Egyptian Informatics Journal 24 (2023) 100417

Contents lists available at ScienceDirect

Egyptian Informatics Journal

journal homepage: www.sciencedirect.com

Multi-fuzzy clustering validity index ensemble: A Dempster-Shafer theory-based parallel and series fusion

Hong-Yu Wang, Jie-Sheng Wang *, Guan Wang

School of Electronic and Information Engineering, University of Science & Technology Liaoning, Anshan 114044, China

A R T I C L E I N F O

Keywords:
Fuzzy clustering validity index
Information ensemble
Fuzzy clustering algorithms
D-S evidence theory
Parallel and serial ensemble

A B S T R A C T

Clustering validity evaluation is a key part in clustering process. To adapt the complex data structure, the traditional fuzzy clustering validity index (FCVI) is designed more complex. The weighted combined validity evaluation method (WCVEM) is simple in structure but difficult in weight selection. Therefore, this paper proposed an ensemble method based on multi-fuzzy clustering algorithms and multi-FCVI. Firstly, multi-FCVI are calculated by using the multiple sets of cluster centers and membership degrees that obtained by multi-fuzzy clustering algorithms. This can improve the robustness of the multi-FCVI. Secondly, multi-FCVI are ensembled by Dempster-Shafer (DS) theory. The validity index basic probability assignment function can be obtained by calculating the credibility of each validity index with different clusters number. Finally, the decision module is used to output the optimal clusters number. This paper ensembles multi-fuzzy clustering algorithms, multi-FCVI, and the DS theory by using series and parallel structure to verify performance of the proposed model and the degree of information retention of the FCVI. The proposed method is simple in structure and does not need to be select weighted. 6 artificial datasets and 12 UCI datasets were selected to simulate and verify the method. When facing different data, the simulation results show that the parallel structure has the highest accuracy, and the series structure is even worse than the weighted method in some datasets. In addition, the paper changes the value of fuzzy weighted, and experimental results show that the ensemble method has better stability than other methods in the face of different fuzzy weighted strategy.

1. Introduction

With the continuous popularization and application of smart devices, a large amount of data is accumulated. How to get useful information from those data has become a major problem facing mankind, and data mining technology has come into being [1]. At present, Data mining has been applied to medical, education, finance and other fields. As one means of data mining, cluster analysis has been widely concerned and studied by scholars [2,3]. Clustering is unsupervised learning, it is applicable to data sets without prior knowledge, such as the inherent structure cognition of biological populations, commercial location based on user location information, traffic recommendation of search engines, and other more fields [4,5].

The core of clustering is similar samples are clustered into one cluster, and dissimilar samples are divided into different clusters [6]. The K-means clustering algorithm [7] is the first membership degree division algorithm, it divides samples into different clusters, that is, either 0 or 1. To overcome this shortcoming and make samples in different classes but have similarities have a better division method [8], Bezdek extending hard clustering like K-means to fuzzy clustering by introduced fuzzy sets [9] and proposed the Fuzzy c-means clustering algorithm (FCM) [10]. The FCM got widely study by scholars because its simple process and low computational complexity [11].

For the same dataset, various clustering algorithms will obtain different results. To improve the adaptability of clustering algorithms, the clustering ensemble method was first proposed by Strehal in 2002 [12] and has since been widely used by scholars to improve traditional clustering algorithms and enhance the ability to process datasets of clustering algorithms. A selective clustering ensemble means based on repeated clustering algorithm proposed by Hong [13] in 2009. Berikov proposed a probability clustering ensemble means based on clustering combination of the fuzzy clustering algorithm and weighted consensus matrix [14]. Bagherinia proposed an ensemble framework that not require the characteristics of samples weighted by fuzzy clustering [15]. Mojarad proposed a clustering ensemble based on iterative Fusion of base clusters [16]. Clustering ensemble technology is to integrate the

* Corresponding author.
E-mail addresses: wanghongyuww@126.com (H.-Y. Wang), wang_jiesheng@126.com (J.-S. Wang), 480433838@qq.com (G. Wang).

https://doi.org/10.1016/j.eij.2023.100417
Received 15 April 2023; Received in revised form 9 July 2023; Accepted 3 November 2023
Available online 11 November 2023
1110-8665/© 2023 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Computers and Artificial Intelligence, Cairo University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

basic clustering algorithms by some information fusion method and obtains better clustering results than the basic clustering algorithm. The scheme of clustering ensemble mainly includes the ensemble by changing different parameters, the ensemble between different algorithms, and the ensemble of dataset feature levels [17,18].

At present, fuzzy clustering needs to face four problems, including dataset preprocessing, clustering analysis, clustering validity evaluation, and post clustering processing [19]. Their relationship is shown in Fig. 1. It is unknown how many clusters these data should be divided into, so clustering is used to classify the data set. Cluster analysis can only give the corresponding cluster situation of different clusters. In order to select the best clustering result, it is necessary to use the validity evaluation to get the value of the best cluster number. Therefore, the validity evaluation directly determines the quality of the clustering results in the whole clustering process.

Research on the fuzzy clustering validity evaluation methods have two directions: fuzzy clustering validity index (FCVI) and weighted combination validity evaluation method (WCVEM). Study on FCVI is based on the following 2 aspects: (1) FCVI only based on membership. Such as PC index [20], Modified Partition Coefficient [21] and P index proposed by Chen [22]. (2) FCVI based with membership and data geometry. Such as Xie-Beni index [23], Wang proposed the VHY validity index [24], Naderipour proposed a fuzzy cluster-validity index based on the topology structure [25], and Tang proposed a fuzzy clustering validity index induced by triple center relation [26]. Study on the WCVEM only based on the way of weighted combination. In 2005, Sheng given the weighted sum validity function (WSVF) from the perspective of hard clustering [27]. Based on Shen's idea, Dong introduced fuzzy clustering into WSVF and proposed the fuzzy weighted sum clustering function (FWSCF) [28]. Wang and Liu proposed component-wise design method of FCVI with different construction method [29,30].

In order to cope with the increase in the structural complexity and number of samples of datasets, the structural design of FCVI is becoming more and more complex, and the calculation amount is also increasing. However, the ability of FCVI to adapt to datasets has not been fundamentally solved, and some FCVI will fail due to the large gap between the intra-class compactness and the inter-class separation degree. WCVEM is simpler and more stable than FCVI. However, the weighting method of WCVEM is difficult to determine and the calculation is very large.

Therefore, this paper proposed a clustering validity ensemble model (DSMFCE) based on the combination of multi-fuzzy clustering algorithms, multi-FCVI and DS theory. Clustering algorithm ensemble uses the membership and clustering center of multi-fuzzy clustering algorithms to improve the generalization ability of the DSMFCE. Then inputs the updated cluster centers and membership degrees into multi-FCVI which can improve the adaptability of DSMFCE. The values of multi-FCVI are ensemble by D-S evidence theory to obtain each validity index basic probability assignment (BPA). Finally, the decision module is used to output the optimal number of the clustering. The paper also discusses the ensemble framework of multi-fuzzy clustering algorithms, multi-FCVI, and DS theory. From the perspective of parallel and serial, three ensemble model frameworks (DSMFCE-I, DSMFCE-II and DSMFCE-III) are proposed. Finally, this paper uses manual datasets and UCI datasets to simulate and verify the proposed DSMFCE. The simulations show that DSMFCE has good stability and accuracy in clustering validity evaluation.

2. Various fuzzy clustering algorithms

2.1. FCM clustering algorithm (FCM)

The FCM is the earliest and the most frequently used algorithms in fuzzy partition algorithm. The basic principles of FCM is classifying n samples $x_1, x_2, \cdots, x_n$ in dataset X into c classes, and calculates the minimum value of loss function shown in Eq. (1).

$$J_{FCM}(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n}\left(u_{ij}\right)^{m}\|x_j - v_i\|^2 \tag{1}$$

As shown in the above function, c is the number of clusters. $m\in(1,\infty)$ is the fuzziness index, which is the measure the fuzziness of the generic degree in each class of X. x_j is the sample, v_i is the cluster center calculation as Eq. (2) and $\|x_j - v_i\|$ is the Euclidean distance between x_j and v_i.

$$v_i = \frac{\sum_{j=1}^{n} u_{ij}{}^{m} \bullet x_i}{\sum_{j=1}^{n} u_{ij}{}^{m}} \tag{2}$$

$u_{ij}(0\leq u_{ij} \leq 1)$ is the membership of x_j to v_i that satisfy Eq. (3). $u_{ij} \in U$, where U is a $c \times n$ matrix made up with u_{ij}. $V = \{v_1, v_2, \cdots, v_c\}$ is a set of v_i. $J_{FCM}(U,V)$ represents the sum of square error of the same x_j in different c, the optimal c can be obtained by the least error value $\min J_{FCM}(U,V)$.

$$u_{ij} = \left[\sum_{k=1}^{c}\left(\frac{\|x_j - v_i\|}{\|x_j - v_k\|}\right)^{2/(m-1)}\right]^{-1} \tag{3}$$

Initialize membership U_0 and cluster center V_0. $\sum_{i=1}^{c} u_{ij} = 1$, $0 < \sum_{j=1}^{n} u_{ij} < n$, where $i = 1, 2, \cdots, c$, $j = 1, 2, \cdots, n$.

The procedure of the FCM is described as follows:

Step 1: Fix clustering parameter c, fuzzy weighted m and convergence threshold ε;

Step 2: Initialize the fuzzy partition membership matrix U_0 and cluster center V_0;

Step 3: Update the $U = \left(u_{ij}\right)_{c\times n}$ according to Eq. (3);

Step 4: Update the $V = \{v_1, v_2, \cdots, v_c\}$ according to Eq. (2);

Step 5: Calculate $e = \|U_{t+1} - U_t\|$. If $e \leq \varepsilon$ (ε is a threshold that usually from 0.001 to 0.01), the FCM stops and outputs optimal c. Else $U_t = U_{t+1}$ turn to Step 2 and repeat the flow.

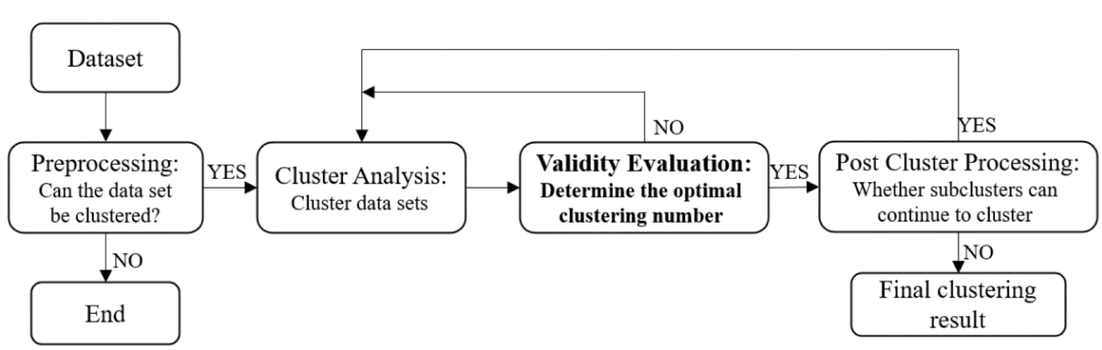


Fig. 1. Clustering process.

2.2. Kernel-based FCM clustering algorithm (KFCM)

Chen proposed an novel the FCM clustering algorithm **(**KFCM**)** [31]. Different with FCM, KFCM adopts the Gauss kernel function to measure the distance between x_j and v_i. KFCM uses the kernel function to map the x_j of the X to the feature space $\Phi(x_j)$, and then analyzes and calculates to obtain the final optimal partition. First, define the mapping from $X = x_1, x_2, \cdots, x_n$ to feature space F: $\Phi : X \rightarrow \Phi(x_j) \in F$. Then the adopted kernel function is defined as Eq. (4).

$$K(x_i, x_j) = \langle \Phi(x_i) \bullet \Phi(x_j) \rangle \tag{4}$$

where, $\langle \bullet \rangle$ is the Euclidean inner product of $\Phi(x_i)$ and $\Phi(x_j)$. The distance calculation method in FCM is changed by the kernel function. The loss function is defined in Eq. (5).

$$J_{KFCM}(U, V) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}{}^{m} \|\Phi(x_j) - \Phi(v_i)\|^2 \tag{5}$$

where, $\|\Phi(x_j) - \Phi(v_i)\|^2 = K(x_j, x_j) + K(v_i, v_i) - 2K(x_j, v_i)$. (x_j, x_j), $K(v_i, v_i)$ and $K(x_j, v_i)$ select the Gaussian radial basis function (RBF) as the operation formula, which is shown in Eq. (6).

$$K(a, b) = \exp\left(-\frac{\|a - b\|^2}{\tau^2} \right) \tag{6}$$

where, $\tau > 0$. The value of $K(x_j, x_j)$ and $K(v_i, v_i)$ is 1, so the $J_{KFCM}(U, V)$ can be simply expressed as Eq. (7).

$$J_{KFCM}(U, V) = 2 \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}{}^{m} \left(1 - K(x_j, v_i)\right) \tag{7}$$

Initialize $U$and$V$, find the minimum value of loss function $J_{KFCM}(U, V)$. The Lagrange extreme condition was used to deduce U as Eq. (8) and V as Eq. (9):

$$u_{ij} = \frac{\left(1 - K(x_j, v_i)\right)^{\frac{-1}{m-1}}}{\sum_{k=1}^{c} \left(1 - K(x_j, v_k)\right)^{\frac{-1}{m-1}}} \tag{8}$$

$$v_i = \frac{\sum_{j=1}^{n} u_{ij}{}^{m} K(x_j, v_i) x_j}{\sum_{j=1}^{n} u_{ij}{}^{m} K(x_j, v_i)} \tag{9}$$

The procedure of the KFCM is shown as follows:
Step 1: Input the parameters of Gaussian radial function τ, c, m and ε;
Step 2: Initialize U_0 and V_0 with the FCM;
Step 3: Update the $U = (u_{ij})_{c \times n}$ according to Eq. (8);
Step 4: Update the $V = \{v_1, v_2, \cdots, v_c\}$ according to Eq. (9);
Step 5: Calculate $e = \|U_{t+1} - U_t\|$. When $e \le \varepsilon$, or there exists $i(1 \le i \le c)$ such that $\sum_{i=1}^{c} u_{ij} = 0$, the KFCM stops and outputs optimal c. Else $U_t = U_{t+1}$ turn to Step 2 and repeat the process.

2.3. Weighted Kernel-based fuzzy clustering algorithm (WKFCM)

The FCM assumes that all the attributes of data are equally related. But in most states, some attributes are equally related and some attributes are not so important. Therefore, sheng proposes WKFCM [32] based on weighted according to the importance of sample features. Given a dataset $X = \{\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_n}\}$, for each sample $\mathbf{x_j}$ has l attributes, $\mathbf{x_j}$ can be expressed as $\mathbf{x_j} = \{x_{j1}, x_{j2}, \cdots, x_{jl}\}$. Define the mapping from dataset X to feature space F: $\Phi : X \rightarrow \Phi(x_{jl}) \in F^l$. The purpose of WKFCM is to calculate the minimum of loss function as shown in Eq. (10).

$$J_{WKFCM}(U, V) = \sum_{i=1}^{c} \sum_{j=1}^{n} \sum_{k=1}^{l} w_{ik}{}^{\beta} u_{ij}{}^{m} \|\Phi(x_{jk}) - \Phi(v_{ik})\|^2 \tag{10}$$

where $u_{ij} \in [0, 1]$, $\sum_{i=1}^{c} u_{ij} = 1$, $1 \le j \le n$. w_{ij} is the weighted of the $k-$ th feature of the $i-$ th cluster, $w_{ik} \in [0, 1]$, $\sum_{k=1}^{l} w_{ik} = 1$, $1 \le i \le c$. $\beta > 0$, $m > 0$. $V = \{\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_c}\}$, each cluster center $\mathbf{v_i}$ also has l attributes $\mathbf{v_i} = \{v_{i1}, v_{i2}, \cdots, v_{il}\}$. According to the KFCM algorithm, Gaussian kernel basis function is selected to get J_{WKFCM}, which is shown in Eq. (11).

$$J_{WKFCM}(U, V) = 2 \sum_{i=1}^{c} \sum_{j=1}^{n} \sum_{k=1}^{l} w_{ik}{}^{\beta} u_{ij}{}^{m} \left(1 - K(x_{jk}, v_{ik})\right) \tag{11}$$

Under the constraints of U and V, the objective function $J_{WKFCM}(U, V)$ is minimized. The Lagrange extremum condition was used to deduce U shown in Eq. (12) and V shown in Eq. (13).

$$u_{ij} = \sum_{r=1}^{c} \frac{\sum_{k=1}^{l} w_{ik}{}^{\beta} \left(1 - K(x_{jk}, v_{ik})\right)^{\frac{-1}{m-1}}}{\sum_{k=1}^{l} w_{ik}{}^{\beta} \left(1 - K(x_{jk}, v_{rk})\right)^{\frac{-1}{m-1}}} \tag{12}$$

$$\begin{cases} v_{ik} = 0, if\, w_{ik} = 0 \\ v_{ik} = \dfrac{\sum_{j=1}^{n} u_{ij}{}^{m} K(x_{jk}, v_{ik}) x_{jk}}{\sum_{j=1}^{n} u_{ij}{}^{m} K(x_{jk}, v_{ik})}, if\, w_{ik} \ne 0 \end{cases} \tag{13}$$

$$w_{ik} = \sum_{t=1}^{l} \frac{\sum_{j=1}^{n} u_{ij}{}^{m} \left(1 - K(x_{jk}, v_{ik})\right)^{\frac{-1}{\beta-1}}}{\sum_{j=1}^{n} u_{ij}{}^{m} \left(1 - K(x_{jt}, v_{it})\right)^{\frac{-1}{\beta-1}}} \tag{14}$$

The procedure of the WKFCM is shown as follows:
Step 1: Input the τ, c, m, ε and weighted $w_{ik} = 1/l$;
Step 2: Initialize U_0 and V_0 with FCM;
Step 3: Update the $U = (u_{ij})_{c \times n}$ according to Eq. (12);
Step 4: Update the $V = \{\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_c}\}$ for $\mathbf{v_i} = \{v_{i1}, v_{i2}, \cdots, v_{il}\}$ according to Eq. (13);
Step 5: Update the weighted coefficient w_{ik} according to Eq. (14);
Step 6: Calculate $e = \|U_{t+1} - U_t\|$. When $e \le \varepsilon$, or there exists $i(1 \le i \le c)$ such that $\sum_{i=1}^{c} u_{ij} = 0$, the WKFCM stops and outputs optimal c. Else $U_t = U_{t+1}$ turn to Step 2 and repeat the process.

2.4. KFCM based on genetic algorithm (GAKFCM)

GAKFCM [33] uses the advantage of global convergence of the Genetic Algorithm [34,35] to optimize the V_0 and then uses the KFCM algorithm to clustering. Genetic Algorithm is introduced to enhance the search speed and accelerate of GAKFCM, will not fall into the local gradient. $V = [v_1, v_2, \cdots, v_c]$ is the clustering center matrix. Randomly selected several samples from $X = x_1, x_2, \cdots, x_n$ as cluster, and their mean values are calculated as the initial cluster centers as c cluster centers are obtained as an individual of the initial population by c calculation. Then, the same method is used to generate n individuals as a population $P = \{V_1, V_2, \cdots, V_n\}$.

The loss function is used to judge the quality of individuals in the population and determine the search direction. In the KFCM, the loss function $J_{KFCM}(U, V)$ of individuals in the population is derived from Eq. (7). According to the $J_{KFCM}(U, V)$, when take the minimum value, the optimal c is obtained. Therefore, the GAKFCM loss function is defined with the influence of $J_{KFCM}(U, V)$, as shown in Eq. (15):

$$f(U, V) = \frac{1}{1 + J_{KFCM}(U, V)} = \frac{1}{1 + 2\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}{}^{m} \left(1 - K(x_j, v_i)\right)} \tag{15}$$

The procedure of the GAKFCM is shown as follows:
Step 1: Input the τ, c, m and ε. Genetic initial crossover probability p_{c0} (generally 0.4–0.99), initial probability of variation p_{m0} (generally 0.0001–0.1), genetic evolution algebra T, genetic stop threshold δ;
Step 2: Randomly generate n cluster center individuals as a population $P = \{V_1, V_2, \cdots, V_n\}$;
Step 3: Calculate $f(U, V)$ of individuals in the population, relative fitness $f(j) = q(1 - q)^{j-1}$, crossover probability $f_c(t) = p_{c0}\left(1 - \frac{t}{T}\right)$ and variation probability $f_m(t) = p_{m0}\left(1 - \frac{t}{T}\right)$, where q is the parameter of

(0,1),j is the sequence number, and t is the current iteration times;

Step 4: The $t+1$ generation population was formed by selection, crossover and mutation of the t generation population;

Step 5: Calculate $\overline{f(U,V)} = \frac{1}{n}\sum_{k=1}^{n} f(k)$. if $|\overline{f(t)} - \overline{f(t+1)}| > \delta$, then $t = t+1$, turn to Step 3, else turn to Step 6;

Step 6: Output the V, and use the KFCM to get the final clustering number.

3. Fuzzy clustering validity evaluation methods

3.1. Fuzzy clustering validity indexes

The FCVI can be divided into two types. The first is FCVI only based with membership as shown in Table1. Bezdek first proposed the partition coefficient (V_{PC}) and partition entropy (V_{PE}) [36], the structure is simple and the calculation is small, but it cannot overcome the monotonicity trend. Silva proposed V_{MPC} to overcome the monotony of V_{PE}. Chen proposed the P index (V_P) by introducing the overlap degree, which can better adapt to the overlap samples.

The other is FCVI based with membership and dataset geometry as shown in Table2. Xie and Beni proposed the XB index (V_{XB}) by the ratio of inter-class separation degree to intra-class compactness. Fukuyam and Sugeno replaced the $\|v_i - v_j\|$ with the $\|v_i - \overline{v}\|$ to calculate the degree of separation between classes, and proposed the FS index (V_{FS}) [37], which gives a better grasp of the overall degree of separation between classes. Know combines V_{XB} and V_{FS} to propose the VK validity function (V_K) [38]. Wu changed the scale operation mode and proposed the new validity function (V_{PCAES}) [39] by using the exponential operation method. V_{PCAES} will normalize the distribution coefficient and index and can cope with noisy data well. However, V_{PCAES} relies too much on the clustering center and falls into local optimization. Pakhiar and Bandyopadhyay proposed the PBMF clustering validity function (V_{PBMF}) in the form of product [40]. Wu introducted $\underset{i \neq j}{\text{median}}\|v_i - v_j\|^2$ to improve V_{XB} and proposed the WL validity function (V_{WL}) [41] so as to better deal with noisy data. Haouas improved V_{WL} by introducing a penalty term and proposed HF index (V_{HF}) [42] to prevent V_{WL} from failing due to too small denominator value. Zhu also introducing a penalty term to suppress the values of compactness and separation, and proposed VZ index (V_Z) [43]. Wang proposed the new index (V_{HY}) by combining compactness, separation and overlap to better deal with overlapping data.

In Table 1 and Table 2, where u_{ij} is the fuzzy membership matrix, x_j is the sample n is the number of x_j, c is the clustering number, m is the fuzziness index, v_i and v_k are two different cluster center, $\|v_i - x_j\|$ is the distance between the v_i and the x_j, $\overline{v} = \frac{\sum_{i=1}^{c} v_i}{c}, u_{mj} = \min_{1\le i\le c}\sum_{j=1}^{n} u_{ij}^2, \beta_T = \frac{1}{c}\sum_{i=1}^{c}\|v_i - \overline{v}\|^2$, and $median\|v_i - v_k\|^2$ is the median distance between v_i and v_k. In the absence of any special explanation in this paper, all the operations are matrix operations.

Table 1
FCVI only based on membership.

Proposer	Validity Index Description	Optimal c
Bezdek and Pal [20]	$V_{PC} = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^2$	Max
Simovici [36]	$V_{PE} = -\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}[u_{ij}\log_a(u_{ij})]$	Min
Silva [21]	$V_{MPC} = 1 - \frac{c}{c-1}(1 - V_{PC})$	Max
Chen and Linkens [22]	$V_P = \frac{1}{n}\sum_{j=1}^{n}\max_i(u_{ij}) - \frac{1}{k}\sum_{i=1}^{c-1}\sum_{k=i+1}^{c}\left[\frac{1}{n}\sum_{j=1}^{n}\min(u_{ij}, u_{kj})\right]$	Max

3.2. Combined validity evaluation method

The study on the WCVEM are not deep at present, mainly using a weighted method for composition. The WCFCVE are shown in Table 3, The WSVF proposed by Sheng is based on the hard clustering algorithm. $\sum_{i=1}^{r} w_i = 1, r = 6$, where r means then number of validity functions formed. $f_i(x)$ is constructed by $1/V_{DB}$, V_{SIL}, V_D, V_{33}, V_{CH}, and V_{PBM} [44–49]. The WSVF takes an averaged-weighted approach.

Dong proposed the FWSVF by extending the idea of weighted combinations to fuzzy clustering through WSVF, where $r = 4$, $f_i(x)$ is $1/V_{XB}$, $1/V_{PE}$, V_{PC} and V_{PBMF}. And w_i needs to change according to the dataset. The FWSVF takes an averaged-weighted approach. Li proposed the WSCVI [50], the weighting method is consistent with the FWSVF, where r represents the number of validity functions, where $r = 10$, $\sum_{i=1}^{r} w_i = 1$, $0 \le w \le 1$, CVI_i is constituted by $1/V_{PC}$, $1/V_{NPC}$, V_{PE}, V_{NPE}, V_{XB}, V_K, $1/V_{PBMF}$, V_{FS}, V_T, and V_{SC}.

Wu proposed the DWSVF [51] based on dynamic weighted, where $w_i = (\text{sum}f_i - f_i)/(\text{sum}f_i)$, which can be changed because of the different validity index, and $f_i(x)$ is $1/V_{MPC}$, V_{XB} and V_{PBMF}. Wang proposed HWCVF [24] based with a hybrid weighted. $w_{hybrid} = \theta w_{object} + (1-\theta) w_{subject}$,$w_{object}$ is experts weighted, $w_{subject}$ is information entropy weighted, θ is the equilibrium factor,$F_i(x)$ is the V_{MPC}, V_{XB}, V_K, V_P, V_{PBMF}, V_{WL}, V_Z, V_{HY} after standardization.

4. DS theory and FCVI's BPA matrix

4.1. DS theory

The embryonic DS theory was proposed by Dempster in 1967 [52]. Shafer introduce the trust functions into the embryonic DS theory proposed an intact uncertainty reasoning theory, that is the Dempster-Shafer (DS) theory [53]. The DS theory is often used to model multi-object information fusion, because it can model uncertainty flexibly and effectively without prior probabilities [54,55].

In the DS theory, Θ is the recognition frame consisted with a finite set of hypotheses, and all hypothesis are mutually exclusive. z is the number of hypotheses, and the subset A of Θ is called proposition. The power set 2^Θ of Θ is composed of all A, can be expressed as Eq. (16):

$$\begin{cases} 2^\Theta = \{A | A \subseteq \Theta\} \\ \Theta = \{\theta_1, \theta_2, \cdots, \theta_z\} \end{cases} \tag{16}$$

If there is a function m whose basic probability distribution is a one-to-one mapping from 2^Θ to [0,1], and satisfy the conditions of Eq. (17), m is the BPA function of Θ:

$$\begin{cases} \sum_{A\subseteq\Theta} m(A) = 1 \\ m(\varnothing) = 0 \end{cases} \tag{17}$$

where, $m(A)$ is the confidence degree of A. When $m(A) > 0$, A is named focus element, and especially $m(\varnothing) = 0$ means that A does not belong to the recognition frame Θ. $m(A)$ represents the confidence that only A belongs to Θ. The uncertainty of event A can be described as $[Bel(A), Pl(A)]$, Bel and Pl are the confidence function and likelihood function of event A, respectively, defined as Eqs. (18) and (19):

$$Bel(A) = \sum_{B\subset A} m(B) \tag{18}$$

$$Pl(A) = \sum_{B\cap A \neq \varnothing} m(B) \tag{19}$$

$[Bel(A), Pl(A)]$ are called confidence intervals, which are the lower and upper bound of the confidence degree of A, respectively. For different sources of evidence that are independent of each other, there are different BPA functions. The DS theory synthesis formula uses an orthogonal sum to synthesize different BPA functions into a new BPA

Table 2
FCVI based on geometric structure and membership.

Proposer	Validity Index Description	Optimal c
Xie and Beni [23]	$V_{XB} = \dfrac{\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}{}^{m}\Vert v_i - x_j\Vert^2}{\min_{i\neq k}\Vert v_i - v_k\Vert^2}$	Min
Fukuyama and Sugeno [37]	$V_{FS} = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m}\left(\Vert x_j - v_i\Vert^2 - \Vert v_i - \overline{v}\Vert^2\right)$	Min
Kwons [38]	$V_K = \dfrac{\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}{}^2\Vert x_j - v_i\Vert^2 + \frac{1}{c}\sum_{i=1}^{c}\Vert v_i - \overline{v}\Vert^2}{\min_{i\neq k}\Vert v_i - v_k\Vert^2}$	Min
Wu and Yang [39]	$V_{PCAES} = \sum_{i=1}^{c}\sum_{j=1}^{n}\dfrac{u_{ij}{}^2}{u_{mj}} - \sum_{i=1}^{c} exp\left(\dfrac{-\min_{k\neq i}\Vert v_i - v_k\Vert^2}{\beta_T}\right)$	Max
Pakhira and Bandyopadhyay [40]	$V_{PBMF} = \left(\dfrac{\sum_{j=1}^{n} u_{ij}\Vert x_j - v_i\Vert \times \max_{i,k=1,\cdots,k}\Vert v_i - v_k\Vert}{k\times\sum_{i=1}^{k}\sum_{j=1}^{n} u_{ij}\Vert x_j - v_i\Vert}\right)^2$	Max
Wu and Li [41]	$V_{WL} = \dfrac{\sum_{i=1}^{c}\left(\dfrac{\sum_{j=1}^{n} u_{ij}{}^2\Vert x_j - v_i\Vert^2}{\sum_{j=1}^{n} u_{ij}}\right)}{\min_{i\neq k}\Vert v_i - v_k\Vert^2 + median\Vert v_i - v_k\Vert^2}$	Min
Haouas [42]	$V_{HF} = \dfrac{\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}\Vert x_j - v_i\Vert^2 + \frac{1}{c(c-1)}\sum_{i=1}^{c}\sum_{k=1,k\neq i}^{c}\Vert v_i - v_k\Vert^2}{\frac{1}{2c}\left(\min_{i\neq k}\Vert v_i - v_k\Vert^2 + median_{i\neq k}\Vert v_i - v_k\Vert^2\right)}$	Min
Zhu [43]	$V_Z = \dfrac{\sum_{j=1}^{n}\dfrac{1-\min_i u_{ij}}{\sum_{i=1}^{c}\Vert x_j - v_i\Vert^2}}{\sum_{k=1}^{c}\sum_{i=1;i\neq k}^{c}\Vert v_i - \overline{v}\Vert / \frac{c(c-1)}{2}}$	Min
Wang [24]	$V_{HY} = \dfrac{\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^2\Vert x_j - v_i\Vert^2 + \frac{1}{c}\sum_{i=1}^{c-1}\sum_{k=i+1}^{c}\left[\frac{1}{n}\sum_{j=1}^{n}\min(u_{ij}, u_{kj})\right]}{\min_{1\le i\le c}\sum_{j=1}^{n} u_{ij}^2 + \min_{i\neq k}\Vert v_i - v_k\Vert^2}$	Min

Table 3
Clustering validity evaluation methods based on weighted combination.

Proposer	Definition	Weighted method	Composed validity index
Sheng [27]	$maxWSVF = \sum_{i=1}^{r} w_i f_i(x)$	Mean weighted	$1/V_{DB}$, V_{SIL}, V_D, V_{33}, V_{CH}, V_{PBM}
Dong [38]	$\max FWSVF = \sum_{i=1}^{r} w_i f_i(x)$	Mean weighted	$1/V_{XB}$, $1/V_{PE}$, V_{PC}, V_{PBMF}
Li [50]	$minWSCVI = \sum_{i=1}^{r} w_i CVI_i$	Mean weighted	$1/V_{PC}$, $1/V_{NPC}$, V_{PE}, V_{NPE}, V_{XB}, V_K, $1/V_{PBMF}$, V_{FS}, V_T, V_{SC}
Wu [51]	$\min DWSVF = \sum_{i=1}^{r} w_i f_i(x)$	Dynamic weighted	$1/V_{MPC}$, V_{XB}, V_{PBMF}
Wang [24]	$\min HWCVF = \sum_{i=1}^{r} w_{hybrid} F_i(x)$	Hybrid weighted	V_{MPC}, V_{XB}, V_K, V_P, V_{PBMF}, V_{WL}, V_Z, V_{HY}

function. For $\forall A \subseteq \Theta, m_1, m_2, \cdots, m_n$ fusion rule [56] is shown in Eq. (20):

$$(m_1 \oplus m_2 \oplus \cdots \oplus m_n)(A) = \frac{1}{1-k}\sum_{A_1\cap A_2\cap\ldots\cap A_n = A} m_1(A_1)\cdot m_2(A_2)\ldots m_n(A_n) \tag{20}$$

where $m_i, i = 1, \cdots, n$ are the BPA Function, k is the conflict coefficient defined as Eq. (21). The closer k is to 1, the more serious the conflict between the evidence sources. The closer k is to 0, the more consistent the evidence sources.

$$k = (m_1 \oplus m_2 \oplus \cdots \oplus m_n)(\varnothing) = \sum_{A_1\cap A_2\cap\cdots\cap A_n = \varnothing} m_1(A_1)\bullet m_2(A_2)\bullet\cdots\bullet m_n(A_n) \tag{21}$$

The DS theory can fusion human subjectivity with the objectivity of function. Its advantage is that the description of uncertain information uses “interval estimation” replace “point estimation”. The uncertain aspects, as well as the accurate reflection evidence collection, show great flexibility, so it has received extensive attention and application. However, there are some disadvantage the DS theory, such as evidential must remain independent, potentially explosive, lack of solid theoretical foundation for evidence synthesis theory, and the controversy of rationality and validity [57].

4.2. BPA matrix of FCVI

By ensemble the FCVI, r types of results can be generated by judging the optimal c in the same dataset by r FCVE. The scope of c is $[2, \sqrt{nmax}]$, where $nmax$ is the samples number in the data X. The BPA of the FCVI is defined as $M_{c\times r}$, as shown in Eq. (22):

$$M = \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_c \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1r} \\ m_{21} & m_{22} & \cdots & \vdots \\ \vdots & \vdots & \cdots & m_{2r} \\ m_{c1} & m_{c2} & \cdots & m_{cr} \end{bmatrix} \tag{22}$$

where m_{11} is the first FCVI belief degree corresponding to $c = 2, r = 1$, and m_{cr} represents the belief degree of the r-th FCVI corresponding to the $c = \sqrt{nmax}$. c should be an integer. When c takes all values, the cumulative confidence of FCVI should be 1, that is, $\sum_{i=2}^{c} m_{ij} = 1$, where $j = 1, 2, \cdots, r$.

Different FCVI have different changes when judging the optimal c, and there is also a gap in the dimension. Therefore, the dimension of the FCVE is normalized first, so that the resulting value range falls within [0,1] as shown in Eq. (23):

$$f_i = \frac{F_i - F_i^{min}(x)}{F_i^{max}(x) - F_i^{min}(x)} \tag{23}$$

where, F_i is the i-th FCVI, $F_i^{min}(x)$ is the minimum of the F_i, and $F_i^{max}(x)$ is the maximum of the F_i, f_i is the value of the F_i after standard normalization. However, inputting f_i into the confidence matrix directly cannot satisfy the condition of the $\sum_{i=1}^{n} m_{ij} = 1$, because the some FCVI selects the maximum as the optimal c, and some FCVI select the minimum value as the optimal c. Therefore, it is discussed in two cases, when the FCVI takes the maximum as the optimal c, the BPA function is $m_{ij} = f_i/sum(f_i)$. When the FCVI takes the minimum as the optimal c, the BPA function is $m_{ij} = \frac{1-f_i}{1-sum(f_i)}$. This can solve the problem of the maximum and minimum value above.

Multiplying the i-th row to rank and the j-th row in M can obtained $A_{m\times m}$ as shown in Eq. (24):

$$A = M_i^T \times M_j = \begin{bmatrix} m_{i1} \times m_{j1} & m_{i1} \times m_{j2} & \cdots & m_{i1} \times m_{jr} \\ m_{i2} \times m_{j1} & m_{i2} \times m_{j2} & \cdots & m_{i2} \times m_{jr} \\ \vdots & \vdots & \vdots & \vdots \\ m_{ir} \times m_{j1} & m_{ir} \times m_{j2} & \cdots & m_{ir} \times m_{jr} \end{bmatrix} \tag{24}$$

where, the main diagonal element is the i-th and j-th FCVI, and the cumulative confidence of the optimal c target recognition is $Z = m_{ip} \times m_{jq}(p = q)$. The sum of the residual diagonals is the conflict coefficient k, as shown in Eq. (25).

$$k = \sum_{p\neq q} m_{ip} \times m_{jq}p, q = (1, 2, 3, \cdots, m) \tag{25}$$

The design of decision module is described as follows. Set $A_i(i = 1, 2, \cdots, n)$ is the size of $, A_\omega$ is the value of FCVI. Then calculating the confidence $m_{ij}(\theta)$ of A_i under the evidence framework Θ. The decision module should satisfy this condition: $m(A_\omega) = \max(m(A_i))$, and the max-belief degree related to the optimal c. $m(A_\omega) - m(A_i) > \delta, (\delta > 0)$, δ is threshold, and the credibility gap between the optimal c and others c should lager than δ.

5. Ensemble method based on multi-fuzzy clustering algorithms and Multi-FCVI

5.1. Ensemble based on FCM clustering algorithm and Multi-FCVI

It can be learned in Section 4.1 that all events should be independent, and the multi-FCVI are also mutually exclusive. The results of v_i and u_{ij} calculated by FCM are input into multi-FCVI respectively to obtain the fuzzy clustering validity ensemble based on DS theory (DS-FCVE) [58], as shown in Fig. 2. The flow of the ensemble model is as follows.

The process of this ensemble model is described as follows:

Step 1: Use the FCM to obtain U, c, V and other parameters and then input them into F_i, and obtain the value of different types of the FCVI;

Step 2: Normalize the value calculated in Step 1 according to Eq. (23);

Step 3: Construct the BPA according to the normalized result;

Step 4: Input the value calculated by the BPA function into the information fusion model of the DS theory to compute the belief degree corresponding to different c;

Step 5: Use the decision-making module to calculate the optimal c.

Based on the FCM, with the ensemble of its various multi-FCVI, DS-FCVE improves its adaptability to different datasets and can better judge the optimal c.

5.2. Ensemble based on multi-fuzzy clustering algorithms and Multi-FCVI

By ensemble single fuzzy clustering algorithm with multi-FCVI, such as DS-FCVE, the ability to adapt to the dataset is still poor due to the different ways of updating U_0 and V_0 of the initial clustering algorithm. In this section, multi-fuzzy clustering algorithms are introduced into DS-FFCVM, and an ensemble model based on multi-fuzzy clustering algorithms and multi-FCVI was proposed (DSMFCE). The stability and accuracy are improved by changing the structure of DS-FFCVM. DSMFCE is makeup with five parts. The first part, multi-fuzzy clustering algorithms and their corresponding updated membership degrees and clustering centers are used. The second part is calculating the value of multi-FCVI. The third part is to construct the validity index BPA function by regarding multi-FCVI as a hypothese. The fourth part is to fuse the information of multi-FCVI through the DS theory. The fifth part is to design the decision module then output the optimal c. Therefore, this section integrates the five parts of DSMFCE by using series and parallel ensemble methods, and finally proposes three clustering ensemble models, DSMFCE-I, DSMFCE-I and DSMFCE-III.

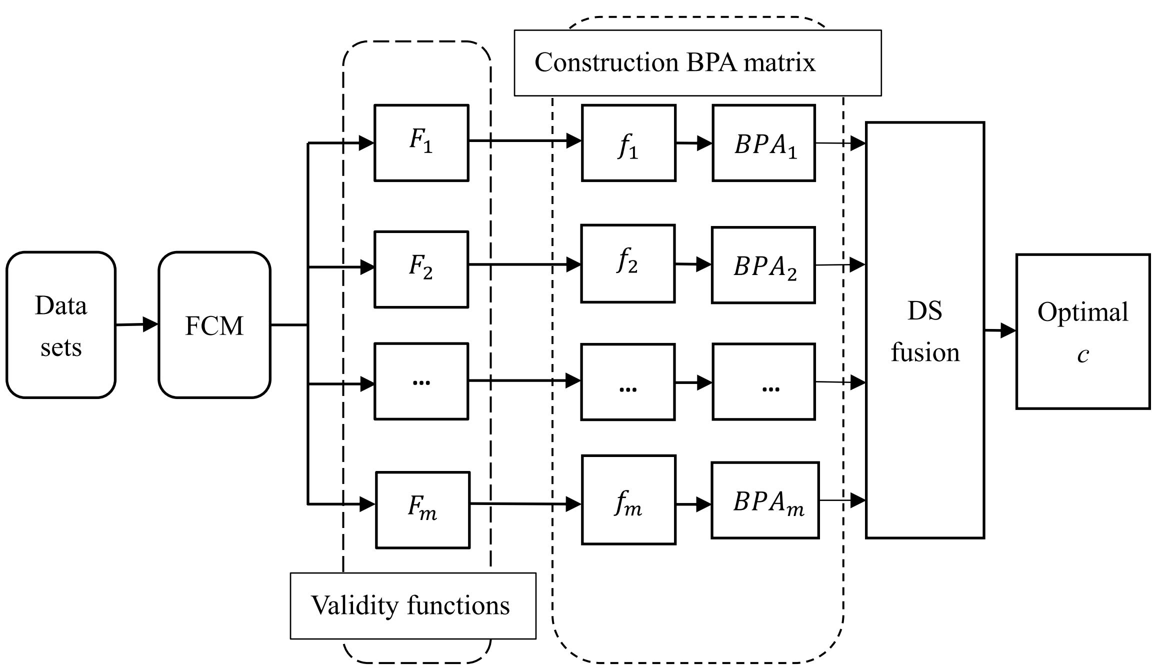


Fig. 2. DS-FCVE based on FCM and multi-FCVI.

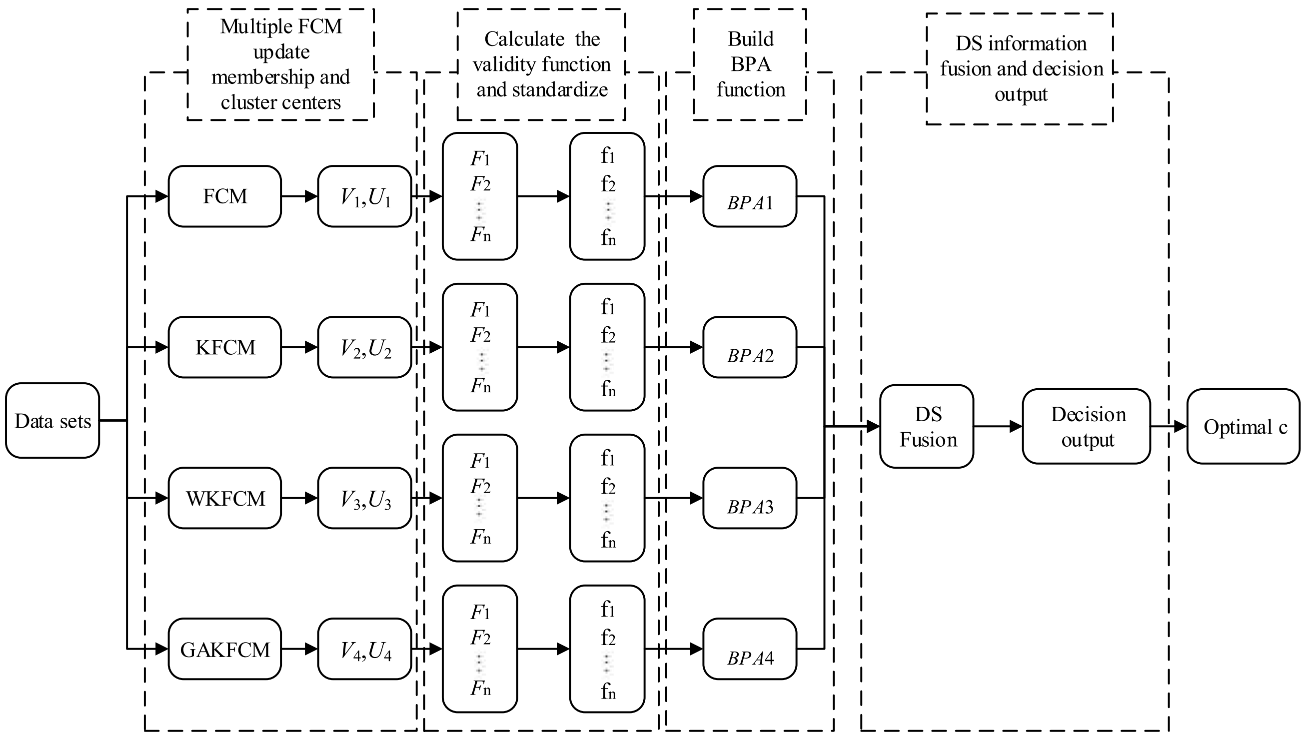


Fig. 3. Structure diagram of DSMFCE-I.

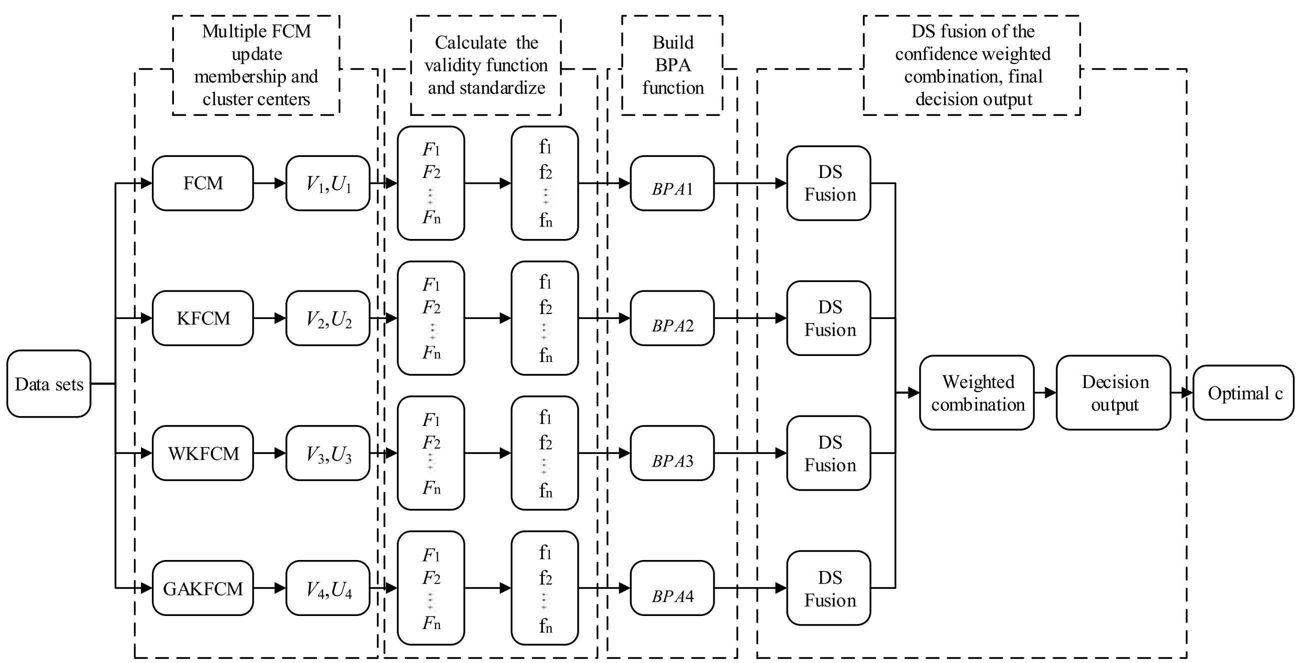


Fig. 4. Structure diagram of DSMFCE-II.

5.2.1. *Parallel clustering ensemble model (DSMFCE-I and DSMFCE-II)*

Fig. 3 shows the parallel ensemble model DSMFCE-I. First updates the U_{ij} and V_i by the different clustering algorithms (FCM, KFCM, WKFCM, GAKFCM). Then all of U_{ij} and V_i are inputted separately inputted to the corresponding FCVI to calculate F_i and the value of F_i is normalized to get f_i. Third, according f_i to construct the BPA. The values of all BPA functions are input into DS theory to obtained the belief degree values of different *c*. Finally, decision module judges the optimal *c* corresponding to the maximum confidence. From the structure diagram, it can be found that the multi-FCM, FCVI and BPA function are independent and parallel before D-S fusion, and they do not interfere with each other. Finally, the overall results are inputted into D-S evidence theory for realizing the clustering ensemble. More details about DSMFCE-I are shown in the Algorithm 1 pseudocode.

Algorithm 1: DSMFCE-I's Pseudocode

Input: FCM's parameters c, ε, and KFCM's parameters τ, and WKFCM's parameters $w_{ik} = 1/l$, and GAKFCM's p_{c0}, p_{m0}, T, δ, P;
Output: optimal c;
Initialize U^0, V^0;
For *iter* = 1: the maximum number of iterations **do**
  For $i = 2{:}c$ **do**
  For $j = 1 : n$ **do**
    Update the FCM's U_{ij}, V_i according to Eq. (2) and Eq. (3);
    Update the KFCM's U_{ij}, V_i according to Eq. (8) and Eq. (9);
    Update the WKFCM's U_{ij}, V_i according to Eq. (12) and Eq. (13);
    Update the GAKFCM's the U_{ij}, V_i according to Eq. (14) and Genetic Algorithm;
    Calculate the $e = \|U^t - U^{t-1}\|$;
    While $e \geq \varepsilon$, **do**
      $U^t = U^{t-1}$
    Else
Return all the values of U_{ij}, V_i;
End For;
End For;
End For;
Return all the values of U_{ij}, V_i;
Calculate the different types of the F_i according to all the values of U_{ij}, V_i;
Calculate the f_i according to Eq. (23);
Calculate the BPA according to the Eq. (24);
Calculate the belief degree corresponding to different c according to the BPA;
Calculate the optimal c according the decision-making module;
Return optimal c;

Fig. 4 shows the parallel ensemble model of DSMFCE-II. DSMFCE-II is similar to DSMFCE-I in structure, which is also based on different values of c, first updated the membership matrix U_i and cluster center matrix V_i by different clustering algorithms input to the corresponding FCVI F_i, then the F_i is normalized to get f_i. According to f_i frames the BPA function. The values of all BPA functions corresponding to different algorithms are inputted into D-S evidence theory. What needs to be focused on is that the values of BPA are input into D-S fusion respectively. Finally, the confidence values of cluster number c corresponding to different algorithms are weighted and combined, and the max-confidence corresponding to c is the optimal through decision module. More details about DSMFCE- II are shown in the Algorithm 2 pseudocode.

It can be found from the structure chart that the cluster validity functions corresponding to different algorithms, BPA function and D-S fusion are independent and parallel before re-weighting ensemble, and they do not interfere with each other. Finally, the c corresponding to the maxi-confidence is calculated by weighted combination of individual results, and the value of c is the optimal.

Algorithm 2: DSMFCE- II's Pseudocode

Input: FCM's parameters c, ε, and KFCM's parameters τ, and WKFCM's parameters $w_{ik} = 1/l$, and GAKFCM's p_{c0}, p_{m0}, T, δ, P;
Output: optimal c;
Initialize U, V;
  For *iter* = 1: the maximum number of iterations **do**
  For $i = 2{:}c$ **do**
  For $j = 1 : n$ **do**
    Update the FCM's U_{ij}, V_i according to Eq. (2) and Eq. (3);
    Update the KFCM's U_{ij}, V_i according to Eq. (8) and Eq. (9);
    Update the WKFCM's U_{ij}, V_i according to Eq. (12) and Eq. (13);
    Update the GAKFCM's the U_{ij}, V_i according to Eq. (14) and Genetic Algorithm;
    Calculate the $e = \|U^t - U^{t-1}\|$;
    While $e \geq \varepsilon$, **do**
    $U^t = U^{t-1}$
  Else
    Return all the values of U_{ij}, V_i;
End For;
End For;
End For;

(continued on next column)

(continued)

Algorithm 2: DSMFCE- II's Pseudocode

Return all the values of U_{ij}, V_i;
Calculate the different types of the F_i according to **all** the values of U_{ij}, V_i;
Calculate the f_i according to Eq. (23);
Calculate the BPA according to the Eq. (24);
Calculate the belief degree corresponding to different c according to the BPA;
Calculate the weighted combination of confidence values of different clustering number c;
Calculate the optimal c according the DS decision-making module;
Return optimal c

5.2.2. Series clustering ensemble model (DSMFCE-III)

Fig. 5 shows the series ensemble model of DSMFCE-III. DSMFCE-III is different from DSMFCE-II and DSMFCE-I that is based on different values of c, updated the membership matrix U_i and cluster center matrix V_i by different clustering algorithms, according to the consensus function (this paper selects weighted average fusion) fusion to get U_{new} and V_{new}, and then U_{new} and V_{new} is input to the FCVI F_i, then the value of F_i is normalized to get f_i. Then according to f_i, the BPA was calculated. The values of all BPA functions are inputted into D-S fusion, and the confidence values of different c are obtained. Finally, the decision module judges that the c corresponding to the maxi-confidence is the optimal. As shown in Fig. 5, it displays that different algorithms are combined with consensus function, and then the membership matrix and cluster center matrix are fused. The ensemble of FCVI and the BPA function are all in serial structuret until the final decision output. Different algorithms restrict each other, and finally the maxi-confidence of the corresponding c is the optimal clusters. More details about DSMFCE- III are shown in the Algorithm 3 pseudocode.

Algorithm 3: DSMFCE- III's Pseudocode

Input: FCM's parameters c, ε, and KFCM's parameters τ, and WKFCM's parameters $w_{ik} = 1/l$, and GAKFCM's p_{c0}, p_{m0}, T, δ, P;
Output: optimal c;
Initialize U, V;
  For *iter* = 1: the maximum number of iterations **do**
  For $i = 2{:}c$ **do**
  For $j = 1 : n$ **do**
    Update the FCM's U_{ij}, V_i according to Eq. (2) and Eq. (3);
    Update the KFCM's U_{ij}, V_i according to Eq. (8) and Eq. (9);
    Update the WKFCM's U_{ij}, V_i according to Eq. (12) and Eq. (13);
    Update the GAKFCM's the U_{ij}, V_i according to Eq. (14) and Genetic Algorithm;
    Calculate the $e = \|U_{new} - U_{last}\|$;
    Calculate the $e = \|U^t - U^{t-1}\|$;
    While $e \geq \varepsilon$, **do**
      $U^t = U^{t-1}$
  Else
  Return all the values of U_{ij}, V_i;
End For;
End For;
End For;
Return all the values of U_{ij}, V_i;
Calculate the U_{new}, V_{new} according to the **all** the values of U_{ij}, V_i;
Calculate the different types of the FCVI F_i according to U_{new}, V_{new}
Calculate the f_i according to Eq. (23);
Calculate the BPA according to Eq. (24);
Calculate the belief degree corresponding to different c according to the BPA;
Calculate the optimal c according the DS decision-making module;
Return optimal c;

6. Simulation experiments and results analysis

6.1. Hyper parameters and experimental datasets

For FCM clustering algorithm, the clustering number c is $2 \leq c \leq 14$, and the fuzzy weighted $m = 2$. Decision threshold $\varepsilon = 0.1$, the parameter of Gaussian radial basis function $\tau = 150$, the initial crossover probability of genetic algorithm $p_{c0} = 0.6$, initial mutation probability $p_{m0} = 0.01$, genetic evolution algebra $T = 100$, genetic stop threshold $\delta = 0.001$. The entire algorithm flow iterations times $t = 100$. The

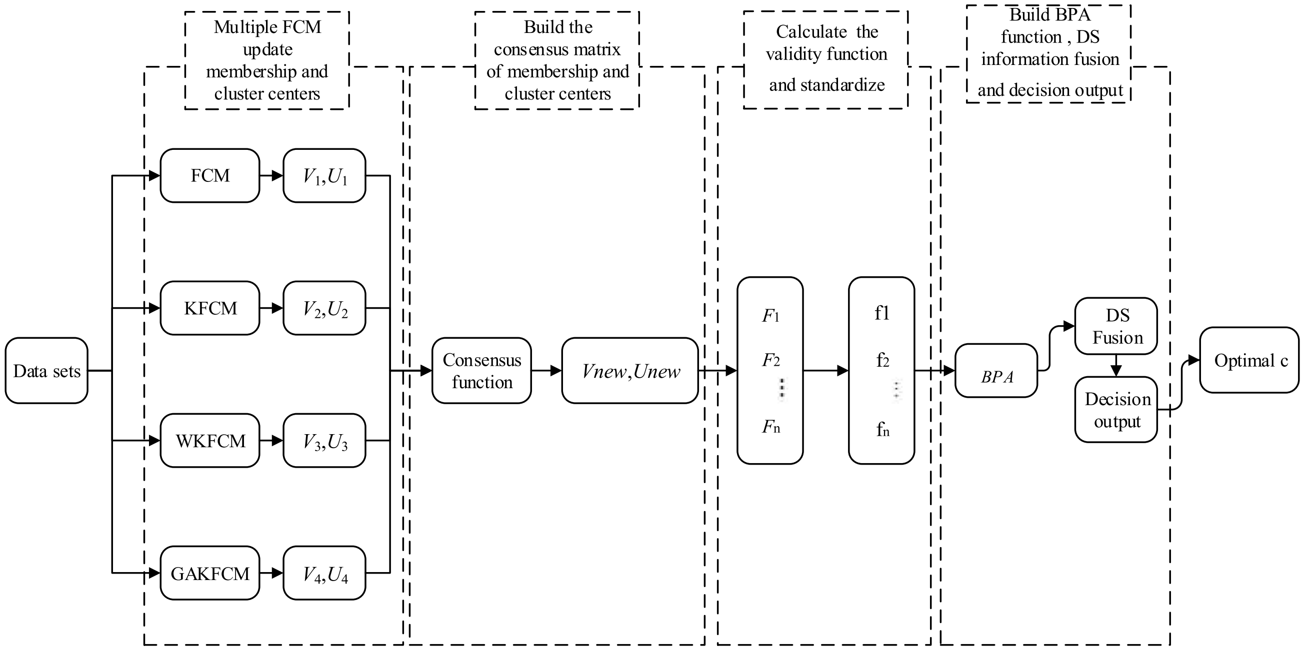


Fig. 5. Structure diagram of DSMFCE-III.

experiment selected 6 manual datasets and 12 UCI datasets. The manual data given in Fig. 6(a) is 2-dim and 3-class dataset, Fig. 6(b) is 2-dim and 3-class with overlapping samples dataset, Fig. 6(c) is 2-dim and 3-class with noise samples dataset, Fig. 6(d) is 2-dim and 5-class dataset, Fig. 6(e) is 3-dim and 3-class dataset, and Fig. 6(f) is 3-dim and 6-class dataset. All samples of manual datasets shown in Fig. 5 obey normal distribution.

The detailed information of UCI datasets is given in Table 4. The column 1 is the datasets name. The column 2 is the datasets samples number. The column 3 represents datasets attributes, the column 4 represents the datasets classes, the column 5 represents the datasets creator, and the last column represents source of the dataset.

More supplement detailed information about the dataset is as follows. The iris dataset collects 3 types of irises, namely Setosa, Versicolor and Virginica and the iris dataset includes 4 attributes, namely the length and width of calyx, the length and the width of petal. The seeds dataset stores different varieties of wheat seeds such as grain size, habitat and more. The wine dataset records the chemical composition analysis of three varieties of wine in the same region of an Italian region. The heart dataset is relative to heart disease contains a total of 4 databases on heart disease diagnosis. The cryotherapy dataset provides information related to patient characteristics, that is, whether the treatment result indicates the cancer level, that is, the malignancy is 0 and the benign is 1. The zoo dataset is a statistical dataset about zoo species information, including attributes of 7 animals such as whether they have tails, legs, fins and other features. The wpbc dataset is following the addition of information to the Wisconsin Breast Cancer Database (wdbc). The balance dataset is designed to simulate psychological experiment results, and the attributes are left weight and distance, right weight and distance. The bupa dataset is on whether adult men will cause liver disease due to alcohol intake. The ibeacon dataset includes 5 attributes of blood tests. The breast dataset is a set of breast cancer datasets. The Led7 dataset indicates that the led display contains 7 light-emitting diodes, and therefore has 7 properties. All the above data sets can be downloaded from UCI website.

The data sets selected in the experiment are different from the categories, features, and the number of samples, so it can well reflect the adaptability of each validity evaluation method in the face of different types of data sets. V_{MPC}, V_{XB}, V_K, V_P, V_{WL}, V_Z and V_{HY} are selected to integrate the model, mainly because the clustering analysis algorithms in the paper are fuzzy clustering algorithms, and the fuzzy membership degree and fuzzy index in the validity function of fuzzy clustering can be better integrated with cluster analysis.

6.2. Comparative experiments among DSMFCE-I, DSMFCE-II, and DSMFCE-III

The simulation experiments on DSMFCE-I, DSMFCE-II, and DSMFCE-III were carried out by using manual datasets, and results are reflected in Figs. 7–9. The DSMFCE-I simulation is reflected in Fig. 7 (a)-(b). DSMFCE-I cannot identify the optimal *c* for data_3_6. The DSMFCE-II simulation is shown in Fig. 8 (a)-(b). DSMFCE-II can identify the optimal *c* for all manual datasets. The DSMFCE-III simulation is shown in Fig. 9 (a)-(b). DSMFCE-III could not identify the optimal *c* for data_3_3, data_2_5, data_3_6. The judgment results of DSMFCE-I, DSMFCE-II, and DSMFCE-III on manual datasets are listed in Table 5.

The simulation experiments on DSMFCE-I, DSMFCE-II, and DSMFCE-III were carried out by using UCI datasets, which results are reflected in Figs. 10–12. The simulation of DSMFCE-I is given in Fig. 10 (a)-(c). DSMFCE-I is unable to find the optimal *c* for wine, heart, cryotherapy, zoo, wpbc, breast, bupa and ibeacon datasets, and even fails to make judgments for bupa and ibeacon data sets, which is because of the evidence conflicts. The simulation of DSMFCE-II is shown in Fig. 11 (a)-(c). DSMFCE-II unable to find the optimal *c* for ibeacon datasets. DSMFCE-III is shown in Fig. 12 (a)-(c). DSMFCE-III can find the optimal *c* for iris, seeds, heart and balance datasets. The judgment results of DSMFCE-I, DSMFCE-II, and DSMFCE-III under UCI datasets are listed in Table 6.

According to Table 5 and 6, it can be found that DSMFCE-II has the highest accuracy. Compared with DSMFCE-III, DSMFCE-II has a large amount of computation, but it can greatly retain the information after the combination of multi-FCM clustering algorithms and multi-FCVI. DSMFCE-III will lose some information when it fuses membership matrix and clustering center matrix. Moreover, compared with DSMFCE-I, DSMFCE-II not only reduces the calculation steps, but also avoids the problem of proposition conflict. In the process of simulation experiments, it can be found that DSMFCE-I not only has a low accuracy, but even produces failure results, which directly fails to provide the final evaluation results of clustering validity. For this reason, only DSMFCE-II

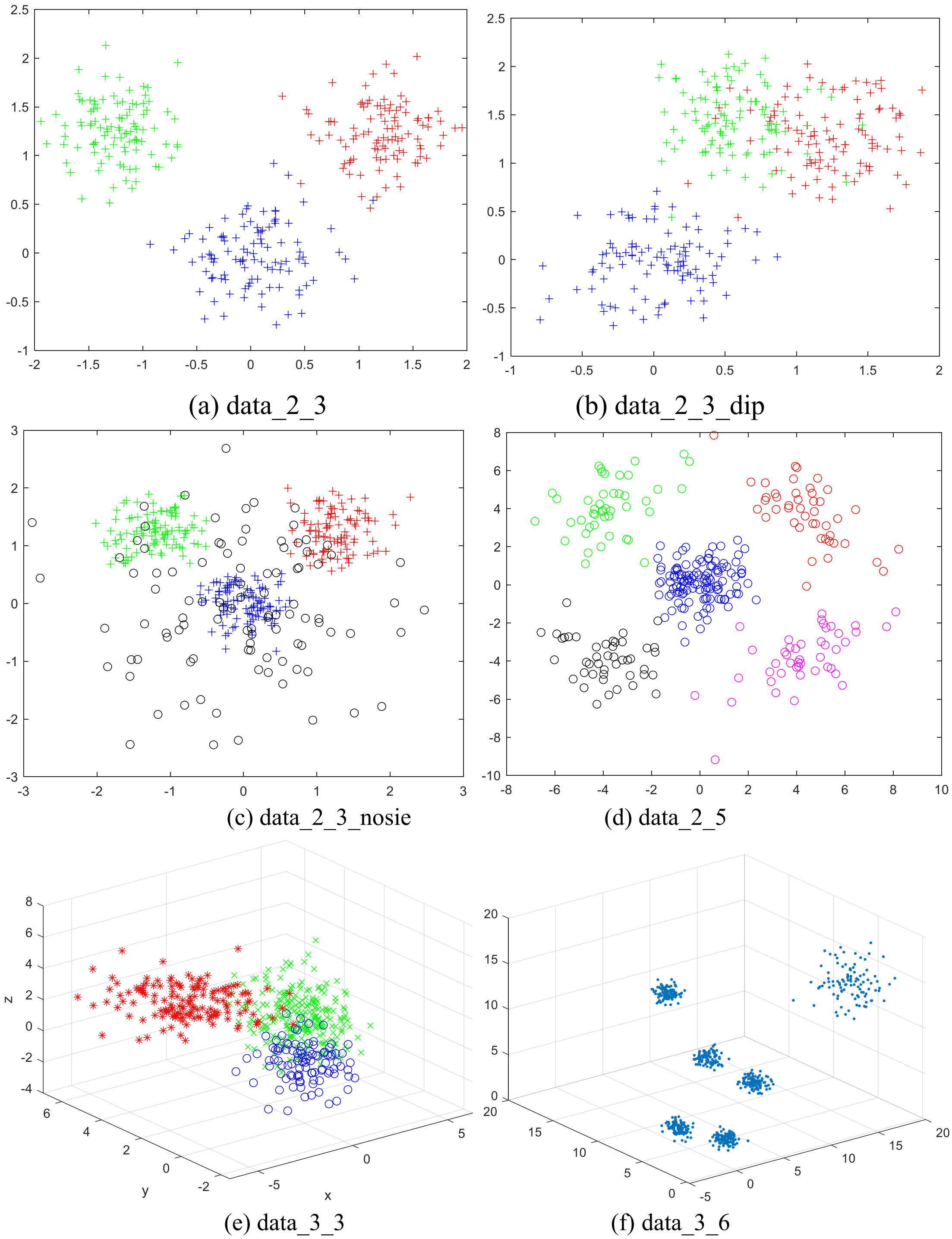


Fig. 6. Manual datasets.

was selected for comparison in subsequent comparative experiments.

6.3. Comparison between DS-FCVE and DSMFCE-II

DS-FCVE is multi-FCVI ensemble based on single Fuzzy clustering algorithm, DSMFCE-II is the ensemble result of multi-fuzzy clustering algorithms and multi-FCVI and then weighted combination to obtain the optimal number of clustering. For this reason, this section compares DSMFCE-II with the DS-FCVE based on GAKFCM, KFCM, WKFCM, and FCM separately. To observe whether the performance of DSMFCE-II improves in the selection of the final cluster number. Validity functions ensemble based on FCM (DS-FVF), validity functions ensemble based on KFCM (DS-KFVF), validity functions ensemble based on WKFCM (DS-WKFVF), validity functions ensemble based on GAKFCM (DS-GAKFVF) are selected to compared with DSMFCE-II and the simulation are given in Fig. 13.

Comparison and simulation of FCVI ensemble based on single fuzzy clustering algorithm and DSMFCE-II is reflected in Fig. 13 (a)-(l). In Fig. 13 (a), only DS-WKFVF cannot get the optimal c for iris dataset. In Fig. 13 (b), the optimal c for seeds dataset can be found by above-mentioned five methods. Fig. 13 (c) shows that DSMFCE-II and DS-GAKFVF can find the correct optimal c for wine dataset and the other

Table 4
UCI datasets.

Data Name	Samples	Features	Classes	Creator	Source
cryotherapy	90	7	2	Fahime K.	https://archive.ics.uci.edu/ml/datasets/cryotherapy
zoo	101	17	7	Richard S. F.	https://archive.ics.uci.edu/dataset/111/zoo
iris	150	4	3	Fisher R. A.	https://archive.ics.uci.edu/ml/datasets/iris
wine	178	13	3	Forina M.	https://archive.ics.uci.edu/dataset/109/wine
wpbc	198	34	2	Olvi M.	https://archive.ics.uci.edu/ml/machine-learning-databases
seeds	210	7	3	MaA O. C.	https://archive.ics.uci.edu/ml/datasets/seeds
heart	270	13	2	CDC	https://archive.ics.uci.edu/ml/datasets/Statlog?%28Heart%29
breast	286	9	2	Matjaz Z.	https://archive.ics.uci.edu/ml/datasets/Breast?Cancer
bupa	345	6	2	BUPA Ltd.	https://archive.ics.uci.edu/ml/datasets/Liver?Disorders
led7	500	7	11	Breiman L.	https://archive.ics.uci.edu/ml/datasets/LED?Display?Domain
balance	625	4	3	Siegler R. S.	https://archive.ics.uci.edu/ml/support/balance?scale
ibeacon	6611	15	13	Apple lcn.	https://archive.ics.uci.edu/ml/machine-learning-databases/00435/

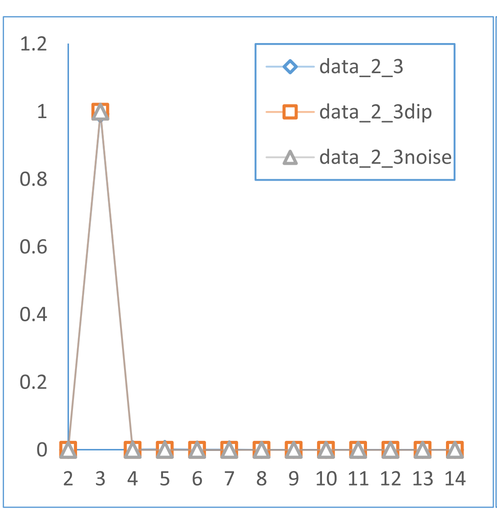
3 methods make mistakes. Fig. 13 (d) shows that DS-GAKFVF could not accurately determine the optimal c found in heart dataset. In Fig. 13 (e), DS-FVF and DS-KFVF cannot get the optimal c for cryotherapy dataset. Fig. 13 (f) shows that DSMFCE-II and DS-FVF can get the optimal c for zoo dataset. In Fig. 13 (g), that above-mentioned five methods can find the optimal c for wpbc dataset. Fig. 13 (h) shows that only DS-GAKFVF cannot get the optimal get the optimal c in balance dataset. Fig. 13 (i) shows that only DS-KFVF cannot get the optimal get the optimal c for breast dataset. Fig. 13 (j) shows that DSMFCE-II and DS-GAKFVF can calculate the optimal c for bupa dataset. In Fig. 13 (k), all five methods cannot calculate the optimal c for ibeacon dataset. Fig. 13 (l) shows that DS-KFVF and DS-GAKFVF wrongly judge the optimal c for led7 dataset. The results of DS-FVF, DS-KFVF, DS-WKFVF, DS-GAKFVF, and DSMFCE-II in calculating the optimal c of the UCI datasets are given in Table 7.

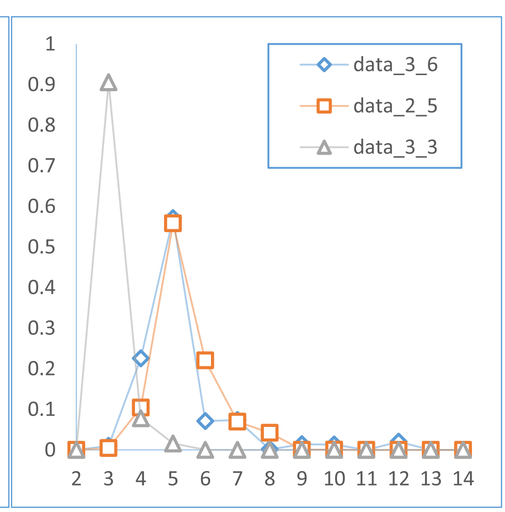
It can be found from Table 7, under the condition of different data sets, the effectiveness of different clustering algorithms in the same function ensemble results are different, that is because the way of different clustering algorithm update membership degree and the clustering center is different. Hence DSMFCE-II in combination multi-fuzzy clustering algorithm information when combines the structure characteristics of multi-FCVI. In this way, even if some algorithms fail, the correct c of the datasets can still be get under the weighted combination of the remaining algorithms. Therefore, DSMFCE-II is more stable than DS-FVF, DS-KFVF, DS-WKFVF, and DS-GAKFVF.

6.4. Comparison experiment between DSMFCE-II and FCVI

DSMFCE-II is the multi-FCVI information ensemble under the support of multi-fuzzy clustering algorithms. For this reason, DSMFCE-II is designed to be contrasted with the common FCVI respectively to observe whether DSMFCE-II has better adaptive capacity than single FCVI in judging the optimal c. In the experiment, V_{MPC}, V_{XB}, V_K, V_P, V_{WL}, V_Z and V_{HY} were selected and compared with DSMFCE-II under the condition of FCM clustering algorithm. The results are given in Table 8. As the FCVI corresponding to each dataset has amount number of values to judge the optimal c, most of which have no comparative significance and have a weak impact on the experimental results.

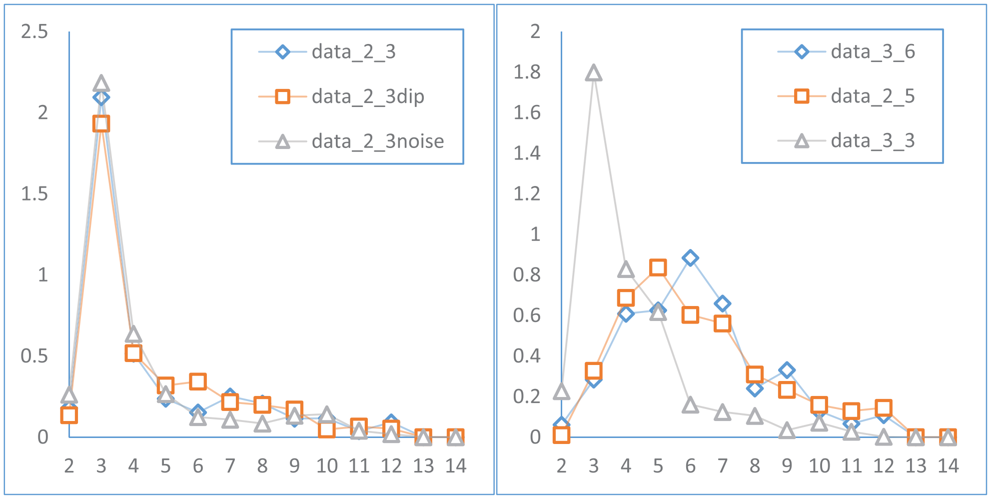
The simulation results on the UCI datasets are shown in Table 8. The optimal c can be found for iris and balance datasets by V_{MPC}, V_{HY} and DSMFCE-II. The optimal c can be got for seeds dataset by V_{MPC}, V_P, V_{HY} and DSMFCE-II. For cryotherapy and bupa datasets only V_{MPC} cannot get the optimal c. For the breast dataset, V_{MPC}, V_P and V_{WL} cannot get the optimal c. For led7 dataset, only DSMFCE-II can get the optimal c. For wine and zoo datasets, only DSMFCE-II can get the optimal c. Only V_{MPC} cannot get the optimal c for heart and wpbc datasets. The optimal c cannot be found for ibeacon. The optimal clustering number judgment





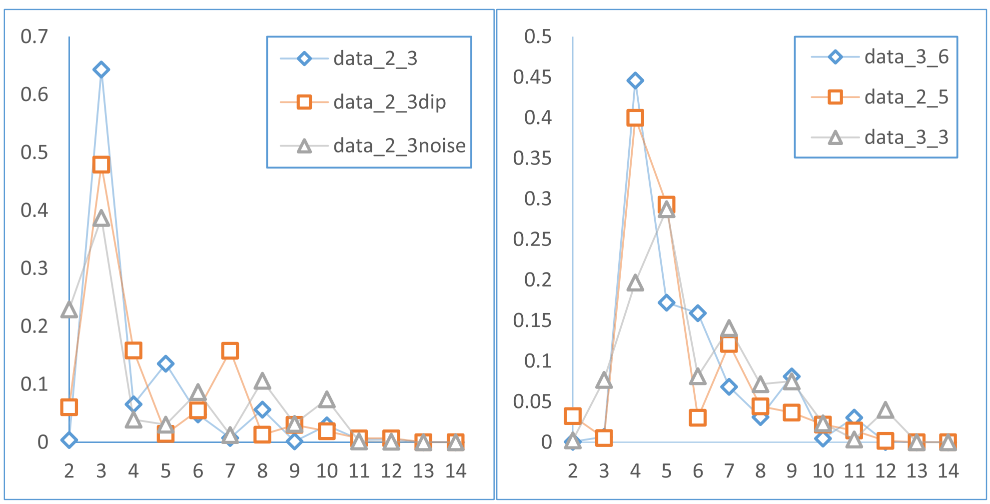
(a) **(b)**

Fig. 7. Simulation results of DSMFCE-I under artificial datasets.



(a) **(b)**

Fig. 8. Simulation results of DSMFCE-II under manual datasets.



(a) **(b)**

Fig. 9. Simulation results of DSMFCE-III under manual datasets.

Table 5
Judgment of optimal c by DSMFCE-I, II, and III under artificial datasets.

Data set	Optimal c	DSMFCE-I	DSMFCE-II	DSMFCE-III
data_2_3	3	**3**	**3**	**3**
data_2_5	5	**5**	**5**	4
data_3_3	3	**3**	**3**	5
data_3_6	6	5	**6**	4
data_2_3_dip	3	**3**	**3**	**3**
data_2_3_noise	3	**3**	**3**	**3**
Accuracy	–	83 %	100 %	50 %

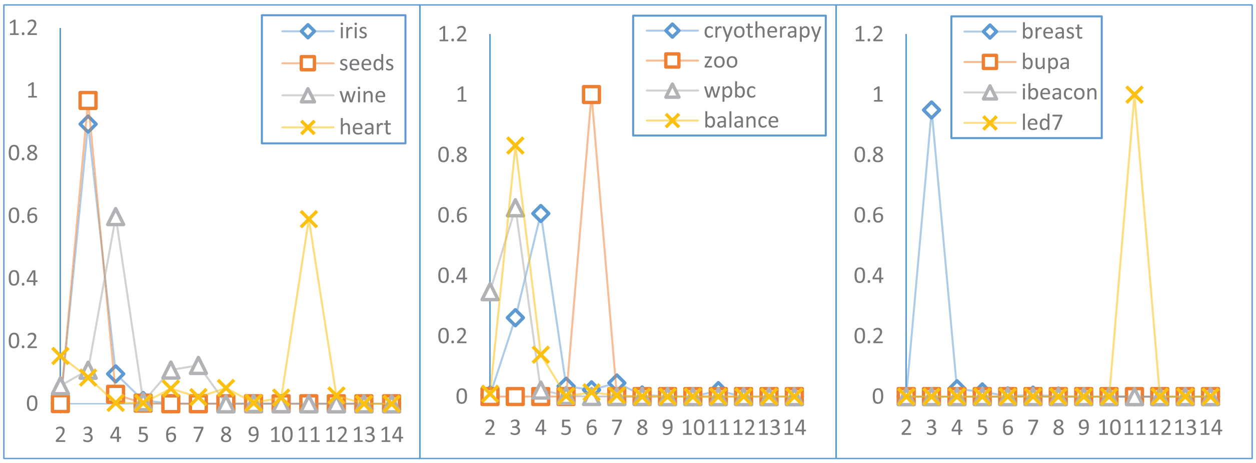
results of V_{MPC}, V_{XB}, V_K, V_P, V_{WL}, V_Z, V_{HY} and DSMFCE-II on 12 UCI datasets are shown in Table 10.

Therefore, DSMFCE-II has a good improvement in judging the optimal c compared with the single FCVI. The effectiveness judgment of the single FCVI under the standard FCM clustering algorithm will lead to errors in the final result due to the shortcomings of FCM clustering algorithm. DSMFCE-II performs information fusion on the validity function under the condition of multiple FCM clustering algorithms to make it more stable.

6.5. Comparative experiment between DSMFCE-II and WCVEM

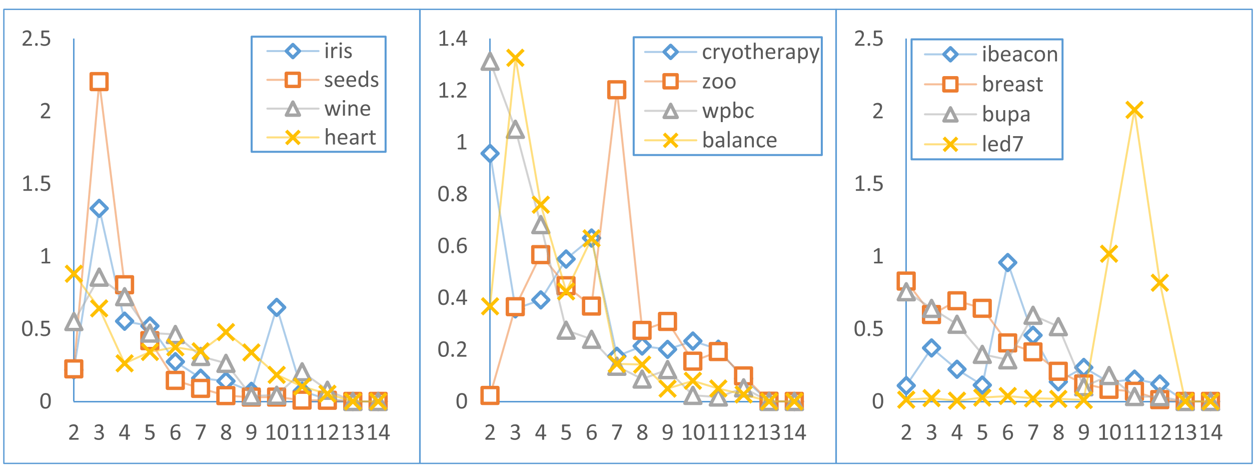
Weighted combination validity evaluation method (WCVEM) is the ensemble of FCVI research of the important direction, but the selection of weighted way tend to be more difficult. Its core is to put single FCVI weighted combined to form a novel ensemble method, then the effectiveness evaluation on the data set. In this paper, four WCVEM methods (HWCVF, DWSVF, WSCVI and FWSCF) were selected to compare with DSMFCE-II to proof whether the ensemble strategy of DSMFCE-II improved stability and accuracy compared with the weighted integration strategy. FCM clustering algorithm was selected to update the validity evaluation method of weighted combination. The simulation results are given in Table 9.

In Table 9, 12 groups of UCI datasets were chosen for compare verification. Only FWSCF could not get the optimal c for iris and seeds datasets. Simulation of wine dataset, only DSMFCE-II can get the optimal c. For heart dataset, DWSVF cannot get the optimal c, and WSCVI was not accurate in its judgment. The optimal c cannot be found for cryotherapy data set only with FWSCF. Only DSMFCE-II can get the optimal c for zoo dataset. DWSVF and WSCVI cannot get the optimal c



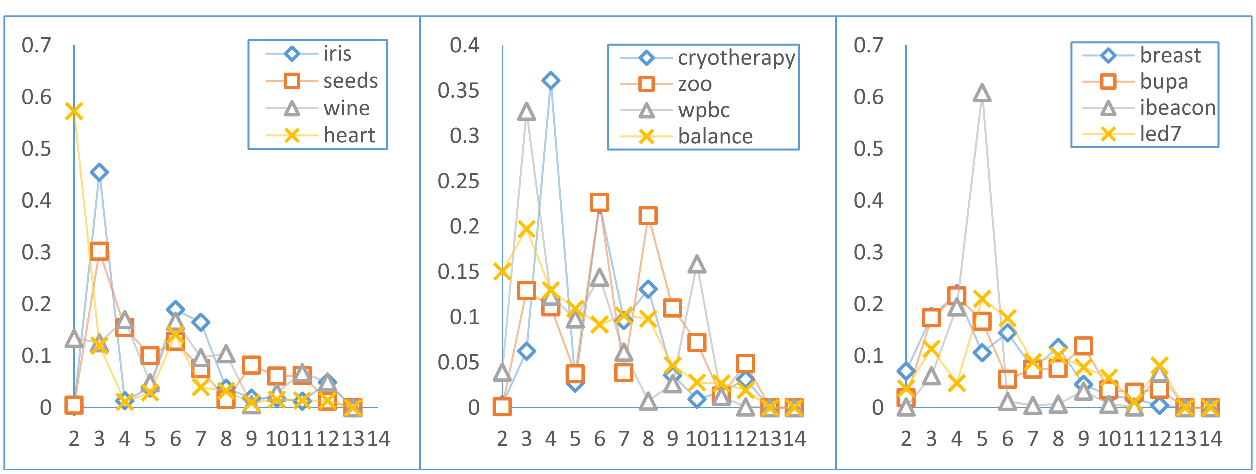
(a) **(b)** **(c)**

Fig. 10. Simulation on DSMFCE-I under UCI datasets.



(a) **(b)** **(c)**

Fig. 11. Simulation on DSMFCE-II under UCI datasets.



(a) **(b)** **(c)**

Fig. 12. Simulation on DSMFCE-III in under UCI datasets.

Table 6
Judgment of optimal *c* by DSMFCE-I, II, and III under UCI datasets.

Data set	optimal c	DSMFCE-I	DSMFCE-II	DSMFCE-III
iris	3	**3**	**3**	**3**
seeds	3	**3**	**3**	**3**
wine	3	4	**3**	4
heart	2	11	**2**	**2**
cryotherapy	2	4	**2**	4
zoo	7	6	**7**	6/8
wpbc	2	3	**2**	3
balance	3	**3**	**3**	**3**
bupa	2	NAN	**2**	4
breast	2	3	**2**	4
ibeacon	13	NAN	6	5
led7	11	**11**	**11**	5
Accuracy	–	33 %	92 %	25 %

for wpbc dataset. Just FWSCF cannot get the optimal c for balance dataset. DWSVF, WSCVI cannot get the optimal c for breast and bupa datasets. None of the schemes could find optimal c for ibeacon dataset. DSMFCE-II can find the optimal c for led7 dataset.

Table 9 listed the simulation value of 4 WCVEM and DSMFCE-II in judging the best c for 12 UCI datasets. It can be found from the experimental results that different weighted ways have a large influence to the final simulation results of the WCVEM.

6.6. Influence of fuzzy weight on DSMFCVE-II

Fuzzy weight m is often the most widely discussed in fuzzy clustering algorithms. When $m = 1$, fuzzy clustering degenerates into hard clustering. At present, most of the selection of m is between 2.0 and 2.5, and $m = 2.0$ is used in the above experimental processes. In order to verify the stability of the proposed model in the face of different m, m=1.5, m=2.0, m=2.5, m=3 and $m = 5$ were selected in this paper to observe the accuracy of DSMFCVE-II under different m. The selected datasets include all the manual and UCI datasets used in the appeal experiment. The accuracy of each validity evaluation method under different m is observed.

Table 10 and Table 11 show the accuracy of FCVI, combination modes and three ensemble modes (DSMFCE-I, DSMFCE-II, DSMFCE-III) under all datasets selected in the experiment, as well as the average and standard deviation of accuracy under different value conditions of m. The experimental results show that the average accuracy of DSMFCE-II is the highest, while the average accuracy of FCVI and WCVEM is almost the same that lower then DSMFCE-II. From the perspective of standard

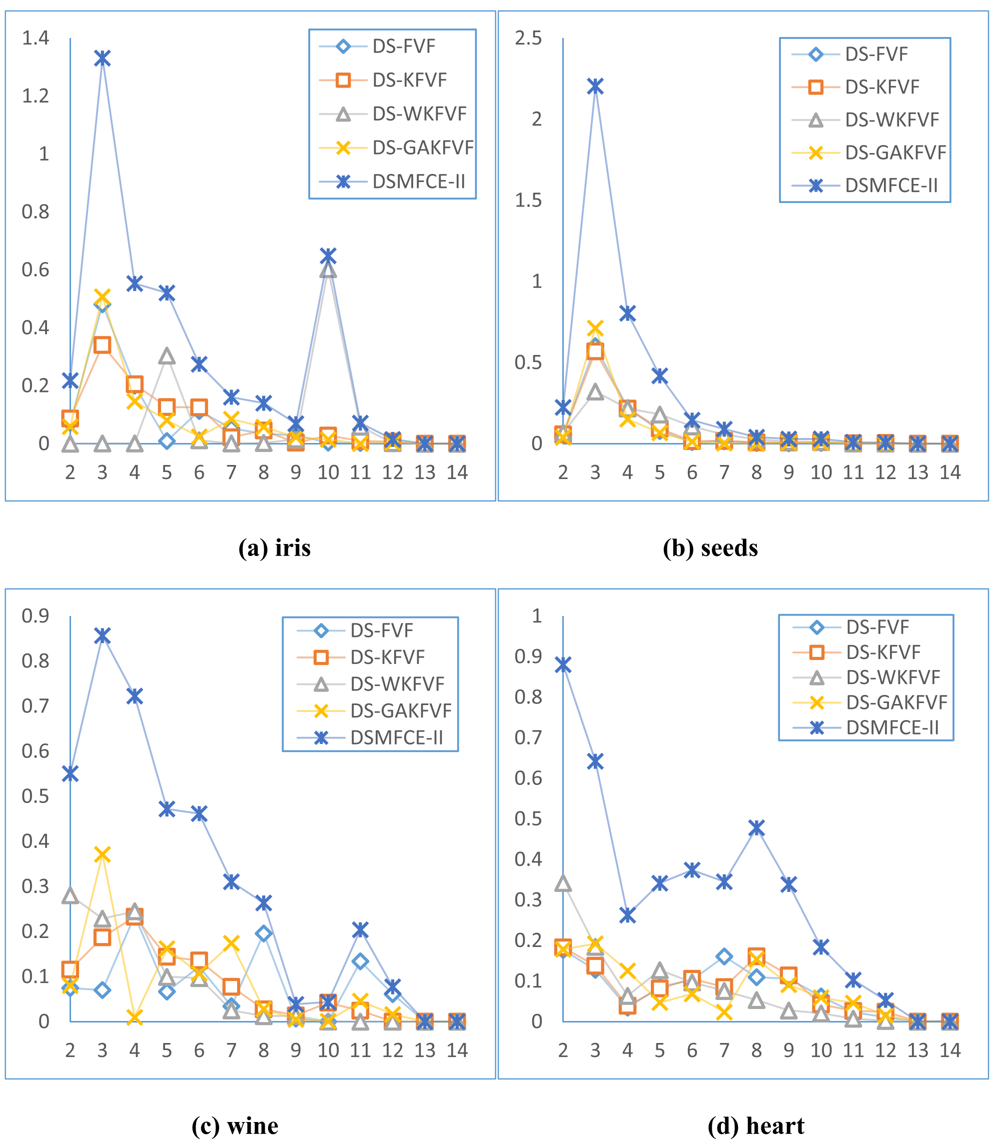


Fig. 13. Comparison between ensemble based on single FCM algorithm and DSMFCE-II.

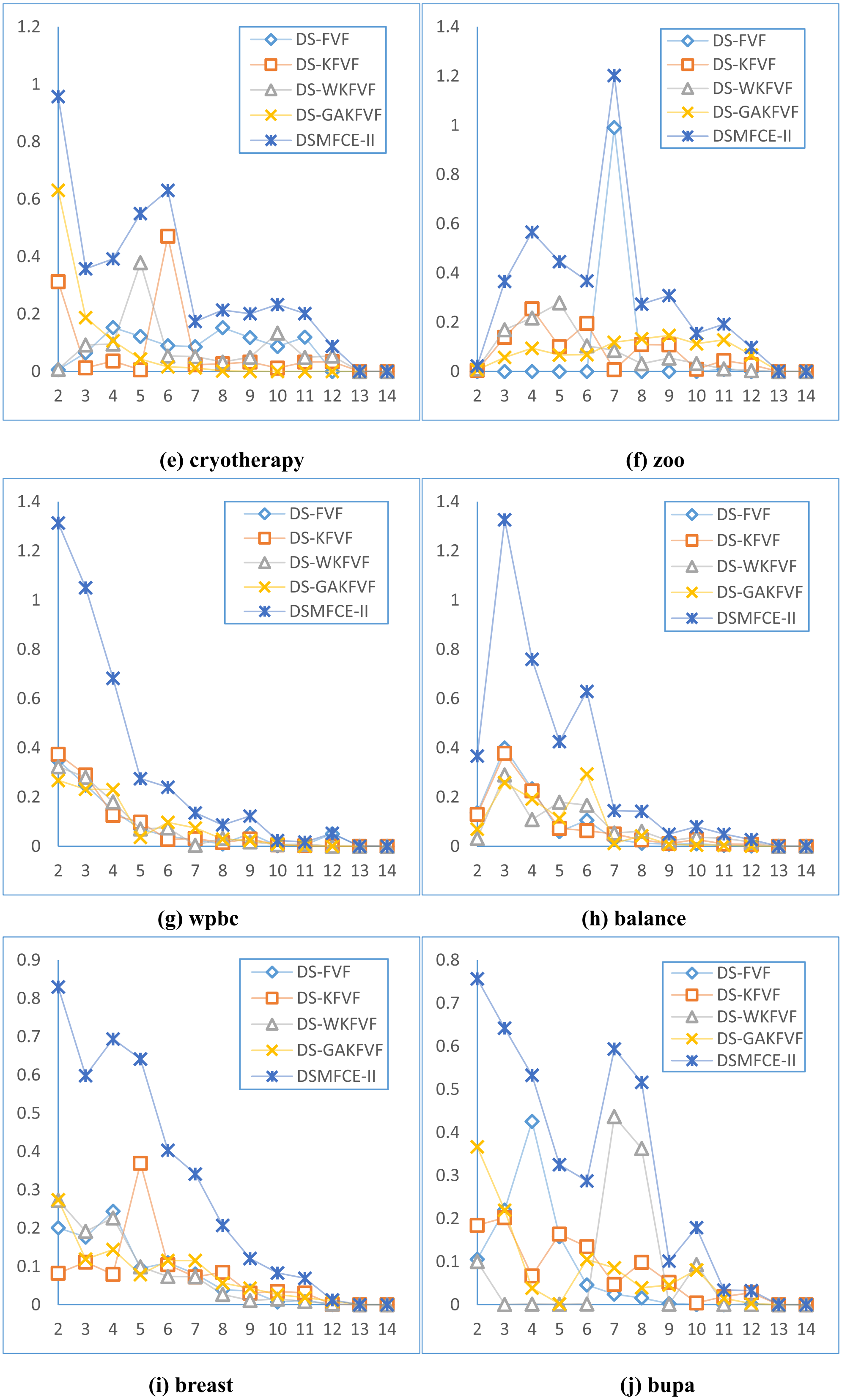
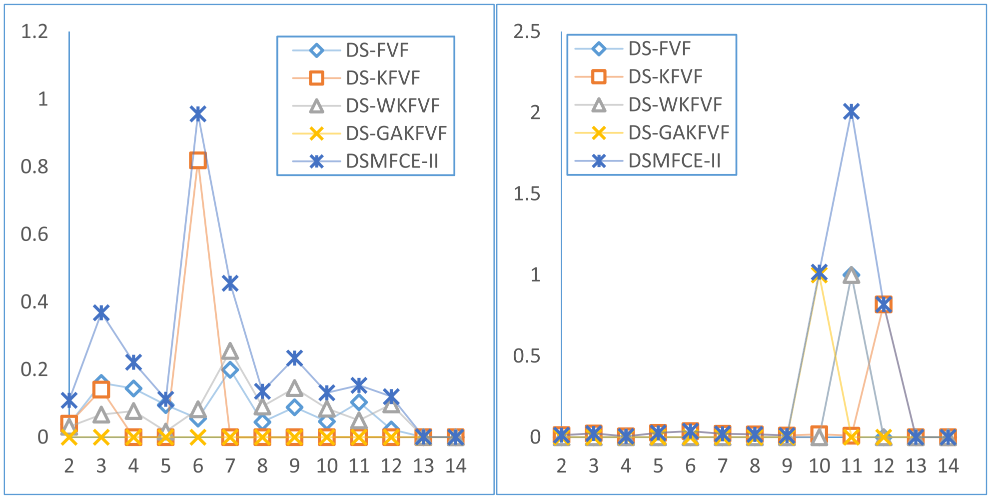


Fig. 13. (*continued*).

deviation, DSMFCE-II is not the lowest. The standard deviation of V_{WL}, V_Z, V_P, DWSVF, FWSCF are all lower than DSMFCE-II. Therefore, the m has a relatively large influence on DSMFCE-II, mainly because DSMFCE-II is composed of multiple FCVI. Each FCVI is affected by m, so the ambiguity is amplified after ensemble. Moreover, it can be found that the accuracy of DSMFCE-II shows a cliff drop when $m = 5.0$.



(k) ibeacon **(l) led7**

Fig. 13. (*continued*).

Table 7
Judging results of optimal c by various DS-FCVE and DSMFCE-II under UCI datasets.

Data set	Optimal c	DS-FVF	DS-KFVF	DS-WKFVF	DS-GAKFVF	DSMFCE-II
iris	3	**3**	**3**	10	**3**	**3**
seeds	3	**3**	**3**	**3**	**3**	**3**
wine	3	4	4	2	**3**	**3**
heart	2	**2**	**2**	**2**	**2/3**	**2**
cryotherapy	2	8	6	5	**2**	**2**
zoo	7	**7**	4	5	9	**7**
wpbc	2	**2**	**2**	**2**	**2**	**2**
balance	3	**3**	**3**	**3**	6	**3**
breast	2	**4**	5	**2**	**2**	**2**
bupa	2	4	2/3	7	**2**	**2**
ibeacon	13	7	6	7	NAN	6
led7	11	**11**	12	**11**	10	**11**
Accuracy	–	64 %	50 %	50 %	67 %	92 %

Figs. 14 and 15 show the boxplot of the accuracy of FCVI, WCVEM and DSMFCE-I, DSMFCE-II, DSMFCE-III under all datasets selected in the experiment under different m value conditions. The boxplot can visually represent the mean value and the degree of dispersion of the data. It can be seen that DSMFCVE-II has a high fitness value on most datasets, and the result distribution is concentrated with fewer outliers, which proves the robustness of the proposed algorithm. In summary, all the above results show that DSMFCE-II has excellent performance in most cases, significantly outperforming the original algorithm and other comparison algorithms.

7. Conclusion

As a means to improve the quality of clustering, fuzzy clustering algorithm integration has been widely studied. In this paper, the integration of fuzzy clustering algorithm is extended to the integration of validity index, and the DSMFCE algorithm is proposed. DSMFCE combines multi-fuzzy clustering algorithm with multi-FCVI algorithm by DS theory. In the module combination of clustering algorithm, validity

Table 9
Judgment of optimal c by four WCVEM and DSMFCE-II under UCI datasets.

Data set	optimal c	DWSVF	FWSCF	WSCVI	HWCVF	DSMFCE-II
iris	3	**3**	2	**3**	**3**	**3**
seeds	3	**3**	2	**3**	**3**	**3**
wine	3	4	2	7	2	**3**
heart	2	3	**2**	**2/3**	**2**	**2**
cryotherapy	2	5	**2**	**2**	**2**	**2**
zoo	7	4	2	4	4	**7**
wpbc	2	3	**2**	**3**	**2**	**2**
balance	3	**3**	2	**3**	**3**	**3**
breast	2	3	**2**	3	**2**	**2**
bupa	2	4	**2**	4	**2**	**2**
ibeacon	13	4	2	3	2	6
led7	11	3	2	3	5	**11**
Accuracy	–	25 %	42 %	50 %	67 %	92 %

Table 8
Judgment of optimal c by V_{MPC}, V_{XB}, V_K, V_P, V_{WL}, V_Z, V_{HY} and DSMFCE-II under UCI datasets.

Data set	optimal c	V_{MPC}	V_{XB}	V_K	V_P	V_{WL}	V_Z	V_{HY}	DSMFCE-II
iris	3	**3**	2	2	2	2	2	**3**	**3**
seeds	3	**3**	2	2	**3**	2	2	**3**	**3**
wine	3	7	2	2	2	7	2	2	**3**
heart	2	3	**2**	**2**	**2**	**2**	**2**	**2**	**2**
cryotherapy	2	5	**2**	**2**	**2**	**2/3**	**2**	**2**	**2**
zoo	7	4	5	5	4	2/3	5	4	**7**
wpbc	2	3	**2**	**2**	**2**	**2**	**2**	**2**	**2**
balance	3	**3**	2	2	2	2	2	2	**3**
breast	2	3	**2**	**2**	3	5	**2**	**2**	**2**
bupa	2	4	**2**	**2**	**2**	**2**	**2**	**2**	**2**
ibeacon	13	3	2	2	2	2	2	2	6
led7	11	5	2	2	2	2	2	2	**11**
Accuracy	–	33 %	42 %	42 %	42 %	25 %	42 %	58 %	92 %

Table 10
Average and standard deviation of ensemble model and WCVEM accuracy under different m.

-	DWSVF	FWSCF	WSCVI	HWCVF	DSMFCE-I	DSMFCE-II	DSMFCE-III
$m = 1.5$	0.19	0.35	0.47	0.62	0.73	0.83	0.32
$m = 2.0$	0.25	0.42	0.50	0.67	0.83	1.00	0.50
$m = 2.5$	0.27	0.42	0.49	0.60	0.70	0.92	0.43
$m = 3.0$	0.15	0.40	0.37	0.50	0.54	0.90	0.30
$m = 5.0$	0.067	0.27	0.15	0.3	0.33	0.70	0.11
AVG	0.19	0.37	0.40	0.54	0.63	0.87	0.33
STD	0.07	0.06	0.13	0.12	0.17	0.10	0.13

Table 11
Average and standard deviation of DSMFCE-II and FCVI accuracy under different..m

-	V_{MPC}	V_{XB}	V_K	V_P	V_{WL}	V_Z	V_{HY}	DSMFCE-II
$m = 1.5$	0.50	0.50	0.67	0.50	0.30	0.37	0.60	0.83
$m = 2.0$	0.33	0.42	0.42	0.42	0.25	0.42	0.58	1.00
$m = 2.5$	0.60	0.50	0.50	0.50	0.42	0.57	0.55	0.92
$m = 3.0$	0.46	0.43	0.44	0.42	0.25	0.44	0.47	0.90
$m = 5.0$	0.15	0.20	0.20	0.30	0.15	0.30	0.32	0.70
AVG	0.41	0.41	0.45	0.43	0.27	0.42	0.50	0.87
STD	0.16	0.11	0.15	0.07	0.09	0.09	0.10	0.10

function and DS theory, three kinds of clustering integration frameworks with serial and parallel structure are constructed. The existing fuzzy clustering effectiveness evaluation methods can be roughly divided into FCVI and WSCVE. FCVI has its own characteristics, for example, V_{HY} has good effect on overlapping data, V_{WL} and V_{PCAE} have good effect on noisy data, and V_{PBMF} is suitable for processing high-dimensional data. Although the method of WSCVE combines several validity functions, its weight method is difficult to determine. By constructing a probability distribution matrix, DSMFCE proposed in this paper regards FCVI as an event and determines the final optimal number of clusters through the probabilities of multiple events. Compared with FCVI, DSMFCE's generalization ability is improved, and no new validity function is required. DSMFCE does not need to set weights and can avoid the influence of human subjective factors.

In this paper, artificial data sets and UCI data sets are selected to compare DSMFCE with existing fuzzy clustering effectiveness evaluation methods. The experimental results show that the parallel DSMFCE-II has the highest accuracy, because DSMFCE-I inputs all BPA results into DS

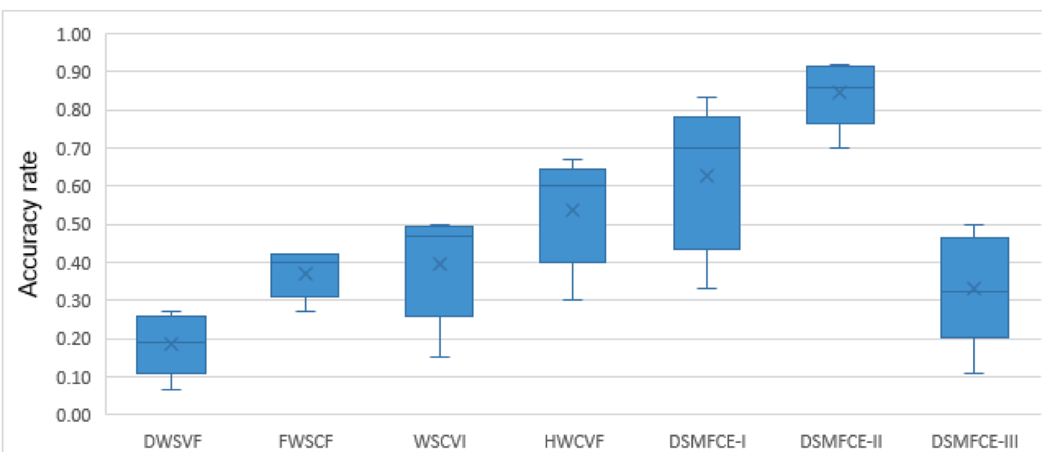


Fig. 14. DSMFCE-II and FCVI accuracy box plot Under different m.

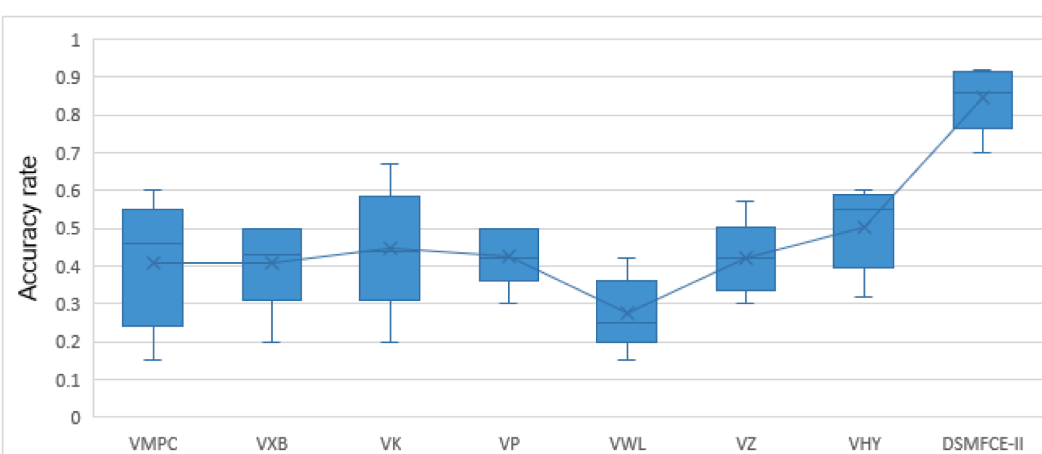


Fig. 15. Ensemble model and WSCVE accuracy box plot Under different m.

evidence theory integration at one time, while DSMFCE-II conducts integration using DS evidence theory separately, which is not prone to evidence conflict, but requires a large amount of calculation. While DSMFCE-III adopts weighted average method when updating cluster center and membership matrix, it loses a lot of information, but its computation is the least. When comparing with FCVI and WSCVE, the results show that DSMFCE is more effective than traditional fuzzy validity evaluation methods. However, the computation of DSMFCE is much larger than FCVI and WSCVE, and there may be problems of evidence conflict. In future work, we plan to propose a clustering algorithm ensemble to optimize DSMFCE-III and reduce the efficiency of DSMFCE. In addition, only fuzzy clustering validity evaluation method is used in this paper. If it can be combined with hard clustering validity evaluation method or external validity function, the generalization ability of DSMFCE may be further improved. Moreover, there are more hyperparameters in DSMFCE, which is what we need to further discuss in our future work.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the Basic Scientific Research Project of Institution of Higher Learning of Liaoning Province (Grant No. LJKZ0293), and the Postgraduate Education Reform Project of Liaoning Province (Grant No. LNYJG2022137).

References

[1] Xu, Haoxiang, Research on clustering algorithms in data mining. 2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE). IEEE, 2022.
[2] D'Orazio P. Mapping the emergence and diffusion of climate-related financial policies: Evidence from a cluster analysis on G20 countries. Int Econ 2022;169:135–47.
[3] Bai L, Liang J, Cao F. A multiple k-means clustering ensemble algorithm to find nonlinearly separable clusters. Inf Fusion 2020;61:36–47.
[4] Kenny G, McCann K, O'Brien C, Savinelli S, Tinago W, Yousif O, et al. Identification of distinct long COVID clinical phenotypes through cluster analysis of self-reported symptoms. Open Forum Infect Dis 2022;9(4).
[5] Jin J, et al. An agent-based traffic recommendation system: revisiting and revising urban traffic management strategies. IEEE Trans Syst, Man, Cybernetics: Syst 2022;52(11):7289–730.
[6] Monshizadeh M, Khatri V, Kantola R, Yan Z. A deep density based and self-determining clustering approach to label unknown traffic. J Netw Comput Appl 2022;207:103513.
[7] Hartigan JA, Wong MA. Algorithm AS 136: a K-means clustering algorithm. Appl Stat 1979;28(1):100.
[8] Zhang L, Zhang H, Cai G. The multiclass fault diagnosis of wind turbine bearing based on multisource signal fusion and deep learning generative model. IEEE Trans Instrum Meas 2022;71:1–12.
[9] Zadeh, Lotfi A. Fuzzy sets. In: Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh. 1996, 394-432.
[10] Bezdek JC, Ehrlich R, Full W. FCM: The fuzzy c-means clustering algorithm. Comput Geosci 1984;10(2-3):191–203.
[11] Chen H, Li C, Mafarja M, Heidari AA, Chen Y, Cai Z. Slime mould algorithm: a comprehensive review of recent variants and applications. Int J Syst Sci 2023;54(1):204–35.
[12] Strehl A, Ghosh J. Cluster ensembles–-a knowledge reuse framework for combining multiple partitions. J Mach Learn Res 2002;3:583–617.
[13] Hong Y, Kwong S, Wang H, Ren Q. Resampling-based selective clustering ensembles. Pattern Recogn Lett 2009;30(3):298–305.
[14] Berikov VB. A probabilistic model of fuzzy clustering ensemble. Pattern Recognit Image Anal 2018;28(1):1–10.
[15] Bagherinia A, Minaei-Bidgoli B, Hosseinzadeh M, Parvin H. Reliability-based fuzzy clustering ensemble. Fuzzy Set Syst 2021;413:1–28.
[16] Mojarad M, Nejatian S, Parvin H, Mohammadpoor M. A fuzzy clustering ensemble based on cluster clustering and iterative Fusion of base clusters. Appl Intell 2019;49(7):2567–81.
[17] Dong X, Yu Z, Cao W, Shi Y, Ma Q. A survey on ensemble learning. Front Computer Sci 2020;14(2):241–58.
[18] Z. Yao, et al. State evaluation of sight control box based on rough set fusion improved FCM clustering algorithm. In: 2020 Prognostics and Health Management Conference (PHM-Besançon). IEEE, 2020. p. 207–213.
[19] Wang H-Y, Wang J-S, Wang G. A survey of fuzzy clustering validity evaluation methods. Inf Sci 2022;618:270–97.
[20] Bezdek JC, Pal NR. Some new indexes of cluster validity. IEEE Trans. Syst., Man Cybernetics 1998;28(3):301–15.
[21] Silva L, Moura R, Canuto AMP, Santiago RHN, Bedregal B. An interval-based framework for fuzzy clustering applications. IEEE Trans Syst 2015;23(6):2174–87.
[22] Chen M-Y, Linkens DA. Rule-base self-generation and simplification for data-driven fuzzy models. Fuzzy Set Syst 2004;142(2):243–65.
[23] Xie, Lisa X, Gerardo B. A validity measure for fuzzy clustering. IEEE Trans Pattern Anal Mach Intell 1991;13(8):841–7.
[24] Wang H-Y, Wang J-S, Wang G. Combination evaluation method of fuzzy c-mean clustering validity based on hybrid weighted strategy. IEEE Access 2021;9:27239–61.
[25] Naderipour M, Fazel Zarandi MH, Bastani S. A fuzzy cluster-validity index based on the topology structure and node attribute in complex networks. Expert Syst Appl 2022;187:115913.
[26] Tang Y, Huang J, Pedrycz W, Li B, Ren F. A fuzzy clustering validity index induced by triple center relation. IEEE Trans. Cybernetics 2023;53(8):5024–36.
[27] Sheng W, Swift S, Zhang L, Liu X. A weighted sum validity function for clustering with a hybrid niching genetic algorithm. IEEE Trans. Syst., Man, Cybernetics, Part B (cybernetics) 2005;35(6):1156–67.
[28] Dong H, Hou W, Yin G. An evolutionary clustering algorithm based on adaptive fuzzy weighted sum validity function. In: 2010 Third International Joint Conference on Computational Science and Optimization; 2010. p. 357–61.
[29] Wang G, Wang J-S, Wang H-Y, Liu J-X. Ratio component-wise design method of fuzzy c-means clustering validity function. J Intell Fuzzy Syst 2022;43(4):4691–707.
[30] Liu JX, Wang JS, Wang G, et al. Exponent and logarithm component-wise construction method of FCM clustering validity function based on subjective and objective weighting. Int J Fuzzy Syst 2023;25:647–69.
[31] Girolami M. Mercer kernel-based clustering in feature space. IEEE Trans Neural Netw 2002;13(3):780–4.
[32] Shen H, Yang J, Wang S, Liu X. Attribute weighted mercer kernel based fuzzy clustering algorithm for general non-spherical datasets. Soft Comput 2006;10(11):1061–73.
[33] Ding Y, Fu X. Kernel-based fuzzy c-means clustering algorithm based on genetic algorithm. Neurocomputing 2016;188:233–8.
[34] Goldberg DE. Genetic algorithms in search, optimization, and machine learning. New York: Addison-Wesley; 2012.
[35] Jennings PC, Lysgaard S, Hummelshøj JS, Vegge T, Bligaard T. Genetic algorithms for computational materials discovery accelerated by machine learning. npj Comput Mater 2019;5(1).
[36] Simovici DA, Jaroszewicz S. An axiomatization of partition entropy. IEEE Trans Inf Theory 2002;48(7):2138–42.
[37] Fukuyama Y, Sugeno M. A new method of choosing the number of clusters for the fuzzy c-means method. 5th Fuzzy Systems Symposium, 1 989: 247–250.
[38] Kwon SH. Cluster validity index for fuzzy clustering. Electron Lett 1998;34(22):2176.
[39] Wu K-L, Yang M-S. A cluster validity index for fuzzy clustering. Pattern Recogn Lett 2005;26(9):1275–91.
[40] Pakhira MK, Bandyopadhyay S, Maulik U, Ujjwal M. Validity index for crisp and fuzzy clusters. Pattern Recogn 2004;37(3):487–501.
[41] Wu C-H, Ouyang C-S, Chen L-W, Lu L-W. A new fuzzy clustering validity index with a median factor for centroid-based clustering. IEEE Trans Fuzzy Syst 2015;23(3):701–18.
[42] F. Haouas, A new efficient fuzzy cluster validity index: Application to images clustering, In: 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). IEEE, 2017, pp. 1–6.
[43] Zhu LF, Wang JS, Wang HY. A novel clustering validity function of FCM clustering algorithm. IEEE Access 2019;7:152289–315.
[44] Vogel MA, Wong AKC. PFS clustering method. IEEE Trans Pattern Anal Mach Intell 1979;PAMI-1(3):237–45.
[45] Hadi AS. Finding groups in data: an introduction to chster analysis. Technometrics 1992;34(1):111–2.
[46] Jain AK, Dubes RC. Algorithms for clustering data. Prentice-Hall Inc; 1988.
[47] Bezdek JC, Pal NR. Some new indexes of cluster validity. IEEE Trans Syst, Man, and Cybernetics, Part B (cybernetics) 1998;28(3):301–15.
[48] Caliński T, Jerzy H. A dendrite method for cluster analysis. Commun Statistics-Theory and Methods 1974;3(1):1–27.
[49] Pakhira MK, Bandyopadhay S, Maulik U. Validity index for crisp and fuzzy clusters. Pattern Recogn 2004;37(3):487–501.
[50] Zhou K, Ding S, Fu C, Yang S. Comparison and weighted summation type of fuzzy cluster validity indices. Int. J. Computers Commun. Control 2014;9(3):370.
[51] Wu Z-F, Huang H-K. A dynamic weighted sum validity function for fuzzy clustering with an adaptive differential evolution algorithm. In: 2010 Third International Joint Conference on Computational Science and Optimization; 2010. p. 362–6.
[52] Dempster AP. Upper and lower probabilities induced by a multivalued mapping. Ann Math Stat 1967;38(2):325–39.
[53] Yager RR. On the dempster-shafer framework and new combination rules. Inf Sci 1987;41(2):93–137.
[54] Li P, Wei C. An emergency decision-making method based on DS evidence theory for probabilistic linguistic term sets. Int J Disaster Risk Reduct 2019;37:101178.

[55] Lu S, Li P, Li M. An improved multi-modal data decision fusion method based on DS evidence theory. In 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) Vol. 1, pp. 1684-1690.2020.June. IEEE.
[56] Yi-Bo L. Based on DS evidence theory of information fusion improved method. In 2010 International Conference on Computer Application and System Modeling (ICCASM 2010) (Vol. 1, pp. V1-416). 2010.october.
[57] Li F, Qian Y, Wang J, Liang J. Multigranulation information fusion: a Dempster-Shafer evidence theory-based clustering ensemble method. Inf Sci 2017;378: 389–409.
[58] Wang HY, Wang JS, Wang G. Clustering validity function fusion method of FCM Clustering algorithm based on Dempster-Shafer evidence theory. Int J Fuzzy Syst 2022;24:650–75.