Hindawi
Complexity
Volume 2021, Article ID 6657849, 12 pages
https://doi.org/10.1155/2021/6657849

Research Article

Neutrosophic Clustering Algorithm Based on Sparse Regular Term Constraint

Dan Zhang, Yingcang Ma, Hu Zhao, and Xiaofei Yang

School of Science, Xi'an Polytechnic Unversity, Xi'an, China

Correspondence should be addressed to Yingcang Ma; mayingcang@xpu.edu.cn

Received 8 November 2020; Revised 8 December 2020; Accepted 8 January 2021; Published 30 January 2021

Academic Editor: Heng Liu

Clustering algorithm is one of the important research topics in the field of machine learning. Neutrosophic clustering is the generalization of fuzzy clustering and has been applied to many fields. This paper presents a new neutrosophic clustering algorithm with the help of regularization. Firstly, the regularization term is introduced into the FC-PFS algorithm to generate sparsity, which can reduce the complexity of the algorithm on large data sets. Secondly, we propose a method to simplify the process of determining regularization parameters. Finally, experiments show that the clustering results of this algorithm on artificial data sets and real data sets are mostly better than other clustering algorithms. Our clustering algorithm is effective in most cases.

1. Introduction

With the increasing development of information technology, the data dimensions on the Internet have increased exponentially. For example, dimensions of various documents, multimedia, and gene expression data can reach hundreds of thousands. Facing these data, scholars have proposed many data processing methods [1–3].

In 1965, Zedah [4] proposed the concept of fuzzy set. Fuzzy theory is applied in many areas, such as multiattribute decision-making [5–7], image processing [8], and cluster analysis [9]. In particular, fuzzy clustering has made considerable progress in the past few decades. Based on fuzzy sets, FCM [10] algorithm is proposed. The quality of the clustering results is good, but there are still some problems for uncertainty problems. Therefore, in recent years, scholars have devoted themselves to propose a variety of methods to improve the fuzzy c-means algorithm of various aspects. Hwang [11] et al. combined the type-2 fuzzy set with the FCM (T2-FCM) clustering algorithm and made an improvement on the uncertainty that affects the final class c classification. Linda [12] et al. improved the general type-2 fuzzy set fuzzy c-means (GT2-FCM) algorithm through the alpha surface representation theorem, described the ambiguity in linguistic terms, and transformed the uncertainty of the language into the uncertain fuzzy positions of the extracted clusters. The algorithm [12] works well when there are noisy samples or insufficient training samples. The T2-FCM and GT2-FCM algorithms are all improved for the uncertainty of fuzzy c-means algorithm.

In 1986, Atanassov [13] proposed the concept of intuitionistic fuzzy sets, which solved some of the drawbacks of traditional fuzzy sets, and is more capable of processing uncertain information. Chaira [14] et al. introduced intuitionistic fuzzy entropy into the traditional fuzzy c-means algorithm, and the new algorithm proposed was used to cluster CT brain scan partial images, which can identify brain abnormalities. Bukiewicz [15] et al. introduced a variable to deal with the uncertainty and similarity measurement between intuitionistic fuzzy sets in the fuzzy c-means algorithm and proposed a data set fuzzy clustering method based on the intuitionistic fuzzy set theory. Zhao [16] et al. constructed the corresponding lambda cutting matrix by calculating the correlation coefficient on the intuitionistic fuzzy set and then clustered on the cutting matrix. Cuong [17] proposed the concept of the picture fuzzy set (PFS), which is a direct extension of a fuzzy set and intuitive fuzzy set. Thong [18] proposed an image fuzzy

clustering algorithm based on image fuzzy sets. The algorithms proposed in literatures [14–18] have better clustering performance than the traditional general algorithm, but they have certain limitations in the application. The generated membership matrix does not have sparseness, which will increase the amount of calculation.

In view of the limitations of the intuitive fuzzy sets, Smarandache [19] proposed the neutrosophic set theory. The basic idea is that everything can be described in three degrees of truth, uncertainty, and distortion. Each object has three degrees of membership function. Each membership function belongs to the standard and nonstandard subsets of $]0^-, 1^+[$. The neutrosophic set theory can not only describe the uncertainty problems better but also solve the existing problems when applying fuzzy theory. Therefore, scholars have done in-depth research on neutrosophic set [20–24] and proposed many neutrosophic clustering algorithms. Ye [25] proposed a single-valued neutrosophic minimum spanning tree (SVNMST) clustering algorithm, which shows great advantages in the clustering of single-valued neutrosophic observation data. In the same year, Ye [26] proposed single-valued neutrosophic clustering methods based on similarity measures between single-valued neutrosophic sets (SVNSs). Guo [27] proposed neutrosophic c-means clustering algorithm (NCM). The NCM algorithm can calculate certainty and uncertainty, and the membership function is not affected by noise. Nowadays, neutrosophic clustering has been applied to many fields such as image segmentation and biology [28–32]. PFS is a standardized form of neutrosophic set. The FC-PFS algorithm proposed in [18] is actually a kind of neutrosophic set type algorithm. However, the algorithm needs to calculate three matrices of the same scale, and the membership matrix is not sparse, which affects the clustering effect to a certain extent.

In order to solve the abovementioned problems, this paper proposes a new algorithm sparse neutrosophic fuzzy clustering algorithm (SNCM). The main idea is to introduce a regularization term into FC-PFS algorithm. The new algorithm can produce sparsity, since it reduces the number of eigenvalue vectors of the sample. Thus, SNCM reduces the complexity of the model. Experiments show that the performance of the proposed algorithm is better than some other clustering algorithms. The experimental results produce a sparse membership matrix, which reflects the effectiveness of the algorithm. The specific arrangements for the rest of this article are as follows.

The second section introduces the related basic concepts and algorithms, the third section presents the new algorithm proposed in this article and the solution process, the fourth section proves the effectiveness of the proposed algorithm through related experiments, and the fifth section gives relevant conclusions.

2. Related Algorithms

In this paper, the data set contains n data points, each point is a d-dimensional feature vector; the purpose of clustering is to obtain c clusters. The following introduces some clustering algorithms FCM and FC-PFS.

2.1. FCM Algorithm. The FCM algorithm proposed in 1984 is a very well-known algorithm. It is not only used in fuzzy engineering but also popular in the fields of medical diagnosis and communication. The FCM algorithm divides each data point x_i into a specific cluster v_j, u_{ij} means the i-th data point x_i belongs to the membership value of the j-th cluster. The cluster center of the cluster is expressed as $v_j \in R^d$, and the objective function of the FCM algorithm is

$$J = \sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij}^{m}\left\|x_i - v_j\right\|^2, \tag{1}$$

where m is a fuzzy parameter and the constraint condition of formula (1) is as follows:

$$\begin{cases} u_{ij} \in [0,1], \\ \sum_{j=1}^{c} u_{ij} = 1. \end{cases} \tag{2}$$

Using Lagrangian multiplier method, the iterative method of membership degree and cluster center is obtained:

$$v_j = \frac{\sum_{i=1}^{n} u_{ij}x_i}{\sum_{i=1}^{n} u_{ij}}; \quad j = 1, 2, \ldots, c,$$

$$u_{ij} = \frac{1}{\sum_{l=1}^{c}\left(\left\|x_i - v_j\right\| / \left\|x_i - v_l\right\|\right)^{2/(m-1)}}; \quad i = 1, 2, \ldots, n. \tag{3}$$

Until the number of iterations reaches the maximum value or $|J^{(t)} - J^{(t-1)}| < \varepsilon$, the iteration terminates, where $J^{(t)}$ and $J^{(t-1)}$are the objective function values of the t and $t-1$ iterations, and ε are the termination thresholds, generally in the range of (0, 0.1). According to the fuzzy membership value, if $u_{il} = \max(u_{i1}, u_{i2}, \ldots, u_{ik})$, then x_i is divided into j-th cluster. It can be proved that the algorithm finally converges to the local optimum or the saddle point of the objective function.

2.2. FC-PFS Algorithm

Definition 1. A picture fuzzy set of nonempty set X is

$$\dot{A} = \{\langle x, \mu_{\dot{A}}(x), \eta_{\dot{A}}(x), \gamma_{\dot{A}}(x)\rangle | x \in X\}, \tag{4}$$

where $\mu_{\dot{A}}(x)$ is the degree of positive membership of each $x \epsilon X$ in A, $\eta_{\dot{A}}(x)$ is the degree of neutral membership of x in A, and $\gamma_{\dot{A}}(x)$ is the degree of negative membership of x in A, and it satisfies the following conditions:

$$\begin{aligned} \mu_{\dot{A}}(x), \eta_{\dot{A}}(x), \gamma_{\dot{A}}(x) &\in [0,1], \quad \forall x \in X, \\ 0 \le \mu_{\dot{A}}(x), \eta_{\dot{A}}(x), \gamma_{\dot{A}}(x) &\le 1, \quad \forall x \in X. \end{aligned} \tag{5}$$

The refusal degree of an element is calculated as

$$\xi_{\dot{A}}(x) = 1 - \left(\mu_{\dot{A}}(x) + \eta_{\dot{A}}(x) + \gamma_{\dot{A}}(x)\right), \quad \forall x \in X. \tag{6}$$

Definition 2. X is an object (point) set, x is an element in X, and the neutrosophic set A on X can be expressed as

$$A = \{[x, (T_A(x), I_A(x), F_A(x))] | x \in X\}, \tag{7}$$

where $T_A(x)$ is the truth membership degree, $I_A(x)$ is the indeterminacy membership degree, and $F_A(x)$ is the falsity membership degree, which belongs to the standard and nonstandard subset of $]0^-, 1^+[$, i.e., $T_A(x), I_A(x), F_A(x) \longrightarrow]0^-, 1^+[$. Because there is no restriction on the sum of $T_A(x)$, $I_A(x)$, $F_A(x)$, there is $0^- \le \sup T_A(x) + \sup I_A(x) + \sup F_A(x) \le 3^+$.

From the abovementioned two definitions, it can be seen that the picture fuzzy set is actually the standard form of the neutrosophic set. Therefore, the FC-PFS algorithm proposed by Thong and Son is based on the neutrosophic set. The objective function of the algorithm is

$$J = \sum_{i=1}^{n}\sum_{j=1}^{c}\left(u_{ij}\left(2 - \xi_{ij}\right)\right)^m \left\|x_i - v_j\right\|^2 + \sum_{i=1}^{n}\sum_{j=1}^{c}\eta_{ij}\left(\log \eta_{ij} + \xi_{ij}\right). \tag{8}$$

Among them, u_{ij}, ξ_{ij}, and η_{ij} are the true membership degree, refusal membership degree, and neutral membership degree of the data points x_i belonging to the j-th cluster, respectively. The constraints of formula (8) are

$$\begin{cases} u_{ij}, \eta_{ij}, \xi_{ij} \in [0,1], \\ u_{ij} + \eta_{ij} + \xi_{ij} \le 1, \\ \sum_{j=1}^{c}\left(u_{ij}\left(2 - \xi_{ij}\right)\right) = 1, \\ \sum_{j=1}^{c}\left(\eta_{ij} + \dfrac{\xi_{ij}}{c}\right) = 1. \end{cases} \tag{9}$$

Using the Lagrangian multiplier method, the iterative method is adopted to obtain the update formulas of u_{ij}, ξ_{ij}, η_{ij}, and v_j:

$$\begin{aligned} \xi_{ij} &= 1 - \left(u_{ij} + \eta_{ij}\right) - \left(1 - \left(u_{ij} + \eta_{ij}\right)^{\alpha}\right)^{1/\alpha}, \quad \alpha \in [0,1], \quad (i = 1, \ldots, n; j = 1, \ldots, c), \\ u_{ij} &= \frac{1}{\sum_{l=1}^{c}\left(2 - \xi_{il}\right)\left(\left\|x_i - v_j\right\| / \left\|x_i - v_l\right\|\right)^{2/(m-1)}}, \quad (i = 1, 2, \ldots, n; j = 1, 2, \ldots, c), \\ \eta_{ij} &= \frac{e^{-\xi_{ij}}}{\sum_{l=1}^{c} e^{-\xi_{il}}}\left(1 - \frac{1}{c}\sum_{l=1}^{c}\xi_{il}\right), \quad (i = 1, \ldots, n; j = 1, \ldots, c), \\ v_j &= \frac{\sum_{i=1}^{n}\left(u_{ij}\left(2 - \xi_{ij}\right)\right)^m x_i}{\sum_{i=1}^{n}\left(u_{ij}\left(2 - \xi_{ij}\right)\right)^m}, \quad (j = 1, \ldots, c). \end{aligned} \tag{10}$$

The iteration is terminated until the number of iterations reaches the maximum or $u^{(t)} - u^{(t-1)} + \eta^{(t)} - \eta^{(t-1)} + \xi^{(t)} - \xi^{(t-1)} \le \varepsilon$.

3. Sparse Neutrosophic Clustering Algorithm

3.1. Determining the Objective Function. In traditional k-means clustering, each row of the membership matrix U contains a 1, and the remaining $c - 1$ elements in this row are 0, so the row sum of U is 1, and each column sum represents the number of sample points in each cluster, and the fuzzy c-means algorithm needs to choose the appropriate fuzzy degree m. Different from the abovementioned three clustering algorithms, the algorithm in this paper relaxes each element of U to a nonnegative value less than 1 under the constraint conditions and presets the ambiguity $m = 1$. Our

goal is to get a sparse U, so we introduce regular terms to get the objective function of the new algorithm:

$$J = \sum_{i=1}^{n}\sum_{j=1}^{c}\left(u_{ij}\left(2-\xi_{ij}\right)\right)\left\|x_i - v_j\right\|^2 + \sum_{i=1}^{n}\sum_{j=1}^{c}\eta_{ij}\left(\log\eta_{ij} + \xi_{ij}\right) + \gamma\sum_{i=1}^{n}\sum_{j=1}^{c}\left(u_{ij}\left(2-\xi_{ij}\right)\right)^2 \longrightarrow \min. \tag{11}$$

The abovementioned formula satisfies the following constraints:

$$\begin{cases} u_{ij}, \eta_{ij}, \xi_{ij} \in [0,1], \\ u_{ij} + \eta_{ij} + \xi_{ij} \le 1, \\ \sum_{j=1}^{c}\left(u_{ij}\left(2-\xi_{ij}\right)\right) = 1, \quad (i = t1, 2n, q \ldots h_n x; 7jC =; 1, 2, \ldots, c). \\ \sum_{j=1}^{c}\left(\eta_{ij} + \dfrac{\xi_{ij}}{c}\right) = 1. \end{cases} \tag{12}$$

We can see that if the sample point is divided into a single cluster, $u_{ij}(2-\xi_{ij})$ is equal to 1. Otherwise, it is a nonnegative value less than 1.

The new algorithm considers the sparsity of the membership degree of each sample point assigned to different clusters in the clustering process. In the process of minimizing equation (11), the importance of each part is controlled by the parameter γ. If the parameter is zero, the membership vector of each sample is not sparse. If the parameter size is constantly adjusted, the sparsity of the member vector will also change. As the parameter gradually increases, the membership vector contains more and more nonzero elements. When the maximum value is reached, all elements of the membership vector are not zero, and the membership vector is nonsparse at this time. Therefore, this parameter controls the sparsity of the membership vector. We will give a method to determine the appropriate parameters in the subsequent part to obtain more accurate clustering results.

3.2. The Proposed Model and Solutions. Solve the abovementioned model using alternating iteration method. First, fix the variable U, ξ, η to find the cluster center V. The derivative of (11) in V is

$$\frac{\partial J}{\partial v_j} = \sum_{i=1}^{n}\left(u_{ij}\left(2-\xi_{ij}\right)\right)\left(-2x_i + 2v_j\right), \cdot (i = 1, 2, \ldots, n; j = 1, 2, \ldots, c). \tag{13}$$

By considering $\partial J/\partial v_j = 0$, we have

$$\sum_{i=1}^{n}\left(u_{ij}\left(2-\xi_{ij}\right)\right)\left(-2x_i + 2v_j\right) = 0, \quad (i = 1, 2, \ldots, n; j = 1, 2, \ldots, c), \tag{14}$$

$$\Leftrightarrow \sum_{i=1}^{n}\left(u_{ij}\left(2-\xi_{ij}\right)\right)x_i = \sum_{i=1}^{n}\left(u_{ij}\left(2-\xi_{ij}\right)\right)v_j, \quad (i = 1, 2, \ldots, n; j = 1, 2, \ldots, c), \tag{15}$$

$$\Leftrightarrow v_j = \frac{\sum_{i=1}^{n}\left(u_{ij}\left(2-\xi_{ij}\right)\right)x_i}{\sum_{i=1}^{n}\left(u_{ij}\left(2-\xi_{ij}\right)\right)}, \quad (j = 1, \ldots, c). \tag{16}$$

Solve U with fixed V, ξ, and η. In order to facilitate the solution, we make the following deformation of the objective function:

$$\min_{s_i^T 1=1, s_i>0}\sum_{i=1}^{n}\sum_{j=1}^{c}\left(s_{ij}d_{ij} + \gamma s_{ij}^2\right), \tag{17}$$

where $s_{ij} = u_{ij}(2-\xi_{ij})$ is an element of matrix S, s_{ij} is the i-th row of matrix S, $d_{ij} = \|x_i - v_j\|^2$is an element of distance matrix D For each x_i, problem (17) can be divided into n subproblems:

$$\min_{s_i^T 1=1, s_i>0}\sum_{j=1}^{c}\left(s_{ij}d_{ij} + \gamma_i s_{ij}^2\right). \tag{18}$$

Then, (18) is written in the following vector form:

$$\min_{s_i^T 1=1, s_i>0}\left\|s_i + \frac{d_i}{2\gamma_i}\right\|^2. \tag{19}$$

By solving problem (19), the solution of S can be obtained, and the update formula of U can be further obtained

$$u_{ij} = \begin{cases} \dfrac{d_{i,k+1} - d_{ij}}{\left(kd_{i,k+1} - \sum_{l=1}^{k} d_{il}\right)/\left(2-\xi_{ij}\right)}, & j \le k, \\ 0, & j > k. \end{cases} \tag{20}$$

The specific solution process for problem (19) is given in Section 3.3. Fixed variables U, ξ and V, use the Lagrange multiplier method to solve η:

$$L(\eta) = \sum_{i=1}^{n}\sum_{j=1}^{c}\eta_{ij}\left(\log\eta_{ij} + \xi_{ij}\right) - \lambda_i\left(\sum_{j=1}^{c}\left(\eta_{ij} + \frac{\xi_{ij}}{c}\right) - 1\right). \tag{21}$$

We use the function L to derive η to make it equal to zero, that is,

$$\frac{\partial L(\eta)}{\partial \eta_{ij}} = \log \eta_{ij} + 1 - \lambda_i + \xi_{ij} = 0, \quad (i = 1, 2, \ldots, n; j = 1, 2, \ldots, c), \tag{22}$$

$$\Leftrightarrow \eta_{ij} = \exp\left(\lambda_i - 1 - \xi_{ij}\right), \quad (i = 1, 2, \ldots, n; j = 1, 2, \ldots, c), \tag{23}$$

$$\sum_{j=1}^{c} e^{\lambda_i - 1 - \xi_{ij}} + \frac{1}{c}\sum_{j=1}^{c} \xi_{ij} = 1, \quad (i = 1, 2, \ldots, n; j = 1, 2, \ldots, c), \tag{24}$$

$$\Leftrightarrow e^{\lambda_i - 1} \sum_{j=1}^{c} e^{-\xi_{ij}} = 1 - \frac{1}{c}\sum_{j=1}^{c} \xi_{ij}, \quad (i = 1, 2, \ldots, n; j = 1, 2, \ldots, c), \tag{25}$$

$$\Leftrightarrow e^{\lambda_i - 1} = \frac{1 - (1/c)\sum_{j=1}^{c} \xi_{ij}}{\sum_{j=1}^{c} e^{-\xi_{ij}}}, \quad (i = 1, 2, \ldots, n; j = 1, 2, \ldots, c), \tag{26}$$

$$\eta_{ij} = \frac{e^{-\xi_{ij}}}{\sum_{l=1}^{c} e^{-\xi_{il}}}\left(1 - \frac{1}{c}\sum_{l=1}^{c} \xi_{il}\right), \quad (i = 1, \ldots, n; j = 1, \ldots, c). \tag{27}$$

Finally, using the similar technique of Yager [33] to generate operators, we modify the hesitation of the intuitionistic fuzzy set $\pi_{\widehat{A}}(x) = 1 - \mu_{\widehat{A}}(x) - (1 - \mu_{\widehat{A}}(x)^{\alpha})^{1/\alpha}$ to obtain the value of element rejection degree by replacing $u_{ij} + \eta_{ij}$ with $\mu_{\widehat{A}}(x)$, as follows:

$$\xi_{ij} = 1 - \left(u_{ij} + \eta_{ij}\right) - \left(1 - \left(u_{ij} + \eta_{ij}\right)^{\alpha}\right)^{1/\alpha}, \quad \alpha \in [0, 1], \cdot (i = 1, \ldots, n; j = 1, \ldots, c). \tag{28}$$

3.3. Optimization Method for γ. In specific practice, the regularization parameter in question (19) is difficult to determine, and its value can be from zero to infinity. In this section, a method for determining the regularization parameter γ is given. The Lagrangian function of question (19) is

$$L\left(s_i, \lambda, \beta_i\right) = \frac{1}{2}\left\| s_i + \frac{d_i}{2\gamma_i}\right\|^2 - \lambda\left(s_i^T \cdot 1 - 1\right) - \beta_i^T s_i, \tag{29}$$

where λ and β_i are greater than zero and are Lagrange multipliers.

According to the KKT condition, the optimal solution of is the following form

$$s_{ij} = \left(-\frac{d_{ij}}{2\gamma_i} + \lambda\right)_+. \tag{30}$$

In practice, if we focus on the locality of the data, usually we can get better performance. Therefore, it is best to learn a sparse s_i. Another advantage of learning sparse matrix S is that it can greatly reduce the computational burden of subsequent processing. Without loss of generality, it is assumed $d_{i1}, d_{i2}, \ldots, d_{ic}$ to be sorted from small to large. If the optimal s_i only has k nonzero elements, then according to equation (30), we know $s_{ik} > 0$ and $s_{i,j+1} = 0$. So, we have

$$\begin{cases} -\dfrac{d_{ik}}{2\gamma_i} + \lambda > 0, \\ \\ -\dfrac{d_{i,k+1}}{2\gamma_i} + \lambda \le 0. \end{cases} \tag{31}$$

According to equation (30) and constraint $s_i^T 1 = 1$, we have

$$\sum_{j=1}^{k}\left(-\frac{d_{ij}}{2\gamma_i} + \lambda\right) = 1 \Rightarrow \lambda = \frac{1}{k} + \frac{1}{2k\gamma_i}\sum_{j=1}^{k} d_{ij}. \tag{32}$$

According to equations (38) and (39), we have an inequality of γ_i

$$\frac{k}{2}d_{ik} - \frac{1}{2}\sum_{j=1}^{k} d_{ij} < \gamma_i \le \frac{k}{2}d_{i,k+1} - \frac{1}{2}\sum_{j=1}^{k} d_{ij}. \tag{33}$$

Therefore, in order to obtain the optimal solution of problem (19) with precise k nonzero values, we can make

$$\gamma_i = \frac{k}{2}d_{i,k+1} - \frac{1}{2}\sum_{j=1}^{k} d_{ij}. \tag{34}$$

Taking the average of $\gamma_1, \gamma_2, \ldots, \gamma_n$, the calculation formula is as follows:

$$\gamma = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{k}{2}d_{i,k+1} - \frac{1}{2}\sum_{j=1}^{k} d_{ij}\right). \tag{35}$$

Equation (35) gives a method to determine the regularization parameters.

According to equations (31), (33), and (35), the following optimal solution can be obtained,

$$\widehat{s}_{ij} = \begin{cases} \dfrac{d_{i,k+1} - d_{ij}}{k d_{i,k+1} - \sum_{l=1}^{k} d_{il}}, & j \le k, \\ \\ 0, & j > k. \end{cases} \tag{36}$$

3.3.1. Sparse Neutrosophic Clustering Algorithm

(i) Input: data set X, number of clusters c, and parameters α and k

(a) Initialization: set $t = 0$, random initialization meets the restriction condition (12)

(b) Set iteration:
(ii) for $i = 1, 2, \ldots$ maxSteps do
(c) Update V
(iii) calculate $v_j^{(t)}$ $(j = 1, 2, \ldots, c)$ by equation (16)
(d) Update U
(iv) for $j = 1, 2, \ldots, n$ do
(v) calculate S by equation (36)
(vi) end for
(vii) By solving problem (19), calculate $u_{ij}^{(t)}$ $(i = 1, 2, \ldots, n; j = 1, 2, \ldots, c)$ by equation (18)
(e) Update η
(viii) Calculate $\eta_{ij}^{(t)}$ $(i = 1, 2, \ldots, n; j = 1, 2, \ldots, c)$ by equation (27)
(f) Update ξ
(ix) Calculate $\xi_{ij}^{(t)}$ $(i = 1, 2, \ldots, n; j = 1, 2, \ldots, c)$ by equation (28)
(g) Set the conditions for jumping out of the iteration $u^{(t)} - u^{(t-1)} + \eta^{(t)} - \eta^{(t-1)} + \xi^{(t)} - \xi^{(t-1)} \le \varepsilon$ or $t >$ max steps
(x) end for
(xi) Output: clustering result y

Below, we analyze the algorithm complexity. First, we analyze the time complexity of the algorithm. From the algorithm steps, the basic sentence of the algorithm is the loop body of the algorithm iterative calculation variable, and the loop body for calculating u is embedded, so the time complexity of the algorithm is $O(nt)$, t is the number of iterations and n is the number of sample points. Secondly, the space complexity of the analysis algorithm is related to the data scale, so the space complexity is $O(nm)$, n is the number of sample points, and m is the dimension.

4. Results and Discussion

In order to verify the feasibility of the clustering algorithm SNCM proposed in this paper, classic clustering algorithms are selected: FCM [10], K-means [34], Ncut [35], Rcut [36], FC-PFS, and an effective clustering method based on data indeterminacy in neutrosophic set domain (INCM) [37], as comparison algorithms. A variety of evaluation indicators such as accuracy (ACC) and normalized mutual information (NMI) are used to evaluate the clustering results.

In terms of parameters, due to the instability of the K-means, FCM, and FS-PFC, a method of averaging them is adopted for 50 runs. For the Rcut and the Ncut, the experiment used the widely used self-tuning Gaussian method to construct the affinity matrix (the value is self-tuning). Take 0.9 for the parameter in FC-PFS algorithm. The parameter values in INCM algorithm are the best values found in literature [37]. The parameter in SNCM algorithm is 0.9, the value of parameter k is self-adjusted, and maxSteps is 1000.

In terms of experimental environment, all the experimental environments in this article are Microsoft Windows 10 system, the processor is Intel(R) Core(TM) i5-7200U CUP @ 250 GHz 2.70 GHz, memory 8.00 GB, programming software used is MATLAB R2016a.

4.1. SNCM Algorithm Descriptions. First, we illustrate the process of the proposed algorithm SNCM clustering the WBC data set; at this time, $n = 683$ and $c = 2$. The initial membership matrix, uncertainty matrix, and rejection matrix are as follows:

$$u^{(0)} = \begin{pmatrix} 0.413425 & 0.13977 \\ 0.030999 & 0.411379 \\ \cdots & \\ 0.237233 & 0.029844 \end{pmatrix},$$

$$\eta^{(0)} = \begin{pmatrix} 0.193085 & 0.035479 \\ 0.269577 & 0.070174 \\ \cdots & \\ 0.103972 & 0.127943 \end{pmatrix}, \quad (37)$$

$$\xi^{(0)} = \begin{pmatrix} 0.076247 & 0.112293 \\ 0.165006 & 0.024957 \\ \cdots & \\ 0.12897 & 0.127931 \end{pmatrix}.$$

The distribution of data points according to these initializations is illustrated in Figure 1(a) in which the SNCM algorithm is used to calculate the cluster centers using equation (19):

$$v = \begin{pmatrix} 4.443498 & 4.4813 \\ 3.214702 & 3.165621 \\ \cdots & \\ 1.649762 & 1.546717 \end{pmatrix}. \quad (38)$$

Then, we calculate the new membership matrix, uncertainty matrix, and rejection matrix:

$$u^{(1)} = \begin{pmatrix} 0 & 0.529743 \\ 0.544961 & 0 \\ \cdots & \\ 0.534465 & 0 \end{pmatrix},$$

$$\eta^{(1)} = \begin{pmatrix} 0.461026 & 0.444704 \\ 0.420874 & 0.484144 \\ \cdots & \\ 0.435548 & 0.436001 \end{pmatrix}, \quad (39)$$

$$\xi^{(1)} = \begin{pmatrix} 0.074124 & 0.010408 \\ 0.013241 & 0.074029 \\ \cdots & \\ 0.01189 & 0.073968 \end{pmatrix}.$$

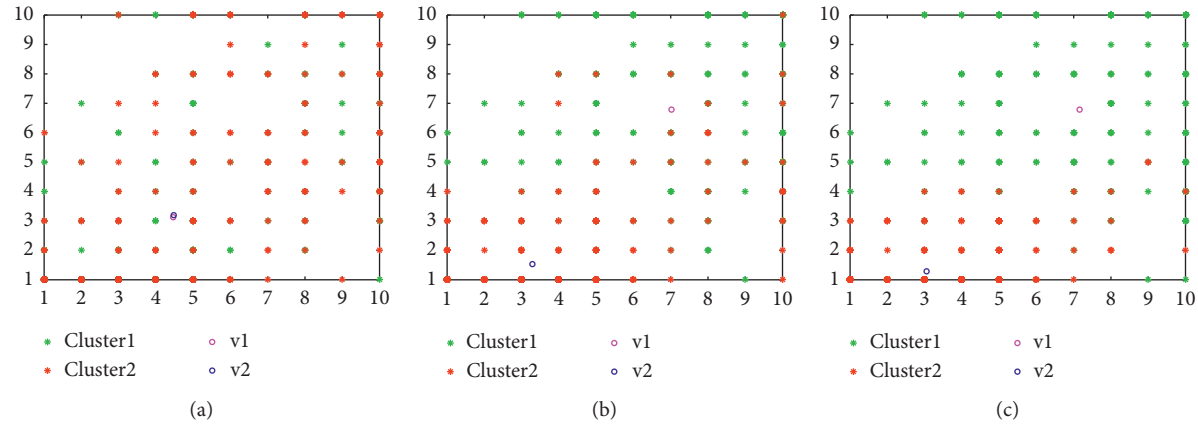


Figure 1: SNCM iteration diagram. (a) The initialized clustering result graph. (b) The clustering result graph after the first iteration. (c) The clustering result diagram of the final iteration.

According to the abovementioned matrix, the calculated value of $u^{(t)} - u^{(t-1)} + \eta^{(t)} - \eta^{(t-1)} + \xi^{(t)} - \xi^{(t-1)}$ is greater than ε, so the iterative step will continue. Figure 1(b) shows the distribution of clusters after the first iteration.

Through a similar process, we continue to calculate the cluster center, membership degree, uncertainty degree, and rejection matrix until the stopping condition is met. The final membership degree, hesitation degree, and rejection degree matrix is as follows:

$$u^* = \begin{pmatrix} 0 & 0.500122 \\ 0.500122 & 0 \\ \cdots & \\ 0.500122 & 0 \end{pmatrix},$$

$$\eta^* = \begin{pmatrix} 0.463632 & 0.499062 \\ 0.499062 & 0.463632 \\ \cdots & \\ 0.499062 & 0.463632 \end{pmatrix}, \quad (40)$$

$$\xi^* = \begin{pmatrix} 0.074125 & 0.000487 \\ 0.000487 & 0.074125 \\ \cdots & \\ 0.000487 & 0.074125 \end{pmatrix}$$

The calculated final cluster centers are expressed as follows, and the distribution of clusters and cluster centers is shown in Figure 1(c):

$$v^* = \begin{pmatrix} 3.050885 & 7.164502 \\ 1.29646 & 6.779221 \\ \cdots & \\ 1.112832 & 2.562771 \end{pmatrix}. \quad (41)$$

4.2. Verification of Sparsity. First of all, experiments are carried out using artificial aggregation data sets and real Wine data sets. The aggregation data set is a data set composed of 7 clusters of 788 2-dimensional data points. The Wine data set is a data set composed of 3 clusters of 178 12-dimensional data points. The parameter k satisfies $k \le c$. The goal of the experiment is to show that the membership matrix generated by the SNCM algorithm which is sparse compared to the FCM algorithm. Due to the large number of sample points, it is inconvenient to present the complete membership matrix in the article, so we select some sample points for display. Tables 1–4 are the membership matrix results obtained by the SNCM algorithm and the FCM algorithm on the two data sets. It can be seen from the experimental results that the SNCM algorithm effectively reduces the complexity of the model.

Next, we perform experiments on the artificial data set. Figures 2(a) and 2(b) show the distribution of the two data sets, where data set (a) has four clusters and data set (b) has three clusters. Clustering is performed using the proposed algorithm, and the clustering results and the weighted connection graph are shown in Figures 2(c)–2(f). Figures 2(d) and 2(f) use the final degree of membership as the connection weight between the data point and the cluster center. The data point is connected to the cluster center. It can be seen that the points within the cluster are closely connected to the cluster center, and the points between the clusters are separated from the cluster center. It is separated, so the proposed algorithm can effectively cluster the aforementioned data sets and can effectively divide clusters with few categories.

4.3. Real Data Set. In addition, WBC, Vote, Dermatology, Dnatest, Pima, Vowel, TOX-171, and Abalone are used for experiments. These data sets are in the UCI Machine Learning Library Data Set. They cover the characteristics of various data sets such as high-dimensional and low-dimensional, multiple samples, and a few samples. The information of the night real data sets is shown in Table 5.

The experimental results on the real data set are shown in Tables 6 and 7. The folded data represent the best result, followed by the italic. Table 6 shows the ACC comparison of different algorithms under each data set. Table 7 shows the

Table 1: The membership results of the SNCM algorithm on the Wine data set.

0	0	0	0	0.4565	0	0	0	0	0
0.3568	0.3675	0.3043	0.2567	0.5435	0.2595	0.2805	0.2796	0.3714	0.3714
0.6432	0.6325	0.6957	0.7433	0	0.7405	0.7195	0.7204	0.6286	0.6286

Table 2: The membership results of the FCM algorithm on the Wine data set.

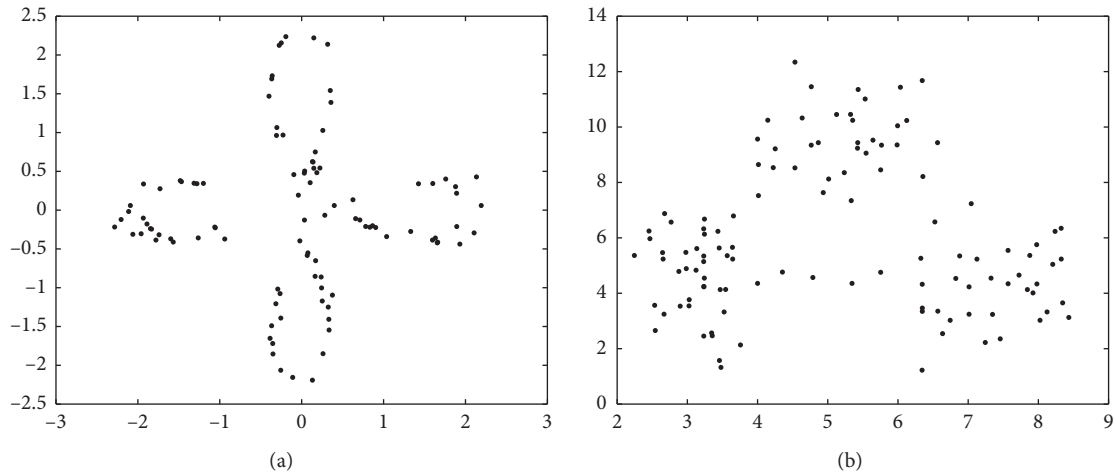
0.766196	0.717242	0.990821	0.841821	0.001096	0.863302	0.977306	0.973795	0.699161	0.699258
0.182098	0.222463	0.006692	0.103933	0.995516	0.090537	0.015821	0.018236	0.237473	0.237411
0.051706	0.060295	0.002487	0.054246	0.003388	0.046161	0.006873	0.007968	0.063366	0.063331

Table 3: The membership results of the SNCM algorithm on the aggregation data set.

0.0940	0.1643	0.1630	0.1641	0.1923	0.1954	0.1920	0.2095	0.2231	0.2245
0	0	0	0	0	0	0	0	0	0
0.8121	0.6713	0.6740	0.6718	0.6155	0.6092	0.6161	0.5810	0.5538	0.5510
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0.0940	0.1643	0.1630	0.1641	0.1923	0.1954	0.1920	0.2095	0.2231	0.2245

Table 4: The membership results of the FCM algorithm on the aggregation data set.

0.035813	0.029956	0.029049	0.028797	0.024013	0.023168	0.024526	0.020445	0.019447	0.015821
0.198913	0.194136	0.18083	0.175156	0.162905	0.154372	0.153976	0.151758	0.162358	0.126154
0.055645	0.047094	0.044304	0.043175	0.036205	0.034272	0.035478	0.030937	0.030262	0.023555
0.56412	0.617044	0.639653	0.649182	0.694951	0.710882	0.704882	0.729989	0.725561	0.785736
0.043555	0.035687	0.03367	0.032841	0.02696	0.025497	0.026545	0.022652	0.021771	0.016962
0.032043	0.024869	0.02378	0.023332	0.018517	0.017567	0.018536	0.015159	0.014101	0.011123
0.069911	0.051214	0.048715	0.047517	0.036449	0.034242	0.036057	0.02906	0.0265	0.020648



(a) (b)

Figure 2: Continued.

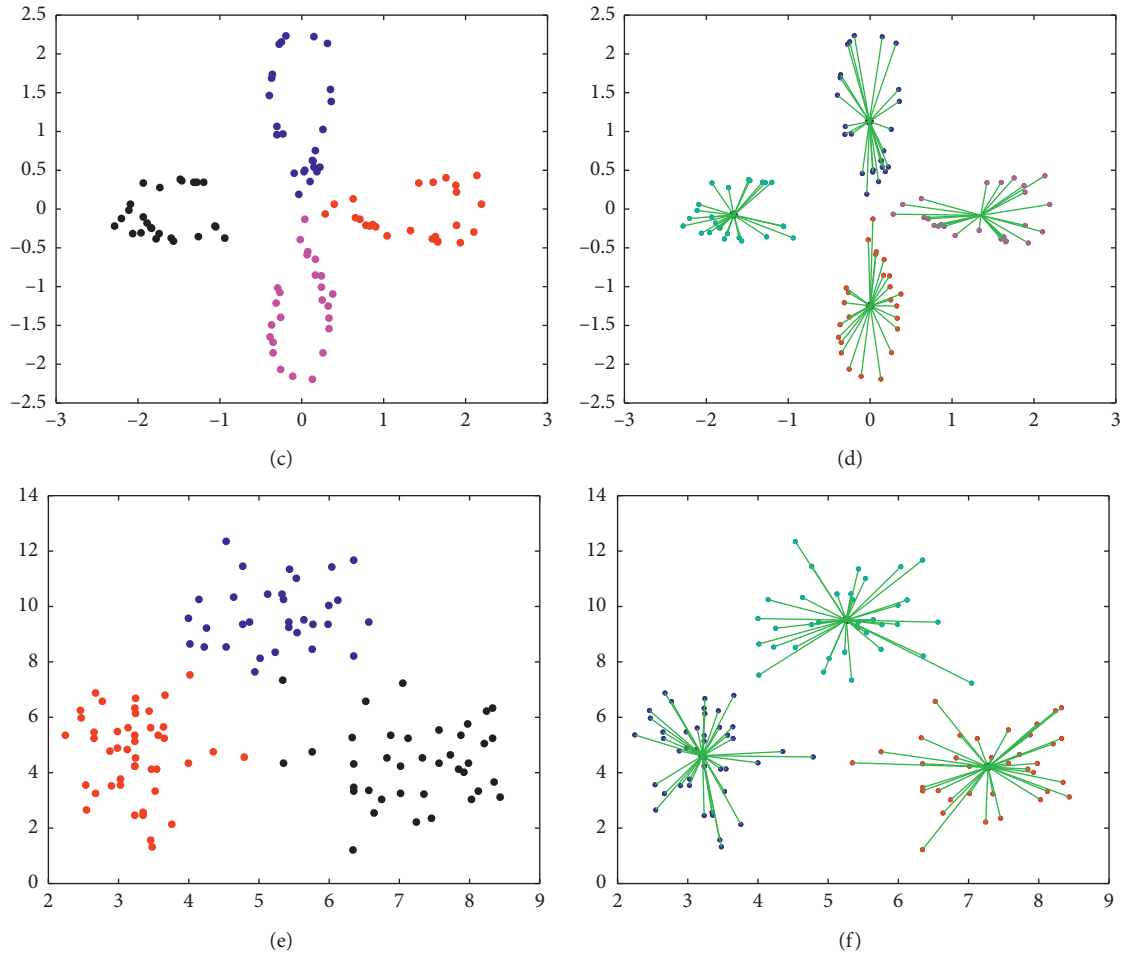


Figure 2: Connection between clustering results and weights. (a), (b) Original graph of artificial data set. (c), (e) Clustering result graph. (d), (f) Weight connection diagram.

Table 5: Real data set information.

Data set	WBC	Vote	Dermatology	Dnatest	Pima	Vowel	TOX-171	Abalone
No. of sample	683	435	366	1186	768	528	171	4177
No. of attribute	9	16	34	180	8	10	5748	7
No. of class	2	2	6	3	2	11	4	28

Table 6: ACC comparison of different algorithms under different data sets.

Algorithm	WBC	Vote	Dermatology	Dnatest	Pima	Vowel	TOX-171	Abalone
K-means	*0.9606*	0.8178	*0.7249*	*0.6830*	**0.6602**	*0.3683*	**0.4261**	0.1436
FCM	0.9561	0.8138	0.5033	0.5735	*0.6589*	0.2387	0.3977	0.1214
Rcut	0.6437	0.6170	0.3138	0.5087	0.6494	0.0975	0.3163	0.1645
Ncut	0.6515	**0.8248**	0.6967	0.5126	0.6445	0.3197	0.2690	0.1386
PS-FCM	0.9561	0.8138	0.5027	0.5698	*0.6589*	0.2321	0.3977	0.1276
INCM	0.9065	0.8000	0.5314	0.6054	0.6510	0.2708	0.3918	*0.1650*
SNCM	**0.9618**	*0.8226*	**0.7530**	**0.6984**	**0.6602**	**0.3743**	*0.4225*	**0.2172**

The bold values indicate the highest clustering accuracy (ACC), and the values in italics indicate the second highest.

NMI comparison of different algorithms under each data set. Experimental results on real data sets show that for different real data sets, the clustering algorithm SNCM is superior to other clustering algorithms in most cases. Therefore, this also confirms the effectiveness of the clustering algorithm SNCM.

Furthermore, taking the average ACC value of SNCM, the average classification performance of the algorithm is

Table 7: NMI comparison of different algorithms under different data sets.

Algorithm	WBC	Vote	Dermatology	Dnatest	Pima	Vowel	TOX-171	Abalone
K-means	*0.7436*	0.3387	**0.8231**	*0.2675*	*0.0267*	**0.4469**	**0.1418**	0.1591
FCM	0.7223	0.3333	0.3203	0.1778	0.0317	0.2397	0.0722	0.1412
Rcut	0.0042	0.0045	0.0201	0.0018	0.0010	0.0203	0.0222	0.0054
Ncut	0.0024	*0.3458*	0.6556	0.0090	0.0043	0.3833	0.0181	**0.1617**
FC-PFS	0.7223	0.3333	0.3193	0.1682	**0.0317**	0.2063	0.0722	0.1456
INCM	0.1148	0.2918	0.0117	0.1787	0.0022	0.2341	0.0685	*0.1596*
SNCM	**0.7494**	**0.3478**	*0.8079*	**0.2786**	*0.0267*	*0.4436*	*0.1355*	0.1286

The bold values indicate the highest NMI, and the values in italics indicate the second highest.

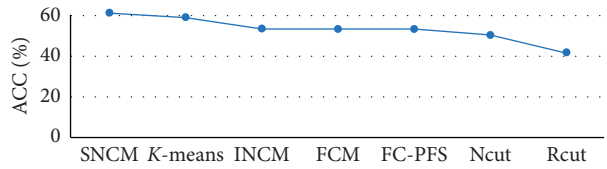


Figure 3: The average accuracy of the algorithm.

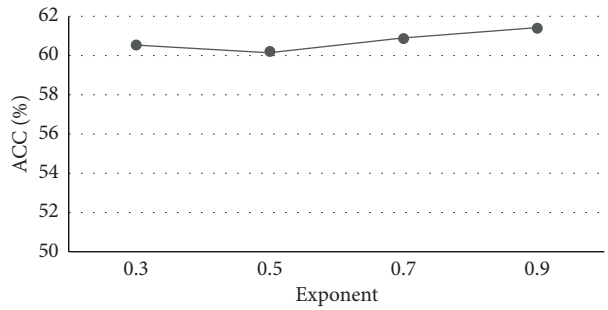


Figure 4: The average accuracy of the algorithm under different indices.

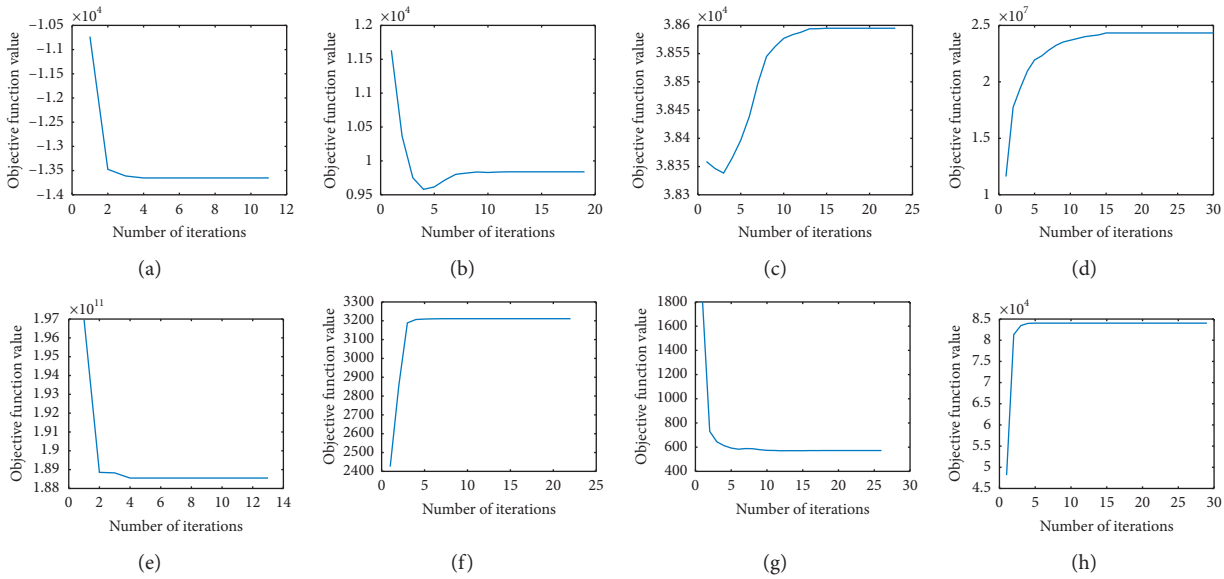


Figure 5: The convergence curves of SNCM on (a) Abalone, (b) Dermatology, (c) Dnatest, (d) Pima, (e) TOX-171, (f) Vote, (g) Vowel, and (h) WBC.

61.38%, which is higher than INCM (54.02%), FCM (53.30%), FC-PFS (53.23%), *K*-means (59.81%), *N*cut (50.72%)), and *R*cut (41.39%). The specific situation is shown in Figure 3.

For the parameters, in Figure 4, different exponents are given to verify the algorithm, and the average clustering accuracy of the proposed algorithm under different exponents is listed in the chart. We find that the clustering quality of SNCM is relatively stable. As the index increases, the accuracy of the SNCM algorithm also tends to increase. Therefore, the parameter value in the experimental part is 0.9 to improve the clustering accuracy of the SNCM algorithm.

Finally, we test the convergence of SNCM on the data sets. The results are shown in Figure 5. It can be seen that SNCM algorithm can absolutely converge with few interaction steps.

The SNCM algorithm improves the generalization ability of the algorithm by introducing regularization terms, so that the membership matrix has sparseness, and the calculation of membership considers the degree of sparseness k. Compared with the comparison algorithm, in most cases, the result of this algorithm is better than that of the comparison algorithm. The experiment of the algorithm on multiple data sets can also illustrate this point, and the parameter k has great influence on results.

5. Conclusion

In this paper, we have proposed a novel method, called neutrosophic clustering algorithm based on sparse regular term constraint. Different from the previous neutrosophic clustering algorithm, the algorithm proposed in this paper can handle the case of ambiguity $m = 1$, not limited to the condition of $m > 1$. Furthermore, the regular term is introduced to make the algorithm sparse, thereby reducing the computational complexity of the algorithm. Moreover, we propose a method to simplify the process of determining regularization parameters and improve the clustering effect. In addition, a large number of experiments show that the clustering results of the proposed algorithm on artificial data sets and real data sets are mostly better than other clustering algorithms. However, the parameter k in the algorithm has a greater impact on the clustering effect. So, we will focus on this in the future.

Data Availability

The data in this article come from the data set in the UCI Machine Learning Library and are available in the official database.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (61976130), Shaanxi Provincial Key Research and Development Program (2018KW-021), Shaanxi Provincial Natural Science Foundation of China (2020JQ-923), and Shaanxi's Scientific and Technological Commission (Project no. 2019KRM072).

References

[1] L. Sun, L. Wang, W. Ding, Y. Qian, and J. Xu, "Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 1, 19 pages, 2021.

[2] L. Sun, T. Yin, W. Ding, and Y. Xu, "Multilabel feature selection using ML-ReliefF and neighborhood mutual information for multilabel neighborhood decision systems," *Information Sciences*, vol. 537, pp. 401–424, 2020.

[3] S. Miyamto, H. Ichihashi, and K. Honda, "Algorithms for fuzzy clustering," *Studies in Fuzziness and Soft Computing*, vol. 229, pp. 157–169, 2008.

[4] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[5] J. ye, J. Zhan, and Z. Xu, "A novel decision-making approach based on three-way decisions in fuzzy information systems," *Information Sciences*, vol. 541, pp. 362–390, 2020.

[6] K. Zhang, J. Zhan, and X. Wang, "TOPSIS-WAA method based on a covering-based fuzzy rough set: an application to rating problem," *Information Sciences*, vol. 539, pp. 397–421, 2020.

[7] J. Zhan, H. Jiang, and Y. Yao, "Covering-based variable precision fuzzy rough sets with PROMETHEE-EDAS methods," *Information Sciences*, vol. 538, pp. 314–336, 2020.

[8] M. Gong, Z. Zhou, and J. Ma, "Change detection in synthetic aperture radar images based on image fus-ion and fuzzy clustering," *Image Processing IEEE Transactions*, vol. 21, no. 4, pp. 2141–2151, 2012.

[9] J. V. d. Oliveira and W. Pedrycz, *Advances in Fuzzy Clustering and its Applications*, John Wiley and Sons, Hoboken, NJ, USA, 2007.

[10] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM.: The fuzzy c-means clustering algorithm," *Computers and Geoences*, vol. 10, no. 2-3, pp. 191–203, 1984.

[11] C. Hwang and 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 and 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] K. T. Atanassov, "Intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 20, no. 1, pp. 87–96, 1986.

[14] T. Chaira, "A novel intuitionistic fuzzy *C* means clustering algorithm and its application to medical images," *Applied Soft Computing*, vol. 11, no. 2, pp. 1711–1717, 2011.

[15] N. Pelekis, D. K. Iakovidis, E. E. Kotsifakos, and I. Kopanakis, "Fuzzy clustering of intuitionistic fuzzy data," *International Journal of Business Intelligence and Data Mining*, vol. 3, no. 1, pp. 45–65, 2008.

[16] H. Zhao, Z. Xu, and Z. Wang, "Intuitionistic fuzzy clustering algorithm based on Boole matrix and association measure," *International Journal of Information Technology and Decision Making*, vol. 12, no. 1, pp. 95–118, 2013.

[17] B. C. Cuong, "Picture fuzzy sets," *Journal of Computer Ence & Cybernetics*, vol. 30, no. 4, pp. 409–420, 2014.

[18] P. H. Thong and L. H. Son, "Picture fuzzy clustering: a new computational intelligence method," *Soft Computing*, vol. 20, no. 9, pp. 3549–3562, 2016.
[19] F. Smarandache, "Neutrosophy, A new branch of pilosophy," *Multiple Valued Logic/An International Journal*, vol. 8, no. 3, 297 pages, 2002.
[20] X. Zhang, Q. Hu, F. Smarandache, and X. An, "On neutrosophic triplet groups: basic properties, NT-subgroups, and some notes," *Symmetry*, vol. 10, no. 7, 289 pages, 2018.
[21] Y. Ma, X. Zhang, X. Yang, and X. Zhou, "Generalized neutrosophic extended triplet group," *Symmetry*, vol. 11, no. 3, 327 pages, 2019.
[22] F. Karaaslan, "Correlation coefficients of single-valued neutrosophic refined soft sets and their applications in clustering analysis," *Neural Computing and Applications*, vol. 28, no. 9, pp. 2781–2793, 2017.
[23] P. Maji, "Neutrosophic soft set," *Annals of Fuzzy Mathematics and Informatics*, vol. 5, no. 1, pp. 157–168, 2013.
[24] J. Ye, "Fault diagnoses of hydraulic turbine using the dimension root similarity measure of single-valued neutrosophic sets," *Intelligent Automation and Soft Computing*, vol. 24, no. 1, pp. 1–8, 2017.
[25] J. Ye, "Single-valued neutrosophic minimum spanning tree and its clustering method," *Journal of Intelligent Systems*, vol. 23, no. 3, pp. 311–324, 2014.
[26] J. Ye, "Clustering methods using distance-based similarity measures of single-valued neutrosophic sets," *Journal of Intelligent Systems*, vol. 23, no. 4, pp. 379–389, 2014.
[27] Y. Guo, A. Sengur, and N. C. M. ", "NCM: neutrosophic c-means clustering algorithm," *Pattern Recognition*, vol. 48, no. 8, pp. 2710–2724, 2015.
[28] J. Shan, H. D. Cheng, and Y. Wang, "A novel segmentation method for breast ultrasound images based on neutrosophic l-means clustering," *Medical Physics*, vol. 39, no. 9, p. 5669, 2012.
[29] S. Ye and J. Ye, "Dice similarity measure between single valued neutrosophic multisets and its application in medical diagnosis," *Neutrosophic Sets and Systems*, vol. 6, pp. 48–53, 2014.
[30] H.-y. Zhang, P. Ji, J.-q. Wang, and X.-h. Chen, "A neutrosophic normal cloud and its application in decision-making," *Cognitive Computation*, vol. 8, no. 4, pp. 649–669, 2016.
[31] D. Koundal, S. Gupta, and S. Singh, "Automated delineation of thyroid nodules in ultrasound images using spatial neutrosophic clustering and level set," *Applied Soft Computing*, vol. 40, pp. 86–97, 2016.
[32] Y. Guo, A. Amira, and S. Florentin, "A novel skin lesion detection approach using neutrosophic clustering and adaptive region growing in dermoscopy images," *Symmetry*, vol. 10, no. 4, 119 pages, 2018.
[33] P. Burillo and H. Bustince, "Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets," *Fuzzy Sets and Systems*, vol. 78, no. 3, pp. 305–316, 1996.
[34] S. J. Nanda, I. Gulati, R. Chauhan, R. Modi, U. Dhaked, and U. Dhaked, "A *K*-Means-Galactic swarm optimization-based clustering algorithm with otsu's entropy for brain tumor detection," *Applied Artificial Intelligence*, vol. 33, no. 2, pp. 152–170, 2018.
[35] J. Shi and J. Malia, "Normalized cuts and image segmentation," *IEEE Trans on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
[36] L. Hagen and A. B. Kahng, "New spectral methods for ratio cut partitioning and clustering," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 11, no. 9, pp. 1074–1085, 1992.
[37] E. Rashno, B. Minaei-Bidgoli, and Y. Guo, "An effective clustering method based on data indeterminacy in neutrosophic set domain," *Engineering Applications of Artificial Intelligence*, vol. 89, 2020.