Adaptively evidential weighted classifier combination

Liguo Fei, Bingyi Kang, Van-Nam Huynh and Yong Deng

Abstract—**Classifier combination plays an important role in classification. Due to the efficiency to handle and fuse uncertain information, Dempster-Shafer evidence theory is widely used in multi-classifiers fusion. In this paper, a method of adaptively evidential weighted classifier combination is presented. In our proposed method, the output of each classifier is modelled by basic probability assignment (BPA). Then, the weights are determined adaptively for individual classifier according to the uncertainty degree of the corresponding BPA. The uncertainty degree is measured by a belief entropy, named as Deng entropy. Discounting-and-combination scheme in D-S theory is used to calculate the weighted BPAs and combine them for the final BPA for classification. The effectiveness of the proposed weighted combination method is illustrated by numerical experimental results.**

Index Terms—**Classifier combination, Dempster-Shafer evidence theory, Deng entropy, Classification, Weight.**

I. Introduction

Classification is a method of integrated learning [1] which belongs to machine learning techniques [2], [3], [4], [5], [6], [7], [8] as one branch. It attracts much attention of researchers along with the perfection of the theoretical basis. And its application is widely published in different fields, such as text classification and retrieval [9], image recognition and speech recognition [10], [11]. There exist a large number of well-known classifiers: support vector machine (SVM) [6], [12], radial basis function (RBF) [13], naive Bayes (NB) [14], decision tree learner (REPTree), multilayer perceptron (MP), 1 nearest neighbor (1NN, or IB1), and RBFnetwork (RBFN). NB and SVM are in the top ten data-mining algorithms [15]. However, it is noteworthy that the ability to collect and deal with information for a single classifier is limited [16]. Moreover, this limitation has a serious impact to the accuracy of the classification results [17]. On the other hand, it's apparent that there exist a lot of patterns that cannot be classified using different learning algorithms or techniques in the classification systems. And these sets of patterns will not overlap necessarily [18]. It means that different classifiers can provide different information from different aspects, which can complement each other for better classification results [19]. In other words, the combination of different classifiers is more beneficial to take advantages of their own strengths to improve the quality of the classification.

Taking notice of the significance and the potential applications of classifiers combination, more and more researches and exploration are done to build an ensemble classifier [20] which could perfect the performance of the individual classifier. Fattah *et al*. [21] presented the comprehensive investigation of different proposed new term weighting schemes for sentiment classification, and exploit the class space density based on the class distribution in the whole documents set as well as in the class documents set. Dłez-Pastor *et al*. [22] proposed a new approach to build ensembles of classifiers for two-class imbalanced data sets which can lead to larger AUC compared to other ensembles of classifiers. Ahmadvand *et al*. [23] applied the combination of multiple classifiers to medical image processing to supervise the segmentation of MRI brain images. Moosavian *et al*. [24] put forword a new method to recognize the spark plug fault based on sensor fusion and classifier combination using Dempster-Shafer evidence theory. Due to the effectiveness to handle uncertainty , D-S theory is paid more and more attention in multi-classifiers fusion. Yager *et al*. [25] proposed the ordered weighted averaging (OWA) to aggregate the information in the uncertainty profile for obtaining representative values in decision-making. Quost *et al*. [26] presented optimized t-norm in the Dempster-Shafer framework based combination rules to combine non independent classifiers. Marek *et al*. [27] built ensemble classifiers using belief functions and OWA operators for classification.

Recently, Huynh *et al*. [28] presented an evidential reasoning based framework for weighted combination of classifiers for word sense disambiguation (WSD). Within this framework, the probability distributions (PD) are obtained from multi-classifiers. Then, the authors presented a method to weight the PDs for discounting their own uncertainty measured by Shannon entropy. Next, the BPAs are determined from each classifier's PD by the discounting operation. Finally, all obtained BPAs are combined using Dempster's rule to obtain the final results as the ensemble classifier for classification. What is certain is that the evidential reasoning based framework conducts itself well for WSD than others congeneric method by their experimental results. However,

L. Fei and B. Kang is with the School of Computer and Information Science, Southwest University, Chongqing, 400715, China; V. Huynh is with the Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan; Y. Deng is with the the School of Computer and Information Science, Southwest University, Chongqing, 400715, China, and also with Institute of Fundamental and Frontier Science, University of Electronic Science and Technology of China, Chengdu, 610054, China. e-mail: (ydeng@swu.edu.cn, prof.deng@hotmail.com.)

this algorithm still has its limitations to handle more general case. The method of Huynh *et al.* [28] obtains PDs firstly. However, the output of each classifier may be BPA due to the high uncertain environment. In other works, the output $\psi_i(\mathbf{x})$ can be BPAs directly instead of a posterior probability distribution on φ in many practical application. In these situations, the method of Huynh *et al.* [28] will be no more applicable. To address this issue, we proposed a new evidential reasoning based framework based on D-S theory. And the process of the two methods are comparing in Figure 1.

Comparing with the method of Huynh *et al.* [28], our proposed method deals with the BPAs from classifiers directly. It is recognized that BPA itself exists uncertainty degree, and the higher uncertainty degree of the BPA, the less information provided by the output of a classifier and then the lower weight it should be assigned. The weighting process is obvious different in the two methods. A new method named Deng entropy [29] is utilized to measure the uncertainty degree of BPAs, and the weight with regard to each classifier are defined adaptively based on the input pattern under classification. Finally, we combine multi-classifiers with Dempster's rule based on the weighed BPAs. In conclusion, there are two major improvements in the proposed method comparing with the method of Huynh *et al.* [28]. The first one is that we use BPAs instead of PDs to represent more uncertain information. The second one is that the Deng entropy is made use of to determine the weights of multi-classifiers. It should be pointed out the proposed method can be seen as the generalization of the method of Huynh *et al.* [28]. If the output of classifiers are PDs, the proposed method degenerated as the method in [28]. From this aspect, the proposed method is more efficient to handle uncertain information. In addition, one of the advantages of the proposed method keeps obtaining the BPAs dynamic with the changes of the output of classifiers. Then the weights and the weighted BPAs also change adaptively. This fully embodies the characteristics of our method adaptive and this recognizes the adaptive quality of our proposed method profoundly.

The organization of the rest of this paper is as follows. Section 2 starts with a brief presentation of the D-S theory and its basic rules and some necessary related concepts. The proposed method for the D-S theory based framework for weighted combination of classifiers is presented in Section 3. Section 4 presents and analyzes the experimental results. Conclusion is presented in Section 5.

A. Dempster-Shafer evidence theory

Dempster-Shafer evidence theory (D-S theory) is proposed by Dempster and developed later by Shafer [30], [31]. This theory extends the elementary event space in probability theory to its power set named as frame of discernment and constructs the basic probability assignment(BPA) on it. In addition, there is a combination rule presented by Dempster to fuse different BPAs. In particular, D-S theory can definitely degenerate to the probability theory if the belief is only assigned to single elements. Therefore, the D-S theory is the generalization of probability theory with the purpose of handling uncertainty and is widely used to uncertainty modeling [32], [33], [34], decision making [35], [36], [37], [38], [39], [40], information fusion [41], [42] and uncertain information processing [43], [44], [45]. The basic definitions about D-S theory is shown as follows:

1) Frame of discernment: D-S theory supposes the definition of a set of elementary hypotheses called the frame of discernment, defined as:

$$\theta = \{H_1, H_2, ..., H_N\} \tag{1}$$

That is, θ is a set of mutually exclusive and collectively exhaustive events. Let us denote 2^θ the power set of θ.

2) Mass functions: When the frame of discernment is determined, a mass function m is defined as follows.

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

which satisfies the following conditions:

$$m(\phi) = 0 \tag{3}$$

$$\sum_{A \in 2^\theta} m(A) = 1 \tag{4}$$

In D-S theory, a mass function is also called a basic probability assignment (BPA).

3) Evidence discounting: The discounting operation is used when an evidence provides a BPA, but the evidence is believed by probability α. In this circumstances, The BPA m^α is redefined based on the probability of reliability α as follows

$$m^\alpha(A) = \alpha \times m(A), \quad A \subset \theta \tag{5}$$

$$m^\alpha(\theta) = (1 - \alpha) + \alpha \times m(\theta) \tag{6}$$

where A is the focal element, and m is the mass function.

4) Dempster's rule of combination: In a real system, there may be many evidence originating from different sensors, so we can get different BPAs. Dempster [31] proposed orthogonal sum to combine these BPAs. Suppose m_1 and m_2 are two mass functions. The Dempster's rule of combination denoted by $m = m_1 \bigoplus m_2$ is defined as follows:

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

with

$$K = \sum_{B \bigcap C = \phi} m_1(B) m_2(C) \tag{8}$$

Note that the Dempster's rule of combination is only applicable to such two BPAs which satisfy the condition $K < 1$.

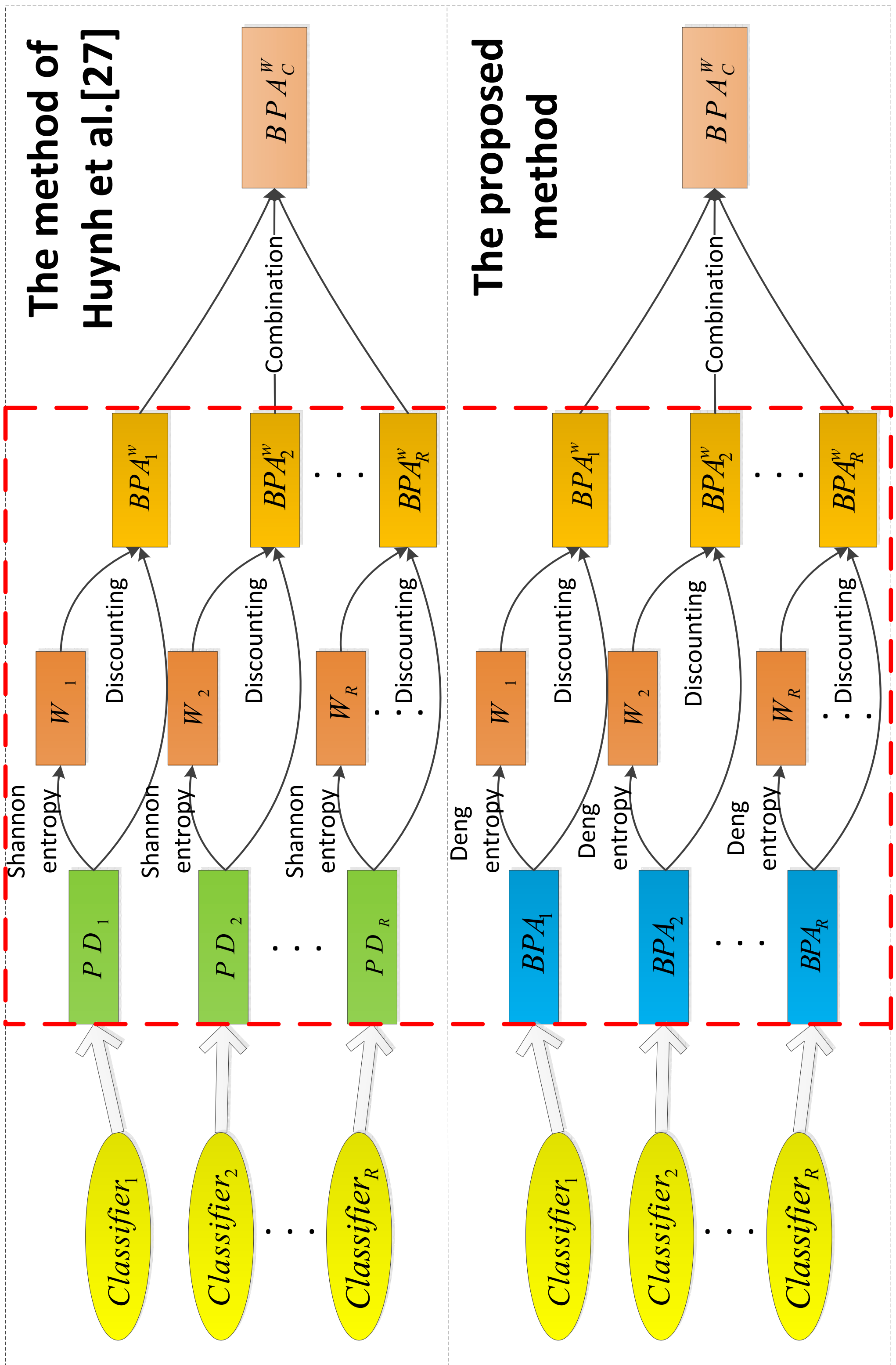

Fig. 1. The comparion between the proposed method and the method in [28]

B. Deng entropy

Deng entropy [29] is presented to measure the uncertainty degree of basic probability assignment as a generalized Shannon entropy in D-S evidence theory. Deng entropy can be described as follows

$$E_d = -\sum_i m(F_i)\log\frac{m(F_i)}{2^{|F_i|}-1} \tag{9}$$

where F_i is a proposition in mass function m, and $|F_i|$ is the cardinality of F_i.

Deng entropy is similar with Shannon entropy in form. The difference is that the belief for each proposition F_i is divided by a term $(2^{F_i}-1)$ which represents the potential number of states in F_i (The empty set is not included). So Deng entropy is the generalization of Shannon entropy, which is used to measure the uncertainty degree of BPA [29].

Specially, Deng entropy can definitely degenerate to the Shannon entropy if the belief is only assigned to single elements. The process is shown as follows [29]

$$E_d = -\sum_i m(\theta_i)\log\frac{m(\theta_i)}{2^{|\theta_i|}-1} = -\sum_i m(\theta_i)\log m(\theta_i) \tag{10}$$

Numerical examples are given to illustrate the computational process of Deng entropy.

Example I.1. *Supposed a frame for discernment $X = \{S,C,V\}$, for a mass function $m(C) = 0.5554, m(C,V) = 0.4420, m(S,C,V) = 0.0026$.*

$$E_d = -0.5554\times\log\frac{0.5554}{2^1-1} - 0.4420\times\log\frac{0.4420}{2^2-1} - 0.0026\times\log\frac{0.0026}{2^3-1} = 1.7220$$

Example I.2. *And another mass function $m(S) = m(C) = m(V) = 1/19, m(S,C) = m(S,V) = m(C,V) = 3/19, m(S,C,V) = 7/19$.*

$$E_d = -\frac{1}{19}\times\log\frac{1/19}{2^1-1} - \frac{1}{19}\times\log\frac{1/19}{2^1-1} - \frac{1}{19}\times\log\frac{1/19}{2^1-1} - \frac{3}{19}\times\log\frac{1/19}{2^2-1} - \frac{3}{19}\times\log\frac{3/19}{2^2-1} - \frac{3}{19}\times\log\frac{3/19}{2^2-1} - \frac{7}{19}\times\log\frac{7/19}{2^3-1} = 4.2479$$

II. The proposed method of weighted combination of classifiers

Let us suppose that there are M classes in the decision system representing as $\varphi = \{c_1, ..., c_M\}$. Also suppose that there are R classifiers $\psi_i (i = 1, ..., R)$ can be used for combination. For each input pattern $\mathbf{X}$, let us denote by

$$\psi_i(\mathbf{x}) = [m_{i1}(\mathbf{x}), ..., m_{i2^M}(\mathbf{x})]$$

the right-hand side of this equality is a mass function obtained from *ith* classifier. We determine the BPA of each classifier from the selected training set using the normal distribution method which is mentioned above.

Each BPA $\psi_i(\mathbf{x})$ is now considered as the belief degree distribution derived from information source provided by classifier ψ_i for classifying $\mathbf{x}$. However, the evidence has a certain extent uncertainty by itself resulting in a decline in the degree of trust. Therefore, it is necessary to quantify somehow the quality of information offering form ψ_i regarding the classification of $\mathbf{x}$ and to consider the uncertainty degree when combining classifiers. Obviously, the greater the uncertainty degree, the lower the accuracy of classification and the larger confusion to us to make classification. Based on these findings we define weights with respect to classifiers according to Deng entropy as follows

$$w_i(BPA) = 1 - \frac{E_d(BPA_i)}{max[E_d(BPA_i)]} \tag{11}$$

where E_d is the Deng entropy expression of the BPA, i.e. The weights are different from one classifier to another depending on how much belief degree the BPA has provided from each classifier.

Based on the mass function and its corresponding weight $w_i(BPA)$, we can obtain the discounted mass function before combining them, expressed as follows

$$m_i^w(A) = w_i(BPA)\times m_i^w(A) \tag{12}$$

$$m_i^w(\theta) = (1 - w_i(BPA)) + w_i(BPA)\times m_i(\theta) \tag{13}$$

where θ is the universal set of mass function.

As of now the weighted BPAs have been determined for individual classifiers. Next, we devote to combine all the evidence' BPAs originating from each classifier ψ_i on the classification of input $\mathbf{x}$, based on the combining rule of D-S theory, to determine an overall mass function for making the final classification decision. The final mass function can be calculated for the expression as follows

$$m_i(BPA) = \bigoplus_{i=1}^{R}(m_i^w(BPA)) \tag{14}$$

where $\otimes$ is a combination operator.

Until now, we have determined BPAs of individual classifier as well as their weights, respectively. Moreover, we also obtain the weighted BPAs by making use of the combination rule of D-S theory. Next, we describe the core algorithm of this paper as follows.

In the following section we will use Iris dataset [46] to conduct some experiments to illustrate our method and demonstrate its effectiveness as well as the dynamic and adaptive nature for classification applying to the combination of multi-classifiers.

III. Experiments and analysis

A large amount of methods for determining the BPA have been proposed by researchers with the more and more application in D-S theory. Zhu *et al.* [47] presented the method using fuzzy membership degrees to obtain the mass function. Within this method, fuzzy c-means (FCM) plays a key role to denote the gray levels as fuzzy

Algorithm 1 The classifier combination algorithm based on Dempster's rule

Initialization:
Determine the BPAs of each classifier expressed as $BPA_i(\mathbf{x})$ (i = 1,....,R)
Iteration:
1: **for** i = 1 to R **do**
2: Calculate $w_i(BPA)$ via (11)
3: Calculate m_i^w via (12) and (13)
4: **endfor**
5: Combine all pieces of m^w via (14)
output:
The final BPA of all the classifiers

sets. Yager *et al*. [48] applied the D-S belief structure to the entire class of fuzzy measures, and studied the entropy from the point of fuzzy measure. Bloch *et al*. [49] associated cluster centers with distance to determine the BPA. Bloch *et al*. used an unsupervised way to obtain the BPA and considered the ambiguity between pixels in medical image processing making use of fuzzy membership functions. It is vagueness instead of randomness leading to the ambiguity. Le Hegarat-Mascle *et al*. [50] and Salzenstein *et al*. [51] used probability density functions (PDFs) to simulate the knowledge derived from all the information source. And then they put forward a subtractive scheme to transform these PDFs into belief degree. Wang *et al*. [52] got mass functions from common multivariate data spaces systematically. In recent years, Xu *et al*. [53] proposed a new method to determine basic probability assignment from training data based on normal distribution assumption. Within his method, normality test is performed for the training set firstly, it will be transformed to an equivalent normal space if training set doesn't meet the normal distribution. And then to construct the models for different attributes. Next, the relationship between the test sample and the normal distribution models will be determined. Finally, the BPA can be calculated on the basis of the intersections of the selected attributes. Comparing with the above-mentioned measures for determining BPA, we consider that the method based on normal distribution is more effective and practical. So in this paper, we will use this method to obtain BPA for each classifier as the preparation for weighted combination of multi-classifiers.

The experiment is based on the Iris data. There are 150 samples of Iris data including 4 attributes for each sample named as Sepal Length (SL), Sepal Width (SW), Petal Length (PL) and Petal Width (PW), respectively. These samples are divided into three classes named as Setosa, Versicolour and Virginica, respectively. There are 50 samples for each of the three classes, and 30 samples are selected randomly as the training set, and the remaining 20 samples regarded as the test set. Each of the four attributes is considered as an information source as well as a classifier, and there are three training sets and three test sets correspondingly. In other words, each attribute is treated as an evidence from a classifier ψ_i. The data can be obtained form the UCI repository of machine learning databases (http://archive.ics.uci.edu/ml/dataset/Iris).

TABLE I
FOUR BPAS OF THE TEST SAMPLE

Attribute (classifier)	BPA		
SL	m({C})	m({C, V})	m({S, C, V})
	0.5554	0.4420	0.0026
SW	m({C})	m({C, V})	m({S, C, V})
	0.6019	0.3771	0.0210
PL	m({V})	m({C, V})	m({S, C, V})
	0.5691	0.4308	0.0001
PW	m({V})	m({C, V})	m({S, C, V})
	0.9555	0.0444	0.0001

Next step, we will determine the BPAs of each attribute of the Iris data, namely, the mass functions of individual classifiers using the above mentioned normal distribution assumption.

Now an example is given to show the process of the classifier combination for classification. Supposing that the training sets and test sets have been obtained from Iris data using normal distribution method. We then select an instance as test sample from the test set of Virginica. The four attribute values are shown as follows:

$$SL = 6.3cm, \quad SW = 2.5cm, \quad PL = 5.0cm, \quad PW = 1.9cm$$

The BPAs of this attributes (classifiers) are show in Table 1, and the S, V and C represent class Setosa, Versicolour and Virginica, respectively.

Taking the attribute SL as an example to explain the calculation of the proposed algorithm.

It is obvious the Deng entropy of attribute SL is 1.7220 from Example 2.1. And it can be proved for three elements in frame of discernment, the Deng entropy gets maximal value when the BPA distributes as Example 2.2. So, the $max[E_d(BPA)]$ is 4.2479 in this exmaple. Then the weight can be obtained by Eq. (12) as follows

$$W_{SL}(BPA) = 1 - \frac{1.7220}{4.2479} = 0.5946$$

Also, the weighted BPA can be calculated by Eqs. (13) and (14).

$$m(\{C\}) = 0.5946 \times 0.5554 = 0.3303$$
$$m(\{C, V\}) = 0.5946 \times 0.4420 = 0.2628$$
$$m(\{S, C, V\}) = (1 - 0.5946) + 0.5946 \times 0.0026 = 0.4069$$

TABLE II
THE FOUR DENG ENTROPY AND WEIGHT OF TEST SAMPLE

Attribute (classifier)	Deng entropy	weight
SL	1.7220	0.5946
SW	1.7451	0.5892
PL	1.6706	0.6067
PW	0.3342	0.9213

TABLE III
FOUR DISCOUNTED BPAS OF THE TEST SAMPLE

Attribute (classifier)	BPA		
SL	m({C})	m({C, V})	m({S, C, V})
	0.3303	0.2628	0.4069
SW	m({C})	m({C, V})	m({S, C, V})
	0.3546	0.2222	0.4232
PL	m({V})	m({C, V})	m({S, C, V})
	0.3452	0.2614	0.3934
PW	m({V})	m({C, V})	m({S, C, V})
	0.8803	0.0409	0.0788

The weights and weighted BPAs of the other three attributes are shown in Table 2 and Table 3.

Now we are committed to combine all the four BPAs from individual classifiers ψ_i on the classification of the test sample by Eq. (18). The result is show as follows

$$m^w(\{C\}) = 0.0933, \quad m^w(\{V\}) = 0.8356$$
$$m^w(\{C,V\}) = 0.0599, \quad m^w(\{S,C,V\}) = 0.0112$$

The process of our experiment is over for this test sample. The combination results illustrate that the belief degree for V (Virginica) is *0.8314* in the combined BPA, and the effectiveness can be demonstrate from this experiment. Other discounted BPAs of rest test samples, namely, the classification of input **x**, can be obtained by this process. In order to demonstrate the results of our experiments more visually and effectively, another parts of experiment results are given based on Iris dataset using the proposed method. The results are shown in Figures 2-4.

In Figure 2-4, the x-coordinate represents 20 test samples of three test sets from Setosa, Versicolour and Virginica, respectively. And the y-coordinate means the probability values of the class which the test sample belong to in the BPAs of individual classifiers. We can find that the all probability values of Setosa test set are close to 1, and for Versicolour the most values exceed 0.9. There exist a few classifiers out of operation in making classification of Virginica test set, but most of the rest part perform a good job. Suppose that 0.5 is the demarcation point deciding whether the the proposed method is effective. Thus, the recognition rate of all pieces of classifiers using our method approaches reaches to 95% approximately in this experiment.

We select a test sample which was worst suitable in classification for class Virginica to analyze the causes of this phenomenon. And we give the four classifiers' BPAs,

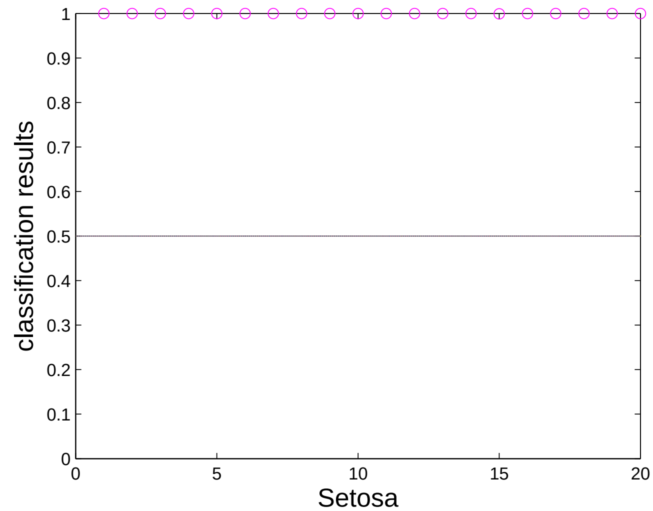

Fig. 2. The classification results based on class Setosa using the proposed method

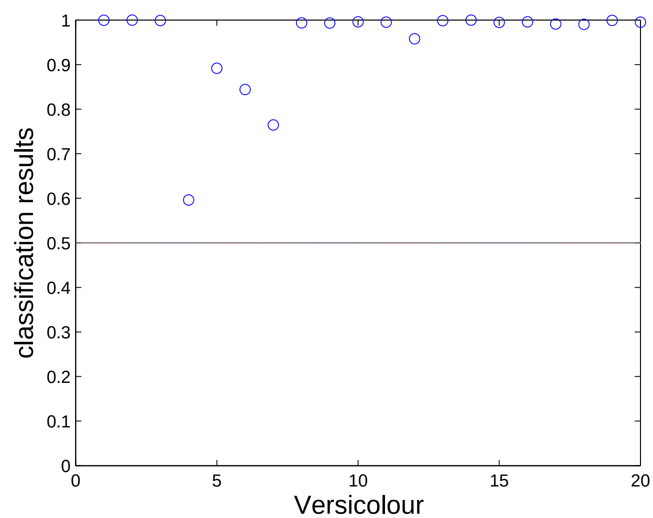

Fig. 3. The classification results based on class Versicolour using the proposed method

their weights, and the four weighted BPAs in Table 4, 5 and 6, respectively. For this input classification, the combination results of corresponding four classifiers are shown as follows.

$$m^w(\{C\}) = 0.7595, \quad m^w(\{V\}) = 0.1093$$
$$m^w(\{C,V\}) = 0.1088, \quad m^w(\{S,C,V\}) = 0.0224$$

From Tables 4-6, we find that there exist three classifiers providing BPAs to support the class Versicolour, so the weighted BPAs still support class Versicolour resulting in the final combination results confirm the input classification belongs to the class Versicolour. By the above examples and analysis we come to a conclusion that the cause of error classification results come from the classifiers instead of our proposed classification method.

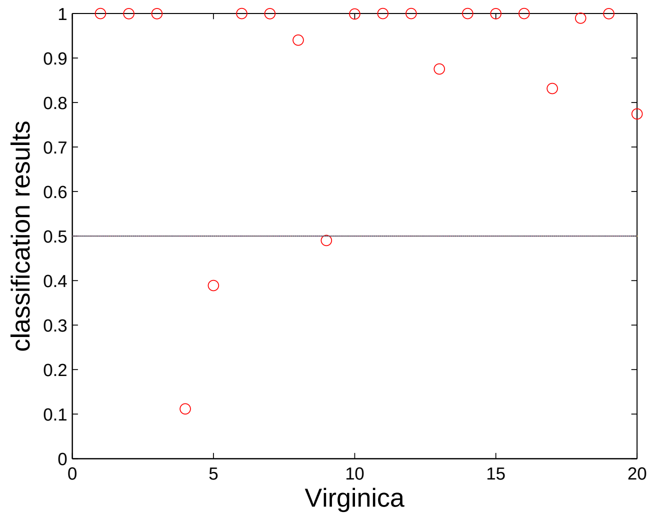

Fig. 4. The classification results based on class Virginica using the proposed method

TABLE IV
FOUR BPAS OF THE TEST SAMPLE

Attribute (classifier)	BPA		
SL	m({C})	m({C, V})	m({S, C, V})
	0.5554	0.4420	0.0026
SW	m({C})	m({C, V})	m({S, C, V})
	0.4821	0.4372	0.0807
PL	m({V})	m({C, V})	m({S, C, V})
	0.6809	0.3189	0.0002
PW	m({C})	m({C, V})	m({S, C, V})
	0.8711	0.1289	0.0000

In addition, we conduct 100 times random experiments with the purpose of further explaining the accuracy of the proposed classification method. We list the average classification accuracy rates for Class Setosa, Versicolour, Virginica and the average of the three classes conducting the random experiments 10, 20,...,100 times, respectively. The results are shown in Table 4.

For expressing the results more unambiguous and visualized, we give the results of the average classification accuracy rates varying the random experiments from 1 to 100 in Figure 5.

From Table 4 and Figure 5, we can find that the average classification accuracy rates of the three classes

TABLE V
THE FOUR DENG ENTROPY AND WEIGHT OF TEST SAMPLE

Attribute (classifier)	Deng entropy	weight
SL	1.7220	0.5946
SW	2.2419	0.4722
PL	1.4118	0.6676
PW	0.7587	0.8214

TABLE VI
FOUR DISCOUNTED BPAS OF THE TEST SAMPLE

Attribute (classifier)	BPA		
SL	m({C})	m({C, V})	m({S, C, V})
	0.3303	0.2628	0.4069
SW	m({C})	m({C, V})	m({S, C, V})
	0.2276	0.2065	0.5659
PL	m({V})	m({C, V})	m({S, C, V})
	0.4546	0.2129	0.3325
PW	m({C})	m({C, V})	m({S, C, V})
	0.7155	0.1059	0.1786

and their average are considerably high. In conclusion, the experiments manifests classification validity of our proposed weighted combination algorithm for multi-classifiers based on D-S theory.

IV. CONCLUSION

In this paper, the basic framework of D-S theory has been constructed for weighted combination of multi-classifiers for classification. A new method has been proposed to define adaptively weights of individual classifier based on Deng entropy which is used to measure the uncertainty degree of BPAs. Then we combine the weighted BPAs derived from individual classifier to obtain the final BPA for the classification decision. It should be pointed out that the proposed method can be seen as the generalization of the method of Huynh *et al.* [28]. If the output of classifiers are PDs, the proposed method is degenerated as the method of Huynh *et al.* [28]. From this point of view, the proposed method is more efficient to handle uncertain information. Moreover, our method can determine corresponding BPAs as the output of classifiers have changed, namely the proposed method has good adaptability.

In the experimental section, we determine BPAs of three test sets using normal distribution method based on Iris dataset. Then, the weights are calculated and weighted BPAs are determined making use of our proposed method for each classification. Finally, these weighted BPAs are combined by Dempster's rule. The experimental results illustrate the effectiveness of our method for classification.

ACKNOWLEDGMENTS

The work is partially supported by National High Technology Research and Development Program of China (863 Program) (Grant No. 2013AA013801), National Natural Science Foundation of China (Grant Nos. 61174022,61573290), China State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (Grant No.BUAA-VR-14KF-02).

REFERENCES

[1] C. Dalton-Puffer, *Discourse in content and language integrated learning (CLIL) classrooms*. John Benjamins Publishing, 2007, vol. 20.

TABLE VII
AVERAGE CLASSIFICATION ACCURACY RATES OF T TIMES

Class	Average classification accuracy(%) of T times									
	10	20	30	40	50	60	70	80	90	100
Setosa	100	100	100	100	100	100	100	100	100	100
Versicolour	94.5	94.25	93.5	93.25	93.2	93.17	93	92.94	93.11	93.2
Virginica	90	90.25	91	90.63	90.3	90.33	90.71	90.75	90.56	90.15
Average of all the classes	94.83	94.83	94.83	94.63	94.5	94.5	94.57	94.56	94.56	94.45

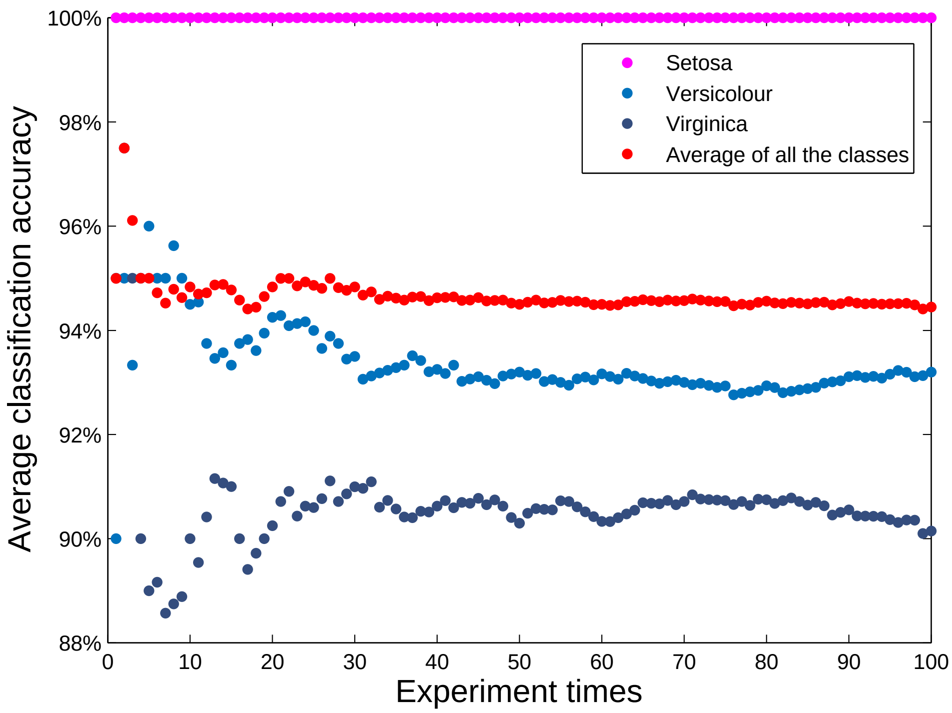

Fig. 5. Average classification accuracy rates varying random experiments from 1 to 100

[2] M. A. Fattah, F. Ren, and S. Kuroiwa, "Sentence alignment using feed forward neural network," *International journal of neural systems*, vol. 16, no. 06, pp. 423–434, 2006.
[3] M. A. Fattah, D. B. Bracewell, F. Ren, and S. Kuroiwa, "Sentence alignment using p-nnt and gmm," *Computer Speech & Language*, vol. 21, no. 4, pp. 594–608, 2007.
[4] M. A. Fattah and F. Ren, "Ga, mr, ffnn, pnn and gmm based models for automatic text summarization," *Computer Speech & Language*, vol. 23, no. 1, pp. 126–144, 2009.
[5] G. Salton, A. Singhal, M. Mitra, and C. Buckley, "Automatic text structuring and summarization," *Information Processing & Management*, vol. 33, no. 2, pp. 193–207, 1997.
[6] M. A. Fattah, "The use of msvm and hmm for sentence alignment," *Journal of Information Processing Systems*, vol. 8, no. 2, pp. 301–314, 2012.
[7] M. A. Fattah, R. Fuji, and S. Kuroiwa, "Effects of phoneme type and frequency on distributed speaker identification and verification," *IEICE transactions on information and systems*, vol. 89, no. 5, pp. 1712–1719, 2006.
[8] M. A. Fattah, F. Ren, and S. Kuroiwa, "Sentence alignment using hybrid model," in *Natural Language Processing and Knowledge Engineering, 2005. IEEE NLP-KE'05. Proceedings of 2005 IEEE International Conference on*. IEEE, 2005, pp. 388–392.
[9] C. D. Manning, P. Raghavan, H. Schütze *et al.*, *Introduction to information retrieval*. Cambridge university press Cambridge, 2008, vol. 1.
[10] K. M. Ting and Z. Zheng, "A study of adaboost with naive bayesian classifiers: Weakness and improvement," *Computational Intelligence*, vol. 19, no. 2, pp. 186–200, 2003.
[11] Z.-g. Liu, Q. Pan, J. Dezert, and A. Martin, "Adaptive imputation of missing values for incomplete pattern classification," *Pattern Recognition*, vol. 52, pp. 85–95, 2016.
[12] M. Joshi and C. Penstein-Rosé, "Generalizing dependency features for opinion mining," in *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*. Association for Computational Linguistics, 2009, pp. 313–316.
[13] M. J. Powell, *The theory of radial basis function approximation in 1990*. University of Cambridge. Department of Applied Mathematics and Theoretical Physics, 1990.
[14] D. D. Lewis, "Naive (bayes) at forty: The independence assumption in information retrieval," in *Machine learning: ECML-98*. Springer, 1998, pp. 4–15.
[15] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, S. Y. Philip *et al.*, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1–37, 2008.
[16] R. Florian, A. Ittycheriah, H. Jing, and T. Zhang, "Named entity recognition through classifier combination," in *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*. Association for Computational Linguistics, 2003, pp. 168–171.
[17] G. H. Rosenfield and K. Fitzpatrick-Lins, "A coefficient of agreement as a measure of thematic classification accuracy." *Photogram-*

metric engineering and remote sensing, vol. 52, no. 2, pp. 223–227, 1986.

[18] J. Kittler, M. Hatef, R. P. Duin, and J. Matas, "On combining classifiers," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 3, pp. 226–239, 1998.

[19] C. Y. Suen and L. Lam, "Multiple classifier combination methodologies for different output levels," in *Multiple Classifier Systems*. Springer, 2000, pp. 52–66.

[20] S. Saha, C. Murthy, and S. K. Pal, "Rough set based ensemble classifier for web page classification," *Fundamenta Informaticae*, vol. 76, no. 1-2, pp. 171–187, 2007.

[21] M. A. Fattah, "New term weighting schemes with combination of multiple classifiers for sentiment analysis," *Neurocomputing*, 2015.

[22] J. F. Díez-Pastor, J. J. Rodríguez, C. García-Osorio, and L. I. Kuncheva, "Random balance: Ensembles of variable priors classifiers for imbalanced data," *Knowledge-Based Systems*, 2015.

[23] A. Ahmadvand, M. Sharififar, and M. R. Daliri, "Supervised segmentation of mri brain images using combination of multiple classifiers," *Australasian Physical & Engineering Sciences in Medicine*, vol. 38, no. 2, pp. 241–253, 2015.

[24] M. Ashkan, K. Meghdad, N. Gholamhassan, K. Maurice, and M. Rizalman, "Spark plug fault recognition based on sensor fusion and classifier combination using dempstervshafer evidence theory," *Applied Acoustics*, vol. 93, pp. 120–129, 2015.

[25] R. R. Yager and N. Alajlan, "Probabilistically weighted owa aggregation," *IEEE Transactions On Fuzzy Systems*, vol. 22, no. 1, pp. 46–56, 2014.

[26] B. Quost, M.-H. Masson, and T. Denœux, "Classifier fusion in the dempster–shafer framework using optimized t-norm based combination rules," *International Journal of Approximate Reasoning*, vol. 52, no. 3, pp. 353–374, 2011.

[27] M. Reformat and R. R. Yager, "Building ensemble classifiers using belief functions and owa operators," *Soft Computing*, vol. 12, no. 6, pp. 543–558, 2008.

[28] V.-N. Huynh, T. T. Nguyen, and C. A. Le, "Adaptively entropy-based weighting classifiers in combination using dempster–shafer theory for word sense disambiguation," *Computer Speech & Language*, vol. 24, no. 3, pp. 461–473, 2010.

[29] Y. Deng, "Deng entropy," *Chaos, Solitons & Fractals*, vol. 91, pp. 549–553, 2016.

[30] G. Shafer *et al.*, *A mathematical theory of evidence*. Princeton university press Princeton, 1976, vol. 1.

[31] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *The annals of mathematical statistics*, pp. 325–339, 1967.

[32] P. Smets and R. Kennes, "The transferable belief model," *Artificial intelligence*, vol. 66, no. 2, pp. 191–234, 1994.

[33] T. L. Wickramarathne, K. Premaratne, and M. N. Murthi, "Toward efficient computation of the dempster–shafer belief theoretic conditionals," *Cybernetics, IEEE Transactions on*, vol. 43, no. 2, pp. 712–724, 2013.

[34] W. Jiang, Y. Yang, Y. Luo, and X. Qin, "Determining basic probability assignment based on the improved similarity measures of generalized fuzzy numbers," *International Journal of Computers Communications & Control*, vol. 10, no. 3, pp. 333–347, 2015.

[35] R. R. Yager and N. Alajlan, "Decision making with ordinal payoffs under dempster–shafer type uncertainty," *International Journal of Intelligent Systems*, vol. 28, no. 11, pp. 1039–1053, 2013.

[36] X. Su, S. Mahadevan, P. Xu, and Y. Deng, "Dependence assessment in Human Reliability Analysis using evidence theory and AHP," *Risk Analysis*, vol. 35, pp. 1296–1316, 2015.

[37] L. V. Utkin, "A new ranking procedure by incomplete pairwise comparisons using preference subsets," *Intelligent Data Analysis*, vol. 13, no. 2, p. 229, 2009.

[38] Y. Deng, "Generalized evidence theory," *Applied Intelligence*, vol. 43, no. 3, pp. 530–543, 2015.

[39] Y. Deng, W. Jiang, and R. Sadiq, "Modeling contaminant intrusion in water distribution networks: A new similarity-based dst method," *Expert Systems with Applications*, vol. 38, no. 1, pp. 571–578, 2011.

[40] Y. Deng, R. Sadiq, W. Jiang, and S. Tesfamariam, "Risk analysis in a linguistic environment: a fuzzy evidential reasoning-based approach," *Expert Systems with Applications*, vol. 38, no. 12, pp. 15 438–15 446, 2011.

[41] F. Cuzzolin, "A geometric approach to the theory of evidence," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 38, no. 4, pp. 522–534, 2008.

[42] J. Ma, W. Liu, D. Dubois, and H. Prade, "Revision rules in the theory of evidence," in *Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on*, vol. 1. IEEE, 2010, pp. 295–302.

[43] S. Rao and K. Annamdas, "A comparative study of evidence theories in the modeling, analysis, and design of engineering systems," *Journal of Mechanical Design*, vol. 135, no. 6, p. 061006, 2013.

[44] W. Jiang, Y. Luo, X. Qin, and J. Zhan, "An improved method to rank generalized fuzzy numbers with different left heights and right heights," *Journal of Intelligent & Fuzzy Systems*, vol. 28, 2015.

[45] Y. Deng, "A threat assessment model under uncertain environment," *Mathematical Problems in Engineering*, vol. 2015, pp. 878 024,http://dx.doi.org/10.1155/2015/878 024, 2015.

[46] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[47] Y. M. Zhu, L. Bentabet, O. Dupuis, D. Babot, M. Rombaut *et al.*, "Automatic determination of mass functions in dempster-shafer theory using fuzzy c-means and spatial neighborhood information for image segmentation," *Optical Engineering*, vol. 41, no. 4, pp. 760–770, 2002.

[48] R. R. Yager, "A class of fuzzy measures generated from a dempster–shafer belief structure," *International Journal of Intelligent Systems*, vol. 14, no. 12, pp. 1239–1247, 1999.

[49] I. Bloch, "Some aspects of dempster-shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account," *Pattern Recognition Letters*, vol. 17, no. 8, pp. 905–919, 1996.

[50] L. Hégarat-Mascle, I. Bloch, D. Vidal-Madjar *et al.*, "Application of dempster-shafer evidence theory to unsupervised classification in multisource remote sensing," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 35, no. 4, pp. 1018–1031, 1997.

[51] F. Salzenstein and A.-O. Boudraa, "Unsupervised multisensor data fusion approach." in *ISSPA*, 2001, pp. 152–155.

[52] P. Xu, Y. Deng, X. Su, and S. Mahadevan, "A new method to determine basic probability assignment from training data," *Knowledge-Based Systems*, vol. 46, pp. 69–80, 2013.

[53] H. Wang and S. McClean, "Deriving evidence theoretical functions in multivariate data spaces: a systematic approach," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 38, no. 2, pp. 455–465, 2008.