means for clustering noisy data and its convergence proof", ***Neurocomputing***, Vol. 219, (2017), 186-202.

27. Askari, S., Montazerin, N. and Zarandi, M.F., "Generalized possibilistic fuzzy c-means with novel cluster validity indices for clustering noisy data", ***Applied Soft Computing***, Vol. 53, (2017), 262-283.

Oil Reservoirs Classification Using Fuzzy Clustering

RESEARCH NOTE

S. Askari

Mechanical Engineering Department, Amirkabir University of Technology, Tehran, Iran

P A P E R I N F O

Paper history:
Received 13 February 2017
Received in revised form 18 March 2017
Accepted 07 July 2017

Keywords:
Enhanced Oil Recovery
Oil Reservoirs
Fuzzy C-Means
Fuzzy Clustering
Outlier
Possibilistic C-Means

چکیده

EOR یک روش شناخته شده برای ازدیاد تولید نفت از مخازن است و اعمال آن به یک مخزن جدید زمانبر و پرهزینه است. افزودن اطلاعات موجود مخازن به EOR باعث صرفه جویی در هزینه ها، زمان عملیات، و کار لازم می شود. این مقاله یک روش فراگیر برای اعمال EOR به مخازن براساس خوشه بندی داده های موجود به گروههای فشرده و متمایز ارایه می کند. سپس به هر خوشه یک برچسب اختصاص داده می شود که در واقع رده داده های متعلق به آن خوشه است. وقتی قرار است EOR به یک مخزن جدید اعمال شود، رده آن مخزن تعیین شده و سپس روش EOR ی که برای مخازن آن رده بکار رفته است بدون نیاز به مطالعات و عملیات بیشتر به این مخزن جدید اعمال می شود. برخلاف رده بندی، خوشه بندی بدون سرپرستی است و تعداد خوشه های موجود در داده ها نامعلوم است. تعدادی از روشهای معروف تعیین تعداد خوشه ها در این مقاله مورد استفاده قرار می گیرند اما نمی توانند تعداد خوشه ها در داده های مخازن را پیدا کنند. برای این منظور، یک روش جدید براساس تفاضل درجه عضویت داده ها در خوشه های مختلف ارایه شده و بر داده های فرضی و داده های واقعی مخازن اعمال شده و نشان داده می شود که این روش تعداد خوشه ها را بدرستی پیدا می کند. همچنین نشان داده می شود که برای درک و تفسیر بهتر داده های خام، می توان آنها را بسهولت به پایگاه قوانین فازی تبدیل کرد.

***doi**: 10.5829/idosi.ije.2017.30.09c.12*