Hindawi

Research Article

Determination of Evidence Weights Based on Convolutional Neural Network for Classification Problem

Shijun Xu (ID)**, Yi Hou** (ID)**, Xinpu Deng** (ID)**, Peibo Chen** (ID)**, and Shilin Zhou** (ID)

College of Electronic Science and Technology, National University of Defense Technology, Changsha, China

Correspondence should be addressed to Yi Hou; yihou@nudt.edu.cn

The weighted average is an efficient way to address conflicting evidence combination in the Dempster-Shafer evidence theory. However, it is an open issue how to determine the evidence weights reasonably. Although many traditional conflicting evidence combination solutions based on evidence distance or entropy have been put forward, the evidence weights are determined with a single aspect, and no comprehensive consideration of other useful information affects the weights. Thus, it does not ensure that determination of weights is the most reasonable. By introducing deep learning into conflicting evidence combination, this paper proposes a comprehensive method for determining the evidence weights based on a convolutional neural network. Taking the evidence as the network input and the corresponding weight as the output, it utilizes convolutional neural network to fully mine potentially useful information that affects the evidence weights, in order to determine the weights comprehensively. Additionally, we define a weight loss function. The weights are continuously optimized through back propagation and achieve the optimal when the weight loss function value is the minimum. Classification experimental results demonstrate that the proposed method outperforms traditional ones based on evidence distance or entropy and can be flexibly extended to other application fields as a decision-making fusion method.

1. Introduction

In practical applications, information from different sources is often uncertain, inconsistent, and vague [1]. How to deal with uncertain information effectively is an open problem. Several mathematical theories have been discussed for solving this problem, such as Bayesian theory [2], fuzzy set theory [3, 4], Dempster-Shafer (D-S) evidence [5, 6], possibility theory [7, 8], D-numbers [9, 10], Z-numbers [11, 12], rough set theory [13, 14], and fractal theory [15, 16]. Especially D-S evidence theory, as a rational and effective method for dealing with uncertain information, has the following three advantages. First, it does not require prior probabilities or satisfy probability additivity compared with Bayesian theory. Secondly, information from different experts and data sources can be fused by Dempster's combination rule [17] to obtain more reasonable results. Thirdly, it can describe uncertainty more flexibly and conviently than other mathematical theories. Therefore, it has been widely applied to

classification [18, 19], risk evaluation [20, 21], fault diagnosis [22, 23], decision-making [24, 25], and so on.

However, Zadeh [26] pointed out that evidence combination would produce counterintuitive results, when there is a conflict between evidences. To overcome this defect, scholars have conducted in-depth research and proposed various improved methods. In general, the existing methods can be divided into two categories. One is to modify Dempster's combination rule; the other is to modify the original evidence.

For the first class, some scholars point out that conflicting information is lost in the process of evidence combination. So the key to modifying Dempster's combination rule is how to allocate the conflict, that is, to which subset the conflict is allocated and in what proportion. Smets et al. suggest that the conflict should be allocated to empty set [27]. Lefevre et al. propose a novel modified approach which proportionally assign the conflict information to the focal element sets [28]. Smardndache et al. propose the conflicting

proportional distribution rule named PCR3 [29]. The method of modifying Dempster's combination rule can solve the conflict problem to a certain extent, whereas the drawback of this method is that the good performance is destroyed, like commutativity and associativity. So this paper focuses on the second method.

For the second class, the initial evidences are corrected with the weights to obtain the weighted average evidences. Then, the weighted average evidences are fused by using Dempster's combination rule to get reasonable results. Therefore, the weighted average is an efficient way to address conflicting evidence combination. Nevertheless, this is a challenging problem how to determinate the evidence weights reasonably. For this reason, scholars have proposed some novel methods. Deng defined a new uncertainty measure Deng entropy to construct the weight coefficient bodies of evidence [30]. Tang et al. propose a weighted belief entropy which measure the uncertainty by using the information of the mass function and the scale of the FOD, in order to obtain the weight of evidence [31]. Qin et al. use a novel belief entropy which is an improved version of Dubois–Prade entropy and Nguyen entropy, to allocate the weights of evidence [32]. Yan et al. use an improved belief entropy based on Deng entropy to determine the weights of evidence [33]. Liu et al. propose a novel weighted evidence combination based on MaxDiff distance [34]. Han et al. introduced the concept of evidence support based on the Jousselme distance function and took a weighted average of all the evidences [35]. Liu et al. design an improved weighted evidence combination method by combining probability distance and conflict coefficient [36]. Xiao generalizes the traditional Jousselme distance to the complex evidence distance to measure the conflicts of complex the basic probability assignment (BPA) functions [37], and used it as a weighted factor to revise the original evidence [38]. The above methods only use entropy or distance information to determine the weights from the perspective of uncertainty and evidence conflict. The disadvantage is that determination of the weights is rather one-sided, and no comprehensive consideration of other useful information that affects the weights, such as importance, reliability, relativity, and unknown information hidden within or between evidence. Consequently, these methods based on evidence distance or entropy do not ensure that determination of weights is the most reasonable.

In view of the powerful adaptive learning and information mining capabilities of convolutional neural networks (CNN), we introduce deep learning into conflicting evidence combination and propose a comprehensive method for determining the evidence weights based on a CNN to solve this problem. Taking the evidence as the network input and the corresponding weight as the output, we define a weight loss function. Through back propagation, update the network parameters and fully mine potentially useful information that affects the weights. So the evidence weights are determined comprehensively and optimized continuously, and finally achieve the optimal when the weight loss function value is the minimum. Compared with traditional algorithms based on evidence distance or entropy, the pro-

posed method makes the determination of weights more reasonable and can achieve higher accuracy rate in classification application.

In summary, the primary contributions in this study are summarized as follows:

(i) Different from traditional methods, this paper proposes a comprehensive method for determining the evidence weights based on CNN

(ii) The evidence weights are not determined for a certain aspect related to the evidence and can reflect the relationship among evidence comprehensively

(iii) Compared with the traditional algorithms based on evidence distance or entropy, the proposed method can achieve higher accuracy rate in classification application

The rest part of this paper is organized below. Section 2 introduces some relevant basic theoretical knowledge about the D-S evidence theory. Section 3 proposes a comprehensive method for determining the evidence weights based on a CNN and introduces overview of proposed method, CNN architecture, and the weight loss function. Section 4 presents the classification application of the proposed method, and analyzes and discusses its results. Conclusions are in Section 5.

2. D-S Evidence Theory

D-S evidence theory is a reasoning system theory first put forward by Dempster in 1967 and further developed by Shafer in 1976. Compared with Bayesian probability theory, it can more flexibly and effectively deal with uncertain information without prior probabilities. Thus, it is an extension of Bayesian probability theory. In the frame of D-S evidence theory, Dempster's combination rule can be used to combinate evidences collected from different sources, which satisfies the commutative and associative laws. Some basic concepts about evidence theory is introduced as follows.

Definition 1 (Frame of discernment). Assuming that Θ is a set of mutually exclusive and exhaustive elements $F_i (i = 1, 2, 3, 4, \cdots, N)$, and it can be defined as [31].

$$\Theta = \{F_1, F_2, F_3, F_4, \cdots, F_N\}, \tag{1}$$

where Θ is called the frame of discernment (FOD), and F_i is named single-element proposition or subset. We define 2^Θ as a power set which contains 2^N elements and can be described as

$$2^\Theta = \{\varnothing, F_1, \cdots, F_N, \{F_1, F_2\}, \{F_1, F_3\}, \cdots, \Theta\}, \tag{2}$$

where $\varnothing$ is an empty set in Eq.(2).

Definition 2 (The basic probability assignment function). The BPA function m is also called mass function and is

defined as a mapping of the power set 2^Θ to $[0,1]$ [33].

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

which satisfies

$$
\begin{aligned}
m(\varnothing) &= 0 \\
0 &\leq m(A) \leq 1, \\
\sum_{A \subseteq \Theta} m(A) &= 1,
\end{aligned} \tag{4}
$$

where mass function $m(A)$ represents the degree of support to A, and A is called focal element or proposition. The mass function $m(\varnothing)$ is equal to 0 in classical D-S evidence theory.

Definition 3 (Dempster' s combination rule). In D-S evidence theory, two BPAs can be combined with Dempster' s rule of combination, defined as followers [17]:

$$m(A) = (m_1 \oplus m_2)(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B)m_2(C), \tag{5}$$

in which

$$K = \sum_{B \cap C = \varnothing} m_1(B)m_2(C), \tag{6}$$

where $\oplus$ represents Dempster's combination rule. K is called conflict coefficient, and it has values between 0 and 1. The bigger K is, the more conflict between two evidences is.

Definition 4 (Weighted average evidence). Suppose that $m_1, m_{2,\cdots}, m_n$ is the evidence collected from different data sources, and $w_1, w_{2,\cdots}, w_n$ is the corresponding weights of evidence. Then, original evidence can be modified by the weight to obtain the weighted average evidence:

$$m' = \sum_{i=1}^{n} w_i \times m_i, \tag{7}$$

in which

$$\sum_{i=1}^{n} w_i = 1. \tag{8}$$

The weight w_i is equal to 1, indicating that the corresponding evidence m_i is a piece of information that can be fully reliable. Relatively small weight indicates that it plays little role in evidence combination, and the weight is equal to 0, which means that it can be discarded directly.

3. Determination of Evidence Weights Based on Convolutional Neural Network

3.1. Overview of Proposed Method. To determine the evidence weights more comprehensively and reasonably, we propose a novel method for determining the evidence weights based on a CNN. First, we take initial BPAs, namely, evidences, as the input of CNN, and the corresponding weights are defined as output. Then, we define a weight loss function and the weights are continuously optimized through back propagation. Finally, the optimal evidence weights can be obtained when the weight loss function value is the minimum.

3.2. CNN Architecture. In this section, a CNN is introduced, which consists of a convolution layer, two fully connected layer, and a softmax function output layer, as shown in Figure 1. The initial BPA m_i which is a $1 \times N$ vector is defined as the input to the network. N represents the number of elements within FOD, namely, the number of categories in the dataset. Before a convolution operation, these input evidence vectors are combined into $k \times N$ matrix. After that, we choose $k \times k$ convolution kernels to mine the potential information contained within and between the evidences. Finally, the outputs of two fully connected layer are fed into a softmax function output layer, in order to produce probability weights corresponding to the input evidences. The size of the convolution kernel and softmax function output lay depends on the amount of input evidences.

3.3. The Weight Loss Function. During the conflicting evidence combination, the input evidence m_i is modified by the output weight w_i:

$$m' = \sum_{i=1}^{k} w_i \times m_i, \tag{9}$$

where w_i ranges from 0 to 1 and m' is the weighted average evidence. After the input evidences is corrected, the final combination results $\hat{y}$ can be got by using Dempster's combination rule to combine the weighted averaged evidence m' for $k-1$ times:

$$
\begin{aligned}
\hat{y}(A) &= \left(m' \oplus m' \oplus \cdots \oplus m' \right)_{k-1}(A) \\
&= \frac{\sum_{A_1,\cdots,A_k \subseteq \Theta, A_1 \cap \cdots \cap A_k = A} m'(A_1) \cdots m'(A_k)}{1 - \sum_{A_1,\cdots,A_k \subseteq \Theta, A_1 \cap \cdots \cap A_k = \varnothing} m'(A_1) \cdots m'(A_k)},
\end{aligned} \tag{10}
$$

where $\oplus$ represents Dempster's combination rule which satisfies the polarizability. It means that the total belief degree of a single element increases and the total belief degree of multiple elements decreases when multiple identical evidences are fused. In addition, provided the basic probability value of a single element of the evidence is the largest, the basic probability value of this single element is still the largest when two identical evidences are combined by Dempster's combination rule.

The output weights corresponding to the input evidences should be determined according to the following rules: (a) The higher the relativity between the input evidence vector and the true category one-hot vector is, the larger its corresponding output weight will be. (b) The higher the conflict
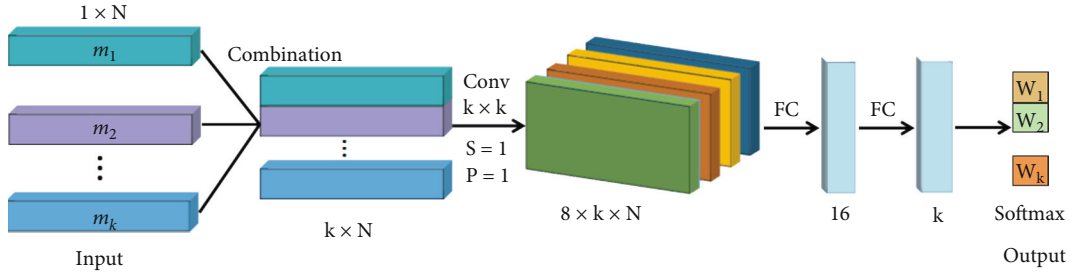
FIGURE 1: Architecture of the proposed method (S=Stride, P=Padding).

between the input evidence vector and the true category one-hot vector is, the smaller its corresponding output weight will be, that is, the evidence will contribute little to the combination results.

Based on the above rules, we define a weight loss function L_w as the sum of the cross entropy between all pairs of y_i and $\widehat{y}_i$ to obtain the optimal evidence weights:

$$L_w = -\frac{1}{n}\sum_i y_i \ln \widehat{y}_i. \tag{11}$$

In the above formula, n represents the number of training samples. $y_i \in R^{1 \times N}$ is the true category one-hot vector. $\widehat{y}_i$ is the final combination result. In the process of seeking the minimized loss function value L_w, when L_w does not meet the condition, network parameters are constantly updated and output weights are continuously optimized through feedback. Since Dempster's combination rule satisfies the polarizability, the weight of normal evidence will be larger and larger, and the weight of conflicting evidence will be smaller and smaller. When L_w satisfies the condition, namely, reaches the minimum value, the output weights achieve the optimal. The flow of the evidence weights optimization is showed in Figure 2.

4. Experimental Results and Discussion

4.1. Experimental Setup. This section gives the introduction of UCR datasets, and the datasets generation, parameter settings, and experiment procedure are presented.

4.1.1. UCR. UCR datasets are obtained from UCR time series data mining archive (http://www.cs.ucr.edu/eamonn/timeseriesdata/) which is a publicly available and real-world time series dataset and has always been used for classification. To verify the classification ability of the proposed method, we conduct the experiments on 6 UCR datasets, including ElectricDevices (ED), UWaveGestureLibraryY (UY), UWaveGestureLibraryZ (UZ), CricketX, CricketY, and CricketZ. ElectricDevices consists of 7711 test samples and contains 7 categories. Both UWaveGestureLibraryY and UWaveGestureLibraryZ consist of 3582 test samples, and contain 8 categories. Each of CricketX, CricketY, and CricketZ consists of 390 test samples and contains 12 categories.

4.1.2. Datasets Generation. UCR datasets are the original data containing attribute information and cannot be directly used as the input of the model which determines the evidence weights. Therefore, to obtain the initial BPAs from these datasets, we utilize the BPA generation method that is the probability output of neural network [39]. Each class in UCR test set can be regarded as one element within the frame of discernment $\Theta = \{F_1, F_2, F_3, \cdots, F_N\}$. These elements are exclusive and independent. We take the probability output of a UCR test set on a single neural network as BPAs. Since the MultiLayer Perceptron (MLP), Fully Convolutional Network (FCN), and Residual Network (ResNet) are defined as the standard baseline in the time series classification [40], these three single neural network are used to generate BPAs in this paper. A test sample will generate three pieces of evidence on three single neural networks. For the same test set, the prediction accuracy rates obtained on three single neural networks which differ in structure and performance will certainly not be all the same, so the BPAs generated by three neural networks will be different. In other words, there is a conflict between the BPAs. Hence, it is reasonable that we use the probability output of a test set on the MLP, FCN, and ResNet as a dataset for conflicting evidence combination research. However, the BPAs generated by the neural network only contains singleton element proposition. Therefore, the artificial simulation method is adopted to obtain the BPAs which contain multi-element propositions. The artificial synthesis dataset is formed by manually collecting classical numerical examples from papers with conflicting evidence fusion. It consists of 21 training samples and 42 test samples. Each sample in this dataset contains 5 pieces of evidences, and 3 categories which are denoted by $\{F_1, F_2, F_3\}$. Assuming that the five pieces of evidence in the first sample are defined as $m_{11}(\cdot)$, $m_{12}(\cdot)$, $m_{13}(\cdot)$, $m_{14}(\cdot)$, and $m_{15}(\cdot)$, the BPAs are shown as follows.

$$\begin{aligned}
&m_{11}(F_1) = 0.41, \quad m_{11}(F_2) = 0.29, m_{11}(F_3) = 0.30 \\
&m_{12}(F_2) = 0.90, \quad\quad\quad m_{12}(F_1, F_3) = 0.10 \\
&m_{13}(F_1) = 0.58, \quad m_{13}(F_2) = 0.07, m_{13}(F_1, F_3) = 0.35 \cdot \\
&m_{14}(F_1) = 0.55, \quad m_{14}(F_2) = 0.10, m_{14}(F_1, F_3) = 0.35 \\
&m_{15}(F_1) = 0.60, \quad m_{15}(F_2) = 0.10, m_{15}(F_1, F_3) = 0.30
\end{aligned} \tag{12}$$

4.1.3. Parameter Settings. The proposed method is trained in

TensorFlow using back propagation with Adam to update the network. The learning rate of Adam is 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 1e-8$ [40]. The second type of dataset is randomly divided into the training set and test set according to the ratio of 2 : 8, and the number of training sets is 1542, 716, 716, 78, 78, and 78, respectively. Therefore, the batch sizes of 13 and 19 are opted. Since the proposed model is not complicated, we set the number of epochs to 2000.

4.1.4. Experiment Procedure. This paper uses two different types of datasets as database to introduce the application of the proposed method in classification. A detailed description of the procedure is depicted as follows:

(1) The first type of dataset containing multi-element propositions is derived from the artificial synthesis dataset. Each sample in this dataset contains 5 pieces of evidences. Correspondingly, the size of the convolution kernel and the softmax function output lay is 5. The second type of dataset contains only single-element propositions. It is derived from the predicted results of 6 UCR test sets on the MLP, FCN, and Resnet. One sample in these test sets will produce 3 pieces of evidences. The second type of dataset is randomly divided into the training set and test set according to the ratio of 2 : 8. In order to verify the robustness of model, each dataset is divided randomly for 5 times, and the same random state is repeated for 5 times to record the average value

(2) Get the optimal evidence weights when the proposed cross entropy loss function value satisfies the minimum

(3) Original evidences are corrected with the weights to obtain the weighted average evidences

(4) Then, the weighted average evidences are fused by using Dempster's rule to get the final combination evidences

(5) The predicted category is determined by the final combination evidence. As to the first class, the category corresponding to the maximum value in the singleton element is the predicted result. As to the second class, the category corresponding to the maximum probability value is the predicted result

To further verify the feasibility of the proposed method, we have compared with four well-known traditional methods, namely, classical D-S theory method [17], two belief entropy-based methods [30, 33], and a distance-based method [35]. Accuracy and processing time for classification are adopted as the evaluation index.

4.2. Results and Discussion

4.2.1. Classification Results. The classification accuracy on the two types of datasets are shown in Tables 1 and 2, respectively. According to experimental results, we can draw the following conclusions: (a) For the first type of dataset,
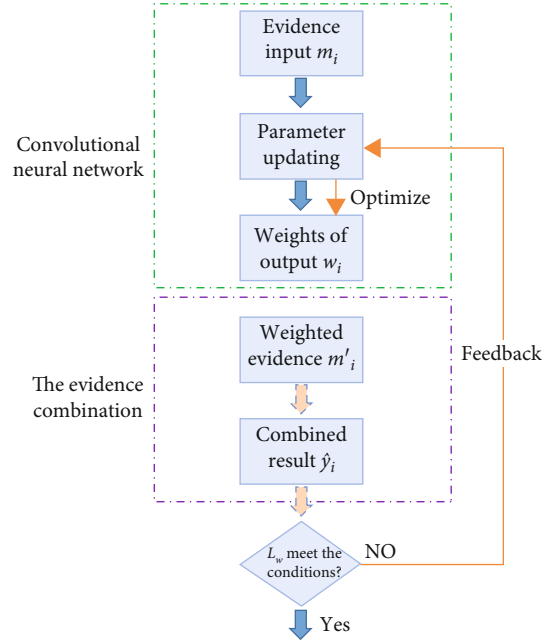

FIGURE 2: Flow of the evidence weights optimization.

Yan et al.'s method, Han et al.'s method, and our proposed method similarly achieve the highest classification accuracy of 97.6%, while Dempster's has a classification accuracy of 64.2% and Deng's method has a classification accuracy of 95.2%. It indicates that the proposed method retains competitive performance. (b) The classification accuracy of the proposed method is always higher compared with a single best-performing neural network classification algorithm, and the average improvement on the second type of datasets is 2.47%, 4.20%, 1.59%, 0.74%, 3.21%, and 1.73%, respectively. It illustrates that the proposed method which combines the predicted results of three single neural networks can obtain better classification results than a single network. In addition, proposed method is to get the final decision by fusing multiple prediction results, which is regarded as the decision level fusion, so it can be flexibly extended to other application fields. (c) For the same dataset, the classification accuracy of the proposed method under five random states has little fluctuation. It proves the robustness of our model. (d) Compared with classic traditional method, two belief entropy-based traditional methods and a distance-based traditional method on the second type of datasets, the classification accuracy of the proposed method is the highest, and the total average improvement is 4.06%, 10.11%, 10.11%, and 16.19%, respectively. These indicate that the determination of weights based on CNN is more reasonable, which proves the validity of proposed method.

4.2.2. Processing Time Results. The experiment of processing time is implemented on an ordinary personal computer with a Intel Core i7-9750H CPU at 2.60 GHz and 8 GB RAM. According to the time complexity calculation presented in [41], the time complexity of four traditional algorithms is $O(x \times 2^n)$, $O(x \times 2^{2n})$, $O(x \times 2^{2n})$, and $O(x^2 \times 2^n)$, where x and n represent the number of evidences and elements in

Table 1: The classification performance comparison on the first type of dataset.

Dataset	Dempster [17]	Deng[30]	Yan and Deng [33]	Han et al. [35]	Proposed method
Artificial synthesis	64.2%	95.2%	**97.6%**	**97.6%**	**97.6%**

Table 2: The classification performance comparison on the second type of datasets.

Dataset	Random state	MLP	FCN	ResNet	Dempster [17]	Deng [30]	Yan and Deng [33]	Han et al. [35]	Proposed method
ED	1	59.1%	69.2%	74.1%	71.4%	61.6%	61.6%	37.9%	**76.5%**
	2	59.2%	68.9%	74.2%	71.6%	61.8%	61.8%	37.3%	**76.4%**
	3	59.6%	69.6%	74.0%	71.9%	62.0%	62.0%	37.7%	**77.1%**
	4	59.8%	69.3%	74.2%	71.6%	62.3%	62.3%	37.2%	**76.7%**
	5	59.3%	69.4%	74.3%	71.4%	62.2%	62.2%	37.1%	**76.4%**
UY	1	68.1%	60.0%	64.1%	71.9%	62.8%	62.8%	70.4%	**72.4%**
	2	69.2%	59.9%	64.6%	72.0%	62.6%	62.6%	71.4%	**73.5%**
	3	68.0%	59.6%	64.4%	71.3%	62.5%	62.5%	71.1%	**72.3%**
	4	69.1%	60.3%	65.0%	72.5%	63.3%	63.3%	71.4%	**73.7%**
	5	69.2%	59.9%	64.4%	72.5%	63.1%	63.1%	71.2%	**72.7%**
UZ	1	68.3%	69.6%	75.8%	76.6%	70.1%	70.1%	64.5%	**77.4%**
	2	69.5%	69.4%	75.8%	76.9%	70.5%	70.5%	64.4%	**77.5%**
	3	68.8%	69.3%	75.8%	76.6%	70.2%	70.2%	64.0%	**77.7%**
	4	69.1%	69.9%	75.5%	76.4%	70.5%	70.5%	64.2%	**76.5%**
	5	68.4%	69.4%	75.2%	76.3%	70.4%	70.4%	64.2%	**76.9%**
CricketX	1	56.0%	79.1%	78.2%	72.1%	74.0%	74.0%	63.4%	**80.4%**
	2	54.8%	79.8%	78.5%	73.0%	72.7%	72.7%	63.4%	**80.4%**
	3	58.9%	80.7%	78.8%	73.7%	74.3%	74.3%	62.1%	**80.8%**
	4	57.0%	81.0%	79.4%	74.3%	73.7%	73.7%	62.1%	**82.1%**
	5	58.9%	80.7%	80.1%	73.0%	74.6%	74.6%	63.1%	**81.4%**
CricketY	1	58.6%	78.2%	78.5%	76.9%	68.9%	68.9%	70.1%	**82.6%**
	2	60.2%	79.4%	80.1%	77.8%	69.8%	69.8%	69.5%	**83.4%**
	3	60.8%	79.1%	79.4%	76.6%	70.1%	70.1%	69.8%	**83.5%**
	4	60.2%	78.5%	79.8%	76.6%	68.5%	68.5%	69.2%	**81.6%**
	5	57.6%	78.2%	79.1%	75.0%	68.5%	68.5%	70.8%	**81.9%**
CricketZ	1	61.2%	79.4%	80.1%	78.2%	73.3%	73.3%	69.5%	**82.5%**
	2	60.2%	80.4%	79.4%	77.2%	73.0%	73.0%	70.1%	**81.4%**
	3	60.2%	81.0%	80.4%	79.4%	74.3%	74.3%	69.8%	**83.3%**
	4	60.5%	79.1%	80.4%	78.2%	73.7%	73.7%	70.5%	**82.6%**
	5	61.8%	80.1%	81.7%	79.4%	75.3%	75.3%	70.8%	**82.5%**

Table 3: The processing time of different approaches.

Dataset	Dempster [17]	Deng [30]	Yan and Deng [33]	Han et al. [35]	Proposed method	
					Training time (s)	Testing time (s)
ED	0.27	1.08	1.08	0.67	252.33	0.05
UY	0.14	0.59	0.59	0.35	117.43	0.02
UZ	0.14	0.59	0.59	0.35	117.68	0.02
CricketX	0.02	0.09	0.09	0.06	16.76	0.01
CricketY	0.02	0.09	0.09	0.06	16.59	0.01
CricketZ	0.02	0.09	0.09	0.06	17.17	0.01
Artificial synthesis	0.01	0.01	0.01	0.01	235.53	0.04

the FOD, respectively. The time complexity of Deng and Yan's algorithms is higher. For all datasets, we take the average processing time of 5 random states for comparison, and the results are illustrated in Table 3. The proposed method spends a lot of time in the training phase, but after training, the processing time of our approach is comparable to that of the other four methods.

5. Conclusion

In this paper, by introducing deep learning into conflicting evidence combination, we propose a comprehensive method for determining the evidence weights based on a CNN. Taking the evidence as the network input and the corresponding weight as the output, it utilizes CNN to fully mine potentially useful information that affects the evidence weights, in order to determine the weights comprehensively. Besides, the weights are continuously optimized through back propagation and achieve the optimal when the weight loss function value satisfies the minimum. The classification experimental results show that the proposed method makes the determination of weights more reasonable and obtains a higher classification accuracy compared with traditional ones based on evidence distance or entropy and can be flexibly extended to other application fields as a decision-making fusion method.

Data Availability

UCR datasets are obtained from UCR time series data mining archive (http://www.cs.ucr.edu/eamonn/timeseriesdata/).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

In this research activity, all the authors were involved in the data collection and preprocessing phase, developing the theoretical concept of the model, empirical research, results analysis and discussion, and manuscript preparation. All authors have agreed to the published version of the manuscript.

Acknowledgments

References

[1] F. Xiao, "CaFtR: a fuzzy complex event processing method," *International Journal of Fuzzy Systems*, vol. 24, no. 2, pp. 1098–1111, 2022.

[2] K. Weise and W. Woger, "A Bayesian theory of measurement uncertainty," *Measurement Science and Technology*, vol. 4, no. 1, pp. 1–11, 1993.

[3] M. B. Ghalia, "Modelling and robust control of uncertain dynamical systems using fuzzy set theory," *International Journal of Control*, vol. 68, no. 6, pp. 1367–1396, 1997.

[4] S. Nestic, A. Djordjevic, H. Puskaric, M. Z. Djordjevic, D. Tadic, and M. Stefanovic, "The evaluation and improvement of process quality by using the fuzzy sets theory and genetic algorithm approach," *Journal of Intelligent & Fuzzy Systems*, vol. 29, no. 5, pp. 2017–2028, 2015.

[5] L. Zadeh, "A simple view of the Dempster–Shafer theory of evidence and its implication for the rule of combination," *AI Magazine*, vol. 29, pp. 85–90, 1986.

[6] M. A. Boujelben, Y. D. Smet, A. Frikha, and H. Chabchoub, "Building a binary outranking relation in uncertain, imprecise and multi- experts contexts: the application of evidence theory," *International Journal of Approximate Reasoning*, vol. 50, no. 8, pp. 1259–1278, 2009.

[7] W. Jiang, C. Huang, and X. Deng, "A new probability transformation method based on a correlation coefficient of belief functions," *Journal of Intelligent & Fuzzy Systems*, vol. 9, pp. 1337–1347, 2019.

[8] E. De la Rosa and W. Yu, "Data-driven fuzzy modeling using restricted boltzmann machines and probability theory," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 7, pp. 2316–2326, 2020.

[9] X. Deng, Y. Hu, Y. Deng, and S. Mahadevan, "Supplier selection using AHP methodology extended by D numbers," *Expert Systems with Applications*, vol. 41, no. 1, pp. 156–167, 2014.

[10] F. Xiao, "A multiple-criteria decision-making method based on D numbers and belief entropy," *International Journal of Fuzzy Systems*, vol. 21, no. 4, pp. 1144–1153, 2019.

[11] X. F. Ding, L. X. Zhu, and Y. Q. Feng, "A novel linguistic Z-number qualiflex method and its application to large group emergency decision making," *Scientific Programming*, vol. 2020, Article ID 1631869, 12 pages, 2020.

[12] A. Azadeh and R. Kokabi, "Z-number DEA: a new possibilistic DEA in the context of Z-numbers," *Advanced Engineering Informatics*, vol. 30, no. 3, pp. 604–617, 2016.

[13] H. Zhang, H. Liang, and D. Liu, "Two new operators in rough set theory with applications to fuzzy sets," *Information Sciences*, vol. 166, no. 1-4, pp. 147–165, 2004.

[14] T. F. Fan, C. J. Liau, and D. R. Liu, "Dominance-based fuzzy rough set analysis of uncertain and possibilistic data tables," *International Journal of Approximate Reasoning*, vol. 52, no. 9, pp. 1283–1297, 2011.

[15] A. M. Erkmen and H. E. Stephanou, "Information fractals for evidential pattern classification," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 20, no. 5, pp. 1103–1114, 1990.

[16] A. E. Jacquin, "Image coding based on a fractal theory of iterated contractive image transformations," *IEEE Transactions on Image Processing*, vol. 1, no. 1, pp. 18–30, 1992.

[17] A. P. Dempster, "A generalization of Bayesian inference," *Journal of the Royal Statistical Society Series B, Statistical methodology*, vol. 30, no. 2, pp. 205–232, 1968.

[18] Z. Liu, Y. Liu, J. Dezert, and F. Cuzzolin, "Evidence combination based on credal belief redistribution for pattern classification," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 4, pp. 618–631, 2020.

[19] Z. G. Liu, Q. Pan, J. Dezert, and G. Mercier, "Hybrid classification system for uncertain data," *IEEE Transactions on Systems,*

Man, and Cybernetics: Systems, vol. 47, no. 10, pp. 2783–2790, 2017.

[20] L. Zhang, X. Wu, H. Zhu, and S. M. Abou Rizk, "Perceiving safety risk of buildings adjacent to tunneling excavation: an information fusion approach," *Automation in Construction*, vol. 73, pp. 88–101, 2017.

[21] H. Zheng and Y. Tang, "A novel failure mode and effects analysis model using triangular distribution-based basic probability assignment in the evidence theory," *IEEE Access*, vol. 8, pp. 66813–66827, 2020.

[22] Y. Kaijuan, X. Fuyuan, and D. Yong, "Modeling sensor reliability in fault diagnosis based on evidence theory," *Sensors*, vol. 16, no. 1, pp. 113–121, 2016.

[23] L. Wan, H. Li, Y. Chen, and C. Li, "Rolling bearing fault prediction method based on QPSO-BP neural network and Dempster–Shafer evidence theory," *Energies*, vol. 13, no. 5, p. 1094, 2020.

[24] W. Li, K. Guo, and Y. Huo, "Study on decision making method under uncertain information based on d-s evidence theory," *International journal of innovative computings*, vol. 6, pp. 3737–3749, 2010.

[25] J. Liu and B. Gong, "Analysis of uncertainty multi-attribute group decision making process based on D-S evidence theory," *Journal of Computers*, vol. 6, no. 4, pp. 711–717, 2011.

[26] L. A. Zadeh, "Reviews of books, a mathematical theory of evidence," *Ai Magazine*, vol. 5, 1984.

[27] P. Smets, "The combination of evidence in the transferable belief model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 447–458, 1990.

[28] E. Lefevre and Z. Elouedi, "How to preserve the conflict as an alarm in the combination of belief functions?," *Decision Support Systems*, vol. 56, pp. 326–333, 2013.

[29] F. Smarandache and J. Dezert, "A simple proportional conflict redistribution rule," *International Journal of Applied Mathematics and Statistics*, vol. 3, pp. 1–36, 2004.

[30] Y. Deng, Z. Deng, and S. Li, "Study on methods to optimize laser-sharpening quality, efficiency and topography," *Chaos*, vol. 46, pp. 409–416, 2016.

[31] Y. Tang, D. Zhou, S. Xu, and Z. He, "A weighted belief entropy-based uncertainty measure for multi-sensor data fusion," *Sensors (Basel)*, vol. 17, no. 4, pp. 928–939, 2017.

[32] M. Qin, Y. Tang, and J. Wen, "An improved total uncertainty measure in the evidence theory and its application in decision making," *Entropy*, vol. 22, no. 4, pp. 487–496, 2020.

[33] H. Yan and Y. Deng, "An improved belief entropy in evidence theory," *IEEE Access*, vol. 8, pp. 57505–57516, 2020.

[34] W. Liu, "Analyzing the degree of conflict among belief functions," *Artificial Intelligence*, vol. 170, no. 11, pp. 909–924, 2006.

[35] D. Yong, S. WenKang, Z. ZhenFu, and L. Qi, "Combining belief functions based on distance of evidence," *Decision Support Systems*, vol. 38, no. 1, pp. 489–493, 2010.

[36] Z. Liu, J. Dezert, Q. Pan, and G. Mercier, "Combination of sources of evidence with different discounting factors based on a new dissimilarity measure," *Decision Support Systems*, vol. 52, no. 1, pp. 133–141, 2011.

[37] F. Xiao, "CEQD: a complex mass function to predict interference effects," *IEEE Transactions on Cybernetics*, vol. 52, no. 8, pp. 7402–7414, 2021.

[38] F. Xiao, "CED: a distance for complex mass functions," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 23, pp. 1–11, 2020.

[39] R. Fay, F. Schwenker, and G. Paim, "Hierarchical neural networks utilising Dempster–Shafer evidence theory," *Artificial Neural Networks in Pattern Recognition*, vol. 4087, pp. 198–209, 2006.

[40] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: a strong baseline," *International Joint Conference on Neural Networks*, vol. 12, pp. 1578–1585, 2017.

[41] H. Li and F. Xiao, "A method for combining conflicting evidences with improved distance function and Tsallis entropy," *International Journal of Intelligent Systems*, vol. 35, no. 11, pp. 1814–1830, 2020.