Rough Set Classifier Based on DSmT

Yilin Dong, Xinde Li, Jean Dezert

Rough Set Classifier Based on DSmT

Yilin Dong
Key Laboratory of Measurement and Control of CSE
School of Automation
Southeast University
Email:dyl@seu.edu.cn

Xinde Li
Corresponding Author
Key Laboratory of Measurement and Control of CSE
School of Automation, Southeast University
Email: xindeli@seu.edu.cn

Jean Dezert
The French Aerospace Lab
ONERA
DTIM/MSDA
F-91123 Palaiseau, France.
Email:jean.dezert@onera.fr

***Abstract*—The classifier based on rough sets is widely used in pattern recognition. However, in the implementation of rough set-based classifiers, there always exist the problems of uncertainty. Generally, information decision table in Rough Set Theory (RST) always contains many attributes, and the classification performance of each attribute is different. It is necessary to determine which attribute needs to be used according to the specific problem. In RST, such problem is regarded as attribute reduction problems which aims to select proper *candidates*. Therefore, the uncertainty problem occurs for the classification caused by the choice of attributes. In addition, the voting strategy is usually adopted to determine the category of target concept in the final decision making. However, some classes of targets cannot be determined when multiple categories cannot be easily distinguished (for example, the number of votes of different classes is the same). Thus, the uncertainty occurs for the classification caused by the choice of classes. In this paper, we use the theory of belief functions to solve two above mentioned uncertainties in rough set classification and rough set classifier based on Dezert-Smarandache Theory (DSmT) is proposed. It can be experimentally verified that our proposed approach can deal efficiently with the uncertainty in rough set classifiers.**

Keywords: Classification, Rough Set, Uncertainty, Evidence Reasoning, DSmT, Belief Functions.

I. Introduction

a) Motivation: In recent years, we have witnessed the rapid development of Rough Set Theory (RST) [1]. There are many practical applications of this theory [2],[3],[4],[5]. Among these, Rough Set Classifier (RSC) has been widely used in the real classification problems [6], [7], [8], [9].

b) Challenges: However, in the practical use of RSC, there always exists uncertainty. In the literature [10] and [11], the discussions of the uncertainty in RST mainly focus on the following points of view: Chen [10] proposed several uncertainty measures of neighborhood granules, which are neighborhood accuracy, information quantity, neighborhood entropy and information granularity in the neighborhood RST; Zheng [11] estimated the uncertainty of rough set originated from two parts of boundary region. Although the uncertainties discussed in the above literature are of certain significance, however, the uncertainties discussed in this paper are shown in two aspects:

1) The choice of attributes: for example, in the decision information table, some attributes are not significant in a representation and deleting of these attributes has no real impact on the classification results. However, such concept of *significancy* is relative, for different problems, the role of each attribute is quite different. Thus, the problems of attribute selection are always ad hoc and depending on the user's preference. Obviously, different attribute selections correspond to different strategies, which generally yield different results. For example, in [12], authors attempted to select the most information-rich attributes from a dataset by incorporating a controlled degree of misclassification into approximations of rough sets. Gao et.al [13] proposed a new uncertainty measure, named maximum decision entropy, for attribute reduction in the decision-theoretic rough set model. Although many robust and efficient reduction algorithms have been proposed, most of them concentrate on the properties of data or user preference in the definition of attribute reduction, which result in the difficulties of choosing appropriate attribute reductions for specific applications. For the same data, different users can define different reductions and obtain their interested results according to their applications. Jia et.al [14] reviewed nearly twenty two different attribute reduction methods, but to design of a robust attribute reduction method is not the focus of this paper. We emphasize the uncertainty caused by the choice of attributes, which is not discussed in details in the recent development of RST. For this aim, one typically seeks a policy for avoiding choosing attributes, and we propose to emphasize the importance of each attribute for the specific problems.

2) The choice of classes: besides, in RST, the category of target concept is determined according to the element composition of its corresponding approximate set: if the number of elements belonging to one class is the largest, the concept of target is labelled as this class. However, this kind of voting method often leads to uncertainty in making decisions, which affects the final precision of RSC. In order to illustrate this problem more vividly, we explain it through Figure 1: in case one, the approximate set of target concept (**red five-poited star**) has four elements (**plus**) belonging to class 1, three elements (**plus**) belonging to class 2 and two elements (**plus**) belonging to class 3. Thus, in case one, we can easily draw the conclusion that the target belongs to class 1. However, in case two or case three, the target cannot be labelled with single category because there are some classes (class 1 and class 2 in case two, class 1, class 2 and class 3 in case three) that have the same number of votes. More specifically, if

the approximate set of such target is empty-set (case four), which category should be allocated to the target concept? As aforementioned, for the RSCs, there are two mentioned neglected uncertainty issues. The theory of belief functions [15] is widely used in uncertainty management and uncertainty reasoning for decision-making. In this paper, we attempt to use it to model and manage the uncertainty incorporated in RSCs.

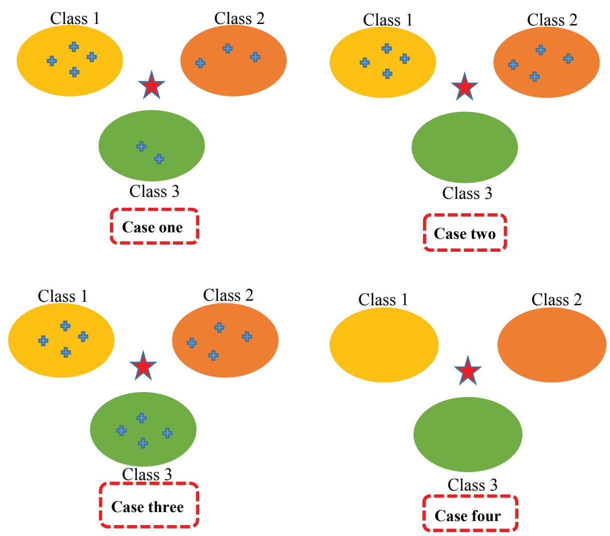

Fig. 1: **Uncertainty in Voting Strategy**.

c) Contributions: Because a certain attribute does not have the ability to distinguish items on a particular problem, but there may be a discriminative performance on another problem. Thus, according to the classification performance of each attribute, the corresponding weights of all attributes in information decision table are calculated, which are used as the evaluation index of the importance of an attribute. At the same time, we do not directly delete unimportant attributes which the classical reduction algorithms have done. We just consider all the attributes in the final classification, after all, we consider that all existing attributes must play a role in the decision. For the uncertainty of the voting strategy in traditional RSC, we have no statistics of the number of votes of each class in approximate sets. Instead, we first calculate the coordinate of each class with respect to each attribute and then get the distance between the target concept and each class in every attribute, in order to calculate the Basic Belief Assignment (BBA) of the target in each attribute. Then, we use the classical combination rule (PCR5 is used in this paper) proposed in DSmT [16] to sequentially [1] combine all BBAs (each attribute has a corresponding BBA). Finally, according to the principle of maximum belief mass, we can obtain the final class of the target concept.

[1]Because PCR5 rule is not associative, which means that the fusion results depend on the order you have chosen. Here, our default way of combination is to combine BBAs in order from small to large. For example, if there are three BBAs: m_1, m_2, m_3, the way of fusion is $m_{12} = PCR5(m_1, m_2) \rightarrow m_{123} = PCR5(m_{12}, m_3) \rightarrow m_{fusion} = m_{123}$.

This paper is organized as follows. Section 2 reviews some basic concepts of Dempster-Shafer Theory (DST), and DSmT. The new rough set classifier based on DSmT (RSCD) is proposed in section 3. Section 4 gives the summary of the proposed classifier. In section 5, we give some experimental results to show the performances of our new method. Also, some meaningful discussions about the extension of RSCD are given in section 6. Section 7 concludes the paper with a summary and direction for future.

II. Preliminaries

This section provides a brief reminder of the basics of DST and DSmT, which is necessary for the presentation and understanding of the more general fusion of evidence.

In DST framework, the Frame of Discernment (FoD)[2] $\Theta \triangleq \{\theta_1, \ldots, \theta_n\}$ ($n \geq 2$) is a set of exhaustive and exclusive elements (hypotheses) which represent the possible solutions of the problem under consideration and thus Shafer's model assumes $\theta_i \cap \theta_j = \emptyset$ for $i \neq j$ in $\{1, \ldots, n\}$. A BBA $m(\cdot)$ is defined by the mapping: $2^\Theta \mapsto [0, 1]$, verifying $m(\emptyset) = 0$ and $\sum_{A \in 2^\Theta} m(A) = 1$. In DSmT, one can abandon Shafer's model (if Shafer's model doesn't fit with the problem) and refute the principle of the third excluded middle. The third excluded middle principle assumes the existence of the complement for any elements/propositions belonging to the power set 2^Θ. Instead of defining the BBAs on the power set $2^\Theta \triangleq (\Theta, \cup)$ of the FoD, the BBAs are defined on the so-called *hyper-power set* (or Dedekind's lattice) denoted $D^\Theta \triangleq (\Theta, \cup, \cap)$ whose cardinalities follows Dedekind's numbers sequence, see [17], Vol.1 for details and examples. A (generalized) BBA, called a mass function, $m(\cdot)$ is defined by the mapping: $D^\Theta \mapsto [0, 1]$, verifying $m(\emptyset) = 0$ and $\sum_{A \in D^\Theta} m(A) = 1$. The DSmT framework encompasses DST framework because $2^\Theta \subset D^\Theta$. In DSmT, we can take into account also a set of *integrity constraints* on the FoD (if known), by specifying all the pairs of elements which are really disjoint. Stated otherwise, Shafer's model is a specific DSm model where all elements are deemed to be disjoint. $A \in D^\Theta$ is called a focal element of $m(.)$ if $m(A) > 0$. A BBA is called a Bayesian BBA if all of its focal elements are singletons and Shafer's model is assumed, otherwise it is called non-Bayesian [18]. A full ignorance source is represented by the vacuous BBA $m_v(\Theta) = 1$. The belief (or credibility) and plausibility functions are respectively defined by $Bel(X) \triangleq \sum_{Y \in D^\Theta | Y \subseteq X} m(Y)$ and $Pl(X) \triangleq \sum_{Y \in D^\Theta | Y \cap X \neq \emptyset} m(Y)$. $BI(X) \triangleq [Bel(X), Pl(X)]$ is called the belief interval of X. Its length $U(X) \triangleq Pl(X) - Bel(X)$ measures the degree of uncertainty of X.

In 1976, Shafer did propose Dempster's rule and we use DS index to refer to Dempster-Shafer's rule (DS rule) because Shafer did really promote Dempster's rule in in his milestone book [18]) to combine BBAs in DST framework. DS rule for combining two distinct sources of evidence characterized

[2]Here, we use the symbol $\triangleq$ to mean equals by definition.

by BBAs $m_1(\cdot)$ and $m_2(\cdot)$ is defined by $m_{DS}(\emptyset) = 0$ and $\forall A \in 2^\Theta \setminus \{\emptyset\}$:

$$m_{DS}(A) = \frac{\sum_{B,C\in 2^\Theta | B\cap C = A} m_1(B)m_2(C)}{1 - \sum_{B,C\in 2^\Theta | B\cap C=\emptyset} m_1(B)m_2(C)} \quad (1)$$

The DS rule formula is commutative and associative and can be easily extended to the fusion of $S > 2$ BBAs. Unfortunately, DS rule has been highly disputed during the last decades by many authors because of its counter-intuitive behavior in high or even low conflict situations, and that is why many rules of combination were proposed in literature to combine BBAs [19]. To palliate DS rule drawbacks, the very interesting PCR5 was proposed in DSmT and it is usually adopted [3] in recent applications of DSmT. The fusion of two BBAs $m_1(.)$ and $m_2(.)$ by the PCR5 rule is obtained by $m_{PCR5}(\emptyset) = 0$ and $\forall A \in D^\Theta \setminus \{\emptyset\}$

$$m_{PCR5}(A) = m_{12}(A) + \sum_{B\in D^\Theta \setminus\{A\} | A\cap B=\emptyset} [\frac{m_1(A)^2 m_2(B)}{m_1(A)+m_2(B)} + \frac{m_2(A)^2 m_1(B)}{m_2(A)+m_1(B)}] \quad (2)$$

where $m_{12}(A) = \sum_{B,C\in D^\Theta | B\cap C=A} m_1(B)m_2(C)$ is the conjunctive operator, and each element A and B are expressed in their disjunctive normal form. If the denominator involved in the fraction is zero, then this fraction is discarded. The general PCR5 formula for combining more than two BBAs altogether is given in [17], Vol. 3. We adopt the generic notation $m_{12}^{PCR5}(.) = PCR5(m_1(.), m_2(.))$ to denote the fusion of $m_1(.)$ and $m_2(.)$ by PCR5 rule. PCR5 is not associative and PCR5 rule can also be applied in DST framework (with Shafer's model of FoD) by replacing D^Θ by 2^Θ in Eq (2).

III. New Rough Set Classifier Based on DSmT (RSCD)

A. Weights of each attribute

RST is a mathematical tool to deal with vagueness and uncertainty [1], which can effectively analyse the incomplete information and does not need additional data beyond the prior information. Next, we briefly give several relevant definitions to show how to calculate the weights of attributes:

Definition 1: An information decision system $\mathbf{S}$ is $\mathbf{S} = (U, A, D)$, where $U = \{x_1, x_2, \cdots, x_n\}$ is non-empty finite set of samples, $A = \{a_1, a_2, \cdots, a_m\}$ is a non-empty finite set of attributes, D is a non-empty set of finite decision classes.

Definition 2: Each attribute $a \in A$ defines an information function $f_a : U \to V_a$, and V_a is the set value of the attribute a. We further extend these notations for a set of attributes $B \subseteq A$, an indiscernibility relation $Ind(B)$ can be defined as follows:

$$Ind(B) = \{(\mathbf{x_i}, \mathbf{x_j}) \in U^2 | f_i(a) = f_j(a), \forall a \in B\} \quad (3)$$

[3]Recently, a new combination rule PCR6 was proposed to combine all the BBAs altogether in a single fusion step, which can be found in [20]. Because PCR6 rule coincides with PCR5 when combining only two BBAs [17], we just use PCR5 rule to combine BBAs in this paper.

where $\mathbf{x_i}$ and $\mathbf{x_j}$ are indiscernible when $(\mathbf{x_i}, \mathbf{x_j}) \in Ind(B)$. Some equivalence classes or elementary sets are generated by $Ind(B)$. The elementary set of $\mathbf{x_i}$ is represented by $[\mathbf{x_i}]_B$. Any finite union of elementary sets is called a B-definable set. For pattern classification, elements have the same class label consisting of a concept X so that $X \in U/D$, where $U/D = \{[\mathbf{x_i}]_D | \mathbf{x_i} \in U\}$ and $[\mathbf{x_i}]_D$ represents the elementary sets of $\mathbf{x_i}$ with respect to decision attribute D. Sometimes $X \subseteq U$ is not B-definable. In other words, there exists elements that are in the same elementary set, but have different class labels, so that X becomes a vague concept. For this, we give the following definitions of approximation sets of such vague concept:

Definition 3: The B-upper approximation $\overline{B}X$ and the B-lower approximation $\underline{B}X$ of the vague concept X is defined as follows:

$$\underline{B}X = \{\mathbf{x_i} \in U | [\mathbf{x_i}]_B \subseteq X\}, \quad (4)$$

$$\overline{B}X = \{\mathbf{x_i} \in U | [\mathbf{x_i}]_B \cap X \neq \emptyset\}. \quad (5)$$

$\underline{B}X \subseteq \overline{B}X$, and $\underline{B}X$ consists of elements that certainly belong to X, whereas $\overline{B}X$ consists of elements that possibly belong to X. The set $BN_B(X) = \overline{B}X - \underline{B}X$ is called the B-*boundary region* of X, and thus consists of those objects that we cannot decisively classify into X on the basis of knowledge in B.

Definition 4: $POS_B(D)$ is a positive region of the partition U/D with respect to B and is defined as follows:

$$POS_B(D) = \bigcup_{X\in U/D} \underline{B}X \quad (6)$$

$$= \bigcup \{Y | Y \subseteq X, Y \in U/B, X \in U/D\}. \quad (7)$$

Definition 5: The degree of support of the condition attributes B with respect to the decision attribute D is defined as follows:

$$\zeta_B^D = \frac{|POS_B(D)|}{|U|}. \quad (8)$$

Here, ζ is regarded as the degree of importance of each attribute in the information decision table $\mathbf{S}$. In order to illustrate how to calculate the weight of a particular attribute based on the aforementioned five definitions, we give a simple example below:

Example 1: Table I is an information decision table with $U = \{x_1, x_2, \cdots, x_{12}\}$, $A = \{a_1, a_2, a_3, a_4\}$, $D = \{d_1 = 1, d_2 = 2, d_3 = 3\}$. According to the decision attribute d and Eq.(3), if $\mathbf{x_i}$ is set to U and B is equal to d, we can get the $[\mathbf{x_i}]_B = [U]_d = U/D = \{\{x_1, x_4, x_7, x_8, x_{12}\}, \{x_2, x_3, x_9, x_{10}, x_{11}\}, \{x_5, x_6\}\}$. Meanwhile, we can also partition U by using each attribute $a_i, i = 1, \cdots, m$ based on the indiscernibility relation $Ind(B)$, which are illustrated in Table II. Thus, each element X in $[U]_d$ can be approximated by each condition attribute $a_i, i = 1, \cdots, m$, and then we can obtain $\underline{a_i}X$ in Table III according to Definition 3. Based on Eq.(7), we can get the positive domain of D with respect to each attribute a_i, which is also given in Table IV. In order to explain how positive domains are calculated in detail, we take $POS_{a_1}(D)$ as an example: $U/D = [U]_D = \{\{x_1, x_4, x_7, x_8, x_{12}\}, \{x_2, x_3, x_9, x_{10}, x_{11}\}, \{x_5, x_6\}\}$, $U/a_1 = [U]_{a_1} = \{\{x_1, x_4\}, \{x_2\}, \{x_3\}, \{x_5\}, \{x_6\}, \{x_7\}, \{x_8, x_9\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}\}$, for any elements Y, where $Y \in U/a_1$, if Y meets the condition: $Y \subseteq X$, where $X \in U/D$, then Y belongs to the domain $POS_{a_1}(D)$, for example, when $Y = \{x_1, x_4\}$ and $X = \{x_1, x_4, x_7, x_8, x_{12}\}$, it satisfies $Y \subseteq X$, so $\{x_1, x_4\}$ belongs to $POS_{a_1}(D)$. However, if $Y = \{x_8, x_9\}$, Y is not a subset of any elements in U/D, so $\{x_8, x_9\}$ does not belong to $POS_{a_1}(D)$. Thus, according to Eq.(8), we can obtain the degree of support of a_i with respect to the decision attribute D in Table IV, which will be regarded as the weights of each attribute in the classification problem.

TABLE I: Information decision table.

U	a_1	a_2	a_3	a_4	d
x_1	5.1	3.5	1.4	0.2	1
x_2	6.6	2.9	4.6	1.3	2
x_3	5.2	2.7	3.9	1.4	2
x_4	5.1	3.8	1.5	0.3	1
x_5	6.4	2.7	5.3	1.9	3
x_6	6.8	3.0	5.5	2.1	3
x_7	5.5	4.2	1.4	0.2	1
x_8	5.0	3.3	1.4	0.2	1
x_9	5.0	2.0	3.5	1.0	2
x_{10}	5.9	3.0	4.2	1.5	2
x_{11}	5.7	2.6	3.5	1.0	2
x_{12}	4.6	3.6	1.0	0.2	1

TABLE II: Results of partitioning the domain U using each attribute.

	The partitioning domain
$[U]_{a_1}$	$\{\{x_1, x_4\}, \{x_2\}, \{x_3\}, \{x_5\}, \{x_6\}, \{x_7\}$ $\{x_8, x_9\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}\}$
$[U]_{a_2}$	$\{\{x_1\}, \{x_2\}, \{x_3, x_5\}, \{x_4\}, \{x_6, x_{10}\}$ $\{x_7\}, \{x_8\}, \{x_9\}, \{x_{11}\}, \{x_{12}\}\}$
$[U]_{a_3}$	$\{\{x_1, x_7, x_8\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}$ $\{x_6\}, \{x_9, x_{11}\}, \{x_{10}\}, \{x_{12}\}\}$
$[U]_{a_4}$	$\{\{x_1, x_7, x_8, x_{12}\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}$ $\{x_6\}, \{x_9, x_{11}\}, \{x_{10}\}\}$

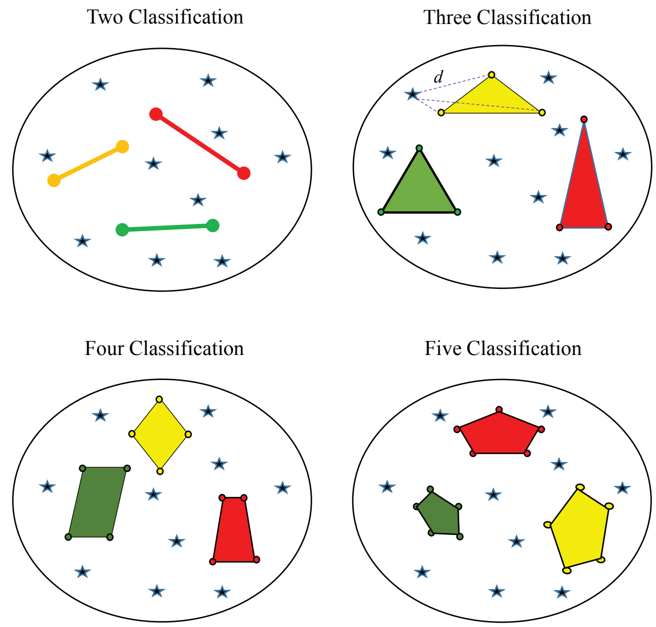

Fig. 2: **Attribute polygon**. Pentagram represents the test example and for example, in three classification, the distances (dotted line) are calculated between the value of one attribute of pentagram and the vertices of one attribute triangle.

B. Construction of BBA of Target Concept

As discussed in the introduction section, the traditional way of voting decision will cause uncertainty when using RSC, and directly affect the final classification accuracy. The evidence theory has a good ability to deal with the uncertainty problem, and evidence theory generally describes such concept of uncertainty through BBAs. However, the BBAs in evidence theory are always given by experts depending on their own experience, which cannot be obtained directly in practical problems. Thus, this requires that, when solving such problems, the corresponding BBAs are first constructed and calculated before using them to make decisions. Referring to the construction methods of BBAs in [21], [22], [23], we propose in this paper a new construction method for the BBA based on so-called attribute polygon in RST. Each polygon represents an attribute and each vertice in a polygon represents one category. That is to say, if it is a two-classification problem, the attribute polygon is the line segment; Similarly, if it is the three-classification problem, such polygon is the triangle, and so on. Figure 2 illustrates the corresponding four polygons which represent for two, three, four and five classification problems. Besides, the coordinates of all vertices in all attribute polygons are calculated according to $[U]_d = \{\{x_1, x_4, x_7, x_8, x_{12}\}, \{x_2, x_3, x_9, x_{10}, x_{11}\}, \{x_5, x_6\}\}$. Then, the Euclidean distance is used to calculate the distance between test example and each attribute polygon. Finally, we can get the belief mass value of this example belonging to each class with respect to one attribute by using Eq.(9) and Eq.(10).

$$m_{a_i}^{x^*}(\theta_s) = \alpha e^{\gamma_s d^\beta}. \tag{9}$$

$$m_{a_i}^{x^*}(\Theta) = 1 - \alpha e^{\gamma_s d^\beta}. \tag{10}$$

where α, γ_s and β are turning parameters and according to the recommendations given in [24], these parameters are set to $\alpha = 0.95, \gamma_s = -2$ and $\beta = 1$. Besides, d is the distance between the vertices of a_i attribute polygon and each attribute value of text example x^*. Next, we will show how to calculate

TABLE III: The lower approximation of elements in $[U]_d$ using each attribute.

$\underline{B}X$	The B-lower approximation
$\underline{a_1}\{x_1, x_4, x_7, x_8, x_{12}\}$	$\{\{x_1, x_4\}, \{x_7\}, \{x_{12}\}\}$
$\underline{a_1}\{x_2, x_3, x_9, x_{10}, x_{11}\}$	$\{\{x_2\}, \{x_3\}, \{x_{10}\}, \{x_{11}\}\}$
$\underline{a_1}\{x_5, x_6\}$	$\{\{x_5\}, \{x_6\}\}$
$\underline{a_2}\{x_1, x_4, x_7, x_8, x_{12}\}$	$\{\{x_1\}, \{x_4\}, \{x_7\}, \{x_8\}, \{x_{12}\}\}$
$\underline{a_2}\{x_2, x_3, x_9, x_{10}, x_{11}\}$	$\{\{x_2\}, \{x_9\}, \{x_{11}\}\}$
$\underline{a_2}\{x_5, x_6\}$	$\emptyset$
$\underline{a_3}\{x_1, x_4, x_7, x_8, x_{12}\}$	$\{\{x_1, x_7, x_8\}, \{x_4\}, \{x_{12}\}\}$
$\underline{a_3}\{x_2, x_3, x_9, x_{10}, x_{11}\}$	$\{\{x_2\}, \{x_3\}, \{x_9, x_{11}\}, \{x_{10}\}\}$
$\underline{a_3}\{x_5, x_6\}$	$\{\{x_5\}, \{x_6\}\}$
$\underline{a_4}\{x_1, x_4, x_7, x_8, x_{12}\}$	$\{\{x_1, x_7, x_8, x_{12}\}, \{x_4\}\}$
$\underline{a_4}\{x_2, x_3, x_9, x_{10}, x_{11}\}$	$\{\{x_2\}, \{x_3\}, \{x_9, x_{11}\}, \{x_{10}\}\}$
$\underline{a_4}\{x_5, x_6\}$	$\{\{x_5\}, \{x_6\}\}$

TABLE IV: The positive domain of $[U]_d$ with respect to each attribute and weights of each attribute according to Eq.(8).

	Domain	ζ
$POS_{a_1}(D)$	$\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_{10}, x_{11}, x_{12}\}$	$\frac{10}{12}$
$POS_{a_2}(D)$	$\{x_1, x_2, x_4, x_7, x_8, x_9, x_{11}, x_{12}\}$	$\frac{8}{12}$
$POS_{a_3}(D)$	$\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\}$	$\frac{12}{12}$
$POS_{a_4}(D)$	$\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\}$	$\frac{12}{12}$

BBAs through Example 1.

Example 1 revisited:

According to the decision attribute d in Table I, we know that this simple example is a three-class problem because $D = \{d_1, d_2, d_3\}$, so we need to construct the triangles. Because the decision table has four condition attributes, we need to construct four triangles. In order to show how to calculate the coordinates of vertices in each attribute triangle, we give the calculation steps as follows: Based on the partitions of the decision attribute d:$\{\{x_1, x_4, x_7, x_8, x_{12}\}, \{x_2, x_3, x_9, x_{10}, x_{11}\}, \{x_5, x_6\}\}$, we can obtain the coordinates of each category with respect to attribute a_1:

- the coordinate of class one with respect to a_1:

$$\frac{1}{|X(a_1)|} \sum_{x \in X(a_1)} f(x, a_1) = 5.06$$

  where $X(a_1) = \{x_1, x_4, x_7, x_8, x_{12}\}$ and $|\cdot|$ denotes the cardinality;
- the coordinate of class two with respect to a_1:

$$\frac{1}{|X(a_1)|} \sum_{x \in X(a_1)} f(x, a_1) = 5.6800$$

  where $X(a_1) = \{x_2, x_3, x_9, x_{10}, x_{11}\}$;
- the coordinate of class three with respect to a_1:

$$\frac{1}{|X(a_1)|} \sum_{x \in X(a_1)} f(x, a_1) = 6.6000$$

  where $X(a_1) = \{x_5, x_6\}$.

where $f(x_i, a_j)$ is the value of the cell of the Table I corresponding to value x_i and attribute a_j.

TABLE V: All coordinates of three classes in each attribute.

Attribute	Class 1	Class 2	Class 3
a_1	5.0600	5.6800	6.600
a_2	3.6800	2.6400	2.8500
a_3	1.3400	3.9400	5.4000
a_4	0.2200	1.2400	2.0000

Similarly, we can calculate all the coordinates of three classes of four attributes, which is given in Table V as follows. Then, we randomly select a test example, which is denoted as $x^* = \{5.1000, 3.5000, 1.4000, 0.2000\}$. Based on the Euclidean distance [4], the corresponding distances between x^* and each attribute polygon is given in Table VI. Based on

[4]The Euclidean distance $d_{ij} = d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j)}$ is used here.

TABLE VI: Distances between target x^* and all vertices of attribute polygons.

Distance	Class 1	Class 2	Class 3
$a_1 \leftrightarrow x^*$	0.0400	0.5800	1.5000
$a_2 \leftrightarrow x^*$	0.1800	0.8600	0.6500
$a_3 \leftrightarrow x^*$	0.0600	2.5400	4.0000
$a_4 \leftrightarrow x^*$	0.0200	1.0400	1.8000

TABLE VII: BBAs of x^* with respect to each attribute.

$m(\cdot)$	Class 1	Class 2	Class 3	Θ
$m_1(\cdot)$	0.7778	0.0523	0.0005	0.1694
$m_2(\cdot)$	0.3862	0.0129	0.0368	0.5641
$m_3(\cdot)$	0.7038	0.0000	0.0000	0.2962
$m_4(\cdot)$	0.8596	0.0052	0.0043	0.1309

Eq.(9) and Eq.(10), we can transform these values of distances into belief mass so as to obtain the BBAs of each attribute, which is given in Table VII. Finally, we use PCR5 formula Eq.(2) to combine the weight of each attribute and the BBAs of each attribute so as to obtain the final BBA of x^{*}[5]. According to the fusion result, we can draw a conclusion that x^* belong to class 1 based on maximum of belief mass principle, which is consistent with the label of x^* in the original dataset.

$$m_{fusion}(\theta_1) = \mathbf{0.8827}; m_{fusion}(\theta_2) = 0.0009;$$
$$m_{fusion}(\theta_3) = 0.0007; m_{fusion}(\Theta) = 0.1157;$$

IV. The Summary of RSCD

Here, we give a brief pseudo code of RSCD in Algorithm 1. Because RSCD in this paper is a data-driven model, so, first of all, we need to divide original dataset into training datasets and test samples (the experiments in this paper are using ten-fold cross validation). Afterwards, the training datasets are applied to construct attribute polygons and calculate the weights of attributes. Finally, we can obtain the corresponding BBAs of each test samples by calculating the distances between test examples and attribute polygons.

V. Simulations

We have tested the different classifiers on real datasets given in the machine learning repository of the University of California Irvine (UCI) [25] and listed in Table VIII.

In our tests, we do not deal with the missing data problem, all the samples with missing values have been eliminated. Features of the samples are normalized by their means and standard deviations before their classification. As with the artificial datasets, we have evaluated the nearest neighbor (NN) classifier, the nearest class centroid (NC) classifier, two k-NN classifiers (one is with big k (k=40) and the other with

[5] In the final BBA, for the sake of convenience, $\theta_1, \theta_2, \theta_3$ and Θ represent class 1, class 2, class 3 and unknown; And $m_{fusion}(\cdot) = [(m_1(\cdot) \oplus m_2(\cdot)) \oplus m_3(\cdot)] \oplus m_4(\cdot)$, where $\oplus$ denotes PCR5 rule.

TABLE VIII: UCI datasets used in the experiments.

Datasets	Class Num.	Feature Dimention	Sample Num.
Iris	3	4	150
Wine	3	13	178
Pima	2	8	768
Bupa	2	6	345
Ionosphere	2	34	351

Algorithm 1 Solving classification problem by RSCD

Input: $Dataset$, $\alpha = 0.95, \gamma_s = -2$ and $\beta = 1$.
Output: $The\ final\ BBA\ of\ test\ data$: $m_{fusion}(\cdot)$.

1) Calculate the weights of attributes w_B, by
$$POS_B(D) = \bigcup_{X \in U/D} \underline{B}X;\ w_B = \zeta_B^D = \frac{|POS_B(D)|}{|U|}.$$
2) Calculate the BBA of each attribute, by
$$m_{a_i}^{x^*}(\theta_s) = \alpha e^{\gamma_s d^\beta};\ m_{a_i}^{x^*}(\Theta) = 1 - \alpha e^{\gamma_s d^\beta}.$$
3) Combine all BBAs of attributes sequentially, by
$$\begin{aligned}
&m_{fusion}(\cdot) = 1;\\
&i = 1;\\
&while\ \ i <= m:\\
&\qquad m_{fusion}(\cdot) = PCR5(m_{fusion}(\cdot), w_i \cdot m_i(\cdot));\\
&\qquad Normalization(m_{fusion}(\cdot)).\\
&end
\end{aligned}$$

a small k (k=5)), and the ER-NN-NC classifier (both with DS+BetP option, and with PCR5+DSmP option) [26]. The results are listed in Table IX. As we can see in Table IX, RSCD performs better in three datasets (Iris, Pima and Bupa) and the classification results are close to ER-NN-NC on the other two datasets (Wine and Ionosphere).

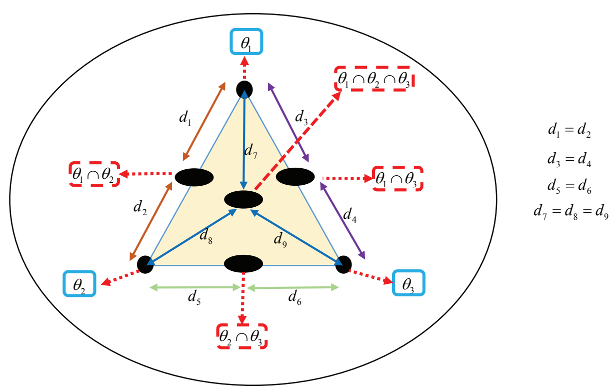

Fig. 3: **The principle of expanded attribute polygon**.

VI. Discussions

In this paper, the Frame of Discernment (FoD) is $\Theta = \{\theta_1, \theta_2, \cdots, \theta_n\}$ where θ_i represents the category and here we

TABLE IX: Classification accuracies of UCI datasets.

Classifiers	Iris(%)	Wine(%)	Pima(%)	Bupa(%)	Ionosphere(%).
NN	93.84	94.76	69.04	60.46	84.41
$NN(Center)$	92.09	95.68	72.70	56.54	79.25
$ER-NN-NC(DSmT+DSmP)$	95.15	**96.42**	73.38	60.96	**87.76**
$k-NN(k=40)$	89.43	95.28	71.60	61.99	67.63
$k-NN(k=5)$	95.65	95.28	72.10	59.57	82.41
$RSCD$	**98.00**	94.17	**74.50**	**62.87**	84.11

just consider singletons without compound focal elements[6]. Actually, some examples are difficult to be divided into a single class, and it may be possible to belong to two categories or several categories at the same time. On the basis of constructing attribute polygons in this paper, we can easily expand the mentioned principle above to more complex circumstances so as to ensure the particular target can belong to several classes simultaneously. The principle is illustrated in Figure 3: In this figure, we give a brief description of the expanded principle by using the three classification problem (triangle). In this triangle, three vertices (light blue and solid frame) represent single class, which is denoted by θ_1, θ_2 and θ_3. The difference is that, the centers of the three edges of such triangle and the center of gravity of this triangle are defined as compound focal elements, respectively. Specifically, the center of θ_1 and θ_2 is denoted as $\theta_1 \cap \theta_2$, in turn, we can define all the centers of all edges of this triangle. Besides, the center of gravity of this triangle is defined as $\theta_1 \cap \theta_2 \cap \theta_3$. Then, we can calculate all the coordinates of these centers and also the corresponding distances so as to obtain the BBAs of all attributes. To illustrate the principle of the expanded attribute polygon, we again revisit **Example 1** as follows: Since the extension method is mainly aimed at constructing BBAs, there is no impact on the calculation of attribute weights, so the following steps are only for BBAs calculation.

- Step 1: Calculate all relevant points in expanded polygon which are given in Table X. In Table X, θ_1, θ_2 and θ_3 represent Class 1, Class 2 and Class 3. $\theta_1 \cap \theta_2$ corresponds to the hypothesis for which the target belongs to two categories simultaneously, and so on. The coordinates of $\theta_1 \cap \theta_2$ and $\theta_1 \cap \theta_2 \cap \theta_3$ are calculated as for example by:

$$a_1(\theta_1 \cap \theta_2) = \frac{a_1(\theta_1) + a_1(\theta_2)}{2} = 5.37.$$

$$a_1(\theta_1 \cap \theta_2 \cap \theta_3) = \frac{a_1(\theta_1) + a_1(\theta_2) + a_1(\theta_3)}{3} = 5.78.$$

- Step 2: Based on Euclidean distance, we can obtain the corresponding distances between the target concept x^* and all relevant points in expanded polygon, which is given in Table XI.
- Step 3: According to Eq.(9) and Eq.(10), BBAs of x^* with respect to each attribute are shown in Table XII.

[6]Here, we do not regard Θ in Eq.(10) as a compound focal element even though Θ can be defined as $\Theta = \theta_1 \cup \theta_2 \cup \cdots \cup \theta_n$. Because Θ represents the ignorance or unknown of category of target concept, however, compound focal elements here mean that this target belongs to two categories or three categories at the same time.

- Step 4: Sequentially combine all four BBAs with PCR5 rule and then, we can get the final BBA as follows.

$$m_{fusion}(\theta_1) = \mathbf{0.8763}; m_{fusion}(\theta_2) = 0.0001;$$
$$m_{fusion}(\theta_3) = 0.0000; m_{fusion}(\theta_1 \cap \theta_2) = 0.0240;$$
$$m_{fusion}(\theta_2 \cap \theta_3) = 0.0000; m_{fusion}(\theta_1 \cap \theta_3) = 0.0006;$$
$$m_{fusion}(\theta_1 \cap \theta_2 \cap \theta_3) = 0.0004; m_{fusion}(\Theta) = 0.0985.$$

Thus, we can also get the result that x^* belongs to Class 1 (θ_1). The biggest difference between the extension method and the RSCD is that the possible category of target is further divided so as to reduce the uncertainty in classification problem, which can be embodied in $m_2(\cdot)$ in Table VII and Table XII. In RSCD, the assignment of x^* to Θ with respect to a_2 is 0.5640 (see the BBA $m_2(\cdot)$ of Table VII), which means the class of x^* cannot be determined if the principle of maximum belief mass is applied. However, in expanded strategy, Θ is further divided into $\theta_1 \cap \theta_2$, $\theta_2 \cap \theta_3$, $\theta_1 \cap \theta_3$, $\theta_1 \cap \theta_2 \cap \theta_3$, which ensure the target can be labelled with the correct class.

VII. Conclusion

In this paper, a new rough set classifier based on DSmT has been proposed to manage uncertainties using belief function theory. Our simulation results show clearly that RSCD performs well and its implementation is relatively simple since the attribute reduction in traditional rough set is avoided. In the implementation of RSCD, different types of combination rules can be used which give some flexibility to the users. In this paper, only one combination rule in DSmT (PCR5) has been tested. Of course many more could be implemented and tested , especially globally combing all BBAs in a single fusion step with PCR6 rule, which is left for future investigations. Also, The way of the attribute weights and BBAs' calculation used in RSCD is an open question and we plan to make investigations on this question, and evaluate the robustness of RSCD in future research works.

VIII. Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 61573097 and 91748106, in part by Key Laboratory of Integrated Automation of Process Industry (PAL-N201704), in part by the Fundamental Research Funds for the Central Universities (3208008401), in part by the Qing Lan Project and Six Major Top-talent Plan, and in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

TABLE X: All relevent points of three classes in each attribute with expanded polygon.

Attribute	θ_1	θ_2	θ_3	$\theta_1 \cap \theta_2$	$\theta_2 \cap \theta_3$	$\theta_1 \cap \theta_3$	$\theta_1 \cap \theta_2 \cap \theta_3$
a_1	5.0600	5.6800	6.6000	5.3700	6.1400	5.8300	5.7800
a_2	3.6800	2.6400	2.8500	3.1600	2.7450	3.2650	3.0567
a_3	1.3400	3.9400	5.4000	2.6400	4.6700	3.3700	3.5600
a_4	0.2200	1.2400	2.0000	0.7300	1.6200	1.1100	1.1533

TABLE XI: Distances between target x^* and all points of expanded attribute polygons.

Distance	θ_1	θ_2	θ_3	$\theta_1 \cap \theta_2$	$\theta_2 \cap \theta_3$	$\theta_1 \cap \theta_3$	$\theta_1 \cap \theta_2 \cap \theta_3$
$a_1 \leftrightarrow x^*$	0.0400	0.5800	1.5000	0.2700	1.0400	0.7300	0.6800
$a_2 \leftrightarrow x^*$	0.1800	0.8600	0.6500	0.3400	0.7550	0.2350	0.4433
$a_3 \leftrightarrow x^*$	0.0599	2.5400	4.0000	1.2400	3.2700	1.9700	2.1600
$a_4 \leftrightarrow x^*$	0.0200	1.0400	1.8000	0.5300	1.4200	0.9100	0.9533

TABLE XII: BBAs of x^* with respect to each attribute in expanded polygon.

$m(\cdot)$	θ_1	θ_2	θ_3	$\theta_1 \cap \theta_2$	$\theta_2 \cap \theta_3$	$\theta_1 \cap \theta_3$	$\theta_1 \cap \theta_2 \cap \theta_3$	Θ
$m_1(\cdot)$	0.7473	0.0293	0.0001	0.1880	0.0019	0.0119	0.0161	0.0054
$m_2(\cdot)$	0.3227	0.0055	0.0192	0.1235	0.0102	0.2319	0.0665	0.2205
$m_3(\cdot)$	0.6628	0.0000	0.0000	0.0006	0.0000	0.0000	0.0000	0.3366
$m_4(\cdot)$	0.8426	0.0019	0.0000	0.0395	0.0002	0.0040	0.0031	0.1087

REFERENCES

[1] Z. Pawlak, *Rough sets: Theoretical aspects of reasoning about data*. Springer Science & Business Media, 2012, vol. 9.

[2] P. Maji and S. Paul, "Rough set based maximum relevance-maximum significance criterion and gene selection from microarray data," *International Journal of Approximate Reasoning*, vol. 52, no. 3, pp. 408–426, 2011.

[3] S. Trabelsi, Z. Elouedi, and P. Lingras, "Classification systems based on rough sets under the belief function framework," *International Journal of Approximate Reasoning*, vol. 52, no. 9, pp. 1409–1432, 2011.

[4] M. He and W. Ren, "Attribute reduction with rough set in context-aware collaborative filtering," *Chinese Journal of Electronics*, vol. 26, no. 5, pp. 973–980, 2017.

[5] D. Liang, Z. Xu, and D. Liu, "A new aggregation method-based error analysis for decision-theoretic rough sets and its application in hesitant fuzzy information systems," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 6, pp. 1685–1697, 2017.

[6] U. Kumar and H. Inbarani, "PSO-based feature selection and neighborhood rough set-based classification for BCI multiclass motor imagery task," *Neural Computing and Applications*, vol. 28, no. 11, pp. 3239–3258, 2017.

[7] Y.-S. Chen and C.-H. Cheng, "Assessing mathematics learning achievement using hybrid rough set classifiers and multiple regression analysis," *Applied Soft Computing*, vol. 13, no. 2, pp. 1183–1192, 2013.

[8] ——, "Evaluating industry performance using extracted RGR rules based on feature selection and rough sets classifier," *Expert Systems with Applications*, vol. 36, no. 5, pp. 9448–9456, 2009.

[9] ——, "Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry," *Knowledge-Based Systems*, vol. 39, pp. 224–239, 2013.

[10] Y. Chen, Y. Xue, Y. Ma, and F. Xu, "Measures of uncertainty for neighborhood rough sets," *Knowledge-Based Systems*, vol. 120, no. C, pp. 226–235, 2017.

[11] T. Zheng and L. Zhu, "Uncertainty measures of neighborhood system-based rough sets," *Knowledge-Based Systems*, vol. 86, no. C, pp. 57–65, 2015.

[12] D. Chen, Y. Yang, and Z. Dong, "An incremental algorithm for attribute reduction with variable precision rough sets," *Applied Soft Computing*, vol. 45, pp. 129–149, 2016.

[13] C. Gao, Z. Lai, J. Zhou, C. Zhao, and D. Miao, "Maximum decision entropy-based attribute reduction in decision-theoretic rough set model," *Knowledge-Based Systems*, 2017.

[14] X. Jia, L. Shang, B. Zhou, and Y. Yao, "Generalized attribute reduct in rough set theory," *Knowledge-Based Systems*, vol. 91, pp. 204–218, 2016.

[15] J. Dezert and F. Smarandache, "An introduction to DSmT," *CoRR*, vol. abs/0903.0279, 2009, http://arxiv.org/abs/0903.0279.

[16] F. Smarandache and J. Dezert, "On the consistency of PCR6 with the averaging rule and its application to probability estimation," in *Proc. of Fusion 2013 Int. Conference on Information Fusion*, 2013, pp. 1119–1126.

[17] ——, "Advances and applications of DSmT for information fusion," vol. 1-4, 2004-2015, https://www.onera.fr/staff/jean-dezert/references.

[18] G. Shafer, *A mathematical theory of evidence*. Princeton university press, 1976, vol. 42.

[19] P. Smets, "Analyzing the combination of conflicting belief functions," *Information fusion*, vol. 8, no. 4, pp. 387–412, 2007.

[20] A. Martin and C. Osswald, "A new generalization of the proportional conflict redistribution rule stable in terms of decision," *Advances and Applications of DSmT for Information Fusion: Collected Works Volume 2*, vol. 2, pp. 69–88, 2006.

[21] Z.-G. Liu, Q. Pan, and J. Dezert, "A new belief-based K-nearest neighbor classification method," *Pattern Recognition*, vol. 46, no. 3, pp. 834–844, 2013.

[22] L. Jiao, Q. Pan, X. Feng, and F. Yang, "An evidential k-nearest neighbor classification method with weighted attributes," in *Proc. of Fusion 2013 Int. Conference on Information Fusion*, 2013, pp. 145–150.

[23] Z.-G. Liu, Q. Pan, and J. Dezert, "Classification of uncertain and imprecise data based on evidence theory," *Neurocomputing*, vol. 133, pp. 459–470, 2014.

[24] T. Denœux, *A k -Nearest Neighbor Classification Rule Based on Dempster-Shafer Theory*. Springer Berlin Heidelberg, 2008.

[25] M. Lichman, "UCI machine learning repository," 2013, http://archive.ics.uci.edu/ml.

[26] D. Han, J. Dezert, Y. Yang, and C. Han, "New neighborhood classifiers based on evidential reasoning," in *Proc. of Fusion 2013 Int. Conference on Information Fusion*, 2013, pp. 158–165.