The Fusion of Discrete Z-Numbers With Application for Fault Diagnosis

Ying Cao, Jian-Bo Yang, Xinyang Deng, and Wen Jiang

***Abstract*— A Z-number, which is a generalization of probability and fuzzy numbers, is a novel model to represent uncertainty in the real world. This article aims to develop discrete Z-numbers in both theory and application with the methodology of information fusion. First, a method for the fusion of discrete Z-numbers is proposed, in which the following problems are solved: 1) the extension principle and linear smoothing operator are used to obtain reasonable possibility distributions and 2) an optimization model based on the maximum entropy is constructed to obtain the most likely underlying probability. Thus, the relationship between the two components of Z-numbers is considered in our method. Second, the fusion of Z-numbers is applied to data-driven fault diagnosis, with which we show the potential of Z-numbers in real data in addition to linguistic information. The effectiveness of the proposed method is illustrated by the numerical experiments, where the two components of Z-numbers are generated according to their interpretations.**

***Index Terms*— Fault diagnosis, information fusion, uncertainty, Z-numbers.**

NOMEMCLATURE

$\mathcal{X}$	Set of points.
$\mu(\cdot)$	Membership degree.
$P(A)$	Probability measure of A.
P_X	Probability distribution of X.
$p(\cdot)$	Probability of an element.
$f(\cdot)$	Mathematical function.
D	Decision matrix.
COG	Center of gravity.
$\bar{X}$	Mean of a set of points.
σ	Standard deviation.
r_{mn}^l	Training pattern.
q^l	Test pattern.
MD	Mahalanobis distance matrix.
S	Matrix of matching degree.
Su	Matrix of support degree.
Cred_{mi}^l	Overall support degree.
NCred_{mi}^l	Normalized overall support degree.
score_i	Final evaluation of a test sample.

Manuscript received 4 February 2022; revised 1 June 2022; accepted 3 July 2022. Date of publication 22 July 2022; date of current version 24 August 2022. This work was supported by the National Natural Science Foundation of China under Program 62173272. The Associate Editor coordinating the review process was Dr. Damodar Reddy Edla. *(Corresponding author: Wen Jiang.)*

Ying Cao, Xinyang Deng, and Wen Jiang are with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: caoying123@mail.nwpu.edu.cn; xinyang.deng@nwpu.edu.cn; jiangwen@nwpu.edu.cn).

Jian-Bo Yang is with the Manchester Business School, The University of Manchester, Manchester M15 6PB, U.K. (e-mail: jian-bo.yang@manchester.ac.uk).

Digital Object Identifier 10.1109/TIM.2022.3193185

I. INTRODUCTION

UNCERTAINTY is a pervasive phenomenon in the real world, so it plays a crucial role when making decisions. Varieties of models, such as the probability [1], Z-numbers [2], [3], [4], fuzzy numbers [5], and evidence theory [6], [7], [8], [9], [10], have been proposed to describe different uncertainties. The focus of this article is the one which covers the following two components when modeling uncertainty: the first one is the uncertainty involved in the description or evaluation of an event, while the second one comes from the imperfection, i.e., the reliability, of the evaluation. In view of the two points, Zadeh [2] proposed the concept of Z-numbers. A Z-number is an ordered pair (A, B), where A is the restriction on the values an uncertain variable X may take, while B is the restriction on the probability measure of A, so it represents how reliable X is A. In this article, we focus on discrete Z-numbers.

Zadeh [2] first suggested some basic rules for the computation of Z-numbers. Thereafter, Yager [11] pointed out that answers of many problems regarding Z-numbers are dependent on the nature of underlying uncertainty of the associated variable, and then, he illustrated basic solutions by assuming some probability distributions. With Zadeh and Yager's interpretation, many scientists have made great theoretical contributions for Z-numbers: the arithmetic [12], approximate reasoning [13], [14], Z-differential equations [15], [16], linear algebra [17], nonlinear system [18], distance measure [19], [20], [21], uncertainty measure [22], utility [23], relative entropy [24], and so on. In addition, Z-numbers have wide applications, such as decision-making [25], [26], [27], [28], [29], failure mode and effect analysis [30], omnidirectional robot [31], and monitor drought [32]. This study aims to solve the following two problems associated with Z-numbers: 1) how to aggregate multiple Z-numbers without ignoring the underlying probability? and 2) can the fusion of Z-numbers be applied to data-driven fault diagnosis?

The first task is to develop the theory for the fusion of Z-numbers. As a general methodology, information fusion [33] for uncertainty refers to aggregating uncertainties from multiple sources to draw more comprehensive and accurate conclusions. The fusion of Z-numbers has great potential in artificial intelligence [34] in which the main task is to simulate human smart behaviors in machine systems. There are three main reasons: 1) people intend to consider information from different aspects to make better decisions and this is actually a process of information fusion; 2) Z-numbers are originally proposed to process the linguistic description provided by human and, thus, they are consistent with human thinking;

and 3) a Z-number, utilizing both information itself and its reliability, is a useful model to represent uncertainty, which plays an important role in artificial intelligence. For example, naive Bayes [35] and fuzzy C-means [36] are proposed to process probability and fuzziness, respectively. The above factors motivate us to study the fusion of Z-numbers.

In this article, we propose a method for the fusion of Z-numbers. Our method consists of two modules: 1) the arithmetic of component A, and in this module, the extension principle is adopted first for the computation of fuzzy numbers and, then, a linear smoothing operator is proposed to get a reasonable possibility distribution satisfying the definition of fuzzy numbers and 2) the computation of component B, where the extension principle and linear smoothing operator also applies. The main difficulty here is to utilize the underlying probability of the associated variable. For this problem, Aliev *et al.* [4] proposed a linear programming model, but they made assumptions about the underlying probability rather than finding it via computational models. With the basis of Aliev *et al.*'s [4] discussion and the maximum entropy principle, we construct an optimization model to find the most likely underlying probability distributions without extra assumption. The possible values of B can be easily computed afterward. A numerical example is provided to illustrate the process of the method and show how Z-numbers work with linguistic estimations.

Another task is the application of the fusion of Z-numbers to data-driven fault diagnosis [37], [38], [39], [40]. As mentioned before, Z-numbers are originally proposed to process the uncertainty involved in linguistic information, also called computing with words [2]. Thus, most previous works [13], [14], [25], [28], [32] study Z-numbers with human natural language (usually the estimations of experts). In this article, we show the following two aspects. First, Z-numbers can deal with not only human linguistic description but also the real data collected from nature. Specifically, the reliability of the information provided by a dataset is considered. Second, aggregating information involved in multiple Z-numbers helps improve the performance of the proposed method than the case without information fusion. To achieve the goals, we propose a new method for fault diagnosis based on the fusion of Z-numbers. Two main problems are solved: Z-number generation and Z-number fusion. For the former one, the two components of Z-numbers are generated from the original data according to their own interpretations. Specifically, component A is generated from the matching degree between the test and training patterns, while B depends on how much multiple training patterns support each other. As for the latter problem, it is consistent with the addition of Z-numbers. The numerical experiments demonstrate the effectiveness of the proposed method.

In this article, we use capital letters X, Y, and Z to represent variables whose values are Z-numbers. Lower case letters x, y, z, and b are used to denote elements belonging to a set. A and B denote the two fuzzy components of Z-numbers. Other main notations are listed in the Nomenclature. The rest of this article is organized as follows. Section II provides the basic knowledge of discrete fuzzy numbers, Z-numbers, and extension principles. Then, a method to achieve the fusion of Z-numbers is proposed in Section III. Section IV discusses how to apply the fusion of Z-numbers to fault diagnosis. Section V draws the conclusion.

II. Preliminaries

This section provides a review of the background knowledge associated with this work:

A. Discrete Fuzzy Numbers

Assume that $\mathcal{X}$ is a space of points (objects), with a generic element of $\mathcal{X}$ denoted by x. Thus, $\mathcal{X} = \{x\}$. In the classic set theory, the possibility that an element x belongs to a set A in $\mathcal{X}$ is 0 or 1, which can be expressed as

$$\mu_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A. \end{cases} \tag{1}$$

However, in fuzzy logic, a fuzzy set A in $\mathcal{X}$ is characterized by a membership function $\mu_A(x) : \mathcal{X} \to [0, 1]$, which maps $x \in \mathcal{X}$ to the interval $[0, 1]$, with the value of $\mu_A(x)$ that is also called possibility representing the grade of membership, also called the membership degree, of x in A. The closer $\mu_A(x)$ to 1, the higher the membership degree of x in A (i.e., the more possible x in A).The support of A is

$$\text{supp}(A) = \{x \in \mathcal{X} | \mu_A(x) \neq 0\} \tag{2}$$

and the α-level set of A is denoted by

$$A^\alpha = [\mu_A(x)]^\alpha = \{x \in X | \mu_A(x) \geq \alpha\}. \tag{3}$$

Thus, fuzzy sets extend the classic sets from Boolean logic $\{0, 1\}$ to the continuous interval $[0, 1]$.

A fuzzy number is a normalized, convex fuzzy set of the real line. There are two categories: discrete fuzzy numbers and the continuous ones. In this article, we only care about discrete fuzzy numbers, which can be defined as follows.

Definition 1 (Discrete Fuzzy Numbers [4]): A fuzzy subset A of the real line $\mathcal{R}$ with membership function $\mu_A : \mathcal{R} \to [0, 1]$ is a discrete fuzzy number if its support is finite, i.e., $\exists\ x_1, \ldots, x_n \in \mathcal{R}$ with $x_1 < x_2 < \cdots < x_n$, such that $\text{supp}(A) = \{x_1, \ldots, x_n\}$ and $\exists\ s$ and t that are natural numbers with $1 \leq s \leq t \leq n$ satisfying the following conditions.

1) $\mu_A(x_i) = 1$ for any natural number i with $s \leq i \leq t$.
2) $\mu_A(x_i) \leq \mu_A(x_j)$ for each natural numbers i and j with $1 \leq i \leq j \leq s$.
3) $\mu_A(x_i) \geq \mu_A(x_j)$ for each natural numbers i and j with $t \leq i \leq j \leq n$.

The three conditions expressed in Definition 1 demonstrate the convexity of discrete fuzzy numbers.

Note that fuzzy numbers and probability are two different models to represent uncertainty. The former one is used to describe fuzziness or ambiguity when there is no clear bound between two objects or classes, especially linguistic descriptions, such as the terms young and old, high and low, and about 45 and about 50, while the latter one describes uncertainty coming from randomness. The probability of a fuzzy number can be measured based on the associated membership and the underlying probability distribution, where the membership plays the role of discount factors.

Definition 2 (Probability Measure of a Discrete Fuzzy Number [4]): Assume a discrete fuzzy number $A = \{(\mu_A(x_i)/x_i)\}$ with $i = 1, \ldots, m$, and its associated probability distribution is $p_X = \{p(x_i)\}$ satisfying

$$\sum_{i=1}^{m} p(x_i) = 1 \tag{4}$$

and then, the probability measure of A denoted as $P(A)$ measuring the reliability of A is defined as

$$P(A) = \sum_{i=1}^{m} \mu_A(x_i) p(x_i). \tag{5}$$

B. Discrete Z-Numbers

Although the concept of fuzzy numbers introduces a new uncertainty, which is not covered by the traditional probability, it is still unclear for the following question: assume that A is the fuzzy evaluation of an event, how sure is it? In our real life, it is possible to receive information like the following descriptions: it is very likely that the anticipated budget deficit is close to 2 billion, it is absolutely that degree of Robert's honesty is very high, and it is likely that the oil price will be higher than 100 dollars. Taking the last one as an example, higher than 100 dollars can be quantified as a fuzzy number, but how to deal with the reliability of the evaluation described by likely? and what is the relationship between the two descriptions? To answer the question, Zadeh [2] generalized the concept of fuzzy numbers and probability to Z-numbers. Similar to fuzzy numbers, Z-numbers include both the discrete Z-numbers and continuous ones. This article focuses on the discrete type.

Definition 3 (Discrete Z-Numbers [4]): A discrete Z-number is an ordered pair of discrete fuzzy numbers denoted as: X is (A, B), where A plays the role of the fuzzy restriction on values that a variable X may take with a membership function $\mu_A : \{x_1, x_2 \ldots, x_m\} \to [0, 1]$: X is A and B is the fuzzy restriction on the probability measure of A with a membership function $\mu_B : \{b_1, \ldots, b_n\} \to [0, 1]$ and $\{b_1, \ldots, b_n\} \subset [0, 1]$: $P(X\ is\ A)$ is B.

The ordered triple (X, A, B) is referred to as a Z-valuation, which is also equivalent to an assignment statement: X is (A, B). Recall that the linguistic information provided by an expert: it is likely that the oil price will be higher than 100 dollars; then, this can be represented by a Z-valuation: (the oil price, higher than 100 dollars, likely), and two fuzzy numbers are used to quantify higher than 100 dollars and likely.

In a Z-number $X = (A, B)$, A is a fuzzy number, which can be represented by

$$\left\{\frac{\mu_A(x_1)}{x_1}, \frac{\mu_A(x_2)}{x_2}, \ldots, \frac{\mu_A(x_m)}{x_m}\right\} \tag{6}$$

where $\{x_i\}$ with $i = 1, \ldots, m$ are possible values of X and $\{\mu_A(x_i)\}$ is the possibility distribution, also called the membership degree, representing to what degree x_i belongs to A. In addition, B is a fuzzy number

$$\left\{\frac{\mu_B(b_1)}{b_1}, \frac{\mu_B(b_2)}{b_2}, \ldots, \frac{\mu_B(b_n)}{b_n}\right\} \tag{7}$$

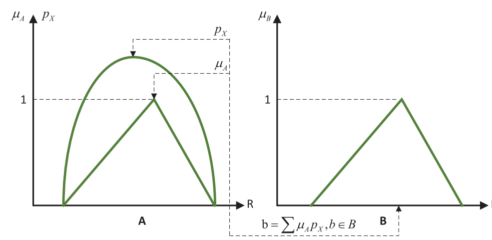


Fig. 1. Meaning of Z-numbers.

which describes the reliability of A, b_j with $j = 1, \ldots, n$ are possible values of $P(A)$, and $\{\mu_B(b_j)\}$ is the associated possibility distribution. Furthermore, the probability measure of A denoted by $P(A)$ can be defined as: $P(A)$ is $\sum_{i=1}^{n} \mu_A(x_i)p(x_i)$. Here, it is important to mention that $p_X = \{p(x_i)\}$ with $i = 1, 2, \ldots, m$ called the underlying (hidden) probability of the variable X is unknown. What is known is p_X satisfies (4) and it has the restriction, which can be expressed as: $\sum_{i=1}^{n} \mu_A(x_i)p(x_i)$ is B from which we know that B is the fuzzy restriction on the probability measure of A. The definition of Z-numbers can be described in Fig. 1, in which the two components of Z-numbers are linked by the underlying probability.

A concept that is closely related to the concept of a Z-numbers is Z^+-numbers. The underlying probability of the variable X, which is unknown in a Z-number, is known in a Z^+-number. A Z^+-number is an ordered pair (A, R), where A plays the same role as it does in a Z-number, while R representing $P(A)$ is a real number as the underlying probability is known with which $P(A)$ can be computed by (5). Similar to Z-numbers, a Z^+-valuation can also be written as: X is (A, p_X). Note that $P(A)$, which follows (5), is the scalar product of μ_A and p_X, and this links the concept of Z-numbers and Z^+-numbers, which can be described as

$$Z(A, B) = Z^+(A, P(A) \text{ is } B)$$

where what is known in a Z-number is not p_X, but the restriction associated with it expressed as: $P(A)$ is B. In fact, a Z^+-number can be viewed as a specific case of a Z-number when the underlying probability is known.

C. Extension Principle

What the computation of fuzzy numbers and Z-numbers could follow is the extension principle of fuzzy logic [41]. Basically, the term extension principle is proposed to describe a rule in which what are known are not the values of arguments but the restrictions on the value of arguments [2]. Here, we list some basic versions of the extension principle.

A basic version related to the computation of fuzzy numbers can be described as

$$\frac{\begin{array}{l} Z = f(X, Y) \\ \mathcal{R}(X) : X \text{ is } \left\{\frac{\mu_{A_1}(x_1)}{x_1}, \ldots, \frac{\mu_{A_1}(x_{n_1})}{x_{n_1}}\right\} \\ \mathcal{R}(Y) : Y \text{ is } \left\{\frac{\mu_{A_2}(y_1)}{y_1}, \ldots, \frac{\mu_{A_2}(y_{n_2})}{y_{n_2}}\right\} \end{array}}{\begin{array}{l} \mathcal{R}(Z) : \mu_A(z_i) = \sup_{x_j, y_k}(\mu_{A_1}(x_j) \wedge \mu_{A_2}(y_k)) \\ \text{s.t. } z_i = f(x_j, y_k) \end{array}} \tag{8}$$

in which $\mathcal{R}(\cdot)$ means the restriction of possible values of a variable, and $\mu_A(z_i)$, $\mu_{A_1}(x_j)$, and $\mu_{A_2}(y_k)$ are the restrictions on the values of Z, X, and Y, respectively.

In probabilistic restrictions, the general extension principle leading to results coincides with standard results relating to the functions of probability distributions [42], which can be described as

$$\frac{\begin{array}{l} Z = f(X, Y) \\ \mathcal{R}(X): X \text{ isp } p_X,\, p_X = p(x_1)\backslash x_1 + \cdots + p(x_{n_1})\backslash x_{n_1} \\ \mathcal{R}(Y): Y \text{ isp } p_Y,\, p_Y = p(y_1)\backslash y_1 + \cdots + p(y_{n_2})\backslash y_{n_2} \end{array}}{\begin{array}{l} \mathcal{R}(Z): p_Z(z_i) = \sum_{j,k} p(x_j)p(y_k) \\ \text{s.t. } z_i = f(x_j, y_k) \end{array}} \tag{9}$$

where isp means probability distribution is, and $p_Z = p_X \circ p_Y$ is the convolution, a mathematical operation on two functions producing the third function, of p_X and p_Y [4], [13].

With the basis of the extension principle of fuzzy and probabilistic restrictions, Zadeh [2] constructed the extension rule for Z-numbers

$$\frac{\begin{array}{l} Z = f(X, Y) \\ X \text{ is } (A_X, B_X) \text{ (restriction on } X) \\ Y \text{ is } (A_Y, B_Y) \text{ (restriction on } Y) \\ Z \text{ is } (A_Z, B_Z) \text{ (induced restriction on } Z) \end{array}}{\begin{array}{l} A_Z = f(A_X, A_Y) \\ B_Z = \mu_{A_Z} \cdot (p_X \circ p_Y) \end{array}} \tag{10}$$

in which $\cdot$ means the scalar product and p_X and p_Y are constrained by

$$\begin{cases} \sum p(x_j) = 1 \\ \sum p(y_k) = 1 \\ \sum p(x_j)\mu_{A_X}(x_j) \text{ is } B_X \\ \sum p(y_k)\mu_{A_Y}(y_k) \text{ is } B_Y. \end{cases} \tag{11}$$

In (10), the possibility distribution of Z can be obtained by

$$\mu_{A_Z}(z_i) = \sup_{x_j, y_k}(\mu_{A_X}(x_j) \wedge \mu_{A_Y}(y_k)) \quad \text{s.t. } z_i = f(x_j, y_k) \tag{12}$$

and the possibility of the probability of Z is

$$\mu_{p_Z}(p(z_i)) = \sup_{p_X, p_Y}\left(\mu_{B_x}\left(\sum \mu_{A_X}(x_j)p(x_j)\right) \wedge \mu_{B_Y}\left(\sum \mu_{B_Y}(y_k)p(y_k)\right)\right). \tag{13}$$

III. Fusion of Z-Numbers

In this section, we propose a method to achieve the fusion of Z-numbers. Variables in classic mathematics are accurate, and then, they can be aggregated in mathematical functions via various basic arithmetic operations, including addition, subtraction, multiplication, division, and power. For example, if the variable X is 1 and Y is 1, then we have: $X + Y = 2$. In this operation, the information involved in X and Y is aggregated by the addition operation, the associated function is: $f(X, Y) = X + Y$. In this section, mathematical functions will be extended from accurate variables to Z-valued ones.

The purpose of the fusion of Z-numbers is to integrate information involved in multiple Z-numbers, and this can be done by Z-valuation-based mathematical functions, that is, the following function is considered:

$$Z = f(Z_1, Z_2, \ldots, Z_K) \tag{14}$$

where $Z_1, Z_2, \ldots, Z_K$ and Z are all variables whose values are Z-numbers. Here, we only discuss the situation when there are two Z-valued variables in f, and multivariable situation follows the same computational rules. Assume two Z-valued variables X and Y: X is (A_1, B_1) and Y is (A_2, B_2), where

$$A_1 = \left\{\frac{\mu_{A_1}(x_1)}{x_1}, \frac{\mu_{A_1}(x_2)}{x_2}, \ldots, \frac{\mu_{A_1}(x_{n_1})}{x_{n_1}}\right\} \tag{15}$$

$$A_2 = \left\{\frac{\mu_{A_2}(y_1)}{y_1}, \frac{\mu_{A_2}(y_2)}{y_2}, \ldots, \frac{\mu_{A_2}(y_{n_2})}{y_{n_2}}\right\} \tag{16}$$

$$B_1 = \left\{\frac{\mu_{B_1}(b_{11})}{b_{11}}, \frac{\mu_{B_1}(b_{12})}{b_{12}}, \ldots, \frac{\mu_{B_1}(b_{1m_1})}{b_{1m_1}}\right\} \tag{17}$$

$$B_2 = \left\{\frac{\mu_{B_2}(b_{21})}{b_{21}}, \frac{\mu_{B_2}(b_{22})}{b_{22}}, \ldots, \frac{\mu_{B_2}(b_{2m_2})}{b_{2m_2}}\right\}. \tag{18}$$

The task is to compute the variable $Z = (A, B)$, which has the following mathematical relationship with X and Y:

$$Z = f(X, Y). \tag{19}$$

The whole framework of the fusion of Z-numbers follows from (10) includes two modules: the fusion of the fuzzy evaluation of events (i.e., the first components of Z-numbers) and the computation of reliability (i.e., the second components of Z-numbers). Details are discussed in the following.

A. Integration of the First Components in Z-Numbers

As the first component in a Z-number is a fuzzy number, the integration of the fuzzy evaluation of events in Z-numbers is equivalent to the arithmetic of fuzzy numbers. The main task in the computation of discrete fuzzy numbers is to obtain possibility distributions (i.e., membership degree). There are two common approaches: the extension principle-based [43] and α-level-set-based [44] methods. In this article, the former one is adopted. The extension principle-based approach is defined as follows.

Definition 4: Assume that two discrete fuzzy numbers denoted as A_1 and A_2 are

$$A_1 = \left\{\frac{\mu_{A_1}(x_1)}{x_1}, \frac{\mu_{A_1}(x_2)}{x_2}, \ldots, \frac{\mu_{A_1}(x_{n_1})}{x_{n_1}}\right\} \tag{20}$$

and

$$A_2 = \left\{\frac{\mu_{A_2}(y_1)}{y_1}, \frac{\mu_{A_2}(y_2)}{y_2}, \ldots, \frac{\mu_{A_2}(y_{n_2})}{y_{n_2}}\right\}. \tag{21}$$

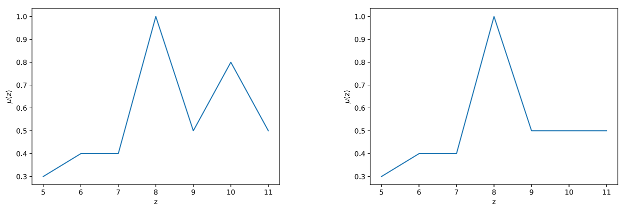
Then, their computational result A satisfying $A = f(A_1, A_2)$ is

$$\text{supp}(A) = \{z \in f(x, y) | \min(f(x, y)) \le z \le \max(f(x, y))\} \tag{22}$$

and the associated membership degree is

$$\mu_A(z) = \sup_{x,y}(\mu_{A_1}(x) \wedge \mu_{A_2}(y)) \quad \text{s.t. } z = f(x, y). \tag{23}$$

As shown in (22), element z in A is obtained by the computational result of pairs (x, y) in which x and y belong



(a) (b)

Fig. 2. (a) Possibility distribution of A without the linear smoothing operator. (b) Possibility distribution of A with the linear smoothing operator.

to $\text{supp}(A_1)$ and $\text{supp}(A_2)$, respectively. The associated membership degree is obtained by (23) that follows the extension rule. However, it might obtain the computational result A not satisfying the convexity of fuzzy numbers. Here, an example is used to illustrate the limitation.

Example 3.1: Assume that we have two fuzzy numbers

$$A_1 = \left\{\frac{0.3}{1}, \frac{1}{2}, \frac{0.5}{3}\right\} \tag{24}$$

$$A_2 = \left\{\frac{0.4}{4}, \frac{1}{6}, \frac{0.8}{8}\right\}. \tag{25}$$

Then, with the basis of (22) and (23), the sum of A_1 and A_2 denoted by A can be obtained

$$A = \left\{\frac{0.3}{5}, \frac{0.4}{6}, \frac{0.4}{7}, \frac{1}{8}, \frac{0.5}{9}, \frac{0.8}{10}, \frac{0.5}{11}\right\}. \tag{26}$$

The possibility distribution of A is shown in Fig. 2(a).

As shown in Fig. 2, A does not satisfy the convexity in Definition 1, and thus, A is not a fuzzy number. To avoid the drawback, Wang *et al.* [44] defined operations on fuzzy numbers based on α-level sets. By using a linear smoothing method, Casasnovas and Riera [45] generalized Wang *et al.*'s [44] result as a family of possibility distributions in which Wang *et al.*'s [44] result is just the lower bound of it. No matter in which approach the key of the arithmetic of fuzzy numbers is to obtain a reasonable possibility distribution satisfying the three conditions in Definition 1. The extension-based approach has been widely adopted in Z-numbers due to its simpleness and practicability [2], [11], [13], [46], [47], but it may cause unreasonable possibility distributions. α-level-set-based approaches can obtain a reasonable membership degree. However, the result is sensitive to the value of α. This makes it not practical in the computation of component B of Z-numbers. In Casasnovas and Riera's [45] method, a family of possible membership degree is involved between the lower and upper bounds, and this makes it not practical in real applications and the computation of component B. To avoid these inconveniences and inspired by Casasnovas and Riera [45], a new method combining the extension principle and linear smoothing operator is proposed in this article. It consists of two steps.

1) Extension Principle-Based Module: This section is totally consistent with Definition 4. The possible values belonging to X and Y are computed with (22). Then, the associated membership degree can be obtained via (23).

2) Linear Smoothing Operator: After step 1), a primary representation of A denoted by A^p can be obtained

$$A^p = \left\{\frac{\mu_A^p(z_1)}{z_1}, \frac{\mu_A^p(z_2)}{z_2}, \dots, \frac{\mu_A^p(z_n)}{z_n}\right\} \tag{27}$$

where $z_1 < z_2 < \cdots < z_n$ and μ_A^p is used to represent the primary possibility distribution (i.e., membership degree). However, A^p may not satisfy the convexity as shown in Example 3.1. Therefore, it is necessary to change the primary extension-based possibility distribution to satisfy the convexity of fuzzy numbers. For this issue, we utilize the original extension-based possibility distribution as much as possible with the constraint that the properties of fuzzy numbers are satisfied. Specifically, the methodology is to find the possibility distribution denoted by μ_A that is closest to μ_A^p and satisfy the convexity of fuzzy numbers at the same time. Here, the similarity between μ_A and μ_A^p is measured by the utility $U = \{u_1, u_2, \dots, u_n\}$ where

$$u_i = \begin{cases} 1, & \text{if } \mu_A(x_i) = \mu_A^p(x_i) \\ 0, & \text{if } \mu_A(x_i) \neq \mu_A^p(x_i) \end{cases} \tag{28}$$

and then, the objective is

$$\begin{aligned} &\arg\max_{\mu_A} \sum_{i=1}^{n} u_i \\ &\text{s.t.} \begin{cases} \mu_A(z_i) \le \mu_A(z_j), & 1 \le i \le j \le s \\ \mu_A(z_i) = 1, & s \le i \le t \\ \mu_A(z_i) \ge \mu_A(z_j), & t \le i \le j \le n. \end{cases} \end{aligned} \tag{29}$$

Before solving (29), a linear smoothing operator, which is inspired by the linear operation in [45], is proposed.

Definition 5: Assume a monotone sequence associated with a fuzzy number

$$A = \left\{\frac{\mu_A(z_1)}{z_1}, \frac{\mu_A(z_2)}{z_2}, \dots, \frac{\mu_A(z_n)}{z_n}\right\} \tag{30}$$

where $z_1 < \cdots < z_n$ and $u_A(z_i)$ satisfy one of the following two monotone conditions:

$$\mu_A(z_1) \le \mu_A(z_2) \le \cdots \le \mu_A(z_n) \tag{31}$$

or

$$\mu_A(z_1) \ge \mu_A(z_2) \ge \cdots \ge \mu_A(z_n). \tag{32}$$

Then, for a new element $z_j \in (z_i, z_{i+1})$, i.e., $z_i < z_j < z_{i+1}$ where $z_i, z_{i+1} \in \text{supp}(A)$, its membership degree $\mu_A(z_j)$ is smoothed as

$$\mu_A(z_j) = \frac{\mu_A(z_{i+1}) - \mu_A(z_i)}{z_{i+1} - z_i}(z_j - z_i) + \mu_A(z_i). \tag{33}$$

Note that with the linear operation (33), it is easy to verify that the new element z_j does not destroy the convexity. Specifically, we have

$$\begin{cases} \mu_A(z_i) \le \mu_A(z_j) \le \mu_A(z_{i+1}), & \text{if } \mu_A(z_i) \le \mu_A(z_{i+1}) \\ \mu_A(z_i) \ge \mu_A(z_j) \ge \mu_A(z_{i+1}), & \text{if } \mu_A(z_i) \ge \mu_A(z_{i+1}). \end{cases} \tag{34}$$

With (33) and (34), (29) can be solved by the following process. Assume that for the fuzzy number (27), $\mu(z_i) = 1$ with $1 \le s \le i \le t \le n$. Taking z_s and z_t as boundary points, then the whole sequence can be divided into three parts: $z_1 \leq z_i \leq z_s$, $z_s \leq z_i \leq z_t$, and $z_t \leq z_i \leq z_n$:

For $z_1 \leq z_i \leq z_s$, we first find the longest nondecreasing subsequence of $\{\mu_A^p(z_1), \ldots, \mu_A^p(z_s)\}$, denoted by $\{\mu_A^p(z_1^*), \ldots, \mu_A^p(z_d^*)\}$ with $d \leq s$, $z_1^* < \cdots < z_d^*$ and $\mu_A^p(z_1^*) \leq \cdots \leq \mu_A^p(z_d^*)$. Note that here, z_i^* is used to distinguish the nondecreasing subsequence and the original possibility sequence, and $\{z_1^*, \ldots, z_d^*\} \subseteq \{z_1, \ldots, z_s\}$. Then, we have

$$\mu_A\left(z_j^*\right) = \mu_A^p\left(z_j^*\right) \tag{35}$$

where $1 \leq j \leq d$, that is, there is no change with the membership degree belonging to the longest nondecreasing subsequence. Then, the possibility of elements satisfying $z_1 \leq z_i \leq z_s$, but $z_i \notin \{z_1^*, z_2^*, \ldots, z_d^*\}$ is smoothed according to (33). After inserting the smoothed elements, $\{\mu_A(z_1^*), \ldots, \mu_A(z_d^*)\}$ can be extended to $\{\mu_A(z_1), \ldots, \mu_A(z_s)\}$. According to (34), we know $z_1 < \cdots < z_s$ and $\mu_A(z_1) \leq \cdots \leq \mu_A(z_s)$. Thus, this operation can make the possibility distribution of the first part monotonically increase.

As for elements $z_i \in [z_t, z_n]$, the associated membership degree follows similar procedures: the longest nonincreasing subsequence of $\{\mu_A^p(z_t), \ldots, \mu_A^p(z_n)\}$ is found at first. Then, the possibility of not belonging to the subsequence is smoothed by (33), and this makes the membership degree monotonically decrease. Therefore, the possibility distribution $\{\mu_A^p(z_t), \ldots, \mu_A^p(z_n)\}$ is smoothed as a nonincreasing sequence: $\{\mu_A(z_t), \ldots, \mu_A(z_n)\}$ with $z_t < z_{t+1} < \cdots < z_n$ and $\mu_A(z_t) \geq \mu_A(z_{t+1}) \geq \cdots \geq \mu_A(z_n)$. In addition, there is no change for $z_s \leq z_i \leq z_t$ as their membership degrees are all equal to 1.

Finally, combining the three smoothed parts, the reasonable possibility distribution $\{\mu_A(z_1), \mu_A(z_2), \ldots, \mu_A(z_n)\}$ satisfying the convexity in Definition 1 can be obtained. In short, the linear smoothing operator guarantees the convexity of the obtained discrete fuzzy number according to (34). The whole process of the arithmetic of fuzzy numbers is shown in Algorithm 1, and the linear smoothing operator is demonstrated by Algorithm 2. With the proposed linear smoothing operator, the computational result in Example 3.1 is

$$A = \left\{\frac{0.3}{5}, \frac{0.4}{6}, \frac{0.4}{7}, \frac{1}{8}, \frac{0.5}{9}, \frac{0.5}{10}, \frac{0.5}{11}\right\} \tag{36}$$

which is shown in Fig. 2(b). Obviously, only $\mu_A(10)$ is modified, and thus, A keeps the original extension principle-based possibility distribution as much as possible on the premise of satisfying convexity.

Algorithm 1 Arithmetic of Fuzzy Numbers

Input: $A_1 = \{\frac{\mu_{A_1}(x_1)}{x_1}, \frac{\mu_{A_1}(x_2)}{x_2}, \ldots, \frac{\mu_{A_1}(x_{n_1})}{x_{n_1}}\}$,
$A_2 = \{\frac{\mu_{A_2}(y_1)}{y_1}, \frac{\mu_{A_2}(y_2)}{y_2}, \ldots, \frac{\mu_{A_2}(y_{n_2})}{y_{n_2}}\}$
Output: $A = f(A_1, A_2) = \{\frac{\mu_A(z_1)}{z_1}, \frac{\mu_A(z_2)}{z_2}, \ldots, \frac{\mu_A(z_n)}{z_n}\}$
Computing the primary $\mathrm{supp}(A) = \{z_1, z_2 \ldots, z_n\}$ and $\mu_A^p = \{\mu_A^p(z_1), \mu_A^p(z_2), \ldots, \mu_A^p(z_n)\}$ with (22) and (23) respectively;
Finding the boundary points satisfying $\mu_A(z_i) = 1$ where $s \leq i \leq t$;
Smoothing $\{\mu_A^p(z_1), \ldots, \mu_A^p(z_s)\}$ as a monotone non-decreasing sequence $\{\mu_A(z_1), \ldots, \mu_A(z_s)\}$ with **algorithm 2**;
Smoothing $\{\mu_A^p(z_t), \ldots, \mu^p(z_n)\}$ as a monotone non-increasing sequence $\{\mu_A^p(z_t), \ldots, \mu^p(z_n)\}$ with **algorithm 2**;
Combining the three parts: $\mu_A = \{\mu_A(z_1), \ldots, \mu_A(z_s)\} \cup \{\mu_A(z_{s+1}), \ldots, \mu_A(z_t)\} \cup \{\mu_A(z_{t+1}), \ldots, \mu_A(_n)\}$;
return $A = \frac{\mu_A}{\mathrm{supp}(A)}$.

Algorithm 2 Linear Smoothing Operator

Input: $A^p = \{\frac{\mu_A^p(z_1)}{z_1}, \ldots, \frac{\mu_A^p(z_n)}{z_n}\}$ with $z_1 < \ldots < z_n$
Output: $A = \{\frac{\mu_A(z_1)}{z_1}, \ldots, \frac{\mu_A(z_n)}{z_n}\}$
if $\mu(z_1) = 1$ **then**
  Finding the longest non-increasing subsequence of μ_A^p: $\{\mu_A^p(z_1^*), \mu_A^p(z_2^*), \ldots, \mu_A^p(z_d^*)\}$ where $d \leq n$
else
  Finding the longest non-decreasing subsequence of μ_A^p: $\{\mu_A^p(z_1^*), \mu_A^p(z_2^*), \ldots, \mu_A^p(z_d^*)\}$ where $d \leq n$
end if
for $i = 1 \rightarrow n$ **do**
  if $z_i \in \{z_1^*, z_2^*, \ldots, z_d^*\}$ **then**
    $\mu_A(z_i) \leftarrow \mu_A^p(z_i)$
  else
    Smoothing $\mu_A^p(z_i)$ as $\mu_A(z_i)$ with (33)
  end if
end for
return $\{\frac{\mu_A(z_1)}{z_1}, \frac{\mu_A(z_2)}{z_2}, \ldots, \frac{\mu_A(z_n)}{z_n}\}$

B. Computation of the Second Component in Z-Numbers

As for the computation of component B, following the definition of Z-numbers, the fusion of reliability includes three steps.

1) Searching for the Underlying Probability: According to Definition 3, both fuzziness and probability should be considered in the computation of Z-numbers. In addition, we know that A and B in Z-numbers are not independent of each other but linked by the underlying probability distribution. Specifically, the elements in B are computed by the underlying probability and the possibility distribution of A according to (5). Inspired by [4], we first obtain (37) for X based on Definition 3 and (4) and (5)

$$\begin{cases} p_1(x_1) + p_1(x_2) + \cdots + p_1(x_{n_1}) = 1 \\ \mu_{A_1}(x_1)p_1(x_1) + \cdots + \mu_{A_1}(x_{n_1})p_1(x_{n_1}) = b_{11} \end{cases}$$
$$\vdots$$
$$\begin{cases} p_{m_1}(x_1) + p_{m_1}(x_2) + \cdots + p_{m_1}(x_{n_1}) = 1 \\ \mu_{A_1}(x_1)p_{m_1}(x_1) + \cdots + \mu_{A_1}(x_{n_1})p_{m_1}(x_{n_1}) = b_{1m_1} \end{cases} \tag{37}$$

where each group of probability distribution of X, denoted by $p_{Xl_1} = \{p_{l_1}(x_1), \ldots, p_{l_1}(x_{n_1})\}$, has a membership degree $\mu(p_{Xl_1})$ that is actually given by b_{1l_1}, i.e.,

$$\mu(p_{Xl_1}) = \mu_{B_1}(b_{1l_1}) \tag{38}$$

with $l_1 = 1, \ldots, m_1$. Similarly, for Y, we have

$$\begin{cases} p_1(y_1) + p_1(y_2) + \cdots + p_1(y_{n_2}) = 1 \\ \mu_{A_2}(y_1)p_1(y_1) + \cdots + \mu_{A_2}(y_{n_2})p_1(y_{n_2}) = b_{21} \end{cases}$$
$$\vdots$$
$$\begin{cases} p_{m_2}(y_1) + p_{m_2}(y_2) + \cdots + p_{m_2}(y_{n_2}) = 1 \\ \mu_{A_2}(y_1)p_{m_2}(y_1) + \cdots + \mu_{A_2}(y_{n_2})p_{m_2}(y_{n_2}) = b_{2m_2} \end{cases} \quad (39)$$

and

$$\mu(p_{Yl_2}) = \mu_{B_2}(b_{2l_2}) \quad (40)$$

with $l_2 = 1, \ldots, m_2$. However, it is clear that infinite solutions satisfy the two equalities in (39) for each b_{2l_2}, which means that infinite probability distributions of Y are possible. The probability of X in (37) also follows the similar case. This makes the computation of Z-numbers difficult, and now, the question is: how to utilize the unknown probabilistic information with the given conditions? To solve the problem, the maximum entropy principle [1], which states that the probability distribution, which best represents the current state of knowledge about a system is the one with the largest entropy, is adopted to find the most likely probability distribution of X for each b_{1l_1}

$$\begin{aligned} &\arg\max_{p_X} \ -\sum_j p_{l_1}(x_j)\log_2 p_{l_1}(x_j) \\ &\text{s.t.} \begin{cases} \sum_{j=1}^{n_1} p_{l_1}(x_j) = 1 \\ \sum_{j=1}^{n_1} p_{l_1}(x_j)\mu_{A_1}(x_j) = b_{1l_1} \leftrightarrow \mu_{B_1}(b_{1l_1}) \\ 0 \le p_{l_1}(x_j) \le 1 \end{cases} \end{aligned} \quad (41)$$

where $\leftrightarrow$ is used to emphasize $\mu_{B_1}(b_{1l_1})$. In (41), the three constraints follow from the definition of Z-numbers, and no extra assumption is made. Note that the maximum entropy model tends to generate the uniform probability distribution, but the second constraint of (41) originated from (5) affects the trend, that is, if b_{1l_1} is improved, elements with larger possibility can be assigned more probability compared with smaller b_{1l_1} and vice versa. Furthermore, (41) has a solution if and only if

$$b_{1l_1} \ge \min_j \mu_A(x_j). \quad (42)$$

As $l_1 = 1, \ldots, m_1$, m_1 groups of probability distribution can be obtained for X, with each group denoted by $p_{Xl_1} = \{p_{l_1}(x_1), \ldots, p_{l_1}(x_{n_1})\}$. What is more, $\mu(p_{Xl_1})$ can be obtained by (38). Similarly, for Y, we have

$$\begin{aligned} &\arg\max_{p_Y} \ -\sum_k p_{l_2}(y_k)\log_2 p_{l_2}(y_k) \\ &\text{s.t.} \begin{cases} \sum_{k=1}^{n_2} p_{l_2}(y_k) = 1 \\ \sum_{k=1}^{n_2} p_{l_1}(y_k)\mu_{A_2}(y_k) = b_{2l_2} \leftrightarrow \mu_{B_2}(b_{2l_2}) \\ 0 \le p_{l_2}(y_k) \le 1 \end{cases} \end{aligned} \quad (43)$$

with $l_2 = 1, \ldots, m_2$. Then, m_2 groups of probability distribution can be obtained for Y, with each group denoted by $p_{Yl_2} = \{p_{l_2}(y_1), \ldots, p_{l_2}(y_{n_2})\}$. Meanwhile, $\mu(p_{Yl_2})$ can be obtained by (40). Then, with (9), the probability of Z can be derived with the convolution of p_{Xl_1} and p_{Yl_2}

$$\begin{aligned} p_l(z_i) &= \sum_j \sum_k p_{l_1}(x_j)p_{l_2}(y_k) \\ &\text{s.t. } z_i = f(x_j, y_k) \end{aligned} \quad (44)$$

where $i = 1, \ldots, n$, $j = 1, \ldots, n_1$, $k = 1, \ldots, n_2$, $l_1 = 1, \ldots, m_1$, and $l_2 = 1, \ldots, m_2$. Note that with (44), m_3 groups of probability distributions for Z denoted by $p_{Zl} = \{p(z_1), p(z_2), \ldots, p(z_n)\}$ can be obtained, where $l = 1, \ldots, m_3$ and $m_3 \le m_1 \times m_2$. The reason why $m_3 \le m_1 \times m_2$ is that it is possible to get the same p_{Zl} with different pairs of p_{Xl_1} and p_{Yl_2}. Moreover, similar to p_X and p_Y, each p_{Zl} has a membership degree that can be derived according to the extension rule

$$\begin{aligned} &\mu(p_{Zl}) = \sup_{p_{Xl_1}, p_{Yl_2}} (\mu(p_{Xl_1}) \wedge \mu(p_{Yl_2})) \\ &\text{s.t.} \begin{cases} \sum_{j=1}^{n_1} p_{l_1}(x_j) = 1 \\ \sum_{k=1}^{n_2} p_{l_2}(y_k) = 1 \\ \mu(p_{Xl_1}) = \mu_{B_1}(b_{1l_1}) \\ \mu(p_{Yl_2}) = \mu_{B_2}(b_{2l_2}) \\ p_{Zl} = p_{Xl_1} \circ p_{Yl_2} \end{cases} \end{aligned} \quad (45)$$

which is extended from (13).

2) Computation of Elements in Component B: With m_3 groups of most likely probability distributions, the possibility distribution of Z, and (5), elements in component B can be computed by

$$b_l = \sum_i \mu_A(z_i)p_l(z_i) \quad (46)$$

where $i = 1, \ldots, n$ and $l = 1, \ldots, m$. As for the value of m which is the total number of b_l, it is originally equal to m_3 as each group of probability distribution can get one element in B. However, with (46), equivalent b_l may be obtained by two different p_{Zl}'s. Therefore, the total number of b_l denoted as m satisfies $m \le m_3$. Furthermore, the possibility of b_l can be obtained with the extension rule

$$\begin{aligned} &\mu_B(b_l) = \sup_{p_{Zl}} \ \mu(p_{Zl}) \\ &\text{s.t. } \sum_{i=1}^{n} p_l(z_i)\mu_A(z_i) = b_l. \end{aligned} \quad (47)$$

Above all, the computation of membership degree is a possibility transmission process, which can be described as

$$\left.\begin{aligned} \mu_{B_1}(b_{1l_1}) \Rightarrow \mu(p_{Xl_1}) \\ \mu_{B_2}(b_{2l_2}) \Rightarrow \mu(p_{Yl_2}) \end{aligned}\right\} \Rightarrow \mu(p_{Zl}) \Rightarrow \mu_B(b_l). \quad (48)$$

As shown in (48), b_{1l_1} and b_{2l_2} transmit their possibility to the underlying probability denoted by p_{Xl_1} and p_{Yl_2}, respectively, and then, p_{Xl_1} and p_{Yl_2} give their membership degree to p_{Zl} via (44). Finally, b_l receives its final possibility from p_{Zl} by (46).

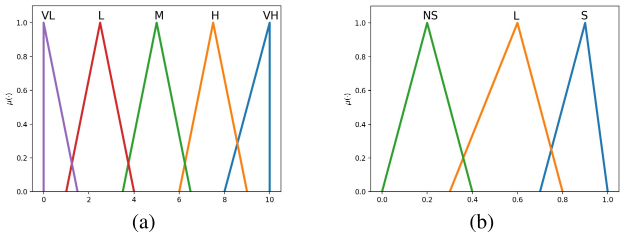


Fig. 3. (a) Fuzzy numbers for the five linguistic terms of reliability. (b) Fuzzy numbers for the five linguistic terms of risk.

3) Smooth the Possibility Distribution of B: It is possible that the possibility distribution of B obtained from step 2) does not satisfy the convexity of fuzzy numbers. To solve the problem, the linear smoothing operator illustrated in Algorithm 2 is adopted again. The smoothing process of the possibility distribution of B is the same as that in the arithmetic of component A: The original B is divided into three parts by taking elements whose possibility is equal to 1 as the boundary points. Then, with Algorithm 2, the possibility of elements less than the boundary points can be smoothed as nondecreasing series, and the membership degree of elements bigger than boundary points can be smoothed as nonincreasing series. This guarantees the convexity of B.

Z-numbers are generally used to quantify the fuzzy linguistic descriptions, and more examples can be found from [2]. Here, we provide a numerical example to illustrate the details about how to aggregate Z-numbers.

Example 3.2: Assume that two experts are invited to estimate the risk of a strategy for a company, and there are five different linguistic levels for the risk: very high (VH), high (H), medium (M), low (L), and very low (VL). Moreover, three reliability levels are provided to show how sure the experts are about their estimations: sure (S), likely (L), and not sure (NS). Assume that the two experts give the following two Z-number-based estimations: (VH, L) and (H, S), where the former one is equivalent to: I am sure that the strategy would bring high risk, and the latter one is: it is likely the strategy can cause very high risk. One general method to integrate the two experts' information is to average them. Here, we use [1, 10] to quantify the five risk levels and [0, 1] to quantify the three reliability levels, and the triangular fuzzy numbers for them are shown in Fig. 3, with which the two estimations can be modeled as the following two discrete Z-numbers:

$$X = (A_1, B_1), \quad Y = (A_2, B_2) \tag{49}$$

where

$$\begin{aligned} A_1 &= \left\{\frac{0.4}{8.8}, \frac{0.6}{9.2}, \frac{0.8}{9.6}, \frac{1}{10}\right\} \\ B_1 &= \left\{\frac{0.6}{0.48}, \frac{0.8}{0.54}, \frac{1}{0.6}, \frac{0.8}{0.64}, \frac{0.6}{0.68}\right\} \\ A_2 &= \left\{\frac{0.6}{6.9}, \frac{0.8}{7.2}, \frac{1}{7.5}, \frac{0.8}{7.8}, \frac{0.6}{8.1}\right\} \\ B_2 &= \left\{\frac{0.6}{0.82}, \frac{0.8}{0.86}, \frac{1}{0.9}, \frac{0.8}{0.92}, \frac{0.6}{0.94}\right\} \end{aligned} \tag{50}$$

and then, the purpose is to compute

$$Z = \frac{1}{2}(X + Y) = (A, B). \tag{51}$$

First, with (22) and (23), we have

$$\begin{aligned} 2A^p = \Big\{ & \frac{0.4}{15.7}, \frac{0.4}{16.0}, \frac{0.6}{16.1}, \frac{0.4}{16.3}, \frac{0.6}{16.4}, \frac{0.6}{16.5}, \frac{0.4}{16.6}, \frac{0.6}{16.7}, \\ & \frac{0.8}{16.8}, \frac{0.6}{16.9}, \frac{0.6}{17.0}, \frac{0.8}{17.1}, \frac{0.8}{17.2}, \frac{0.6}{17.3}, \frac{0.8}{17.4}, \frac{1}{17.5}, \\ & \frac{0.6}{17.7}, \frac{0.8}{17.8}, \frac{0.6}{18.1} \Big\} \end{aligned} \tag{52}$$

which is not a fuzzy number as the convexity is not satisfied at $\mu_A^p(16.3)$, $\mu_A^p(16.8)$, $\mu_A^p(17.3)$, and $\mu_A^p(17.8)$, and then, with Algorithm 2, we have

$$\begin{aligned} 2A = \Big\{ & \frac{0.4}{15.7}, \frac{0.4}{16.0}, \frac{0.6}{16.1}, \frac{0.6}{16.3}, \frac{0.6}{16.4}, \frac{0.6}{16.5}, \frac{0.6}{16.6}, \frac{0.6}{16.7}, \\ & \frac{0.6}{16.8}, \frac{0.6}{16.9}, \frac{0.6}{17.0}, \frac{0.8}{17.1}, \frac{0.8}{17.2}, \frac{0.8}{17.3}, \frac{0.8}{17.4}, \frac{1}{17.5}, \\ & \frac{0.6}{17.7}, \frac{0.6}{17.8}, \frac{0.6}{18.1} \Big\}. \end{aligned} \tag{53}$$

Thus,

$$\begin{aligned} A = \Big\{ & \frac{0.4}{7.85}, \frac{0.4}{8.0}, \frac{0.6}{8.05}, \frac{0.6}{8.15}, \frac{0.6}{8.2}, \frac{0.6}{8.25}, \frac{0.6}{8.3}, \frac{0.6}{8.35}, \frac{0.6}{8.4}, \frac{0.6}{8.45}, \\ & \frac{0.6}{8.5}, \frac{0.8}{8.55}, \frac{0.8}{8.6}, \frac{0.8}{8.65}, \frac{0.8}{8.7}, \frac{1}{8.75}, \frac{0.6}{8.85}, \frac{0.6}{8.9}, \frac{0.6}{9.05} \Big\}. \end{aligned} \tag{54}$$

Second, with (41) and (43), the probability distributions and the associated possibility for X and Y can be obtained. For example, for $b_{13} = 0.6$ from B_1, we have

$$\begin{aligned} p_{X3} &= \{0.4213, 0.2770, 0.1820, 0.1197\} \\ \mu(p_{X3}) &= \mu(b_{13}) = 1 \end{aligned} \tag{55}$$

while for $b_{11} = 0.48$ from B_1, we have

$$\begin{aligned} p_{X1} &= \{0.6818, 0.2363, 0.0819, 0\} \\ \mu(p_{X1}) &= \mu(b_{11}) = 0.6. \end{aligned} \tag{56}$$

Comparing p_{X3} and p_{X1}, we observe that elements with larger possibility are assigned more probability in p_{X3} than p_{X1} since $b_{13} > b_{11}$. As for $b_{23} = 0.9$, we have

$$p_{Y3} = \{0.0443, 0.1614, 0.5886, 0.1614, 0.0443\} \tag{57}$$

$$\mu(p_{X3}) = \mu(b_{23}) = 1. \tag{58}$$

Due to the space limitation, here, we just show p_{X1}, p_{X3}, and p_{Y3}, and probability for other values from B_1 and B_2 follows the same method. After receiving all probability distributions of X and Y from all values of B_1 and B_2, the probability distributions and the associated possibility for Z can be obtained with the convolution operation in (44), e.g.,

$$p_Z(8.05) = p_Z\left(\frac{1}{2}(6.9 + 7.2)\right) = p_X(6.9) * p_Y(7.2) \tag{59}$$

and with p_{X3} and p_{Y3}, we have

$$\begin{aligned} p_Z = \{ & 0.0187, 0.068, 0.248, 0.068, 0.024, 0.0123, 0.0447, \\ & 0.163, 0.0447, 0.0123, 0.0081, 0.0294, 0.1071, 0.0294, \\ & 0.0081, 0.0193, 0.0704, 0.0193, 0.0052\}. \end{aligned} \tag{60}$$

After countering all p_X and p_Y, and with (45)–(47), we have

$$B = \left\{ \frac{0.6}{0.57}, \frac{0.6}{0.58}, \frac{0.6}{0.59}, \frac{0.8}{0.6}, \frac{0.8}{0.61}, \frac{0.8}{0.62}, \frac{1}{0.63}, \frac{0.8}{0.64}, \frac{0.6}{0.65} \right\} \tag{61}$$

which is a discrete fuzzy number as it satisfies the convexity in Definition 1.

C. Discussion

In this section, a method to achieve the fusion of Z-numbers in mathematical functions is proposed. Two main problems are solved.

1) Obtaining the Possibility Distribution Satisfying Convexity: The extension principle and linear smoothing operator are combined since the possibility distribution from the traditional extension rule may be unreasonable (i.e., not satisfying properties in Definition 1). The basic idea of the method is to keep the extension-based sequence of membership degree as much as possible on the premise of satisfying the definition of fuzzy numbers. To implement the idea, a utility function measuring the similarity between the original extension-based sequence and the final sequence of membership degree is constructed. The longest nondecreasing (or nonincreasing) subsequence of the original extension-based possibility distribution is adopted. Thereafter, the least number of elements not belonging to the monotone subsequence is smoothed by the linear smoothing operator to guarantee convexity.

2) Exploring the Underlying Probability of Z-Numbers: According to Definition 3, two components denoted by A and B in a Z-number are not independent of each other but linked by the underlying probability. Therefore, finding the underlying probability is important for the computation of B. In view of this and motivated by the maximum entropy approach, an optimization model is constructed to get the most likely underlying probability. In the whole process, the underlying probability is obtained based on the definition and properties of Z-numbers rather than any extra assumption.

All in all, the whole process is extended from the framework (10) of Z-numbers proposed by Zadeh [2]. On the one hand, the linear smoothing operator is proposed to guarantee the convexity of the two obtained components so that they are all fuzzy numbers. On the other hand, the maximum entropy model is proposed to explore the most likely underlying probability without extra assumption, with which the relationship between the two components in Z-numbers is utilized. Thus, there is no change with the principles of Z-numbers in the computation process.

IV. Application

In this section, we show how the fusion of Z-numbers can be applied to fault diagnosis. As Zadeh mentioned in [2], humans have a remarkable capability to make rational decisions based on information which is uncertain, imprecise, and/or incomplete. Z-number is an effective model to formalize this sort of capability. Similar to Example 3.2, many previous works [2], [14], [47], [48] focused on how to process uncertainty involved in natural language by Z-numbers. The purpose here is to utilize the fusion of Z-numbers in real datasets of fault diagnosis, where the reliability of the datasets is considered.

A. Fault Diagnosis Based on the Fusion of Z-Numbers

The proposed method consists of the following steps:

1) Modeling Uncertainty With Z-Numbers: The first difficulty is to generate Z-numbers to model the uncertainty of the original dataset. For each attribute of an unlabeled sample, a Z-number is generated to measure to what extent and with what reliability the sample matches a class label. Assume that there are M classes, and each class is characterized by L features. Then, the decision matrix denoted by D for a test sample can be constructed as

$$D = \begin{array}{c} \\ C_1 \\ \vdots \\ C_i \\ \vdots \\ C_M \end{array} \begin{array}{c} \omega_1 \cdots \omega_j \cdots \omega_L \\ \left[\begin{array}{c} \\ \\ (A_{ij}, B_{ij}) \\ \\ \\ \end{array} \right] \end{array} \tag{62}$$

where C_i corresponds to a class, ω_j represents an attribute, and the Z-number $Z_{ij} = (A_{ij}, B_{ij})$ indicates the matching degree between the test data and class C_i under feature ω_j.

2) Fusion of Z-Numbers via the Addition Operation: Uncertainty can be further processed after generating Z-numbers in step (1). For this issue, Z-numbers $Z_{ij} = (A_{ij}, B_{ij})$ evaluating how much an unlabeled sample can match a class under different features can be integrated to obtain a global evaluation. In this article, the addition operation is adopted to achieve the goal

$$Z_{\rm i} = f(Z_{i1}, Z_{i2}, \ldots, Z_{iL}) = Z_{i1} + Z_{i2} + \cdots + Z_{iL} \tag{63}$$

where Z_i is the global Z-valuation covering the matching degree between the unlabeled data and class C_i over all features.

3) Decision-Making Based on Defuzzification: After getting the global evaluation over all features in step 2), it is necessary to rank all evaluations to decide the true class of the input pattern. As shown in (64), the true class C_{true} should be the one with the highest evaluation

$$C_{\text{true}} = \arg\max_{C_i}\{\text{score}_i\}. \tag{64}$$

As for this issue, a popular defuzzification method called the COG [49] is adopted. Assume a Z-number (A, B) with $A = \{(\mu_A(x_1)/x_1), \ldots, (\mu_A(x_n)/x_n)\}$ and $B = \{(\mu_B(b_1)/b_1), \ldots, (\mu_B(b_m)/b_m)\}$, and the defuzzification result of it can be computed by

$$\text{score}_i = \frac{\sum x_j \mu_A(x_j)}{\sum \mu_A(x_j)} \times \frac{\sum b_k \mu_B(b_k)}{\sum \mu_B(b_k)} \tag{65}$$

where the COG of B representing the information reliability is used as a discount factor of the COG of A which is the comprehensive matching degree between the test sample and the class C_i. Finally, the label C_i with the highest score$_i$ is assigned to the test sample according to (64).

B. Case Study I

In this section, we simulate the proposed method with the dataset of motor rotor fault diagnosis from [50], in which

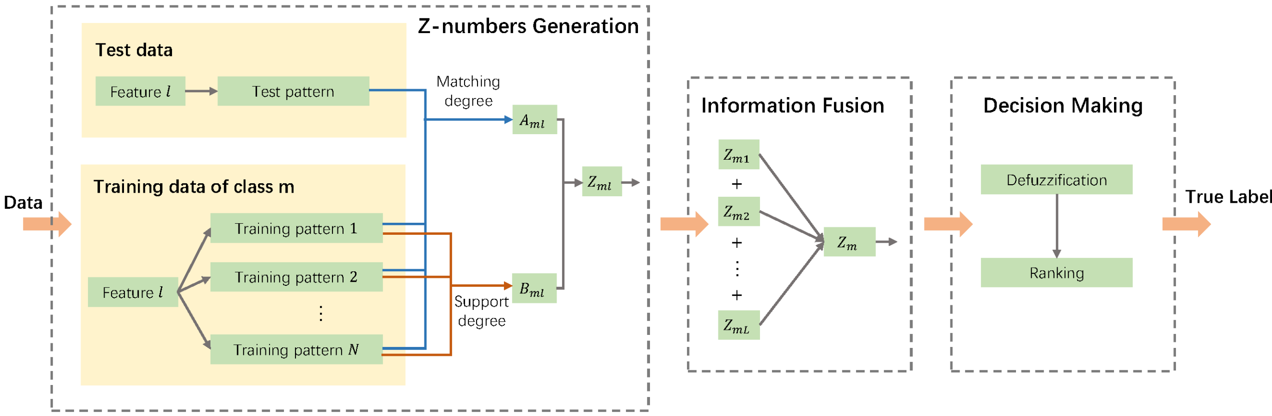


Fig. 4. Fault diagnosis based on the fusion of Z-numbers.

the equipment to collect the fault data is a multifunctional flexible rotor test stand. In the experiments of the flexible rotor, multiple sensors measure the vibration of the rotor system when the motor rotor runs stably in a certain working condition. The details of the dataset are given as follows. All fault data are from the rotor vibration signal. To collect the vibration data, the displacement sensor and the acceleration sensor are installed in the vertical and horizontal directions of the rotor supporting base, respectively. In the experiment, three faults are configured: rotor imbalance (denoted by F_1), rotor misalignment (denoted by F_2), and support base loosening (denoted by F_3). Thus, there are three classes in total. Each fault is characterized by four features. First, the vibration energy of the three faults is mostly concentrated on the basic frequency $1X$, double frequency $2X$, and triple frequency $3X$, so the amplitude of the vibration acceleration from $1X$ to $3X$ frequency is used as three features. Then, the average amplitude of the vibration displacement in the time domain is viewed as the fourth feature. For each feature of each fault class, six groups of fault data are collected. Furthermore, in each data group, 40 data are collected in 16 s for each feature of each class, so there are 2880 points in total across the dataset. In this experiment, we use 6-fold cross validation, that is, five groups of data are used as the training set and, then, the rest group is the test set. Thus, six runs of experiments are done. The final accuracy is the mean of the accuracy of all six runs. With the basis of Fig. 4, the experiments of fault diagnosis with the dataset used in this article consist of the following steps.

1) Generating Z-Numbers From the Original Data: Note that the collected data have a certain degree of ambiguity: data collected for a single feature from the same period (i.e., the same group) fluctuate in a certain range, and data collected for the same feature from different periods overlap with each other. Considering these points, Z-numbers can be generated with the idea, as shown in Fig. 4: the training patterns and test patterns can be built with the training data and test data, respectively, and then, the component A can be generated based on the matching degree (i.e., similarity) between them. Also, as there are multiple groups of training data, the component B can be generated by using the degree of how much they support each other, that is, the higher the degree all groups support each other, the more reliable A is. Details for Z-number generation of Fig. 4 about how to generate Z-numbers with this dataset are described in the following.

a) Modeling the component A of Z-numbers: The fuzzy evaluation measures how much a test sample matches the training templates. The event associated with the Z-numbers we aim to generate in (62) is the similarity, also called the matching degree, between a training pattern and a test pattern. A is the fuzzy evaluation toward this event. As mentioned before, there are $M = 3$ fault classes in the dataset, and each fault is characterized by $L = 4$ features. Sensors collect fault data $k = 40$ times in a time interval $\Delta t = 16$ s. In each experiment round, five groups of training data are collected. Here, we assume that the collected data are normally distributed to construct the training patterns. Then, the training templates and test patterns from the training data and test data can be constructed as follows.

For the training data of each feature of each class, $N = 5$ training patterns can be built by the empirical distribution of the data, i.e., each group of data observed from the same time interval for a feature of a class is built as a Gaussian model. Let $\mathcal{X}_{mn}^l = \{x_1, x_2, \ldots, x_k\}$ be the nth group of the observed data for feature l of class m, and then, a Gaussian distribution (i.e., template) can be built as

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_{mn}^l}\exp\left(-\frac{\left(x-\bar{X}_{mn}^l\right)^2}{2\left(\sigma_{mn}^l\right)^2}\right) \tag{66}$$

where $\bar{X}_{mn}^l$ is the mean of $\mathcal{X}_{mn}^l$ and σ_{mn}^l is the sample standard deviation with $l = 1, \ldots, 4$, $m = 1, \ldots, 3$ and $n = 1, \ldots, 5$, and they can be obtained by

$$\bar{X}_{mn}^l = \frac{x_1 + x_2 + \cdots + x_k}{k} \tag{67}$$

$$\sigma_{mn}^l = \sqrt{\frac{\left(x_1 - \bar{X}_{mn}^l\right)^2 + \cdots + \left(x_k - \bar{X}_{mn}^l\right)^2}{k-1}} \tag{68}$$

from which we know that a training pattern utilizes the empirical distribution of a group of data, so it summarizes the information of the data group. Thus, for the training set consisting of L features and N groups, each fault class has $L * N = 20$ associated training templates. We use

$r_{mn}^l = \mathcal{N}(\bar{X}_{mn}^l, \sigma_{mn}^l)$ to represent the training template associated with feature l of class m from group n. Taking $m = 1$ (i.e., the training data of F_1) and $l = 1$ as an example, the five training patterns for the first feature of F_1 are given as follows:

$$\begin{aligned} r_{11}^1 &= \mathcal{N}(0.1619, 0.02) \\ r_{12}^1 &= \mathcal{N}(0.1596, 0.0073) \\ r_{13}^1 &= \mathcal{N}(0.1644, 0.0009) \\ r_{14}^1 &= \mathcal{N}(0.1617, 0.0006) \\ r_{15}^1 &= \mathcal{N}(0.1598, 0.001). \end{aligned} \tag{69}$$

As for a test sample denoted by q, it includes $L = 4$ features, so it consists of four test patterns denoted by q^l with $l = 1, \ldots, 4$, and each pattern corresponds to a feature. Here, we use $q^1 = 0.1421$ to illustrate the process. The question is: how to measure the matching degree between the training template r_{mn}^l and the test pattern q^l? To solve the problem, the Mahalanobis distance [51], which can be viewed as a measure of the distance between a point and a distribution, is adopted. Here, the Mahalanobis distance of 1-D q^l and the distribution r_{mn}^l can be computed by

$$d\left(r_{mn}^l, q^l\right) = \frac{1}{\sigma_{mn}^l}\left|q^l - \bar{X}_{mn}^l\right| \tag{70}$$

with which the lower $d(r_{mn}^l, q^l)$, the higher the matching degree between the template pattern and the test pattern. Thus, we have

$$d\left(r_{11}^1, q^1\right) = \frac{1}{0.02} \times |0.1619 - 0.1421| = 0.9899. \tag{71}$$

Then, for the test pattern q^l, its distance to N training templates of each class can be computed with (70), and the Mahalanobis distance matrix of feature l denoted by MD_l can be obtained afterward

$$\mathrm{MD}_l = \begin{bmatrix} d\left(r_{11}^l, q^l\right) & \cdots & d\left(r_{1N}^l, q^l\right) \\ \vdots & \ddots & \vdots \\ d\left(r_{M1}^l, q^l\right) & \cdots & d\left(r_{MN}^l, q^l\right) \end{bmatrix} \tag{72}$$

with which the matrix of matching degree for q^l denoted by S_l can be obtained

$$S_l = \begin{bmatrix} s_{11}^l & \cdots & s_{1N}^l \\ \vdots & \ddots & \vdots \\ s_{M1}^l & \cdots & s_{MN}^l \end{bmatrix} \tag{73}$$

where

$$s_{mn}^l = 1 - \frac{d\left(r_{mn}^l, q^l\right)}{\max(\mathrm{MD}_l)} \tag{74}$$

is obtained by normalizing MD_l, and it measures the similarity of q^l and r_{mn}^l. The larger s_{mn}^l is, the more similar q^l and r_{mn}^l are. Then, for the test pattern $q^1 = 0.1421$, we have

$$S_1 = \begin{bmatrix} 0.9878 & 0.9705 & 0.6912 & 0.5853 & 0.7801 \\ 0.9648 & 0.9116 & 0.9544 & 0.8359 & 0.8493 \\ 0.6582 & 0.1006 & 0.6942 & 0.5325 & 0.0 \end{bmatrix} \tag{75}$$

from which we know that the similarity between q^1 and the first group of training data for the first feature of F_1 is 0.9878.

The final step to generate A is to generate fuzzy numbers. With S_l which describes the matching degree between the test pattern and each class under feature l, the triangular fuzzy numbers can be generated as (76) in which all information from N data groups involved in S_l is considered

$$\begin{aligned} A_{ml} = \Bigg\{ &\frac{0.5}{a + (h-a) \times 0.5}, \frac{0.75}{a + (h-a) \times 0.75}, \frac{1}{h}, \\ &\frac{0.75}{u - 0.75 \times (u-h)}, \frac{0.5}{u - 0.5 \times (u-h)} \Bigg\} \end{aligned} \tag{76}$$

where

$$a = \frac{\min\left(s_{m1}^l, \ldots, s_{mN}^l\right)}{\max(S_l)} \tag{77}$$

$$h = \frac{\min\left(s_{m1}^l, \ldots, s_{mN}^l\right) + \max\left(s_{m1}^l, \ldots, s_{mN}^l\right)}{2 \times \max(S_l)} \tag{78}$$

$$u = \frac{\max\left(s_{m1}^l, \ldots, s_{mN}^l\right)}{\max(S_l)} \tag{79}$$

where $m = 1, \ldots, M$. For $q^1 = 0.1421$, and with (75) and (76), we have

$$\begin{aligned} A_{11} &= \left\{\frac{0.5}{0.6944}, \frac{0.75}{0.7453}, \frac{1}{0.7962}, \frac{0.75}{0.8472}, \frac{0.5}{0.8981}\right\} \\ A_{21} &= \left\{\frac{0.5}{0.8788}, \frac{0.75}{0.8952}, \frac{1}{0.9115}, \frac{0.75}{0.9278}, \frac{0.5}{0.9441}\right\} \\ A_{31} &= \left\{\frac{0.5}{0.1757}, \frac{0.75}{0.2635}, \frac{1}{0.3514}, \frac{0.75}{0.4392}, \frac{0.5}{0.5271}\right\}. \end{aligned} \tag{80}$$

The above process demonstrates how to generate the component A of the Z-number for a test pattern q^l. As each class is determined by L features, part A of the decision matrix D in (62) can be obtained as

$$D_A = \begin{bmatrix} A_{11} & \cdots & A_{1L} \\ \vdots & \ddots & \vdots \\ A_{M1} & \cdots & A_{ML} \end{bmatrix} \tag{81}$$

where A_{ml} with $m = 1, \ldots, M$ and $l = 1, 2, \ldots, L$ is the similarity between a test pattern and the feature l of class m from the training set.

b) Generating B measuring the reliability of A: In the dataset of fault diagnosis, data collected by sensors may not be completely reliable. Thus, its reliability should be quantified. As a fuzzy evaluation A is generated based on the matching degree between a test pattern and N training patterns, the reliability of A should come from the reliability of training patterns, that is, the more reliable of the training data, which is used to build training templates, the more reliable of A. In view of this and as shown in Fig. 4, the similarity is also called the support degree among N training templates built from N groups of data for a feature of a class, showing how much they support each other, is adopted to generate component B. Details are demonstrated as follows.

First, quantify the similarity among training patterns. In each run of experiment, $N * M * L = 60$ training patterns denoted by r_{mn}^l are built when modeling A. As shown in Fig. 5, the overlapping area of two gaussian curves, i.e., the green area, is used to quantify how much they support each other. The higher the overlapping area is, the more similar the two

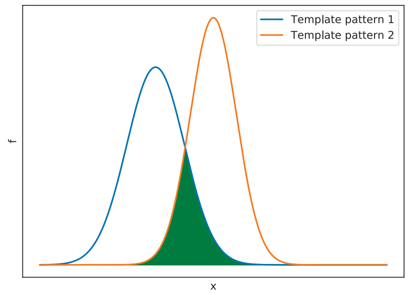


Fig. 5. Overlapping area of two Gaussian curves.

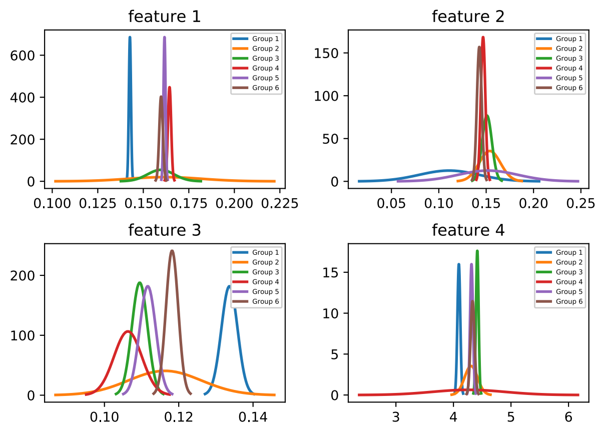


Fig. 6. Visualization for six groups of data of F_1.

distributions are. As mentioned before, we have six groups of data in the training set, and in each run of experiments, five groups of data are used as the training set. In Fig. 6, we draw the Gaussian curves for all six groups of data associated with all four features of fault F_1. Note that in each run, only five curves from five training data groups are utilized to generate B. We observe from Fig. 6 that all curves have different overlapping areas so that they have different reliability.

In each run, after obtaining all overlapping areas of every two training patterns from N groups, the matrix of support degree for feature l of class m denoted by Su_m^l can be obtained

$$\mathrm{Su}_m^l = \begin{bmatrix} \mathrm{su}_m^l\left(r_{m1}^l, r_{m1}^l\right) & \cdots & \mathrm{su}_m^l\left(r_{m1}^l, r_{mN}^l\right) \\ \vdots & \ddots & \vdots \\ \mathrm{su}_m^l\left(r_{mN}^l, r_{m1}^l\right) & \cdots & \mathrm{su}_m^l\left(r_{mN}^l, r_{mN}^l\right) \end{bmatrix} \tag{82}$$

where $\mathrm{su}_m^l(r_{mi}^l, r_{mj}^l)$ is the overlapping area between r_{mi}^l and r_{mj}^l with $i, j = 1, \ldots, N$. Also, Su_m^l is symmetric. Taking the five training patterns in (69) as an example, we have

$$\mathrm{Su}_1^1 = \begin{bmatrix} 1.0 & 0.5473 & 0.1002 & 0.0693 & 0.1100 \\ 0.5472 & 1.0 & 0.1984 & 0.1608 & 0.2585 \\ 0.1002 & 0.1984 & 1.0 & 0.0586 & 0.0131 \\ 0.0693 & 0.1608 & 0.0586 & 1.0 & 0.2168 \\ 0.1100 & 0.2585 & 0.0131 & 0.2168 & 1.0 \end{bmatrix}. \tag{83}$$

The next step is to generate fuzzy numbers with the matrix of support degree. The overall support degree, also called the credibility, of template r_{mi}^l can be defined as the sum of its overlapping area with all data groups

$$\mathrm{Cred}_{mi}^l = \sum_{j=1}^{N} \mathrm{su}_m^l\left(r_{mi}^l, r_{mj}^l\right). \tag{84}$$

Then, the normalized overall support denoted by NCred_{mi}^l can be obtained as

$$\mathrm{NCred}_{mi}^l = \frac{\mathrm{Cred}_{mi}^l}{\sum_{i=1}^{N} \mathrm{Cred}_{mi}^l}. \tag{85}$$

Similar to generating fuzzy numbers for A, fuzzy numbers associated with B can be constructed

$$B_{ml} = \left\{ \frac{0.5}{a + (h-a) \times 0.5}, \frac{0.75}{a + (h-a) \times 0.75}, \frac{1}{h}, \frac{0.75}{u - 0.75 \times (u-h)}, \frac{0.5}{u - 0.5 \times (u-h)} \right\} \tag{86}$$

where

$$a = \frac{\min_{i=1,\ldots,N} \mathrm{NCred}_{mi}^l}{\max_{l=1,\ldots,L} \max_{i=1,\ldots,N} \mathrm{NCred}_{mi}^l} \tag{87}$$

$$h = \frac{\min_{i=1,\ldots,N} \mathrm{NCred}_{mi}^l + \max_{i=1,\ldots,N} \mathrm{NCred}_{mi}^l}{2 * \max_{l=1,\ldots,L} \max_{i=1,\ldots,N} \mathrm{NCred}_{mi}^l} \tag{88}$$

$$u = \frac{\max_{i=1,\ldots,N} \mathrm{NCred}_{mi}^l}{\max_{l=1,\ldots,L} \max_{i=1,\ldots,N} \mathrm{NCred}_{mi}^l}. \tag{89}$$

For example, with Su_1^1, Su_1^2, Su_1^3, and Su_1^4 showing the support degree of all four features of F_1, we have

$$\begin{aligned} B_{11} &= \left\{ \frac{0.5}{0.7247}, \frac{0.75}{0.7706}, \frac{1}{0.8165}, \frac{0.75}{0.8624}, \frac{0.5}{0.9082} \right\} \\ B_{12} &= \left\{ \frac{0.5}{0.7340}, \frac{0.75}{0.7594}, \frac{1}{0.7847}, \frac{0.75}{0.8101}, \frac{0.5}{0.8355} \right\} \\ B_{13} &= \left\{ \frac{0.5}{0.6063}, \frac{0.75}{0.6531}, \frac{1}{0.6999}, \frac{0.75}{0.7467}, \frac{0.5}{0.7936} \right\} \\ B_{14} &= \left\{ \frac{0.5}{0.6665}, \frac{0.75}{0.7128}, \frac{1}{0.7590}, \frac{0.75}{0.8052}, \frac{0.5}{0.8515} \right\}. \end{aligned} \tag{90}$$

B_{ml} represents how N training patterns generated from the training data associated with feature l of class m support each other. Then part B of the decision matrix D in (62) can be generated as:

$$D_B = \begin{bmatrix} B_{11} & \cdots & B_{1L} \\ \vdots & \ddots & \vdots \\ B_{M1} & \cdots & B_{ML} \end{bmatrix}. \tag{91}$$

Since it is training data that are used to construct D_B without test data, D_B generated from the training data for all test samples is always the same. Then, concatenate D_A and D_B, and the decision matrix for a test sample over all L features can be constructed as

$$D = \begin{bmatrix} (A_{11}, B_{11}) & \cdots & (A_{1L}, B_{1L}) \\ \vdots & \ddots & \vdots \\ (A_{M1}, B_{M1}) & \cdots & (A_{ML}, B_{ML}) \end{bmatrix}. \tag{92}$$

TABLE I

CASE STUDY I: ACCURACY OF FAULT DIAGNOSIS BY AGGREGATING DIFFERENT NUMBERS OF FEATURES

	F_1	F_2	F_3	Overall
No Fusion	0.2479	0.7052	0.9229	0.6253
$L_1 + L_2$	0.6764	0.9659	0.9875	0.8766
$L_1 + L_2 + L_3$	0.8583	0.9948	0.9864	0.9465
$L_1 + L_2 + L_3 + L_4$	0.9917	1.0	0.9750	0.9889

2) Aggregating Z-Numbers via the Addition Operation: Z-numbers representing the matching degree between a test sample and each class over all four features can be obtained by step 1), and then, they can be aggregated by (63). For example, for a test sample with $q^1 = 0.1421$, $q^2 = 0.104$, $q^3 = 0.1305$, and $q^4 = 4.125$, we have

$$\begin{aligned} Z_1 &= \left(\left\{\frac{0.5}{2.85}, \ldots, \frac{0.75}{3.19}, \frac{1}{3.2}, \frac{0.75}{3.21}, \ldots, \frac{0.5}{3.57}\right\},\right. \\ &\quad \left.\left\{\frac{0.5}{0.65}, \frac{0.75}{0.66}, \frac{1}{0.67}, \frac{0.5}{0.68}, \frac{0.5}{0.69}\right\}\right) \\ Z_2 &= \left(\left\{\frac{0.5}{2.19}, \ldots, \frac{0.75}{2.59}, \frac{1}{2.6}, \frac{0.75}{2.61}, \ldots, \frac{0.5}{3.02}\right\},\right. \\ &\quad \left.\left\{\frac{0.5}{0.65}, \frac{1}{0.67}, \frac{0.5}{0.68}, \frac{0.5}{0.69}\right\}\right) \\ Z_3 &= \left(\left\{\frac{0.5}{1.6}, \ldots, \frac{0.75}{2.06}, \frac{1}{2.07}, \frac{0.75}{2.08}, \ldots, \frac{0.5}{2.55}\right\},\right. \\ &\quad \left.\left\{\frac{0.5}{0.7}, \frac{0.5}{0.71}, \frac{0.75}{0.72}, \frac{1}{0.73}, \frac{0.75}{0.74}, \frac{0.5}{0.75}, \frac{0.5}{0.76}\right\}\right) \end{aligned} \tag{93}$$

where Z_m is the sum of the four Z-numbers generated for all features from the training data of class m, and it means the overall matching degree of the test sample with class m.

3) Defuzzification for Decision-Making: All the three Z-numbers in (93) can be defuzzified via (65), so we have

$$\text{score}_1 = 2.1474, \quad \text{score}_2 = 1.7440, \quad \text{score}_3 = 1.5148 \tag{94}$$

from which the true class of the example test sample is F_1.

To demonstrate the superiority of information fusion, we compute the accuracy of fault diagnosis by combining different numbers of information sources (i.e., features). Four different strategies are carried out. We first compute the average accuracy without aggregating Z-numbers of different features. Then, only the Z-numbers from any two features are aggregated. In the third comparative experiment, Z-numbers from every three features are combined. Finally, all four features are considered in information fusion. The experimental results of the above four experiments are shown in Table I, where the number of L_i means the number of features used in information fusion.

We observe from Table I that the accuracy of diagnosis increases with the number of sources (i.e., features) used in information fusion. By combining all four features, the overall accuracy is improved by 36.36% than no fusion. Thus, the accuracy of fault diagnosis can be improved by aggregating information from multiple sources. This is because information from some sources is complements of other sources, that is, if the true class of a test sample is F_1, its matching degree with another class under feature 1 may be higher than that of F_1. Then, if only feature 1 is considered, an erroneous decision is made. Information from other features can help mitigate the negative influence of feature 1.

TABLE II

CASE STUDY I: ACCURACY OF FAULT DIAGNOSIS BY USING DIFFERENT METHODS

Methods	F_1	F_2	F_3	Overall
Dempster [52]	0.9850	0.8750	1.0	0.9530
Zhang [53]	1.0	0.9150	1.0	0.9720
Our method	0.9917	1.0	0.9750	0.9889

In addition, we compare our approach with the method based on the Dempster–Shafer evidence theory [52] and the extended one from the evidence theory proposed by Zhang *et al.* [53]. When testing the classic Dempster's combination rule, the method from [53] is used to generate the basic probability assignment. The results are shown in Table II, from which we observe that our method gets better overall accuracy than the other two. Note that our method is based on the fusion of Z-numbers so that the reliability of the information provided by the training set is considered, while Shafer [52] and Zhang *et al.* [53] aggregated basic probability assignments generated from the original dataset by utilizing Dempster–Shafer's evidence theory, which is another useful tool to quantify uncertainty. The comparative results demonstrate the effectiveness of the fusion of Z-numbers in fault diagnosis.

C. Case Study II

In this section, we test the performance of the framework proposed in Section IV-A with the Case Western Reserve University (CWRU) dataset [54], which is a popular and easily accessible dataset for fault diagnosis. CWRU includes 161 records, and they are grouped into four subsets: normal baseline, 12k drive end fault, 48k drive end fault, and 12k fan end fault. In each subset, the vibration data were recorded for motor loads of 0–3 horsepower (motor speeds of 1797–1720 RPM). Also, fault diameters range from 0.007 to 0.040 in. In our experiments, we use the data files from normal baseline and 12k drive fault data with a diameter of 0.007 in, that is, there are six classes of faults in our simulation: normal (F_1), ball bearing fault (F_2), inner race fault (F_3), outer race fault centered at $6:00$ (F_4), outer race fault orthogonal at $3:00$ (F_5), and outer race fault opposite at $12:00$ (F_6). For the sake of convenience, we use the first 120 000 points of the driven-end data from each data file and reshape it to (300, 400), that is, we have 300 fault samples for each class in our dataset and each sample consists of 400 points. Moreover, as the vibration data were collected from multiple motor loads with different horsepower, we aggregate the fault signal from each load to make better decisions.

To generate Z-numbers, we randomly split the original dataset into six groups. Then, similar to case study I, six runs of experiments are done. In each run, five groups of data are used as the training templates, and the rest group is the test set. Different from case study I where each feature of a test sample is a data point, it consists of 400 points in this simulation. Thus, the overlapping area of the training templates

TABLE III
CASE STUDY II: ACCURACY OF FAULT DIAGNOSIS BY AGGREGATING DATA FROM DIFFERENT NUMBERS OF LODES

	F_1	F_2	F_3	F_4	F_5	F_6	Overall
No Fusion	0.9808	1	0.9533	0.7942	0.6125	1	0.8901
2 sources	0.9978	1	0.9967	0.9061	0.7533	1	0.9423
3 sources	0.9992	1	1	0.9508	0.7950	1	0.9575
4 sources	1	1	1	0.98	0.8433	1	0.9706

TABLE IV
CASE STUDY II: ACCURACY OF FAULT DIAGNOSIS BY USING DIFFERENT METHODS

Methods	F_1	F_2	F_3	F_4	F_5	F_6	Overall
Dempster	1	1	0.8333	0.9967	0.9167	1	0.9578
Our method	1	1	1	0.98	0.8433	1	0.9705

and test templates are computed to generate the component A of Z-numbers, and the rest procedures to generate Z-numbers are similar to case study I.

Similar to Table I from case study I, Table III shows the classification accuracy by combining information from a different number of information sources (i.e., motor loads). We observe that the classification accuracy can be improved by information fusion. What is more, in Table IV, we compare our method with the one based on the Dempster–Shafer evidence theory [52], [53]. We see that our method can get better performance.

V. CONCLUSION

This article develops both the theory and application of Z-numbers. In theory, we propose a method for the fusion of Z-numbers. Besides the extension principle, a linear smoothing operator is adopted to guarantee the convexity of the two components of Z-numbers. What is more, without extra assumption, the underlying probability linking the two components of Z-numbers is computed from an optimization model. Then, we apply the fusion of Z-numbers to fault diagnosis. Different from previous works in which Z-numbers were used to process linguistic information, we exhibit the potential of Z-numbers in real data.

Future works could focus on the generation of Z-numbers, which has always been a complicated problem. The picture of the proposed framework is applicable to other pattern classification problems, but the details of modeling Z-numbers are not general for all datasets. In this article, we empirically generate Z-numbers for the specific fault datasets used in the experiments according to the interpretation of the two components of Z-numbers and the characteristics of the datasets themselves. For wider applications, the generation of Z-numbers is still an open issue.

REFERENCES

[1] E. T. Jaynes, *Probability Theory: The Logic of Science*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[2] L. A. Zadeh, "A note on Z-numbers," *Inf. Sci.*, vol. 181, no. 14, pp. 2923–2932, 2011.

[3] R. Banerjee, S. K. Pal, and J. K. Pal, "A decade of the Z-numbers," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 8, pp. 2800–2812, Aug. 2021.

[4] R. A. Aliev, A. V. Alizadeh, and O. H. Huseynov, "The arithmetic of discrete Z-numbers," *Inf. Sci.*, vol. 290, pp. 134–155, Jan. 2015.

[5] R. Bělohlávek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*. Oxford, U.K.: Oxford Univ. Press, 2017.

[6] S.-W. Tang, Z.-J. Zhou, C.-H. Hu, J.-B. Yang, and Y. Cao, "Perturbation analysis of evidential reasoning rule," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 8, pp. 4895–4910, Aug. 2019.

[7] Z. Huang, L. Yang, and W. Jiang, "Uncertainty measurement with belief entropy on the interference effect in the quantum-like Bayesian networks," *Appl. Math. Comput.*, vol. 347, pp. 417–428, Apr. 2019.

[8] Z. He and W. Jiang, "An evidential dynamical model to predict the interference effect of categorization on decision making," *Knowl.-Based Syst.*, vol. 150, pp. 139–149, Jun. 2018.

[9] W. Jiang, "A correlation coefficient for belief functions," *Int. J. Approx. Reasoning*, vol. 103, pp. 94–106, Dec. 2018.

[10] W. Jiang, C. Xie, M. Zhuang, and Y. Tang, "Failure mode and effects analysis based on a novel fuzzy evidential method," *Appl. Soft Comput.*, vol. 57, pp. 672–683, Aug. 2017.

[11] R. R. Yager, "On Z-valuations using Zadeh's Z-numbers," *Int. J. Intell. Syst.*, vol. 27, no. 3, pp. 259–278, Mar. 2012.

[12] R. A. Aliev, O. H. Huseynov, and L. M. Zeinalova, "The arithmetic of continuous Z-numbers," *Inf. Sci.*, vol. 373, pp. 441–460, Dec. 2016.

[13] R. A. Aliev, W. Pedrycz, O. H. Huseynov, and S. Z. Eyupoglu, "Approximate reasoning on a basis of Z-number-valued if–then rules," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 6, pp. 1589–1600, 2016.

[14] W. Jiang, Y. Cao, and X. Deng, "A novel Z-network model based on Bayesian network and Z-number," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 8, pp. 1585–1599, May 2020.

[15] R. A. Aliev and A. V. Alizadeh, "Z-differential equations," in *Proc. Int. Conf. Theory Appl. Soft Comput., Comput. Words Perceptions*. Cham, Switzerland: Springer, 2019, pp. 69–77.

[16] M. Mazandarani and Y. Zhao, "Z-differential equations," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 3, pp. 462–473, Mar. 2020.

[17] R. Aliev, O. Huseynov, and K. Aliyeva, "Toward eigenvalues and eigenvectors of matrices of Z-numbers," in *Proc. Int. Conf. Theory Appl. Soft Comput., Comput. Words Perceptions*. Cham, Switzerland: Springer, 2019, pp. 309–317.

[18] R. Jafari, S. Razvarz, and A. Gegov, "Neural network approach to solving fuzzy nonlinear equations using Z-Numbers," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 7, pp. 1230–1241, Jul. 2020.

[19] R. A. Aliev, W. Pedrycz, and O. H. Huseynov, "Hukuhara difference of Z-numbers," *Inf. Sci.*, vol. 466, pp. 13–24, Oct. 2018.

[20] K. Shen and J. Wang, "Z-VIKOR method based on a new comprehensive weighted distance measure of Z-number and its application," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 6, pp. 3232–3245, Dec. 2018.

[21] S. Das, A. Garg, S. Pal, and J. Maiti, "A weighted similarity measure between Z-numbers and bow-tie quantification," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 9, pp. 2131–2142, Jul. 2019.

[22] B. Kang, Y. Deng, K. Hewage, and R. Sadiq, "A method of measuring uncertainty for Z-number," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 4, pp. 731–738, Apr. 2018.

[23] B. Kang, G. Chhipi-Shrestha, Y. Deng, K. Hewage, and R. Sadiq, "Stable strategies analysis based on the utility of Z-number in the evolutionary games," *Appl. Math. Comput.*, vol. 324, pp. 202–217, May 2018.

[24] Y. Li, D. Pelusi, Y. Deng, and K. H. Cheong, "Relative entropy of Z-numbers," *Inf. Sci.*, vol. 581, pp. 1–17, Dec. 2021.

[25] H. Peng, X. Wang, H. Y. Zhang, and J. Wang, "Group decision-making based on the aggregation of Z-numbers with Archimedean t-norms and t-conorms," *Inf. Sci.*, vol. 569, pp. 264–286, Aug. 2021.

[26] Q. Jia, J. Hu, Q. He, W. Zhang, and E. Safwat, "A multicriteria group decision-making method based on AIVIFSs, Z-numbers, and trapezium clouds," *Inf. Sci.*, vol. 566, pp. 38–56, Aug. 2021.

[27] Y. Tian, L. Liu, X. Mi, and B. Kang, "ZSLF: A new soft likelihood function based on Z-numbers and its application in expert decision system," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 8, pp. 2283–2295, Aug. 2020.

[28] T. Wang, H. Li, X. Zhou, D. Liu, and B. Huang, "Three-way decision based on third-generation prospect theory with Z-numbers," *Inf. Sci.*, vol. 569, pp. 13–38, Aug. 2021.

[29] Z. He and W. Jiang, "An evidential Markov decision making model," *Inf. Sci.*, vol. 467, pp. 357–372, Oct. 2018.

[30] S. J. Ghoushchi, K. Gharibi, E. Osgooei, M. N. Ab Rahman, and M. Khazaeili, "Risk prioritization in failure mode and effects analysis with extended SWARA and MOORA methods based on Z-numbers theory," *Informatica*, vol. 32, no. 1, pp. 41–67, 2021.

[31] R. H. Abiyev, N. Akkaya, and I. Gunsel, "Control of omnidirectional robot using Z-number-based fuzzy system," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 1, pp. 238–252, Jan. 2019.
[32] V. Nourani, H. Najafi, E. Sharghi, and K. Roushangar, "Application of Z-numbers to monitor drought using large-scale oceanic-atmospheric parameters," *J. Hydrol.*, vol. 598, Jul. 2021, Art. no. 126198.
[33] J. Quan, Y. Zhou, X. Wang, and J.-B. Yang, "Information fusion based on reputation and payoff promotes cooperation in spatial public goods game," *Appl. Math. Comput.*, vol. 368, Mar. 2020, Art. no. 124805.
[34] W. Jiang, K. Huang, J. Geng, and X. Deng, "Multi-scale metric learning for few-shot learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 3, pp. 1091–1102, Mar. 2021.
[35] A. Y. Ng and M. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes," in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, pp. 841–848.
[36] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, "A possibilistic fuzzy C-means clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 4, pp. 517–530, Aug. 2005.
[37] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, p. 106587, 2020.
[38] F. Zhou, S. Yang, H. Fujita, D. Chen, and C. Wen, "Deep learning fault diagnosis method based on global optimization GAN for unbalanced data," *Knowl.-Based Syst.*, vol. 187, Jan. 2020, Art. no. 104837.
[39] X. Xu, X. Yan, C. Sheng, C. Yuan, D. Xu, and J. Yang, "A belief rule-based expert system for fault diagnosis of marine diesel engines," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 50, no. 2, pp. 656–672, Feb. 2017.
[40] F. Xiao, Z. Cao, and A. Jolfaei, "A novel conflict measurement in decision-making and its application in fault diagnosis," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 1, pp. 186–197, Jan. 2020.
[41] L. A. Zadeh, "Probability measures of fuzzy events," *J. Math. Anal. Appl.*, vol. 23, no. 2, pp. 421–427, 1968.
[42] R. B. Ash, *Basic Probability Theory*. North Chelmsford, MA, USA: Courier Corporation, 2008.
[43] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, vol. 4. Upper Saddle River, NJ, USA: Prentice-Hall, 1995.
[44] G. Wang, C. Wu, and C. Zhao, "Representation and operations of discrete fuzzy numbers," *Southeast Asian Bull. Math.*, vol. 29, no. 5, pp. 1003–1010, 2005.
[45] J. Casasnovas and J. V. Riera, "On the addition of discrete fuzzy numbers," in *Proc. 5th WSEAS Int. Conf. Telecommun. Informat.*, 2006, pp. 432–437.
[46] N. Marhamati, E. K. Buxton, and S. Rahimi, "Integration of Z-numbers and Bayesian decision theory: A hybrid approach to decision making under uncertainty and imprecision," *Appl. Soft Comput.*, vol. 72, pp. 273–290, Nov. 2018.
[47] P. Patel, E. S. Khorasani, and S. Rahimi, "Modeling and implementation of Z-number," *Soft Comput.*, vol. 20, no. 4, pp. 1341–1364, 2016.
[48] J. Huang, D.-H. Xu, H.-C. Liu, and M.-S. Song, "A new model for failure mode and effect analysis integrating linguistic Z-numbers and projection method," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 3, pp. 530–538, Mar. 2019.
[49] E. Van Broekhoven and B. De Baets, "Fast and accurate center of gravity defuzzification of fuzzy system outputs defined on trapezoidal fuzzy partitions," *Fuzzy sets Syst.*, vol. 157, no. 7, pp. 904–918, 2006.
[50] C. Wen and X. Xu, *Multi-Source Uncertain Information Fusion Theory and Application: Fault Diagnosis and Reliability Assessment*. Beijing, China: Science Press, 2012.
[51] P. C. Mahalanobis, "On the generalized distance in statistics," *Proc. Nat. Inst. Sci. India*, pp. 49–55, 1936.
[52] G. Shafer, *A Mathematical Theory of Evidence*, vol. 42. Princeton, NJ, USA: Princeton Univ. Press, 1976.
[53] Y. Zhang, W. Jiang, and X. Deng, "Fault diagnosis method based on time domain weighted data aggregation and information fusion," *Int. J. Distrib. Sensor Netw.*, vol. 15, no. 9, Sep. 2019, Art. no. 1550147719875629.
[54] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mech. Syst. Signal Process.*, vol. 64, pp. 100–131, Dec. 2015.

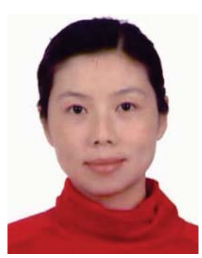
Ying Cao was born in Shaanxi, China, in 1996. She received the bachelor's and master's degrees in electronic and information engineering from Northwestern Polytechnical University, Xi'an, China, in 2017 and 2020, respectively.

Her research interests include information fusion and intelligent information processing.

Jian-Bo Yang is currently the Chair of decision and systems sciences and the Director of the Decision and Cognitive Sciences Research Centre (DCSRC), Alliance Manchester Business School, The University of Manchester, Manchester, U.K. Over the last three decades, he has been conducting research in a range of areas, including the evidential reasoning theory; multiple criteria decision analysis under uncertainty; multiple objective optimization; decision support systems; hybrid quantitative and qualitative decision modeling using techniques from operational research, artificial intelligence, and systems engineering; and complex system modeling, simulation, and control for engineering and management systems. His current applied research is in the areas of artificial intelligence (AI)-powered professional services and financial technologies, in particular legal, insurance, healthcare, and financial services; new product development; system maintenance management; risk and security modeling and analysis; performance analysis and improvement; and the integrated evaluation of products, systems, projects, and policies.

Xinyang Deng received the bachelor's, master's, and Ph.D. degrees from Southwest University, Chongqing, China, in 2010, 2013, and 2016, respectively.

Since 2017, he has been an Associate Professor with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China. His main research interests include multisource information fusion, Dempster–Shafer evidence theory, and uncertain information modeling and processing.



Wen Jiang received the bachelor's and master's degrees from Information Engineering University, Zhengzhou, China, in 1994 and 1997, respectively, and the Ph.D. degree from Northwestern Polytechnical University, Xi'an, China, in 2009.

She is currently a Professor with the School of Electronics and Information, Northwestern Polytechnical University. Her main research interests include multisource information fusion and artificial intelligence with uncertainty.