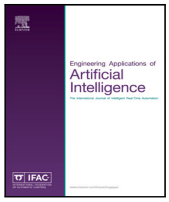
Engineering Applications of Artificial Intelligence 107 (2022) 104521

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Updating incomplete framework of target recognition database based on fuzzy gap statistic

Zichong Chen [a], Rui Cai [a,*]

[a] School of Business College, Southwest University, Chongqing 402460, China

ARTICLE INFO

Keywords:
Dempster–Shafer evidence theory
Fuzzy gap statistic
Fuzzy C-means
Open world
Target recognition

ABSTRACT

Generalized evidence theory (GET) is a generalization of Dempster–Shafer evidence theory. It copes with information in an open world, which makes up for the shortcoming that Dempster–Shafer evidence theory cannot handle information conflict effectively. However, GET also faces an unavoidable problem: how to determine the number of unknown targets in the incomplete frame of discernment (FOD). Fuzzy C-means (FCM) is a clustering algorithm that divides the original data set into different clusters and summarizes similar data into the same cluster. Therefore, determining the number of unknown targets in the open world can be transformed into finding the number of clusters. However, FCM has the disadvantage of subjectively controlling the number of clusters. In order to overcome this shortcoming, we use fuzzy gap statistic algorithm (FGS) to optimize it. FGS can effectively determine the optimal number of clusters in FCM. Therefore, this paper proposes a new method based on FGS to determine the number of unknown targets in the open world. In addition, to verify the method's accuracy, we conducted seven experiments based on the University of California Irvine (UCI) data sets, including Iris, glass, Haberman, Knowledge, Robot, seeds, and WDBC. Finally, the experimental results illustrate that the proposed method to determine the number of unknown targets in the incomplete FOD has high effectiveness.

1. Introduction

Information fusion can integrate data from multiple sources to obtain the optimum decision-making of the targets. Therefore, the technique has been utilized in a variety of fields, including but not limited to risk analysis (Wang et al., 2021), uncertainty problems (Zhou et al., 2020; Abellan and Bosse, 2020; Li et al., 2020), fault diagnosis (Wang et al., 2020; Huang et al., 2021) and multi-criteria decision making (Cao et al., 2019; Fu et al., 2020; Liao et al., 2020; Tang et al., 2020). However, the information quality is uneven and sometimes even harmful to the decision-making. Thus, numerous theories are presented to handle the aforementioned problem, such as intuitionistic fuzzy sets (Pan et al., 2021; Garg and Kumar, 2019; Deng and Deng, 2021), soft likelihood function (Fei et al., 2019), soft sets (Feng et al., 2016), Dempster–Shafer evidence theory (Song et al., 2020; Liu et al., 2020b), and others (Ye et al., 2021; Lai et al., 2020; Fujita and Ko, 2020; Gao et al., 2021).

As an uncertain reasoning method, Dempster–Shafer evidence theory (Dempster, 1967; Shafer, 1978) has the advantage to handle uncertain information (Dutta, 2018) and is widely used in information fusion (Li et al., 2021), network (Dutta and Palash, 2017; Song et al., 2016), expert decision system (Tian et al., 2020), risk analysis (Mo, 2021), decision-making (Fei et al., 2020) and other fields (Xiao, 2021). Compared with Bayesian probability theory, Dempster–Shafer evidence theory can directly express certainty and uncertainty with weaker conditions (Deng and Jiang, 2020). Therefore, a number of theories are extended and combined with Dempster–Shafer evidence theory, such as entropy theory (Song and Deng, 2021; Xue and Deng, 2021; Zhang and Deng, 2021), evidential reasoning (Liu et al., 2020a; Xu et al., 2018), classification (Liu et al., 2020a), and other hybrid models (Deng, 2020; Deng et al., 2021). However, the conventional Dempster–Shafer evidence theory can only apply to the closed world and cannot effectively manage the incomplete information (Haenni, 2002; Murphy, 2000). Therefore, many of researchers have proposed a lot of methods to settle a dispute caused by the incomplete FOD (Jousselme et al., 2001; Liu, 2006; Schubert, 2011). GET (Deng and Yong, 2015; Jiang and Zhan, 2017; Xiao, 2020) is one of the most famous methods among those proposed solutions. It allows an incomplete FOD and deals with conflicts in an open world effectively. Therefore, it also makes up for the shortcomings of the traditional Dempster–Shafer evidence theory.

To be noticed, however, how many unknown targets in the incomplete FOD is still a problem (Su et al., 2018). Many researchers have proposed a number of methods to discuss the incomplete FOD, and here are several representative methods. The first method is that Jiang proposed using evidence distance and evidence similarity to

* Corresponding author.
E-mail address: cairui686@swu.edu.cn (R. Cai).

https://doi.org/10.1016/j.engappai.2021.104521
Received 9 May 2021; Received in revised form 24 August 2021; Accepted 17 October 2021
Available online 29 October 2021

represent the relationship between the two GBPAs to estimate if the FOD is complete (Jiang et al., 2017, 2016). The second method is that Sun judged the integrity of the FOD based on the minimum spanning tree (Sun and Deng, 2019). The third method is that Liu proposed using the traditional Elbow method to determine the number of unknown targets in the incomplete FOD (Liu and Deng, 2021). The fourth method is gap statistic (GS) proposed by Robert et al. (Tibshirani and Hastie, 2001). The fifth method is machine learning (ML) (Abdalzaher et al., 2021). It is a new technology applied to intelligent systems, which helps us deal with the repetitive work and accumulate learning experience.

Deficiently, although the theories developed by these scholars promote the development of Dempster–Shafer evidence theory, there are still some problems. Jiang's methods (Jiang et al., 2017, 2016) only judge the integrity of the FOD, and both of them do not explain how many unknown objectives are there. Sun's method (Sun and Deng, 2019) is based on a minimum spanning tree, which is complex to implement. In addition, Liu uses Elbow method to estimate how many unknown targets in the open world (Liu and Deng, 2021), but the results are not noticeable. The GS method proposed by Robert et al. (Tibshirani and Hastie, 2001) cannot deal with the fuzzy information effectively because it is based on the K-means algorithm, which will produce errors to the results. Although ML has a strong learning performance, it is easy to produce errors because of hyper-parameters (Moustafa et al., 2021).

In order to overcome these issues, this paper develops a novel method based on FGS (Sentelle et al., 2007). In this way, we can find out the number of unknown targets in the open world and deal with fuzzy information. It should be noted that FGS is used to determine the optimal number of clusters in FCM, but FCM is easy to fall into a local minimum. Therefore, we use Monte Carlo sampling to reduce the error. In brief, if the FOD is incomplete judged by the method, the original data is tested by Monte Carlo sampling. Then, we use FCM to cluster the test data. Finally, FGS is applied to find out the actual number of targets in the open world. A large number of experiments show that our method is more comprehensive and improves accuracy.

As far as we know, FGS has not been applied to GET. The main contributions of this paper are as follows:

- This paper applies FGS to GET, which is an innovative work.
- FGS can overcome the subjectivity of FCM.
- The unknown targets in the open world can be easy found by the proposed method.
- Our experiments are supported by UCI data sets, including Iris, glass, Haberman, Knowledge, Robot, seeds, and WDBC. These data sets make the results more convincing.
- We compare the proposed method with some models in the literature. Experiments show that our method is more reasonable than these models.

The core content of this paper consists of 4 parts. In Section 2, this paper introduces some related theories. In Section 3, we propose our method based on FGS. In Section 4, the steps of proposed method are described in detail. Then, we conduct seven experiments. Moreover, several existing methods are compared with the proposed method. In Section 5, we summarize the article.

2. Preliminaries

In this section, Dempster–Shafer evidence theory and generalized evidence theory will be reviewed. And several methods quoted in this paper will be introduced. Table 1 records the abbreviations and notations used in this paper.

2.1. Dempster–Shafer evidence theory

Dempster–Shafer evidence theory (Dempster, 1967; Shafer, 1978) is a kind of inexact reasoning theory, which plays a significant role in information fusion. It was first developed by mathematician Dempster and perfected by his student Shafer, also famous as DS evidence theory.

Definition 2.1. Let Ω be the elaborate set of all possible values of the variable x, and the elements of Ω repel each other, then Ω is called the FOD. Suppose that there are N elements in Ω, then it can be defined as (Dempster, 1967; Shafer, 1978)

$$\Omega = \left\{H_1, H_2, \ldots, H_N\right\} \tag{1}$$

The power set of Ω can be noted as P with the number of 2^N elements. And each element corresponds to a proposition about the value of x.

Definition 2.2. For each subset A which belongs to Ω, let it correspond to a number $m(\cdot) \in [0, 1]$ satisfies (Shafer, 1978):

$$\begin{cases} \sum\limits_{A \subseteq \Omega} m(A) = 1 \\ m(\emptyset) = 0 \end{cases} \tag{2}$$

where $m(\emptyset) = 0$ means that probability is not allowed to be assigned to empty set in DS evidence theory.

2.2. Generalized evidence theory

GET (Deng and Yong, 2015; Jiang and Zhan, 2017) can effectively represent uncertain information such as stochasticity and fuzziness, which many researchers have promoted.

AS the development of DS evidence theory, GET inherits the basic definition of it. And the most important feature is no limit of $m(\emptyset) = 0$ in GET (Deng and Yong, 2015). In other words, if $m(\emptyset)$ is equal to 0, GET will degenerate into classical DS evidence theory.

It abandons the condition of $m(\emptyset) = 0$ in DS evidence theory. If the evidence is conflicting, GET can combine conflicting evidence through the reliability of evidence and will not produce counterintuitive results (Deng and Yong, 2015). Therefore, when information conflict occurs, GET can replace DS evidence theory to deal with conflict. Most definitions of GET are similar to DS evidence theory. Here are some new definitions.

Definition 2.3. Suppose there are two GBPAs, and their conflict coefficients are defined as (Deng and Yong, 2015)

$$K = \sum_{B \cap C = \emptyset} m_1(B)\, m_2(C) \tag{3}$$

The rules of combination between them are as follows (Deng and Yong, 2015):

$$m(A) = \frac{\left(1 - m(\emptyset)\right) \sum_{B \cap C = A} m_1(B)\, m_2(C)}{1 - K} \tag{4}$$

The allocation method for $m(\emptyset)$ is defined as follows (Deng and Yong, 2015):

$$m(\emptyset) = \begin{cases} m_1(\emptyset)\, m_2(\emptyset), K \neq 1 \\ 1, K = 1 \end{cases} \tag{5}$$

2.3. K-means and Fuzzy C-means

K-means (Wong, 1979; Jain, 2010; Aghdaie and Tafreshi, 2018) is a kind of standard clustering algorithm and an unsupervised learning method. Based on the internal distance of each sample point, the optimization function is established, and the adjustment rules of the operation are obtained by using the extremum method. We usually use Euclidean distance or square Euclidean distance. In this paper, the square Euclidean distance is used for the experiment.

Definition 2.4. Let d denotes the internal distance between two observations i and i', the formula is as follows (Wong, 1979):

$$d = \sum_j \left(x_{ij} - x_{i'j}\right)^2 \tag{6}$$

where $i = 1, 2, \ldots, s$, $j = 1, 2, \ldots, t$. s is the number of samples and t is the number of attributes. By using the above formula to iterate continuously, s samples can be divided into k clusters.

Table 1
The table of symbol description.

Category	Abbreviations or notations	Meaning
Dempster–Shafer evidence theory and Generalized evidence theory	Ω	The frame of discernment
	FOD	The frame of discernment
Generalized evidence theory	GET	Generalized evidence theory
	K	Generalized conflict coefficient
	GBPA	Generalized basic probability assignment
	$m(\cdot)$	GBPA
K-means and Fuzzy C-means	d	The square Euclidean distance
	FCM	Fuzzy C-means
	J_m	The objective function
	u	The fuzzy membership
	m	The fuzzy factor index
	k	The number of clusters
Elbow method	F	The cost function
	C	The cluster center
Gap statistic	GS	Gap statistic
	D	The sum of d in K-means
	W	The mean value of D
	B	The number of Monte Carlo sampling
	k	The number of clusters in K-means
Fuzzy gap statistic	FGS	Fuzzy gap statistic
	$J_{m,k}$	The objective function
	u	The fuzzy membership
	v	The cluster center of a class
	m	The fuzzy factor index
	B	The number of Monte Carlo sampling
	sd_k	The standard deviation
	s_k	The simulation error
	k	The number of clusters in FCM

FCM (Bezdek, 1974) is also an unsupervised algorithm. However, unlike K-means, FCM considers the membership of each sample and clustering center and proposes a membership matrix U for clustering samples. In other words, each data point is assigned a fuzzy membership. Therefore, FCM has a better ability to deal with uncertain information.

Definition 2.5. The core formula of the FCM algorithm is the objective function J, and the formula is as follows (Bezdek, 1974):

$$J_m = \sum_{i=1}^{s}\sum_{j=1}^{t}\left(u_{ij}\right)^m d \tag{7}$$

The definition of distance d is the same as K-means. The smaller the distance between two samples, the greater the similarity will be. Thus we can regard the two objects as a cluster. Therefore, we can separate s objects into k clusters in the light of the distance index. In addition, $u_{ij} \in [0,1]$ is the fuzzy membership, which satisfies the following relation (Bezdek, 1974):

$$\sum_{i=1}^{s}\sum_{j=1}^{t} u_{ij} = 1 \tag{8}$$

And m is the fuzzy factor index, which is usually defined as a smooth parameter, and different researchers have different standards for setting smoothing parameters. In this paper, we use Gaussian dispersion as a smoothing parameter. It is defined as (Galadi-Enriquez et al., 1998)

$$m = \left(\frac{4}{\theta+2}\right)^{\frac{1}{\theta+4}} \frac{\sigma}{N^{\frac{1}{\theta+4}}} \tag{9}$$

where θ is the number of attributes of a dataset, N is the number of samples, and σ is the standard deviation of each attribute. The complete FCM algorithm is as follows (Bezdek, 1974):

Algorithm1: FCM
Input: The database and the number of clusters. k
output: Samples of each cluster.
1.Standardize all data.
2.Establishing fuzzy similarity matrix.
3.Cluster the data until the result converges.
End

2.4. Elbow method

Elbow method (Thorndike, 1953) is used in cluster analysis to determine the optimal number of clusters in K-means.

Definition 2.6. The core idea of the cost function shown as follows (Thorndike, 1953):

$$F = \sum_{i=1}^{k}\sum_{x\in C_i}\left|x - C_i\right|^2 \tag{10}$$

where x is the element of each cluster, F is the cost function, and k is the number of clusters $|C_i|$. If the determined value of k is greater than the actual value of k, the change of cost function will not be so apparent for each time k increases by 1. If the determined value of k is less than the real value of k, the curve of cost function will decrease greatly for each time k increases by 1. In this way, the correct value of k will appear at the elbow point.

2.5. Gap statistic

GS method (Tibshirani and Hastie, 2001) is proposed by Robert et al. to find the best cluster number in K-means. The several core formulas are as follows.

Let D be the sum of the square Euclidean distances of every cluster. It is defined as (Tibshirani and Hastie, 2001)

$$D_r = \sum_{i,i'\in C_r} d_{ii'} \tag{11}$$

where C represents the cluster and r represents the serial number of the cluster.

Let W represents the mean value of the D_r and it is defined as (Tibshirani and Hastie, 2001)

$$W_k = \sum_{r=1}^{k} \frac{1}{2\left|C_r\right|} D_r \tag{12}$$

where k is the number of the clusters, $\left|C_r\right|$ is the cardinal of the clusters.

Definition 2.7. The core function in GS is the function $Gap(k)$. The formula is as follows (Tibshirani and Hastie, 2001):

$$Gap(k) = \frac{1}{B}\sum_{b}^{B} \log\left(W_{kb}^{*}\right) - \log\left(W_k\right) \tag{13}$$

where W_{kb}^{*} is a statistic generated by repeated sampling, and B is the number of sampling. The authors use W_k represents the W from the original data set, W_k^{*} represents the W from sampling (Tibshirani and Hastie, 2001).

2.6. Fuzzy gap statistic

FGS is an effective method to determine the optimal number of clusters in FCM, which is the development of GS method (Sentelle et al., 2007). FGS emphasizes the fuzziness of information, so the algorithm is more suitable for the real world. The objective function J in FGS is defined as Sentelle et al. (2007)

$$J_{m,k} = \sum_{r=1}^{k}\sum_{j=1}^{t}\left(u_{rj}\right)^{m}\left(x_j - v_r\right)^2 \tag{14}$$

where k is the number of clusters, t is the number of x, x is the samples belonging to the cluster r, v_r is the cluster center of the rth class.

Considering the uncertainty of information, FGS can effectively optimize FCM and overcome its subjectivity.

Definition 2.8. FGS is defined as (Sentelle et al., 2007)

$$\widetilde{Gap}(k) = \frac{1}{B}\sum_{b}^{B}\log\left(J_{m,k}^{*}\right) - \log\left(J_{m,k}\right) \tag{15}$$

where $J_{m,k}^{*}$ is a statistic. To be noticed, FGS is implemented based on FCM, but FCM is easy to fall into a local minimum. In order to solve this problem, we need to repeatedly change the initial value of cluster center u_r to optimize the results (usually 20 times), and Monte Carlo simulation can effectively complete this process. Therefore, we generate a new cluster center u_r to obtain the statistic $J_{m,k}^{*}$ by Monte Carlo sampling. In addition, B is the number of it, and k is the number of clusters.

Let sd_k be the standard deviation, and it is defined as (Sentelle et al., 2007)

$$sd_k = \left[\frac{1}{B}\sum_{b}^{B}\left(\log\left(J_{kb}^{*}\right) - \overline{\log\left(J_{m,k}\right)}\right)^2\right]^{\frac{1}{2}} \tag{16}$$

where B is the number of Monte Carlo sampling, and k is the number of clusters.

Let s_k be the simulation error of Monte Carlo sampling, which can be defined as (Sentelle et al., 2007)

$$s_k = sd_k\left(1 + \frac{1}{B}\right)^{\frac{1}{2}} \tag{17}$$

where B is the number of Monte Carlo sampling, and k is the number of clusters.

Finally, we can find out the best cluster number in FCM through the following inequality (Sentelle et al., 2007):

$$\widetilde{Gap}(k) \geq \widetilde{Gap}(k+1) - s_{k+1} \tag{18}$$

where k is the number of clusters. And the complete FGS algorithm is as follows (Sentelle et al., 2007):

Algorithm2: FGS
Input: The original database.
output: Optimal cluster number.
1.Monte Carlo sampling according to uniform distribution.
2.Apply FCM clustering to the generated samples.
3.Repeat 1 and 2 many times (usually 20 times) to reduce the error, and record the data every time.
4.Calculate FGS.
5.Determine the optimal number of clusters in FCM.
End

3. Proposed method

In GET, determining the target number of incomplete FOD is still a problem worthy of study. In recent years, a lot of scholars have developed many methods to solve this problem, among which Liu's method (Liu and Deng, 2021) and Robert et al.'s method (Tibshirani and Hastie, 2001) are more representative. However, in the real world, the information we obtain is usually fuzzy, and these two methods cannot deal with fuzzy information effectively. Therefore, after considering the fuzziness of information, we propose a new method to determine the number of unknown targets in the open world.

Our method can be separated into three steps. The first step is to generate GBPA based on the original data set and determine whether the FOD is complete according to the value of $m(\emptyset)$. If the FOD is complete, we do not need to continue working. Otherwise, we need to reconstruct the FOD. Secondly, we need to do Monte Carlo sampling on the original test data to generate the simulated data set. After that, with the help of FCM, the simulated data set can be divided into several clusters. Then we can determine the optimal number of clusters based on the FGS. Finally, we reconstruct the FOD and generate GBPA again to judge the integrity of the FOD. What should be noted is that if the result of FGS is greater than the real number of targets of the known class, the system will be considered to be inoperable.

The detailed steps are as follows:

Step 1: Generate GBPA and judge whether the FOD is complete. In this paper, the method to generate GBPA in Deng and Yong (2015) is adopted.

Step 1.1: Suppose there is a dataset S, whose elements are represented as x_{ij}. Then the triangular fuzzy numbers can be established. The rule of establishing triangular fuzzy numbers is as follows (Deng and Yong, 2015):

$$f(x) = \begin{cases} 0 & x < min \\ \dfrac{x - min}{mean - min} & min \leqslant x \leqslant mean \\ \dfrac{max - x}{max - mean} & mean \leqslant x \leqslant max \\ 0 & x > max \end{cases} \tag{19}$$

Step 1.2: According to the triangular fuzzy model generated in step 1.1, the strong constraint generation algorithm is used to obtain GBPA for each attribute. Here are the details of the generation rules (Deng and Yong, 2015):

(1) If the target is within the triangular fuzzy number of a single proposition, it will be regarded as a single proposition. And the intersection ordinate of sample and fuzzy number is the GBPA of the proposition.
(2) If the target is inside the fuzzy number model of multiple simple subset propositions, it will be regarded as multiple subset propositions. Under this condition, the GBPA of a single proposition is represented by the greater value of the ordinate of the intersection, and the GBPA of multiple subset is represented by the smaller value of the ordinate.

(3) If the target is in the multi proposition triangular fuzzy number model, it will be regarded as a multi subset proposition. In addition, the greater value of the intersection longitudinal coordinates is the value of GBPA of individual subset proposition. In addition, the smaller value of the longitudinal coordinates is the GBPA of each multi subset proposition.
(4) If the GBPA generated by the above rules is greater than 1, then $m(\emptyset) = 0$. Otherwise, the remaining value is assigned to $m(\emptyset)$.

Step 1.3: Calculate all the $m(\emptyset)$ from each attribute of the samples to get the mean value of $\overline{m(\emptyset)}$. Calculate the value of it with following method:

$$\overline{m(\emptyset)} = \frac{1}{t}\sum_{i=1}^{t} m(\emptyset) \tag{20}$$

Then $\overline{m(\emptyset)}$ will be used as the standard to judge whether the Ω is complete or not. A widely used critical value standard is that when the value of $\overline{m(\emptyset)}$ is greater than a critical value p, Ω is incomplete (Deng and Yong, 2015; Jiang et al., 2017).

Step 2: If the FOD is incomplete, this means current information will conflict. Therefore, it is necessary to reconstruct it. Thus, the following steps need to be strictly implemented.

Step 2.1: The original data is divided into n objects, and n is the number of target attributes. Then Monte Carlo sampling is carried out for each attribute according to the uniform distribution on the interval $[\min_{ij}, \max_{ij}]$, where $[\min_{ij}, \max_{ij}]$ is the minimum value of the jth attribute of the ith target, and $[\max_{ij}, \max_{ij}]$ represents the minimum is the maximum value of the jth attribute of the ith target.

Step 2.2: Use the test data obtained from Monte Carlo sampling in step 2.1 to cluster by FCM into k clusters, where k is the current number of clusters. Then we need to calculate the value of FGS for each cluster k. And then calculate the value of $\log\left(J^*_{m,k}\right)$ under each number of k, where $k = 1, 2, \ldots, K$.

Step 2.3: In order to reduce the error, we need to repeat steps 2.1 and 2.2, that is to do Monte Carlo sampling for the original data repeatedly (usually 20 times). Thus, we can obtain the mean value of $\log\left(J^*_{m,k}\right)$ for each cluster k based on each time of Monte Carlo sampling. Finally, the optimal number of clusters can be determined according to the $\log\left(J^*_{m,k}\right)$.

Step 2.4: Reconstruct the FOD. To be noticed, if the k is greater than the number of the real targets in the system, the known objects in the FOD are exact the whole targets.

Step 3: Judge whether the FOD is complete again. In step 2, the optimal number of clusters is obtained, which is the number of objects in the open world. Based on this condition, the FOD can be reconstructed, and we need to regenerate the GBPA to judge its integrity again. If the FOD has been complete, we can finish the program. Otherwise, the algorithm will be implemented until the FOD is complete.

Finally, the method and flowchart can be shown in Fig. 1.

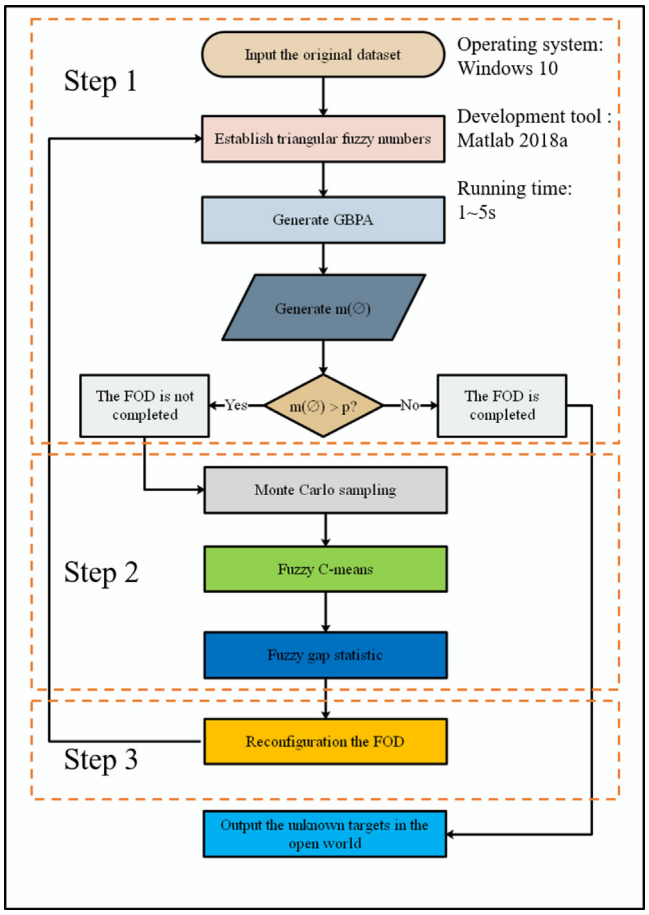


Fig. 1. Three steps to determine the number of unknown targets in the open world.

4. Experimental results

In this subsection, we will use a large number of experiments to show the effectiveness of the proposed method.

4.1. Experimental example

In this section, we will use an experiment based on Iris data set to Zhong and Fukushima (2007) demonstrate the calculation process. There are three categories in Iris dataset: setosa, versicolor, and virginia. Every category contains four attributes, and each attribute consists of 50 samples. They are sepal length (SL), sepal width (SW), petal length (PL), and petal width (PW) respectively. Since Iris has three categories, there are four possibilities for the FOD. They are

$$\Omega_1 = \{setosa, versicolor\} \quad \Omega_2 = \{setosa, virginica\}$$
$$\Omega_3 = \{versicolor, virginica\} \quad \Omega_4 = \{setosa, versicolor, virginica\}$$

In the process of information fusion, assume that the information we have obtained is $\Omega_1 = \{setosa, versicolor\}$, therefore, the known targets in the system are setosa and versicolor. Now the system is facing the problem of whether there is a third category, virginica, or other unknown categories. If the third category exists, information fusion will conflict. Therefore, it is essential to determine the number of unknown categories through experiments.

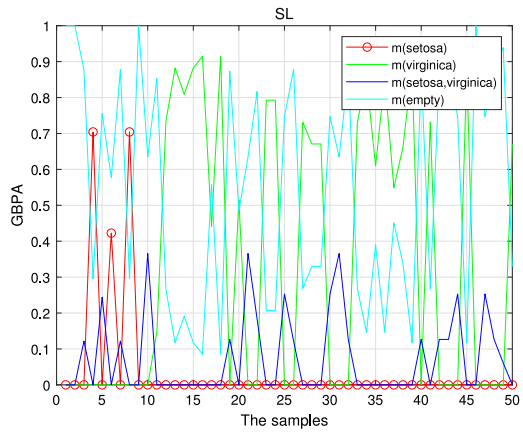
Experiment settings: During the experiment, the data set Iris is divided into test set and training set. Taking attribute SL as an example, the allocation rules are as follows:

(1) In setosa and virginia, 40 samples were randomly selected as the training set. Therefore, a total of 80 samples were used as the training set.
(2) Randomly select 30 samples from versicolor, plus the remaining 10 samples from setosa and virginia, a total of 50 samples as the test set. For the remaining attributes SW, PL, and PW, the same rules are used to process the data.

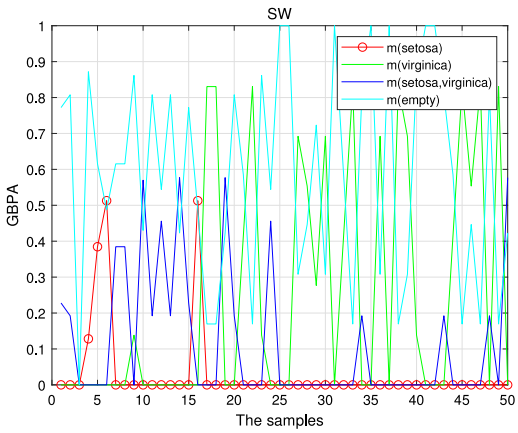
Algorithm demonstration: The experimental process based on Iris data set is as follows:

Step 1: Generate GBPA.

Step 1.1: Establish the triangular fuzzy numbers. First, we need to extract the minimum value, mean value and maximum value of each

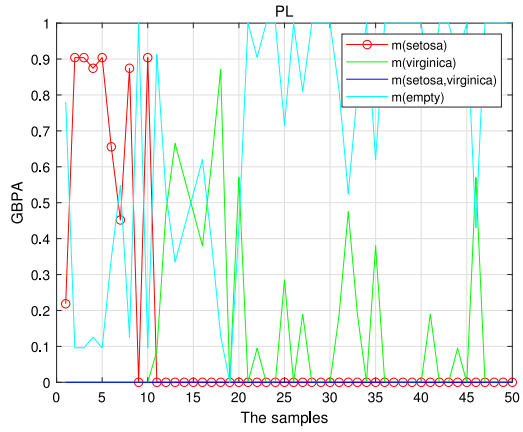


(a) Triangular fuzzy number of SL

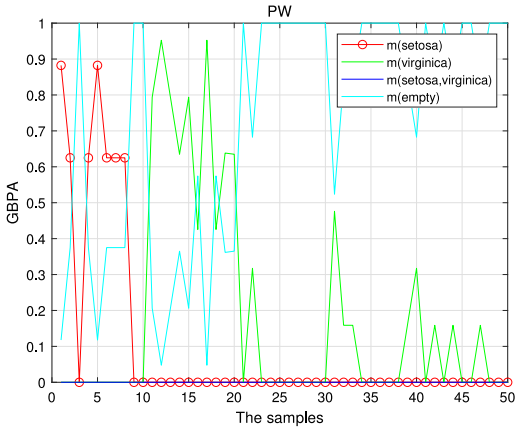


(b) Triangular fuzzy number of SW

Fig. 2. Generate GBPA for SL and SW.



(a) Triangular fuzzy number of PL



(b) Triangular fuzzy number of PW

Fig. 3. Generate GBPA for PL and PW.

Table 2
The average GBPA value after calculating FGS on each attribute.

Attribute	SL	SW	PL	PW	Mean
$m(\emptyset)$	0.4283	0.5451	0.6829	0.6995	0.589

class sample. The results are shown in Table 3. Then the triangular fuzzy models of the training set samples are established based on Eq. (19).

Step 1.2: The GBPA of each attribute can be generated by the strong constraint generation method (Deng and Yong, 2015). The results are shown in Figs. 2 and 3.

Step 1.3: Calculate the value of $\overline{m(\emptyset)}$, and the result is shown in Table 2. In light of the result of the experiment, $\overline{m(\emptyset)} = 0.589$, which is greater than $p = \frac{1}{2}$. Therefore, the FOD is incomplete, and it is necessary to determine how many targets in the FOD.

Step 2: Algorithm 2 is applied to reconstruct the FOD.

Step 2.1: Use Monte Carlo sampling for each attribute based on uniform distribution to generate as many samples as the training set. To reduce the error, we repeat it 20 times.

Step 2.2: FCM is applied to the data generated by each Monte Carlo sampling, and $\log\left(J^*_{m,k}\right)$ can be obtained after each clustering. The results of each sampling are recorded in Table 4. In addition, we also need to calculate $\log\left(J_{m,k}\right)$ of the original data set. The result is recorded in Table 5.

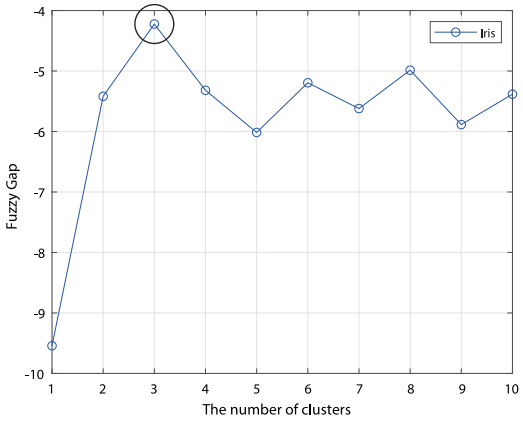
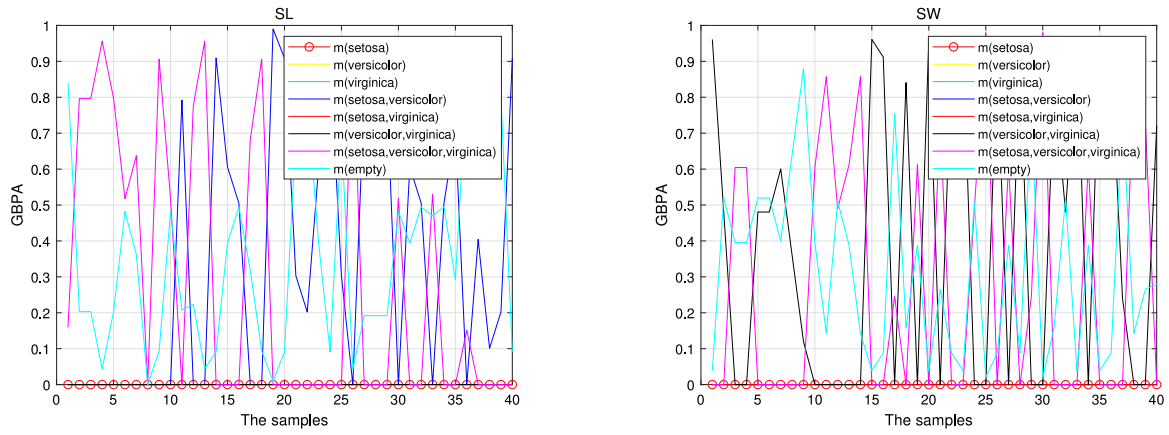


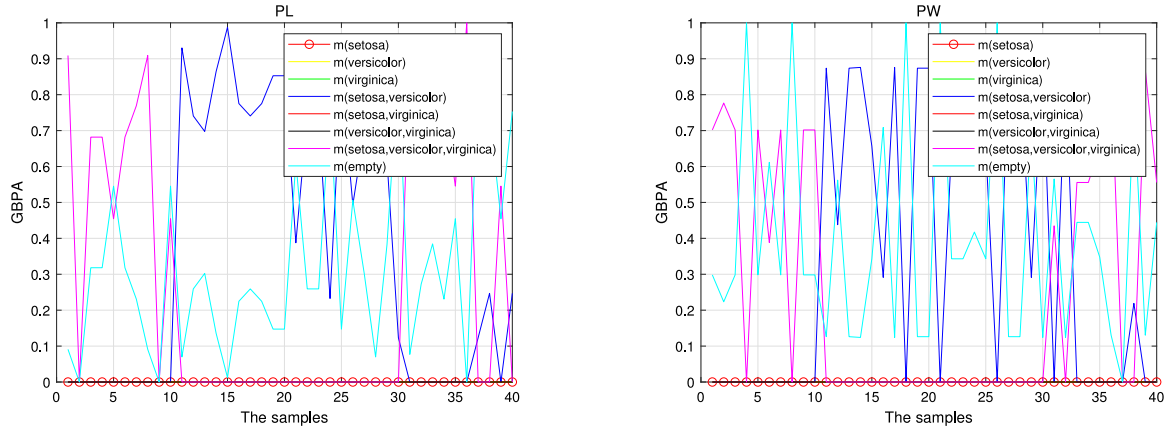
Fig. 4. The FGS with different clusters.

Step 2.3: FGS can be calculated according to the Eq. (15). The result is shown in Fig. 4. It is easy to find out the true number of clusters k based on the highest point of the ordinate. It is the best clustering



(a) Generate GBPA by triangular fuzzy number model (b) Generate GBPA by triangular fuzzy number model

Fig. 5. Generate GBPA for SL and SW.



(a) Generate GBPA by triangular fuzzy number model (b) Generate GBPA by triangular fuzzy number model

Fig. 6. Generate GBPA for PL and PW.

Table 3
Triangular fuzzy number of each attribute.

Category	setosa	virginica
SL	(4.300,5.005,5.800)	(4.900,6.617,7.900)
SW	(2.900,3.450,4.400)	(2.200,2.955,3.800)
PL	(1.000,1.460,1.900)	(4.500,5.540,6.900)
PW	(0.100,0.258,0.600)	(1.400,2.028,2.500)

number of samples, and it is the true number of targets in the open world. Therefore, the number of targets in the current system is 3.

Step 3: Judge again whether the FOD is complete. After determining the number of targets in the incomplete FOD, it needs to be reconstructed. By repeating step 1, the new triangular fuzzy number models can be established. Then generate the GBPAs again. The results are shown in Figs. 5 and 6. Finally calculate the $m(\emptyset)$ again. The result is recorded in Table 6. Now the value of $\overline{m(\emptyset)}$ is less than $p = \frac{1}{2}$. Therefore, the FOD has been complete. Because we have 2 known targets and the optimal number of clusters $k = 3$, the number of unknown targets is $3 - 2 = 1$.

4.2. Further experiments

In order to further illustrate the efficiency of the FGS, this section will use other UCI data sets for experiments (Kahraman et al., 2013; Valentini and Masulli, 2002), including glass, Haberman, knowledge, robot, seeds, and WDBC. The details of the data are in Table 7. Furthermore, we will take some existing methods as the standard of comparison.

Liu used Elbow method (Liu and Deng, 2021) to determine the number of unknown targets in the incomplete FOD. His method can be implemented in two steps. In the first step, GBPAs are generated, and then judge the integrity of the FOD. The second step is to find out the optimal number of clusters with the Elbow method to determine how many unknown targets are in the open world. The results can be calculated by Eq. (10), and they are shown in Figs. 7 and 8. The analysis results show that Liu's method sometimes appears not obvious. For example, for the elbow point of glass, Liu regards $k = 3$ as the elbow point. However, according to the definition of Elbow method, in fact, $k = 2$ also satisfies the condition. The same problems appear in Knowledge, Robot, and WDBC. On the contrary, the proposed method can accurately identify the number of unknown targets in all datasets.

Robert et al. used GS (Tibshirani and Hastie, 2001) to do the same. The results can be calculated by Eq. (13), and they are shown in Figs. 9 and 10. GS overcomes the ambiguity of Elbow method because of the obvious results. However, for some data sets, they are not very effective. For instance, there are 2 targets in the dataset Haberman, but the result of GS is 1. And the real number of targets of glass is 6, but the result of GS is 3.

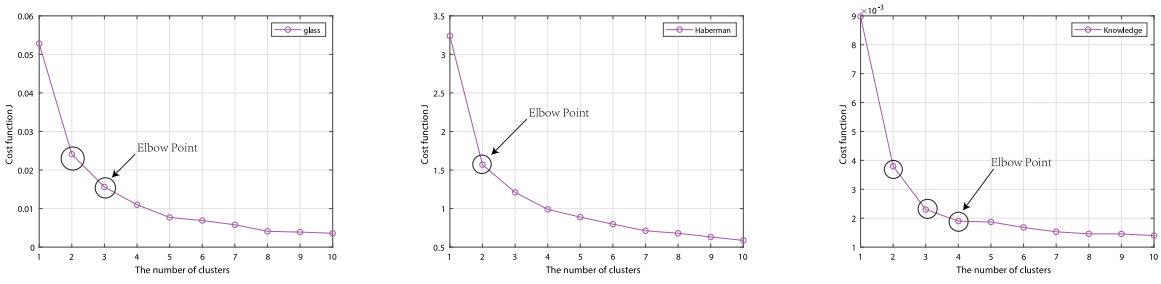
ML is a new technology, which is widely used in optimization (Abdalzaher and Muta, 2020), game theory (Abdalzaher et al., 2016) and other fields because of its efficient learning efficiency. However, ML algorithms always contain some adjustable parameters, called hyper-parameters. The setting of hyper-parameters is the key of ML algorithm.

Table 4
The value of $\log\left(J^*_{m,k}\right)$ corresponding to different cluster numbers k in each Monte Carlo sampling b.

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
$b=1$	10.00686	6.21267	6.79647	6.97014	6.46890	6.74887	6.60922	6.78373	6.26230	6.16481
$b=2$	10.00686	6.21267	6.79647	6.89738	6.80691	6.45116	6.46164	6.47006	6.25679	5.93485
$b=3$	10.00686	7.97449	7.38228	7.21152	6.86908	5.50352	6.46164	6.57936	5.82602	6.33119
$b=4$	10.00686	7.97449	6.96657	7.21152	6.12375	6.81897	6.27725	6.25417	6.82420	5.82286
$b=5$	10.00686	6.21267	6.79647	7.21152	7.46608	5.59574	7.11646	6.78217	6.55439	5.21029
$b=6$	10.00686	7.97449	6.79647	6.66097	6.46890	6.14225	5.74622	6.01913	5.96284	6.36327
$b=7$	10.00686	6.21267	7.38228	7.38144	7.46607	5.85080	5.81813	6.78011	8.38127	6.16489
$b=8$	10.00686	7.97449	6.79647	7.21152	7.53344	7.72870	6.36462	7.23400	5.64595	6.61042
$b=9$	10.00686	7.97449	6.79647	6.50143	6.40672	7.72893	5.81813	6.80186	6.74925	6.70249
$b=10$	10.00686	7.97449	6.79647	5.80181	6.63151	6.39785	6.46164	6.17173	6.97178	5.58596
$b=11$	10.00686	6.21267	8.53937	6.54129	7.21925	5.59574	6.46164	6.17093	6.94995	6.36417
$b=12$	10.00686	6.21267	6.96657	7.14434	6.12377	6.38680	6.62024	6.95449	6.63978	6.36327
$b=13$	10.00686	7.97449	6.96654	6.54130	6.18797	6.14221	6.76185	5.93893	6.82420	6.13658
$b=14$	10.00686	7.97449	7.38227	5.65421	7.53344	5.59574	5.74644	5.84915	5.32765	6.61944
$b=15$	10.00686	6.21267	6.96657	6.54129	5.97621	7.21539	6.88256	8.21144	6.23598	6.97621
$b=16$	10.00686	7.97449	6.96657	7.38144	6.18798	6.39785	5.74624	6.26556	6.25164	5.97802
$b=17$	10.00686	6.21267	6.96657	7.21151	6.18798	6.91950	6.27723	7.14692	6.42314	7.08414
$b=18$	10.00686	7.97449	6.96657	6.54129	6.45374	6.71675	5.87340	6.08146	6.21386	6.57607
$b=19$	10.00686	7.97449	6.96657	7.79674	6.25616	6.23924	5.74644	5.67240	6.01289	5.55840
$b=20$	10.00686	6.21267	6.96657	6.66097	6.98284	6.38683	6.36463	6.57935	7.40392	6.54580

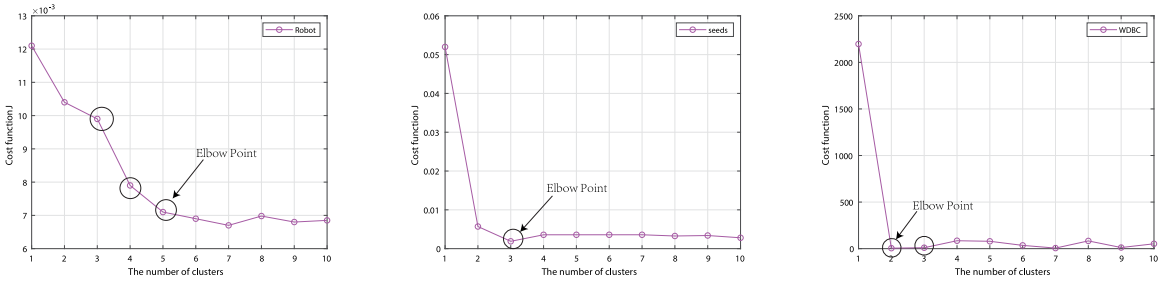
Table 5
The value of $\log\left(J_{m,k}\right)$ of original data set corresponding to different cluster numbers k.

Clusters	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
$\log\left(J_{m,k}\right)$	10.06925	5.70356	6.81652	4.83069	6.64138	4.09659	7.12681	5.63107	5.19248	7.63863



(a) Elbow point of glass (b) Elbow Point of Haberman (c) Elbow point of Knowledge

Fig. 7. Elbow method of glass, Haberman, and Knowledge.



(a) Elbow Point of Robot (b) Elbow point of seeds (c) Elbow Point of WDBC

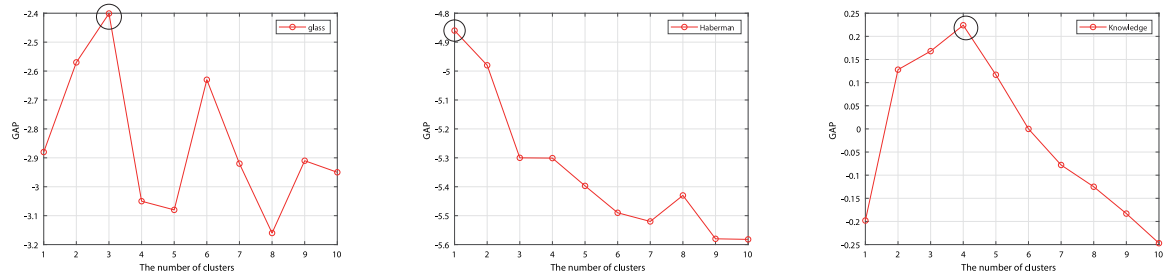
Fig. 8. Elbow method of Robot, seeds, and WDBC.

Table 6
Average value of $m(\emptyset)$ on each attribute.

Attribute	SL	SW	PL	PW	Mean
$m(\emptyset)$	0.3886	0.2361	0.3206	0.3993	0.3661

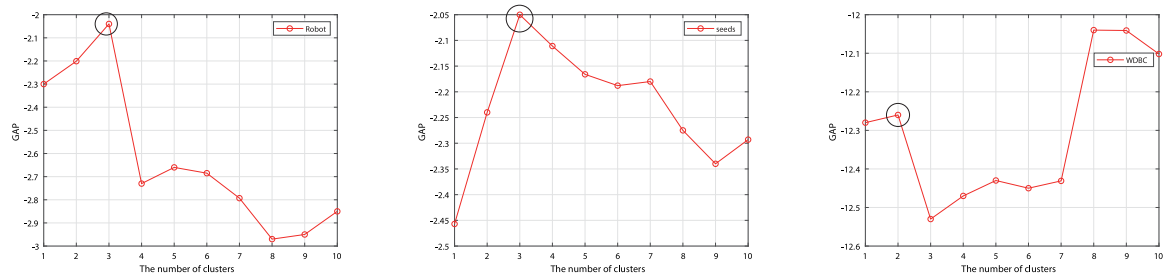
In addition, by adjusting the value of hyper-parameters, the model can meet different practical needs. Many ML classification algorithms are introduced in Abdalzaher et al. (2021). Among them, K-Nearest neighbors (KNN) algorithm is a nonlinear classification model, which can judge the type of data. KNN algorithm can be implemented based on cosine similarity (Abdalzaher et al., 2021). For the problem studied in this paper, it is easy to know that the hyper-parameter of KNN algorithm is the number of unknown targets in the open world. Therefore, when the number of unknown targets is correct, the accuracy of KNN algorithm is the highest.

Figs. 11 and 12 show the results of KNN. To be noticed, for the dataset Knowledge, when the number of categories is 5, the accuracy of KNN is the highest. However, dataset Knowledge has only 4 categories, so we choose 4 as a result. Similarly, we regard 2 as the result of data



(a) GS of glass (b) GS of Haberman (c) GS of Knowledge

Fig. 9. GS of glass, Haberman, and Knowledge.



(a) GS of Robot (b) GS of seeds (c) GS of WDBC

Fig. 10. GS of Robot, seeds, and WDBC.

Table 7
The details of the data sets.

Datasets	The number of instances	The number of attributes	The number of types
glass	214	20	6
Haberman	306	3	2
Knowledge	403	5	4
Robot	5456	24	5
seeds	210	7	3
WDBC	569	30	2

set seeds. Therefore, we can find that even though ML is convenient to use, it is prone to errors.

These three methods have shortcomings because they do not consider the fuzziness of information and the membership relationship between samples and sets. However, FGS can overcome the issues of these methods. It is based on FCM, considering the fuzziness of information. Therefore, using it to determine the number of unknown targets can enhance accuracy. The calculation process of the six data sets is the same as the example in Section 3. Finally, the results of the proposed method are shown in Figs. 13 and 14. We record these results of three methods in Tables 8 and 9. The results are analyzed as follows:

(1) Liu's method can accurately identify Iris, Haberman, seeds, and WDBC, but the glass, Knowledge, and Robot results are not accurate.
(2) GS method accurately identifies Iris, Knowledge seeds, and WDBC, but the other three results for glass, Haberman, and Robot are wrong.
(3) KNN accurately identifies Iris, Haberman, Knowledge, Robot, and WDBC, but the results for glass and seeds are not accurate.
(4) The proposed method can correctly identify the number of unknown targets in all six data sets.

Therefore, through contrastive analysis, it is easy to find that the results becomes more accurate after considering the membership degree of the sample.

Table 8
The optimal k of each method.

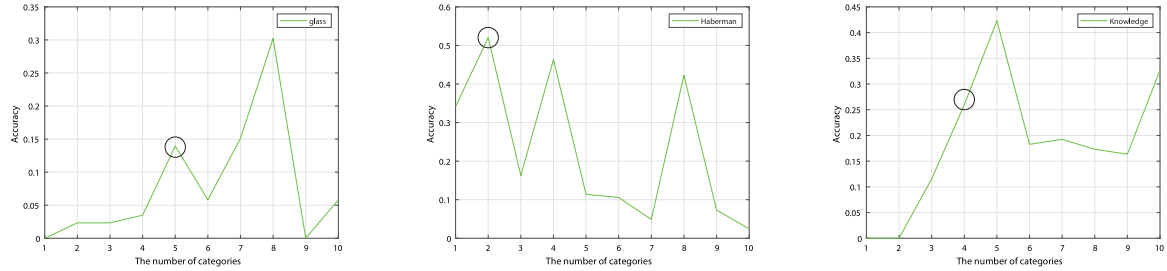
Datasets	Real targets	Liu's method	GS	KNN	Proposed method
Iris	3	3	3	3	3
glass	6	3 (2)	3	5	6
Haberman	2	2	1	2	2
Knowledge	4	4 (2,3)	4	4	4
Robot	5	5 (3,4)	3	5	5
seeds	3	3	3	2	3
WDBC	2	2	2	2	2

Table 9
Is the number of unknown targets determined successfully.

Database	Liu's method	GS	KNN	Proposed method
Iris	Yes	Yes	Yes	Yes
glass	No	No	No	Yes
Haberman	Yes	No	Yes	Yes
Knowledge	No	Yes	Yes	Yes
Robot	No	No	Yes	Yes
seeds	Yes	Yes	No	Yes
WDBC	Yes	Yes	Yes	Yes

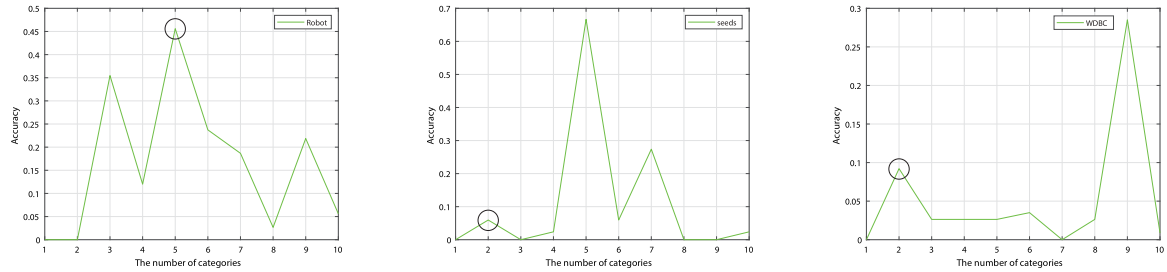
5. Conclusion

The traditional DS evidence theory cannot effectively deal with highly conflicting evidence. As its extension, GET cancels the strict



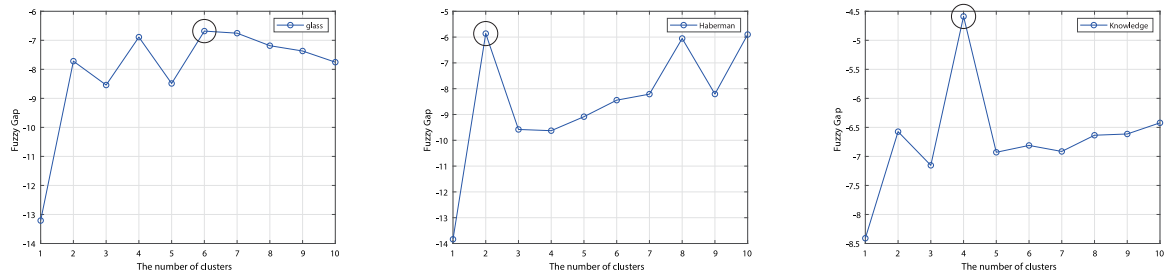
(a) Accuracy of glass (b) Accuracy of Haberman (c) Accuracy of Knowledge

Fig. 11. Accuracy of glass, Haberman, and Knowledge.



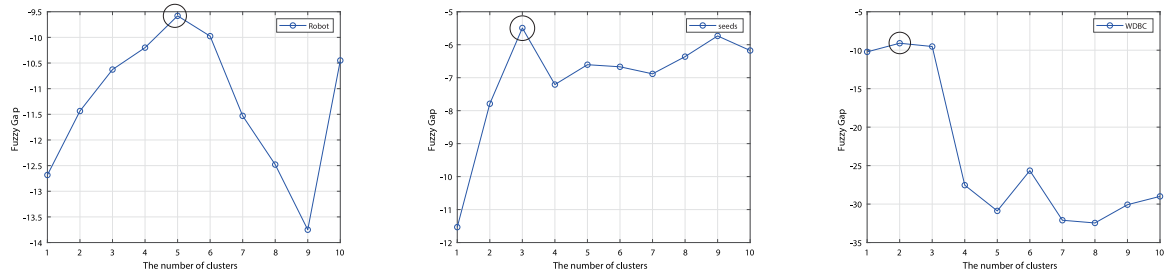
(a) Accuracy of Robot (b) Accuracy of seeds (c) Accuracy of WDBC

Fig. 12. Accuracy of Robot, seeds, and WDBC.



(a) FGS of glass (b) FGS of Haberman (c) FGS of Knowledge

Fig. 13. FGS of glass, Haberman, and Knowledge.



(a) FGS of Robot (b) FGS of seeds (c) FGS of WDBC

Fig. 14. FGS of Robot, seeds, and WDBC.

restriction of $m(\emptyset) = 0$ and has more open advantages. However, how to determine the number of unknown targets in GET is still a problem worthy of discussion.

The clustering algorithm can divide the original data set into several clusters and summarize similar data into the same cluster. Therefore, determining the number of unknown targets in the open world can be transformed into finding the number of clusters. FCM is one of the clustering algorithms. It introduces a membership parameter m, which fully considers the uncertainty in the open world and makes the result more reasonable. However, FCM clustering algorithm has the disadvantage of subjectively controlling the number of clusters. FGS is a valuable tool to determine the optimal number of clusters. In other words, FGS can effectively solve the problems of FCM clustering algorithm. Therefore, this paper combines FCM and FGS to determine the number of unknown targets in the open world. Applying FGS to

GET is an innovative work, which is also one of the main contributions of this article.

Our method is divided into three steps. Firstly, we need to establish triangular fuzzy numbers to generate GBPA. Then, the integrity of the FOD is judged. If it is incomplete, FCM is used to cluster the original data set. Secondly, FGS is used to determine the optimal number of clusters. Finally, the FOD is reconstructed according to the results of FGS. A large number of experiments show that our proposed method is reasonable.

A large amount of data supports our experiment. Based on UCI data sets, including Iris, glass, Haberman, Knowledge, Robot, seeds, and WDBC, we verify the effectiveness of the proposed method. Moreover, we also compare the proposed method with the models in the literature. The comparison results show that our method is more reasonable after considering the fuzziness of data.

In future research, we will improve the ability of our methods to deal with big data and make more contributions to the field of information fusion.

CRediT authorship contribution statement

Zichong Chen: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft, Writing - review & editing. **Rui Cai:** Supervision, Project administration, Funding acquisition.

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.

Acknowledgments

The authors greatly appreciate the reviews' suggestions and the editor's encouragement.

References

Abdalzaher, Mohamed S., Moustafa, Sayed S.R., Abd-Elnaby, Mohammed, Elwekeil, Mohamed, 2021. Comparative performance assessments of machine-learning methods for artificial seismic sources discrimination. IEEE Access 9, 65524–65535.

Abdalzaher, M.S., Muta, O., 2020. A game-theoretic approach for enhancing security and data trustworthiness in IoT applications. IEEE Internet Things J. 7 (11), 11250–11261.

Abdalzaher, M.S., Seddik, K., Muta, O., Abdelrahman, A., 2016. Using Stackelberg game to enhance node protection in WSNs. In: 2016 13th IEEE Annual Consumer Communications Networking Conference (CCNC), pp. 853–856.

Abellan, J., Bosse, E., 2020. Critique of recent uncertainty measures developed under the evidence theory and belief intervals. IEEE Trans. Syst. Man Cybern. 50 (3), 1186–1192.

Aghdaie, M.H., Tafreshi, P.F., 2018. A new perspective on RFM analysis. In: Intelligent Systems: Concepts, Methodologies, Tools, and Applications. pp. 1458–1477.

Bezdek, J.C., 1974. Numerical taxonomy with fuzzy sets. J. Math. Biol. 1 (1), 57–71.

Cao, Z., Chuang, C.-H., King, J.-K., Lin, C.-T., 2019. Multi-channel EEG recordings during a sustained-attention driving task. Sci. Data 6 (19).

Dempster, A.P., 1967. Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat. 38 (2), 325–339.

Deng, Y., 2020. Information volume of mass function. Int. J. Comput. Commun. Control 15 (6), 3983.

Deng, J., Deng, Y., 2021. Information volume of fuzzy membership function. Int. J. Comput. Commun. Control 16 (1), 4106.

Deng, J., Deng, Y., Cheong, K.H., 2021. Combining conflicting evidence based on pearson correlation coefficient and weighted graph. Int. J. Intell. Syst..

Deng, X., Jiang, W., 2020. On the negation of a Dempster–Shafer belief structure based on maximum uncertainty allocation. Inform. Sci. 516, 346–352.

Deng, Yong, 2015. Generalized evidence theory. Appl. Intell. 43 (3), 530–543.

Dutta, P., 2018. An uncertainty measure and fusion rule for conflict evidences of big data via Dempster-Shafer theory. Int. J. Image Data Fusion 9 (1–4), 152–169.

Dutta, Palash, 2017. An uncertainty measure and fusion rule for conflict evidences of big data via Dempster–Shafer theory. Int. J. Image Data Fusion 1–18.

Fei, Liguo, Feng, Yuqiang, Liu, Luning, 2019. Evidence combination using OWA-based soft likelihood functions. Int. J. Intell. Syst. 34 (9), 2269–2290.

Fei, L., Lu, J., Feng, Y., 2020. An extended best-worst multi-criteria decision-making method by belief functions and its applications in hospital service evaluation. Comput. Ind. Eng. 142, 106355.

Feng, F., Cho, J., Pedrycz, W., Fujita, H., Herawan, T., 2016. Soft set based association rule mining. Knowl.-Based Syst. 111, 268–282.

Fu, C., Chang, W., Yang, S., 2020. Multiple criteria group decision making based on group satisfaction. Inform. Sci. 518, 309–329.

Fujita, H., Ko, Y.-C., 2020. A heuristic representation learning based on evidential memberships: Case study of UCI-SPECTF. Internat. J. Approx. Reason. 120, 125–137.

Galadi-Enriquez, D., Jordi, C., Trullols, E., 1998. The overlapping open clusters NGC 1750 and NGC 1758. III. Cluster-field segregation and clusters physical parameters. Astron. Astrophys. -Berlin.

Gao, X., Pan, L., Deng, Y., 2021. Quantum pythagorean fuzzy evidence theory (QPFET): A negation of quantum mass function view. IEEE Trans. Fuzzy Syst. 99, 1.

Garg, H., Kumar, K., 2019. Linguistic interval-valued atanassov intuitionistic fuzzy sets and their applications to group decision making problems. IEEE Trans. Fuzzy Syst. 27 (12), 2302–2311.

Haenni, R., 2002. Are alternatives to Dempster's rule of combination real alternatives? Inf. Fusion 3 (3), 237–239.

Huang, Z., Wang, T., Liu, W., Valencia-Cabrera, L., Pérez-Jiménez, M.J., Li, P., 2021. A fault analysis method for three-phase induction motors based on spiking neural p systems. Complexity 2021 (2087027), 19.

Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern Recognit. Lett. 31 (8), 651–666.

Jiang, W., Yue, C., Wang, S., 2017. A method to identify the incomplete framework of discernment in evidence theory. Math. Probl. Eng. 2017, 1–15.

Jiang, W., Zhan, J., 2017. A modified combination rule in generalized evidence theory. Appl. Intell. 46, 630–640.

Jiang, W., Zhan, J., Zhou, D., Li, X., 2016. A method to determine generalized basic probability assignment in the open world. Math. Probl. Eng. 2016 (5), 1–11.

Jousselme, A.L., Grenier, D., Bossé, éloi, 2001. A new distance between two bodies of evidence. Inf. Fusion 2 (2), 91–101.

Kahraman, H.T., Sagiroglu, S., Colak, I., 2013. The development of intuitive knowledge classifier and the modeling of domain dependent data. Knowl.-Based Syst. 37 (2), 283–295.

Lai, J.W., Chang, J., Ang, L., Cheong, K.H., 2020. Multi-level information fusion to alleviate network congestion. Inf. Fusion 63, 248–255.

Li, D., Deng, Y., Cheong, K.H., 2021. Multisource basic probability assignment fusion based on information qualitys. Int. J. Intell. Syst. 36 (4), 1851–1875.

Li, M., Huang, S., De Bock, J., De Cooman, G., Pižurica, A., 2020. A robust dynamic classifier selection approach for hyperspectral images with imprecise label information. Sensors 20 (18), 5262.

Liao, H., Ren, Z., Fang, R., 2020. A Deng-entropy-based evidential reasoning approach for multi-expert multi-criterion decision-making with uncertainty. Int. J. Comput. Intell. Syst. 13 (1), 1281–1294.

Liu, W., 2006. Analyzing the degree of conflict among belief functions. Artificial Intelligence 170 (11), 909–924.

Liu, F., Deng, Y., 2021. Determine the number of unknown targets in open world based on elbow method. IEEE Trans. Fuzzy Syst. 29 (5), 986–995.

Liu, Z., Zhang, X., Niu, J., Dezert, J., 2020a. Combination of classifiers with different frames of discernment based on belief functions. IEEE Trans. Fuzzy Syst. 29 (7), 1764–1774.

Liu, P., Zhang, X., Pedrycz, W., 2020b. A consensus model for hesitant fuzzy linguistic group decision-making in the framework of Dempster-Shafer evidence theory. Knowl.-Based Syst. 212, 106559.

Mo, H., 2021. A SWOT method to evaluate safety risks in life cycle of wind turbine extended by d number theory. J. Intell. Fuzzy Systems 40 (3), 4439–4452.

Moustafa, S.S.R., Abdalzaher, M.S., Yassien, M.H., Wang, T., Elwekeil, M., Hafiez, H.E.A., 2021. Development of an optimized regression model to predict blast-driven ground vibrations. IEEE Access 9, 31826–31841.

Murphy, C.K., 2000. Combining belief functions when evidence conflicts. Decis. Support Syst. 29 (1), 1–9.

Pan, L., Gao, X., Deng, Y., Cheong, K.H., 2021. The constrained Pythagorean fuzzy sets and its similarity measure. IEEE Trans. Fuzzy Syst. 1.

Schubert, J., 2011. Conflict management in Dempster–Shafer theory using the degree of falsity. Internat. J. Approx. Reason. 52 (3), 449–460.

Sentelle, C., Hong, S.L., Georgiopoulos, M., Anagnostopoulos, G.C., 2007. A fuzzy gap statistic for fuzzy c-means. In: IASTED International Conference on Artificial Intelligence and Soft Computing. Vol. 2007, ASC, pp. 68–73.

Shafer, G.A., 1978. A mathematical theory of evidence. Technometrics 20 (1), 106.

Song, Y., Deng, Y., 2021. Entropic explanation of power set. Int. J. Comput. Commun. Control 16 (4), 4413.

Song, Y., Wang, X., Lei, L., Yue, S., 2016. Uncertainty measure for interval-valued belief structures. Measurement 80, 241–250.

Song, Y., Zhu, J., Lei, L., Wang, X., 2020. A self-adaptive combination method for temporal evidence based on negotiation strategy. Sci. China Inf. Sci. 63, 210204.

Su, X., Li, L., Shi, F., Qian, H., 2018. Research on the fusion of dependent evidence based on mutual information. IEEE Access 6, 71839–71845.

Sun, R., Deng, Y., 2019. A new method to identify incomplete frame of discernment in evidence theory. IEEE Access PP, 1.

Tang, M., Liao, H., Herrera-Viedma, E., Chen, C.P., Pedrycz, W., 2020. A dynamic adaptive subgroup-to-subgroup compatibility-based conflict detection and resolution model for multicriteria large-scale group decision making. IEEE Trans. Cybern. 1–12.

Thorndike, R., 1953. Who belongs in the family? Psychometrika 18 (4), 267–276.

Tian, Y., Liu, L., Mi, X., Kang, B., 2020. ZSLF: A new soft likelihood function based on Z-numbers and its application in expert decision system. IEEE Trans. Fuzzy Syst. 29 (8), 2283–2295.

Tibshirani, R., Hastie, W.T., 2001. Estimating the number of clusters in a data set via the gap statistic. J. R Stat. Society B 63 (2), 411–423.

Valentini, G., Masulli, F., 2002. Neurobjects: an object-oriented library for neural network development. Neurocomputing 48 (1–4), 623–646.

Wang, H., Fang, Y.-P., Zio, E., 2021. Risk assessment of an electrical power system considering the influence of traffic congestion on a hypothetical scenario of electrified transportation system in new york state. IEEE Trans. Intell. Transp. Syst. 22 (1), 142–155.

Wang, T., Liu, W., Zhao, J., Guo, X., Terzijae, V., 2020. A rough set- based bio-inspired fault diagnosis method for electrical substations. Int. J. Electr. Power Energy Syst. 119 (105961), 10.

Wong, J.A.H.A, 1979. Algorithm AS 136: A K-means clustering algorithm. J. R Stat. Soc. 28 (1), 100–108.

Xiao, F., 2020. CEQD: A complex mass function to predict interference effects. IEEE Trans. Cybern. 99, 1–13.

Xiao, F., 2021. CaFtR: A fuzzy complex event processing method. Int. J. Fuzzy Syst. 39 (4).

Xu, X., Li, Z., Li, G., Li, J., Wen, C., 2018. An acoustic resonance-based liquid level detector with error compensation. IEEE Trans. Instrum. Meas. 68 (4), 963–971.

Xue, Y., Deng, Y., 2021. Interval-valued belief entropies for Dempster Shafer structures. Soft Comput. 25, 8063–8071.

Ye, J., Zhan, J., Ding, W., Fujita, H., 2021. A novel fuzzy rough set model with fuzzy neighborhood operators. Inform. Sci. 544, 266–297.

Zhang, H., Deng, Y., 2021. Entropy measure for orderable sets. Inf. Sci. 561, 141–151.

Zhong, P., Fukushima, M., 2007. Regularized nonsmooth Newton method for multi-class support vector machines. Optim. Methods Softw. 22 (1), 225–236.

Zhou, M., Liu, X.-B., Chen, Y.-W., Qian, X.-F., Yang, J.-B., Wu, J., 2020. Assignment of attribute weights with belief distributions for MADM under uncertainties. Knowl.-Based Syst. 189, 105110.