INTEGRATION OF GUSTAFSON-KESSEL ALGORITHM AND KOHONEN'S SELF-ORGANIZING MAPS FOR UNSUPERVISED CLUSTERING OF SEISMIC ATTRIBUTES

MEHDI EFTEKHARIFAR[1], M. ALI RIAHI[2] and R. KHARRAT[3]

[1] *Petroleum University of Technology (IFP-School France). mehdieftekharifar@gmail.com*
[2] *Institute of Geophysics, University of Tehran, P.O. Box 14155-6466, Tehran, Iran. mariachi@ut.ac.ir*
[3] *Petroleum University of Technology, Tehran, Iran. kharrat@put.ac.ir*

(Received January 5, 2008; revised version accepted June 8, 2009)

ABSTRACT

Eftekharifar, M., Riahi, M.A. and Kharrat, R., 2009. Integration of Gustafson-Kessel algorithm and Kohonen's self-organizing maps for unsupervised clustering of seismic attributes. *Journal of Seismic Exploration*, 18: 315-328.

The goal of different methods for clustering of seismic attributes has been to analyze the discrimination ability of the chosen set of attributes. Kohonen clustering networks or Kohonen's self organizing maps are well known for cluster analysis (unsupervised learning). This class of algorithms is a set of heuristic procedures that suffers from some major problems.

In this paper we propose the use of an unsupervised method by integrating the fuzzy c-means clustering and Gustafson-Kessel algorithms into the learning rate and updating strategies of the Kohonen clustering network. Using a new fuzzy calibration method, the different attribute classes are calibrated to lithology classes and the appropriate attributes and classes are determined. In this paper we propose a robust clustering algorithm which can be quality-controlled by using fuzzy modeling. That means using the clustering results, the available but limited log data is rebuilt after clustering to see the performance of the clustering technique. Classification and modeling results tested on a real data set show reasonable accuracy when compared to well logs.

KEYWORDS: unsupervised clustering, Kohonen's self-organizing maps (SOM), fuzzy self-organizing maps, fuzzy logic, Gustafson-Kessel algorithm, complex seismic attributes, reservoir characterization, inversion.

0963-0651/09/$5.00

INTRODUCTION

The effectiveness of the interpretive use of seismic attributes depends on the discrimination ability of a set of seismic attributes. For determining different lithologies from seismic data, some attributes may be necessary; a few may be sufficient. This could be determined by experimenting with logical combinations of various mixtures of attributes (Taner, 1997). In cases where well logs or lithological columns are available, feed-forward fully connected neural networks may be trained in a supervised manner. In cases where insufficient well information is available, one must use an unsupervised methodology to cluster the data and check to see if it will have any relation to our experience based interpretation. In our case, the data is the computed set of complex attributes. The user selects a set of attributes to be used in clustering or in classification. There are several unsupervised clustering methods that may be applied to this problem. Thus far, Kohonen's self-organizing maps (Kohonen, 1989) have been applied for this purpose (Taner, 1997).

Fuzzy logic is a very powerful tool, when information regarding the uncertainty of estimation is desired. It deals with membership functions, and each data point plays a role in the classification of the entire dataset. It means despite other crisp classification methods that, even very small probabilities will be taken into account.

Clustering techniques try to assess the interaction among patterns by organizing the patterns into clusters such that patterns of a cluster are more similar to each other than they are to patterns of different clusters (Tsao et al., 1992). Treatments of many classical approaches to this kind of problem have been proposed by several researchers (Kohonen, 1989; Bezdek, 1981; Duda and Hart, 1973; Tou and Gonzalez, 1974; Hartigan, 1975; Dubes and Jain, 1988). Kohonen's algorithm has become timely in recent years because of the widespread interest in the applications of neural network (Pao, 1989). However, Kohonen's self organizing maps or KCN suffer from several problems (Tsao et al. 1992):

1. KCNs are heuristic procedures, therefore termination of clustering procedure is not based on optimizing any model of the process or its data.

2. Weight vectors usually depend on the input data sequence (the order of feeding the data).

3. Different initial conditions will yield different results.

4. Several parameters of the KCN algorithms, like learning rate and the size of the updating neighborhood, must be different from one data set to another to obtain good results.

It is widely accepted that KCN clustering is closely related to the c-Means (CM) algorithms (Lippman, 1987). But the question is that what "closely related" really means. Since CM algorithms are optimization procedures, integration of fuzzy c-means and KCN is one way to address several problems of Kohonen's clustering networks (Tsao et al., 1992). This was first considered by Huntsberger and Ajjimarangsee (1989).

In Fuzzy Kohonen Clustering Network (FKCN) algorithm, the ideas of fuzzy membership values for learning rates, the parallelism of fuzzy c-means, and the structure and update rules of KCNs are used simultaneously. FKCN or F-SOM is self-organizing, since the size of the update neighborhood is automatically adjusted during learning, and FKCN usually terminates in such a way that the fuzzy c-means objective function is approximately minimized (Tsao et al. 1992). FKCN is non-sequential, meaning that it is independent of the sequence in which the data are input to the algorithm. The neighborhood constraint of Kohonen's clustering network can be embedded in the learning rate strategy of FKCN (Tsao et al., 1992).

METHODOLOGY

The inputs to this method are sonic log data from one well and a set of complex seismic attributes obtained from the raw seismic data. We input these attributes to the clustering network, and then based on the clustering results and calibration information, attempt to estimate the log data. Before doing the clustering, the input data should be preprocessed. The results of calibration and consequently of modeling depend greatly on the quality of the well to seismic tie. This implies that extracting a proper wavelet is of crucial importance. In this study, each attribute is chosen to vary in one dimension (in depth along the well). This method of choosing the data has some advantages. First, the data size becomes as small as possible. Therefore, computation time will be greatly reduced. The main seismic attributes used in this study were envelope, first derivative of amplitude, second derivative, quadrature trace and instantaneous frequency.

Before clustering the data, in order to avoid the *curse of dimensionality*, it is important to perform a *Principal Component Analysis* (PCA). PCA is useful for studying multi dimensional data, for example, in reducing the effective dimensionality of a data set. PCA has been used for multi-attribute analysis (Taner, 2002). The PCA process determines an orthogonal set of axes that lie in the directions of largest, second largest, and so on down to the smallest variance in the data volume. PCA is of crucial importance because it allows one to recognize the most valuable part of the data. It avoids overtraining, and decrease computation time by reducing the size of data. PCA determines the directions in which the data has most variations and replaces the axes of present data with new axes which are in the direction of the maximum variance. Then those directions that the data variation has minimum values, can be ignored (usually they keep those axes that contain 95% of the original data). In order to avoid the dominance of some of the attributes, all input data should be normalized so that their mean equals zero and root mean square (RMS) of all the data is equal to one (RMS = 1).

Gustafson-Kessel Algorithm (G-K)

The fuzzy c-means (FCM) algorithm is a fuzzified version of the classical k-means algorithm. The minimized loss functions of the two algorithms are almost identical. The difference is that in FCM, a membership function or degree is introduced (μ) which defines the degree by which a data sample belongs to a specific class. Therefore sum of all membership degrees of a data sample is equal to one.

$$I = \sum_{j=1}^{C} \sum_{i=1}^{N} \mu_{ji}^{\upsilon} \| u(i) - c_j \|^2 \Sigma_j \quad , \tag{1}$$

with

$$\sum_{j=1}^{C} \mu_{ji} = 1 \quad , \tag{2}$$

where j = 1,...,K or C is the number of clusters and i = 1,...,N is the dimension of the input data (number of attributes).

The variable υ controls the fuzziness of the clustering. It means that the smaller υ, the sharper the boundaries of the clusters. One should find a compromise between the number of clusters and the fuzziness parameter υ. Depending on the size of the data, the variable υ can vary but usually values between 1.5-1.8 yield reasonable results.

The Gustafson-Kessel algorithm (Gustafson, 1979; Babuska, 1996; Hoppner et al., 1997) is the extended version of the c-means clustering algorithm. Here each cluster possesses its individual distance measure Σ_j. Furthermore, not only the cluster centers c_j (centroids), but also the norm matrices Σ_j are subject to minimization of the loss function (1).

The distance of the data sample l from cluster center j is given by (Nelles, 2001):

$$D^2_{lj,\Sigma_j}(t) = \| u_{li} - c_{ij} \|^2_{\Sigma_j} = \sum_{i=1}^{N} [u_{li} - c_{ij}(t)]^T \Sigma_j [u_{li} - c_{ij}(t)] \quad , \tag{3}$$

where $l = 1,\ldots,M$ is the number of data points, $j = 1,\ldots,K$ or C is the number of clusters and $i = 1,\ldots,N$ is the dimension of the input data (number of attributes).

The norm matrices can be calculated by (Nelles, 2001):

$$\Sigma_j = F_j^{-1} \quad , \tag{4}$$

where

$$F_j = \sum_{l=1}^{M} \sum_{i=1}^{N} \mu^{\upsilon}_{lj} [u_{li} - c_{ij}(t)][u_{li} - c_{ij}(t)]^T / \sum_{l=1}^{M} \mu^{\upsilon}_{lj} \quad , \tag{5}$$

where

$$\mu_{ij}(t) = (1/D^2_{lj,\Sigma_j}) / \sum_{m=1}^{K} [1/D^2_{lm,\Sigma_j}(t)] \quad , \tag{6}$$

and

$$c_{ij} = \sum_{l=1}^{M} \mu^{\upsilon}_{lj} u_{li} / \sum_{l=1}^{M} \mu^{\upsilon}_{lj} \quad , \tag{7}$$

One difficulty is that the Euclidian distance of the data sample l from cluster center j (3) and consequently the loss function (1) can be reduced simply by making the determinant of Σ_j small. To prevent this effect, the norm matrices or the fuzzy covariance matrices must be normalized. Typically, the determinant of the norm matrices is constrained to a user defined constant:

$$\det(\Sigma_j) = v_j \quad . \tag{8}$$

Therefore the norm matrices are defined as:

$$\Sigma_j = F_j^{-1} \cdot [v_j \det(F_j)]^{1/N} \quad . \tag{9}$$

With N being the dimensionality of the input space (number of attributes, used). By this normalization the volume of the clusters will be restricted to equal v_j. Thus, the Gustafson-Kessel algorithm searches for clusters with given, and equal, volumes if no prior information is available. If prior knowledge about the expected cluster volumes is available, the v_j can be chosen individually for each cluster.

Fuzzy Self Organizing Maps (FSOM)

The unsupervised clustering method used in this paper is a kind of Fuzzy-Self Organizing Map (FSOM) in which the Gustafson-Kessel algorithm is integrated in the learning rate and updating strategies of the Kohonen Self Organizing Maps. Tsao et al. (1992) introduced a FSOM method by combining the FCM clustering in the learning rate and updating strategies of the SOM and showed the superiority of this method with respect compared to the SOM. Also, Hu et al. (2004) show that F-SOM (FCM + SOM) is much more efficient than the SOM and vector quantization in both speed and accuracy.

In this method, eq. (10) is devised for updating the centers of clusters instead of eq. (7)

$$c_{ij}(t+1) = c_{ij}(t) + \sum_{l=1}^{M} \mu_{lj}^{v}(t)\cdot[u_{li} - c_{ij}(t)] \Big/ \sum_{l=1}^{M} \mu_{lj}^{v}(t) \quad . \tag{10}$$

This will solve the problem of label dependency in SOM (in FSOM scheme, during each iteration, the whole data is scanned. In SOM, data samples are fed one by one, in a random manner and if we do clustering n times, we will get n different results. This means that the clustering algorithm depends on the order by which the data samples are fed into the clustering network (label dependency). In addition, we use the G-K algorithm instead of FCM and the performance of these two methods will be compared. In the G-K algorithm, the volume of the clusters are controlled by Σ_j. This algorithm can be summarized as follows:

Step 1. Randomize the initial values of the weight vectors.
Step 2. Input all samples.
Step 3. Calculate the Euclidean distances from each sample u_l to all output eurons (D) using eq. (3).
Step 4. Compute the membership degrees of each data sample to all neurons μ.
Step 5. Adjust the weights of each neuron according to the computed memberships using eq. (10).
Step 6. Determine the stability condition of the clustering network (since the data has been normalized, ϵ should have a small value of about 0.01 to 0.001.

$$\max\{|c_{ij}(t+1) - c_{ij}(t)|\} < \epsilon$$

$$1 \le i \le N$$

$$1 \le j \le K$$

The process of clustering schematically is equivalent to perturbing and fitting a polygon (like the one in Fig. 1) and moving the centers so as to minimize the objective function and obtain a new arrangement which best clusters the data (like the one shown in Fig. 2).

The membership degrees of the data points to each class are shown in Fig. 3. The horizontal axis shows the number of the class or cluster (in this Fig. 16 classes), and the vertical axis shows the number of the data points with increasing depth (about 550 data points). The color bar shows the membership degrees ranging from zero to one (dark blue and dark red represent zero and one respectively). In other words, if you imagine a horizontal line in this Fig. and add all values (membership degrees) up, you will get the result equal to one. This is equivalent to saying that the sum of membership degrees of a data point to different classes is equal to unity.

Fuzzy Calibration

The unsupervised clustering methods seek an organization in the dataset and form relational organized clusters. However, these clusters may or may not have any physical analogues in the real world. In order to relate these clusters to the real world, we have to develop some form of calibration method that not only defines the relationship between the clusters and real world physical properties, but also provides us with an estimate of the validity of these relationships. With the development of a calibrated relationship, the whole dataset can be classified. Taner (2000) proposed a method for doing such a calibration. A fuzzy calibration procedure is proposed in order to quality-control the clustering performance. To do such a calibration, the log data should be classified using fuzzy logic as well. This kind of calibration is done using the IF-THEN fuzzy rules and Mamdani Implications (Wang, 1996). Mamdani implications are the most widely used implications applied in the fuzzy systems and fuzzy control.

For *connective* "and" we use fuzzy intersections. We let x and y be linguistic variables in the physical domains U and V and, A and B be fuzzy sets in U and V, respectively. Then the compound fuzzy proposition

x is A and y is B

is interpreted as the fuzzy relation $A \cap B$ in $U \times V$ with membership function

$$\mu_{A \cap B}(x,y) = t[\mu_A(x),\mu_B(y)] \quad , \tag{11}$$

where $t:[0,1]\times[0,1]\rightarrow[0,1]$ is any t-norm.

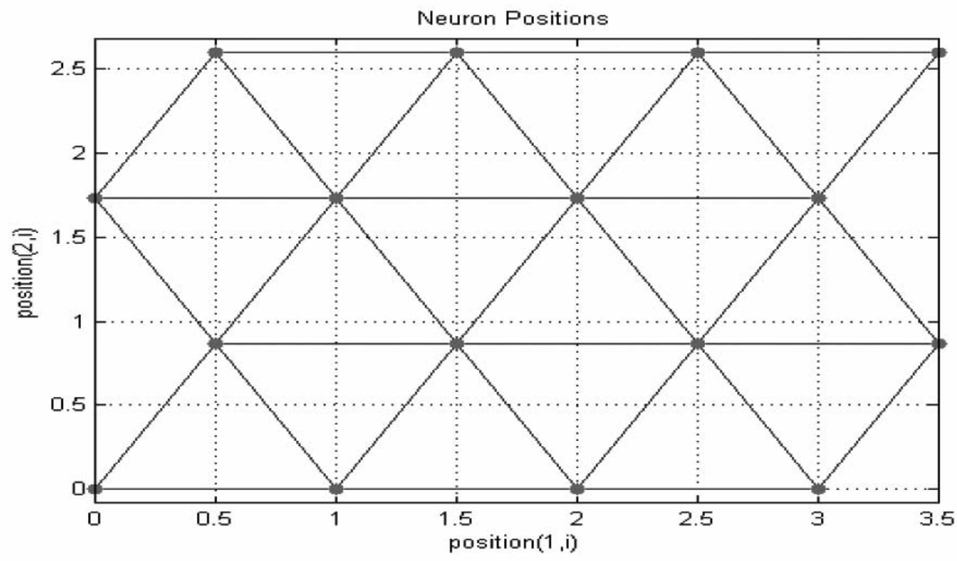

Fig. 1. The network to be fitted to the data.

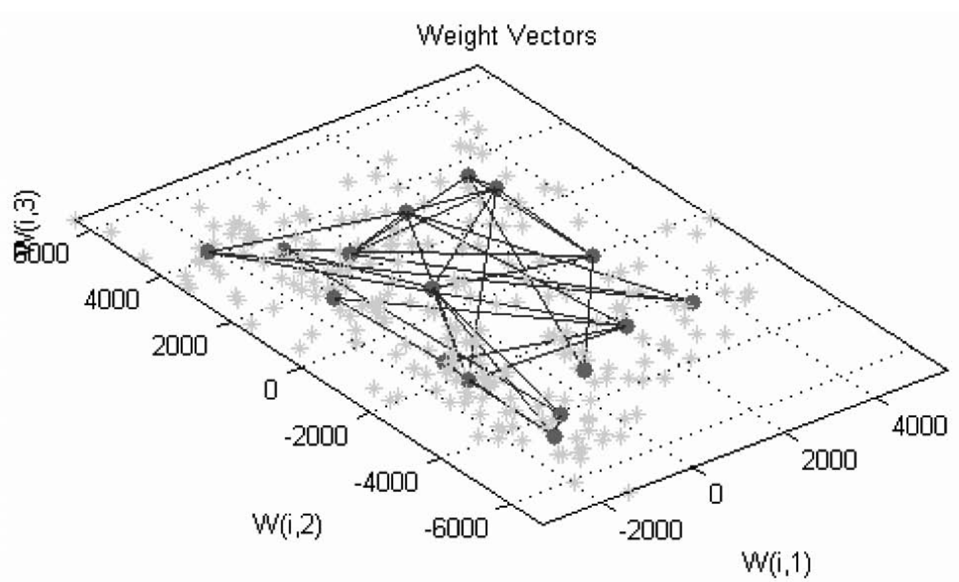

Fig. 2. The network after training.

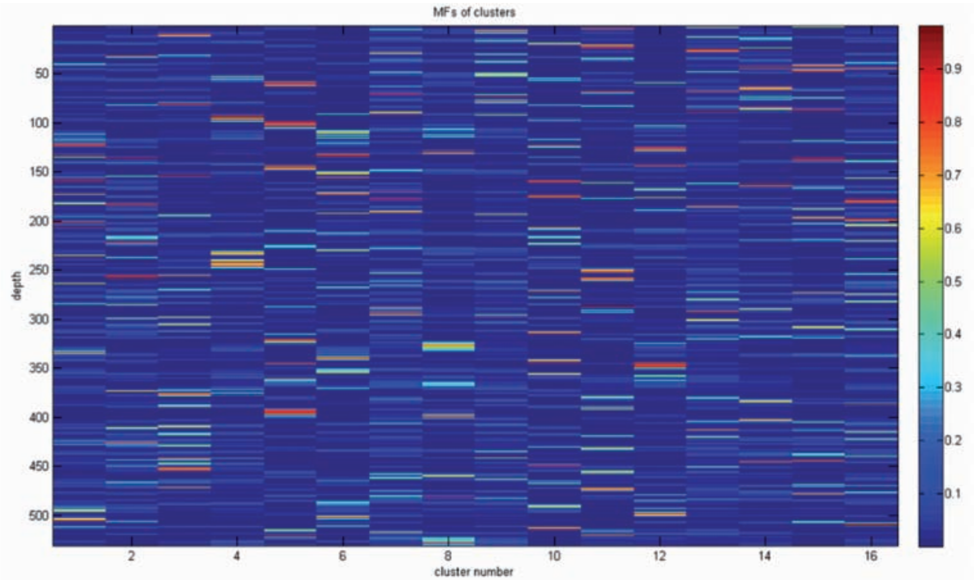

Fig. 3. Membership degrees which indicate the degree by which the data points (each data point is a set of attributes) belong to each class (The horizontal axis shows the number of the class or cluster and the vertical axis shows the number of the data points with increasing depth).

If we use the algebraic product for the t-norm in (11), then the fuzzy proposition:

P: IF x is A THEN y is B

is a fuzzy relation in U × V with the membership function:

$$\mu_P = \mu_A \times \mu_B \quad . \tag{12}$$

In terms of geophysics, this rule can be expressed as: *IF the attributes set belongs to the class A THEN the log value belongs to the class B*. Therefore, according to eqs. (11) and (12), the greater value of μ_P means the greater similarity between A and B. To make it more clear, suppose we want to cluster our data set (seismic attributes) into 16 different classes. In order to do the calibration we should classify the desired output data (log data or lithology) into the same number of classes as seismic attributes. We are trying to find out which seismic attribute class corresponds to which log value or lithology class.

In order to do this we multiply the membership degrees of input data points to one specific class with the membership degrees of output data points to all classes. In other word, if we have for example 500 data points (500 vectors that each of them has a length equal to the number of the attributes we are using), and we want to have 16 clusters, after fuzzy clustering we will have a matrix of the size 500×16. We will have the same matrix for output or log data. That means we will have 16 log value classes for 500 log value data points. To better understand these matrices, take an element a(280,13) from the clustered input data. This element shows the degree by which the 280-th input data sample which is a vector of size 5 for example, (5 is the number of the attributes), belongs to the 13-th cluster. We have the same definition for the matrix of the log values after clustering. Now, to find the best candidate among 16 log class values which corresponds to the first class of attributes, we multiply the column one a(i,1) of the input matrix with all 16 columns of the output or log matrix b(i,1),b(I,2),...b(I,16), (element by element). Therefore we get 16 columns. If we add up all elements in one column, we will get 16 numbers. The biggest number among these 16 numbers shows the greatest similarity between the first class of input data with the corresponding output class. That means if for example we get the greatest value by multiplying the first column of input matrix with the 7-th column of the output matrix, then the first class of the attributes is equivalent to the 7th class in the log classes. We do the same procedure for the second, third,... input classes.

Having the calibration results (fuzzy rules), and using the membership functions obtained by F-SOM clustering, one can do fuzzy modeling. This can be thought of as a weighted averaging operation, or Kriging, in geostatistics in which the unknown value is estimated using all other data points. Every log sample is calculated by addition of the log class values, which are weighted by their corresponding membership degrees. In fact this can be considered as QC for this method where we rebuild the well log by using clusters we made. The result of modeling, using the F-SOM (FCM+SOM) is shown in Fig. 4, and the result obtained by F-SOM (G-K+SOM) is shown in Fig. 5, where the modeled log (blue) and the actual log (red) are compared. The process is depicted in flowchart form in Fig. 5. The correlation between real (actual) log values and the modeled log values (calculated by weighted averaging) indicate that clustering by G-K+SOM yields better results than FCM+SOM (Correlation Coefficient for G-K+SOM: 54% and for FCM+SOM: 36%).

CONCLUSIONS

In this paper, a fuzzy Kohonen clustering network (FKCN) algorithm, based on the integration of Gustafson-Kessel (G-K) and Kohonen's clustering network (KCN), is proposed. This algorithm addresses some intrinsic problems of KCN.

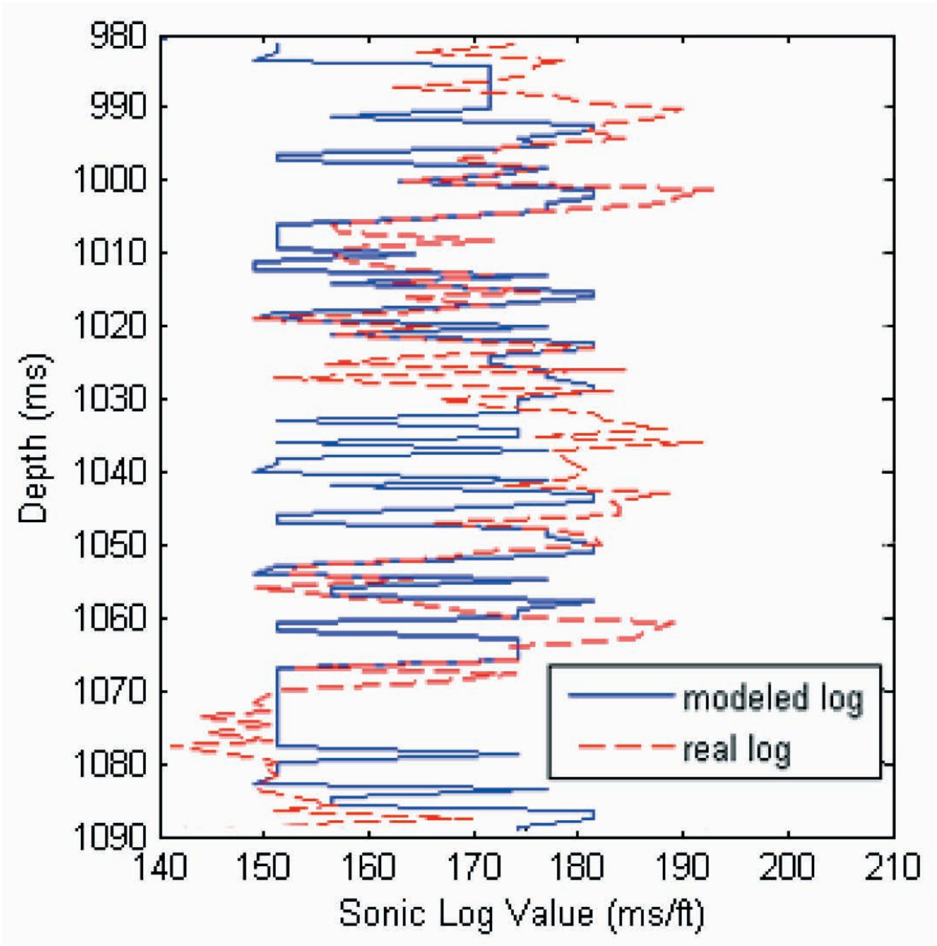

Fig. 4. The modeled log using fuzzy modeling and F-SOM (C-Means), in blue and the real log in red (Correlation Coefficient: 36%).

C-Means algorithms are optimization procedures, but KCN is not. Therefore, integration of G-K and KCN is one way to address several problems of KCNs. The proposed method is a robust clustering technique which can be quality-controlled, using fuzzy modeling. That means using the clustering results, the available but limited log data is rebuilt after clustering to see the performance of the clustering technique. Using a number of complex seismic attributes and a limited log data (single well), log properties can be extrapolated throughout the 3D seismic cube. The data selection strategy used in this paper makes the proposed method a very fast tool for clustering. Classification and modeling results tested on a real data set show reasonable accuracy when compared to well logs.

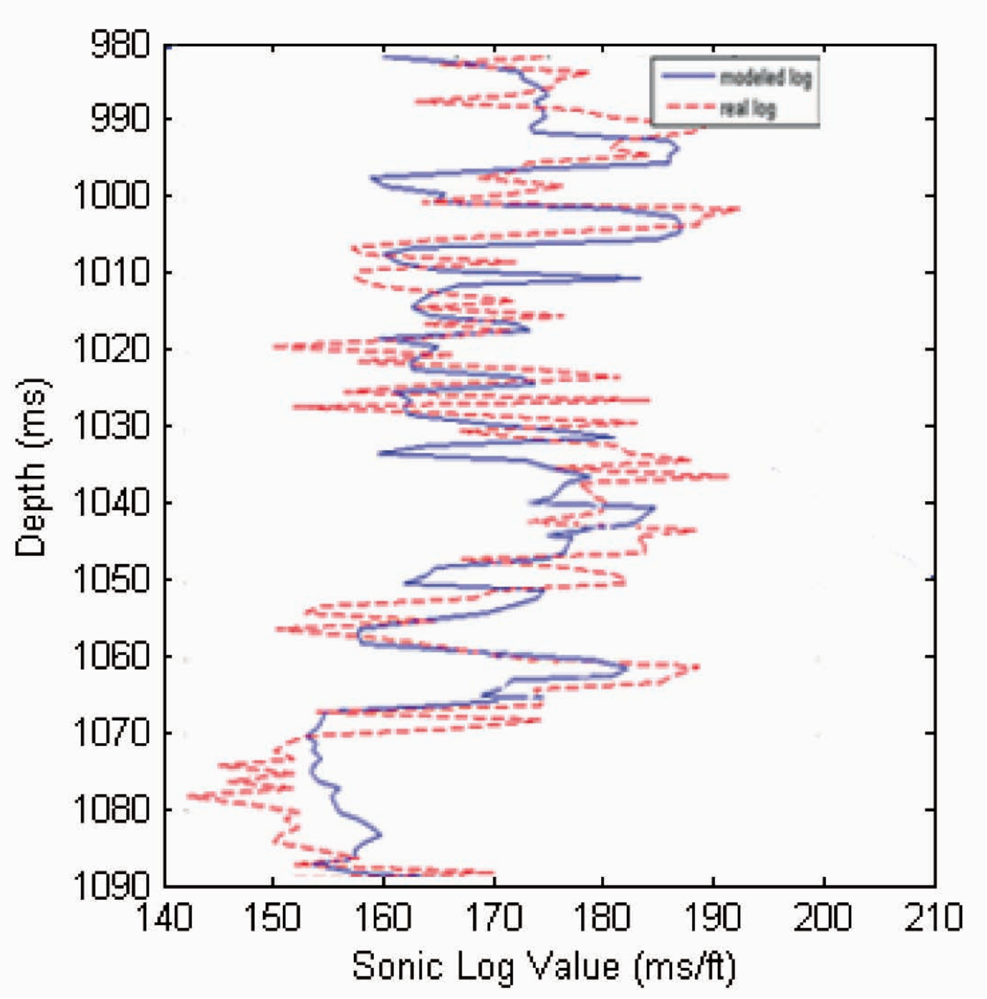

Fig. 5. The modeled log using fuzzy modeling and F-SOM (G-K), in blue and the real log in red (Correlation Coefficient: 54%).

ACKNOWLEDGMENT

The authors wish to thank the Petroleum University of Tehran and the Institute of Geophysics, University of Tehran, for their adequate and complete assistance in this research. We also appreciate the critical reading by the arbitration committee, and we acknowledge our gratitude for enlightening suggestions and insightful comments.

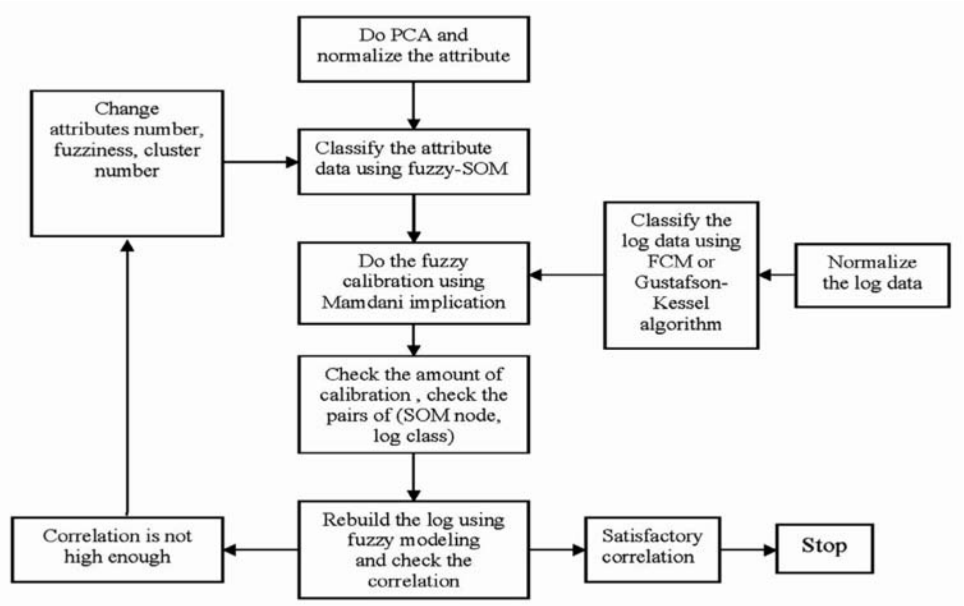

Fig. 6. F-SOM Clustering and fuzzy modeling flowchart. Where is the flowchart?

REFERENCES

Babuska, R., 1996. Fuzzy Modeling and Identification. PhD thesis, Dept. of Control Engineering, Delft University of Technology, Delft, The Netherlands.

Bezdek, J.C., 1981; Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.

Tsao, E.C.-K., Bezdek, J.C. and Pal, N.R., 1992. Fuzzy Kohonen Clustering Networks. IEEE: 1035-1043.

Dubes, R. and Jain, A., 1988. Algorithms that Cluster Data. Prentice Hall, Englewood Cliffs.

Duda, R. and Hart, P., 1973. Pattern Classification and Scene Analysis, Wiley & Sons, New York.

Gustafson, D.E. and Kessel, W.C., 1979. Fuzzy clustering with a fuzzy covariance matrix. In: Proc IEEE CDC, San Diego, CA: 761-766.

Hartigan, J., 1975. Clustering Algorithms. Wiley & Sons, New York.

Hoppner, F., Kalwoon, F. and Kruse, R., 1997. Fuzzy Clustering Analysis. Vieweg Verlag, Braunschweig, Germany.

Hu, W., Xie, D., Tan, T. and Maybank, S., 2004. Learning activity patterns using fuzzy self-organizing neural network. IEEE Trans. Syst. Man. Cybern B, 34: 1618-1626.

Huntsberger, T. and Ajjimarangsee, P., 1989. Parallel self-organizing feature maps for unsupervised pattern recognition, Internat. J. Gen. Systems, 16: 357-372.

Kohonen, T., 1989. Self-Organization and Associative Memory, 3rd Ed. Springer-Verlag, Berlin.

Lippman, R., 1987. An Introduction to Neural Computing, IEEE ASSP Magazine, April: 4-22.

Pao, Y.H., 1989. Adaptive Pattern Recognition and Neural Networks. Addison-Wesley, Reading.

Taner, M.T., 1997. Kohonen's Self Organizing Network with "Conscience", Rock Solid Images Technical Report.

Taner, M.T., 2000. Calibration of Self-Organized Map Clusters. November, Rock Solid Images Technical Report.

Taner, M.T., 2002. Principal Component Analysis Neural Networks; Linear and Non-Linear, January, Rock Solid Images Technical Report.

Tou, J. and Gonzalez, R., 1974. Pattern Recognition Principles. Addison-Wesley, Reading.

Wang, L.X., 1996. A Course In Fuzzy Systems and Control. Prentice Hall, Englewood Cliffs.