IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received December 21, 2018, accepted January 5, 2019, date of publication February 1, 2019, date of current version February 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2893884

A New Method to Identify Incomplete Frame of Discernment in Evidence Theory

RENLIANG SUN[1] AND YONG DENG[2,3]

[1]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[2]Institution of Fundamental and Frontier Science, University of Electronic Science and Technology of China, Chengdu 610054, China
[3]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN 37235, USA

Corresponding author: Yong Deng (prof.deng@hotmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61573290 and Grant 61503237.

ABSTRACT One assumption of Dempster–Shafer evidence theory (D–S theory) is the closed world. However, how to determine whether the frame of discernment is incomplete or not is still an open issue. In this paper, a new method is proposed based on a minimum spanning tree. For each new sample collected in the system, the mass function of the empty set is determined based on the proposed method. The value of the mass function of the empty set is used to determine whether the frame of discernment is incomplete or not. The experimental results on several benchmark datasets illustrate that the proposed method has a high accuracy in identifying the incomplete frame of discernment while the error rate is low.

INDEX TERMS Dempster-Shafer evidence theory, generalized evidence theory, open world, incomplete frame of discernment, minimum spanning tree covering model, classification.

I. INTRODUCTION

Information fusion is an extensive technology of information process that makes use of multi-source information to obtain a more fundamental understanding of things or targets. It is one of the key technologies to improve the intelligence of intelligent system.

Dempster-shafer evidence theory (D-S theory) expands the basic event space in probability theory into the power set of basic events (also known as the framework of discernment(FOD)), and establishes the basic probability assignment function (BPA) on the framework of discernment [1], [2]. In addition, D-S theory provides a Dempster combination rule, which can achieve evidence fusion without prior information. From this perspective, D-S theory can represent and process uncertain information more effectively than probability theory, which contributes to its widely use in the field of information fusion.

Although D-S theory is widely used in many fields such as information fusion systems [3]–[9], complex network [10]–[15], target identification [16], [17], decision-making method [18]–[25], fuzzy systems [26]–[30], there are still many problems with D-S theory, which restrict its application in real situation. One of its key problems is that it does not take into account the incompleteness of the FOD [31], [32]. For example, assume that a system knows target A and target B. When the system identifies a new unknown target C, traditional D-S theory will treat it as A or B. In this situation, reports from sensors often conflict highly with each other and system often reach conclusions that defy common sense.

Many scholars have proposed varieties of methods to solve the problem. Smets and Kennes [33] put forward the transferable belief model (TBM) and introduced the concept of closed world and open world. The TBM thought that the classical evidence theory only assumed that the information fusion environment was a closed world, which led to a high degree of confliction among various evidences when the FOD was incomplete, and the system would often get wrong fusion results. Unfortunately, TBM did not propose how to determine whether the system was incomplete in the FOD.

Based on the concepts of closed world and open world, Deng [34] proposed the basic framework of generalized evidence theory (GET). The theory defines the generalized basic probability assignment function (GBPA), $m(\emptyset)$ indicates the degree of support for the proposition that the framework of discernment is incomplete.

Deng also put forward a method to determine whether the system was incomplete in the FOD according to the GBPA assigned to $\emptyset$. In this method, each attribute of a sample should be treated as an evidence. If the average of $m(\emptyset)$ of evidences from one sample is larger than the threshold value 0.5 and the $m(\emptyset)$ which the attributes are fused from one sample according to GCR is larger than 0.8, the system is incomplete in the FOD. Unfortunately, the method of determining whether the system was incomplete in the FOD still has some shortcomings. Firstly, the threshold value 0.5 is a value set subjectively. It remains to be seen whether it is reasonable in different cases. Secondly, the fusion process of attributes according to GCR is in low efficiency especially when a sample has lots of attributes. Thirdly, the accuracy of recognizing a new target is still not too high.

In this paper, a new method to identify incomplete frame of discernment is proposed. The rest of this paper is organized as follows: Section 2 begins with a brief introduction to D-S theory and some related concepts of GET. Then we introduce the algorithm with reject option based on adaptive minimum spanning tree covering model(MST covering model). In Section 3, we adapt the algorithm in Section 2 and set up a formula to generate m(∅) according to the adapted algorithm in order to determine whether the system is incomplete in the FOD. Some experiments on the common datasets are presented in Section 4 to show the effectiveness of the method we put forward. Conclusions are given in the last section of this paper.

II. PRELIMINARIES

In this section, some preliminary information including D-S theory, GET and a classification algorithm based on MST covering model, are briefly provided.

A. DEMPSTER-SHAFER EVIDENCE THEORY

D-S theory satisfies weaker conditions than the bayesian probability theory, that is, the theory can represent uncertain information more effectively than probability theory, which makes it widely used in many fields [35]–[46]. Formally, the definitions of D-S theory are shown as follows:

Definition 1: A set of hypotheses Θ is the exhaustive hypotheses of variable, X. The elements are mutually exclusive in Θ. Then Θ is called the frame of discernment, defined as follows [1], [2]:

$$\Theta = \{H_1, H_2, \cdots, H_i, \cdots, H_N\} \tag{1}$$

The power set of Θ is denoted by 2^Θ, and

$$2^\Theta = \{\emptyset, \{H_1\}, \cdots, \{H_N\}, \{H_1, H_2\}, \cdots, \{H_1, H_2, \cdots, H_i\}, \cdots, \Theta\} \tag{2}$$

where $\emptyset$ is an empty set.

Definition 2: A basic probability assignment function m is a mapping of 2^Θ to a probability interval [0, 1], formally defined by [1], [2]:

$$m : 2^\Theta \to [0, 1] \tag{3}$$

which satisfies the following conditions:

$$m(\emptyset) = 0 \quad \sum_{A \in 2^\Theta} m(A) = 1$$
$$0 \le m(A) \le 1 \quad A \in 2^\Theta \tag{4}$$

The mass $m(A)$ represents how strongly the evidence supports A.

Given two BPAs, the Dempster's combination rule can be used to obtain the final results [1], [2]:

$$\begin{cases} m(\emptyset) = 0 \\ m(A) = \dfrac{\sum_{B \bigcap C = A} m_1(B)m_2(C)}{1 - K} \end{cases} \tag{5}$$

where $K = \sum_{B \bigcap C = \emptyset} m_1(B)m_2(C)$.

B. GENERALIZED EVIDENCE THEORY

Generalized evidence theory (GET) defines the generalized basic probability assignment function (GBPA), $m(\emptyset)$ indicates the degree of support for the proposition that the framework of discernment is incomplete. A generalized combination rule which can fuse GBPA was also proposed. GET is the extension of D-S theory. When $m(\emptyset)$ is assigned to zero, which means the information fusion environment is a closed world, GET degenerates to D-S theory. The preliminary definitions of GET are shown as follows:

Definition 3: Suppose that U is a FOD in an open world. Its power set, 2_G^U, is composed of 2^U propositions, $\forall A \subset U$. A GBPA is a mapping $m : 2_G^U \to [0, 1]$ that satisfies [34]:

$$\sum_{A \subset 2_G^U} m_G(A) = 1 \tag{6}$$

then m_G is the GBPA of the FOD U. Note that $m_G(\emptyset) = 0$ is not necessary in GBPA. If $m_G(\emptyset) = 0$, the GBPA reduces to a traditional BPA.

Definition 4: Given two GPBAs(m_1 and m_2), the generalized combination rule (GCR) is defined as follows [34]:

$$m(A) = \frac{(1 - m(\emptyset)) \sum_{B \bigcap C = A} m_1(B)m_2(C)}{1 - K} \tag{7}$$

with

$$K = \sum_{B \bigcap C = \emptyset} m_1(B)m_2(C) \tag{8}$$
$$m(\emptyset) = m_1(\emptyset)m_2(\emptyset) \tag{9}$$
$$m(\emptyset) = 1 \quad \textit{if} \quad \textit{and} \quad \textit{only} \quad \textit{if} \quad K = 1 \tag{10}$$

C. A CLASSIFICATION ALGORITHM WITH REJECT OPTION BASED ON ADAPTIVE MINIMUM SPANNING TREE COVERING MODEL

The algorithm is very effective in solving the problem of rejected classification by accepting the idea of boundary cover and virtual sample [47], [48]. It describes the known class using Minimum Spanning Tree (MST) with the assumption that the edges of the graph are also basic elements of the classifier which offers additional virtual training data for a better coverage. By this model, similar samples from the same class are divided into a connected geometric coverage area, and similar samples from different classes are divided into different geometric coverage areas. Furthermore, in order to reduce the degradation of the rejection performance due to the existence of unreasonable additional virtual training data, an adjustable coverage radius strategy is presented in coverage construction. Then the test pattern of non-training classes could be rejected by the coverage decision boundary. The preliminary definitions of algorithm are shown as follows:

Definition 5: Let $X_{tr}^k = \{x_i\}_{i=1}^{n_k} \in R^N$ represent the set of w_k training samples containing n_k samples [49]. Define a weighted undirected connection graph $G = \{V, E\}$, where $V = X$ represents the vertex set of G, $E = \{e_{ik} = (x_i, x_j)\}$ represents the set of sides of G. By using Euclidean distance measure, the length d_{ij} of edge (x_i, x_j) in G will be $d_{ij} = \|e_{ij}\| = \|x_i - x_j\|$. Search for a subtree of G that contains $(n_k - 1)$ edges, which meets the following conditions :(1) connect all vertices, (2) have no loop, and (3) minimize the sum of all edge weights. Such a subgraph is called the minimum spanning tree of graph G.

Definition 6: Let $X_{tr}^k = \{x_i\}_{i=1}^{n_k} \in R^N$ represent the set of w_k training samples. Then establish minimum spanning tree covering model for each class of training sample. The coverage area is represented as $\{\Omega_k, k = 1, 2, \cdots, C\}$. For the test sample x, whether it falls within the coverage area of each class is determined according to equation as follows [48]:

$$L_k(x) = g(d(x, X_{tr}^k) \leq r_k) = \begin{cases} 1, & x \in \Omega_k \\ 0, & x \notin \Omega_k \end{cases} \tag{11}$$

In Eq.(11), r_k represents the covering radius of Ω_k, $d(x, X_{tr}^k)$ is the distance between x and class w_k, $g(\bullet)$ represents a correspondence.

Assuming that the rejected samples belong to class w_r and the accepted samples belong to class w_a, the rejection rule is shown as follows [48]:

$$h(x) = \begin{cases} w_r, & \sum_{i=1}^{C} L_i(x) = 0 \\ w_a, & Others \end{cases} \tag{12}$$

Definition 7: Define the distance $d(x, X_{tr}^k)$ between training sample x and class $\{w_k\}_{k=1}^C$ as follows [48]:

$$d(x, X_{tr}^k) = min(d(x, e_{ij})), \quad \forall e_{ij} \in EMST(w_k) \tag{13}$$

In Eq.(13), EMST represents the $(n_k - 1)$ edges of MST of class $\{w_k\}_{k=1}^C$.

According to the geometry of higher dimensional space, the definition of projection $Pr(x, e_{ij})$ of x onto the edge e_{ij} is shown as follows [48]:

$$Pr(x, e_{ij}) = \left[\left(\frac{x_i - x_j}{\|x_i - x_j\|}\right)'(x - x_i)\right]\left(\frac{x_i - x_j}{\|x_i - x_j\|}\right) + x_i \tag{14}$$

According to geometry, $d(x, e_{ij})$ is actually the minimum value of the distance between x and x_i, x and x_j, x and $Pr(x, e_{ij})$.

$$d(x, e_{ij}) = min(\|x - x_i\|, \|x - x_j\|, \|x - Pr(x, e_{ij})\|) \tag{15}$$

Definition 8: Define that for class w_k, the adaptive covering radius is shown as follows [48]:

$$r_{ij} = r_0 \cdot (exp(-\frac{d_{mean}^2}{h}))^{-1} \cdot exp(-\frac{d_{ij}^2}{h}) \tag{16}$$

In Eq.(16), r_0 represents the reference radius. Define that r_0 is the maximum length, d_{mean} is the average length of the remaining edge of MST after the top 10% is removed. h is radius attenuation control parameters, the value of which is calculated as follows:

$$h = -\frac{(d_{max}^2)}{ln(\beta)}, \quad 0 < \beta < 1 \tag{17}$$

In Eq.(17), β is the cover confidence, which depends on the degree of dispersion of the data distribution. By comparing d_{max} and d_{mean}, the degree of dispersion of the data distribution is defined as follows [48]:

$$\lambda = \frac{d_{max}}{d_{mean}} \tag{18}$$

When $d_{max} >> d_{mean}$, β should approach 0. In contrast, β should approach 1. It is generally recognized that when $\lambda = 3 \sim 5$, $d_{max} >> d_{mean}$.

III. PROPOSED METHOD TO IDENTIFY INCOMPLETE FOD

The method proposed below can effectively judge whether FOD is complete or not, thus helping to resolve the situation of highly conflicting or contradictory conclusions in information fusion.

A. IMPROVEMENT TO EXISTING METHODS

In the rejection rule established in Def.6, covering radius r is an important factor affects the rejection performance. In order for our method to be able to better judge whether the system is in the condition of incomplete FOD, we need to modify the Eq.(16) in Def.8 for the minimum spanning tree overlays adaptive radius. The modified equation is shown below:

$$r_{ij} = r_0 \cdot \beta \cdot exp(\frac{d_{mean}^2 - d_{ij}^2}{d_{max}^2}) \tag{19}$$

In Eq.(19), r_0 represents the reference radius. d_{max} is the maximum length of the length of MST. Define that r_0 is the maximum length, d_{mean} is the average length of the remaining edge of MST after the top 10% is removed. β is the cover confidence, which depends on the degree of dispersion of the data distribution. According to Eq.(18), When $d_{max} >> d_{mean}$, β should approach 0. In contrast, β should approach 1. It is generally recognized that when $\lambda = 3 \sim 5$, $d_{max} >> d_{mean}$.

From Eq.(19), we know that when $d_{ij} > d_{mean}$, the covering radius r_{ij} should be reduced to reduce coverage redundancy. When $d_{ij} < d_{mean}$, the covering radius r_{ij} should be increased to reduce the risk of error rejection of known samples to FOD.

B. A NEW METHOD TO GENERATE $m(\emptyset)$

According to the GET, the GBPA of $\emptyset$ directly reflects the possibility that the system is incomplete in FOD. In this paper, we will still choose $m(\emptyset)$ as the measurement to identify whether the framework is complete.

For a new sample, the GBPA for the $\emptyset$ firstly depends on whether it is in the MST coverage of each known class. If the new sample is not in the MST coverage of each known class, then $m(\emptyset) = 1$, which means the system is completely ignorant of the new sample.

If the new sample is in the MST coverage of a class, $m(\emptyset)$ will be generated based on the relationship between the distance between the new sample and the known sample of the class and the covering radius.

Define that if the distance between the new sample and the known sample of the class is less than $d_{mean}/2$, then $m(\emptyset) = 0$. That is, it is reasonable to think that the new sample is the sample of the known class. If the distance is larger than $d_{mean}/2$, $m(\emptyset)$ will between 0 and 1, indicating that the new sample is likely to belong to the known class.

For each known class w_k, we define $\lambda_k = 1 - m(\emptyset)$. λ_k indicates how much certainty the new sample belongs to class w_k. Finally, the probability measure of the new sample does not belong to the FOD can be expressed as $m(\emptyset) = 1 - \sum_{k=1}^{N} \lambda_k$. $m(\emptyset)$ ranges from 0 to 1, if $m(\emptyset)$ is less than 0, then we let $m(\emptyset) = 0$. To sum up, the formulas to generate the $m(\emptyset)$ are:

$$m_k(\emptyset) = \begin{cases} 1, & \forall e_{ij}, \quad d(x, e_{ij}) > r_{ij} \\ 0, & \exists e_{ij}, \quad d(x, e_{ij}) < \dfrac{d_{mean}}{2} \\ min\left(\dfrac{d(x, e_{ij}) - \dfrac{d_{mean}}{2}}{r_{ij} - \dfrac{d_{mean}}{2}}\right), & \exists e_{ij}, \quad r_{ij} > d(x, e_{ij}) > \dfrac{d_{mean}}{2} \end{cases} \tag{20}$$

$$\lambda_k = 1 - m_k(\emptyset) \tag{21}$$

$$m(\emptyset) = 1 - \sum_{k=1}^{N} \lambda_k, \quad 0 \le m(\emptyset) \le 1 \tag{22}$$

$$if \quad m(\emptyset) < 0, \quad m(\emptyset) = 0 \tag{23}$$

C. IDENTIFICATION OF OPEN WORLD WITH $m(\emptyset)$

Whether FOD is complete or not should be considered from two aspects. One is single new sample, the other is a set of new samples from one class. For single new sample, according to our method of generating $m(\emptyset)$, if $m(\emptyset) = 1$, there is a great chance that this new sample is a sample unknown to the system. That is, FOD is incomplete. For a set of new samples from one class, we define a threshold. If the average value of $m(\emptyset)$ of this set of samples is greater than the threshold value, it is reasonable to think that the current FOD is incomplete. Without loss of generality, the threshold is set as 0.5 in this paper. If the degree of data dispersion is too high, it is feasible to take a value greater than 0.5.

D. PROCUDURE

A flow chart of our method is shown as Fig.1. Details of steps are shown as follows:

STEP 1

For each kind of known class w_k, a MST is established.

STEP 2

According to the improved radius generating formula proposed in Section 3.1, the covering radius of each edge of each MST is obtained. Then the MST coverage is established.

STEP 3

For each new sample entering the system, $m(\emptyset)$ is generated using the proposed method in Section 3.2.

STEP 4

Judge if FOD is incomplete using the proposed method in Section 3.3.

IV. EXPERIMENT

In this section, we will conduct experiments using three commonly used UCI datasets namely Iris, Wine and Car evaluation, to show that our method can effectively identify situations where FOD is incomplete. According to the steps in Section 3.4, an example on the Iris dataset will be given to illustrate in detail how to identify incomplete FOD.

A. EXPERIMENTS ON THE IRIS DATASET

Iris dataset contains 3 classes of 50 instances each, where each class refers to a type of iris plant [50]. In this experiment, we selected Iris *Versicolor* and Iris *Virginica* as known classes and Iris *Setosa* as unknown class. Among 50 instances of Iris *Versicolor* and Iris *Virginica*, 30 instances are randomly selected as the training set, and the remaining 20 instances serve as the test set. 50 instances of Iris *Setosa* serve as the unknown test set.

STEP 1

For the current FOD, known classes are *Versicolor* and *Virginica*. Therefore, for these two classes, establish their minimum spanning trees m_1 and m_2 respectively. Since both classes have 30 instances, the number of vertexes in m_1 and m_2 is 30, and the number of edges is 29.

STEP 2

According to the improved radius generating formula proposed in Section 3.1, the covering radius of each edge in m_1 and m_2 can be obtained.

$$r_{ij} = r_0 \cdot \beta \cdot exp(\frac{d_{mean}^2 - d_{ij}^2}{d_{max}^2}) \tag{24}$$

Since the training instances and test instances are randomly selected, we select the result of one of the experiments. In the edges of the minimum spanning tree m_1, the length of the longest edge d_{max} is 0.8246. After removing 10% of the longest edges in m_1, the length of the longest edge r_0 is

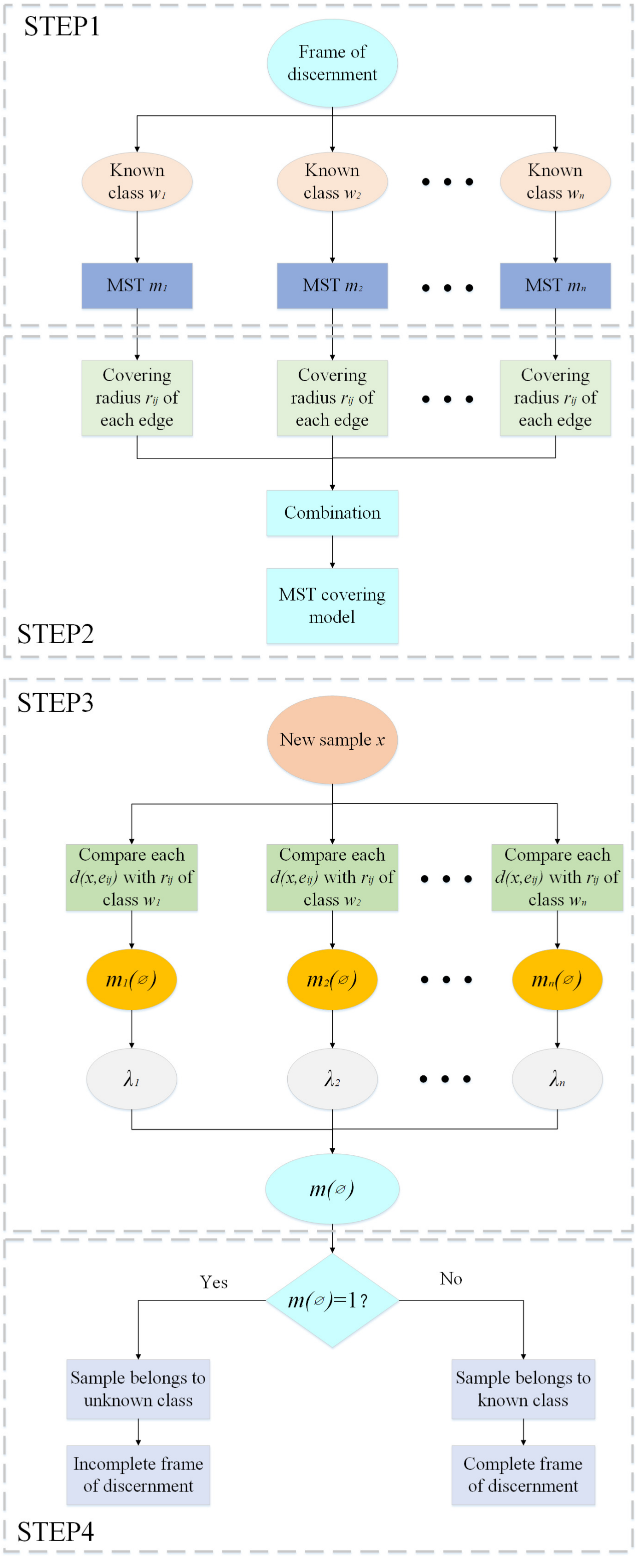

FIGURE 1. The four steps of our proposed method.

0.4899, and the average length of the remaining edges d_{mean} is 0.3219.

Similarly, in the edges of the minimum spanning tree m_2, the length of the longest edge d_{max} is 0.9110. After removing 10% of the longest edges in m_2, the length of the longest edge r_0 is 0.6164, and the average length of the remaining edges d_{mean} is 0.4008.

TABLE 1. The values of attributes of m_1 and m_2.

m_1	m_2
$r_0 = 0.4899$	$r_0 = 0.6164$
$d_{max} = 0.8246$	$d_{max} = 0.9110$
$d_{mean} = 0.3219$	$d_{mean} = 0.4008$
$\lambda = 2.5617$	$\lambda = 2.2730$
$\beta = 0.65$	$\beta = 0.65$

Since the values of d_{max}/d_{mean} of m_1 and m_2 are between 2 and 3, it is reasonable that the value of the empirical constant β is 0.65. d_{ij} is the length of the edge connecting vertex i and vertex j, which has been calculated in *step* 1. We use the same approach for m_2. Finally, the values of attributes of m_1 and m_2 are shown in the table below:

After the coverage radius of each edge is obtained, the minimum spanning tree coverage of m_1 and m_2 is established.

STEP 3

For each new sample entering the system, use the formula proposed in Section 3.2 to generate $m(\emptyset)$ of the new sample.

$$m_k(\emptyset) = \begin{cases} 1, & \forall e_{ij}, \quad d(x, e_{ij}) > r_{ij} \\ 0, & \exists e_{ij}, \quad d(x, e_{ij}) < \dfrac{d_{mean}}{2} \\ min\left(\dfrac{d(x, e_{ij}) - \dfrac{d_{mean}}{2}}{r_{ij} - \dfrac{d_{mean}}{2}}\right), \\ \exists e_{ij}, \quad r_{ij} > d(x, e_{ij}) > \dfrac{d_{mean}}{2} \end{cases} \tag{25}$$

$$\lambda_k = 1 - m_k(\emptyset) \tag{26}$$

$$m(\emptyset) = 1 - \sum_{k=1}^{N} \lambda_k, \quad 0 \le m(\emptyset) \le 1 \tag{27}$$

$$if \quad m(\emptyset) < 0, \quad m(\emptyset) = 0 \tag{28}$$

Here we present a test instance belonging to the *Versicolor* class (known class) and a test instance belonging to the *Setosa* class (unknown class) to illustrate the calculation steps.

The first is the test instance belonging to the *Versicolor* class. According to the distance between it and each edge in m_1, it can be seen that there exists an edge, the distance between it and the edge is less than $d_{mean}/2$. Therefore, $m_1(\emptyset)$ is equal to zero. According to the distance between it and each edge in m_2, it can be seen that there exists an edge, and the distance between it and the edge is greater than $d_{mean}/2$ and less than r_{ij}. Then the value of $m(\emptyset)$ can be obtained. Select the minimum value of $m(\emptyset)$ as the value of $m_2(\emptyset)$. Therefore, $\lambda_1 = 1$, $\lambda_2 = 0.434$, $m(\emptyset) = 1 - \lambda_1 - \lambda_2 = 0$.

Then is the test instance belonging to the *Setosa* class. According to distance between it and each edge in m_1, it can be seen that for every edge, the distance between it and the edge is far greater than r_{ij}. Then, $m_1(\emptyset) = 1$, $\lambda_1 = 0$. The same is true of the distance between it and each edge in m_2.

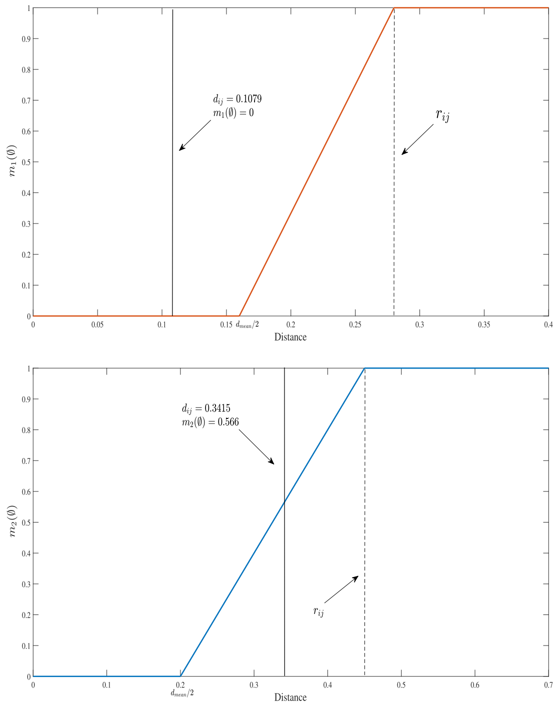

FIGURE 2. $m_1(\emptyset)$ **and** $m_2(\emptyset)$ **of a test instance of** ***Versicolor*** **class.**

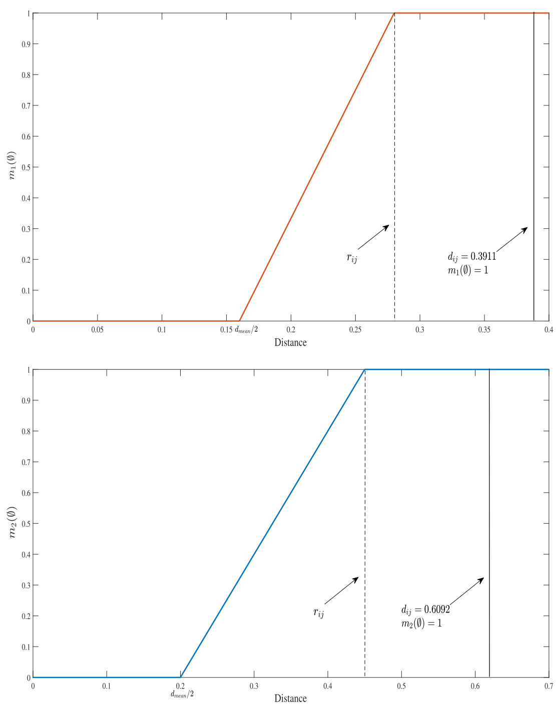

FIGURE 3. $m_1(\emptyset)$ **and** $m_2(\emptyset)$ **of a test instance of** ***Setosa*** **class.**

Then, $m_2(\emptyset) = 1, \lambda_2 = 0$. Therefore, $m(\emptyset) = 1 - \lambda_1 - \lambda_2 = 1$.

STEP 4

According to *STEP* 3, we obtain the $m(\emptyset)$ of 40 known samples and 50 unknown samples respectively. Here, each sample is treated as single new sample. Thus, if $m(\emptyset) = 1$, there is a great chance that this new sample is a sample unknown to the system, which means FOD is incomplete.

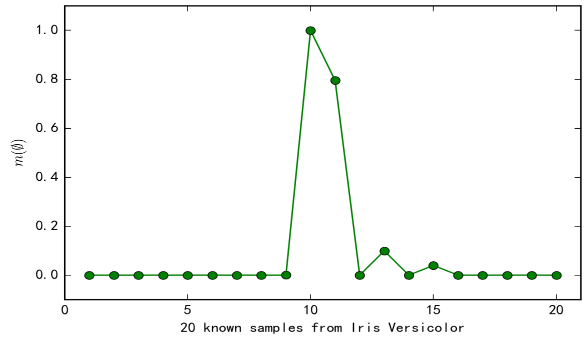

FIGURE 4. $m(\emptyset)$ **generated by known samples from Iris** ***Versicolor*****.**

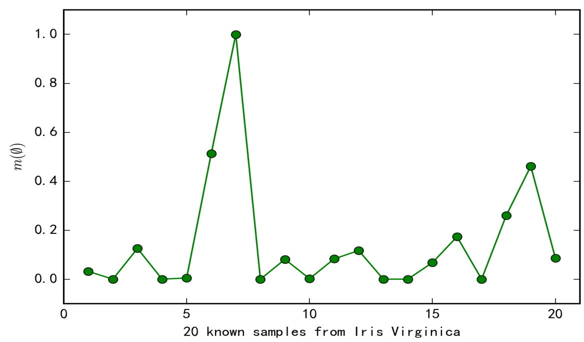

FIGURE 5. $m(\emptyset)$ **generated by known samples from Iris** ***Virginica*****.**

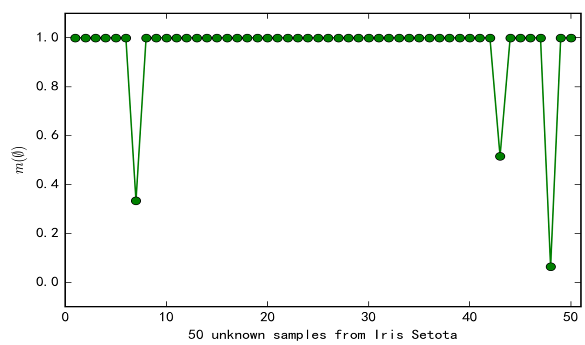

FIGURE 6. $m(\emptyset)$ **generated by unknown samples from Iris** ***Setosa*****.**

TABLE 2. The result on Iris dataset.

Dataset	Recognition accuracy	Rejection accuracy
Iris	92.03%	95.7%

By comparing the result of the system judgment with the class of the new sample itself, the accuracy rate of the system to identify the known sample (recognition accuracy) and the accuracy rate of the system to reject the unknown sample (rejection accuracy) can be obtained. After the experiment is conducted 100 times, $m(\emptyset)$ generated by one of the experiments and the average of accuracy of 100 trials is shown as Fig.4, Fig.5 and Fig.6.

As can be seen from Fig.4, for new samples of the known *Versicolor* class, only one sample's $m(\emptyset) = 1$. That is,

TABLE 3. The result on wine dataset.

Dataset	Recognition accuracy	Rejection accuracy
Wine	77.69%	81%

the system rejects it as an unknown sample. $m(\emptyset)$ of most other samples is equal to 0, indicating that the system believes that most new samples are in FOD. Similarly, from Fig.5 we can see that for new samples of the known class *Virginica*, only one sample 's $m(\emptyset) = 1$. $m(\emptyset)$ of most other samples is very low.

Fig.4 and Fig.5 show that for new samples of known classes, their $m(\emptyset)$ will be very low according to our method. That is, the system thinks they are in FOD. It directly reflects the effectiveness of our method.

As can be seen from Fig.6, for new samples of unknown class *Setosa*, $m(\emptyset)$ of most samples is equal to 1, while $m(\emptyset)$ of only a few samples is less than 1. This means that the system believes that the vast majority of new samples are not in FOD, which also reflects the effectiveness of our approach.

If new samples of known classes are considered as a group of samples, it can be seen from Fig.4 and Fig.5 that the average value of their $m(\emptyset)$ is far less than the threshold value of 0.5, indicating that both groups of new samples are in FOD. If new samples of unknown classes are considered as a group of samples, it can be seen from Fig.6 that the average value of their $m(\emptyset)$ is far greater than the threshold value of 0.5, indicating that this group of new samples is not in FOD, and there is a good chance that this group of new samples belongs to a new unknown class.

B. EXPERIMENTS ON THE WINE DATASET

The Wine dataset is also a widely used dataset in scientific research [51]. It contains the data of three classes of wines and thirteen attributes.

There are 59 instances in class 1, 71 in class 2 and 48 in class 3. In this experiment, we select class 1 and class 3 as known class and class 2 as unknown class. 30 samples are randomly selected from class 1 and class 3 as training samples, 15 samples are randomly selected as test samples respectively, and 60 samples are randomly selected from class 2 as test samples.

According to the steps to determine whether FOD is incomplete proposed in Section 3.4, we obtain the $m(\emptyset)$ of 30 known samples and 60 unknown samples respectively, and the result whether system consider this new sample to be an unknown. By comparing the result of the system judgment with the class of the new sample itself, the accuracy rate of the system to identify the known sample (recognition accuracy) and the accuracy rate of the system to reject the unknown sample (rejection accuracy) can be obtained. After the experiment is conducted 100 times, the average of accuracy of 100 trials is shown as follows:

From Table 3, we can see that both the recognition accuracy and the rejection accuracy are lower than those of Iris data. After analyzing, we think that the method itself has the problem that the performance of high-dimensional space is inferior to that of low-dimensional space. However, the accuracy rate of 80% in identifying unknown samples is sufficient to show that our method can still identify whether FOD is in an incomplete state in high-dimensional space.

TABLE 4. The result on car evaluation dataset.

Dataset	Recognition accuracy	Rejection accuracy
Car Evaluation	93.75%	94.27%

TABLE 5. The summarized results on three datasets.

Dataset	Iris	Wine	Car Evaluation
Recognition accuracy	92.03%	77.69%	93.75%
Rejection accuracy	95.7%	81%	94.27%
Training samples	60	60	60
Test samples	40+50	30+60	40+500
number of tests	100	100	100

Moreover, the traditional method is to combine the evidence of various attributes, and then obtain the total $m(\emptyset)$. Such calculation quantity will increase exponentially with the increase of dimensions. Our method is able to maintain a high accuracy rate even when the computational amount is greatly reduced, so it is more advantageous than the original method.

C. EXPERIMENTS ON THE CAR EVALUATION DATASET

Car Evaluation Database contains the data of four classes of instances and six attributes [52] . Different from the previous two datasets, the values of its six attributes are all discrete variables and also contain the values of qualitative description.

We define the qualitative description value for each attribute which increase from small to large. For example, for buying attributes, we set *low*=1, *med*=2, *high*=3, *vhigh*=4. Because the number of instances varies greatly from class to class, we choose to select *good* and *vgood* as known classes, *unacc* and *acc* as unknown classes. 30 samples are randomly selected from class *vgood* and class *good* as training samples, 20 samples are randomly selected as test samples respectively, and 250 samples are randomly selected from class *unacc* and *acc* as test samples respectively.

According to the steps to determine whether FOD is incomplete proposed in Section 3.4, we obtain $m(\emptyset)$ of 40 known samples and 500 unknown samples respectively, and the result whether system consider this new sample to be an unknown. By comparing the result of the system judgment with the class of the new sample itself, the accuracy rate of the system to identify the known sample (recognition accuracy) and the accuracy rate of the system to reject the unknown sample (rejection accuracy) can be obtained. After the experiment is conducted 100 times, the average of accuracy of 100 trials is shown as follows:

By analyzing Table 4, we think the experimental results on Car Evaluation dataset can represent that in the case of discrete data, our method can still identify whether FOD is incomplete.

Experimental results on the three commonly used datasets are summarized in the following table.

V. CONCLUSION

Although evidence theory has many advantages that makes it widely used in a lot of fields, how to determine whether FOD is incomplete is still worth exploring. In this paper, a method is proposed to judge if FOD is incomplete. Firstly, for each kind of known class, a MST is established. Then, according to the improved radius generating formula, the covering radius of each edge of each MST is obtained. Thus the MST coverage is established. For each new sample entering the system, $m(\emptyset)$ is generated using the proposed method. Finally, judge if FOD is incomplete according to $m(\emptyset)$ generated.

The advantages of the proposed method are:

1. The proposed method has a high accuracy in identifying the incomplete FOD.

2. The error rate of classifying known samples into unknown class is very low.

3. Comparing to existing method, the proposed method does not need complex calculation, which means it can be used in many situations that require real-time computation.

4. The proposed method can still perform well while using a small number of training data.

5. The proposed method is data-driven, which avoids the problems caused by wrong subjective judgment.

The proposed method also has some limitations. From the experimental results of Wine dataset, we can see that the method itself has the problem that the performance of high-dimensional space is inferior to that of low-dimensional space, which may limits the application in real situations. Therefore, our future research direction is to build new model to improve the recognition accuracy and rejection accuracy, especially in the situation of high-dimensional space.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests regarding the publication of this paper.

REFERENCES

[1] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Ann. Math. Statist.*, vol. 38, no. 2, pp. 325–339, Apr. 1967.
[2] G. Shafer, *A Mathematical Theory of Evidence*, vol. 1. Princeton, NJ, USA: Princeton Univ. press, 1976.
[3] C. Zhang, K. Chen, X. Zeng, and X. Xue, "Misbehavior detection based on support vector machine and Dempster-Shafer theory of evidence in VANETs," *IEEE Access*, vol. 6, pp. 59860–59870, 2018.
[4] Y. Li and Y. Deng, "Generalized ordered propositions fusion based on belief entropy," *Int. J. Comput. Commun. Control*, vol. 13, no. 5, pp. 792–807, Oct. 2018.
[5] Y. Gao, C. Sun, R. Li, Q. Li, L. Cui, and B. Gong, "An efficient fraud identification method combining manifold learning and outliers detection in mobile healthcare services," *IEEE Access*, vol. 6, pp. 60059–60068, 2018.
[6] H. Seiti and A. Hafezalkotob, "Developing pessimistic–optimistic risk-based methods for multi-sensor fusion: An interval-valued evidence theory approach," *Appl. Soft Comput.*, vol. 72, pp. 609–623, Nov. 2018.
[7] L. Fei and Y. Deng, "A new divergence measure for basic probability assignment and its applications in extremely uncertain environments," *Int. J. Intell. Syst.*, to be published, doi: 10.1002/int.22066.
[8] L. Yin, X. Deng, and Y. Deng, "The negation of a basic probability assignment," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 1, pp. 135–143, Jan. 2018.
[9] F. Xiao, "Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy," *Inf. Fusion*, vol. 46, pp. 23–32, Mar. 2019.
[10] X. Deng and W. Jiang, "Dependence assessment in human reliability analysis using an evidential network approach extended by belief rules and uncertainty measures," *Ann. Nucl. Energy*, vol. 117, pp. 183–193, Jul. 2018.
[11] I. Ghafir, K. G. Kyriakopoulos, F. J. Aparicio-Navarro, S. Lambotharan, B. Assadhan, and H. Binsalleeh, "A basic probability assignment methodology for unsupervised wireless intrusion detection," *IEEE Access*, vol. 6, pp. 40008–40023, 2018.
[12] M. Li, Q. Zhang, and Y. Deng, "Evidential identification of influential nodes in network of networks," *Chaos, Solitons Fractals*, vol. 117, pp. 283–296, Dec. 2018.
[13] H. Yang, Y. Deng, and J. Jones, "Network division method based on cellular growth and physarum-inspired network adaptation," *Int. J. Unconventional Comput.*, vol. 13, no. 6, pp. 477–491, 2018.
[14] L. Yin and Y. Deng, "Toward uncertainty of weighted networks: An entropy-based model," *Phys. A, Stat. Mech. Appl.*, vol. 508, pp. 176–186, Oct. 2018.
[15] L. Fei, Q. Zhang, and Y. Deng, "Identifying influential nodes in complex networks based on the inverse-square law," *Phys. A, Stat. Mech. Appl.*, vol. 512, pp. 1044–1059, Dec. 2018.
[16] Y. Han and Y. Deng, "An enhanced fuzzy evidential DEMATEL method with its application to identify critical success factors," *Soft Comput.*, vol. 22, no. 15, pp. 5073–5090, Aug. 2018.
[17] Y. Wang, K. Zhang, and Y. Deng, "Base belief function: An efficient method of conflict management," *J. Ambient Intell. Humanized Comput.*, pp. 1–11, Oct. 2018, doi: 10.1007/s12652-018-0882-4.
[18] P. Liu and X. Zhang, "Approach to multi-attributes decision making with intuitionistic linguistic information based on dempster-shafer evidence theory," *IEEE Access*, vol. 6, pp. 52969–52981, 2018.
[19] L. Chen and X. Deng, "A modified method for evaluating sustainable transport solutions based on ahp and Dempster–Shafer evidence theory," *Appl. Sci.*, vol. 8, no. 4, Apr. 2018, Art. no. 563.
[20] F. Xiao, "A hybrid fuzzy soft sets decision making method in medical diagnosis," *IEEE Access*, vol. 6, pp. 25300–25312, 2018.
[21] J. H. Dahooie, E. K. Zavadskas, M. Abolhasani, A. Vanaki, and Z. Turskis, "A novel approach for evaluation of projects using an interval–valued fuzzy additive ratio assessment (ARAS) method: A case study of oil and gas well drilling projects," *Symmetry*, vol. 10, no. 2, p. 45, Jan. 2018.
[22] X. Deng and Y. Deng, "D-AHP method with different credibility of information," *Soft Comput.*, vol. 23, no. 2, pp. 683–691, Jan. 2019.
[23] L. Fei, Y. Deng, and Y. Hu, "DS-VIKOR: A new multi-criteria decision-making method for supplier selection," *Int. J. Fuzzy Syst.*, pp. 1–19, Sep. 2018, doi: 10.1007/s40815-018-0543-y.
[24] X. Zhou, Y. Hu, Y. Deng, F. T. S. Chan, and A. Ishizaka, "A DEMATEL-based completion method for incomplete pairwise comparison matrix in AHP," *Ann. Oper. Res.*, vol. 271, no. 2, pp. 1045–1066, Dec. 2018.
[25] Z. He and W. Jiang, "An evidential dynamical model to predict the interference effect of categorization on decision making results," *Knowl.-Based Syst.*, vol. 150, pp. 139–149, Jun. 2018.
[26] B. Kang, Y. Deng, K. Hewage, and R. Sadiq, "A method of measuring uncertainty for Z-number," *IEEE Trans. Fuzzy Syst.*, to be published, doi: 10.1109/TFUZZ.2018.2868496.2018.
[27] P. Dutta, "Modeling of variability and uncertainty in human health risk assessment," *MethodsX*, vol. 4, pp. 76–85, 2017.
[28] H. Zheng and Y. Deng, "Evaluation method based on fuzzy relations between Dempster–Shafer belief structure," *Int. J. Intell. Syst.*, vol. 33, no. 7, pp. 1343–1363, Jul. 2018.
[29] H. Mo and Y. Deng, "A new MADA methodology based on D numbers," *Int. J. Fuzzy Syst.*, vol. 20, no. 8, pp. 2458–2469, Dec. 2018.
[30] W. Deng and Y. Deng, "Entropic methodology for entanglement measures," *Phys. A, Stat. Mech. Appl.*, vol. 512, pp. 693–697, Dec. 2018.
[31] W. Jiang, "A correlation coefficient for belief functions," *Int. J. Approx. Reasoning*, vol. 103, pp. 94–106, Dec. 2018.
[32] W. Jiang, J. Zhan, D. Zhou, and X. Li, "A method to determine generalized basic probability assignment in the open world," *Math. Problems Eng.*, vol. 2016, Apr. 2016, Feb. 3878634.
[33] P. Smets and R. Kennes, "The transferable belief model," *Appl. Intell.*, vol. 66, no. 2, pp. 191–234, Apr. 1994.

[34] Y. Deng, "Generalized evidence theory," *Appl. Intell.*, vol. 43, no. 3, pp. 530–543, Oct. 2015.
[35] R. R. Yager, "Combining various types of belief structures," *Inf. Sci.*, vol. 303, pp. 83–100, May 2015.
[36] P. Dutta, "An uncertainty measure and fusion rule for conflict evidences of big data via Dempster–Shafer theory," *Int. J. Image Data Fusion*, vol. 9, no. 2, pp. 152–169, Oct. 2018.
[37] D. Palash and H. G. Hazarika, "Construction of families of probability boxes and corresponding membership functions at different fractiles," *Expert Syst.*, vol. 34, no. 3, p. e12202, Jun. 2017.
[38] K. Chatterjee, E. K. Zavadskas, J. Tamosaitiene, K. Adhikary, and S. Kar, "A hybrid MCDM technique for risk management in construction projects," *Symmetry*, vol. 10, no. 2, p. 46, Feb. 2018.
[39] K. Chatterjee, D. Pamucar, and E. K. Zavadskas, "Evaluating the performance of suppliers based on using the R'AMATEL-MAIRCA method for green supply chain implementation in electronics industry," *J. Cleaner Prod.*, vol. 184, pp. 101–129, May 2018.
[40] J. Abellán, "Analyzing properties of Deng entropy in the theory of evidence," *Chaos, Solitons Fractals*, vol. 95, pp. 195–199, Feb. 2017.
[41] J. Abellán and E. Bossé, "Drawbacks of uncertainty measures based on the pignistic transformation," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 3, pp. 382–388, Mar. 2018.
[42] L. Chen and Y. Deng, "A new failure mode and effects analysis model using Dempster–Shafer evidence theory and grey relational projection method," *Eng. Appl. Artif. Intell.*, vol. 76, pp. 13–20, Nov. 2018.
[43] X. Deng, W. Jiang, and Z. Wang, "Zero-sum polymatrix games with link uncertainty: A Dempster-Shafer theory solution," *Appl. Math. Comput.*, vol. 340, pp. 101–112, Jan. 2019.
[44] X. Deng, "Analyzing the monotonicity of belief interval based uncertainty measures in belief function theory," *Int. J. Intell. Syst.*, vol. 33, no. 9, pp. 1869–1879, Sep. 2018.
[45] F. Xiao, "A novel multi-criteria decision making method for assessing health-care waste treatment technologies based on D numbers," *Eng. Appl. Artif. Intell.*, vol. 71, no. 2018, pp. 216–225, May 2018.
[46] Z. Li and L. Chen, "A novel evidential FMEA method by integrating fuzzy belief structure and grey relational projection method," *Eng. Appl. Artif. Intell.*, vol. 77, pp. 136–147, Jan. 2019.
[47] J. Lin, D. Ye, C. Chen, and M. Gao, "Minimum spanning tree based spatial outlier mining and its applications," in *Proc. Int. Conf. Rough Sets Knowl. Technol.*, May 2008, pp. 508–515.
[48] Z. P. Hu, "A classification algorithm with reject option based on adaptive minimum spanning tree covering model in high-dimensional space," *J. Electron. Inf. Technol.*, vol. 32, no. 12, pp. 2895–2900, 2010.
[49] *Minimum Spanning Tree*. Accessed: Oct. 15, 2018. [Online]. Available: https://en.wikipedia.org/wiki/Minimum_spanning_tree
[50] R. A. Fisher and M. Marshall. *Iris Data Set*. Accessed: Oct. 24, 2018. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Iris
[51] M. E. A. Forina and S. Aeberhard. *Wine Data Set*. Accessed: Oct. 24, 2018. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Wine
[52] M. Bohanec, M. Bohanec, and B. Zupan. *Car Evaluation Data Set*. Accessed: Oct. 24, 2018. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Car+Evaluation

RENLIANG SUN is currently pursuing the master's degree with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, where he joined Prof. Deng's Lab, Institute of Fundamental and Frontier Science, in 2017. In 2018, he successfully hosted an innovative training program for college students. In summer 2018, he was with the Summer School, University of Electronic Science and Technology of China, where he was involved in a research on machine learning and graph theory. His current research interests involve evidence theory and machine learning. He has received the National Encouragement Scholarship, in 2017. He has received the WAC Scholarship, in 2018.

YONG DENG received the Ph.D. degree in precise instrumentation from Shanghai Jiao Tong University, Shanghai, China, in 2003. From 2005 to 2011, he was an Associate Professor with the Department of Instrument Science and Technology, Shanghai Jiao Tong University. Since 2010, he was a Professor with the School of Computer and Information Science, Southwest University, Chongqing, China. Since 2012, he has been a Visiting Professor with Vanderbilt University, Nashville, TN, USA. Since 2016, he has been a Professor with the School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China. Since 2017, he has been a Full Professor with the Institute of Fundamental and Frontier Science, University of Electronic Science and Technology of China, Chengdu, China. He has published over 100 papers in referred journals such as *Decision Support Systems*, the *European Journal of Operational Research*, and *Scientific Reports*. His research interests include evidence theory, decision making, information fusion, and complex system modeling. He has severed as a Program Member of many conferences such as the International Conference on Belief Functions. He has served as an Editorial Board Member such as an Academic Editor of PLOS ONE. He has served as a Reviewer for over 30 journals such as the IEEE TRANSACTION ON FUZZY SYSTEMS. He has received numerous honors and awards, including the Elsevier Highly Cited Scientist, in China, from 2014 to 2017.