Probability Transform Based on the Ordered Weighted Averaging and Entropy Difference

L. Pan, Y. Deng

Lipeng Pan
Institute of Fundamental and Frontier Science
University of Electronic Science and Technology of China, Chengdu, 611731

Yong Deng*
1. Institute of Fundamental and Frontier Science
University of Electronic Science and Technology of China, Chengdu, 611731
2. Medical Center, Vanderbilt University, Nashville, TN. USA, 37205
*Corresponding author: dengentropy@uestc.edu.cn, prof.deng@hotmail.com

Abstract

Dempster-Shafer evidence theory can handle imprecise and unknown information, which has attracted many people. In most cases, the mass function can be translated into the probability, which is useful to expand the applications of the D-S evidence theory. However, how to reasonably transfer the mass function to the probability distribution is still an open issue. Hence, the paper proposed a new probability transform method based on the ordered weighted averaging and entropy difference. The new method calculates weights by ordered weighted averaging, and adds entropy difference as one of the measurement indicators. Then achieved the transformation of the minimum entropy difference by adjusting the parameter \( r \) of the weight function. Finally, some numerical examples are given to prove that new method is more reasonable and effective.

Keywords: Dempster-Shafer evidence theory, probability transform, mass function, ordered weighted averaging, entropy difference.

1 Introduction

In engineering practice and scientific research, imprecise information often appears, which increases the difficulty of decision-making [21, 28, 50, 60]. The imprecise information generally has two characteristics: fuzziness and probabilistic. Some researchers have proposed a lot of theories on the fuzziness of information. Such as Fuzzy sets [14, 76], Intuitionistic fuzzy sets [4], Z number [30, 77], Pythagorean fuzzy sets [72], and so on [24, 27]. These theories are applied to decision-making [15, 32, 52, 64], expert systems [41, 48, 49, 66], pattern recognition [19, 55, 59, 81]. On the other hand, some theories have been proposed based on the probabilistic of information, such as Probability theory [57, 75], Dempster-Shafer (D-S) evidence theory [12, 43], D number [9, 10], Evidence reasoning [34, 80], complex evidence theory [58, 62, 63], and so on [33, 71, 74]. Among them, D-S evidence theory has been
used in many fields due it has stronger processing ability in uncertain information. Such as pattern recognition[7, 35, 61, 78], decision making[17, 26, 46, 79], expert system[16, 18, 51, 65], fault diagnosis[25, 31, 53, 54] and so on[20, 23, 36, 37].

In order to enhance the application of Dempster-Shafer evidence theory, there are studies about the mass function translated into probability distribution. Smets proposed transferable belief model(TBM), it can transform mass functions to probability distribution and provide the expected utility theory for decision-making[44]. Cobb and Sheony proposed the plausibility transformation method, which translates each independent mass function to a corresponding probability distribution by plausibility function[6]. Jiang proposed a probabilistic transformation method of adjustable parameters, it relies on the subjective judgment of decision makers[13]. However, existing methods can not consider the loss information after transformation. Hence, it is still a open issue.

In D-S evidence theory, entropy plays an important role. Deng entropy is the extension of Shannon entropy[8], which has attracted many people due it can measure the total nonspecificity and discord [2, 29]. Hence, using Deng entropy can handle the information measurement in this work.

The paper proposed a new method to transform mass functions to probability distributions, which is based on ordered weighted averaging(OWA) operator. The method can flexibly transform different probability distributions by adjusting the parameters \( r \) in OWA operator. Besides, the probability distribution with the minimum information loss can be obtained by using entropy difference of Deng entropy.

The rest of the paper is organized as follows. Main preliminaries about Dempster-Shafer evidence theory, Deng entropy and OWA operator are introduced in section 2. In section 3, new method is presented, and some numerical examples are used to illustrate it. Application in target recognition system is introduced in section 4. Conclusion is given in section 5.

2 Preliminaries

In this section, the preliminaries about D-S evidence theory, Deng entropy and ordered weighted averaging (OWA) will be briefly introduced.

2.1 D-S evidence theory

D-S evidence theory assigns probabilities to the power set of events [12, 43], so as to better grasp the unknown and uncertainty of the things, it offers a useful fusion tool for uncertain information [1, 3, 42, 69, 73]. Some preliminaries in D-S theory are introduced as follows. For additional details about D-S evidence theory, refer to [12, 43].

**Definition 1. (Frame of discernment)**

For a mutually exclusive set \( \Theta \) composed of \( A_i \), it is defined as

\[
\Theta = \{ A_1, A_2, \cdots, A_n \}
\]  

(1)

The power set of \( \Theta \) is indicated by \( 2^{\Theta} \):

\[
2^{\Theta} = \{ \phi, \{ A_1 \}, \{ A_2 \}, \cdots, \{ A_1, A_2 \}, \cdots, \Theta \}
\]  

(2)

**Definition 2. (Mass Function)**

For a frame of discernment \( \Theta = \{ A_1, A_2, \cdots, A_n \} \), a mapping of \( m \) from 0 to 1 is defined as :

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

(3)

which satisfies

\[
m(\phi) = 0
\]  

(4)

\[
\sum_{A \subseteq \Theta} m(A) = 1
\]  

(5)

where \( m(A) \) represents the degree of evidence supports \( A \).
Definition 3. (Belief Function and Plausibility Function)
For any $A \subseteq \Theta$, the belief function $\text{Bel} : 2^\Theta \rightarrow [0, 1]$ is defined as

$$\text{Bel}(A) = \sum_{B \subseteq A} m(B)$$  \hspace{1cm} (6)

The plausibility function $\text{Pl} : 2^\Theta \rightarrow [0, 1]$ is defined as

$$\text{Pl}(A) = 1 - \text{Bel}(\overline{A}) = \sum_{B \cap A \neq \emptyset} m(B)$$  \hspace{1cm} (7)

Obviously, $\text{Pl}(A) \geq \text{Bel}(A)$, the Bel($A$) and Pl($A$) are lower limit function and upper limit function of $A$. When $A$ is a single set in power set $2^\Theta$, the equal sign is established.

Definition 4. (Pignistic Probability Transformation)
For a mass function $m$ in the frame of discernment $\Theta$, its associated pignistic probability transformation (PPT) function $\text{BetP}_m : \Theta \rightarrow [0, 1]$ is defined as[44]:

$$\text{BetP}_m(w) = \sum_{A \subseteq P(\Theta), w \in A} \frac{1}{|A|} \frac{m(A)}{1 - m(\phi)}$$  \hspace{1cm} (8)

where $m(\phi) \neq 1, |A|$ is the cardinality of subset $A$. In this way, mass functions can be transferred to probability distribution.

Definition 5. (Distance Between Betting Commitments of the Two mass functions)
Let $m_1$ and $m_2$ be two mass functions on frame $\Theta$ and let $\text{BetP}_{m_1}$ and $\text{BetP}_{m_2}$ be the results of two pignistic transformations from them respectively. Then distance between betting commitments of the two mass functions is described as follows[5]:

$$\text{difBetP}_{m_2}^m = \max_{A \subseteq \Theta} \left( |\text{BetP}_{m_1(w)} - \text{BetP}_{m_2(w)}| \right)$$  \hspace{1cm} (9)

Thus, similarity between betting commitments of the two mass functions is described as follows:

$$\text{Sim}_{m_1,m_2} = 1 - \text{difBetP}_{m_2}^m$$  \hspace{1cm} (10)

Definition 6. (Deng Entropy)
In D-S evidence theory, Deng entropy can be used to measure uncertainty, it is defined as[8]:

$$E_d = \sum m(A) \log \frac{m(A)}{2^{|A|} - 1}$$  \hspace{1cm} (11)

where $m(A)$ is a mass function in the frame of discernment $\Theta$, $|A|$ is the cardinality of subset $A$.

2.2 Ordered Weighted Averaging

Yager proposed a kind of information settlement operator bounded between "and" and "or", that is, the ordered weighted averaging (OWA) operator[67, 70]. The operator reorders the data in the order from large to small, and then carries out weight processing combine with the location of the data, it can eliminate unreasonable situation well. Thus, OWA operator has been widely applied to complex network[56], fuzzy theory[45], information fusion[39], expert system[38], decision-making[68]. One of the core of OWA opeartor theory is solution method of OWA weight, which usually uses the function $f : [0, 1] \rightarrow [0, 1]$. The function satisfies the following properties[68].

- $\circled{1} f(0) = 0; f(1) = 1$.
- $\circled{2}$ $f$ is monotonic, $f(a) \geq f(b)$ if $a > b$.

In particular using this function, The weight of OWA $w_k (k = 1, 2, \cdots, q)$ can be obtain as[68]:

$$w_k = f\left(\frac{k}{q}\right) - f\left(\frac{k-1}{q}\right)$$  \hspace{1cm} (12)

Since the function $f$ is monotonic, then $w_k = f\left(\frac{k}{q}\right) - f\left(\frac{k-1}{q}\right) \geq 0$. 
3 The proposed method

In the section, a new method is presented, which translates the multielement subsets to single subset, and an example to illustrate the method.

Definition 7. (Mean Function)

For a frame of discernment \( \Theta \), \( m \) is a mass function. The mean function of \( \text{Bel}(A_i) \) and \( \text{Pl}(A_i) \) is defined as:

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

where \( A_i \) is a single subset.

Definition 8. (Normalized Function)

\[
\varrho(A_i) = \frac{M_{[\text{Bel,Pl}]}(A_i)}{\sum_{F_j \subseteq P(\Theta), A_j \in F_j} M_{[\text{Bel,Pl}]}(A_i)}
\]

where \( \varrho(A_i) \) is the result of \( M_{[\text{Bel,Pl}]}(A_i) \) normalization, satisfying \( \sum \varrho(A_i) = 1 \).

Definition 9. (Weight Function)

One classical of function that can be used to generate the weight, it is defined as[68]:

\[
f(x) = x^r, \ (r > 0)
\]

Then, denoting \( T_i = \sum_{k=1}^{i} \varrho(A_k) \), and \( T_0 = 0 \).

Finally,

\[
P_m(A_i) = (T_i)^r - (T_{i-1})^r
\]

Obviously, \( P_m(A_i) \) is the result of translating the multielement subsets to single subset.

Hence, a example is used to illustrate the transformation process.

Example 1 : For a frame of discernment \( \Theta = \{\theta_1, \theta_2, \theta_3\} \), \( m \) is a mass function in \( \Theta \), there is \( m(\theta_1) = 0.2, m(\theta_2) = 0.1, m(\theta_1, \theta_2) = 0.3, m(\theta_2, \theta_3) = 0.25, m(\theta_1, \theta_2, \theta_3) = 0.15 \).

Firstly, \( E_q13 \) is used to calculate \( M_{[\text{Bel,Pl}]}(\{i\}) \):

\[
M_{[\text{Bel,Pl}]}(\{\theta_1\}) = 0.425, M_{[\text{Bel,Pl}]}(\{\theta_2\}) = 0.45, M_{[\text{Bel,Pl}]}(\{\theta_3\}) = 0.2.
\]

Then, standard normalization is realized by \( E_q14 \):

\[
\varrho(\theta_1) = 0.395, \varrho(\theta_2) = 0.419, \varrho(\theta_3) = 0.186.
\]

Thus,

\[
T_0 = 0, T_1 = 0.395, T_2 = 0.814, T_3 = 1.
\]

Assuming \( r = 0.5 \), the final result is as follows:

\[
P_m(\theta_1) = 0.628, P_m(\theta_2) = 0.274, P_m(\theta_3) = 0.098.
\]

Or \( r = 1.5 \), the result is as follows:

\[
P_m(\theta_1) = 0.248, P_m(\theta_2) = 0.486, P_m(\theta_3) = 0.266.
\]

As can be from the results of the case, a change in the value of \( r \) will cause a change in the result of transformation. Obviously, when \( r = 1 \), then \( P_m(\theta_1) = \varrho(\theta_1), P_m(\theta_2) = \varrho(\theta_2), P_m(\theta_3) = \varrho(\theta_3) \).

Therefore, the transformation result will be more diverse under the adjustment of \( r \) value. But can \( r \) be any number?

In Example 1, when \( r \to 0 \), then \( P_m(\theta_1) = 1 \). If \( r \to +\infty \), the \( P_m(\theta_3) = 1 \). Obviously, they are unreasonable. Since in the frame of discernment \( \Theta \), \( \text{Bel}(\theta_i) \leq m(\theta_i) \leq \text{Pl}(\theta_i) \) must be satisfied, that is, \( \text{Bel}(\theta_i) \leq P_m(\theta_i) \leq \text{Pl}(\theta_i) \) is also satisfied, but \( P_m(\theta_1) = 1 \) or \( P_m(\theta_3) = 1 \) violated the rule. Therefore, this conclusion cannot be admitted.

In order to make the transformation result conform to the rule, belief function and plausibility function are considered in the transformation process, result in the following equations:

\[
\begin{align*}
&f(T_k) - f(T_{k-1}) \\
&\text{Bel}(\theta_i) \leq P_m(\theta_i) \leq \text{Pl}(\theta_i)
\end{align*}
\]
Figure 1: The result of the probability transformation when \( r \) changes

The range of \( r \) is \([0.464, 1.732]\) by using Eq.13 in Example 1. Under the condition, the transformation probability distribution is shown in Figure.1.

As can be seen from Figure.1, the probability distributions of three element \((\theta_1, \theta_2, \theta_3)\) are between belief function and plausibility function. \( r \) starts to change from 0.464, when \( r \) increases, \( \theta_1 \) can decrease, and \( \theta_2, \theta_3 \) can increase. And \( r = 1.025 \), then \( \theta_1 = \theta_2 \), after that, \( \theta_2 \) continued to increase. When \( r = 1.732 \), \( \theta_2 \) reaches a maximum value of 0.5001.

During the process of translating mass functions to probability distributions, the change of information is also a noteworthy indicator. In this section, a loss function is used to measure the change of information, it is defined as follows:

**Definition 10. (Loss Function)**

\[
d_E = |E_d - H|
\]

where \( d_E \) is the difference between the Deng entropy of mass functions before transformation and Shannon entropy of the probability distribution after transformation. The smaller the value of \( d_E \), the less information is lost. Using the **Definition 10**, Eq. 17 is reconstructed as follows:

\[
\begin{align*}
&\min \{d_E\} \\
&f(T_k) - f(T_{k-1}) \\
&\text{Bel}(\theta_i) \leq P_m(\theta_i) \leq P_l(\theta_i)
\end{align*}
\]

By using Eq. 19, we can get that when \( r = 1.244 \), \( d_E \) is the minimum 1.9956, and \( P_m(\theta_1) = 0.3149, P_m(\theta_2) = 0.4592, P_m(\theta_3) = 0.2259 \).

In addition, PPT, PFT, IPT and DSmP1 are used in Example 1, and transformation results are shown in Table 1. As can seen from Table 1, the transformation result of the new method is relatively clear. Figure. 2 shows the entropy difference before and after the transformation in this case. It can be seen from Figure.2 that the entropy difference of the new method before and after the transformation is smallest, indicating that the information loss of the proposed method after the transformation is relatively small. This shows that the new method is relatively reasonable.
Table 1: Result of various methods

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_m(\theta_1)$</td>
<td>0. 4</td>
<td>0. 34</td>
<td>0. 4</td>
<td>0. 41</td>
<td>0. 3149</td>
</tr>
<tr>
<td>$P_m(\theta_2)$</td>
<td>0. 425</td>
<td>0. 42</td>
<td>0. 42</td>
<td>0. 42</td>
<td>0. 4592</td>
</tr>
<tr>
<td>$P_m(\theta_3)$</td>
<td>0. 175</td>
<td>0. 237</td>
<td>0. 18</td>
<td>0. 17</td>
<td>0. 2259</td>
</tr>
</tbody>
</table>

The difference between the entropy($dE$)
Different models
(1.9956)

Figure 2: The entropy difference in various methods

4 Application in target recognition system

In the section, an example is used to illustrate the application about new method. A numerical example about multisensor-based automatic target recognition system was proposed[11]. In multisensor-based automatic target recognition system, there are three known targets, $\theta_1, \theta_2, \theta_3$. The output results of sensors are shown in Table 2. The main process of target recognition is described as follows:

**Step 1**: Obtained evidence from sensors.
Independent evidences from different sensors, denoted as $m_1, m_2, ..., m_n$, respectively.

**Step 2**: Probability transform by new method.
The probability distribution corresponding to mass function can be obtained by Eq.19.

**Step 3**: Measured the weight of evidence $W_{m_i}$.
Firstly, the similarity between probability distributions was measured by Eq.9 and Eq.10, denoted as $Sim_{m_i,m_j}$. Secondly, sum over each row of the similar matrix by $Crd_{m_i} = \sum_{j=1}^{k} Sim_{m_i,m_j}$. Finally, using the following equation to measure the weight of evidence:

$$W_{m_i} = \frac{Crd_{m_i}}{\sum_{i=1}^{n} Crd_{m_i}}$$  \hspace{1cm} (20)

**Step 4**: Modified the mass functions of evidence.
For event $A$, its mass function is modified as follows:

$$m(A) = \sum_{i=1}^{n} W_{m_i} \times m_i(A)$$  \hspace{1cm} (21)

**Step 5**: Used Dempster rule by N-1.
The mass function can be combined N-1 times by Dempster combination rule to get the result of recognition.

Step 6: Probability transform by new method and decision making.

<table>
<thead>
<tr>
<th>Table 2: Output of the sensors[11]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_i )</td>
</tr>
<tr>
<td>( m_1 )</td>
</tr>
<tr>
<td>( m_2 )</td>
</tr>
<tr>
<td>( m_3 )</td>
</tr>
<tr>
<td>( m_4 )</td>
</tr>
<tr>
<td>( m_5 )</td>
</tr>
</tbody>
</table>

The recognition results of various method are listed in Table 3. As can be seen from Table 3 that the final recognition target is \( \theta_1 \), but the recognition process is different. After the first combination, target \( \theta_2 \) was recognized in all methods. In the subsequent recognition process, the support of \( \theta_1 \) in the new method is higher, it is more effective than other methods. The maximum support for \( \theta_1 \) is 0.9852. Therefore, the result of proposed method is more reasonable relatively.

<table>
<thead>
<tr>
<th>Table 3: Results of different combination rules of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_1, m_2 )</td>
</tr>
<tr>
<td>( m (\theta_1) = 0 )</td>
</tr>
<tr>
<td>( m (\theta_2) = 0.8571 )</td>
</tr>
<tr>
<td>( m (\theta_3) = 0.1429 )</td>
</tr>
<tr>
<td>( m (\theta_1) = 0.1543 )</td>
</tr>
<tr>
<td>( m (\theta_2) = 0.7469 )</td>
</tr>
<tr>
<td>( m (\theta_3) = 0.0988 )</td>
</tr>
<tr>
<td>( m (\theta_1, \theta_3) = 0.0093 )</td>
</tr>
<tr>
<td>( m (\theta_2) = 0.299 )</td>
</tr>
<tr>
<td>( m (\theta_3) = 0.174 )</td>
</tr>
<tr>
<td>( m (\theta_1) = 0.2624 )</td>
</tr>
<tr>
<td>( m (\theta_2) = 0.6455 )</td>
</tr>
<tr>
<td>( m (\theta_3) = 0.0921 )</td>
</tr>
</tbody>
</table>

5 Conclusion

In the paper, a transformation method is proposed, it based on OWA, which can be transformation flexibly by adjusting the parameter \( r \) of weight function. The proposed method is constrained by belief function and plausibility function. In addition, the entropy difference is taken as the measurement standard, and the optimal solution of transformation can be finally obtained under the condition. Examples are used to prove that the new method is relatively reasonable. However, there are some issues that need to be solved in the future. For example, \( m (\theta_1) = \frac{1}{5}, m (\theta_2, \theta_3) = \frac{2}{5}, \) when entropy
difference is the smallest, the result of the new method is a uniform probability distribution, and other methods also can produce the same results. It increases the difficulty of decision-making. For this case, the current approaches are that reasonably improve the method of generating mass functions or collecting data again. Thus, the mass function transformation to the probability distribution is still a lot of potential work to be done.

Acknowledgment

The work is partially supported by National Natural Science Foundation of China (Grant No. 61973332).

References


