Research Article
Determination Method of Nonlinear Membership Function Based on the Density Function of the Square Error

[1]Liu Qi, [2]Liu Kuili, [3]Liu Xuelin, [4]Wang Jie, [5]Chen Zhanwei and [6]Huang Zhenzhen
[1]Department of Physics and Electronic Engineering,
[2]Deparement of Laboratory and Equipment Management, Zhoukou Normal University,
Zhoukou, 466001, China
[3]Medical Instrument Testing Institute of Henan Province, Zhengzhou, 450003, China
[4]School of Electric Engineering, Zhengzhou University, Zhengzhou, 450001, China
[5]Department of Computer Science,
[6]Department of Politics and Law, Zhoukou Normal University, Zhoukou, 466001, China

Abstract: In this study, membership function is used to express the uncertainty in the petroleum drilling, then the fuzzy density clustering method based on density function of square error is put forward, which has avoided the defect of artificial clustering parameter and initial centroid, by using this fast algorithm, the clustering center, the number of clustering center and the parameters used for describing the nonlinear membership function can be obtained. In the simulation, the nonlinear membership function of change percentage of flow out rate EB_Da36 is determined, which has solved the problem that the membership function is hard to be defined.

Keywords: Density function of square error, flow out rate, function approximation, fuzzy density cluster, membership function, petroleum drilling

INTRODUCTION

The research of uncertainty phenomenon and things has great significance (Li *et al*., 1995, 2011; Li, 2000; Lu *et al*., 2008; Luo *et al*., 2007). Petroleum drilling is a risky and costly system engineering (Zhao *et al*., 2009; Fred *et al*., 2005; Li *et al*., 2009). During the process of drilling, there is much uncertainty, such as fuzziness and randomness. In the process, all kinds of engineering accidents may happen and dealing with the troublesome condition and drilling accidents will cost lots of manpower and fund. There have been lots of research on petroleum drilling accidents at home and aboard and have got rich achievements, in which the fuzzy inference method is always applied. When using the fuzzy inference to forecast the accident in the petroleum drilling, there is a difficulty that is how to express the uncertain data detected by the sensor.

Doctor (Zadeh, 1965) proposed a method using membership to signify the uncertainty, there are two forms of membership function: piecewise linear and nonlinear. However, how to determine the membership function has not been resolved essentially. In the application, it is unreliable to determine the value of membership subjectively, when using the statistical method, it costs too much, at the same time and it cannot be achieved sometimes.

Fuzzy clustering is the foundation of classification and system model (Ni *et al*., 2005; Zhou and Zhou *et al*., 2000; Ma *et al*., 2003; Zheng *et al*., 2011). The purpose of clustering is to extract the inherent characteristics from larges of data and get representation of the system behavior. Fuzzy C-means clustering (Pedrycz and Reformat, 2005; Liu *et al*., 2009; Wang *et al*., 2012; Liu and Liu, 2012) put forward by Doctor Bezdek is a mature method at present, but it has disadvantage, that is to say, its clustering result seriously depends on the initial cluster center. For mass data, there are two classic clustering methods: density clustering and grid clustering. The grid clustering method has small calculation and high speed, but the processing result is rough, only the rectangle and the square can be found. The density clustering method can avoid defect of artificial cluster parameter and the initial centroid and can find the round cluster, but its speed is slow. In this study, a clustering method based on the density function of square error is proposed, which can obtain the membership function fast and objectively express the uncertainty of the condition in the petroleum drilling.

Corresponding Author: Liu Qi, Department of Physics and Electronic Engineering, Zhoukou Normal University, Zhoukou 466001, China
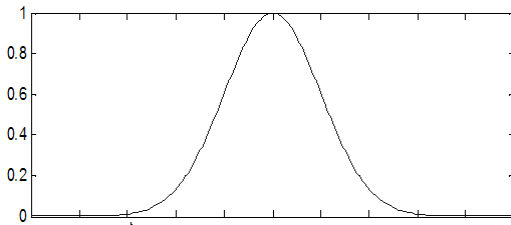
Fig. 1: Gaussian membership function
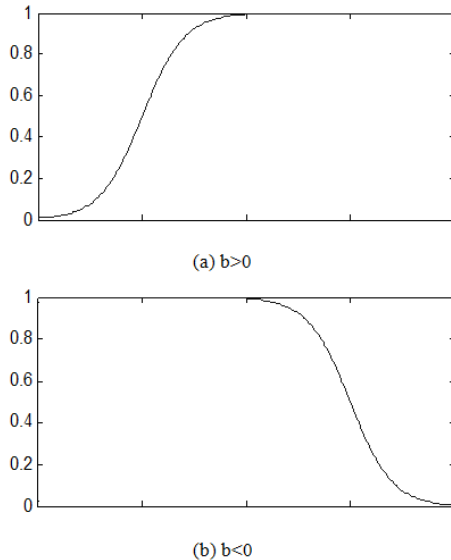

(a) b>0

(b) b<0

Fig. 2: Sigmoid membership function

Nonlinear membership function: The main nonlinear membership functions are Gaussian, Sigmoid, S and parabola. Gaussian and Sigmoid membership function have well smoothness, that means, they have clear physical meaning and no zeros in the figures and both of them are commonly used in the research work. In this study, how to determine Gaussian and Sigmoid function is mainly studied.

The expression of Gaussian membership function (3) is shown as Eq. (1), which is suitable for intermediate language value, in this equation, the main parameters are function center a and width of curve σ, shown as Fig. 1:

$$y = \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right] \tag{1}$$

The expression of sigmoid membership function (3) is shown as Eq. (2), which is suitable for bipolar language value. This equation is determined by b and c, if b is positive, the curve opens to right, if not, the curve opens to left, shown as Fig. 2:

$$y = \frac{1}{1 + \exp\left[-b(x-c)\right]} \tag{2}$$

FUZZY DENSITY CLUSTER METHOD BASED ON THE DENSITY FUNCTION OF SQUARE ERROR

In this study, a fuzzy density cluster method based on the density function of square error is proposed, first, according to the data distance, the initial cluster center is obtained, then by using the density function, the cluster center is modified, at last the final center and the membership is got. This method can avoid falling into the local minimum and the speed is fast.

Density function of the square error: In fact, there is a defect in the traditional density cluster method when evaluating standard parameters in the connected domain, in this study, a new evaluation criterion of the density is brought forward, that is the density function of the square error, shown as Eq. (3):

$$E_\rho = \frac{\sum_{i=1}^{k} \sum_{a_i \in C_j}^{c} \left|a_i - O_j\right|^2}{n\pi d^2 f(\lambda, d)} \tag{3}$$

In the above equation,
d : The radius of the similarity range
n : The number of the nodes in a cluster

In the adjacency domain with radius d, E_ρ is the density of the square error, which reflects the hypo dispersion of each sample to the center particle, when the distance is small, the density of the particle is large. f (λ, d) Is a penalty function of factor $\lambda(\lambda>1)$ in the index grade, which represents the intensive degree of a cluster and can be set by the user. With the increasing of d f (λ, d) will grow in the index degree, as a result, if the slope of the function curve becomes larger, the curve is steeper. Meanwhile, proportion of the area is larger, the cluster is sparser and vice versa. In this study, Eq. (4) is used as regulatory factor:

$$f(\lambda, d) = m \cdot \lambda^{kd} \cdot k \tag{4}$$

Clustering method: For the data set X = {x₁, x₂, …, xₙ} ⊂ R, the process of a new cluster method is shown as follows:

Step 1: Set a range value θ of the square error density, mean of the neighborhood radius of the core object and parameter D of the neighborhood radius of the extended node, here the calculation is shown as Eq. (5) and (6):

$$\theta = a \cdot \frac{\sum_{i=1}^{k} \sum_{a_i \in C_j}^{c} |a_i - O_j|^2}{n\pi d^2 f(\lambda, d)} \tag{5}$$

$$D = \beta \cdot \sum_{i=1}^{n} E_i + (1-\beta)E_r \tag{6}$$

In the above equation,

E : A neighborhood radius of the centroid, in this region, the number of samples means a sample density

E_i : A radius of the core object

E_p : Radius of the object waiting for extended

Step 2: Calculate the distance of any two data:

$$D_i = x_{i+1} - x_i \tag{7}$$

Set two nearest data as a class and select the middle point of the data as the initial cluster data $w_k(1)$ here $k = 1$.

Step 3: Delete the field whose center is $w_k(1)$ and radius is α from the data set and calculate the distance of the residual data, then definy the nearest data as a cluster, at last select the middle point of the two data as a new initial cluster data $w_k(1)$.

Step 4: Repeat Step 3, then the q cluster center can be obtained; this center is defined as the initial center $w_k(l)$, here $l = 1$ $k = 1, 2, \ldots, q$.

Step 5: Calculate the membership degree u_{ik} of x_i to $w_k(l)$:

$$u_{ik}(l) = \frac{\left(1/\|x_i - w_k(l)\|^2\right)^{1/m-1}}{\sum_{i=1}^{q} \left(1/\|x_i - w_k(l)\|^2\right)^{1/m-1}} \tag{8}$$

Step 6: Correct the cluster center:

$$w_k(l+1) = \frac{\sum_{i=1}^{n} [u_{ik}(l)]^m x_i}{\sum_{i=1}^{n} [u_{ik}(l)]^m} \quad 1 \le k \le q \tag{9}$$

Step 7: According to Eq. (3), calculate the density function of square error.

If

$$E_\rho < \frac{\theta}{a} \tag{10}$$

Stop the iteration and export the cluster result, if not, set $l = l + 1$ and return to step 5.
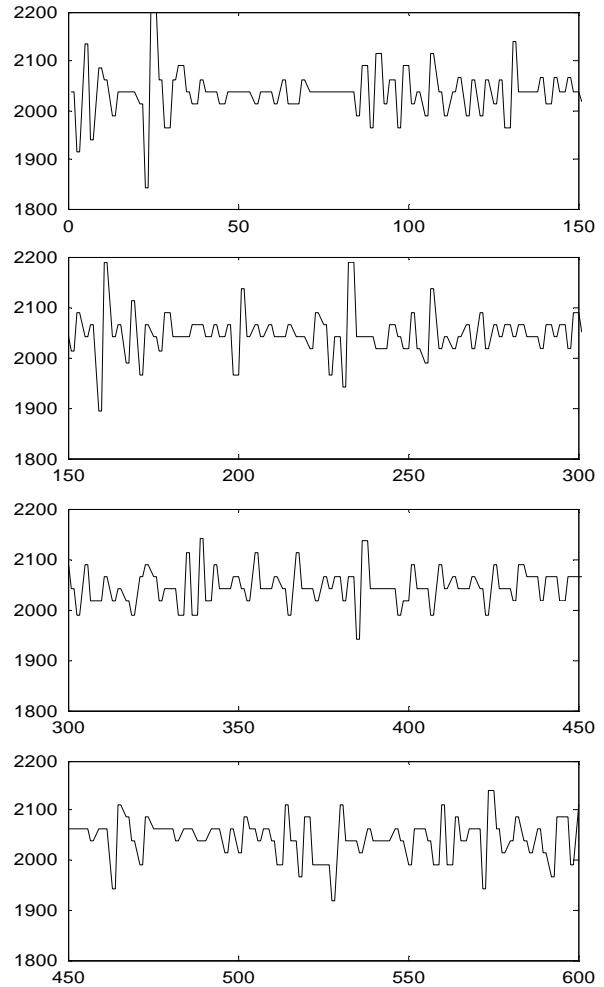

Fig. 3: The flow out rate in the petroleum drilling

SIMULATION

By using the fuzzy density clustering method, the clustering center, the number of center and the membership can be obtained, according to these parameters, the nonlinear membership function can be determined. In this simulation, the data EB-Da36 of flow out rate is selected from a petroleum drilling in the north of Hubei (Wang *et al.*, 2007, 2010), as the change percentage of the flow out rate is a main parameter for kick and lost circulation, the nonlinear membership function of it is defined. The distribution of 600 data from the total volume EB-Da36 is shown as Fig. 3.

By using these parameters: the number of the cluster center q, the cluster center and the membership, the nonlinear membership function can be obtained by the following method:

Step 1: Converse the sample data and get the change percentage of the flow out rate, then map it to the unified universe space.
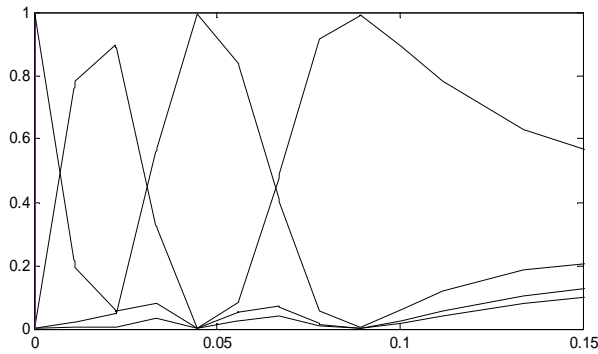
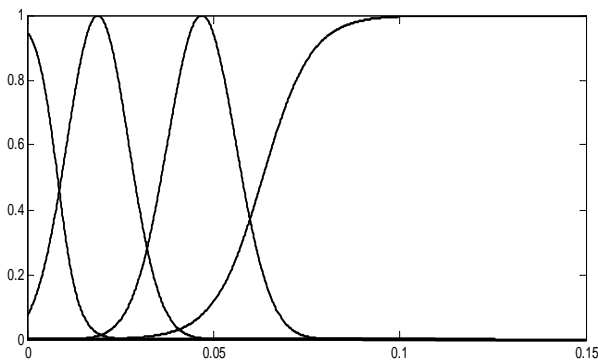Fig. 4: Membership of the change of the flow out rate

Fig. 5: Membership function of the change of flow out rate

Table 1: Parameters of membership function

	Function types	Function models	Parameters
1	Sigmoid	$y = \dfrac{1}{1 + \exp[-b(x-c)]}$	$b = -362.1$ $c = -0.0081$
2	Gaussian	$y = \exp\left[-\dfrac{(x-a)^2}{2\sigma^2}\right]$	$\alpha = 0.0189$ $\sigma = 0.0083$
3	Gaussian	$y = \exp\left[-\dfrac{(x-a)^2}{2\sigma^2}\right]$	$\alpha = 0.0468$ $\sigma = 0.0092$
4	Sigmoid	$y = \dfrac{1}{1 + \exp[-b(x-c)]}$	$b = 153.6$ $c = 0.0632$

Step 2: By using the fuzzy density clustering algorithm, the cluster center w_k the number of cluster center q and membership degree u_{ik} are determined. Seen from Fig. 4 it can be known that the middle linguistic value basically obeys to the normal distribution, which can be represented by Gaussian membership function, but there is an edge distortion in the bipolar values, which can be well avoided by the sigmoid membership function.

Step 3: By using Gaussian membership function, the universe of the middle linguistic value is approached a curve composed by membership degree u_{ik} of 600 sample data, that is to say, the object function shown in Eq. (11) is made to be the smallest, the parameters obtained are shown in Table 1:

$$g(a_k, \sigma_k) = \sum_{i=1}^{600} \left\{ \exp\left[-\frac{(x_i - a_k)^2}{2\sigma_k^2} \right] - u_{ik} \right\}^2 \qquad (11)$$

Step 4: By using Sigmoid membership function, the universe of the bipolar linguistic value is approached a curve composed by membership degree u_{ik} of 600 sample data, that means, the object function shown in Eq. (12) is made to be the smallest, the parameters obtained are shown in Table 1:

$$g(b_k, c_k) = \sum_{i=1}^{600} \left\{ \frac{1}{1 + \exp[-b(x-c)]} - u_{ik} \right\}^2 \qquad (12)$$

Step 5: According to the parameters in Table 1, the membership function can be defined, which is shown as Fig. 5.

CONCLUSION

According to the case study, the fuzzy clustering method based to the density function of square error put forward in this study has advantages, which avoids the defect of artificial clustering parameter and initial centroid, by using this method, the clustering center, clustering number and the membership can be obtained. By function approximation, the Gaussian and Sigmoid membership function of flow out rate in petroleum drilling is determined, which solves the problem that nonlinear membership function is hard to be obtained.

ACKNOWLEDGMENT

REFERENCES

Fred, A., 2005. Applications of AI and soft computing for challenging problems in the oil industry (J). J. Petroleum. Sci. Eng., 47(1-2): 5-14.

Li, D., 2000. Uncertainty in knowledge representation (J). Eng. Sci., 2(10): 73-79.

Li, D., H. Meng and X. Shi, 1995. Membership clouds and membership cloud generators (J). J. Comput. Res. Dev., 32(6): 51-20.

Li, Q., D. Chang and Y. Xu, 2009. Drilling risk management system based on knowledge integration (J). Acta Petrolei Sinica, 5: 755-759.

Li, D., C. Liu and W. Gan, 2011. Proof of the heavy-tailed property of normal cloud model (J). Eng. Sci., 4: 20-23, (In Chinese).

Liu, Q. and X. Liu, 2012. Membership function of the normal cloud determined by FSM method (J). Process Automat. Instrumentation, 2: 16-18, (In Chinese).

Liu, Q., H. Zhang and Z. Huang, 2009. Analysis of oil drilling working state based on Fuzzy Clustering Method (FCM) (J). J. Zhoukou Normal Univ., 5: 47-49, (In Chinese).

Lu, J., Y. Zhang and Z. Song, 2008. Research and application on the fuzzy association rules (M). Sci. Press, 2: 43-51.

Luo, Z., G. Zhang and D. Li, 2007. Probability statistics analysis of one-dimensional normal cloud (J). Info. Control, 8:471-475.

Ma, S., T. Wang and S. Tang, 2003. A fast clustering algorithm based on reference and density (J). J. Software, 14(6): 1089-1095, (In Chinese).

Ni, W., Z. Sun and J. Lu, 2005. K-LDCHD A local density based-neighborhood clustering algorithm for high dimensional space (J). J. Comput. Res. Dev., 42(5): 784-79l, (In Chinese).

Pedrycz, A. and M. Reformat 2005. Hierarchical FCM in a stepwise discovery of structure in data (J). Soft. Comput., 10(3): 244-256.

Wang, J., D. Chen and X. Zhu, 2007. Oil Drilling Accident Pre-Warning System Based on Hierarchical Fuzzy System (P). China Patent: 200710055178.6.

Wang, J., G. Wu and X. Zhu, 2010. Dynamic processing of the right border of real time logging signal in wavelet analysis (J). J. Zhengzhou Univ. Natural Sci. Edn., 3: 84-89, (In Chinese).

Wang J, Q. Liu and Q. Dang, 2012. Nonlinear membership function established by single-shot clustering method (J). J. Zhengzhou Univ. Eng. Sci., 2: 28-30, (In Chinese).

Zadeh, L.A., 1965. Fuzzy sets (J). Info. Control, 8: 338-353.

Zhao, S., J. Yan and Y. Shu, 2009. Prediction model for rheological parameters of oil-based drilling fluids at high temperature and high pressure (J). Acta. Petrolei. Sinica., 4: 603-606.

Zheng, G, J. Xiao and Q. Jiang, 2011. Similarity and inclusion measures between IT-FSs (J). Control Decision, 6: 861-866, (In Chinese).

Zhou, S. and A. Zhou, 2000. A data-partitioning based DBSCAN algorithm (J). J. Comput. Res. Develop., 37(10): 1153-1159, (In Chinese).