An Interval Type-2 Possibilistic C-Means Clustering Algorithm and Its Application

Haihua Xing 1, Huannan Chen 2, Hongyan Lin 1 and Xinghui Wu 1*

1College of Information Science and Technology, Hainan Normal University, Haikou, Hainan, China, 571158
2College of business administration, Hainan College of Economics and Business, Haikou, Hainan, China, 571127
*Corresponding author’s e-mail: hhxing@hainnu.edu.cn

Abstract. In this paper, we aim at the fuzzy uncertainty caused by noise in pattern data. The advantages of PCM algorithm to deal with noise and interval type-2 fuzzy sets to deal with high-order uncertainties are used, respectively. An interval type-2 probability C-means clustering (IT2-PCM) based on penalty factor is proposed. The performance of the algorithm is evaluated by two sets of data sets and two groups of images segmentation experiments. The results show that IT2-PCM algorithm can assign proper membership degrees to clustering samples with noise, and it can detect noise points effectively, and it has good performance in image segmentation.

1. Introduction
In practice, uncertain, inaccurate or incomplete data are often involved. Fuzzy theory is usually used to deal with this kind of fuzzy uncertainty. Fuzzy C-means clustering (FCM) is a widely used algorithm in fuzzy theory, but FCM algorithm is very sensitive to noise and singular value[1]. To solve this problem, some improved algorithms have been proposed: Possibilistic C-Means Clustering (PCM) proposed by Krishnapuram and Keller in 1996[2] and the Possibilistic Fuzzy C-Means Clustering (PFCM) proposed by Pal in 2005[3]. PCM gets rid of the constraint that the sum of membership is one, and it is more robust in dealing with noise. In 2008, Wu et al. proposed an improved possibilistic fuzzy C-means clustering algorithm (IPFCM) to solve the consistency clustering problem of PCM[4]. In 2011, Zhang proposed a sample weighted possibilities clustering algorithm (SWPCM)[5], which determines the initial clustering center according to the joint sample weights. In 2015, Wen proposed an anti-consistency possibilistic C-means clustering algorithm (ACPCM)[6]. In 2016, Peng proposed a new clustering algorithm called maximum the between-cluster (MPCM)[7]. In 2018, Xia et al. proposed Central-Constraints Possibilistic C-Means algorithms based on the Source Domain (CCSD_PCM)[8]. The above algorithms are proposed on the basis of type-1 fuzzy C-means clustering, and have limited ability to deal with high-order fuzzy uncertainties. In recent years, type-2 fuzzy sets have stronger ability to deal with uncertainty than type-1 fuzzy sets, and have been gradually applied to various fields [9-10]. In 2007, Rhee constructed membership intervals on the basis of their research in 2001, and proposed an interval type-2 fuzzy C-means clustering method (IT2FCM)[11]. In 2012, Linda proposed a general type-2 fuzzy C-means clustering algorithm (GT2 FCM)[12]. In 2015, Long proposed a semi-supervised interval type-2 fuzzy C-means clustering algorithm based on spatial
information for multi-spectral satellite image classification [13]. In 2019, Xing et al. proposed a method for generating interval type-2 fuzzy sets that incorporates neighborhood information [14].

In order to deal with fuzzy uncertainty caused by noise data, the probability penalty factor is used as a rule item to improve the objective function of the traditional IT2FCM, and an interval type-2 probability C-means clustering algorithm (IT2-PCM) is proposed. The performance of the proposed method is evaluated by two data sets and two groups of image segmentation experiments.

2. An interval type-2 possibilistic c-means clustering algorithm

Interval type-2 fuzzy sets have strong descriptive ability to uncertainty, and the primary uncertainty of a type-2 fuzzy set can be illustrated using its Footprint of Uncertainty (FOU). The FOU is delimited by a lower membership function (LMF) and by an upper membership function (UMF) [15]. Therefore, the membership function of an interval type-2 fuzzy sets is determined when the corresponding UMF and LMF are defined.

We construct the interval type-2 possibility partition function with two different possibility penalty factors, and propose an interval type-2 possibility C-means clustering algorithm (IT2-PCM) based on possibility penalty factor. The upper bound and the lower bound of the interval type-2 membership function $\mu(x) = [\underline{\mu}(x), \overline{\mu}(x)]$ can be expressed as:

$$
\mu^u_k = \frac{1}{1 + \left( \frac{d^2_k + \sum_{i=1}^{n} \frac{1}{\eta_k} ||x_i - v_k||_2}{\eta_k} \right)^{\frac{1}{m-1}}} \\
\mu^l_k = \frac{1}{1 + \left( \frac{d^2_k + \sum_{i=1}^{n} \frac{1}{\eta_k} ||x_i - v_k||_2}{\eta_k} \right)^{\frac{1}{m-1}}},
$$

UMF: $\pi_k = \max(\mu^u_k, \mu^l_k)$,

LMF: $\mu_k = \min(\mu^u_k, \mu^l_k)$

(1)

where, $\mu_k$ represents the degree of possibility that the $i$th sample belongs to the $k$th class in PCM algorithm, $d_k$ represents the distance between a sample and a certain cluster center, and the possibility penalty factors $\eta_k^u$ and $\eta_k^l$ are defined as follows:

$$
\eta_k^u = K \frac{\sum_{i=1}^{n} \mu_k^l d_k^2}{\sum_{i=1}^{n} \mu_k^l} \\
\eta_k^l = K \frac{\sum_{i=1}^{n} \mu_k^u d_k^2}{\sum_{i=1}^{n} \mu_k^u}
$$

(2)

The calculation method of clustering center $v_k$ is shown at Eq. (3):

$$
v_k = \frac{\sum_{i=1}^{n} \mu_k^u x_i}{\sum_{i=1}^{n} \mu_k^u}
$$

(3)

Based on the above analysis, the main steps of IT2-PCM clustering method are given as follows:
3. Performance evaluation of the proposed algorithm
Experiments on different data sets classification, noise point recognition and image segmentation are carried out. The effectiveness of the proposed algorithm is verified in different applications, and the experimental results are compared with the results of FCM, IT2FCM, PCM clustering method. The algorithm parameters are set as follows: (1) FCM and PCM with $m = 2$, (2) IT2FCM and IT2-PCM with $m_1 = 2$, $m_2 = 5$.

3.1. Experiments and results analysis of artificial data sets
The purpose of the experiment with two data sets is to verify whether the proposed algorithm IT2-PCM can correctly identify the noise points and correctly classify the data. The operation steps are as follows: generating noiseless data sets; adding noise points to the dataset; applying FCM, IT2FCM, PCM and IT2-PCM algorithm to calculate the membership degree of each sample point; According to the membership value, the sample points with lower membership value of each category are recognized as noise points.

The first experiment involves two well-separated clusters of seven data points each and two noise points. These data points are numbered in the order in which they would be encountered in the left to right, bottom to top scan of the figure (fig. 1 (a)), and data points 8 and 9 are the two noise points. Fig. 1 (b)-(e) contain the membership degrees of each data point obtained from FCM, IT2FCM, PCM and IT2-PCM methods.

From fig. 1, we can see that the proposed algorithm IT2-PCM can correctly distinguish two categories, and the membership degrees of the noise points are almost zero in both clusters, so it is easy to detect and eliminate the noise points. However, FCM and IT2FCM algorithms can not identify the noise points effectively because the membership degrees of the two noise points is 0.5.
The second experiment involves three clusters of 200 points each and 100 noise points are added to this noise-free data set. The three clusters of 200 points each are generated based on multivariate normal distribution with parameters $\mu_1 = [4,6]$, $\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\mu_2 = [-1,-2]$, $\Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$, and $\mu_3 = [-3,6]$, $\Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$. The 100 noise points are then generated based on $\mu_4 = [0,4]$, $\Sigma_4 = \begin{bmatrix} 8 & 0 \\ 0 & 8 \end{bmatrix}$ and are
added to the noise-free data set. These three clusters are demonstrated in Fig. 2(a). FCM, IT2FCM, PCM, IT2-PCM methods are applied and the obtained membership functions are depicted in (Fig.2(b)–(e), respectively.

It is obvious from Fig. 2 that the membership degrees of noise points in IT2-PCM is very low, and it is easy to detect noise points. For the non-noise sample data, the membership degrees can be given as well as the FCM, IT2FCM and PCM algorithms, so as to correctly distinguish the categories of the data. From the experimental results, it can be seen that IT2-PCM algorithm is not disturbed by noise points and can assign the membership degrees of noise points reasonably. However, the membership degrees of FCM and IT2FCM clustering is easily influenced by noise.

![FIGURE 2](image_url)

**FIGURE 2.** Two-dimensional data set with noise points and the membership degrees of the 2st experiment obtained by FCM, IT2FCM, PCM and IT2-PCM
3.2. Image segmentation experiments and results analysis

During segmentation, an image is divided into different homogeneous regions of similar attribute such as image luminance, color components edges and texture. In fact, as definitions and characteristics of these attributes are imprecise and uncertain. In this paper we provide two images, aircraft (Fig. 3(a)) and moonlight forest (Fig. 4(a)), and segment them based on their color component or pixels’ intensity level into different clusters to verify the performance of the proposed algorithm. The results are compared with FCM, IT2FCM and PCM methods.

![Aircraft and results of image segmentation obtained by FCM, IT2FCM, PCM and IT2-PCM](image1)

From the results shown in Fig. 3 (b) - (e) and Fig. 4 (b) - (e), it can be seen that IT2-PCM can detect the main objects in the images more effectively than the other methods. FCM and IT2FCM segment aircraft with blurred edges (see Fig. 3 (b) - (c)), while PCM and IT2-PCM can segment complete aircraft with clear edges (see Fig. 3 (d) - (e)). It can be observed from the results (see Fig. 4 (d) - (e)) that IT2-PCM can segment the trees and the moon from the background completely, while the FCM and IT2FCM methods fail to segment the moon from the background. One of the reasons why IT2-PCM method has a good performance is that it cancels the condition that the sum of membership degrees of samples is one, so IT2-PCM can deal with noise points and singular values better. Moreover, the interval type-2 fuzzy sets can describe the uncertainties caused by noise better.

4. Conclusion

Traditional IT2FCM algorithm can not detect and eliminate noise points effectively. In this paper, we construct the interval type-2 possibility partition function by using the possibility penalty factor in PCM algorithm, and propose an interval type-2 possibility C-means clustering algorithm based on the penalty factor (IT2-PCM). The performance of the algorithm is evaluated by two sets of data sets, two groups of images segmentation experiments. The results show that IT2-PCM algorithm can assign proper membership degrees to clustering samples with noise, and it can detect noise points effectively. And it has good performance in image segmentation.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China under Grant 62066013, and in part by Hainan Provincial Natural Science Foundation of China under Grant 2019RC182 and 618MS058.
References

[1] R. Krishnapuram, J. M. Keller, “A possibilistic approach to clustering,” IEEE Trans on Fuzzy Systems, vol. 1, no. 2, pp. 98-110, 1993.

[2] R. Krishnapuram, J. M. Keller, “The possibilistic C-means algorithms: insights and recommendation,” IEEE Trans on Fuzzy Systems, vol. 4, no. 3, pp. 385-393, 1996.

[3] N. R. Pal, K. Pal, J. M. Keller, “A possibilistic fuzzy c-means clustering algorithm,” IEEE Transactions on Fuzzy Systems, vol. 13, no. 4, pp. 517-530, 2005.

[4] X. H. Wu, J. J. Zhou, “A Novel Possibilistic Fuzzy C-Means Clustering,” Acta Electronica Sinica, vol. 36, no. 10, pp. 1996-2000, 2008.

[5] C. Zhang, B. Liu, “Possibilistic fuzzy clustering algorithm based on sample weighted,” In: Proceedings of International Workshop on Intelligent Systems and Applications, Wuhan, China, 2011, pp. 1-4.

[6] C. J. Wen, Q. M. Wang, et al. “Anti-consistency possibilistic C-means clustering algorithm,” Computer Science, vol. 4, no. 1, pp. 290-293, 2015.

[7] X. Peng, L. Di, W. J. Yang, “Image segmentation method based on maximum between-cluster of PCM algorithm,” Computer Engineering and Applications, vol. 52, no. 16, pp. 142-148, 2016.

[8] Y. Y. Xia, Y. Liu, Y. D. Huang. “Central-constraints possibilistic C-means algorithms based on source domain,” Computer Engineering and Applications, vol. 54, no. 5, pp. 72-78, 2018.

[9] J. M. Mendel, “Type-2 fuzzy sets and systems: an overview,” Computational Intelligence Magazine, IEEE. Vol. 2, no. 1, pp. 20-29, 2007.

[10] Q. Liang, J. M. Mendel, “Equalization of nonlinear time-varying channels using type-2 fuzzy adaptive filters,” IEEE Transactions on Fuzzy Systems, vol. 8, no. 5, pp. 551-563, 2000.

[11] C. Hwang, F. C. H. Rhee, “Uncertain fuzzy clustering: Interval type-2 fuzzy approach to C-means,” IEEE Transactions on Fuzzy Systems, vol. 15, no. 1: pp. 107-120, 2007.

[12] O. Linda, M. Manic, “General type-2 fuzzy C-means algorithm for uncertain fuzzy clustering,” IEEE Transactions on Fuzzy Systems, vol. 20, no. 5, pp. 883-897, 2012.

[13] T. N Long, D. S Mai, W. Pedrycz, “Semi-supervising Interval Type-2 Fuzzy C-Means clustering with spatial information for multi-spectral satellite image classification and change detection,” Computers & Geosciences, vol. 83, pp. 1-16, 2015.

[14] XING H H, HE H, YU X C. An interval Type-2 fuzzy sets generation method for remote sensing imagery classification[J]. Computers and Geosciences, 2019, 33.10.1016/j.cageo.2019.06.008

[15] J. M. Mendel, “Type-2 fuzzy sets and systems: an overview,” IEEE Comput. Intell. Mag., vol. 2, no. 1, pp. 20–29, 2007.