Research Journal of Applied Sciences, Engineering and Technology 5(8): 2527-2530, 2013
DOI:10.19026/rjaset.5.4691
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: July 27, 2012 Accepted: September 03, 2012 Published: March 15, 2013

Research Article

Determination Method of Nonlinear Membership Function Based on the Fuzzy Density Means Cluster

[1]Liu Qi, [1, 2]Chen Xi, [3]Wu Zhenjie, [4]Liu Xuelin, [5]Wang Jie and [6]Chen Zhanwei
[1]Department of Physics and Electronic Engineering,
[2]Deparement of Laboratory and Equipment Management, Zhoukou Normal University,
Zhoukou 466001, China
[3]Henan Pinggao Toshiba High Voltage Switchgear Co. Ltd., Pingdingshan 467000, China
[4]Medical Instrument Testing Institute of Henan Province, Zhengzhou 450003, China
[5]School of Electric Engineering, Zhengzhou University, Zhengzhou 450001, China,
[6]Department of Computer Science, Zhoukou Normal University, Zhoukou 466001, China

Abstract: Gaussian and Sigmoid membership function are commonly used, both of which have well smoothness, clear physical meaning and have no zeros in their figures. In this paper, fuzzy density means clustering method based on data distance is put forward, by using this method, the initial cluster center is obtained objectively, which has avoided the cluster results falling into a local minimum, finally the parameters used for describe the membership function are got. Through simulation, the membership function of data EB_Da51 in petroleum drilling is determined, which solves the problem that the function is hard to be defined.

Keywords: Function approximation, fuzzy density means cluster, initial cluster center, membership function, petroleum drilling, variation of total volume

INTRODUCTION

The phenomenon and things with uncertainty are widespread in the nature and human society. In the carrier of human thinking and understanding, how to express and deal with the uncertainty is a hot-spot and key point in the research of nature science, which is also a blockage (Li *et al*., 1995, 2011; Li, 2000; Lu *et al*., 2008; Luo *et al*., 2007) at the same time. In 1965, Doctor Zadeh published a seminal paper on fuzzy set (Zadeh, 1965) and put forward a method using membership to signify the uncertainty. The main forms of membership function are piecewise linear and nonlinear (Zheng *et al*., 2011).

Until now, method for determining the membership function (Liu and Yuan-Dong, 2011) has not been fundamentally resolved, but in practice, it is unreliable to determine the membership value subjectively, if using the statistical method, it not only costs too much, but also cannot be achieved sometimes. Fuzzy clustering method is the base of many classification problems and system modeling, the purpose of fuzzy clustering is to extract the inherent characteristics from a large number, to obtain the compact representation of the system behavior. Up to present, fuzzy clustering has been used in the field of picture transmission, voice recognition and data mining (Ni *et al*., 2005; Zhou and Zhou, 2000; Ma and Tang, 2003; Zheng *et al*., 2011; Zhou *et al*., 2007; Yu and Yu, 2010; Nasibov and Ulutagay, 2009). Bezdek proposed a Fuzzy C-mean Clustering method, which can classify a data point into a cluster degree, in a word, it is a more mature method, but the cluster result deeply depends on initial cluster center randomly generated (Pedrycz and Reformat, 2005; Liu *et al*., 2009; Wang *et al*., 2012; Liu and Xiaoqing, 2012). In this paper, a fuzzy density-mean clustering method based on data distance is proposed, which is simple and easy to spread, by using this method, nonlinear membership function of variable of total volume in petroleum drilling (Zhao *et al*., 2009; Fred, 2005; Li *et al*., 2009) is determined.

NONLINEAR MEMBERSHIP FUNCTION

There are several nonlinear membership functions, such as Gaussian, Sigmoid, S and Parabola. Membership function of Gaussian and Sigmoid have well smoothness, that is to say, the figure of them have clear physical meaning and no zeros, which are commonly used in the research work. In this paper, the

Corresponding Author: Liu Qi, Department of Physics and Electronic Engineering, Zhoukou Normal University, Zhoukou 466001, China

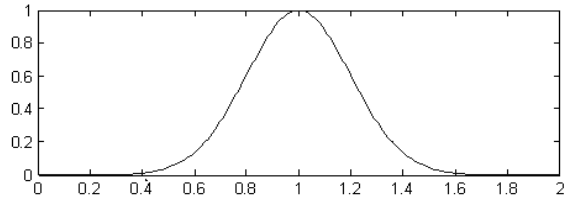
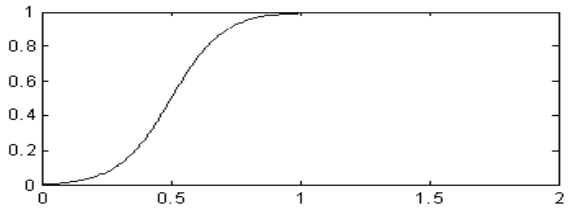
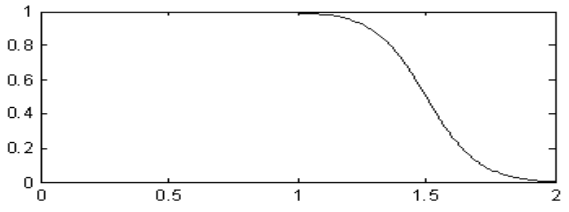


Fig. 1: Gaussian membership function



(a) b>0



(b) b<0

Fig. 2: Sigmoid membership function

determination of Gaussian and Sigmoid is mainly studied.

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

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

The expression of sigmoid membership function is shown as Eq. (2), which is suitable for a membership function with bipolar language value. The equation is determined by b and c, if b is positive, the curve of it opens to right, but if b is negative, the curve opens to left, such as Fig. 2:

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

FUZZY DENSITY MEANS CLUSTERING METHOD

In this paper, a fuzzy density means clustering method based on data distance is proposed. In this paper, first according to data distance, the initial cluster centers are obtained, then on the basis of value function and error function, the cluster center is corrected, last, membership degree of the cluster center and data is determined, which has avoided the cluster results falling into a local minimum.

For data $X = \{x_1, x_2, \ldots, x_n\} \subset R$, the process of fuzzy density means clustering method is shown as follows:

Step 1: Set threshold α of the smallest distances between classes, fuzzy index m and number of the cluster center q

Step 2: Calculate the distance of two arbitrary data:

$$D_i = x_{i+1} - x_i \tag{3}$$

Definite the most nearest data as a cluster and select the middle point of the two data as the first initial cluster data $w_K(l)$, K = 1

Step 3: Delete the field whose center is $w_K(l)$ and radius is α from the data set and calculate the distance of the last data, then definite the nearest data as a cluster and select the middle point of the two data as the first initial cluster data $w_K(l)$

Step 4: Repeat step 3 until the cluster center q is obtained and define the cluster center as the initial center $W_K(l)$, l = 1, k = 1, 2,…, q

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

$$u_{ik}(l) = \frac{\left(1/\left\|x_i - w_k(l)\right\|^2\right)^{1/m-1}}{\sum_{i=1}^{q}\left(1/\left\|x_i - w_k(l)\right\|^2\right)^{1/m-1}} \tag{4}$$

Step 6: Adjust the cluster center:

$$w_k(l+1) = \frac{\sum_{i=1}^{n}[u_{ik}(l)]^m x_i}{\sum_{i=1}^{n}[u_{ik}(l)]^m} \quad 1 \le k \le q \tag{5}$$

Step 7: Calculate the value function j(l) and calculation error δ:

$$J(l) = \sum_{k=1}^{q}\sum_{i}^{n} u_{ik}^m \left\|x_i - w_k(l)\right\|^2 \tag{6}$$

$$\delta = \left\|J(l+1) - J(l)\right\| \tag{7}$$

If $\delta < E_{max}$, stop the iteration and export the cluster result, if not, set l = l + 1 and return to step 5.

In the above method, α is a threshold of the smallest distance between the clusters, which represents the dense radius of the data and is determined on the basis of the data distribution? In practical analysis, as long as the cluster center q is determined, α has no effect on the cluster result, but if α is too large, the number of the cluster center obtained is smaller than q. The determination of fuzzy index m is a little complex, according to the practical analysis, the cluster result is not sensitive to variation of m and generally, the best selection interval of m is (1.5, 2.5), but for convenience m = 2.

SIMULATION RESULTS

By using fuzzy density means clustering method, the cluster center and the membership degree to the center can be obtained, according to these parameters, the membership function can be determined. In this paper, the total volume data EB_Da51 from the petroleum drilling (Wang *et al*., 2007, 2010) in Hubei province is selected and the membership function of variation of the volume are defined. Some data of EB_Da51 is shown in Fig. 3, in which the ordinate represents the total volume with unit m^3.

On the basis of the initial cluster center q obtained by the fuzzy density means clustering method, the classified center and the membership degree of data, the concrete processes of determining the membership function are shown as follows:

Step 1: Unify universe of the data and determine the number of the linguistic value. In this paper, membership function of variation of the total volume in petroleum drilling is searched, then data EB_Da51 is mapped to a unified universe space and there are four linguistic values: none, small, middle and large. That is to say, q = 4, $1 \le k \le q$.

Step 2: By using the fuzzy density means clustering method, the cluster center w_k and membership degree u_{ik} to w_k are determined. Seen from Fig. 4, 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, luckily sigmoid membership function can avoid this distortion.

Step 3: Universe of the middle linguistic value, such as Small and Middle in this paper, is approached to the curve of membership degree u_{ik} by Gaussian membership function, which is composed by the 500 samples, that is to say the objective function in Eq. (8) is made to be the smallest and the parameters are shown in Table 1:

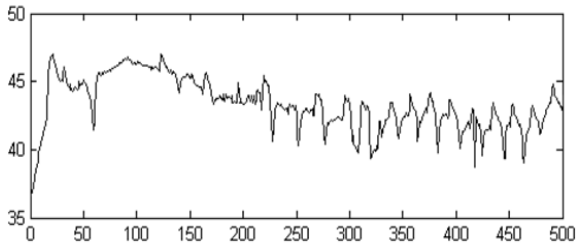


Fig. 3: Change of volume data EB_Da51 from the petroleum drilling

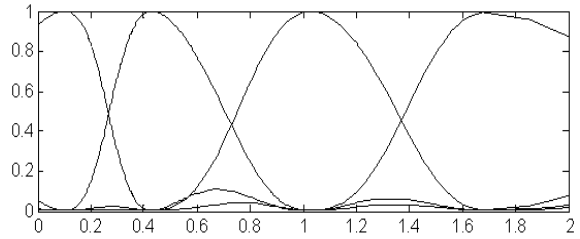


Fig. 4: Membership degree of the change of the total volume

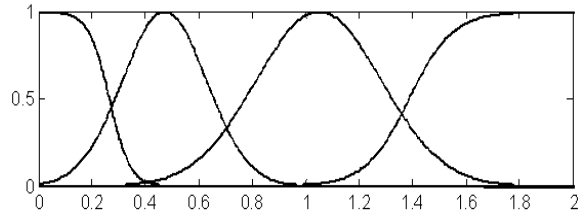


Fig. 5: Membership function of data EB_Da51

Table 1: Parameters of nonlinear membership function

Linguistic value	Function types	Function models	Parameters
None	Sigmoid	$y = \frac{1}{1+\exp[-b(x-c)]}$	$b = -26.12$ $c = 0.266$
Small	Gaussian	$y = \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right]$	$a = 0.4718$ $\sigma = 0.2227$
Middle	Gaussian	$y = \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right]$	$a = 1.049$ $\sigma = 0.3307$
Large	Sigmoid	$y = \frac{1}{1+\exp[-b(x-c)]}$	$b = 12.21$ $c = 1.387$

$$g(a_k, \sigma_k) = \sum_{i=1}^{500} \left\{ \exp\left[-\frac{(x_i - a_k)^2}{2\sigma_k^{\ 2}} \right] - u_{ik} \right\}^2 \quad (8)$$

Step 4: Universe of the bipolar linguistic value, such as None and Large in this paper, is approached to the curve of membership degree u_{ik} by Sigmoid membership function, which is composed by the 500 samples, that is to say that the objective function in Eq. (9) is made to be the smallest and the parameters are shown in Table 1:

$$g(b_k, c_k) = \sum_{i=1}^{500} \left\{ \frac{1}{1+\exp\left[-b(x-c)\right]} - u_{ik} \right\}^2 \quad (9)$$

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

CONCLUSION

According to the case study, the fuzzy density means clustering method put forward in this paper has got objective effects, in this method; the initial cluster center is obtained by the distance of data, which can effectively avoid the clustering results falling into a local minimum. By using this method, the Gaussian and Sigmoid membership function of variation of total volume in petroleum drilling is determined, which solves the problem that nonlinear membership function is hard to be obtained.

ACKNOWLEDGMENT

The research of this paper has been sponsored by Henan Provincial Research Foundation for Basic Research, China (Grant No: 122300410168), Natural Science Foundation of He'nan Educational Committee, China (Grant No. 2011B510021) and the Scientific Research Innovation Foundation for youth teachers of Zhoukou Normal University, China (Grant No. 2012QNA02).

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, 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).

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

Liu, Q. and D.U Yuan-Dong, 2011. Application research of PID control optimized by GBLB-PSO in HVAC system (C). Adv. Mat. Res., 219-220: 1325-1328.

Liu, Q. and L.I.U Xiaoqing, 2012. Membership functions of the normal cloud determined by FSM method (J). Proc. Automat. Instru., 2: 16-18, (In Chinese).

Liu, Q., Z. Hong-Hui and H. Zhen-Zhen, 2009. Analysis of oil drilling working state based on Fuzzy Clustering Method (FCM) (J). J. Zhoukou Norm. 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). Inform. Control, 8:471-475.

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

Nasibov, E.N. and G. Ulutagay, 2009. Robustness of density-based clustering methods with various neighborhood relations (J). Fuzzy Set. Syst., 160(24): 3601-3615.

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. Develop., 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., W.U. Guo-Zeng and Z.H.U. Xiao-Dong, 2010. Dynamic processing of the right border of real time logging signal in wavelet analysis (J). J. Zhengzhou Univ. Nat. Sci. Edn., 3: 84-89, (In Chinese).

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

Yu, X. and X. Yu, 2010. Research of un-supervisory algorithms based on distance and density (J). Comput. Appl. Soft., 27(7): 122-125.

Zadeh, L.A., 1965. Fuzzy sets (J). Inform. 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 IT2 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).

Zhou, Y., X. Kong, R. Zhang, 2007. Research of clustering algorithms in data mining (J). Fujian Comput., 8: 9-10.