A Type-2 Fuzzy Subtractive Clustering Algorithm

Long Thanh Ngo and Binh Huy Pham

Department of Information Systems, Faculty of Information Technology,
Le Quy Don Technical University, No 100, Hoang Quoc Viet, Hanoi, Vietnam
{ngotlong,huybinhhvhc}@gmail.com

Abstract. The paper introduces a new approach to subtractive clustering algorithm (SC) with the fuzzifier parameter m which controls the clustering results in SC. And to manage the uncertainty of the parameter m, we have expanded the SC algorithm to interval type-2 fuzzy subtractive clustering algorithms (IT2-SC) using two fuzzifiers parameters m_1 and m_2 which creates a footprint of uncertainty (FOU) for the fuzzifier. The experiments are done based on image segmentation with the statistics show that the depends greatly on the parameter m of SC and stability and accuracy of our IT2-SC.

Keywords: Subtractive clustering, type-2 fuzzy subtractive clustering, type-2 fuzzy sets.

1 Introduction

Clustering is a common technique in many areas such as data mining, pattern recognition, image processing,.... Clustering is the division of data into clusters so that objects in a cluster have the biggest similarity. In particular, many clustering algorithms have been studied and applied, including k-mean [14], fuzzy c-mean [13] and mountain clustering [12]. However, in most real data exists uncertainty and vaguenesses which cannot be appropriately managed by type-1 fuzzy sets. Meanwhile, type-2 fuzzy sets allows us to obtain desirable results in designing and managing uncertainty. Therefore, type-2 fuzzy sets are studied and widely applied in many fields [7], [8], [9], especially pattern recognitions. On that basis, clustering algorithms has been extended and developed into type-2 fuzzy clustering algorithms to identify FOU of fuzzifiers, resulting to managing uncertainty is better.

Subtractive clustering algorithm, proposed in 1993 by Chiu [1],[2], is an extension of the mountain clustering methods by improving the mountain function to calculate potential of becoming a cluster center for each data point based on the location of the data point with all the other data points. The subtractive clustering algorithm only consider data points, not a grid points, which reduces the computational complexities and gives better distribution of cluster centers in comparison with the Mountain clustering algorithm and other algorithms. By 2005, Kim et al. have improved subtractive clustering algorithm by proposing

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 395–402.
springerlink.com

a kernel-induced distance instead of the conventional distance when calculating the mountain value of data point [6]. In 2008, J. Chen et al proposed a weighted mean subtractive clustering [5].

In subtractive clustering algorithms,except uncertainty have been created in the data, setting subtractive clustering's parameters are very influential to the results of clustering [3],[4]. So in this paper, we extend the subtractive clustering algorithm by putting a fuzzifier parameter m into mountain function to caculate for data points. The fuzzifier parameter m can alter the results of the mountain function so it has great influence to the results of clustering. Through fuzzifier parameter m, we can reduce the dependence of clustering results in initial values for the parameters of the algorithm. As with the adjustment parameter m, we can obtain better clustering results without interesting into setting the initial values of the parameters of the algorithm. Therefore, fuzzifier parameter m created uncertainties for the SC. To design and manage uncertainties of fuzzifier m , we extend to interval type-2 fuzzy subtractive clustering algorithm (IT2-SC) by using two fuzzifiers m_1 and m_2 which creates a footprint of uncertainty (FOU) for the fuzzifier m. Then, comparative experiments between IT2-SC and other subtractive clustering to showing the validity of our proposed method.

The remainder of this paper is organized as follows. In Section 2.1 introduces briefly type-2 fuzzy set and interval type-2 fuzzy set. Subtractive clustering is presented in section 2.2. In section 3, we discuss how to extend subtractive clustering method with fuzzifier parameter m and propose interval type-2 fuzzy subtractive clustering algorithm. In section 4, we provide several experiments showing the validity of our proposed method. Finally, Section 5 gives the summary and conclusions.

2 Preliminaries

2.1 Type-2 Fuzzy Sets

A type-2 fuzzy set in X is denoted $\tilde{A}$, and its membership grade of $x \in X$ is $\mu_{\tilde{A}}(x, u), u \in J_x \subseteq [0, 1]$, which is a type-1 fuzzy set in [0, 1]. The elements of domain of $\mu_{\tilde{A}}(x, u)$ are called primary memberships of x in $\tilde{A}$ and memberships of primary memberships in $\mu_{\tilde{A}}(x, u)$ are called secondary memberships of x in $\tilde{A}$.

Definition 1. *A $type - 2$ fuzzy set, denoted $\tilde{A}$, is characterized by a type-2 membership function $\mu_{\tilde{A}}(x, u)$ where $x \in X$ and $u \in J_x \subseteq [0, 1]$, i.e.,*

$$\tilde{A} = \{((x, u), \mu_{\tilde{A}}(x, u)) | \forall x \in X, \forall u \in J_x \subseteq [0, 1]\} \tag{1}$$

in which $0 \leq \mu_{\tilde{A}}(x, u) \leq 1$.

At each value of x, say $x = x'$, the 2-D plane whose axes are u and $\mu_{\tilde{A}}(x', u)$ is called a *vertical slice* of $\mu_{\tilde{A}}(x, u)$. A *secondary membership function* is a vertical slice of $\mu_{\tilde{A}}(x, u)$. It is $\mu_{\tilde{A}}(x = x', u)$ for $x \in X$ and $\forall u \in J_{x'} \subseteq [0, 1]$, i.e.

$$\mu_{\tilde{A}}(x = x', u) \equiv \mu_{\tilde{A}}(x') = \int_{u \in J_{x'}} f_{x'}(u)/u, J_{x'} \subseteq [0, 1] \tag{2}$$

in which $0 \leq f_{x'}(u) \leq 1$.

Type-2 fuzzy sets are called an interval type-2 fuzzy sets if the secondary membership function $f_{x'}(u) = 1\ \forall u \in J_x$ i.e. a type-2 fuzzy set are defined as follows:

Definition 2. *An interval type-2 fuzzy set $\tilde{A}$ is characterized by an interval type-2 membership function $\mu_{\tilde{A}}(x, u) = 1$ where $x \in X$ and $u \in J_x \subseteq [0, 1]$, i.e.,*

$$\tilde{A} = \{((x, u), 1) | \forall x \in X, \forall u \in J_x \subseteq [0, 1]\} \tag{3}$$

Uncertainty of $\tilde{A}$, denoted FOU, is union of primary functions i.e. $FOU(\tilde{A}) = \bigcup_{x \in X} J_x$. Upper/lower bounds of membership function (UMF/LMF), denoted $\overline{\mu}_{\tilde{A}}(x)$ and $\underline{\mu}_{\tilde{A}}(x)$, of $\tilde{A}$ are two type-1 membership function and bounds of FOU.

2.2 Subtractive Clustering Algorithm

Subtractive clustering finds the optimal data point to define a cluster center based on the density of surrounding data points, see [1,2] for more details. Consider a collection of n data points: $X = \{x_1, x_2, ..., x_{\mathrm{n}}\}$, where, x_i is a vector in an M-dimensional space. The subtractive clustering algorithm includes the following steps:

Step 1: Initialization, r_a, η with $\eta = \frac{r_b}{r_a}$, $\overline{\varepsilon}$ and $\underline{\varepsilon}$
Step 2: Caculating density for all data points by using formula bellow:

$$P_i = \sum_{j=1}^{n} e^{-\frac{4}{r_a^2}\|x_i - x_j\|^2} \tag{4}$$

Where P_i denotes the density of ith data point, r_a is a positive constant defining a neighborhood radius and $\|.\|$ denotes the Euclidean distance.
Data point with the highest density is selected as the first cluster center.
Step 3: The density of all data points is revised by using formula bellow:

$$P_i = P_i - P_k^* e^{-\frac{4}{r_b^2}\|x_i - x_k^*\|^2}; i = 1, ..., n \tag{5}$$

Where r_b is is a positive constant and $r_b = \eta * r_a$ with a good choise is $\eta = 1.5$.
Step 4: Let x^* is a data point with its density is highest and equal P^*.

-If $P^* > \overline{\varepsilon} P^{ref}$: x^* is a new cluster center and back to Step 3.

- Else if $P^* < \underline{\varepsilon} P^{ref}$: back to Step 5.

- Else:

+ Let d_{min} is shortest of the distances between x^* and all previously found cluster centers.

+ If $\frac{d_{\min}}{r_a} + \frac{P^*}{P^{\mathrm{ref}}} \geq 1$: x^* is a new cluster center and back to Step 3.

+ Else: $P(x^*) = 0$ and select x^* with the next highest density, $P(x^*)$, v back to step 4.

Step 5: Output the results of clustering.

When the membership degree of data point in each cluster is as follows:

$$\mu_{ik} = e^{-\frac{4}{r_a^2}\|x_i - x_k\|^2} \tag{6}$$

3 Interval Type-2 Fuzzy Subtractive Clustering

3.1 Extending Subtractive Clustering Algorithm

In the subtractive clustering algorithm, we must set four parameters: acceptance ratio $\overline{\varepsilon}$, reflection ratio $\underline{\varepsilon}$, cluster radius r_a and squash factor η (or r_b). The choice of parameters have greatly influences to results of clustering. If values of $\overline{\varepsilon}$ and $\underline{\varepsilon}$ are large, the number of cluster centers will be reduced. The influences of the four parameters to clusters have been described detail in papers of Demirli [3],[4]. Therefore, these parameters are uncertainties in the subtractive clustering algorithm.

On the other hand, subtractive clustering estimated the potential of a data point as a cluster center based on the density of surrounding data points, which is actually based on the distance between the data point with the remaining data points. Therefore, SC includes various types of uncertainty as distance measure, parameters Initialization...So we consider a fuzzifier parameter that control the distribution of data points into clusters by making the parameter m in the density function to calculate the potential of a data point as follows:

$$P_i = \sum_{j=1}^{n} e^{-\frac{4}{r_a^2}(x_j - x_i)^{\frac{2}{m-1}}} \tag{7}$$

If x_k is the k^{th} cluster position, has potential P_k^*, then the potential of each data point is revised by the following formula:

$$P_i = P_i - P_k^* e^{-\frac{4}{r_b^2}(x_i - x_k)^{\frac{2}{m-1}}}; i = 1, ..., n \tag{8}$$

Then the choice of the value of the parameter m has greatly influence to results of clustering. If m is small, the number of cluster centers will be reduced. Conversely, If m is too large, too many cluster centers will be generated. In addtion, through the adjustment of fuzzifier parameters m, it is easy to obtain the good results of clustering that is not dependent on the setting of the parameters for SC.

Fig.1 illustrates the influence of fuzzifier parameters m to initial parameters of SC. In Fig.1(a) represents the best results of clustering with the value of the initial parameters are $\overline{\varepsilon} = 0.5$, $\underline{\varepsilon} = 0.15$, $r_a = 0.25$, $\eta = 1.5$, respectively (Chiu [1,2]). In Fig.1(b) describes the results of clustering of expansion SC algorithm with the value of the initial parameters are $\overline{\varepsilon} = 0.5$, $\underline{\varepsilon} = 0.15$, $r_a = 0.4$, $\eta = 1.35$ and $m = 2.47$, respectively. We see that the result of clustering in the two cases are quite similar.

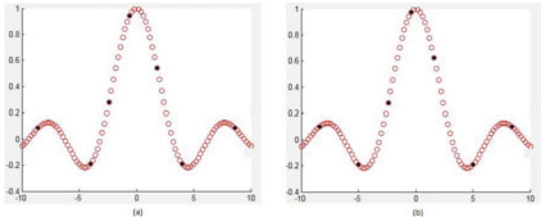

Fig. 1. The influence of fuzzifier parameters m to initial parameters of SC

3.2 Interval Type-2 Fuzzy Subtractive Clustering Algorithm

In expansion subtractive clustering algorithm, the degree of membership of a data point in k^{th} cluster center is defined as following formula:

$$\mu_{ik} = e^{-\frac{4}{r_a^2}(x_i - x_k)^{\frac{2}{m-1}}} \tag{9}$$

Where x_k is k^{th} cluster center According to formula 9, the value of membership of a data point in k^{th} cluster center depends on the position of k^{th} cluster and the fuzzifier parameter m. On the other hand, the position of the k^{th} cluster also depends on the fuzzifier parameter m.Thus, the fuzzifier parameter m is the most uncertainty element in expansion subtractive clustering algorithm. Therefore, to design and manage the uncertainty for fuzzifier parameter m , we extend a pattern set to interval type-2 fuzzy sets using two fuzzifiers m_1 and m_2 which creates a footprint of uncertainty (FOU) for the fuzzifier parameter m. Then the degree of membership of k^{th} cluster center is defined as following formula:

$$\begin{cases} \overline{\mu}_{ik} = e^{-\frac{4}{r_a^2}(x_i - x_k)^{\frac{2}{m_1-1}}} \\ \underline{\mu}_{ik} = e^{-\frac{4}{r_a^2}(x_i - x_k)^{\frac{2}{m_2-1}}} \end{cases} \tag{10}$$

Thence, we have two density functions to caculate potential of each data point as bellows:

$$\begin{cases} \overline{P}_i = \sum_{j=1}^{n} e^{-\frac{4}{r_a^2}(x_j - x_i)^{\frac{2}{m_1-1}}} \\ \underline{P}_i = \sum_{j=1}^{n} e^{-\frac{4}{r_a^2}(x_j - x_i)^{\frac{2}{m_2-1}}} \end{cases} \tag{11}$$

If the centroids are identified by the formula (11), we will have centroids v_L and v_R. Thus, we will do type reduction for centroids as bellows:

$$P_i = \frac{\overline{P}_i * m_1 + \underline{P}_i * m_2}{m_1 + m_2} \tag{12}$$

And when we identified k^{th} cluster center, the density of all data points is revised by using following formula:

$$\begin{cases} \underline{P}_i^{sub} = P_k^* \sum\limits_{j=1}^{n} e^{-\frac{4}{r_b^2} d_{\text{ij}}^{\frac{2}{m_1-1}}} \\ \overline{P}_i^{sub} = P_k^* \sum\limits_{j=1}^{n} e^{-\frac{4}{r_b^2} d_{\text{ij}}^{\frac{2}{m_2-1}}} \\ P_i^{sub} = \frac{\underline{P}_i^{sub} * m_1 + \overline{P}_i^{sub} * m_2}{m_1 + m_2} \\ P_i = P_i - P_i^{sub} \end{cases} \tag{13}$$

The interval type-2 fuzzy subtractive clustering algorithm includes the following steps:

Step 1: Initialization, r_a, η with $\eta = \frac{r_b}{r_a}$, $\overline{\varepsilon}$ and $\underline{\varepsilon}$, m_1 and m_2 $(1 < m_1 < m_2)$
Step 2: Calculating density for all data points with two fuzzifiers m_1 and m_2 by using formulas (11) and (12). Data point with the highest density is selected as the first cluster center: $P_k^* = \max\limits_{i=1}^{n} P_i$ where $k = 1$ and P_k^* is the density of the first cluster center.
Step 3: The density of all data points is revised by using formula (13).
Step 4: The identification of the next cluster centers are as similar as SC.
Step 5: Output the results of clustering.

4 Application to Image Segmentation

In this section, we compare the results between subtractive clustering algorithm and interval type-2 fuzzy subtractive clustering algorithm through the tests of image segmentation. Here, multiple levels of image segmentation is a combination between FCM clustering algorithm and subtractive clustering algorithms. In particular, we use the result of SC or IT2-SC to initialize cluster centers for FCM algorithm, then through FCM given the results of simage segment and iterations of the FCM. Finally, we compare the validity between SC and our proposed IT2-SC on based the results of image segmentation and iterations of the FCM. The steps are as follows:

Step 1: Using subtractive clustering algorithm (in section 3.1) or interval type-2 fuzzy subtractive clustering algorithm (in section 3.2) initializes for the initial cluster centers matrix $V^{(0)} = [v_{ij}], V^{(0)} \in R^{d \times c}$.
Step 2: Using FCM clustering with cluster center matrix is initialized in step 1 and gives results of clustering $V^{(j)}$ and iterations of FCM j.
Step 3: Comparison of iterations to be performed and the results of clustering to rate the effectiveness of subtractive clustering algorithms.

As the result of experiments, IT2-SC makes clustering results more efficiently than SC when the results of its clustering applied into FCM to converge quickly to clustering results of FCM with better clusters and the number of iterations to perform also better than the SC.

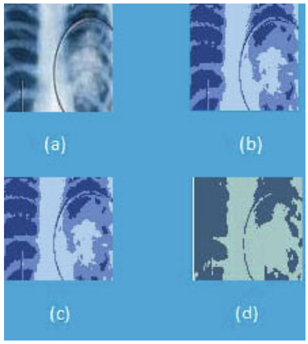

Fig. 2. Segmented results for chest trauma image for SC and IT2-SC

In this experiment, we conduct image segmentation for image of chest trauma area. In Fig.2(a), we only consider to contain two regions, namely, bone and damaged domain.

Tests was conducted with the original initialization parameters, respectively, $\overline{\varepsilon} = 0.5$, $\underline{\varepsilon} = 0.15$, $r_a = 0.25$ and $\eta = 1.25$. Fig.2(b) shows the image segmentation result of SC and FCM ($m = 2$) and Fig.2(c) shows the image segmentation result of IT2-SC and FCM with $m_1 = 1.65$ and $m_2 = 2.35$. In both of these test results, image is divided into three quite distinct regions. Fig.2(d) shows the best result of image segmentation with two distinct regions in which we easily observe the damaged area.

The statistical results of the comparison between two algorithms in Tab. 4 shows that the IT2-SC algorithm gives better results than the SC algorithm. The results show that FCM terminals in 38 iterations with FSC (m=2) whereas FCM terminals in 22 or 20 iterations with IT2-SC.

Table 1. Comparison between FSC and IT2-FSC

Algorithm	$\overline{\varepsilon} = 0.5, \underline{\varepsilon} = 0.15, r_a = 0.5, \eta = 1.25$		
	Fuzzifier parameter	Iterations of FCM	Number of clusters
FSC	m=2	38	3
IT2-FSC	$m_1 = 1.65$ $m_2 = 2.35$	22	3
	$m_1 = 1.42$ $m_2 = 2.58$	20	2

5 Conclusion

The article presents a new approach for subtractive clustering algorithm with fuzzifier parameter m which has a great influence on the clustering process. Through fuzzifier parameter m, we can reduce the uncertainty of the algorithm in the initialization for parameters. Since then, we have expanded into interval type-2 fuzzy subtractive clustering algorithm by using two different values of fuzzifier parameter m. Through experiments on image segmentation have shown the effectiveness of our proposed IT2-SC.

For the future works, we will apply our proposed IT2-SC method to model type-2 and interval type-2 TSK fuzzy logic systems by the construction of the type-2 and interval type-2 fuzzy rules.

Acknowledgments. This paper is sponsored by Vietnam's National Foundation for Science and Technology Development (NAFOSTED).

References

1. Chiu, S.L.: Fuzzy Model Identification Based on Cluster Estimation. Journal on Intelligent Fuzzy Systems 2, 267–278 (1994)
2. Chiu, S.L.: Extracting Fuzzy Rules from Data for Function Approximation and Pattern Classification. In: Dubois, D., Prade, H., Yager, R.R. (eds.) Fuzzy Information Engineering: a Guide Tour of Applications, pp. 149–162. Wiley (1997)
3. Demirli, K., Cheng, S.X., Muthukumaran, P.: Subtractive Clustering Based Modeling of Job Sequencing with Parametric Search. Fuzzy Sets and Systems 137, 235–270 (2003)
4. Demirli, K., Muthukumaran, P.: Higher Order Fuzzy System identification Using Subtractive Clustering. J. of Intelligent and Fuzzy Systems 9, 129–158 (2000)
5. Chen, J.Y., Qin, Z., Jia, J.: A Weighted Mean Subtractive Clustering Algorithm. Information Technology Journal 7(2), 356–360 (2008)
6. Kim, D.W., Lee, K.Y., Lee, K.H.: A Kernel Based Subtractive Clustering Method. Pattern Recognition Letter 26(7), 879–891 (2005)
7. Mendel, J.M., John, R.I., Liu, F.: Interval Type-2 Fuzzy Logic Systems Made Simple. IEEE Trans. on Fuzzy Systems 14(6), 808–821 (2006)
8. Karnik, N., Mendel, J.M.: Operations on Type-2 Fuzzy Sets. Fuzzy Sets and Systems 122, 327–348 (2001)
9. Liang, Q., Mendel, J.M.: Interval Type-2 Fuzzy Logic Systems: Theory and Design. IEEE Trans. on Fuzzy Systems 8(5), 535–550 (2000)
10. Hwang, C., Rhee, F.C.H.: Uncertain Fuzzy clustering: Interval Type-2 Fuzzy Approach to C-means. IEEE Trans. on Fuzzy Systems 15, 107–120 (2007)
11. Zarandi, M.H.F., Zarinbal, M., Trksen, I.B.: Type-II Fuzzy Possibilistic C-Mean Clustering. In: IFSA/EUSFLAT Conf., pp. 30–35 (2009)
12. Yager, R.R., Filev, D.P.: Approximate clustering via the mountain method. IEEE Trans. on Systems, Man and Cybernetics 24(8), 1279–1284 (1994)
13. Bezdek, J.: Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers (1981)
14. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y.: An Efficient k-Means Clustering Algorithm: Analysis and Implementation. IEEE Trans. Pattern Anal. Mach. Intell., 881–892 (2002)