A Decision Probability Transformation Method Based on the Neural Network

Junwei Li, Aoxiang Zhao * and Huanyu Liu

School of Artificial Intelligence, Henan University, Zhengzhou 450046, China
* Correspondence: zhaoaoxiang08@126.com

Abstract: When the Dempster–Shafer evidence theory is applied to the field of information fusion, how to reasonably transform the basic probability assignment (BPA) into probability to improve decision-making efficiency has been a key challenge. To address this challenge, this paper proposes an efficient probability transformation method based on neural network to achieve the transformation from the BPA to the probabilistic decision. First, a neural network is constructed based on the BPA of propositions in the mass function. Next, the average information content and the interval information content are used to quantify the information contained in each proposition subset and combined to construct the weighting function with parameter \( r \). Then, the BPA of the input layer and the bias units are allocated to the proposition subset in each hidden layer according to the weight factors until the probability of each single-element proposition with the variable is output. Finally, the parameter \( r \) and the optimal transform results are obtained under the premise of maximizing the probabilistic information content. The proposed method satisfies the consistency of the upper and lower boundaries of each proposition. Extensive examples and a practical application show that, compared with the other methods, the proposed method not only has higher applicability, but also has lower uncertainty regarding the transformation result information.

Keywords: Dempster–Shafer evidence theory; average information content; interval information content; neural network; probabilistic information content

1. Introduction

Uncertain information [1] plays a significant role in many engineering applications, including multi-attribute decision-making [2–4], fault diagnosis [5–7], image processing [8,9], knowledge inference [10,11], risk assessment [12–15], and pattern classification [16,17]. Recent studies have focused on how to measure and handle information uncertainty, and a series of theories have been introduced, such as modeling information uncertainty based on the entropy function [18–21] and using Dempster–Shafer (DS) evidence theory [22–30] to deal with uncertain information. Compared with the traditional probability theory, the DS evidence theory can directly use the basic probability assignment (BPA) of multi-subset focal elements to express information uncertainty. However, a large number of elements in a proposition can make the credibility assignment too fragmented, thus causing certain complications in the decision-making process and reducing its precision. The BPA of multi-element propositions can respond to the support for each single-element proposition, which can be reasonably mapped to form probability in the frame of discernment [31–37].

Sudano [31] mapped the BPA of multi-element propositions to the probability of each single-element proposition according to a certain ratio using the belief and plausibility functions. Smets [32] equally distributed the BPA of multi-element propositions to each single-element proposition according to cardinality. However, this method is too conservative and does not fully use the known information. Pan [33] assigned the BPA of multi-element propositions according to the ordered weighted average operator (OWA) and then determined the final probability of each single-element proposition using the
minimum entropy difference as a constraint. When a single-element proposition was not contained in any multi-element proposition, the allocation could still affect the single-element proposition, leading to anti-intuitive results. Deng [34] proposed a probability transformation method based on the belief interval, which first obtains the preference of each single-element proposition by the possibility degree, then quantifies the data information about the belief interval of each singleton based on the continuous interval argument ordered weighted average (C-OWA) operator, and finally calculates the support degree of the singleton based on quantitative data information to reasonably allocate the BPA of multi-subset focal elements. However, this method can result in extreme cases where the preference degree of a single-element proposition can be zero, which will further lead to a support degree of zero, thus meaning that the single-subset proposition is not assigned by the BPA to a multi-element proposition. Huang [35] proposed a probabilistic transformation method based on the Shapley value. This method requires that the degree to which each single-element proposition contributes to the multi-element propositions is determined; then, these are transformed according to the contribution degree. Since the marginal probability of a single-element proposition may be zero, which can mean that the proposition is not considered in the assignment of multi-element propositions, this may lead to inaccurate transformation results. Li [36] proposed a probability transformation method based on the ordered visibility graph (OVG). The OVG network was constructed according to the BPA order, and then the weight of each single-element proposition was obtained according to the proposition edges and cardinality. According to the weight, the BPA of multi-element propositions was transformed into the probability of each single-element proposition. This method only uses the out-degree and in-degree of the nodes to determine the weight and does not make full use of the influence of nodes with larger shadows. Chen [37] improved Li’s method as follows. First, the OVG network was constructed by each proposition order of information volume, and the weighted adjacency matrix was constructed with proposition edges or belief entropy. Then, the weight of each single-element proposition was obtained from the matrix and cardinality. Finally, the probability of each single-element proposition was obtained by assigning the BPA of multi-element propositions according to weight.

The key to the above methods is obtaining the distribution weight of the multi-element proposition. Inspired by these methods, under the premise of obtaining reasonable transformation results and minimizing the uncertainty of information, an efficient probability transformation method based on a neural network is proposed. The neural network is a computational model [38,39] designed to simulate the neural network of the human brain. Similar to human brain neurons, it consists of multiple nodes (neurons), which are interconnected to model the complex relationship between data. The connections between different nodes are given different weights, representing the influence of one node on another. Each node represents a specific function, which is calculated by combining the input and weight of other nodes. The result of the calculation is input into the activation function, which, in turn, provides the final output of the neuron and is passed on to the next neuron. The main question is how to construct the neuronal network and changing weight, assigning the BPA of a proposition subset layer by layer according to weight until the probability of each single-element proposition containing the parameter is output. Then, the probabilistic information content (PIC) [31] is determined (with PIC being the dual form of Shannon entropy [40], which was first proposed by Sudano [31] and is widely used to measure the uncertainty of probability distribution) and, finally, optimal probability transformation results are obtained under the premise of PIC maximization. The main contributions of this study can be summarized as follows:

(1). The BPA of a multi-element proposition with the largest cardinality is used as an input layer of a neural network, and the BPAs of the remaining existing propositions are used as bias units of the neural network. The hidden network layers are constructed according to decreasing order of cardinality for the proposition subset of the input
layer and bias units. Finally, the probability of each single-element proposition is output by the output layer.

(2). The interval information content (IIC) and average information content (AIC) are introduced to quantify the information contained in each proposition subset and combined to construct a weighting function containing the parameter. The weighting function reaches its extreme value at a certain point, in preparation for the use of the constraints below. After obtaining the PIC of the change based on the probability of each single-element proposition, the maximum PIC value is taken as a constraint to obtain the optimal transformation results.

The remainder of this paper is organized as follows. Section 2 briefly introduces the DS evidence theory and decision probability transformation methods. Section 3 describes the properties of belief entropy and PIC. Section 4 proposes a probability transformation method based on a neural network and proves its properties. Section 5 presents numerical simulation results to demonstrate the rationality and superiority of the proposed method compared to the existing methods. A practical application of our proposed method is described in Section 6. Section 7 concludes the paper.

2. Preliminaries

In this section, preliminaries such as the DS evidence theory [24] and some existing probability transformation methods are briefly introduced.

2.1. DS Evidence Theory

Assume a set $\Theta$ is composed of $k$ mutually exclusive elements, which can be expressed as: $\Theta = \{\theta_1, \theta_2, \ldots, \theta_k\}$, where $\Theta$ is a frame of discernment; these elements are randomly combined to form the power set as follows:

$2^\Theta = \{\emptyset, \theta_1, \theta_2, \ldots, \theta_k, \theta_1 \cup \theta_2, \ldots, \theta_1 \cup \theta_2 \cup \ldots \cup \theta_k, \Theta\}$,

where $\emptyset$ denotes the empty set. If there is a function: $m : 2^\Theta \rightarrow [0, 1]$ satisfying the conditions of $m(\emptyset) = 0$ and $\sum_{A \subseteq \Theta} m(A) = 1$, then $m$ is called the mass function, and $m(A) > 0$ is the BPA of a proposition $A$.

If $\text{Bel} : 2^\Theta \rightarrow [0, 1]$ holds for $A \in 2^\Theta$, then:

$$\text{Bel}(A) = \sum_{B \subseteq A} m(B),$$

(1)

where $\text{Bel}(A)$ denotes the belief function, which represents the degree of trust that proposition $A$ is true.

If $\text{Pl} : 2^\Theta \rightarrow [0, 1]$ holds for $A \in 2^\Theta$, then:

$$\text{Pl}(A) = \sum_{B: A \cap B = \emptyset} m(B),$$

(2)

where $\text{Pl}(A)$ denotes the plausibility function, which represents the degree of trust that proposition $A$ is non-false.

Since the belief function $\text{Bel}$ represents a degree of agreement with a proposition, and the plausibility function $\text{Pl}$ indicates a degree of disagreement with the proposition; the interval formed by the belief and plausibility functions represents the uncertainty of the proposition. The intervals are formed by $\text{Bel}$ and $\text{Pl}$, as shown in Figure 1.

2.2. Probability Transformation Methods

In the frame of discernment, $\Theta = \{\theta_1, \theta_2, \ldots, \theta_k\}$, where $\theta_i \in \Theta$, when there is a multi-element proposition $D$ in evidence, which has a large BPA, then the information assignment is too scattered, with large uncertainty, so it is difficult to make decisions directly based on
the BPA. To facilitate the decision-making process and improve its accuracy, the following improvements have been introduced in recent studies.

Sudano [31] used Bel and Pl to obtain the probability of a single-element proposition by:

$$\text{Pr aPl}(\theta_i) = \text{Bel}(\theta_i) + \epsilon \cdot \text{Pl}(\theta_i),$$

where $\epsilon = \frac{1 - \sum_{\theta_i \in \Theta} \text{Bel}(\theta_i)}{\sum_{\theta_i \in \Theta} \text{Pl}(\theta_i)}$.

When only Pl is used, the probability of a single-element proposition is calculated by:

$$\text{Pr Pl}(\theta_i) = \sum_{D \subseteq \Theta, \theta_i \in D} \frac{\text{Pl}(\theta_i)}{\sum_{D_i \subseteq \Theta, |D_i| = 1} \text{Pl}(D_i) m(D_i).}$$

In contrast, when only Bel is used, the probability of a single-element proposition is obtained by:

$$\text{Pr Bel}(\theta_i) = \sum_{D \subseteq \Theta, \theta_i \in D} \frac{\text{Bel}(\theta_i)}{\sum_{D_i \subseteq \Theta, |D_i| = 1} \text{Bel}(D_i) m(D_i).}$$

Smets [32] provided an in-depth analysis of the reasonability of probability transformation in the decision-making domain and proposed the Pignistic probability transformation method, which is given by:

$$\text{BetP}(\theta_i) = \sum_{\theta_i \in \Theta} \frac{m(D)}{|D|},$$

where $|D|$ is the cardinality of a multi-element proposition $D$.

Pan [33] proposed a probability transformation method based on the OWA operator and entropy difference, and the average function can be expressed by:

$$M_{[\text{Bel}, \text{Pl}]}(\theta_i) = \frac{\text{Bel}(\theta_i) + \text{Pl}(\theta_i)}{2}.$$  

The adopted normalization function is as follows:

$$\rho(\theta_i) = \frac{M_{[\text{Bel}, \text{Pl}]}(\theta_i)}{\sum_{\theta_j \in \Theta} M_{[\text{Bel}, \text{Pl}]}(\theta_j).}$$

The probability of each single-element proposition obtained using the OWA operator is given by:

$$\text{OWAP}_m(\theta_i) = (T_i)' - (T_{i-1})',$$

where $T_i = \sum_{k=1}^i \rho(\theta_k)$ and $T_0 = 0$. Since the probability of a single-element proposition contains an unknown variable $r$, $\min\{|E_d - H|\}$ is used to determine the variable $r$, where $E_d$ is the Deng entropy of the BPA, and $H$ is the Shannon entropy of probability.
Deng [34] defined the belief interval and preference degree of a single-element proposition based on the belief and plausibility function, as follows:

\[ l_i = Pl(\theta_i) - Bel(\theta_i), \]

\[ p_{ij} = \max \left\{ 1 - \max \left( \frac{Pl(\theta_i) - Bel(\theta_j)}{l_i + l_j}, 0 \right), 0 \right\}, \]

where \( i, j = 1, 2, \ldots, k \).

\[ p(\theta_i) = \frac{1}{n-1} \sum_{j=1, j\neq i}^{k} p_{ij}, \]

where quantization of belief interval data was performed according to the C-OWA operator as follows:

\[ \beta_i = Pl(\theta_i) + Bel(\theta_i) \cdot 2^|\Theta| \frac{2}{2^|\Theta| + 1} . \]

The preference degree was used to modify the quantized belief interval data to obtain the support degree of the single-element proposition as follows:

\[ Sup(\theta_i) = \beta_i \cdot p(\theta_i). \]

The probability of a single-element proposition is given by:

\[ ITP(\theta_i) = m(\theta_i) + \sum_{\theta_j \in B \subseteq \Theta} \epsilon_{\theta_j} m(B), \]

where \( \epsilon_{\theta_j} = \frac{Sup(\theta_j)}{\sum_{j=1, j \neq i}^{k} Sup(\theta_j)} \).

Huang [35] proposed a probability transformation method based on the Shapley value, where the marginal probability of a proposition \( \theta_i \) for a proposition \( D \) is expressed by:

\[ MP(\theta_i) = m(D) - m(D \setminus \{ \theta_i \}), \]

where \( \theta_i \in D \), and \( D \subseteq \Theta; D \setminus \{ \theta_i \} \) denotes a subset of propositions \( D \) excluding \( \theta_i \). The average marginal probability contribution of \( \theta_i \) in \( D \) is calculated by:

\[ AMP_D(\theta_i) = \frac{1}{|D|!} \sum_{D \subseteq \Theta} \left[ m(D) - m(D \setminus \{ \theta_i \}) \right]. \]

The probability of each single-element proposition is obtained by:

\[ MPSV(\theta_i) = \sum_{D \subseteq \Theta, \theta_i \in D} AMP_D(\theta_i). \]

After reordering each proposition according to the BPA from the largest to the smallest, Li [36] obtained a set of edges of each proposition based on the ordered visibility graph and obtained the weight of each single-element proposition according to the cardinality as follows:

\[ g(\theta_i) = \frac{K(\theta_i)}{\sum_{D \subseteq \Theta, \theta_i \in D} K(\theta_i)^\prime}, \]

\[ K(\theta_i) = \frac{1}{|D|} \left( \frac{1}{|D|} K(D) \right). \]
The probability of a single-element proposition is calculated by:

\[ OVG_P_m(\theta_i) = \sum_{D \subseteq \Theta_i \in D} \left( g(\theta_i) \frac{m(D)}{1 - m(\emptyset)} \right), m(\emptyset) \neq 1. \]  

(21)

Chen [37] ordered propositions according to the Deng entropy magnitude, and the Deng entropy of a proposition is calculated by:

\[ IV(A_i) = -m(A_i) \log_2 \frac{m(A_i)}{|A_i| - 1}. \]  

(22)

After ordering, which is denoted as \([1, IV(A'_1)], (2, IV(A'_2)), \ldots, (s, IV(A'_s))\], and where \(A'_1\) is the ordered proposition, the network is constructed by the OVG. Then, the weighted adjacency matrix is obtained based on the set of edges with an internal element \(b_{ij}\). When there is a connection between \(A'_i\) and \(A'_j\), then \(b_{ij} = 1\); otherwise, \(b_{ij} = 0\). The two edge weights are obtained using the node distance and belief entropy, respectively, as follows:

\[ w_{ij} = \frac{b_{ij}}{|i-j|}, \]  

\[ w_{ij} = \frac{1}{2} \left( \frac{IV(A'_i)}{\sum_{a_{k=1}} IV(A'_k)} + \frac{IV(A'_j)}{\sum_{a_{k=1}} IV(A'_k)} \right). \]  

(24)

Then, the degree of a focal element is calculated based on the edge weights as follows:

\[ D(A'_i) = \sum_{j=1}^s w_{ij}. \]  

(25)

The weight of a single-element proposition is calculated by:

\[ w(\theta_i) = \sum_{\theta_j \in A'} \frac{D(A'_i)}{|A_i|}. \]  

(26)

Finally, the probability of each single-element proposition is given by:

\[ OVGWP(\theta_i) = m(\theta_i) + \sum_{\theta_j \in A \subseteq \emptyset} \frac{w(\theta_i)}{\sum_{\theta_j \in A} w(\theta_j)} m(A). \]  

(27)

3. Shannon Entropy and Probabilistic Information Content

This section describes how to evaluate the performance of a transformation method after the probabilistic transformation is completed.

3.1. Shannon Entropy

Shannon [40] first introduced the concept of entropy into the field of information theory, defining the entropy of discrete finite sets. Assume a finite discrete set is defined as \(F = \{f_1, f_2, \ldots, f_n\}\), and its probability distribution is denoted by \(P = \{p(f_1), p(f_2), \ldots, p(f_n)\}\). Then, the Shannon entropy of this set is obtained by:

\[ H(F) = -\sum_{i=1}^n p(f_i) \log_a p(f_i), \]  

(28)
where \( \sum_{i=1}^{n} p(f_i) = 1 \); in this study, \( a = e \), where \( e \) is Euler’s number. In a discrete set, when the probability of each element is equally distributed, i.e., \( p(f_1) = p(f_2) = \ldots = p(f_n) = \frac{1}{n} \), then, the Shannon entropy is maximal, and it is given by \( H_{\text{max}}(F) = -\sum_{i=1}^{n} \frac{1}{n} \ln \frac{1}{n} \). The Shannon entropy is used to measure information uncertainty, and the greater information uncertainty is, the greater the entropy will be and vice versa.

3.2. Probabilistic Information Content

Sudano [31] developed a PIC method to evaluate transformation results. The PIC for the probability distribution \( P = \{p(f_1), p(f_2), \ldots, p(f_n)\} \) is calculated by:

\[
\text{PIC}(F) = 1 + \frac{1}{H_{\text{max}}(F)} \left( \sum_{i=1}^{n} p(f_i) \ln p(f_i) \right). 
\]

The PIC is expressed as a dual of the normalized Shannon entropy, and varies between zero and one. The smaller the PIC value is, the greater the information uncertainty will be, and vice versa. When \( \text{PIC} = 1 \), there is no interference of uncertain information in decision-making, but when \( \text{PIC} = 0 \), it will be impossible to make a decision based on the information. The PIC has often been used to evaluate the performance of probabilistic transformation methods.

4. Probability Transformation Based on the Neural Network

The existing methods assign the BPA of multi-element propositions to single-element propositions according to certain weights, but these methods have the defects of not fully considering the relationship between propositions and using the existing information, which will lead to inaccuracies in the weight factors generated in some special cases, and the probability transformation results that are obtained are counter-intuitive. To overcome the shortcomings of the existing methods, this study proposes a probability transformation method based on a neural network. In this section, the neural network construction and weighting function are introduced. The optimal transformation result is described in detail.

4.1. Neural Network Construction

A neural network has one input layer and one output layer, and the number of hidden layers and bias nodes depends on the actual situation of the evidence. In this paper, a neural network uses the ReLu function [41] as the activation function.

If the frame of discernment \( \Theta = \{\theta_1, \theta_2, \ldots, \theta_k\} \) exists, the BPA \( m(\theta_1, \theta_2, \ldots, \theta_n) \) of the proposition \( \{\theta_1, \theta_2, \ldots, \theta_n\} \) with the maximum cardinality is used as the input network layer. The BPAs of the remaining propositions are denoted as bias units, the propositions in the hidden layer are subsets of the input layer and the bias unit propositions, and the size of the cardinality decreases with the layer number. From the combination of elements in probability statistics, it is known that when the number of proposition subsets of the first hidden layer is \( C_n^{n-1} = n \), the BPA of proposition subsets is expressed as \( \text{N}m(\theta_1, \theta_2, \ldots, \theta_{n-1}) \), \( \text{N}\text{m}(\theta_1, \theta_2, \ldots, \theta_{n-2}, \theta_n) \), \ldots, \( \text{N}\text{m}(\theta_2, \theta_3, \ldots, \theta_n) \), where each proposition subset cardinality is \( (n-1) \). The number of proposition subsets of the second hidden layer is \( C_n^{n-2} = \frac{n \times (n-1)}{2} \), and the number of proposition subsets of the \( j \)th hidden layer is \( C_n^{n-j} = \frac{n \times (n-1) \times \ldots \times (n-j+1)}{j!} \), where \( j! = j \times (j-1) \times (j-2) \times \ldots \times 2 \times 1 \). There are \( (n-2) \) hidden layers; the output layer gives the probability values of all single-element propositions in the frame of discernment as follows: \( \text{PN}\text{m}(\theta_1), \text{PN}\text{m}(\theta_2), \ldots, \text{PN}\text{m}(\theta_k) \). The neural network structure is presented in Figure 2.

Each neuron in the hidden layer is composed of multiple parts, and the initial value of a neuron is obtained based on the accumulation of weights and bias units and then activated using the activation function before obtaining the BPA of each focal proposition, as shown in Figure 3.
Figure 2. Schematic representation of the neural network structure.
4.2. AIC and IIC Values

Assume the set of discernment \( \Theta = \{ \theta_1, \theta_2, \ldots, \theta_k \} \), and a proposition \( A_i \subseteq \Theta \). In order to accurately quantify the information content of the proposition \( A_i \), the average information content (AIC) and interval information content (IIC) based on belief and plausibility functions are defined as follows:

\[
AIC(A_i) = e^{\left| \Theta \right| \left( \frac{\text{Bel}(A_i)}{r} + \text{Pl}(A_i) \right)},
\]

\[
IIC(A_i) = e^{\left| \Theta \right| \left( \frac{\text{Pl}(A_i)}{r} - \text{Bel}(A_i) \right)}.
\]

4.3. Weighting Function with Variable

This paper proposes a weighting function containing variable parameter \( r \), which can obtain the weight of each proposition by combining the AIC and IIC. The weighting function is defined as follows:

\[
W(A_i) = rAIC(A_i)^{1-r} + (1-r)IIC(A_i),
\]

where \( 0 \leq r \leq 1 \).

**Theorem 1.** The weight varies with a variable \( r \); the function curve is neither monotonically increasing nor monotonically decreasing but has an extreme value point at a certain point.

**Proof.** Let \( W(A_i) = f(r) \), \( AIC(A_i) = X \), and \( IIC(A_i) = Y \) so that \( f(r) = rX^{1-r} + (1-r)Y^r \), where \( 0 < r < 1 \), \( X \geq 1 \) and \( Y \geq 1 \). When \( r = 0 \), then \( f(0) = 1 \); when \( r = 1 \), then \( f(1) = 1 \). The derivative of the weighting function is given by:

\[
f'(r) = X^{1-r} - rX^{1-r}\ln X - Y^r + (1-r)Y^r \ln Y,
\]

since \( f'(0) = X + \ln Y - 1 \) and \( f'(1) = 1 - \ln X - Y \), regardless of the values of \( X \) and \( Y \), \( f'(0) \) and \( f'(1) \) will always have opposite signs. According to the first sufficient condition for determining an extreme value, there exists \( Z \in [0,1] \) so that \( f'(Z) = 0 \); this point represents an extreme value of the function.

If \( A_i \) is a proposition of the \( p \)th hidden layer, assign its BPA to the subset proposition \( a_i \) of the \((p+1)\)th layer as follows:

\[
w(a_i) = \frac{W(a_i)}{\sum_{a_j \subseteq A_i} W(a_j)},
\]
Then, the corresponding BP A is given by:

\[Nm'(a_i) = m(a_i) + \sum_{a_i \subseteq A_i \subseteq \Theta} w(a_i)Nm(A_i), \quad (35)\]

\[Nm(a_i) = \max(0, Nm'(a_i)), \quad (36)\]

where \(Nm(A_i)\) is the BPA of the proposition \(A_i\); \(m(a_i)\) is the bias unit; when there is no bias unit, then \(Nm'(a_i) = \sum_{a_i \subseteq A_i \subseteq \Theta} w(a_i)Nm(A_i)\). If \(A_i\) is a proposition subset of the last hidden layer, its BPA is assigned to a single proposition subset \(\theta_i\) as follows:

\[PNm'(\theta_i) = m(\theta_i) + \sum_{\theta_i \in A_i \subseteq \Theta} w(\theta_i)Nm(A_i), \quad (37)\]

\[PNm(\theta_i) = \max\left(0, PNm'(\theta_i)\right). \quad (38)\]

At this point, the output probability contains variable parameter \(r\).

**Theorem 2.** According to the related literature [34], this transformation method can be justified by verifying the consistency of the upper and lower bounds: \(Bel(\theta_i) \leq PNm(\theta_i) \leq Pl(\theta_i)\).

**Proof.** Because \(PNm(\theta_i) = m(\theta_i) + \sum_{\theta_i \in A_i \subseteq \Theta} w(\theta_i)Nm(A_i)\), \(PNm(\theta_i) \geq m(\theta_i)\), \(m(\theta_i) = Bel(\theta_i)\), so \(PNm(\theta_i) \geq Bel(\theta_i)\) holds; because \(Nm(A_i)\) is jointly assigned by the BPA and bias nodes of the previous subset of propositions, so \(Pl(\theta_i) \geq PNm(\theta_i)\); finally, inequality \(Bel(\theta_i) \leq PNM(\theta_i) \leq Pl(\theta_i)\) holds. \(\square\)

The above proof shows that the proposed method is a reasonable decision probability transformation method and, in the following, a specific example with large uncertainty is used to illustrate the calculation process of the proposed method.

**Example 1.** Assume the frame of discernment is \(\Theta = \{\theta_1, \theta_2, \theta_3\}\), and \(m\) is a mass function in \(\Theta\). Then, the corresponding BPA is given by:

\[m(\theta_1) = 0.2, m(\theta_2) = 0.1, m(\theta_1, \theta_2) = 0.3, m(\theta_2, \theta_3) = 0.25, m(\theta_1, \theta_2, \theta_3) = 0.15\]

First, a neural network is constructed based on the evidence, as follows: \(m(\theta_1, \theta_2, \theta_3)\) is the input layer; \(m(\theta_1)\), \(m(\theta_2)\), \(m(\theta_1, \theta_2)\) and \(m(\theta_2, \theta_3)\) are the bias units; \{\(\theta_1\), \(\theta_2\), \(\theta_3\)\}, \{\(\theta_1, \theta_2\)\}, \{\(\theta_1, \theta_3\)\}, and \{\(\theta_2, \theta_3\)\} are the proposition subsets in the hidden layer.

Then, the AIC and IIC values of the hidden-layer propositions are obtained and used to construct the weight factors as follows:

\[AIC(\theta_1, \theta_2) = e^{\|\theta_2\|\left(\frac{Bel(\theta_1, \theta_2) + Pl(\theta_1, \theta_2)}{2}\right)} = 601.84,\]

\[IIC(\theta_1, \theta_2) = e^{\|\theta_2\|\left(Bel(\theta_1, \theta_2) - Pl(\theta_1, \theta_2)\right)} = 24.53,\]

\[W(\theta_1, \theta_2) = r \times 601.84^{1-r} + (1 - r) \times 4.95^r.\]

Similarly,

\[W(\theta_1, \theta_3) = r \times 81.45^{1-r} + (1 - r) \times 270.43^r,\]

\[W(\theta_2, \theta_3) = r \times 403.43^{1-r} + (1 - r) \times 601.85^r.\]
Then, assigning \( m(\theta_1, \theta_2, \theta_3) \) to proposition subsets in the hidden layer according to the weights and combining the bias units, the BPA of each proposition in the hidden layer can be obtained by:

\[
Nm'(\theta_1, \theta_2) = m(\theta_1, \theta_2) + \frac{W(\theta_1, \theta_2)}{W(\theta_1) + W(\theta_2)} \times m(\theta_1, \theta_2, \theta_3)
\]

\[
= 0.3 + \frac{r \times 601.85^{1-r} + (1-r) \times 4.95^r + r \times 81.45^{1-r} + (1-r) \times 16.44^r + r \times 99.48^{1-r} + (1-r) \times 6.05^r}{N(\theta_1, \theta_2) = \max(0, Nm'(\theta_1, \theta_2))},
\]

\[
Nm'(\theta_1, \theta_3) = m(\theta_1, \theta_3) + \frac{W(\theta_1, \theta_3)}{W(\theta_1) + W(\theta_3)} \times m(\theta_1, \theta_2, \theta_3)
\]

\[
= 0.25 + \frac{r \times 601.85^{1-r} + (1-r) \times 4.95^r + r \times 81.45^{1-r} + (1-r) \times 16.44^r + r \times 99.48^{1-r} + (1-r) \times 6.05^r}{N(\theta_1, \theta_3) = \max(0, Nm'(\theta_1, \theta_3))},
\]

For simplicity, in the following, \( Nm(\theta_1, \theta_2) \), \( Nm(\theta_1, \theta_3) \), and \( Nm(\theta_2, \theta_3) \) are denoted by \( g_{12}(r), g_{13}(r), \) and \( g_{23}(r) \), respectively. Next, the weight of each single-element proposition can be obtained by:

\[
W(\theta_1) = r \times \left( e^{4(g_{12}(r)+g_{13}(r)+0.4)} \right)^{1-r} + (1-r) \times \left( e^{8(g_{12}(r)+g_{13}(r))} \right)^{1-r},
\]

\[
W(\theta_2) = r \times \left( e^{4(g_{12}(r)+g_{23}(r)+0.2)} \right)^{1-r} + (1-r) \times \left( e^{8(g_{12}(r)+g_{23}(r))} \right)^{r},
\]

\[
W(\theta_3) = r \times \left( e^{4(g_{13}(r)+g_{23}(r))} \right)^{1-r} + (1-r) \times \left( e^{8(g_{13}(r)+g_{23}(r))} \right)^{r}.
\]

Similarly, \( W(\theta_1), W(\theta_2), \) and \( W(\theta_3) \) are denoted by \( q_1(r), q_2(r), \) and \( q_3(r) \), respectively. The bias units are combined to obtain the probability of a single-element proposition by:

\[
P(\theta_1) = 0.2 + \frac{W(\theta_1)}{W(\theta_1) + W(\theta_3)} \times Nm(\theta_1, \theta_2) + \frac{W(\theta_1)}{W(\theta_1) + W(\theta_3)} \times Nm(\theta_1, \theta_3)
\]

\[
P(\theta_1) = 0.2 + \frac{q_1(r)}{q_1(r)+q_3(r)} \times g_{12}(r) + \frac{q_1(r)}{q_1(r)+q_3(r)} \times g_{13}(r)
\]

\[
P(\theta_1) = \max(0, PNm'(\theta_1)).
\]

\[
P(\theta_2) = 0.1 + \frac{W(\theta_2)}{W(\theta_2) + W(\theta_3)} \times Nm(\theta_1, \theta_2) + \frac{W(\theta_2)}{W(\theta_2) + W(\theta_3)} \times Nm(\theta_2, \theta_3)
\]

\[
P(\theta_2) = 0.2 + \frac{q_2(r)}{q_1(r)+q_3(r)} \times g_{12}(r) + \frac{q_2(r)}{q_1(r)+q_3(r)} \times g_{23}(r)
\]

\[
P(\theta_2) = \max(0, PNm'(\theta_2)).
\]

\[
P(\theta_3) = 0.2 + \frac{W(\theta_1)}{W(\theta_1) + W(\theta_3)} \times Nm(\theta_1, \theta_3) + \frac{W(\theta_1)}{W(\theta_1) + W(\theta_3)} \times Nm(\theta_2, \theta_3)
\]

\[
P(\theta_3) = \max(0, PNm'(\theta_3)).
\]
The changing trends of the probability and the PIC value of single-element propositions with \( r \) are shown in Figure 4.

**Figure 4.** The changing trend of the single-element proposition probability (a) and the changing trend of PIC value (b).

### 4.4. Optimal Probability Calculation of Single-Element Propositions

The larger the PIC value is, the lower the information uncertainty in the decision-making, the better the transformation results, and the more favorable the decision-making. Additionally, the higher the information uncertainty, the worse the transformation results and the less favorable the decision-making. According to [33], to obtain optimal transformation results, it is necessary to determine parameter \( r \) by maximizing the PIC as follows:

\[
\arg \max_r \left( \text{PIC}(PNm(\Theta)) \right)
\]

s.t \( \text{PIC}(PNm(\Theta)) = 1 + \frac{1}{H_{\text{max}}(PNm(\Theta))} \left( \sum_{i=1}^{k} PNm(\theta_i) \ln PNm(\theta_i) \right) \)  

where \( H_{\text{max}}(PNm(\Theta)) = -\sum_{i=1}^{k} \frac{1}{k} \ln \frac{1}{k} \). For Example 1, it holds that \( \max(\text{PIC}(PNm(\Theta))) = 0.2230 \), \( r = 0.91 \), so \( PNm(\theta_1) = 0.2755 \), \( PNm(\theta_2) = 0.6380 \), and \( PNm(\theta_3) = 0.0865 \). This transformation process is illustrated in Figure 5.

Next, a brief overview of the proposed method steps is given. The exact content of each step varies with the actual situation, but the main operations of each of the steps are as follows:

**Step 1:** Different propositions are combined to construct a neural network model;

**Step 2:** The AIC and IIC of proposition are obtained by Equations (30) and (31), weights are initiated combining the AIC and IIC by Equation (32), and the BPA of each proposition is assigned according to the weights until the single-element proposition probability containing the variable parameter \( r \) is output by Equations (34)–(38);

**Step 3:** The parameter \( r \) is determined according to the constraints and the optimal transformation results are obtained by Equation (39).
5. Analysis and Numerical Examples

In this section, a few examples are given to compare the transformation results of the proposed method with those of the other methods [31–37] to verify the rationality and accuracy of the method proposed in this paper. The method performances are evaluated based on the PIC value.

Example 2. The frame of discernment is $\Theta = \{A, B, C\}$, and $m$ is a mass function in $\Theta$; then, the BPA is given by:

\[ m(A) = 0.2, m(B, C) = 0.8 \]

The probability transformation results of different methods for Example 2 are shown in Table 1.

In the probability transformation, since there is no information on single-element proposition $A$ in the multi-element proposition, the assignment should be independent of the proposition $A$, and it is impossible to discern the difference between propositions $B$
and C based on the known conditions. Intuitively, the BPA of a multi-element proposition should equally be assigned to the single-element propositions B and C. However, the PeBel transformation method cannot be used to obtain the transformation results, and the OWA method is influenced by the single-element proposition A when assigning the BPA of the multi-element proposition, which leads to counter-intuitive transformation results. The other probability transformation methods distribute the BPA of multi-element propositions equally according to the elemental cardinality to obtain a single-element proposition probability, which is consistent with the intuitive results.

Table 1. Probability transformation results of different methods for Example 2.

| Method  | A   | B   | C   |
|---------|-----|-----|-----|
| PraPl [31] | 0.2 | 0.4 | 0.4 |
| PpPI [31] | 0.2 | 0.4 | 0.4 |
| PrBel [31] | 0.2 | NaN | NaN |
| BetP [32]  | 0.2 | 0.4 | 0.4 |
| ITP [34]   | 0.2 | 0.4 | 0.4 |
| MPSV [35]  | 0.2 | 0.4 | 0.4 |
| OWAP [33]  | 0.3230 | 0.3756 | 0.3014 |
| OVG [36]   | 0.2 | 0.4 | 0.4 |
| OVGWP1 [37]| 0.2 | 0.4 | 0.4 |
| OVGWP2 [37]| 0.2 | 0.4 | 0.4 |
| PNm       | 0.2 | 0.4 | 0.4 |

Example 3. The frame of discernment is \( \Theta = \{ A, B \} \), and \( m \) is a mass function in \( \Theta \); then, the BPA is given by:

\[
m(A) = 0.5, \quad m(A, B) = 0.5
\]

Step 1: According to the cardinality of propositions, \( m(A, B) \) is used as the input layer, \( m(A) \) denoted the bias unit, and the probability of a single-element proposition is obtained by the output layer.

Step 2: The AIC and IIC of the single-element proposition A and B are obtained by Equations (30) and (31), respectively, as follows:

\[
AIC(A) = 20.0855, \\
AIC(B) = 2.7183, \\
IIC(A) = 7.3891, \\
IIC(B) = 7.3891
\]

The weights of a single-element are calculated by Equation (32) as follows:

\[
W(A) = r \times 20.0855^{1-r} + (1 - r) \times 7.3891', \\
W(B) = r \times 2.7183^{1-r} + (1 - r) \times 7.3891'.
\]

The probabilities of a single-element proposition are obtained by Equations (34), (37), and (38):

\[
PNm'(A) = 0.5 + \frac{W(A)}{W(A) + W(B)} \times 0.5 \\
PNm(A) = \max\left(0, PNm'(A)\right), \\
PNm'(B) = \frac{W(B)}{W(A) + W(B)} \times 0.5
\]
\[ P_{Nm}(B) = \max \left(0, P_{Nm}'(B)\right). \]

The changing trends of the probability and PIC value with \( r \) are shown in Figure 6.

\[ P_{Nm}(A) = 0.837, \quad P_{Nm}(B) = 0.163. \]

Step 3: Obtain the optimal probability by Equation (39) as follows:

\[ r = 0.19, \quad P_{Nm}(A) = 0.837, \quad P_{Nm}(B) = 0.163. \]

The transformation results of different methods for Example 3 are shown in Table 2 and Figure 7.

| Method      | A   | B   | PIC  |
|-------------|-----|-----|------|
| PraPI [31]  | 0.75| 0.25| 0.1877 |
| PrPI [31]   | 0.75| 0.25| 0.1877 |
| PrBel [31]  | 1   | NaN | NaN  |
| BetP [32]   | 0.75| 0.25| 0.1877 |
| ITP [34]    | 1   | 0   | NaN  |
| MPSV [35]   | 1   | 0   | NaN  |
| OWAP [33]   | 0.6 | 0.4 | 0.0290 |
| OVG [36]    | 0.75| 0.25| 0.1887 |
| OVGWP1 [37] | 0.875| 0.125| 0.4564 |
| OVGWP2 [37] | 0.875| 0.125| 0.4564 |
| PNM         | 0.837| 0.163| 0.3594 |

Since the BPA of a single-subset proposition \( A \), but not of a single-subset proposition \( B \), directly exists in the evidence, when assigning multi-element propositions, intuitively, the proposition \( A \) should have a larger weight; however, in this test, the OWA method assigned a larger weight to the proposition \( B \). The OVG, PraPI, PrPI, and BetP methods assigned the BPA of multi-element proposition equally to propositions \( A \) and \( B \). The OVGWP method and the proposed method considered the prior information of \( m(A) = 0.5 \), thus considering
the connection between multi-element propositions and each single-element proposition, and assigned a larger weight to the single-subset proposition $A$. The transformation results are reasonable, and the PIC is relatively larger, which is more favorable for decision-making.

![Figure 7. The probabilistic transformation results (a) and the PIC values (b) of different methods.](image)

**Example 4.** The frame of discernment is $\Theta = \{A, B, C\}$, and $m$ is a mass function in $\Theta$; then, the BPA is given by:

$$m(A) = 0.1, m(A, B) = 0.2, m(B, C) = 0.3, m(A, B, C) = 0.4$$

Step 1: The proposition cardinality defines $m(A, B, C) = 0.4$ as the input layer of the neural network; the hidden-layer proposition subsets are denoted by $\{A, B\}$, $\{A, C\}$, and $\{B, C\}$; $m(A) = 0.1$ and $m(B, C) = 0.3$ are bias units; $PNm(A)$, $PNm(B)$, and $PNm(C)$ denote the output layer.

Step 2: Calculate the AIC and IIC values of propositions $\{A, B\}$, $\{A, C\}$, and $\{B, C\}$ by Equations (30) and (31) as follows:

$$AIC(A, B) = 181.2722,$$
$$AIC(A, C) = 81.4509,$$
$$AIC(B, C) = 121.5104,$$
$$IIC(A, B) = 270.4264,$$
$$IIC(A, C) = 1339.4308,$$
$$IIC(B, C) = 121.5104.$$  

The weights of each proposition are obtained by Equation (32) as follows:

$$W(A, B) = r \times 181.2722^{1-r} + (1-r) \times 270.4264,$$
$$W(A, C) = r \times 81.4509^{1-r} + (1-r) \times 1339.4308,$$
$$W(B, C) = r \times 121.5104^{1-r} + (1-r) \times 121.5104.$$  

The BPAs of proposition subsets are obtained by Equations (34) and (35) as follows:

$$Nm'(A, B) = 0.2 + \frac{W(A, B)}{W(A, B) + W(A, C) + W(B, C)} \times 0.4 = g_{AB}(r),$$
$$Nm(A, B) = \max(0, g_{AB}(r)),$$
$$Nm'(A, C) = \frac{W(A, C)}{W(A, B) + W(A, C) + W(B, C)} \times 0.4 = g_{AC}(r),$$
$$Nm(A, C) = \max(0, g_{AC}(r)).$$
\[ Nm'(B, C) = 0.3 + \frac{W(B, C)}{W(A, B) + W(A, C) + W(B, C)} \times 0.4 \times g_{BC}(r), \]
\[ Nm(B, C) = \max(0, g_{BC}(r)). \]

The AIC and IIC values of each single-element proposition are calculated by Equations (30) and (31) as follows:

\[ AIC(A) = e^{A \times (0.1 + g_{AB}(r) + g_{AC}(r))}, \]
\[ AIC(B) = e^{A \times (g_{AB}(r) + g_{BC}(r))}, \]
\[ AIC(C) = e^{A \times (g_{AC}(r) + g_{BC}(r))}, \]
\[ IIC(A) = e^{A \times (g_{AB}(r) + g_{AC}(r))}, \]
\[ IIC(B) = e^{A \times (g_{AB}(r) + g_{BC}(r))}, \]
\[ IIC(C) = e^{A \times (g_{AC}(r) + g_{BC}(r))}. \]

The weights of each single-element proposition are obtained by Equation (32) as follows:

\[ W(A) = r \times AIC(A)^{1-r} + (1 - r) \times IIC(A) = q_A(r), \]
\[ W(B) = r \times AIC(B)^{1-r} + (1 - r) \times IIC(B) = q_B(r), \]
\[ W(C) = r \times AIC(C)^{1-r} + (1 - r) \times IIC(C) = q_C(r). \]

The probabilities of each proposition are calculated by Equations (37) and (38):

\[ PNm'(A) = 0.1 + \frac{q_A(r)}{q_A(r) + q_B(r)} \times g_{AB}(r) + \frac{q_A(r)}{q_A(r) + q_C(r)} \times g_{AC}(r), \]
\[ PNm(A) = \max(0, PNm'(A)), \]
\[ PNm'(B) = \frac{q_B(r)}{q_A(r) + q_B(r)} \times g_{AB}(r) + \frac{q_B(r)}{q_B(r) + q_C(r)} \times g_{BC}(r), \]
\[ PNm(B) = \max(0, PNm'(B)), \]
\[ PNm'(C) = \frac{q_C(r)}{q_A(r) + q_C(r)} \times g_{AC}(r) + \frac{q_C(r)}{q_B(r) + q_C(r)} \times g_{BC}(r), \]
\[ PNm(C) = \max(0, PNm'(C)). \]

When the value of parameter \( r \) varies in the range of \([0, 1]\), the probability and PIC curves change, as shown in Figure 8.

Step 3: Determine the optimal probability by Equation (39):

\[ r = 0.16, \]
\[ PNm(A) = 0.305, \]
\[ PNm(B) = 0.494, \]
\[ PNm(C) = 0.201. \]

The transformation results of different methods are shown in Table 3 and Figure 9.

The PrBel method is limited and cannot obtain the correct transformation results. The PraPI, ITP, and OWA methods assign a larger weight to proposition \( A \) than to proposition \( B \), due to the direct presence of the BPA of a single-element proposition \( A \) in the evidence. However, it is not considered that the multi-element propositions contain single-element proposition \( B \) when \( PI(B) > PI(A) \). Intuitively, a single-element proposition \( B \) should have...
a larger weight. The transformation results of the PrPl, ITP, MPSV, OVGP, and OVGPWP methods and the proposed method are reasonable. However, compared with the other methods, the proposed method has the largest PIC value of 0.0596, lower information uncertainty after transformation, and is more beneficial to the decision-making process.

Figure 8. The changing trend of the single-element proposition probability (a) and the changing trend of PIC value (b).

Table 3. Probability transformation results of different methods for Example 4.

| Method      | A   | B   | C   | PIC  |
|-------------|-----|-----|-----|------|
| PrPl [31]   | 0.374| 0.352| 0.274| 0.0078 |
| PrPl [31]   | 0.309| 0.438| 0.253| 0.0240 |
| PrPb [31]   | 0.7  | NaN | NaN | NaN  |
| BetP [32]   | 0.333| 0.383| 0.284| 0.0067 |
| ITP [34]    | 0.414| 0.388| 0.198| 0.0414 |
| MPSV [35]   | 0.333| 0.400| 0.267| 0.0122 |
| OWAP [33]   | 0.384| 0.356| 0.260| 0.0120 |
| OVGP [36]   | 0.393| 0.418| 0.189| 0.0474 |
| OVGWP1 [37] | 0.364| 0.445| 0.191| 0.0494 |
| OVGWP2 [37] | 0.357| 0.451| 0.192| 0.0500 |
| PNm         | 0.305| 0.494| 0.201| 0.0596 |

Figure 9. The probabilistic transformation results (a) and the PIC values (b) of different methods for Example 4.
6. Practical Application

In this section, the newly proposed method is applied to the practical problem of target recognition to further verify its effectiveness.

**Example 5.** The Iris dataset can be divided into three categories: \{Setosa, Versicolor, Virginica\} as FOD \( \Theta = \{Se, Ve, Vi\} \), each category contains four attributes: \{SL, SW, PL, PW\}; then, the Iris objects are assigned as BPAs \( \{m_{SL, m_{SW}, m_{PL}, m_{PW}}\} \) in light of the attributes given in Table 4.

**Table 4.** Modeled BPA of an object from Iris dataset.

| BPA          | \( m_{SL} \) | \( m_{SW} \) | \( m_{PL} \) | \( m_{PW} \) |
|--------------|--------------|--------------|--------------|--------------|
| \( m(Se) \) | 0.0437       | 0.0865       | 1.40 \times 10^{-9} | 8.20 \times 10^{-6} |
| \( m(Ve) \) | 0.3346       | 0.2879       | 0.6570        | 0.6616       |
| \( m(Vi) \) | 0.2916       | 0.1839       | 0.1726        | 0.1692       |
| \( m(Se, Ve) \) | 0.0437       | 0.0863       | 1.30 \times 10^{-9} | 8.20 \times 10^{-6} |
| \( m(Se, Vi) \) | 0.0239       | 0.0865       | 1.40 \times 10^{-11} | 3.80 \times 10^{-6} |
| \( m(Ve, Vi) \) | 0.2385       | 0.1825       | 0.1704        | 0.1692       |
| \( m(Se, Ve, Vi) \) | 0.0239       | 0.0863       | 1.40 \times 10^{-11} | 3.80 \times 10^{-6} |

The specific steps can be described as follows:

Step 1: The degree of credibility between evidence is measured according to the literature [42].

Step 2: The credibility degree of each evidence is modified.

Step 3: The weighted average evidence is obtained by taking into account the relationship between the bodies of evidence and the relative importance of the collected evidence.

Step 4: According to the Dempster combination rule, the combination result is obtained by combining the weighted average evidence three times.

Step 5: Due to the large information uncertainty in the combination result, the combination result is transformed into probability distribution by the proposed method.

For the target recognition problem in the Iris dataset, the proposed method is compared with some existing methods [35, 42–44], as shown in Table 5.

**Table 5.** Results of the target recognition problem in Iris dataset.

| Method          | \( m(Se) \) | \( m(Ve) \) | \( m(Vi) \) | Target |
|-----------------|--------------|--------------|--------------|--------|
| Xiao’s method [43] | 0.0053       | 0.7390       | 0.2407       | Ve     |
| Jiang’s method [44] | 4.90 \times 10^{-4} | 0.8798       | 0.1130       | Ve     |
| MSDF [42]      | 6.88 \times 10^{-5} | 0.9163       | 0.0790       | Ve     |
| MPSV [35]      | 7.24 \times 10^{-5} | 0.9186       | 0.0813       | Ve     |
| PNm            | 0.0001       | 0.9201       | 0.0798       | Ve     |

The recognition results of the five methods are the same, and the category of Iris is identified as Ve. It can be seen from Table 5 that this method outperformed the comparison methods, and the belief in Ve is as high as 92.01%, the belief degrees of Xiao’s method [43], Jiang’s method [44], MSDF [42], MPSV [35] for the category Ve are 73.90%, 87.98%, 91.63%, and 91.86%, respectively.

7. Conclusions

How to reasonably transform the BPA under the DS evidence theory into probability before decision-making has become a major research hotspot. In this paper, a probability transformation method is proposed by combining the BPA with a neural network. The BPA of multi-element propositions with maximum cardinality is used as the input network layer, and the BPAs of the remaining propositions are used as bias nodes, which are assigned
to the proposition subsets in each hidden layer of the neural network according to the weights. The probability of each single-element proposition is obtained as the network output. The AIC and IIC values of each proposition subset are determined using the belief and plausibility functions, respectively, and then combined to obtain the weight factors of the contained variables to output the probability containing variables. Finally, the PIC is maximized as a constraint to determine the variables to obtain the optimal probability transformation result. The proposed method is verified by numerical examples and compared with the other methods. The results indicate that the proposed method is more reasonable and has better generalizability and lower information uncertainty regarding the transformed results than the other method, which makes it more beneficial for decision-making. However, the proposed method may generate a large computational effort when the cardinality of a multi-element proposition is too large. In the future, a more comprehensive evaluation index for probabilistic transformation results could be explored to further verify the rationality and superiority of this method and apply this method to more practical scenarios.

**Author Contributions:** Methodology, J.L. and A.Z.; validation, J.L.; formal analysis, H.L.; writing—original draft preparation, A.Z.; writing—review and editing, J.L. and A.Z.; funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the National Natural Science Foundation of China (No. 61976080), the Programs for Science and Technology Development of Henan Province, China (No. 222102210004), the Key Research Projects of the University in Henan Province, China (No. 20B510001).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**

1. Xu, J.; Deng, Z.; Song, Q.; Chi, Q.; Wu, T.; Huang, Y.; Liu, D.; Gao, M. Multi-UAV counter-game model based on uncertain information. *Appl. Math. Comput.* **2020**, *366*, 124684. [CrossRef]
2. Xiao, F.; Wen, J.; Pedrycz, W. Generalized divergence-based decision making method with an application to pattern classification. *IEEE Trans. Knowl. Data Eng.* **2022**, *34*, 4035–4047. [CrossRef]
3. Liu, P.; Diao, H.; Zou, L.; Deng, A. Uncertain multi-attribute group decision making based on linguistic-valued intuitionistic fuzzy preference relations. *Inf. Sci.* **2019**, *398*, 293–308. [CrossRef]
4. Yan, M.; Wang, J.; Dai, Y.; Han, H. A method of multiple-attribute group decision making problem for 2-dimension uncertain linguistic variables based on cloud model. *Optim. Eng.* **2021**, *22*, 2403–2427. [CrossRef]
5. Zhu, D.; Cheng, X.; Yang, L.; Chen, Y.; Yang, S. Information fusion fault diagnosis method for deep-sea human occupied vehicle thruster based on deep belief network. *IEEE Trans. Cybern.* **2021**, *52*, 9414–9427. [CrossRef]
6. Yao, L.; Wu, Y. Robust fault diagnosis and fault-tolerant control for uncertain multiagent systems. *Int. J. Robust Nonlinear Control* **2020**, *30*, 8192–8205. [CrossRef]
7. Chen, Y.; Tang, Y. An improved approach of incomplete information fusion and its application in sensor data-based fault diagnosis. *Mathematics* **2021**, *9*, 1292. [CrossRef]
8. Zhu, Z.; Wei, H.; Hu, G.; Li, Y.; Qi, G.; Mazur, N. A novel fast single image dehazing algorithm based on artificial multiexposure image Fusion. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 500153. [CrossRef]
9. Zhang, Y.; Hua, C. A new adaptive visual tracking scheme for robotic system without image-space velocity information. *IEEE Trans. Syst. Man Cybern. Syst.* **2021**, *52*, 5249–5258. [CrossRef]
10. Li, L.; Xie, Y.; Chen, X.; Yue, W.; Zeng, Z. Dynamic uncertain causality graph based on cloud model theory for knowledge representation and reasoning. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 1781–1799. [CrossRef]
11. Legner, C.; Pentek, T.; Otto, B. Accumulating design knowledge with reference models: Insights from 12 years’ research into data management. *J. Assoc. Inf. Syst.* **2020**, *21*, 735–770. [CrossRef]
12. Chen, J.; Wang, C.; He, K.; Zhao, Z.; Chen, M.; Du, R.; Ahn, G. Semantics-aware privacy risk assessment using self-learning weight assignment for mobile apps. *IEEE Trans. Dependable Secur. Comput.* **2021**, *18*, 15–29. [CrossRef]
13. Zhang, Z.; Tang, Q.; Ruiz, R.; Zhang, L. Ergonomic risk and cycle time minimization for the U-shaped worker assignment assembly line balancing problem: A multi-objective approach. *Comput. Oper. Res.* **2020**, *118*, 104905. [CrossRef]
14. Liu, M.; Liang, B.; Zheng, F.; Chu, F. Stochastic airline fleet assignment with risk aversion. *IEEE Trans. Intell. Transp. Syst.* 2019, 20, 3081–3090. [CrossRef]

15. Mahmood, N.; Butalia, T.; Qin, R.; Manasrah, M. Concurrent events risk assessment generic models with enhanced reliability using Fault tree analysis and expanded rotational fuzzy sets. *Expert Syst. Appl.* 2022, 197, 116681. [CrossRef]

16. Xiao, F.; Pedrycz, W. Negation of the quantum mass function for multisource quantum information fusion with its application to pattern classification. *IEEE Trans. Pattern Anal. Mach. Intell.* 2022. [CrossRef]

17. Xiao, F.; Cao, Z.; Lin, T. A complex weighted discounting multisource information fusion with its application in pattern classification. *IEEE Trans. Knowl. Data Eng.* 2022. [CrossRef]

18. Xiao, F. On the maximum entropy negation of a complex-valued distribution. *IEEE Trans. Fuzzy Syst.* 2021, 29, 3259–3269. [CrossRef]

19. Cui, H.; Zhou, L.; Li, Y.; Kang, B. Belief entropy-of-entropy and its application in the cardiac interbeat interval time series analysis. *Chaos Solitons Fractals* 2022, 155, 111736. [CrossRef]

20. Zhang, H.; Deng, Y. Entropy measure for orderable sets. *Inf. Sci.* 2021, 561, 141–151. [CrossRef]

21. Xiao, F. GIQ: A generalized intelligent quality-based approach for fusing multi-source information. *IEEE Trans. Fuzzy Syst.* 2021, 29, 2018–2031. [CrossRef]

22. Deng, X.; Jiang, W.; Wang, Z. Zero-sum polymatrix games with link uncertainty: A Dempster-Shafer theory solution. *Appl. Math. Comput.* 2019, 340, 101–112. [CrossRef]

23. Xue, Y.; Deng, Y. Mobius transformation in generalized evidence theory. *Appl. Intell.* 2021, 52, 7818–7831. [CrossRef]

24. Xiao, F. A new divergence measure for belief functions in D–S evidence theory for multisensor data fusion. *Inf. Sci.* 2020, 514, 462–483. [CrossRef]

25. Deng, J.; Deng, Y.; Cheong, K. Combining conflicting evidence based on Pearson correlation coefficient and weighted graph. *Int. J. Intell. Syst.* 2020, 36, 7443–7460. [CrossRef]

26. Wang, Z.; Xiao, F. An improved multi-source data fusion method based on the belief entropy and divergence measure. *Entropy* 2019, 21, 611. [CrossRef] [PubMed]

27. Deng, Z.; Wang, J. Measuring total uncertainty in evidence theory. *Int. J. Intell. Syst.* 2021, 36, 1721–1745. [CrossRef]

28. Zhao, K.; Li, L.; Chen, Z.; Sun, R.; Yuan, G.; Li, J. A survey: Optimization and applications of evidence fusion algorithm based on Dempster-Shafer theory. *Appl. Soft Comput.* 2022, 124, 106075. [CrossRef]

29. Li, R.; Chen, Z.; Li, H.; Tang, Y. A new distance-based total uncertainty measure in Dempster-Shafer evidence theory. *Appl. Intell.* 2021, 52, 1209–1237. [CrossRef]

30. Qiang, C.; Deng, Y. A new correlation coefficient of mass function in evidence theory and its application in fault diagnosis. *Appl. Intell.* 2021, 52, 7832–7842. [CrossRef]

31. Sudano J. Pignistic probability transforms for mixes of low- and high-probability events. *Comput. Sci.* 2015, 23–27.

32. Smets, P.; Kennes, R. The transferable belief model. *Artif. Intell.* 1994, 66, 191–234. [CrossRef]

33. Pan, L.; Deng, Y. Probability transform based on the ordered weighted averaging and entropy difference. *Int. J. Comput. Commun. Control.* 2020, 15, 3743. [CrossRef]

34. Deng, Z.; Wang, J. A novel decision probability transformation method based on belief interval. *Knowl.-Based Syst.* 2020, 208, 106427. [CrossRef]

35. Huang, C.; Mi, X.; Kang, B. Basic probability assignment to probability distribution function based on the Shapley value approach. *Int. J. Intell. Syst.* 2021, 36, 4210–4236. [CrossRef]

36. Li, M.; Zhang, Q.; Deng, Y. A new probability transformation based on the ordered visibility graph. *Int. J. Comput. Int. Syst.* 2016, 31, 44–67. [CrossRef]

37. Chen, L.; Deng, Y.; Cheong, K. Probability transformation of mass function: A weighted network method based on the ordered visibility graph. *Eng. Artif. Intell.* 2021, 105, 104438. [CrossRef]

38. Rathipriya, R.; Rahman, A.; Dhamodharavadhani, S.; Meero, A.; Yoganandan, G. Demand forecasting model for time-series pharmaceutical data using shallow and deep neural network model. *Neural. Comput. Appl.* 2022, 1–13. [CrossRef]

39. Jagtap, A.; Shinn, Y.; Kawaguchi, K.; Kurniadakis, G. Deep Kronecker neural networks: A general framework for neural networks with adaptive activation functions. *Neurocomputing* 2021, 468, 165–180. [CrossRef]

40. Shannon, C.E. A mathematical theory of communication. *Sigmobile Mob. Comput. Commun. Rev.* 2001, 5, 3–55. [CrossRef]

41. Boob, D.; Dey, S.; Lan, G. Complexity of training ReLU neural network. *Discret. Optim* 2022, 44, 100620. [CrossRef] [PubMed]

42. Xiao, F. Evidence combination based on prospect theory for multi-sensor data fusion. *ISA Trans.* 2020, 106, 253–261. [CrossRef] [PubMed]

43. Xiao, F. Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy. *Inf. Fusion* 2019, 46, 23–32. [CrossRef]

44. Jiang, W.; Hu, W.; Xie, C. A new engine fault diagnosis method based on multi-sensor data fusion. *Appl. Sci.* 2017, 7, 280. [CrossRef]