Received: 6 October 2020 | Revised: 29 June 2021 | Accepted: 19 July 2021

DOI: 10.1002/int.22593

RESEARCH ARTICLE

WILEY

Combining conflicting evidence based on Pearson correlation coefficient and weighted graph

Jixiang Deng[1] | **Yong Deng**[1,2,3,4] | **Kang Hao Cheong**[5,6]

[1]Institute of Fundamental and Frontier Science, University of Electronic Science and Technology of China, Chengdu, China

[2]School of Education, Shannxi Normal University, Xi'an, China

[3]School of Knowledge Science, Japan Advanced Institute of Science and Technology, Nomi, Japan

[4]Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

[5]Science, Mathematics and Technology Cluster, Singapore University of Technology and Design (SUTD), Singapore, Singapore

[6]SUTD-Massachusetts Institute of Technology International Design Centre, Singapore, Singapore

Correspondence

Yong Deng, Institute of Fundamental and Frontier Science, University of Electronic Science and Technology of China, 610054 Chengdu, China.
Email: dengentropy@uestc.edu.cn

Kang Hao Cheong, Science, Mathematics and Technology Cluster, Singapore University of Technology and Design, Singapore S487372, Singapore.
Email: kanghao_cheong@sutd.edu.sg

Funding information

National Natural Science Foundation of China, Grant/Award Number: 61973332

Abstract

Dempster–Shafer evidence theory (evidence theory) has been widely used as an efficient method for dealing with uncertainty. In evidence theory, Dempster's rule is the most well-known evidence combination method but it does not work well when the evidence is in high conflict. To improve the performance of combining conflicting evidence, an original and novel evidence combination method is presented based on the Pearson correlation coefficient and weighted graph. The proposed method can correctly recognize the alternative situation with a high accuracy. Besides, the convergence performance of this method is better when compared with other combination rules. In addition, the weighted graph generated by the proposed method can directly represent the relationship between different evidence, which can help researchers estimate the reliability of different body of evidence. Our experimental results indicate the advantages of our proposed evidence combination rule over existing methods, and the results are analyzed and discussed.

KEYWORDS

Dempster–Shafer evidence theory, evidence combination, Pearson correlation coefficient, target recognition, weighted graph

1 | INTRODUCTION

Over the past decades, many theories have been developed for dealing with uncertainty, for instance, probability theory,[1] fuzzy sets,[2] Dempster–Shafer evidence theory (evidence theory),[3,4] rough sets,[5] and Z-numbers.[6]

Evidence theory has a variety of real-world application, like uncertainty measurements,[7–9] data fusion[10,11] decision making,[12] complex networks,[13,14] risk analysis,[15,16] and in many other fields.[17,18] In evidence theory, Dempster's rule is widely used for combining evidence, but it does not work well when the evidence is in high conflict. To close this critical gap, a lot of improved evidence combination methods has been developed. Murphy[19] proposed a modified combination method by averaging the basic probability assignment (BPA) of every evidence. Then, Deng[20] improved Murphy's method by weighted averaging BPA based on distance of evidence. Taking independent degree as a discounting factor, Yager[21] proposed a improved combination method of belief function. Some other works have also been presented.[22,23] Most of these evidence combinations use the technique of averaging BPA to reduce the influence of conflicting evidence, namely, averaging BPA based on the reliability of every evidence, which can be calculated by similarity,[24,25] distance,[20] information volume,[26,27] and eXtropy.[28]

In statistics, Pearson correlation coefficient[29] is a linear correlation coefficient for measuring the relationship, or association, of two variables, which is developed by Karl Pearson with wide applications.[30,31] Additionally, for modeling the nonlinear relationship of two variables, nonlinear correlation coefficient should be used, such as Spearman's rank correlation coefficient[32] and Kendall's rank correlation coefficient.[33] Since Pearson correlation coefficient is based on the covariance, it can be introduced into evidence theory to calculate the reliability of different evidence. Inspired by this, Xu[34] proposed a method based on Pearson coefficient, and it shows a good accuracy of recognizing objects. Nevertheless, Xu's method cannot directly reflect which evidence is in conflict. Moreover, its accuracy can be improved.

In discrete mathematics, graph theory is one of the prime research field, and graph is a useful mathematical tool for directly modeling relationship between objects. In a certain graph, the objects can be represented by nodes and linked by edges. If the edges have sense of direction, the graph is called directed graph; if not, the graph is called undirected graph. Due to the good performance of representing objects, graphs can be used to represent the nervous system, where nerve cells and nerve fibers can be represented by nodes and edges[35] respectively. The social relationships of people can also be represented by graphs.[36]

Recently, based on graph and complex networks, a new technique of identifying conflicting evidence is proposed,[14] which provides a feasible way to solve the issues of evidence theory with the help of graphs. Based on this technique, Liu proposes a new evidence combination method,[23] which shows good performance in combining conflicting evidence. However, Liu's method does not use averaging technique, and it just use simple graph to identify conflicting evidence, which can be further improved to enhance the performance of combining evidence in conflict.

To overcome the problems discussed above, we propose a novel evidence combination method based on Pearson correlation coefficient and weighted graph, which can combine evidence in conflict and correctly recognize the alternative situation with a high accuracy. Besides, the convergence rate of this combination method is better, in comparison with other conventional methods. Moreover, our proposed method is able to generate a weighted graph to illustrate the relationship between different evidence, and can directly show the reliability for every evidence.

The main contributions of this paper are as follows:

(i) An original and novel evidence combination method based on the Pearson correlation coefficient and weighted graph, is proposed and it can combine evidence in high conflict and accurately recognize the correct target.
(ii) Our proposed method has a good convergence rate, and is well-suited for real world scenarios compared with other existing methods.
(iii) The weighted graph generated by the proposed method can directly show the relationship between different evidence, and can be used to determine the reliability of evidence, as well as identifying conflicting evidence.

Section 2 reviews the preliminaries. Section 3 presents a new evidence combination method based on the Pearson correlation coefficient and weighted graph. We show an experiment in Section 4 to illustrate the advantages of our proposed method. In Section 5, the results of the experiment are discussed and we conclude in Section 6.

2 | PRELIMINARIES

In this section, we will review some preliminaries about evidence theory, Pearson correlation coefficient and graph theory.

2.1 | Dempster–Shafer evidence theory

Dempster-Shafer evidence theory[3,4] can be used to deal with uncertainty. Besides, compared with probability theory, the conditions that evidence theory contents is weaker, which provides it with the ability to express uncertain information directly. Then, we will introduce several conceptions about evidence theory.

Definition 2.1 (Frame of discernment and its power set). Let Θ be an exhaustive nonempty set of N hypotheses, where N elements are mutually exclusive. Θ is called frame of discernment (FOD) denoted as follows:

$$\Theta = \{\theta_1, \theta_2, \theta_3, \ldots, \theta_N\} \tag{1}$$

2^Θ denotes the power set of Θ, and has 2^N elements which are the subsets of Θ. 2^Θ is represented by

$$2^\Theta = \{A_1, A_2, A_3, \ldots, A_{2^N}\} \tag{2}$$

where $A_1 = \varnothing$ and $A_{2^N} = \Theta$.

Definition 2.2 (Basic probability assignment [BPA]). A BPA is defined as a mass function:

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

constrained by these conditions:

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

$$m(\emptyset) = 0 \tag{5}$$

Definition 2.3 (Dempster's rule of combination). Given two BPAs m_1 and m_2, Dempster's rule of combination is defined as follows:

$$\begin{cases} m(A) = \dfrac{\sum_{B \cap C = A} m_1(B) \cdot m_2(C)}{1 - K_{12}} & A \neq \emptyset \\ m(\emptyset) = 0 \end{cases} \tag{6}$$

where K_{12} can be calculated by

$$K_{12} = \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C). \tag{7}$$

2.2 | Pearson correlation coefficient

Pearson correlation coefficient can represent the linear correlation of two variables. The definition of Pearson correlation coefficient is as follows[29]:

Definition 2.4 (Pearson correlation coefficient). Assume two samples X and Y which can be denoted as vectors: $\overrightarrow{\boldsymbol{X}}$ and $\overrightarrow{\boldsymbol{Y}}$. Each sample contains N sample observations which can be denoted as the components of the vectors, namely, $\overrightarrow{\boldsymbol{X}} = [x_1, x_2, ..., x_N]$ and $\overrightarrow{\boldsymbol{Y}} = [y_1, y_2, ..., y_N]$. Then the Pearson correlation coefficient of $\overrightarrow{\boldsymbol{X}}$ and $\overrightarrow{\boldsymbol{Y}}$ is defined as follows:

$$r_{\overrightarrow{\boldsymbol{X}}\overrightarrow{\boldsymbol{Y}}} = \frac{N\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{\sqrt{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}\sqrt{N\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2}} \tag{8}$$

The main properties of Pearson correlation coefficient is that:

(i) The value range of $r_{\overrightarrow{\boldsymbol{X}}\overrightarrow{\boldsymbol{Y}}}$ is $[-1, 1]$.

(ii) If $r_{\overrightarrow{\boldsymbol{X}}\overrightarrow{\boldsymbol{Y}}} > 0$, $\overrightarrow{\boldsymbol{X}}$ and $\overrightarrow{\boldsymbol{Y}}$ is positive correlation.
(iii) If $r_{\overrightarrow{\boldsymbol{X}}\overrightarrow{\boldsymbol{Y}}} = 0$, the linear correlation of $\overrightarrow{\boldsymbol{X}}$ and $\overrightarrow{\boldsymbol{Y}}$ is not obvious.
(iv) If $r_{\overrightarrow{\boldsymbol{X}}\overrightarrow{\boldsymbol{Y}}} < 0$, $\overrightarrow{\boldsymbol{X}}$ and $\overrightarrow{\boldsymbol{Y}}$ is negative correlation.
(vi) The greater $|r_{\overrightarrow{\boldsymbol{X}}\overrightarrow{\boldsymbol{Y}}}|$ is, the higher linear correlation rate of $\overrightarrow{\boldsymbol{X}}$ and $\overrightarrow{\boldsymbol{Y}}$ will be.

It should be noted that Pearson correlation coefficient is for handling the linear relationship of two variables. If the relationship is nonlinear, nonlinear correlation coefficient should be used, such as Spearman's rank correlation coefficient[32] and Kendall's rank correlation coefficient.[33]

2.3 | Graph theory

In graph theory, the graph is a useful mathematical tool for dealing with the relationships among objects. Some basic conceptions of graph theory are listed as follows[37]:

Definition 2.5 (Weighted graph). A weighted graph is defined as $G = (V, E, W)$, where $V = \{v_1, v_2, ..., v_N\}$ is called the node set whose element v_i is node, $E = \{\{v_i, v_j\} | \{v_i, v_j\} \in V \wedge V\}$ is called the edge set whose element $\{v_i, v_j\}$ is edge which connects two nodes v_i and v_j, and $W = \{w_{ij} | i, j = 1, ..., N\}$ is called the weight set whose element w_{ij} is the weight assigned to the edge $\{v_i, v_j\}$.

Definition 2.6 (Adjacency matrix of weighted graph). The adjacency matrix A of a weighted graph $G = (V, E, W)$ is defined as a $|V| * |V|$ matrix, whose elements $a_{mn} = w_{mn}$ if and only if $\{v_m, v_n\} \in E$, otherwise, $a_{mn} = 0$.

It worth noting that if a graph is undirected, its adjacent matrix will be symmetric.

3 | PROPOSED METHOD

Since evidence theory has been proposed, different kinds of evidence combination methods have been proposed. Among them, Dempster's method[3] is the most popular evidence combination rule, and it has been widely used. However, when evidence is in high conflict, the result calculated by Dempster's method may be illogical.

Hence, we propose a novel method to combine evidence in conflict. The main idea is that different evidence has different reliability. The conflicting evidence should be identified and treated cautiously, and the reliable evidence should be trusted and given a high credibility. For the purpose of estimating the reliability of every evidence, we apply following techniques in the proposed method:

(i) Weighted averaging the BPA of every evidence. The reliability (or the weight) of every evidence is calculated based on Pearson correlation coefficient. If an evidence is supported by other evidence, its reliability is considered to be high, which can be calculated by Pearson correlation coefficient. After that, the reliability can be used to weighted average the BPA of evidence, enhancing the accuracy of the algorithm.

(ii) Representing the relationship of every evidence based on weighted graph. Graph is a tool to represent the relationship of objects. The relationship of evidence can illustrated by weighted graph which can help researchers to directly identify the evidence in conflict or the relatively unreliable evidence.

In the rest of this section, first, some basic definitions are proposed. And then, a new evidence combination is present.

3.1 | Basic definitions

Pearson correlation coefficient is the linear correlation of two samples. When the number of samples that we are dealing with is larger than two, we need a efficient way to reorganize and represent the linear correlation of them. Inspired of this, Pearson correlation coefficient matrix (PCCM) is proposed.

Definition 3.1 (Pearson correlation coefficient matrix). Assume there are K samples denoted as vectors: $\vec{\boldsymbol{M}}_1, \vec{\boldsymbol{M}}_2, \ldots, \vec{\boldsymbol{M}}_K$. Then the PCCM of these samples is defined as follows:

$$PCCM = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1K} \\ r_{21} & r_{22} & & r_{2K} \\ \vdots & & \ddots & \vdots \\ r_{K1} & r_{K2} & \cdots & r_{KK} \end{bmatrix} \tag{9}$$

where r_{ij} is the Pearson correlation coefficient of sample $\overrightarrow{\boldsymbol{M}_i}$ and sample $\overrightarrow{\boldsymbol{M}_j}$.

PCCM has N^2 Pearson correlation coefficient so that this matrix describes the linear correlation of all the samples. Assume the samples can be seen as nodes and the Pearson correlation coefficient as weight. Then we can convert the PCCM into a weighted graph and its adjacent matrix.

Definition 3.2 (PCCM-based weighted graph). A PCCM-based weighted graph is a undirected weighted graph defined as $G_{PCCM} = (V, E, W)$, where $V = \{\vec{\boldsymbol{M}}_1, \vec{\boldsymbol{M}}_2, \ldots, \vec{\boldsymbol{M}}_K\}$ is the node set, $E = \{\{\overrightarrow{\boldsymbol{M}_i}, \overrightarrow{\boldsymbol{M}_j}\} | \{\overrightarrow{\boldsymbol{M}_i}, \overrightarrow{\boldsymbol{M}_j}\} \in V \wedge V\}$ is the edge set, and $W = \{w_{ij} | i, j = 1, \ldots, K\}$ is the weight set whose element w_{ij} is defined as follows:

$$w_{ij} = \begin{cases} r_{ij} & (r_{ij} > 0 \text{ and } i \neq j) \\ 0 & (r_{ij} \leq 0 \text{ or } i = j) \end{cases} \tag{10}$$

where r_{ij} is the Pearson correlation coefficient of PCCM. If $w_{ij} > 0$, node $\overrightarrow{\boldsymbol{M}_i}$ and node $\overrightarrow{\boldsymbol{M}_j}$ are connected, namely, $\{\overrightarrow{\boldsymbol{M}_i}, \overrightarrow{\boldsymbol{M}_j}\} \in E$. If $w_{ij} = 0$, node $\overrightarrow{\boldsymbol{M}_i}$ and node $\overrightarrow{\boldsymbol{M}_j}$ are unconnected, namely, $\{\overrightarrow{\boldsymbol{M}_i}, \overrightarrow{\boldsymbol{M}_j}\} \notin E$.

It should be noted that the PCCM-based weighted graph does not have self-loop. As a result, $w_{ij} = 0$ when $i = j$.

Definition 3.3 (Adjacent matrix of PCCM-based weighted graph). A adjacent matrix of PCCM-based weighted graph is defined as follows:

$$A_{PCCM} = \begin{bmatrix} 0 & w_{12} & \cdots & w_{1K} \\ w_{21} & 0 & & w_{2K} \\ \vdots & & \ddots & \vdots \\ w_{K1} & w_{K2} & \cdots & 0 \end{bmatrix} \tag{11}$$

where w_{ij} is the weight of PCCM-based weighted graph.

Because the PCCM-based weighted graph is undirected, its adjacent matrix is symmetric, namely, $w_{ij} = w_{ji}$.

3.2 | Evidence combination algorithm

Assume that there are K evidence $m_1, m_2, ..., m_K$ and N alternatives defined in FOD $\Theta = \{\theta_1, \theta_2, ..., \theta_N\}$. The corresponding power set of FOD is $2^\Theta = \{A_1, A_2, ..., A_{2^N}\}$ and the BPA of these K evidence is $m_i(A_j)$ for $i = 1, 2, ..., K$ and $j = 1, 2, ..., 2^N$. Then the proposed evidence combination algorithm is detailed as follows:

Step 1: Convert K evidence m_i $(i = 1, 2, ..., K)$ into vectors:

$$\overrightarrow{\boldsymbol{M}_i} = [m_i(A_1), m_i(A_2), ..., m_i(A_{2^N})] \tag{12}$$

Step 2: Calculate the Pearson correlation coefficient matrix (PCCM) of K evidence vectors:

$$PCCM = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1K} \\ r_{21} & r_{22} & & r_{2K} \\ \vdots & & \ddots & \vdots \\ r_{K1} & r_{K2} & \cdots & r_{KK} \end{bmatrix} \tag{13}$$

Step 3: Convert PCCM into PCCM-based weighted graph $G_{PCCM} = (V, E, W)$.
Step 4: Obtain the adjacent matrix of PCCM-based weighted graph:

$$A_{PCCM} = \begin{bmatrix} 0 & w_{12} & \cdots & w_{1K} \\ w_{21} & 0 & & w_{2K} \\ \vdots & & \ddots & \vdots \\ w_{K1} & w_{K2} & \cdots & 0 \end{bmatrix} \tag{14}$$

Step 5: Calculate the total weight TW_i of evidence m_i based on the adjacent matrix of PCCM-based weighted graph:

$$TW_i = \sum_{j=1,\, j\neq i}^{K} w_{ij} \tag{15}$$

Step 6: Normalize the total weight to achieve the normalized weight NW_i of evidence m_i:

$$NW_i = \frac{TW_i}{\sum_{i=1}^{K} TW_i} \tag{16}$$

Step 7: Based on NW_i, calculate the weighted averaged evidence WAE:

$$WAE = \{m(A_j)|j = 1, 2, ..., 2^N\} \tag{17}$$

$$m(A_j) = \sum_{i=1}^{K} m_i(A_j)NW_i \tag{18}$$

where $m_i(A_j)$ is the BPA for evidence m_i of A_j.

Step 8: Use Dempster's rule to combine the weighted averaged evidence $K - 1$ times and get the combination result of K evidence.

4 | APPLICATION OF THE PROPOSED METHOD IN TARGET RECOGNITION

In this section, we apply the proposed method in target recognition to illustrate its calculating procedure and performance.

4.1 | Problem statement

Target recognition refers to the ability to recognize or identify targets based on the information generated by various sensors, such as radar, photosensitive device, or vibration detector. By using target recognition algorithm, researcheres are able to recognize the objects in different environment, such as humans, vehicles, or planes.

In the scenario of aircraft navigation detection, assume there are three types of plane (alternatives) represented by $\{A, B, C\}$ and five radars (evidence sources) denoted by $m_1, m_2, ..., m_5$. These five radars detect the planes and report what plane is it most likely to be in the form of BPA. The BPA reported by five evidence sources are collected in Table 1.

4.2 | Experimental result

These five evidence sources are combined by the proposed evidence combination, and then we can recognize the exact alternative of the three based on the result. The calculating steps are detailed as follows.

TABLE 1 The BPA reported by five evidence sources

	$\{A\}$	$\{B\}$	$\{C\}$	$\{A, C\}$
m_1	0.50	0.20	0.30	0.00
m_2	0.00	0.90	0.10	0.00
m_3	0.45	0.20	0.00	0.35
m_4	0.50	0.20	0.00	0.30
m_5	0.45	0.25	0.00	0.30

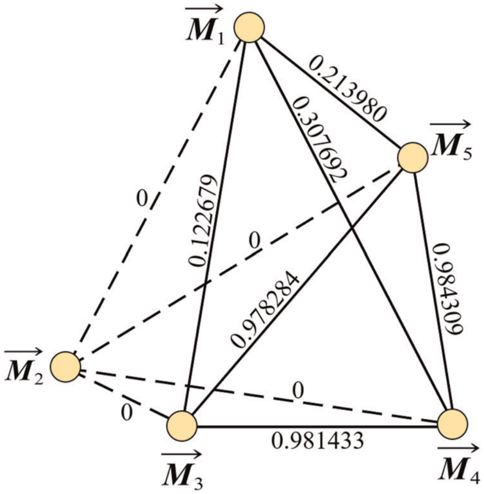

FIGURE 1 PCCM based weighted graph [Color figure can be viewed at wileyonlinelibrary.com]

Step 1: Convert these five evidence into vectors:

$$\overrightarrow{M_1} = [0.50, 0.20, 0.30, 0.00]$$
$$\overrightarrow{M_2} = [0.00, 0.90, 0.10, 0.00]$$
$$\overrightarrow{M_3} = [0.45, 0.20, 0.00, 0.35]$$
$$\overrightarrow{M_4} = [0.50, 0.20, 0.00, 0.30]$$
$$\overrightarrow{M_5} = [0.45, 0.25, 0.00, 0.30]$$

Step 2: Calculate the PCCM of these five evidence vectors:

$$PCCM = \begin{bmatrix} 1 & -0.146944 & 0.122679 & 0.307692 & 0.213980 \\ -0.146944 & 1 & -0.273408 & -0.257151 & -0.102190 \\ 0.122679 & -0.273408 & 1 & 0.981433 & 0.978284 \\ 0.307692 & -0.257151 & 0.981433 & 1 & 0.984309 \\ 0.213980 & -0.102190 & 0.978284 & 0.984309 & 1 \end{bmatrix} \tag{19}$$

Step 3: Convert PCCM into PCCM-based weighted graph $G_{PCCM} = (V, E, W)$ as Figure 1, where the full lines with number represent the weight w_{ij} of two nodes, and the dash lines indicate that two nodes are unconnected, namely, $w_{ij} = 0$.

Step 4: Obtain the adjacent matrix of PCCM-based weighted graph:

$$A_{PCCM} = \begin{bmatrix} 0 & 0 & 0.122679 & 0.307692 & 0.213980 \\ 0 & 0 & 0 & 0 & 0 \\ 0.122679 & 0 & 0 & 0.981433 & 0.978284 \\ 0.307692 & 0 & 0.981433 & 0 & 0.984309 \\ 0.213980 & 0 & 0.978284 & 0.984309 & 0 \end{bmatrix} \tag{20}$$

Step 5: Calculate the total weight TW_i of evidence m_i based on the adjacent matrix of PCCM-based weighted graph:

$$\begin{aligned} TW_1 &= 0 + 0.122679 + 0.307692 + 0.213980 = 0.644351 \\ TW_2 &= 0 + 0 + 0 + 0 = 0 \\ TW_3 &= 0.122679 + 0 + 0.981433 + 0.978284 = 2.082396 \\ TW_4 &= 0.307692 + 0 + 0.981433 + 0.984309 = 2.273434 \\ TW_5 &= 0.213980 + 0 + 0.978284 + 0.984309 = 2.176573 \end{aligned} \tag{21}$$

Step 6: Normalize the total weight to achieve the normalized weight NW_i of evidence m_i:

$$\begin{aligned} NW_1 &= \frac{0.644351}{7.176754} = 0.089783 \\ NW_2 &= \frac{0}{7.176754} = 0 \\ NW_3 &= \frac{2.082396}{7.176754} = 0.290158 \\ NW_4 &= \frac{2.273434}{7.176754} = 0.316777 \\ NW_5 &= \frac{2.176573}{7.176754} = 0.303281 \end{aligned} \tag{22}$$

Step 7: Based on NW_i, calculate the weighted averaged evidence WAE:

$$\begin{aligned} m(A) &= \sum_{i=1}^{5} m_i(A)NW_i = 0.470328 \\ m(B) &= \sum_{i=1}^{5} m_i(B)NW_i = 0.215164 \\ m(C) &= \sum_{i=1}^{5} m_i(C)NW_i = 0.026935 \\ m(A, C) &= \sum_{i=1}^{5} m_i(A, C)NW_i = 0.287573 \end{aligned} \tag{23}$$

Step 8: Use Dempster's rule to combine the weighted averaged evidence four times and get the final combination result:

$$\begin{aligned} m(A) &= 0.985939 \\ m(B) &= 0.001833 \\ m(C) &= 0.004413 \\ m(A, C) &= 0.007816 \end{aligned} \tag{24}$$

In this experiment, we choose the other four typical evidence combination method, including Dempster's method,[3] Murphy's method,[19] Liu et al.'s method,[23] and Deng et al.'s method[20] to compare with the proposed method. The experiment results of these five methods are shown in Table 2.

For the convenience of discussion, the calculating procedure of the BPA $m(A)$ is shown in Table 3. It should be noted that, the total times of combining the five evidence is $5 - 1 = 4$, except for Liu et al.'s method. The reason is that it removes conflicting evidence m_2 and combines the rest of the four evidence by three times.

In the next section, the five evidence combination methods are analyzed based on the experimental result.

5 | ANALYSIS AND DISCUSSION

In this section, comparisons between the proposed method and the other four methods are analyzed and discussed.

In general, Dempster's method[3] is widely used to combine data from sensors based on evidence theory. Murphy's method[19] is an efficient and typical tool to combine conflicting evidence by simple averaging the BPA of evidence. Deng et al.'s method[20] applies evidence distance to calculate credibility of every evidence, which is a weighted-averaging-based method for dealing with conflicting evidence. Liu et al.'s method[23] is a novel evidence combination based on generalized belief entropy, and this method uses graph model to improve the performance of combining evidence. The proposed method is based on Pearson correlation

TABLE 2 Results of five evidence combination methods

Method	$m(A)$	$m(B)$	$m(C)$	$m(A, C)$	Target
Dempster's method[3]	0.00000000	0.65573770	0.34426230	0.00000000	B
Murphy's method[19]	0.89960650	0.07885140	0.01782472	0.00371738	A
Liu et al.'s method[23]	0.95446411	0.00795387	0.03758202	0.00000000	A
Deng et al.'s method[20]	0.96571592	0.01600922	0.01394744	0.00432741	A
Proposed method	0.98593887	0.00183259	0.00441303	0.00781550	A

TABLE 3 The value of $m(A)$ by different times of combining

Method	Times = 1	Times = 2	Times = 3	Times = 4
Dempster's method	0.000000	0.000000	0.000000	0.000000
Murphy's method	0.596448	0.740307	0.836883	0.899607
Liu et al.'s method	0.733945	0.890125	0.954464	/
Deng et al.'s method	0.689934	0.844120	0.925710	0.965716
Proposed method	0.772013	0.909264	0.964400	0.985939

coefficient and weighted graph, which takes both weighted averaging technique and graph model into consideration. The techniques of these five methods are summarized in Table 4.

It is shown in Table 2 that, the proposed method has the best performance because it successfully recognizes the correct alternative A based on the conflicting evidence, and $m(A)$ calculated by proposed combination method is the highest (0.985939) compared with other methods.

As is illustrated in Figure 2, although the difference of Murphy's, Liu et al.'s, Deng et al.'s and the proposed method is not large, proposed method can also identify the alternatives correctly under the condition that the threshold is 0.97.

When confronting extreme environment, sensors would be influenced by many factors such as radiation, temperature or design defects which cause the sensors to report evidence in high conflict with each other. Under this circumstance, the threshold of identifying target will be higher than common. With the highest accuracy, the proposed method is more reliable to combine the conflicting evidence, at least its result will not worse than the other four methods. As a result, the proposed method has the efficiency to handle conflict in a environment with high uncertainty.

It is worth noting that the proposed method is based on Pearson correlation coefficient, which considers the linear relationship between the conflicting evidence and alternatives. When facing the situation of nonlinear relationship, the proposed method should be extended by nonlinear correlation coefficient, such as Spearman's rank correlation coefficient[32] and Kendall's rank correlation coefficient.[33] This paper only focuses on the linear relationship based

TABLE 4 The techniques of five evidence combination methods

Method	Graph-based method	Averaging-based method
Dempster's method	×	×
Murphy's method	×	✓
Liu et al.'s method	✓	×
Deng et al.'s method	×	✓
Proposed method	✓	✓

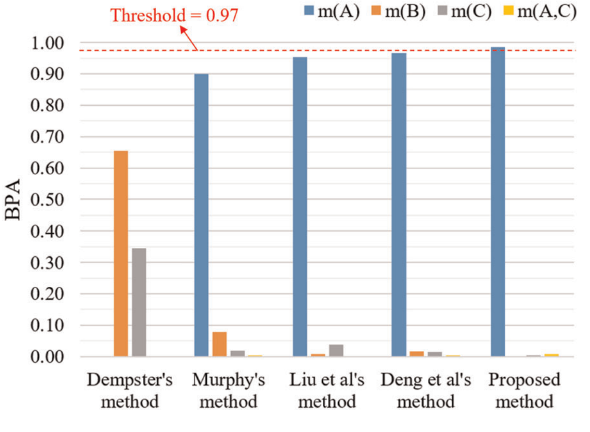

FIGURE 2 Results of five methods when the threshold is 0.97 [Color figure can be viewed at wileyonlinelibrary.com]

on Pearson correlation coefficient. In future work, we will introduce nonlinear correlation coefficient into the proposed method under the condition of nonlinear situation.

Apart from above discussions, more detailed comparisons are expounded in the following three sections based on the techniques that the method uses.

5.1 | Compared with Dempster's method

Dempster's method is neither an averaging-based method nor a graph-based method. When the evidence reported by sensors are in conflict with each other, Dempster's method may yield counter-intuitive results. In this experiment, four evidence support alternative A, while evidence m_2 supports B which is conflicted with other evidence. The result of Dempster's method shows that, even though more evidence support A, Dempster's method supports alternative B ($m(B) = 0.655738$) and is totally against A ($m(A) = 0$), which is illogical.

By contrast, as is both an averaging-based and a graph-based method, the proposed method draws the correct conclusion with high accuracy ($m(A) = 0.985939$), which can be a great alternative of Dempster's method to combine conflicting evidence.

5.2 | Compared with averaging-based methods

Murphy's method, Deng et al.'s method and the proposed method are averaging-based methods. In specific, Murphy's method is a simple-averaging method, namely, every weight of evidence is equal to each other. Deng et al.'s method is a weighted-averaging method, which means that, the weight of evidence can be modified. The proposed method is actually a weighted-averaging method, whose weight is based on Pearson correlation coefficient.

All of the three averaging-based methods get the correct conclusion. However, in contrast to other averaging-based methods, the proposed method is more efficient. The advantages of it are analyzed as follows:

5.2.1 | Better performance of convergence

According to Table 3, the calculating procedure of $m(A)$ based on averaging-based methods is shown in Figure 3. Obviously, at every time of combining, the BPA $m(A)$ of the proposed method is the highest. Besides, the speed of convergence of the proposed method is the best, since the value of $m(A)$ reaches more than 0.9 only by two times of combining.

5.2.2 | More efficient to identify the reliability of evidence

Beyond being a weighted-averaging-based method, the proposed method is also a graph-based method. According to Step 5 of the proposed method, the algorithm generates a weighted graph. Compared with other averaging-based methods, the proposed can directly reflect the relationship of the evidence based on the weighted graph.

As is illustrated in Figure 4, the node $\overrightarrow{\boldsymbol{M}_2}$ has no edge connected with other node, which means that m_2 is not supported by other evidence. As a result, we can directly identify that m_2 is

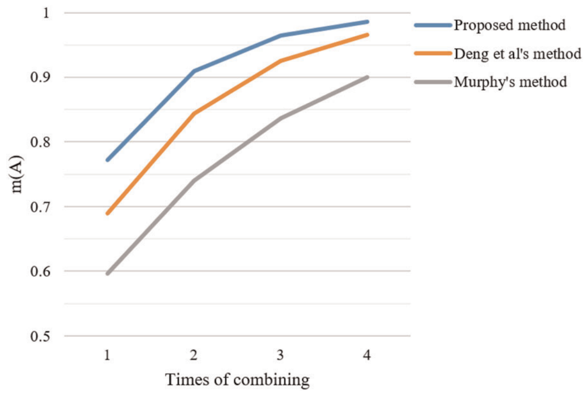

FIGURE 3 The calculating procedure of $m(A)$ based on averaging-based methods [Color figure can be viewed at wileyonlinelibrary.com]

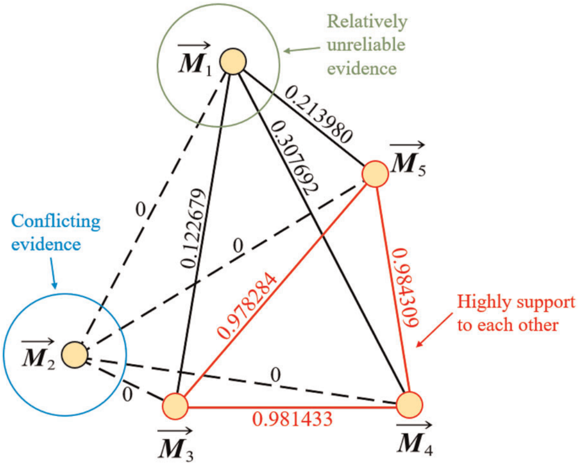

FIGURE 4 Identify the reliability of evidence based on weighted graph [Color figure can be viewed at wileyonlinelibrary.com]

the conflicting evidence which should be carefully checked. Besides, the nodes $\overrightarrow{M_3}$, $\overrightarrow{M_4}$, and $\overrightarrow{M_5}$ connect with each other with high weight more than 0.97, which indicates that, m_3, m_4, and m_5 highly support to each other. Hence, we can trust these three evidence. Moreover, three weights of the node $\overrightarrow{M_1}$ are relatively low, which alerts us that, m_1 is relatively unreliable, and the BPA reported by m_1 should be taken with a grain of salt.

To summarize, compared with other averaging based methods, the advantages of the proposed method are the great performance of convergence and the efficiency of identifying evidence in conflict.

5.3 | Compared with graph-based methods

Liu et al.'s method and the proposed method are graph-based methods. Both of the two methods recognize the correct alternative A. However, the proposed method is better than Liu et al.'s method. The reasons are as follows.

5.3.1 | Better performance of convergence

The calculating procedure of $m(A)$ based on graph-based methods is shown in Figure 5. It worth noting that, since Liu et al.'s method removes the conflicting evidence m_2 and combine the rest of the 4 evidence, the times of combining of it are $4-1=3$. As for the proposed method, because the number of the weighted averaged evidence is the same as that of the original evidence, which equals to 5, the times of combining of the proposed method are $5-1=4$. It is obviously that, at every time of combining, the BPA $m(A)$ of the proposed method is higher than that of Liu et al.'s method, which means that, the performance of convergence of the proposed method is better than Liu et al.'s method. Although the the BPA $m(A)$ of the proposed method is close to that of Liu et al.'s method at the the first three times of combining, the proposed method can still enhance the value of $m(A)$ at the 4th times, improving the ability to identify the correct target. According to Table 3 and Figure 5, even if the proposed method is limited to combining for three times, the value of $m(A)$ produced by the proposed method (0.964400) is still higher than that produced by Liu et al.'s method (0.954464), which further proves the better performance of convergence of the proposed method compared with Liu et al.'s method.

5.3.2 | More efficient to identify the reliability of evidence

Both of the two graph-based methods can generate a graph to identify conflicting evidence. As is illustrated in Figure 6, based on the simple graph generated by Liu et al.'s method shown in Figure 6A, we can draw the conclusion that, m_2 is the conflicting evidence since it is not connected to any nodes, and the remaining evidence supports to each other because m_1, m_3, m_4, and m_5 are connected together. Similarly, based on the weighted graph produced by the proposed method shown in Figure 6B, we can also identify that m_2 is in conflict and the rest of the evidence supports to each other.

However, the big difference is that, Liu et al.'s method generates a simple graph whose edge does not have weight, and the connection state is just true (1) or false (0). By contrast, the proposed method generates a weighted graph with weight attached to the edges which can better represent the relationship between nodes compared with simple graph.

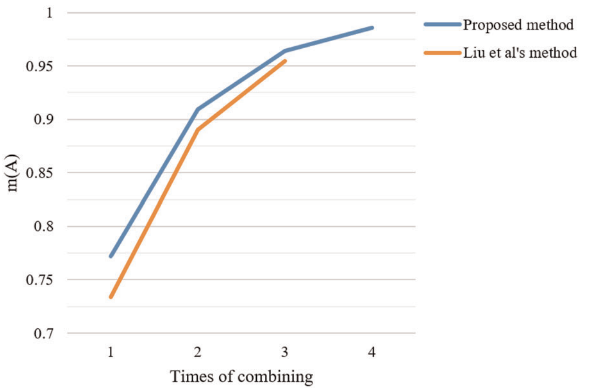

FIGURE 5 The calculating procedure of $m(A)$ based on graph-based methods [Color figure can be viewed at wileyonlinelibrary.com]

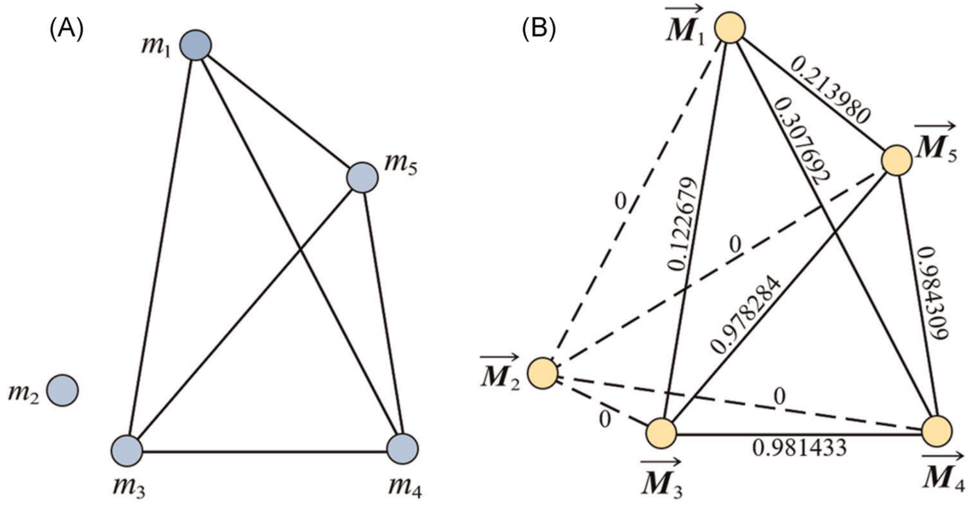

FIGURE 6 Comparison between (A) simple graph based on Liu et al.'s method and (B) weighted graph based on the proposed method [Color figure can be viewed at wileyonlinelibrary.com]

For example, in Figure 6B generated by the proposed method, there are three edges connected to $\overrightarrow{M_1}$ (the dash line means two nodes are unconnected); the weight of these edges are 0.122679, 0.307692, and 0.213980 which represents that the relationship of node $\overrightarrow{M_1}$ between other nodes is not close. While the weights attached to the edges of $\overrightarrow{M_3}$, $\overrightarrow{M_4}$, and $\overrightarrow{M_5}$ are higher than that of $\overrightarrow{M_1}$. As a result, we can draw the conclusion that, evidence m_1 is relatively unreliable compared with m_3, m_4, and m_5, and the BPA reports of evidence m_1 should be treated cautiously. On the contrary, in Figure 6A generated by Liu et al.'s method, node m_1, m_3, m_4, and m_5 are connected to each other, from which we cannot distinguish the difference of reliability of evidence m_1 from m_3, m_4, and m_5.

In addition, based on the weighted graph produced by the proposed method, we can obtain and make full use of the information about the supporting degrees (correlations) of $\overrightarrow{M_3}$, $\overrightarrow{M_4}$, and $\overrightarrow{M_5}$. As is shown in Figure 6B, the supporting degree are 0.978284, 0.984309, and 0.981433, which can be further used in other applications, such as data preprocessing or outlier detection. However, from the simple graph generated by Liu et al.'s method, we can just get the information that two evidence sources are support (connected) or not support (unconnected) to each other, which limits the further processing of BPA.

In conclusion, the proposed method is better than Liu et al.'s method in terms of performance of convergence and the efficiency of identifying reliability of evidence.

6 | CONCLUSION

In this paper, an original and novel evidence combination method based on the Pearson correlation coefficient and weighted graph has been proposed. Our proposed method is both an averaging-based combination method and a graph-based method, and has shown to improve the performance of combining evidence in conflict. Our experimental results indicate the advantages of our proposed evidence combination rule over existing methods. The advantages are summarized as follows:

(i) Our proposed method is able to correctly identify the target among other alternatives, and its accuracy is better than other methods.
(ii) In contrast to existing combination methods, our proposed method has the best performance of convergence.
(iii) As a graph-based method, our proposed method can generate a weighted graph to directly reflect the relationship of different evidence, giving us an ideal approach in estimating the reliability of every evidence.

In future work, we will focus on applying the proposed method to practical fields, such as data combination in fault analysis and information fusion of automatic drive sensors. Besides, Pearson correlation coefficient is for dealing with the linear correlation of two variables. However, the relationship of the conflicting evidence and the alternatives might be nonlinear in the real practice. Hence, with respect to nonlinear situation, it is also worth studying to extend the proposed method based on the nonlinear coefficient, such as Spearman's rank correlation coefficient and Kendall's rank correlation coefficient.

ACKNOWLEDGMENTS
The authors greatly appreciate the editor's encouragement and the reviews' suggestions. The work is partially supported by National Natural Science Foundation of China (Grant no. 61973332) and Invitational Fellowships for Research in Japan (Short-term).

CONFLICT OF INTERESTS
The authors declare that there are no conflict of interests.

ORCID
Jixiang Deng https://orcid.org/0000-0002-1521-1770
Yong Deng https://orcid.org/0000-0001-9286-2123
Kang Hao Cheong https://orcid.org/0000-0002-4475-5451

REFERENCES
1. Lee P. Book review. *Bulletin London Math Society*. 1980;12(4):318.
2. Zadeh LA. Fuzzy sets. *Inform Control*. 1965;8(3):338.
3. Dempster AP. Upper and lower probabilities induced by a multivalued mapping. *Annal Math Stat*. 1967;38(2):325.
4. Shafer G. *A mathematical theory of evidence*. Vol 1. Princeton University Press; 1976.
5. Pawlak Z. Rough sets. *Int J Comput Inform Sci*. 1982;11(5):341.
6. Zadeh LA. A note on Z-numbers. *Inform Sci*. 2011;181(14):2923.
7. Yager RR, Alajlan N, Bazi Y. Uncertain database retrieval with measure-based belief function attribute values. *Inform Sci*. 2019;501:761.
8. Deng Y. Uncertainty measure in evidence theory. *Sci China Inform Sci*. 2020;63(11):210201.
9. Zhang H, Deng Y. Entropy measure for orderable sets. *Inform Sci*. 2021;561:141.
10. Xiao F. A multiple-criteria decision-making method based on D numbers and belief entropy. *Int J Fuzzy Syst*. 2019;21(4):1144.
11. Koksalmis E, Kabak O. Sensor fusion based on Dempster-Shafer theory of evidence using a large scale group decision making approach. *Int J Intell Syst*. 2020;35(7):1126.
12. Xu H, Deng Y. Dependent evidence combination based on decision-making trial and evaluation laboratory method. *Int J Intell Syst*. 2019;34(7):1555.
13. Wen T, Pelusi D, Deng Y. Vital spreaders identification in complex net- works with multi-local dimension. *Knowl-Based Syst*. 2020:105717.

14. Zhao J, Deng Y. Complex network modeling of evidence theory. *IEEE Transactions on Fuzzy Systems*. IEEE; 2020:1. http://doi.org/10.1109/tfuzz.2020.3023760
15. Zhang L, Wu X, Zhu H, AbouRizk SM. Perceiving safety risk of buildings adjacent to tunneling excavation: an information fusion approach. *Autom Construct*. 2017;73:88
16. Zhang L, Ding L, Wu X, Skibniewski MJ. An improved Dempster–Shafer approach to construction safety risk perception. *Knowl-Based Syst*. 2017;132:30.
17. Xiao F. Generalization of Dempster–Shafer theory: a complex mass function. *Appl Intell*. 2020;50:3266.
18. Yager RR. A class of belief structures based on possibility measures. *Soft Comput*. 2018;22(23):7909.
19. Murphy CK. Combining belief functions when evidence conflicts. *Decision Support Syst*. 2000;29(1):1.
20. Deng Y, Shi W, Zhu Z, Liu Q. Combining belief functions based on distance of evidence. *Decision Support Syst*. 2004;38(3):489.
21. Yager RR. On the fusion of non-independent belief structures. *Int J General Syst*. 2009;38(5):505.
22. Li H, Xiao F. A method for combining conflicting evidences with improved distance function and Tsallis entropy. *Int J Intell Syst*. 2020;35(11):1814.
23. Liu F, Gao X, Zhao J, Deng Y. Generalized belief entropy and its application in identifying conflict evidence. *IEEE Access*. 2019;7(1):126625.
24. Li P, Fei L. On combination rule in Dempster-Shafer theory using OWA-based soft likelihood functions and its applications in environmental impact assessment. *Int J Intell Syst*. 2019;34(12):3168.
25. Pan L, Deng Y. An association coefficient of a belief function and its application in a target recognition system. *Int J Intell Syst*. 2020;35(1):85-104.
26. Deng Y. Information volume of mass function. *Int J Comput Commun Control*. 2020;15(6):3983.
27. Deng J, Deng Y. Information volume of fuzzy membership function. *Int J Comput Commun Control*. 2021; 16(1):4106.
28. Xue Y, Deng Y. Tsallis eXtropy. *Commun Statistics-Theory Methods*. 2021:1-14.
29. Benesty J, Chen J, Huang Y, Cohen I. Pearson correlation coefficient. *In Noise Reduction in Speech Processing*. Springer; 2009:1-4.
30. Monedero I, Biscarri F, León C, Guerrero JI, Biscarri J, Millán R. Detection of frauds and other non-technical losses in a power utility using Pearson coefficient, Bayesian networks and decision trees. *Int J Electrical Power Energy Syst*. 2012;34(1):90.
31. Wang G-J, Xie C, Stanley HE. Correlation structure and evolution of world stock markets: Evidence from Pearson and partial correlation-based networks. *Computat Econ*. 2018;51(3):607.
32. Myers JL, Well A, Lorch RF. *Research Design and Statistical Analysis*. Routledge; 2010.
33. Kendall MG. A new measure of rank correlation. *Biometrika*. 1938;30(1/2):81.
34. Xu H, Deng Y. Dependent evidence combination based on shearman coefficient and pearson coefficient. *IEEE Access*. 2017;6:11634.
35. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature*. 1998;393(6684):440.
36. Hofman JM, Sharma A, Watts DJ. Prediction and explanation in social systems. *Science*. 2017;355(6324): 486.
37. Gross JL, Yellen J. *Graph Theory and Its Applications*. CRC Press; 2005.

How to cite this article: Deng J, Deng Y, Cheong KH. Combining conflicting evidence based on pearson correlation coefficient and weighted graph. *Int J Intell Syst*. 2021;1-18. https://doi.org/10.1002/int.22593