Is Entropy Enough to Evaluate the Probability Transformation Approach of Belief Function?

Deqiang Han
Jean Dezert
Chongzhao Han
Yi Yang

Originally published as Han D., Dezert J., Han C., *Is Entropy Enough to Evaluate the Probability Transformation Approach of Belief Function?*, Fusion 2010, Edinburgh, Scotland, UK, 26-29 July 2010, and reprinted with permission.

Abstract – *In Dempster-Shafer Theory (DST) of evidencee and transferable belief model (TBM), the probability transformation is necessary and crucial for decision-making. The evaluation of the quality of the probability transformation is usually based on the entropy or the probabilistic information content (PIC) measures, which are questioned in this paper. Another alternative of probability transformation approach is proposed based on the uncertainty minimization to verify the rationality of the entropy or PIC as the evaluation criteria for the probability transformation. According to the experimental results based on the comparisons among different probability transformation approaches, the rationality of using entropy or Probabilistic Information Content (PIC) measures to evaluate probability transformation approaches is analyzed and discussed.*

Keywords: TBM, uncertainty, pignistic probability transformation, evidence theory, decision-making.

1 Introduction

Evidence theory, as known as Dempster-Shafer Theory (DST) [1, 2] can reason with imperfect information including imprecision, uncertainty, incompleteness, etc. It is widely used in many fields in information fusion. There are also some drawbacks and problems in evidence theory, i.e. the high computational complexity, the counter-intuitive behaviors of Dempster's combination rule and the decision-making in evidence theory, etc. Several modified, refined or extended models were proposed to resolve the problems aforementioned, such as transferable belief model (TBM) [3] proposed by Philippe Smets and Dezert-Smarandache Theory (DSmT) [4] proposed by Jean Dezert and Florentin Smarandache, etc.

The goal of uncertainty reasoning is the decision-making. To take a decision, the belief assignment values for a compound focal element should be at first assigned to the singletons. So the probability transformation from belief function is crucial for the decision-making in evidence theory. The research on probability transformation has attracted more attention in recent years.

The most famous probability transformation in evidence theory is the pignistic probability transformation (PPT) in TBM. TBM has two levels including credal level and pignistic level. At the credal level, beliefs are entertained, combined and updated while at the pignistic level, the PPT maps the beliefs defined on subsets to the probability defined on singletons, then a classical probabilistic decision can be made. In PPT, belief assignment values for a compound focal element are equally assigned to the singletons belonging to the focal element. In fact, PPT is designed according to principle of minimal commitment, which is somehow related with uncertainty maximization. But the goal of information fusion at decision-level is to reduce the uncertainty degree. That is to say more uncertainty might not be helpful for the decision. PPT uses equal weights when splitting masses of belief of partial uncertainties and redistributing them back to singletons included in them. Other researchers also proposed some modified probability transformation approaches [5–13] to assign the belief assignment values of compound focal elements to the singletons according to some ratio constructed based on some available information. The typical approaches include the Sudano's probabilities [8] and the Cuzzolin's intersection probability [13], etc. In the framework of DSmT, another probability transformation approach was proposed, which is called DSmP [9]. DSmP takes into account both the values of the masses and the cardinality of focal elements in the proportional redistribution process. DSmP can also be used in Shafer's model within DST framework.

In almost all the research works on probability transformations, the entropy or Probabilistic Information Content (PIC) criteria are used to evaluate the probability transformation approaches. Definitely for the purpose of decision, less uncertainty should be better to make a more clear and solid decision. But does the probability distribution generated from belief functions with less uncertainty always rational or always be benefit to the decision? We do not think so. In this paper, an alternative probability transformation approach based on the uncertainty minimization is proposed. The objective function is established based on the Shannon entropy and the constraints are established based on the given belief and plausibility functions. The experimental results based on some provided numerical examples show that the probability distributions generated based on the proposed alternative approach have the least uncertainty degree when compared with other approaches. When using the entropy or PIC to evaluate the proposed probability transformation approach, the probability distribution with the least uncertainty seemingly should be the optimal one. But some risky and strange results can be derived in some cases, which are illustrated in some numerical examples. It can be concluded that the entropy or PIC, i.e. the uncertainty degree might not be enough to evaluate the probability transformation approach. In another word, the entropy or PIC might not be used as the only criterion to make the evaluation.

2 Basics of evidence theory and probability transformation

2.1 Basics of evidence theory

In Dempster-Shafer theory [2], the elements in the frame of discernment (FOD) Θ are mutually exclusive. Define the function $m : 2^\Theta \to [0, 1]$ as the basic probability assignment (BPA, also called mass function), which satisfies:

$$\sum\nolimits_{A \subseteq \Theta} m(A) = 1,\ m(\emptyset) = 0 \tag{1}$$

Belief function and plausibility function are defined respectively in (2) and (3):

$$Bel(A) = \sum\nolimits_{B \subseteq A} m(B) \tag{2}$$

$$pl(A) = \sum\nolimits_{A \cap B \neq \emptyset} m(B) \tag{3}$$

and Dempster's rule of combination is defined as follows: $m_1, m_2, ..., m_n$ are n mass functions, the new combined evidence can be derived based on (4)

$$m(A) = \begin{cases} 0, \quad A = \emptyset \\ \dfrac{\sum\limits_{\cap A_i = A} \prod\limits_{1 \leq i \leq n} m_i(A_i)}{\sum\limits_{\cap A_i \neq \emptyset} \prod\limits_{1 \leq i \leq n} m_i(A_i)}, \quad A \neq \emptyset \end{cases} \tag{4}$$

Dempster's rule of combination is used in DST to accomplish the fusion of bodies of evidence. But the final goal of the information fusion at decision-level is to make the decision. The belief function (or BPA, plausibility function) should be transformed to the probability, before the probability-based decision-making. Although there are also some research works on making decision directly based on belief function or BPA [14], probability-based decision methods are the development trends of uncertainty reasoning and theories [15]. This is because the two-level reasoning and decision structure proposed by Smets in his TBM is appealing.

2.2 Pignistic transformation

As a type of probability transformation approach, the classical pignistic probability in TBM framework was coined by Philippe Smets. TBM is a subjective and non probabilistic interpretation of evidence theory. It extends the evidence theory to the open–world propositions and it has a range of tools for handling belief functions including discounting and conditioning, etc. At the credal level of TBM, beliefs are entertained, combined and updated while at the pignistic level, beliefs are used to make decisions by transforming beliefs to probability distribution based on pignistic probability transformation (PPT). The basic idea of the pignistic transformation consists in transferring the positive belief of each compound (or nonspecific) element onto the singletons involved in that element split by the cardinality of the proposition when working with normalized BPAs.

Suppose that $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$ is the FOD. The PPT for the singletons is illustrated as follows [3]:

$$\mathrm{BetP}_m(\theta_i) = \sum_{\theta_i \in B,\ B \subseteq 2^\Theta} \frac{m(B)}{|B|} \tag{5}$$

where 2^Θ is the power set of the FOD. Based on the pignistic probability derived, the corresponding decision can be made.

But in fact, PPT is designed according to the idea being similar to uncertainty maximization. In general, the PPT is just a simple averaging operation. The mass value is not assigned discriminately to the different singletons involved. But for information fusion, the aim is to reduce the degree of uncertainty and to gain a more consolidated and reliable decision result. The high uncertainty in PPT might be not helpful for the decision. Several researchers aim to modify the traditional PPT. Some typical modified probability transformation approaches are as follows.

1) Sudano's probabilities: Sudano [8] proposed some interesting alternatives to PPT denoted by PrPl, PrNPl, PraPl, PrBel and PrHyb, respectively. Sudano uses different kinds of mappings either proportional to the plausibility, to the normalized plausibility, to all plausibilities and to the belief, respectively or a hybrid mapping.

2) Cuzzolin's intersection probability: In the framework of DST, Fabio Cuzzolin [13] proposed another type of transformation. From a geometric interpretation of Dempster's combination rule, an intersection probability measure was proposed from the proportional repartition of the total non specific mass (TNSM) by each contribution of the non-specific masses involved in it.

3) DSmP: Dezert and Smarandache proposed the DSmP as follows: Suppose that the FOD is $\Theta = \{\theta_1, ..., \theta_n\}$, the $\mathrm{DSmP}_\varepsilon(\theta_i)$can be directly obtained by:

$$\mathrm{DSmP}_\varepsilon(\theta_i) = m(\{\theta_i\}) + (m(\{\theta_i\}) + \varepsilon)\cdot \Big(\sum_{\substack{X\in 2^\Theta \\ \theta_i \subset X \\ |X|\geq 2}} \frac{m(X)}{\sum\limits_{\substack{Y\in 2^\Theta \\ Y\subset X \\ |Y|=1}} m(Y) + \varepsilon\cdot |X|} \Big) \tag{6}$$

In DSmP, both the values of the mass assignment and the cardinality of focal elements are used in the proportional redistribution process. DSmP does an *improvement* of all Sudano, Cuzzolin, and BetP formulas, in the sense that DSmP mathematically makes a more accurate redistribution of the ignorance masses to the singletons involved in ignorance. DSmP works in both theories: DST and DSmT as well.

There are still some other definitions on modified PPT such as the iterative and self-consistent approach PrScP proposed by Sudano in [5] and a modified PrScP in [12]. Although the approaches aforementioned are different, all the probability transformation approaches are evaluated based on the degree of uncertainty. Less uncertainty means that the corresponding probability transformation result is better. According to such a idea, the probability transformation approach should attempt to enlarge the belief differences among all the propositions and thus to derive a more reliable decision result. Is this definitely rational? Is the uncertainty degree always proper or enough to evaluate the probability transformation? In the following section, some uncertainty measures are analyzed and an alternative probability transformation approach based on uncertainty minimization is proposed to verify the rationality of the uncertainty degree as the criteria for evaluating the probability transformation.

3 An alternative probability transformation based on uncertainty minimization

3.1 Evaluation criteria for probability transformation

The metrics depicting the strength of a critical decision by a specific probability distribution are introduced as follows:

1) Normalized Shannon entropy

Suppose that p_θ is a probability distribution, where $\theta \in \Theta$, $|\Theta| = N$ and the $|\Theta|$ represents the cardinality of the FOD Θ. The evaluation criterion for the probability distribution derived based on different probability transformation is as follows [12].

$$\mathrm{E_H} = \frac{-\sum\limits_{\theta\in\Theta} p_\theta \log_2(p_\theta)}{\log_2 N} \tag{7}$$

The dividend in (7) is the Shannon entropy and the divisor in (7) is maximum value of the Shannon entropy for $\{p_\theta|\theta \in \Theta\}$,$|\Theta| = N$. Obviously E_H is normalized. The larger the E_H is, the larger the degree of uncertainty is. The less the E_H is, the less the degree of uncertainty is. When $E_H = 0$, there is only one hypothesis has a probability value of 1 and the rest has 0, the agent or system can make decision correctly. When $E_H = 1$, it is impossible to make a correct decision, because all the $p_\theta, \forall\theta \in \Theta$ are equal.

2) Probabilistic Information Content

Probabilistic Information Content (PIC) criterion is an essential measure in any threshold-driven automated decision system. A PIC value of one indicates the total knowledge to make a correct decision.

$$\mathrm{PIC}(P) = 1 + \frac{1}{\log_2 N} \cdot \sum_{\theta\in\Theta} p_\theta \log_2(p_\theta) \tag{8}$$

Obviously, $\mathrm{PIC} = 1 - \mathrm{E_H}$. The PIC is the dual of the normalized Shannon entropy. A PIC value of zero indicates that the knowledge to make a correct decision does not exist (all the hypotheses have an equal probability value), i.e. one has the maximal entropy.

As referred above, for information fusion at decision-level, the uncertainty seemingly should be reduced as much as possible. The less the uncertainty in probability measure is, the more consolidated and reliable decision can be made. Suppose such a viewpoint is always right and according to such an idea, an alternative probability transformation of belief function is proposed.

3.2 Probability transformation of belief function based on uncertainty minimization

To accomplish the probability transformation, the belief function (or the BPA, the plausibility function) should be available. The relationship between the probability and the belief function are analyzed as follows.

Based on the viewpoint of Dempster and Shafer, the belief function can be considered as a lower probability and the plausibility can be considered as an upper probability. Suppose that $p_\theta \in [0,1]$ is a probability distribution, where $\theta \in \Theta$. For a belief function defined on FOD Θ, suppose that $B \in 2^\Theta$, the inequality (9) is satisfied:

$$Bel(B) \leq \sum\nolimits_{\theta \in B} p_\theta \leq Pl(B) \tag{9}$$

This inequality can be proved according to the properties of the upper and lower probability.

Probability distributions $\langle p_\theta | \theta \in \Theta \rangle$ also must meet the usual requirements for probability distributions, i.e.

$$\begin{cases} 0 \leq p_\theta \leq 1, \forall \theta \in \Theta \\ \sum_{\theta \in \Theta} p_\theta = 1 \end{cases} \tag{10}$$

It can be taken for granted that there are several probability distributions $\{p_\theta | \theta \in \Theta\}$ consistent with the given belief function according to the relationships defined in (9) and (10). This is a multi-answer problem or one-to-many mapping relation. As referred above, the probability is used for decision, so the uncertainty seemingly should be as little as possible. We can select one probability distribution from all the consistent alternatives according to the uncertainty minimization criterion and use the corresponding probability distribution as the result of the probability transformation.

The Shannon entropy is used here to establish the objective function. The equations and inequalities in (9) and (10) are used to establish the constraints. The problem of probability transform of belief function here is converted to an optimization problem under constraints as follows:

$$\begin{aligned} &\underset{\{p_\theta | \theta \in \Theta\}}{\text{Min}} \left\{ -\sum_{\theta \in \Theta} p_\theta \log_2(p_\theta) \right\} \\ &s.t. \begin{cases} Bel(B) \leq \sum_{\theta \in B} p_\theta \leq Pl(B) \\ 0 \leq p_\theta \leq 1, \forall \theta \in \Theta \\ \sum_{\theta \in \Theta} p_\theta = 1 \end{cases} \end{aligned} \tag{11}$$

Given belief function (or the BPA, the plausibility), by solving (11), a probability distribution can be derived, which has least uncertainty measured by Shannon entropy and thus is seemingly more proper to be used in decision procedure.

It is clear that the problem of finding a minimum entropy probability distribution does not admit a unique solution in general. The optimization algorithm used is the Quasi-Newton followed by a global optimization algorithm [16] to alleviate the effect of the local extremum problem. Other intelligent optimization algorithms [17, 18] can also be used,such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), etc.

4 Analysis based on examples

At first, two numerical examples are first provided to illustrate some probability transformation approaches. To make the different approaches reviewed and proposed in this paper more comparable, the examples in [6, 12] are directly used here. The PIC is used to evaluate the probability transformation.

4.1 Example 1

For FOD $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$, the corresponding BPA is as follows:

$m(\{\theta_1\}) = 0.16,\ m(\{\theta_2\}) = 0.14,\ m(\{\theta_3\}) = 0.01,$
$m(\{\theta_4\}) = 0.02,$
$m(\{\theta_1, \theta_2\}) = 0.20,\ m(\{\theta_1, \theta_3\}) = 0.09,$
$m(\{\theta_1, \theta_4\}) = 0.04,\ m(\{\theta_2, \theta_3\}) = 0.04,$
$m(\{\theta_2, \theta_4\}) = 0.02,\ m(\{\theta_3, \theta_4\}) = 0.01,$
$m(\{\theta_1, \theta_2, \theta_3\}) = 0.10,\ m(\{\theta_1, \theta_2, \theta_4\}) = 0.03,$
$m(\{\theta_1, \theta_3, \theta_4\}) = 0.03,\ m(\{\theta_2, \theta_3, \theta_4\}) = 0.03,$
$m(\Theta) = 0.08.$

The corresponding belief functions are calculated and listed as follows:

$Bel(\{\theta_1\}) = 0.16,\ Bel(\{\theta_2\}) = 0.14,$
$Bel(\{\theta_3\}) = 0.01,\ Bel(\{\theta_4\}) = 0.02,$
$Bel(\{\theta_1, \theta_2\}) = 0.50,\ Bel(\{\theta_1, \theta_3\}) = 0.26,$
$Bel(\{\theta_1, \theta_4\}) = 0.22,\ Bel(\{\theta_2, \theta_3\}) = 0.19,$
$Bel(\{\theta_2, \theta_4\}) = 0.18,\ Bel(\{\theta_3, \theta_4\}) = 0.04,$
$Bel(\{\theta_1, \theta_2, \theta_3\}) = 0.74,\ Bel(\{\theta_1, \theta_2, \theta_4\}) = 0.61,$
$Bel(\{\theta_1, \theta_3, \theta_4\}) = 0.36,\ Bel(\{\theta_2, \theta_3, \theta_4\}) = 0.27,$
$Bel(\Theta) = 1.00.$

The corresponding plausibility functions are calculated and listed as follows:

$Pl(\{\theta_1\}) = 0.73,\ Pl(\{\theta_2\}) = 0.64,\ Pl(\{\theta_3\}) = 0.39,$
$Pl(\{\theta_4\}) = 0.26,$
$Pl(\{\theta_1, \theta_2\}) = 0.96,\ Pl(\{\theta_1, \theta_3\}) = 0.82,$
$Pl(\{\theta_1, \theta_4\}) = 0.81,\ Pl(\{\theta_2, \theta_3\}) = 0.78,$
$Pl(\{\theta_2, \theta_4\}) = 0.74,\ Pl(\{\theta_3, \theta_4\}) = 0.50,$
$Pl(\{\theta_1, \theta_2, \theta_3\}) = 0.98,\ Pl(\{\theta_1, \theta_2, \theta_4\}) = 0.99,$
$Pl(\{\theta_1, \theta_3, \theta_4\}) = 0.86,\ Pl(\{\theta_2, \theta_3, \theta_4\}) = 0.84,$
$Pl(\Theta) = 1.00.$

Suppose the probability distribution as the unknown variables. Based on the plausibility functions and the belief functions, the constraints and the objective function can be established according to (11). The probability distribution can be derived based on the minimization. The results of some other probability transformation approaches are also calculated. All the results are listed in Table 1 (on the next page) to make the comparison between the approach proposed in this paper (denoted by Un_min) and other available approaches.

4.2 Example 2

For FOD $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$, the corresponding BBA is as follows:

$m(\{\theta_1\}) = 0.05,\ m(\{\theta_2\}) = 0.00,\ m(\{\theta_3\}) = 0.00,$
$m(\{\theta_4\}) = 0.00,$

Table **1** Probability Transformation Results of Example 1 based on Different Approaches

	θ_1	θ_2	θ_3	θ_4	PIC
BetP [3]	0.3983	0.3433	0.1533	0.1050	0.0926
PraPl [8]	0.4021	0.3523	0.1394	0.1062	0.1007
PrPl [8]	0.4544	0.3609	0.1176	0.0671	0.1638
PrHyb [8]	0.4749	0.3749	0.0904	0.0598	0.2014
PrBel [8]	0.5176	0.4051	0.0303	0.0470	0.3100
FPT[11]	0.5176	0.4051	0.0303	0.0470	0.3100
DSmP_0[9]	0.5176	0.4051	0.0303	0.0470	0.3100
PrScP [10]	0.5403	0.3883	0.0316	0.0393	0.3247
PrBP1 [12]	0.5419	0.3998	0.0243	0.0340	0.3480
PrBP2 [12]	0.5578	0.3842	0.0226	0.0353	0.3529
PrBP3 [12]	0.0605	0.3391	0.0255	0.0309	0.3710
Un_min	0.7300	0.2300	0.0100	0.0300	0.4813

$m(\{\theta_1,\theta_2\}) = 0.39,\ m(\{\theta_1,\theta_3\}) = 0.19,$
$m(\{\theta_1,\theta_4\}) = 0.18, m(\{\theta_2,\theta_3\}) = 0.04,$
$m(\{\theta_2,\theta_4\}) = 0.02,\ m(\{\theta_3,\theta_4\}) = 0.01,$
$m(\{\theta_1,\theta_2,\theta_3\}) = 0.04,\ m(\{\theta_1,\theta_2,\theta_4\}) = 0.02,$
$m(\{\theta_1,\theta_3,\theta_4\}) = 0.03,\ m(\{\theta_2,\theta_3,\theta_4\}) = 0.03,$
$m(\Theta) = 0.00.$

The corresponding belief functions are calculated and listed as follows:

$Bel(\{\theta_1\}) = 0.05,\ Bel(\{\theta_2\}) = 0.00,$
$Bel(\{\theta_3\}) = 0.00,\ Bel(\{\theta_4\}) = 0.00,$
$Bel(\{\theta_1,\theta_2\}) = 0.44,\ Bel(\{\theta_1,\theta_3\}) = 0.24,$
$Bel(\{\theta_1,\theta_4\}) = 0.23,\ Bel(\{\theta_2,\theta_3\}) = 0.04,$
$Bel(\{\theta_2,\theta_4\}) = 0.02,\ Bel(\{\theta_3,\theta_4\}) = 0.01,$
$Bel(\{\theta_1,\theta_2,\theta_3\}) = 0.71,\ Bel(\{\theta_1,\theta_2,\theta_4\}) = 0.66,$
$Bel(\{\theta_1,\theta_3,\theta_4\}) = 0.46,\ Bel(\{\theta_2,\theta_3,\theta_4\}) = 0.10,$
$Bel(\Theta) = 1.00.$

The corresponding plausibility functions are calculated and listed as follows:

$Pl(\{\theta_1\}) = 0.90,\ Pl(\{\theta_2\}) = 0.54,\ Pl(\{\theta_3\}) = 0.34,$
$Pl(\{\theta_4\}) = 0.29,$
$Pl(\{\theta_1,\theta_2\}) = 0.99,\ Pl(\{\theta_1,\theta_3\}) = 0.98,$
$Pl(\{\theta_1,\theta_4\}) = 0.96,\ Pl(\{\theta_2,\theta_3\}) = 0.77,$
$Pl(\{\theta_2,\theta_4\}) = 0.76,\ Pl(\{\theta_3,\theta_4\}) = 0.56,$
$Pl(\{\theta_1,\theta_2,\theta_3\}) = 1.00,\ Pl(\{\theta_1,\theta_2,\theta_4\}) = 1.00,$
$Pl(\{\theta_1,\theta_3,\theta_4\}) = 1.00,\ Pl(\{\theta_2,\theta_3,\theta_4\}) = 0.95,$
$Pl(\Theta) = 1.00.$

Suppose the probability distribution as the unknown variables. Based on the plausibility functions and the belief functions, the constraints and the objective function can be established according to (11). The probability distribution can be derived based on the minimization. The results of some other probability transformation approaches are also calculated. All the results are listed in Table 2 to make the comparison between the approach proposed in this paper and other available approaches.

N/A in Table 2 means "Not available". DSmP_0 means the parameter ε in DSmP is 0.

Table **2** Probability Transformation Results of Example 2 based on Different Approaches

	θ_1	θ_2	θ_3	θ_4	PIC
PrBel [8]	N/A due to 0 value of singletons				
FPT[11]	N/A due to 0 value of singletons				
PrScP [10]	N/A due to 0 value of singletons				
PrBP1 [12]	N/A due to 0 value of singletons				
PraPl [8]	0.4630	0.2478	0.1561	0.1331	0.0907
BetP [3]	0.4600	0.2550	0.1533	0.1317	0.0910
PrPl [8]	0.6161	0.2160	0.0960	0.0719	0.2471
PrBP2 [12]	0.6255	0.2109	0.0936	0.0700	0.2572
PrHyb [8]	0.6368	0.2047	0.0909	0.0677	0.2698
DSmP_0[9]	0.5162	0.4043	0.0319	0.0477	0.3058
PrBP3 [12]	0.8823	0.0830	0.0233	0.0114	0.5449
Un_min	0.9000	0.0900	0.0000	0.0100	0.7420

Based on the experimental results listed in Table 1 and Table 2, it can be concluded that the probability derived based on the proposed approach (denoted by Un_min) has significantly lower uncertainty when compared with the other probability transformation approaches. The difference among all the propositions can be further enlarged, which is seemingly helpful for the more consolidated and reliable decision.

Important remark: In fact, there exist fatal deficiencies in the probability transformation based uncertainty minimization, which are illustrated in following examples.

4.3 Example 3

The FOD and BPA are as follows [4]:

$\Theta = \{\theta_1,\theta_2\},\ m(\{\theta_1\}) = 0.3,$
$m(\{\theta_2\}) = 0.1,\ ,m(\{\theta_1,\theta_2\}) = 0.6$

Based on different approaches, the experimental results are derived as listed in Table 3

Table **3** Probability Transformation Results of Example 3 based on Different Approaches

	θ_1	θ_2	PIC
BetP	0.6000	0.4000	0.0291
PrPl	0.6375	0.3625	0.0553
PraPl	0.6375	0.3625	0.0553
PrHyb	0.6825	0.3175	0.0984
DSmP_0.001	0.7492	0.2508	0.1875
PrBel	0.7500	0.2500	0.1887
DSmP_0	0.7500	0.2500	0.1887
Un_min	0.9000	0.1000	0.5310

DSmP_0 means the parameter ε in DSmP is 0 and DSmP_0.001 means the parameter ε in DSmP is 0.001.

Is the probability transformation based on PIC maximization (i.e. entropy minimization) rational ?

It can be observed, in our very simple example 3, that all the mass of belief 0.6 committed $\{\theta_1,\theta_2\}$ is actually redistributed only to the singleton $\{\theta_1\}$

using the Un_min transformation in order to get the maximum of PIC.

A deeper analysis shows that with Un_min transformation, the mass of belief $m\{\theta_1, \theta_2\} > 0$ is always fully distributed back to $\{\theta_1\}$ as soon as $m(\{\theta_1\}) > m(\{\theta_2\})$ in order to obtain the maximum of PIC (i.e. the minimum of entropy). Even in very particular situations where the difference between masses of singletons is very small like in the following example:

$\Theta = \{\theta_1, \theta_2\},\ m(\{\theta_1\}) = 0.1000001,$
$m(\{\theta_2\}) = 0.1,\ m(\{\theta_1, \theta_2\}) = 0.7999999.$

This previous modified example shows that the probability obtained from the minimum entropy principle yields a counter-intuitive result, because $m(\{\theta_1\})$ is almost the same as $m(\{\theta_2\})$ and so there is no solid reason to obtain a very high probability for θ_1 and a small probability for θ_2. Therefore, the decision based on the result derived from Un_min transformation is too risky. Sometimes uncertainty can be useful, and sometimes it is better to not take a decision than to take the wrong decision. So the criterion of uncertainty minimization is not sufficient for evaluating the quality/efficiency of a probability transformation. There are also other problems in the probability transformation based on uncertainty minimization principle, which are illustrated in our next example.

4.4 Example 4

The FOD and BPA are as follows: $\Theta = \{\theta_1, \theta_2, \theta_3\}$, with ,

$$m(\{\theta_1, \theta_2\}) = m(\{\theta_2, \theta_3\}) = m(\{\theta_1, \theta_3\}) = 1/3.$$

Using the probability transformation based on uncertainty minimization, we can derive six different probability distributions yielding the same minimal entropy, which are listed as follows:

$$\begin{array}{lll}
P(\{\theta_1\}) = 1/3, & P(\{\theta_2\}) = 2/3, & P(\{\theta_3\}) = 0; \\
P(\{\theta_1\}) = 1/3, & P(\{\theta_2\}) = 0, & P(\{\theta_3\}) = 2/3; \\
P(\{\theta_1\}) = 0, & P(\{\theta_2\}) = 1/3, & P(\{\theta_3\}) = 2/3; \\
P(\{\theta_1\}) = 0, & P(\{\theta_2\}) = 2/3, & P(\{\theta_3\}) = 1/3; \\
P(\{\theta_1\}) = 2/3, & P(\{\theta_2\}) = 1/3, & P(\{\theta_3\}) = 0; \\
P(\{\theta_1\}) = 2/3, & P(\{\theta_2\}) = 0, & P(\{\theta_3\}) = 1/3.
\end{array}$$

It is clear that the problem of finding a probability distribution with minimal entropy does not admit a unique solution in general. So if we use the probability transformation based on uncertainty minimization, there might exist several probability distributions derived as illustrated in this Example 4. How to choose a unique one? In Example 4, depending on the choice of the admissible probability distribution, the decision results derived are totally different which is a serious problem for decision-making support.

From our analysis, it can be concluded that the maximization of PIC criteria (or equivalently the minimization of Shannon entropy) is not sufficient for evaluating the quality of a probability transformation and other criteria have to be found to give more acceptable probability distribution from belief functions. The search for new criteria for developing new transformations is a very open and challenging problem. Until finding new better probability transformation, we suggest to use DSmP as one of the most useful probability transformation. Based on the experimental results shown in Examples 1–3, we see that the DSmP can always be computed and generate a probability distribution with less uncertainty and it is also not too risky, i.e. DSmP can achieve a better tradeoff between a high PIC value (i.e. low uncertainty) and the risk in decision-making.

5 Conclusion

Probability transformation of belief function can be considered as a probabilistic approximation of belief assignment, which aims to gain more reliable decision results. In this paper, we focus on the evaluation criteria of the probability transformation function. Experimental results based on numerical examples show that the maximization of PIC criteria proposed by Sudano is insufficient for evaluating the quality of a probability transformation. More rational criteria have to be found and to better justify the use of a probability transformation with respect to another one.

All the current probability transformations developed so far redistribute the mass of partial ignorances to the belief of singletons included in it. The redistribution is based either only on the cardinality of partial ignorances, or eventually also on a proportionalization using the masses of singletons involved in partial ignorances. However when the mass of a singleton involved in a partial ignorance is zero, some probability transformations, like Cuzzolin's transformation by example, do not work at all and that's why the ε parameter has been introduced in DSmP transformation to make it working in all cases. In future, we plan to develop a more comprehensive and rational criterion, which can take both the risk and the uncertainty degree into consideration, to evaluate the quality of a probability transformation and to find an optimal probability distribution from any basic belief assignment.

Acknowledgment

This work is supported by Grant for State Key Program for Basic Research of China (973) No.2007CB311006 and the Research Fund of Shaanxi Key Laboratory of Electronic Information System Integration (No.200910A).

References

[1] A. P. Dempster, "Upper and lower probabilities induced by a multi-valued mapping", *J. Annual Math Statist*, Vol 38, No. 4, pp. 325–339, 1967.

[2] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, 1976.

[3] P. Smets, R. Kennes, "The transferable belief model", *Artificial Intelligence*, Vol 66, No. 2, pp. 191-234, 1994.

[4] F. Smarandache, J. Dezert (Editors), *Applications and Advances of DSmT for Information Fusion*, Vol 3, American Research Press, 2009. http://www.gallup.unm.edu/~smarandache/DSmT-book3.pdf.

[5] J. Sudano, "Pignistic probability transforms for mixes of low- and high-probability events", *Proc. of International Conference on Information Fusion 2001*, Montreal, Canada, August 2001, TUB3, pp. 23-27.

[6] J. Sudano, "The system probability information content (PIC) relationship to contributing components, combining independent multisource beliefs, hybrid and pedigree pignistic probabilities". *Proc. of International Conference on Information Fusion 2002*, Annapolis, July 2002. Vol 2, pp. 1277-1283.

[7] J. Sudano, "Equivalence between belief theories and naive Bayesian fusion for systems with independent evidential data - Part I, The theory", *Proc. of International Conference on Information Fusion 2003*, Cairns, Australia, July 2003, Vol 2, pp. 1239-1243.

[8] J. Sudano, "Yet another paradigm illustrating evidence fusion (YAPIEF)", *Proc. of International Conference on Information Fusion 2006*, Florence, July 2006, pp. 1-7.

[9] J. Dezert, F. Smarandache, "A new probabilistic transformation of belief mass assignment", *International Conference on Information Fusion 2008*, Germany, Cologne, June 30^{th} – July 3rd 2008, pp. 1-8.

[10] J. Sudano, "Belief fusion, pignistic probabilities, and information content in fusing tracking attributes", *Radar Conference 2004*, April 26th-29th 2004, pp. 218 – 224.

[11] Y. Deng, W. Jiang, Q. Li, "Probability transformation in transferable belief model based on fractal theory (in Chinese)", *Proc. of Chinese Conference on Information Fusion 2009*, Yantai, China, Nov. 2009, pp. 10-13, 79.

[12] W. Pan, H. J. Yang, "New methods of transforming belief functions to pignistic probability functions in evidence theory", *Proc. of International Workshop on Intelligent Systems and Applications*, Wuhan, China, May 2009, pp. 1-5.

[13] F. Cuzzolin, "On the properties of the Intersection probability", submitted to the *Annals of Mathematics and AI*, Feb. 2007.

[14] W. H. Lv, "Decision-making rules based on belief interval with D-S evidence theory, (book section)", *Fuzzy Information and Engineering*, pp. 619-627, 2007.

[15] P. Smets, "Decision making in the TBM: the necessity of the pignistic transformation", *International Journal of Approximate Reasoning*, Vol 38, pp. 133–147, 2005.

[16] R. Horst, P. M. Pardalos, N. V. Thoai, *Introduction to Global Optimization (2nd Edition)*, Kluwer Academic Publishers, New York, 2000.

[17] D. T. Pham, *Intelligent Optimisation Techniques (1st Edition)*, Springer-Verlag, London, 2000.

[18] J. Kennedy, R. C. Eberhart, Y. H. Shi, *Swarm Intelligence*, Morgan Kaufmann Publisher, Academic Press, USA, 2001