

RFCM clustering algorithm based on adaptive weights for radar signal sorting

Qiang Zhang, Hongwei Wang, Yuanzhi Yang, Wenzhe Wang

Aeronautics and Astronautics Engineering College, Air Force Engineering University, Xi'an 710038, China

ABSTRACT: Rough Fuzzy C-Means (RFCM) clustering algorithm is a valid algorithm to process the inseparability of border of clusters, which handles isolated data well and reduces influence of noise on clusters results. However, different data objects are unified with fixed weights, which influences clusters results greatly. In order to solve the problem, RFCM clustering algorithm based on self-adaptive weights is proposed. According to different distance between every data object and clustering center, using improved arc cotangent function to redistribute the distance, acquiring adaptive weights through the equivalence factor in each iteration, then carrying out RFCM clustering algorithm. The simulations are performed on UCI data sets, and the results show the validity of the proposed algorithm. Furthermore, the proposed algorithm is applied in sorting radar signal, and the results show the practicability of the proposed algorithm.

KEYWORD: clustering; adaptive weights; equivalence factor; radar signal sorting.

1 INTRODUCTION

According to the increasingly complex and intense electronic warfare pulse environment of, radar signal sorting has been the key technology of the modern high technology war and the future information war (Qiang GUO, 2008). Clustering algorithm is one of the most important methods of radar signal sorting. As the battlefield electromagnetic environment becomes more dense and complex, facing with high density, poor separability, large noise of radar signal, the traditional clustering algorithms such as k -means cannot perform the sorting task effectively.

The rough sets and fuzzy sets are two different mathematical tools to deal with uncertain information, and a lot of researches have used it in clustering analysis, e.g. fuzzy C-Means clustering algorithm (FCM), rough C-Means clustering algorithm (RCM), fuzzy rough C-Means clustering algorithm (FRCM) and rough fuzzy C-Means clustering algorithm (RFCM) and so on. RFCM clustering algorithm, compared with other algorithms, is more suitable for the complex and heterogeneous distribution of data clustering analysis, and can deal with fuzzy clustering boundary problem effectively, as well as can better deal with isolated point and reduce the clustering result affected by noise, thus optimizing clustering results. Therefore RFCM clustering algorithm can analyze highly dense complex radar signal quite well.

The problem existing in the RFCM clustering algorithm is the ignorance of the imbalance of the different data object distribution, uniformly fixing weights for all objects. For example in lower approximation set, each object is different from distance clustering center, suggesting that the closeness between each of the object and the center of them are different, thus if the same all objects are given equally fixed weights, the clustering center will have greater extent in the iterative process of change, seriously influencing the clustering performance. Therefore aiming at the problems, it is quite necessary to improve the RFCM clustering algorithm, and to apply it to the radar signal sorting, making it more suitable for radar signal sorting.

Aiming at the problem the RFCM clustering algorithm is faced with, a RFCM clustering algorithm based on adaptive weights is proposed, and applied to radar signal sorting in this paper. The specific ideology of the algorithm is: in the process of each iteration, according to the different Euclidean distances between the clustering center and each data object, the equivalence factor of each data object is ascertained after redistributing the Euclidean distance using improved positive-axis arc cotangent function, to dynamically adjust the fixed weights to acquire adaptive weights, and to further renew the cluster center according to the object weights and the current distribution state. Finally, simulation is carried out on UCI data and radar signal data, vali-

dating the effectiveness of the proposed algorithm and its practical applicability on radar signal sorting.

2 RFCM CLUSTERING ALGORITHM

Lingrus came up with the RCM clustering algorithm, the main idea is using the upper approximation set and the lower approximation set's concepts of rough set to improve C-means clustering algorithm. RFCM clustering algorithm is added the weights of fuzzy membership degree based on RCM clustering algorithm, which takes the membership degree as the thin judgment criterion. At present, there are two main RFCM clustering algorithms proposed by Mitra and Maji (MITRA S, 2006, MAJI P, 2007), Maji's RFCM clustering algorithm is adapted in this paper.

Let $X = \{x_1, \dots, x_N\}$ as sample data set, x_j denotes a data object, C_i denotes the i th classes, \underline{C}_i and \overline{C}_i denote respectively the approximation set and the lower approximation set of i th classes; $V = \{v_1, \dots, v_k\}$, v_i denotes the clustering center of C_i ; u_{ij} denotes the degree of membership of j th object belong to i th classes. The calculation formula of RFCM algorithm's clustering center is as follows:

$$v_i = \begin{cases} w_l A + w_u B & \underline{C}_i \neq \emptyset \wedge (\overline{C}_i - \underline{C}_i) \neq \emptyset \\ A & \underline{C}_i \neq \emptyset \wedge (\overline{C}_i - \underline{C}_i) = \emptyset \\ B & \underline{C}_i = \emptyset \wedge (\overline{C}_i - \underline{C}_i) \neq \emptyset \end{cases} \quad (1)$$

$$A = \frac{\sum_{x_j \in \underline{C}_i} x_j}{|\underline{C}_i|} \quad (2)$$

$$B = \frac{\sum_{x_j \in (\overline{C}_i - \underline{C}_i)} u_{ij}^m x_j}{\sum_{x_j \in (\overline{C}_i - \underline{C}_i)} u_{ij}^m} \quad (3)$$

Where w_l and w_u denote respectively the weights of the approximation set and the lower approximation set, and $w_l > w_u$, $w_l + w_u = 1$, and are both preplanned fixed weight.

The updated formula of fuzzy membership is as follows:

$$u_{ij} = \begin{cases} 1 & x_j \in \underline{C}_i \\ u & x_j \in \overline{C}_i - \underline{C}_i \\ 0 & x_j \notin \overline{C}_i \end{cases} \quad (4)$$

$$u = 1 / \left(\sum_{z=1}^k \left(\frac{d_{ij}}{d_{zj}} \right)^{\frac{2}{m-1}} \right) \quad (5)$$

Where $d_{ij} = \|x_j - v_i\|$ denotes the Euclidean distance between data object x_j and clustering center v_i , $d_{zj} = \|x_j - v_z\|$ denotes the Euclidean distance between data object x_j and clustering center v_z .

3 RFCM CLUSTERING ALGORITHM BASED ON ADAPTIVE WEIGHT

Equivalence factor. Equivalence factor is used to describe the equivalence degree of different data object in the belonged classes, the smaller the equilibrium factor is, the smaller the equilibrium degree, i.e. the unbalanced degree is bigger. Distance is the most remarkable measure weather objects are balanced in the classes, and distance is selected to calculate equivalence factor; the farther the distance between data object and clustering center is, the smaller the equivalence degree and equivalence factor.

While distance cannot be calculated directly, because the distance after normalization is small and linear distribution, which cannot perfectly shows the differences of various data objects. In order to better depict the degree of equilibrium data distribution, distance needs to be recalculated and calculates equivalence factor.

Considering that the Euclidean distance between different data objects and clustering center is $[0, +\infty)$, and the Euclidean distance and equivalence factor have inverse relationship. So the function whose value range of variables is $[0, +\infty)$ and monotone decreasing is needed to be found. In the half shaft parts of improved cotangent function $y = 2 / \pi \times \text{arc cot } x$, the X axis value range is $[0, +\infty)$, Y axis value range is $(0, 1]$, The function is monotone decreasing during value range which can meet the needs perfectly and simply calculated. So the improved half shaft cotangent function is used to redistribute distance and then calculate equivalence factor. The curve of function is in figure 1.

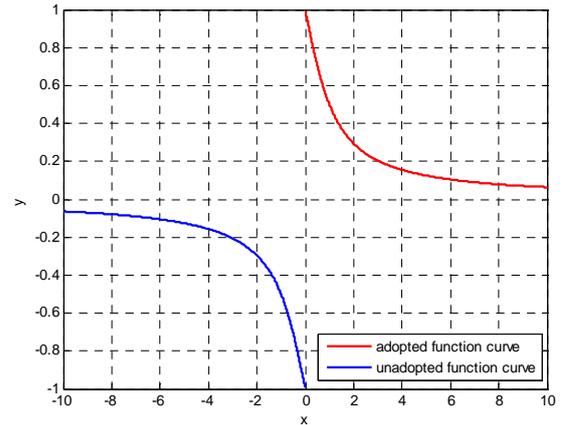


Fig. 1 improved cotangent function curve

The distance equivalence factor after redistribution M_{ij} is as follows:

$$M_{ij} = \frac{2 / \pi \times \text{arc cot } d_{ij}^2}{\sum_{x_n \in C_i} (2 / \pi \times \text{arc cot } d_{in}^2)} \quad (6)$$

Using equilibrium factor can dynamically adjust the fixed weight and gain adaptive weight, the farther between data object and clustering center is, the smaller equivalence factor, the adaptive weight which is less than fixed weight is got via using equivalence factor to adjust fixed weight. the nearer between data object and clustering center is, the bigger equivalence factor, the adaptive weight which is greater than fixed weight is got via using equivalence factor to adjust fixed weight. The value of adaptive weight affects compact degree between data objects and cluster center.

RFCM clustering algorithm based on adaptive weight. Iterative formula of RFCM clustering algorithm based on adaptive weight:

$$v_i = \begin{cases} A & \underline{C}_i \neq \emptyset \wedge (\overline{C}_i - \underline{C}_i) \neq \emptyset \\ B & \underline{C}_i = \emptyset \wedge (\overline{C}_i - \underline{C}_i) \neq \emptyset \end{cases} \quad (7)$$

Where equivalence factor is M_{ij} when $x_j \in \underline{C}_i$ and equivalence factor is M'_{ij} when $x_j \in (\overline{C}_i - \underline{C}_i)$.

The procedures of RFCM clustering algorithm based on adaptive weight are as follows:

Step1: initialization. Select data set whose amount is N ; set number k of objects clustering, value w_u of the upper approximation set, value w_l of the lower approximation set, judgment threshold τ of membership, convergence threshold ε of membership, fuzzy index m and initial membership u_{ij} ; calculate initial clustering center v_i ;

Step2: calculate data object x_j until the maximum value u_{ij} ($1 \leq i \leq k$) and secondary value u_{ij} ($1 \leq l \leq k$) of each membership; $x_j \in \underline{C}_i$ when $u_{ij} - u_{lj} \leq \tau$, or $x_j \in \underline{C}_i$;

Step3: calculate equivalence factor of data object x_j with formula (6);

Step4: update clustering center with formula (2), (3), (7) and fuzzy membership function with formula (4), (5);

Step5: repeat step 2, step 3 and step 4 until $u_{ij} < \varepsilon$ meets condition of convergence.

4 SIMULATION

In the experiment, UCI data set was firstly used to test algorithm's performance, then indexes of DB (Davies-Bouldin) and XB (Xie-Beni) were used to evaluate the effectiveness of algorithm (DAMES D L, 1979, XIE X L, 1991), at the same time, index of Rand was used to verify the accuracy of clustering's result. The smaller DB and XB shows better clustering's result and higher compactness. The greater Rand means higher clustering accuracy.

Use RCM clustering algorithm, RFCM clustering algorithm and RCFM (marked as SRCFM in experiment) clustering algorithm based on adaptive weight for clustering experiment. Clustering pa-

rameters adopted in the experiment of three algorithms are shown in table 1. Value τ_d is judgment threshold of distance, ε_d is convergence threshold of distance.

Tab.1 parameters of clustering

	k	m	w_u	w_l	τ	τ_d	ε	ε_d
RCM	3		0.1	0.9		0.2		0.01
RFCM	3	2	0.1	0.9	0.2		0.01	
SRFCM	3	2	0.1	0.9	0.2		0.01	

UCI data set is used in experiments, the one is Iris data set and the other is Wine. Adapting three clustering algorithm to do clustering experiment with each UCI data set. Clustering results are shown in table 2 and table 3. Results shows that SRFCM clustering algorithm has better performance than RCM and RFCM. Although the running time has increased, there is no great sacrifice, and verify the effectiveness of SRFCM clustering algorithm.

Tab.2 clustering's result of Iris data set

algorithm	XB index	DB index	Rand index	run time/s
RCM	0.1275	0.8264	0.872	0.047
RFCM	0.1218	0.6353	0.879	0.119
SRFCM	0.1141	0.5985	0.884	0.146

Tab.3 clustering's result of Wine data set

algorithm	XB index	DB index	Rand index	run time/s
RCM	0.2947	2.9310	0.941	0.127
RFCM	0.2618	2.1450	0.947	0.156
SRFCM	0.2523	1.8655	0.958	0.189

Select simulation data of four 3 cm airborne pulse doppler radar from the tab.1 in literature 1. Respectively adopt RFCM clustering algorithm and SRFCM clustering algorithm to cluster, sort simulation data in table 4, and both use 50 times Monte - Carlo experiments. Set number of clustering as four and initial clustering center is selected at random. Change of τ has great influence on clustering result. Set the range of τ is [0,1], in order to check the influence, various τ is used to do clustering and sorting experiment, while the other parameters are constant as table 1. Select sorting accuracy to evaluate clustering and sorting result. The results are shown in figure 2.

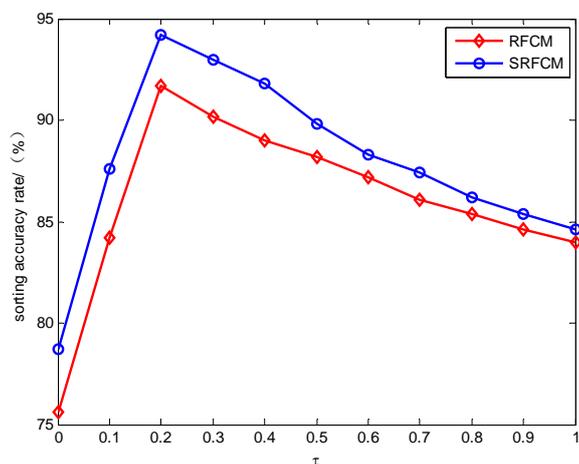


Fig.2 the result of clustering and sorting

Figure 2 shows, on the various τ values, SRFCM clustering algorithm has higher than RFCM clustering algorithm has achieved higher sorting accuracy than RFCM clustering algorithm, which proves the practicability of SRFCM clustering algorithm. When τ is increased to a certain value, the sorting accuracy rate begin to fall. So an appropriate τ value must be set when choose SRFCM clustering algorithm for radar signal sorting.

5 SUMMARY

This paper aims at the problem that different data objects are uniformly given fixed weight in RFCM clustering algorithm, RFCM clustering algorithm based on adaptive weight is put forward. Equivalence factor is calculated via redistribution's distance after choosing the improved inverse cotangent function, fixed weight is dynamically adjusted with equivalence factor to get adaptive weight, and then run the clustering arithmetic until clustering result is stable. Simulation shows that RFCM clustering algorithm based on adaptive weight has more efficient clustering results on UCI data set, which is applied to sort radar signal and achieves higher separation accuracy.

REFERENCES

- DAMES D L, BOULDIN D W. A cluster separation measure [J]. IEEE Transactions on Pattern Analysis and Machine intelligence, 1979, 1(2): 224-227.
- HUBERT L, ARABIE P. Comparing partitions [J]. Journal of Classification, 1985, 2(2/3): 193-218.
- MITRA S, BANKA H, PEDRYCZ W. Rough-fuzzy collaborative clustering [J]. IEEE Transactions on Systems, Man and Cybernetics; Part B, 2006, 36(4):795-805.
- MAJI P, PAI. S K. Rough set based generalized fuzzy C-means algorithm and quantitative indices [J]. IEEE Transactions on Systems, Man, and Cybernetics; Part B, 2007, 37(6):1529-1540.
- Qiang GUO, Chang-Hong WANG, Li-Min GUO, et al. Application of Segment Clustering in Radar Signal Sorting [J].