Journal of Information Hiding and Multimedia Signal Processing ©2018 ISSN 2073-4212
Ubiquitous International Volume **9**, Number **2**, March **2018**

C-means Clustering Algorithm Based on Intuitionistic Fuzzy Sets and Its Application in Satisfaction Evaluation

Xiao-Guang Zhou[1]*, Mi Lu[2] and Xiao-Xia Huang[1]

[1]Donlinks School of Economics and Management
University of Science and Technology Beijing
30 Xueyuan Road, Haidian District, Beijing, P.C 100083 - China
xiaoguang@ustb.edu.cn; mlu@ece.tamu.edu; hxiaoxia@manage.ustb.edu.cn

[2]Department of Electrical and Computer Engineering
Texas A&M University
188 Bizzell St, College Station, Texas, 77840 - USA
*Corresponding author: xiaoguang@ustb.edu.cn (X. Zhou)

Received September, 2017; revised December, 2017

Abstract. *Due to the uncertainty and fuzziness of information, the traditional clustering analysis method sometimes cannot meet the requirement in practice. The clustering method based on intuitionistic fuzzy set has attracted more and more scholars' attention nowadays. This paper discusses the intuitionistic fuzzy C-means clustering algorithm. The partition matrix is initialized by given conditions, and the cluster center matrix is obtained through the iterative computation between the object matrix and the partition matrix. The final results are achieved according to the membership degrees and non-membership degrees of the objects to the partition matrix. Several important parameters during the intuitionistic fuzzy C-means clustering process, such as the initial form of the partition matrix, the number of classification and the threshold of terminating the iteration, which significantly affect the clustering results, are analyzed and discussed. Finally, a case of customer satisfaction evaluation is illustrated by the intuitionistic fuzzy C-means clustering method, and the method is compared with the fuzzy C-means clustering method as well.*

Keywords: Intuitionistic fuzzy set, Fuzzy clustering, C-means algorithm, Satisfaction evaluation

1. **Introduction.** Since human judgments including preferences are often vague and they cannot express their preferences with exact numerical values. A more realistic approach is to use the fuzzy values instead of the exact numerical values. Atanassov [1] introduced the definition and operation rules of intuitionistic fuzzy set. Intuitionistic fuzzy set assigns to each element in the universe both a membership degree and a non-membership degree, thus relaxing the enforced duality from fuzzy set theory. Intuitionistic fuzzy set can describe the ambiguity and uncertainty of problems better than that of fuzzy set.

Clustering analysis is an important branch of data mining. It classifies objects according to the characteristics of objects, the affinity levels and the similarity degrees. Pan et al. [2] presented a fast clustering algorithm for vector quantization. For traditional hard clustering, each sample is strictly assigned to a class. It is not applicable to the complicated and rapidly changing problems. Fuzzy mathematics provides a mathematical basis for fuzzy clustering. Fuzzy clustering can obtain the uncertainty degree of each object in the set. Khalilia et al. [3] proposed an improved algorithm of relational fuzzy C-means (FCM) clustering. Fuzzy clustering analysis has been widely used in many fields, such as truck backer-upper problem [4], supplier selection [5], multi-objective optimization [6], distributed picture fuzzy clustering [7] and forecasting enrollments [8].

In recent years, more and more scholars have studied the clustering method based on intuitionistic fuzzy sets. Xu et al. [9] defined the concepts of association matrix and the equivalent matrix of intuitionistic fuzzy sets, and firstly presented a clustering algorithm based on intuitionistic fuzzy sets. Wang et al. [10] proposed a netting method to make clustering analysis via intuitionistic fuzzy similarity matrix. To improve the accurateness of CT scan brain images, a new objective function which is the intuitionistic fuzzy entropy is incorporated in the conventional fuzzy C-means clustering algorithm [11]. Son et al. [12] proposed a novel clustering algorithm for geo-demographic analysis based on the results regarding intuitionistic fuzzy sets and the possibilistic fuzzy C-means. Zhao et al. [13] introduced the concepts of graph, minimum spanning tree, intuitionistic fuzzy set and intuitionistic fuzzy distance, and then presented an intuitionistic fuzzy and an interval-valued intuitionistic fuzzy minimum spanning tree clustering algorithms respectively. Xu et al. [14] defined two new methods of intuitionistic fuzzy similarity measures, and used them to construct the intuitionistic fuzzy similarity degree matrix, by which they presented a spectral algorithm to cluster intuitionistic fuzzy information. Wang et al. [15] analyzed the alternatives in multiple attribute decision making with intuitionistic fuzzy triangle product, and constructed an intuitionistic fuzzy similarity matrix by the intuitionistic fuzzy square product, and then developed an intuitionistic fuzzy clustering analysis method. Son [16] presented a novel fuzzy clustering algorithm named as Kernel Fuzzy Geographically Clustering that utilized both the kernel similarity function and the new update mechanism of the spatial interaction – modification model. Thong and Son [17] proposed a hybrid model between picture fuzzy clustering and intuitionistic fuzzy recommender systems for medical diagnosis called as Hybrid Intuitionistic Fuzzy Collaborative Filtering. Verma et al. [18] developed an improved intuitionistic fuzzy C-means (IFCM) algorithm, which considered the local spatial information in the intuitionistic fuzzy way. A rough set- based intuitionistic fuzzy C-means clustering algorithm is proposed for the segmentation of the magnetic resonance brain images in reference [19]. A novel similarity measure of intuitionistic fuzzy set is presented based on the equivalence relation in the intuitionistic fuzzy set - interpolative Boolean algebra approach. The proposed similarity measure can be combined with various intuitionistic fuzzy aggregation operators [20]. D'Urso [21] presented a systematic literature review of different uncertainty-based clustering approaches -i.e. fuzzy clustering, rough set-based clustering, intuitionistic fuzzy clustering, type-2 fuzzy clustering, and picture fuzzy clustering.

Intuitionistic fuzzy C-means algorithm is a clustering algorithm for Euclidean distances among sample points in Euclidean space. It is the extension of fuzzy C-means clustering algorithm. According to whether the cluster center, the clustered objects and the relationships between them are intuitionistic fuzzy sets or not, IFCM method can generally be divided into three types: (1) Only the cluster centers are represented by intuitionistic fuzzy sets. (2) The relationships between clustered objects and cluster centers are extended to intuitionistic fuzzy sets, while the clustered objects are denoted by common sets. (3) The clustered objects, the cluster centers and the relationships between them are represented by intuitionistic fuzzy sets simultaneously.

In the existing studies, although some literatures discussed the clustering method when the clustered objects, cluster centers and the relationships between them are intuitionistic fuzzy sets simultaneously. There is no discussion about the factors that influence the intuitionistic fuzzy clustering results. Therefore, this paper makes an in-depth study on the intuitionistic fuzzy $C-$means clustering algorithm. Several important parameters during the IFCM clustering process, such as the initial form of the partition matrix, the number of classification and the threshold of terminating the iteration, which significantly affect the clustering results, are analyzed and discussed. A case of customer satisfaction evaluation is given by the intuitionistic fuzzy $C-$means clustering method, and the method is compared with the fuzzy C-means clustering method as well.

2. **Preliminary knowledge.**

2.1. **Intuitionistic fuzzy set.**

Definition 1 Let $X = \{x_1, x_2, \ldots, x_n\}$ be a fixed non-empty universe set, an intuitionistic fuzzy set A in X is defined as $A = \{< x, \mu_A(x), v_A(x) \,|\, x \in X >\}$, which is characterized by a membership function $\mu_A(x){:}X \to [0,1]$ and a non-membership function $v_A(x){:}X \to [0,1]$ with the condition $0 \le \mu_A(x) + v_A(x) \le 1$ for all $x \in X$, where μ_A and v_A represent, respectively, the degree of membership and non-membership of element x to set A [1].

In addition, for all $x \in X$, $\pi_A(x) = 1$ - $\mu_A(x) - v_A(x)$ denotes the hesitation degree of the element x to the set A. Especially, if $\pi_A(x) = 0$, for all $x \in X$, then the intuitionistic fuzzy set A is reduced to a fuzzy set.

2.2. **Fuzzy *C*-partition.**

The fuzzy C-partition is the generalization of the ordinary C-partition, which extends the sample space from the ordinary set to the fuzzy set. The definition of fuzzy C-partition is as follows.

Definition 2 Let$A = \{A_1, A_2, \ldots, A_n\}, A_j = \{A_{j1}, A_{j2}, \ldots, A_{jm}\}$,

$$R = \begin{pmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{c1} & \cdots & r_{cn} \end{pmatrix}, (r_{ij} \in [0,1], i = 1, 2, \ldots, c, j = 1, 2, \ldots, n)\,,$$

if the fuzzy matrix R satisfies the following two conditions, then R is called fuzzy C-partition matrix of A.

(1) $\sum\limits_{i=1}^{c} r_{ij} = 1, (j = 1, 2, \ldots, n)$, which indicates that the sum of each sample A_j belongs to all fuzzy subsets R_i is 1.

(2) $0 < \sum\limits_{j=1}^{n} r_{ij} < n, (i = 1, 2, \ldots, c)$, which indicates that each R_i is between the empty set and the universal set.

A fuzzy C-partition matrix gives a way of the fuzzy C-partition of matrix A. Matrix A can be divided into c fuzzy subsets by fuzzy C-partition matrices, named as V_1, V_2, …, V_c. The whole fuzzy C-partition is fuzzy C-partition space.

Let

$$V_i = \frac{\sum\limits_{j=1}^{n} (r_{ij})^q A_j}{\sum\limits_{j=1}^{n} (r_{ij})^q},$$

then V_i is the cluster center of the ith class.

Let

$$J(R, V) = \sum_{i=1}^{c} \sum_{j=1}^{\mathrm{n}} (r_{ij})^q D^2(V_i, A_j),$$

where q >1, $D^2(V_i, A_j)$ is the distance from A_j to the cluster center V_i. If $J(R, V)$ is the minimum, then R is the ideal fuzzy C-partition.

2.3. **The fuzzy *C*-means algorithm.** Suppose that the set of clustered objects $A = \{A_1, A_2, \ldots, A_n\}$, and each object A_j has m characteristic indicators, that is, $A_j = \{x_{j1}, x_{j2}, \ldots, x_{jm}\}$. A will be divided into c types, and the cluster center vectors $V = \{V_1, V_2, \ldots, V_n\}^T$, $V_i = \{v_{i1}, v_{i2}, \ldots, v_{im}\}$, i =1, 2, ..., c. The FCM algorithm is as follows.

Step 1: Select c and the initial value $R(0)$, $2 \le c \le n$, $R(0)$ is the initial fuzzy C-partition matrix.

Step 2: Calculate the cluster center matrix $V^{(l)}$, l =0, 1, 2, ...,

$$V^{(l)} = \begin{pmatrix} V_1^{(l)} \\ \vdots \\ V_c^{(l)} \end{pmatrix} = \begin{pmatrix} V_{11}^{(l)} & \cdots & V_{1m}^{(l)} \\ \vdots & \ddots & \vdots \\ V_{c1}^{(l)} & \cdots & V_{cm}^{(l)} \end{pmatrix},$$

where

$$V_i^{(l)} = \frac{\sum_{j=1}^{n} (r_{ij}^{(l)})^q A_j}{\sum_{j=1}^{n} (r_{ij}^{(l)})^q}.$$

Step 3: Revise $R^{(l)}$,

$$r_{ij}^{(l+1)} = \left(\sum_{k=1}^{c} \left(\frac{D(V_i^{(l)}, A_j)}{D(V_k^{(l)}, A_j)}\right)^{\frac{2}{q-1}}\right)^{-1}.$$

Step 4: Compare $R^{(l)}$ with $R^{(l+1)}$, if for a given $\varepsilon > 0$, $D^2(R^{(l)}, R^{(l+1)}) \le \varepsilon$, then stop; otherwise, let $l = l$+1, return step 2.

3. ***C*-means clustering algorithm based on intuitionistic fuzzy sets.** In this paper, the clustered objects, the cluster centers, and the relationships between them are all expressed by intuitionistic fuzzy sets. The clustered objects are represented as

$$A_j = (\langle \mu_{Aj}(x_1), \gamma_{Aj}(x_1)\rangle, \langle \mu_{Aj}(x_2), \gamma_{Aj}(x_2)\rangle, \ldots, \langle \mu_{Aj}(x_m), \gamma_{Aj}(x_m)\rangle),$$
$$A_{jk} = (\mu_{Aj}(x_k), \gamma_{Aj}(x_k)), \ (1 \le j \le n, 1 \le k \le m).$$

The cluster centers are expressed as

$$V_i = (\langle \mu_{Vi}(x_1), \gamma_{vi}(x_1)\rangle, \langle \mu_{Vi}(x_2), \gamma_{Vi}(x_2)\rangle, \ldots, \langle \mu_{Vi}(x_m), \gamma_{Vi}(x_m)\rangle),$$
$$V_{ik} = (\mu_{Vi}(x_k), \gamma_{Vi}(x_k)), (1 \le i \le c, 1 \le k \le m).$$

3.1. **The intuitionistic fuzzy *C*-means algorithm.** Intuitionistic fuzzy C-means clustering method belongs to the clustering method based on the objective function, and the result is obtained by optimizing the objective function. The clustering rule is to minimize the objective function value. It is classified by minimizing the square error of distance with iterative calculation. In order to avoid the generation of the ordinary solutions, the in-class error square and the objective function are extended to the weighted in-class error square and the weighted objective function [3]. Therefore, the objective function of the IFCM clustering algorithm is as follows:

$$\begin{cases} \min J(R,V) = \sum_{j=1}^{n} \sum_{i=1}^{c} \frac{(\mu_{ij})^q + (1-\gamma_{ij})^q}{2} D(A_j, V_i)^2 \quad q \in [1, \infty] \\ \mu_{ij} \in [0,1], \gamma_{ij} \in [0,1], \\ 0 < \sum_{k=1}^{n} \mu_{ik} \le n, \ 0 < \sum_{k=1}^{n} \gamma_{ik} \le n, \forall\, i, \forall\, k. \end{cases} \tag{1}$$

Where q indicates the smoothing parameter, usually $q = 2$; μ_{ij} is the membership degree of the clustered object j belongs to the class i, and γ_{ij} is the non-membership degree of j to the class i; $D(A_j,V_i)$ denotes the distance from the clustered objectA_j to the cluster center V_i, and the Euclidean distance formula [22] is adopted.

$$D(A_j, V_i) =(\sum_{k=1}^{m} w_k[\alpha(\mu_{Aj}(x_k) - \mu_{vi}(x_k))^2 + \beta(\gamma_{Aj}(x_k) - \gamma_{Vi}(x_k))^2 + \lambda(\pi_{Aj}(x_k) - \pi_{Vi}(x_k))^2])^{1/2}.$$

When the clustered objects are divided into different categories, the object A_j has a specific membership degree and non-membership degree for each type, that is, the possibilities of the object A_j belonging to the category are expressed by intuitionistic fuzzy sets. The intuitionistic fuzzy matrix for the possibilities is the association matrix between the clustered object set A and the cluster center V. Each clustering will generate an intuitionistic fuzzy association matrix, which is called the intuitionistic fuzzy partition matrix, named as $R = \mu_{ij}, \gamma_{ij} c \times n$, and R satisfies:

(1) $\mu_{ij} + \gamma_{ij} + \pi_{ij} = 1$;

(2) $\sum_{i=1}^{c} \mu_{ij} = 1$, which indicates that each clustered object must belong to one of the categories;

(3) $\sum_{j=1}^{n} \mu_{ij} > 1$, $(1 \leq i \leq c)$, which indicates that there are always some objects belong to the category with different levels.

The algorithm of intuitionistic fuzzy C-means is as below.

Step 1: Define the attribute indices matrix A of the clustered objects, and select the number of classification and the initial intuitionistic fuzzy partition matrix $R(0)$.

Step 2: Calculate the characteristic matrix of the cluster center $V(I)$ according to the intuitionistic fuzzy partition matrix $R(l)$, $V(l) = (V_{ik}(l))_{c\times m}$, $l = 0, 1, 2, \ldots$

Where,

$$\mu_{Vi}(x_k)^{(l)} = \frac{\sum_{j=1}^{n} \frac{(\mu_{ij}^{(l)})^q + (1-\gamma_{ij}^{(l)})^q}{2} \mu_{Aj}(x_k)}{\sum_{j=1}^{n} \frac{(\mu_{ij}^{(l)})^q + (1-\gamma_{ij}^{(l)})^q}{2}} \ (1 \leq i \leq c, 1 \leq k \leq m), \tag{2}$$

$$\gamma_{Vi}(x_k)^{(l)} = \frac{\sum_{j=1}^{n} \frac{(\mu_{ij}^{(l)})^q + (1-\gamma_{ij}^{(l)})^q}{2} \gamma_{Aj}(x_k)}{\sum_{j=1}^{n} \frac{(\mu_{ij}^{(l)})^q + (1-\gamma_{ij}^{(l)})^q}{2}} \ (1 \leq i \leq c, 1 \leq k \leq m). \tag{3}$$

Step 3: Revise the intuitionistic fuzzy partition matrix.

(1) For $\forall\, i$, $i = 1, 2, \ldots, c$, if $D(A_j, V_i) > 0$, then

$$\mu_{ij}^{(l+1)} = \frac{1}{\sum_{k=1}^{c} \left(\frac{D\left(A_j, V_i^{(l)}\right)}{D\left(A_j, V_k^{(l)}\right)} \right)^{\frac{2}{q-1}}} \ (1 \leq i \leq c, 1 \leq j \leq n), \tag{4}$$

$$\gamma_{ij}^{(l+1)} = \mu_{ij}^{(l)} + \gamma_{ij}^{(l)} - \frac{1}{\sum_{k=1}^{c} \left(\frac{D\left(A_j, V_i^{(l)}\right)}{D\left(A_j, V_k^{(l)}\right)} \right)^{\frac{2}{q-1}}} \ (1 \leq i \leq c, 1 \leq j \leq n). \tag{5}$$

(2) If $\exists k$, k =1, 2, ..., c, makes $D(A_j,V_i) = 0$, then

$$\begin{cases} \mu_{ij} = 1, \gamma_{ij} = 0, i = k; \\ \mu_{ij} = 0, \gamma_{ij} = 1, i \neq k. \end{cases}$$

Step 4: Compare $R(l)$ with $R(l+1)$, if for the threshold ε >0,

$$\max \left\{ \left| \mu_{ij}^{(l+1)} - \mu_{ij}^{(l)} \right| \mid i = 1, \ldots, c; j = 1, \ldots, n \right\} \leq \varepsilon,$$

then the optimal solution is achieved. Otherwise, let $l = l$+1, return to step 2.

Property 1. *When intuitionistic fuzzy set degenerates into fuzzy set, the IFCM algorithm degrades into FCM algorithm.*

When the intuitionistic fuzzy set degenerates into fuzzy set in the IFCM algorithm, formula (1) degrades as,

$$\begin{cases} \min J(R,V) = \sum\limits_{j=1}^{n} \sum\limits_{i=1}^{c} (\mu_{ij})^q D(A_j, V_i)^2, & q \in [1, \infty] \\ \sum\limits_{i=1}^{c} \mu_{ij} = 1, & 1 \leq j \leq n. \end{cases}$$

Formula (2) degrades as,

$$\mu_{Vi}(x_k)^{(l)} = \frac{\sum\limits_{j=1}^{n} \left(\mu_{ij}^{(l)}\right)^q \mu_{Aj}(x_k)}{\sum\limits_{j=1}^{n} (\mu_{ij}^{(l)})^q}.$$

Formula (4) remains unchanged,

$$\mu_{ij}^{(l+1)} = \frac{1}{\sum\limits_{k=1}^{c} \left(\frac{D\left(A_j, V_i^{(l)}\right)}{D\left(A_j, V_k^{(l)}\right)} \right)^{\frac{2}{q-1}}} \ (1 \leq i \leq c, 1 \leq j \leq n)$$

Formula (3) and (5) are gone. Hence IFCM algorithm degenerates into FCM algorithm.

For the computation complexity, although the intuitionistic fuzzy C-means clustering method is the extension of fuzzy C-means clustering method, the IFCM algorithm is not more complicated than FCM algorithm. They have the same computation complexity. In the case study, IFCM algorithm is compared with FCM algorithm.

3.2. **The influences of several parameters on the clustering results in IFCM algorithm.**

3.2.1. *The initial intuitionistic fuzzy partition matrix.* The partition matrix represents the membership degrees and non-membership degrees of the clustered objects to the categories. Although the initial partition matrix $R(0)$ has the following constraints:

(1) $\mu_{ij} + \gamma_{ij} + \pi_{ij} = 1$;

(2) $\sum\limits_{i=1}^{c} \mu_{ij} = 1$;

(3) $\sum\limits_{j=1}^{n} \mu_{ij} > 1 (1 \leq i \leq c)$.

However, there is no specific method to determine $R(0)$ and no numerical specifications as well. Therefore, whether the initial partition matrix affects the iterations and the final clustering results or not, is a question worth considering. In the case study, different initial partition matrices will be selected to test their influences on the iterations and clustering results.

3.2.2. *The number of classification.* The number of classification c is an important parameter in the IFCM algorithm. When the number of classification increases (or decreases), the clustering results will change. In general, the number of classification depends on the requirements of the decision maker. If the decision maker does not have specific requirements, it is necessary to determine the appropriate number of category by some means, such as historical data and experience. In the case study, the influence of the number of classification on the clustering results will be discussed.

3.2.3. *The threshold of iteration termination.* Whether the IFCM algorithm continues or not, the threshold ε is the judgment criterion. That is, if $\max\{|\mu_{ij}^{(n+1)} - \mu_{ij}^{(n)}||\} \leq \varepsilon$ holds, then we stop calculation and the optimal partition matrix is achieved. The accuracy of the algorithm and the size of the error are determined by the threshold. A smaller threshold can generally get a more accurate result. Meanwhile, the times of iteration and amount of calculation will increase together. Whether the larger threshold will affect the final clustering is also worth discussing.

4. Satisfaction evaluation based on the IFCM algorithm.

4.1. **Problem description.** In order to illustrate the effectiveness of the proposed method, this paper has adopted the case in reference [23]. Suppose there are eight commodities, named as $A_i\,(i = 1, 2, \ldots, 8)$. We need to classify these goods according to the results of customer satisfaction evaluation. Each commodity has four evaluable attributes, namely price (G_1), shape (G_2), quality (G_3) and function (G_4). The characteristic information of the commodities under the evaluation indices are represented by intuitionistic fuzzy values, as shown in Table 1.

Table 1. Customer satisfaction evaluation for the eight commodities

Goods	G_1	G_2	G_3	G_4
A_1	(0.56, 0.34)	(0.40, 0.50)	(0.30, 0.40)	(0.70, 0.10)
A_2	(0.41, 0.40)	(0.08, 0.80)	(0.05, 0.75)	(0.20, 0.50)
A_3	(0.38, 0.52)	(0.90, 0.10)	(0.80, 0.10)	(0.01, 0.80)
A_4	(0.31, 0.60)	(0.40, 0.50)	(0.30, 0.50)	(0.63, 0.15)
A_5	(0.31, 0.61)	(0.74, 0.22)	(0.70, 0.25)	(0.00, 0.90)
A_6	(0.44, 0.45)	(0.11, 0.80)	(0.06, 0.80)	(0.31, 0.52)
A_7	(0.58, 0.30)	(0.37, 0.52)	(0.30, 0.50)	(0.45, 0.35)
A_8	(0.43, 0.45)	(0.14, 0.72)	(0.07, 0.70)	(0.25, 0.55)

In Table 1, the intuitionistic fuzzy values represent the ratios of the customers supporting or opposing the attributes of commodities. For example, the membership degree of goods A_1 under the price attribute G_1 is 0.56, which indicates that the customers in favor of the price of the goods A_1 is 0.56, in other words, 56% of all the customs are satisfied with price; the non-membership degree is 0.34, which indicates that 34% of all the customs are dissatisfied with price; and the hesitation degree is 0.10 ($1 - 0.56 - 0.34 = 0.10$), which indicates that 10% of all the customs are keeping neutral with price. The meanings of other intuitionistic fuzzy values are similar. The optimal solutions in the final partition matrix are achieved by the membership degrees and non-membership degrees of intuitionistic fuzzy values.

According to Table 1, the characteristic matrix of the clustered objects is

$$A = \begin{pmatrix} (0.56, 0.34) & (0.40, 0.50) & (0.30, 0.40) & (0.70, 0.10) \\ (0.41, 0.40) & (0.08, 0.80) & (0.05, 0.75) & (0.20, 0.50) \\ (0.38, 0.52) & (0.90, 0.10) & (0.80, 0.10) & (0.01, 0.80) \\ (0.31, 0.60) & (0.40, 0.50) & (0.30, 0.50) & (0.63, 0.15) \\ (0.31, 0.61) & (0.74, 0.22) & (0.70, 0.25) & (0.00, 0.90) \\ (0.44, 0.45) & (0.11, 0.80) & (0.06, 0.80) & (0.31, 0.52) \\ (0.58, 0.30) & (0.37, 0.52) & (0.30, 0.50) & (0.45, 0.35) \\ (0.43, 0.45) & (0.14, 0.72) & (0.07, 0.70) & (0.25, 0.55) \end{pmatrix}.$$

4.2. **The influence of the initial partition matrix on clustering results.** The partition matrix R indicates the membership degrees and non-membership degrees of the classified objects to each category. The following are discussed with two different initial partition matrices.

(1) Let the cluster center $V = \{V_1, V_2, V_3\}$, n =8, c =3, q =2, ε =0.01, and the initial partition matrix adopts formula (6).

$$R(0) = \begin{pmatrix} (0.7, 0.1) & (0.1, 0.5) & (0.2, 0.3) & (0.7, 0.1) & (0.2, 0.3) & (0.1, 0.5) & (0.7, 0.1) & (0.2, 0.3) \\ (0.2, 0.3) & (0.7, 0.1) & (0.1, 0.5) & (0.2, 0.3) & (0.1, 0.5) & (0.7, 0.1) & (0.1, 0.5) & (0.7, 0.1) \\ (0.1, 0.5) & (0.2, 0.3) & (0.7, 0.1) & (0.1, 0.5) & (0.7, 0.1) & (0.2, 0.3) & (0.2, 0.3) & (0.1, 0.5) \end{pmatrix} \tag{6}$$

According to the equation (2) and (3), the characteristic matrix $V(0)$ of the cluster center can be calculated by Matlab software,

$$V(0) = \begin{pmatrix} (0.467, 0.428) & (0.403, 0.499) & (0.318, 0.465) & (0.517, 0.283 \\ (0.427, 0.439) & (0.172, 0.717) & (0.116, 0.694) & (0.283, 0.499 \\ (0.375, 0.525) & (0.668, 0.289) & (0.600, 0.293) & (0.104, 0.733 \end{pmatrix}.$$

The iterative operation is performed according to section 3.1, and it is stopped until $\max\{|\mu_{ij}^{(4)} - \mu_{ij}^{(3)}||\} \leq 0.01$. Then the optimal partition matrix

$$R(4) = \begin{pmatrix} (0.939, 0.051) & (0.012, 0.978) & (0.015, 0.975) & (0.858, 0.132) & (0.017, 0.973) & (0.013, 0.977) & (0.786, 0.204) & (0.013, 0.977) \\ (0.043, 0.948) & (0.984, 0.006) & (0.011, 0.979) & (0.102, 0.888) & (0.014, 0.976) & (0.984, 0.007) & (0.166, 0.825) & (0.983, 0.007) \\ (0.019, 0.972) & (0.004, 0.986) & (0.974, 0.017) & (0.041, 0.949) & (0.969, 0.021) & (0.004, 0.987) & (0.048, 0.942) & (0.004, 0.986) \end{pmatrix} \tag{7}$$

According to formula (7) and the principle of maximum membership degree, the classified goods can be divided into three categories: $\{A_1, A_4, A_7\}$, $\{A_2, A_6, A_8\}$ and $\{A_3, A_5\}$.

(2) The initial partition matrix adopts formula (8), and the other conditions are the same as above. Compared with formula (6), the membership degrees and non-membership degrees of the classified goods belonging to a category are vaguer in formula (8).

$$R(0) = \begin{pmatrix} (0.4, 0.3) & (0.4, 0.1) & (0.4, 0.1) & (0.2, 0.5) & (0.4, 0.3) & (0.4, 0.3) & (0.2, 0.5) & (0.2, 0.5) \\ (0.4, 0.1) & (0.4, 0.3) & (0.2, 0.5) & (0.4, 0.1) & (0.4, 0.1) & (0.2, 0.5) & (0.4, 0.3) & (0.4, 0.1) \\ (0.2, 0.5) & (0.2, 0.5) & (0.4, 0.3) & (0.4, 0.3) & (0.2, 0.5) & (0.4, 0.1) & (0.4, 0.1) & (0.4, 0.3) \end{pmatrix} \tag{8}$$

The characteristic matrix of the cluster center is calculated according to the equation (2) and (3),

$$V(0) = \begin{pmatrix} (0.420, 0.461) & (0.426, 0.497) & (0.359, 0.475) & (0.263, 0.536) \\ (0.424, 0.465) & (0.385, 0.521) & (0.314, 0.498) & (0.357, 0.449) \\ (0.439, 0.450) & (0.369, 0.542) & (0.297, 0.528) & (0.332, 0.474) \end{pmatrix}$$

The optimal partition matrix can be acquired in the same manner. The optimal partition matrix

$$R(8) = \begin{pmatrix} (0.018, 0.972) & (0.004, 0.986) & (0.974, 0.016) & (0.039, 0.951) & (0.969, 0.021) & (0.004, 0.987) & (0.051, 0.940) & (0.004, 0.986) \\ (0.941, 0.049) & (0.012, 0.978) & (0.015, 0.975) & (0.863, 0.127) & (0.017, 0.973) & (0.013, 0.977) & (0.776, 0.214) & (0.013, 0.977) \\ (0.041, 0.949) & (0.984, 0.006) & (0.011, 0.979) & (0.098, 0.892) & (0.014, 0.976) & (0.984, 0.006) & (0.174, 0.816) & (0.983, 0.007) \end{pmatrix} \tag{9}$$

According to formula (9), the classified goods can be divided into three categories: $\{A_1, A_4, A_7\}$, $\{A_2, A_6, A_8\}$ and $\{A_3, A_5\}$. The clustering results are the same as that of formula (7). Different initial partition matrix does not affect the final clustering results, although the membership degrees and non-membership degrees of the classified objects to each category are different. However, we need to note that, the first experiment only performs four iterations to get the optimal clustering results, but the second experiment needs eight times. Therefore, if the membership degrees and non-membership degrees of the classified objects to the categories are obviously different in the initial partition matrix, the iterations will be reduced accordingly.

4.3. **The influence of the classification number on clustering results.** The classified goods are clustered with the classification number $c = 3$ in section 4.2. Suppose the cluster centers are $V = \{V_1, V_2, V_3, V_4\}$, let $c = 4$, $q = 2$, $\varepsilon = 0.01$, and the initial partition matrix

$$R(0) = \begin{pmatrix} (0.4, 0.1) & (0.1, 0.7) & (0.2, 0.5) & (0.3, 0.6) & (0.2, 0.5) & (0.1, 0.7) & (0.4, 0.1) & (0.2, 0.5) \\ (0.3, 0.6) & (0.3, 0.6) & (0.1, 0.7) & (0.4, 0.1) & (0.1, 0.7) & (0.4, 0.1) & (0.3, 0.6) & (0.4, 0.1) \\ (0.2, 0.5) & (0.4, 0.1) & (0.1, 0.7) & (0.2, 0.5) & (0.4, 0.1) & (0.3, 0.6) & (0.1, 0.7) & (0.3, 0.6) \\ (0.1, 0.7) & (0.2, 0.5) & (0.4, 0.1) & (0.1, 0.7) & (0.3, 0.6) & (0.2, 0.5) & (0.2, 0.5) & (0.1, 0.7) \end{pmatrix} \tag{10}$$

According to section 3.1, the characteristic matrix of the cluster center can be calculated as,

$$V(0) = \begin{pmatrix} (0.489, 0.403) & (0.424, 0.483) & (0.345, 0.447) & (0.432, 0.377) \\ (0.415, 0.474) & (0.261, 0.634) & (0.189, 0.619) & (0.387, 0.412) \\ (0.392, 0.486) & (0.401, 0.515) & (0.345, 0.499) & (0.229, 0.574) \\ (0.425, 0.473) & (0.550, 0.393) & (0.476, 0.382) & (0.187, 0.619) \end{pmatrix}.$$

And the optimal partition matrix

$$R(23) = \begin{pmatrix} (0.578, 0.412) & (0.013, 0.977) & (0.016, 0.974) & (0.010, 0.981) & (0.018, 0.972) & (0.014, 0.976) & (0.904, 0.086) & (0.018, 0.972) \\ (0.054, 0.936) & (0.975, 0.015) & (0.010, 0.980) & (0.003, 0.987) & (0.014, 0.976) & (0.971, 0.019) & (0.032, 0.958) & (0.965, 0.025) \\ (0.344, 0.646) & (0.010, 0.981) & (0.014, 0.976) & (0.987, 0.004) & (0.017, 0.973) & (0.011, 0.979) & (0.055, 0.935) & (0.013, 0.977) \\ (0.024, 0.966) & (0.003, 0.987) & (0.960, 0.030) & (0.001, 0.989) & (0.951, 0.039) & (0.003, 0.987) & (0.010, 0.981) & (0.004, 0.986) \end{pmatrix} \tag{11}$$

The classified goods can be divided into four categories: $\{A_1, A_7\}$, $\{A_2, A_6, A_8\}$, $\{A_3, A_5\}$ and $\{A_4\}$. Therefore, when the number of classification changes, the clustering results change. If the number of classification increases, then some objects belonging to a certain category will be separated from the original class.

4.4. **The influence of the threshold on clustering results.** Whether the IFCM algorithm terminates or not depends on the threshold. Compared with section 4.3, except for the threshold, other conditions remain the same in this section. We select $\varepsilon = 0.1$ and $\varepsilon = 0.3$ respectively, to analyze the influences of different thresholds on the final clustering results.

(1) If ε =0.1, then the optimal partition matrix$R(6)$ is calculated as,

$$\begin{pmatrix} (0.802,0.188) & (0.011,0.979) & (0.015,0.976) & (0.005,0.985) & (0.017,0.974) & (0.012,0.978) & (0.717,0.273) & (0.014,0.976) \\ (0.028,0.962) & (0.976,0.014) & (0.010,0.980) & (0.001,0.989) & (0.014,0.976) & (0.973,0.017) & (0.095,0.896) & (0.969,0.021) \\ (0.158,0.832) & (0.010,0.980) & (0.014,0.976) & (0.987,0.004) & (0.017,0.973) & (0.012,0.978) & (0.161,0.830) & (0.013,0.977) \\ (0.012,0.978) & (0.003,0.987) & (0.961,0.029) & (0.001,0.990) & (0.952,0.038) & (0.003,0.987) & (0.028,0.962) & (0.004,0.986) \end{pmatrix} \tag{12}$$

The membership degrees and non-membership degrees of the classified goods in the optimal partition matrix, that is formula (12), are obviously different with formula (11). However, the classified goods can be divided into the identical four categories: $\{A_1,\ A_7\}$, $\{A_2,\ A_6,\ A_8\}$, $\{A_3,\ A_5\}$ and $\{A_4\}$. The clustering results are exactly the same as that of ε =0.01.

(2) If ε =0.3, then the optimal partition matrix $R(1)$ is calculated as,

$$\begin{pmatrix} (0.480,0.510) & (0.160,0.830) & (0.191,0.799) & (0.393,0.597) & (0.157,0.833) & (0.151,0.839) & (0.659,0.331) & (0.136,0.854) \\ (0.239,0.751) & (0.519,0.471) & (0.127,0.863) & (0.301,0.690) & (0.112,0.878) & (0.566,0.424) & (0.164,0.826) & (0.575,0.415) \\ (0.156,0.834) & (0.217,0.774) & (0.233,0.757) & (0.181,0.809) & (0.239,0.751) & (0.191,0,799) & (0.111,0.879) & (0.204,0.786) \\ (0.125,0.865) & (0.104,0.886) & (0.449,0.541) & (0.126,0.864) & (0.491,0.499) & (0.092,0.898) & (0.066,0.924) & (0.085,0.905) \end{pmatrix} \tag{13}$$

The classified goods can only be divided into three categories: $\{A_1,\ A_4,\ A_7\}$, $\{A_2,\ A_6,\ A_8\}$ and $\{A_3,\ A_5\}$. At this point we cannot achieve the fourth category.

Therefore, the threshold does not have an effect on the clustering results within a certain range, but it affects the membership degrees and non-membership degrees of the final partition matrix. When the threshold exceeds the certain range, it is impossible to obtain the clustering results satisfying the conditions.

5. **Compared with fuzzy *C*-means clustering algorithm.** Intuitionistic fuzzy set is the extension of fuzzy set. When intuitionistic fuzzy set degrades to fuzzy set, the IFCM algorithm reduces to the FCM algorithm. The case in section 4.1 is clustered by the FCM clustering algorithm at the following.

The characteristic matrix of the classified objects and the partition matrix in FCM are fuzzy values. The fuzzy characteristic matrixA' can be formed by the membership degrees of the classified objects in matrix A.

$$A' = \begin{pmatrix} 0.56 & 0.40 & 0.30 & 0.70 \\ 0.41 & 0.08 & 0.05 & 0.20 \\ 0.38 & 0.90 & 0.80 & 0.01 \\ 0.31 & 0.40 & 0.30 & 0.63 \\ 0.31 & 0.74 & 0.70 & 0.00 \\ 0.44 & 0.11 & 0.06 & 0.31 \\ 0.58 & 0.37 & 0.30 & 0.45 \\ 0.43 & 0.14 & 0.07 & 0.25 \end{pmatrix}$$

Let the cluster center $V = \{V_1,\ V_2,\ V_3\}$, n =8, c =3, q =2, ε =0.01, and the initial partition matrix

$$R = \begin{pmatrix} 0.6 & 0.4 & 0.2 & 0.3 & 0.6 & 0.4 & 0.1 & 0.4 \\ 0.3 & 0.4 & 0.4 & 0.6 & 0.1 & 0.2 & 0.3 & 0.4 \\ 0.1 & 0.2 & 0.4 & 0.1 & 0.3 & 0.4 & 0.6 & 0.2 \end{pmatrix}.$$

The characteristic matrix of the cluster center can be calculated as,

$$V' = \begin{pmatrix} 0.423 & 0.402 & 0.336 & 0.325 \\ 0.402 & 0.378 & 0.298 & 0.389 \\ 0.472 & 0.435 & 0.367 & 0.281 \end{pmatrix}.$$

According to section 2.3, the optimal fuzzy partition matrix

$$R' = \begin{pmatrix} 0.939 & 0.012 & 0.112 & 0.468 & 0.017 & 0.013 & 0.786 & 0.013 \\ 0.043 & 0.984 & 0.011 & 0.402 & 0.014 & 0.584 & 0.166 & 0.983 \\ 0.019 & 0.004 & 0.877 & 0.131 & 0.969 & 0.404 & 0.048 & 0.004 \end{pmatrix} \tag{14}$$

The classified goods can be divided into three categories: $\{A_1, A_4, A_7\}$, $\{A_2, A_6, A_8\}$ and $\{A_3, A_5\}$. The clustering results of FCM are the same as that of IFCM. Compared with formula (7), it is obvious that the membership degrees of $\{A_2, A_5, A_6, A_8\}$ are more similar, but the membership degrees of $\{A_3, A_5\}$ are less similar in formula (14). This will result in the confused classification for the FCM algorithm. However, as far as IFCM algorithm is concerned, the membership degrees of $\{A_2, A_6, A_8\}$ and $\{A_3, A_5\}$ are individually similar, and the divisions of each goods in the optimal partition matrix are very clear. That is, IFCM algorithm can avoid the confusion of classification. The reason is that intuitionistic fuzzy set express more abundant and richer information than that of fuzzy set. Therefore, Intuitionistic fuzzy set can describe the ambiguity and uncertainty of problems better than that of fuzzy set.

6. **Conclusions.** This paper studies the intuitionistic fuzzy C-means clustering algorithm. Several important parameters during the IFCM clustering process, such as the initial form of the partition matrix, the number of classification and the threshold of terminating the iteration, which significantly affect the clustering results, are analyzed and discussed. A case of customer satisfaction evaluation is given by the IFCM clustering method, and the method is compared with the FCM clustering method.

Since the intuitionistic fuzzy C-means clustering algorithm is a non-reference classification method, the initial partition matrix does not affect the final clustering results, but it affects the iteration times. For the initial partition matrix, if the membership degrees and non-membership degrees of the classified objects to the categories are obviously different, the iteration times will be reduced accordingly. When the number of classification increases, the clustering results change, and some objects are separated from the original category. The threshold value does not affect the clustering results in a certain range, but it has an impact on the membership degrees and non-membership degrees of the optimal partition matrix. When the threshold value is out of certain range, it is impossible to obtain the clustering results satisfying the condition. Compared with FCM clustering algorithm, IFCM algorithm can classify objects more accurately. This study can help to understand and master the factors that affect the intuitionistic fuzzy clustering results, and help to promote the further research and application of IFCM clustering method.

Acknowledgments. The authors are very grateful to the two anonymous referees for their insightful and constructive comments and suggestions, which have been very helpful in improving the paper. This work was supported by "the National Natural Science Foundation of China (No. 71771023)" and "the National Social Science Foundation of China (No. 17BGL052)".

REFERENCES

[1] K. T. Atanassov, Intuitionistic fuzzy sets *Journal of Fuzzy sets and Systems*, 20, no. 1, pp. 87-96, 1986.

[2] J. S. Pan, F. R. McInnes, M. A.Jack, Fast Clustering Algorithms for Vector Quantization, *Journal of Pattern Recognition*, vol. 29, no. 3, pp. 511-518, 1996.

[3] M. A. Khalilia, J. Bezdek, M. Popescu, et al. Improvements to the relational fuzzy c-means clustering algorithm, *Journal of Pattern Recognition*, vol. 47, pp. 3920-3930, 2014.

[4] Y. Li, Y. Li, Neural-fuzzy control of truck backer-upper system using a clustering method, *Journal of Neurocomputing*, vol. 70, pp. 680-688, 2007.

[5] S. Khaleie, M, Fasanghari, E. Tavassoli, Supplier selection using a novel intuitionist fuzzy clustering approach, *Journal of Applied Soft Computing*, vol. 12, pp. 1741-1754, 2012.

[6] S. Wikaisuksakul, A multi-objective genetic algorithm with fuzzy c-means for automatic data clustering, *Journal of Applied Soft Computing*, vol. 24, pp. 679-691, 2014.

[7] L. H. Son, DPFCM: A novel distributed picture fuzzy clustering method on picture fuzzy sets, *Journal of Expert Systems with Applications*, vol. 42, pp. 51-66, 2015.

[8] S. M. Chen, N. Y. Wang, J. S. Pan, Forecasting enrollments using automatic clustering techniques and fuzzy logical relationships, *Journal of Expert Systems with Applications*, vol. 36, no. 8, pp. 11070-11076, 2009.

[9] Z. Xu, J. Chen, J. Wu, Clustering algorithm for intuitionistic fuzzy sets, *Journal of Information Sciences*, vol. 178, pp. 3775-3790, 2008.

[10] Z. Wang, Z. Xu, S. Liu, et al. Netting clustering analysis method under intuitionistic fuzzy environment, *Journal of Applied Soft Computing*, vol. 11, pp. 5558-5564, 2011.

[11] T. Chaira, A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images, *Journal of Applied Soft Computing*, vol. 11, pp. 1711-1717, 2011.

[12] L. H. Son, B. C. Cuong, P. L. Lanzi, et al, A novel intuitionistic fuzzy clustering method for geo-demographic analysis, *Journal of Expert Systems with Applications*, vol. 39, pp. 9848-9859, 2012.

[13] H. Zhao, Z. Xu, Liu S, et al. Intuitionistic fuzzy MST clustering algorithms, *Journal of Computers & Industrial Engineering*, vol. 62, pp. 1130-1140, 2012.

[14] D. Xu, Z. Xu, S. Liu, et al. A spectral clustering algorithm based on intuitionistic fuzzy information, *Journal of Knowledge-Based Systems*, vol. 53, pp. 20-26, 2013.

[15] Z. Wang, Z. Xu, S. Liu, et al. Direct clustering analysis based on intuitionistic fuzzy implication, *Journal of Applied Soft Computing*, vol. 23, pp. 1-8, 2014.

[16] L. H. Son, A novel kernel fuzzy clustering algorithm for Geo-Demographic Analysis, *Journal of Information Sciences*, vol. 317, pp. 202-223, 2015.

[17] N. T. Thong, L. H. Son, HIFCF: An effective hybrid model between picture fuzzy clustering and intuitionistic fuzzy recommender systems for medical diagnosis, *Journal of Expert Systems with Applications*, vol. 42, pp. 3682-3701, 2015.

[18] H. Verma, R. K. Agrawal, A. Sharan An improved intuitionistic fuzzy c-means clustering algorithm incorporating local information for brain image segmentation, *Journal of Applied Soft Computing*, vol. 46, pp. 543-557, 2016.

[19] Y. K. Dubey, M. M. Mushrif, K. Mitra, Segmentation of brain MR images using rough set based intuitionistic fuzzy clustering, *Journal of Biocybernetics and Biomedical Engineering*, vol. 36, pp. 413-426, 2016.

[20] P. Milošević, B. Petrović, Jeremić V. IFS-IBA similarity measure in machine learning algorithms, *Journal of Expert Systems with Applications*, vol. 89, pp. 296-305, 2017.

[21] P. D'Urso, Informational Paradigm, management of uncertainty and theoretical formalisms in the clustering framework: A review, *Journal of Information Sciences*, vol. 400-401, pp. 30-62, 2017.

[22] W. Wang, X. Xin, Distance measure between intuitionistic fuzzy sets, *Journal of Pattern Recognition Letters*, vol. 26, pp. 2063-2069, 2005.

[23] Z. H. He, Y. J. Lei, Research on intuitionistic fuzzy C-means clustering algorithm, *Journal of Control and Decision*, vol. 26, no. 6, pp. 847-850, 2011.