Evaluation of Probability Transformations of Belief Functions for Decision Making

Deqiang Han, *Member, IEEE*, Jean Dezert, and Zhansheng Duan, *Member, IEEE*,

***Abstract*—The transformation of belief function into probability is one of the most important and common ways for decision making under the framework of evidence theory. In this paper, we focus on the evaluation of such probability transformations (PTs), which are crucial for their proper applications and the design of new ones. Shannon entropy or probabilistic information content (PIC) measure is traditionally used in evaluating PTs. The transformation having the lowest entropy or highest PIC is considered as the best one. This standpoint is questioned in this paper by comparing a PT based on uncertainty minimization with other available PTs. It shows experimentally that entropy or PIC is not comprehensive to evaluate a PT. To make a comprehensive evaluation, some new approaches are proposed by the joint use of PIC and the distance of evidence according to the value- and rank-based fusion. A pattern classification application oriented evaluation approach for PTs is also proposed. Some desired properties for PTs are also discussed. Experimental results and related analysis are provided to show the rationality of the new evaluation approaches.**

***Index Terms*—Decision making, entropy, evidence theory, probabilistic information content (PIC), probability transformation (PT).**

I. Introduction

DEMPSTER–Shafer theory (DST), also known as the theory of belief functions [1], [2], provides a way to reason with imprecise, uncertain, and incomplete information. DST can distinguish "unknown" and "imprecision" and provides a method to fuse different evidences by using the commutative and associative Dempster's rule of combination. That is why the DST is widely used in information fusion. There are, however, some drawbacks [3], [4] of the DST, e.g., counter-intuitive combination results, high computational cost, and lack of evaluation criteria. So some modified models were proposed, e.g., the transferable belief model (TBM) [3] and Dezert–Smarandache theory (DSmT) [4].

The final goal of uncertainty reasoning is usually decision making. To make decision easier, the mass assignment for a compound focal element is usually assigned to each singleton by a probability transformation (PT). The PT aims to approximate a basic belief assignment (BBA) by a probabilistic measure. The pursuit of efficient PTs has attracted great attention in recent years and many PTs have been proposed [4]–[13].

The most well-known PT is the pignistic PT (PPT) [4] in TBM. PPT maps a belief defined on subsets to a probability measure defined on singletons, based on which a classical decision under probabilistic framework can be readily applied. PPT uses equal weights when splitting mass assignments of the compound focal elements and redistributing them to singletons included in them. Other modified PTs were also proposed [5]–[13], which assign the mass assignments of compound focal elements to singletons according to some ratio constructed from the available information (e.g., the belief and plausibility). Typical examples include Sudano's probabilities [8], Cuzzolin's intersection probability [12], etc. Under the DSmT framework, other PTs called Dezert–Smarandache probability (DSmP) [9] and hierarchical Dezert–Smarandache probability (HDSmP) [13] were proposed. DSmP takes into account both masses and cardinality of focal elements in the proportional redistribution process and HDSmP is a hierarchical version of DSmP. They can also be used in the DST framework.

To compare all the available transformations for the purpose of appropriate application and design of new transformations, evaluation is required. In almost all the existing works on PTs, Shannon entropy or its dual, probabilistic information content (PIC) criterion, is used to evaluate them. Definitely, less uncertainty should be preferred for decision making. However, is the probability measure generated from a belief function with less uncertainty always rational or beneficial for decision making? To answer this, i.e., to illustrate the irrationality of the over-emphasize of PIC or entropy, another PT based on a constrained entropy minimization [14] is used and analyzed through examples. When using entropy or PIC for evaluation, the probability measure with the least uncertainty seems the best one. Unfortunately, some risky and unexpected results may also be obtained. Han *et al.* [14] show that either entropy or PIC is not a comprehensive measure.

Manuscript received November 27, 2014; accepted February 7, 2015. Date of publication April 28, 2015; date of current version December 14, 2015. This work was supported in part by the State Key Program for Basic Research of China (973) under Grant 2013CB329405, in part by the National Natural Science Foundation under Grant 61104214, Grant 61203222, and Grant 61174138, in part by the Foundation for Innovative Research Groups of the National Natural Science Foundation of China under Grant 61221063, in part by the Science and Technology Project of Shaanxi Province under Grant 2013KJXX-46, in part by the Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20120201120036, and in part by the Fundamental Research Funds for the Central Universities under Grant xjj2012104 and Grant xjj2014122. This paper was recommended by Associate Editor L. Cao.

D. Han is with the Ministry of Education Key Laboratory for Intelligent Networks and Network Security, Institute of Integrated Automation, School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: deqhan@gmail.com).

J. Dezert is with ONERA, The French Aerospace Laboratory, Palaiseau F-91761, France.

Z. Duan is with the Center for Information Engineering Science Research, Xi'an Jiaotong University, Xi'an 710049, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMC.2015.2421883

Comprehensive evaluation of PT is desired and motivates this paper. PIC only emphasizes the clarity of the transformed probability, which is only from the aspect of the clarity of decision making. On the other hand, the transformed probability should be consistent with the original belief function in some sense meaning that comprehensive evaluation should also consider the fidelity of the transformed probability to the original BBA. Higher degree of fidelity means the less loss of information caused by PT. Our comprehensive evaluation aims to make a balance between clarity and fidelity, i.e., a probability with higher clarity and bigger fidelity should be preferred. Then, how to quantify the degree of fidelity? The distance of evidence [15] is used to measure the dissimilarity between two BBAs. Since the probability can be considered as a particular BBA, we can simply use the distance of evidence [15] between the original BBA and transformed probability to quantify the degree of fidelity (smaller distance means higher degree of fidelity). So, in this paper, we evaluate the PTs jointly by PIC and the distance of evidence. This joint use of the two criteria is implemented by using value-based fusion (via the values of PIC and distance) and rank-based fusion (via the ranks of the values of PIC and distance). We also propose an application-oriented evaluation approach for PTs, such as the application of pattern classification. Besides the evaluation criteria, some desired properties (qualitative evaluations) of PTs are also helpful. In [16], some desired properties of PTs were proposed and analyzed including upper and lower bound consistency and combination consistency. In this paper, some new desired properties of a PT are also proposed. This paper extends our previous ideas briefly introduced in [14], where we preliminarily pointed out that entropy or PIC is not enough to evaluate a PT. However, in [14], comprehensive evaluation was not proposed, which are the main contribution of this paper.

The rest of this paper is organized as follows. In Section II, evidence theory is briefly introduced. The decision-making methods in evidence theory including belief based approaches and PTs are briefly summarized in Section III. The definitions and pertinent analysis of the commonly used PTs are given in Section IV. In Section V, the evaluation of the PT is discussed. The irrationality of using entropy alone as an evaluation criterion is clearly shown by simple examples. In Section VI, we propose to evaluate a PT based on two criteria (PIC and distance). The joint use of them is implemented either directly at their values, or at their ranks. Some supporting examples are provided in Section VII. In Section VIII, an application-oriented evaluation approach is proposed. In Section IX, some desired properties of PTs are proposed and analyzed. The conclusion is drawn in Section X.

II. Basics of Evidence Theory

In DST [2], the elements in the frame of discernment (FOD) Θ, which is a discrete finite set, are mutually exclusive and exhaustive. Let 2^Θ be the power set of the FOD. The function $m : 2^\Theta \to [0, 1]$ defines a BBA, also called a mass function, which satisfies

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

Then, the belief and plausibility functions are defined as in (2) and (3), respectively, $\forall A \in 2^\Theta$

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

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

where $\mathrm{Bel}(A)$ and $\mathrm{Pl}(A)$ can be interpreted as the lower and upper bounds of the probability $P(A)$.

Dempster's rule of combination, which is used to fuse n distinct[1] bodies of evidence (BOEs), is

$$m(A) = \begin{cases} 0, & \forall A = \emptyset \\ \dfrac{\sum_{\cap A_i = A} \prod_{1\le i\le n} m_i(A_i)}{\sum_{\cap A_i \neq \emptyset} \prod_{1\le i\le n} m_i(A_i)}, & \forall A \neq \emptyset \end{cases} \tag{4}$$

where $m_1, m_2, \ldots, m_n$ are n BBAs.

Distances of evidence [15], [17] measures the dissimilarity between BOEs. One of the most commonly used distances of evidence is Jousselme's distance $d_J(\cdot,\cdot)$ [15]

$$d_J(m_1, m_2) = \sqrt{\frac{1}{2}(m_1 - m_2)^T \mathbf{Jac}\ (m_1 - m_2)} \tag{5}$$

where the element $J_{ij} = \mathbf{Jac}(A_i, B_j)$ of Jaccard's weighting matrix $\mathbf{Jac}$ is defined as

$$\mathbf{Jac}(A_i, B_j) = \frac{|A_i \cap B_j|}{|A_i \cup B_j|}. \tag{6}$$

For example, two BBAs are defined over the FOD $\Theta = \{\theta_1, \theta_2\}$

$$m_1(A_1) = 0.2, \quad m_1(A_2) = 0.8$$
$$m_2(B_1) = 0.5, \quad m_2(B_2) = 0.5$$

where

$$A_1 = \{\theta_1\}, \quad A_2 = \{\theta_1, \theta_2\}$$
$$B_1 = \{\theta_2\}, \quad B_2 = \{\theta_1, \theta_2\}.$$

We have

$$\mathbf{Jac}(A_1, B_1) = |\emptyset|/|\{\theta_1, \theta_2\}| = 0$$
$$\mathbf{Jac}(A_2, B_1) = |\{\theta_2\}|/|\{\theta_1, \theta_2\}| = 0.5$$
$$\mathbf{Jac}(A_1, B_2) = |\{\theta_1\}|/|\{\theta_1, \theta_2\}| = 0.5$$
$$\mathbf{Jac}(A_2, B_2) = |\{\theta_1, \theta_2\}|/|\{\theta_1, \theta_2\}| = 1.$$

Although there are other distance definitions for belief functions, they either have some limitations or are not strict distance metrics [18]. Jousselme's distance has been proved to be a strict distance metric [19].

The aim of the evidential reasoning is for decision making. Several decision-making approaches in evidence theory are briefly reviewed next.

III. Decision Making in Evidence Theory

There are two major types of decision-making approaches under the evidence theory framework: directly using belief functions [20], [21] and using PTs of belief functions [22].

[1] That is, cognitively independent.

A. Decision Making Using Belief Functions

There exist three main decision-making rules using Bel and Pl.

1) Max Bel: One chooses the proposition A with the maximum $\mathrm{Bel}(A)$. $\mathrm{Bel}(\cdot)$ describes the lowest trust degree of a given proposition. So it is also called pessimistic decision making in DST [20].

2) Max Pl: One chooses the proposition A with the maximum $\mathrm{Pl}(A)$. $\mathrm{Pl}(\cdot)$ describes the highest trust degree of a given proposition. So it is also called optimistic decision making in DST [20].

3) Joint Use of Bel and Pl: Bel and Pl measure the degree of trust of a given proposition from two points of view. So it is not comprehensive to make a decision based on only one of them. An extension is the "final belief" defined below [21]

$$\mathrm{FB}(A) = \mathrm{Bel}(A) + \alpha(\mathrm{Pl}(A) - \mathrm{Bel}(A)) \tag{7}$$

where $\alpha = \mathrm{Bel}(A)/(\mathrm{Bel}(A) + \mathrm{Bel}(\bar{A}))$. The proposition A with the maximum $\mathrm{FB}(A)$ is preferred.

Note that the proposition A can be either a singleton or compound proposition (containing more than one singleton).

B. Decision Making Using Probability Transformations

Probability-based decision rules are the main stream of decision-making based on evidence theory [21], because the two-level reasoning and decision structure (i.e., the credal and pignistic levels) proposed by Smets in his TBM is quite appealing. In this type of decision-making approach, the belief function (or BBA, plausibility function) is transformed into a probability measure P first and then the decision can be made as $\theta_i^* = \arg\max_{\theta_i} P(\theta_i)$, where θ_i is a singleton of the FOD. As we will see next, the PT is crucial for this type of decision making.

IV. Probability Transformations

A PT is a mapping PT : $\mathrm{Bel}_\Theta \to \mathrm{Pr}_\Theta$. Bel_Θ is a belief function defined on Θ and Pr_Θ is a probability measure [in fact a probability mass function (PMF)] defined on Θ. Major PTs are summarized below.

1) Pignistic Transformation: The classical pignistic probability was proposed in TBM framework [3], which is a subjective and nonprobabilistic interpretation of evidence theory. At the credal level of TBM, beliefs are entertained [3], combined, and updated. While at the pignistic level, decisions are made by applying the PPT.

Suppose that FOD is $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$ in the sequel. The PPT [3] for singletons is defined as

$$\mathrm{BetP}(\theta_i) = \sum_{\theta_i \in B,\ B \in 2^\Theta} \frac{m(B)}{|B|}. \tag{8}$$

PPT is designed according to an idea similar to uncertainty maximization [14]. In PPT, masses are assigned uniformly to different singletons involved.

2) Sudano's Probabilities: Sudano [5]–[7] proposed PT proportional to plausibilities (PrPl) [5], PT proportional to beliefs (PrBel) [5], PT proportional to all plausibilities (PraPl) [5], hybrid PT (PrHyb) [6], and an iterative version of PT (PrScP) [5].

They are defined by different types of mappings as follows:

$$\mathrm{PrPl}(\theta_i) = \mathrm{Pl}(\{\theta_i\}) \cdot \sum_{Y \in 2^\Theta, \theta_i \in Y} \frac{m(Y)}{\sum_{\cup_j \theta_j = Y} \mathrm{Pl}(\{\theta_j\})} \tag{9}$$

$$\mathrm{PrBel}(\theta_i) = \mathrm{Bel}(\{\theta_i\}) \cdot \sum_{Y \in 2^\Theta, \theta_i \in Y} \frac{m(Y)}{\sum_{\cup_j \theta_j = Y} \mathrm{Bel}(\{\theta_j\})} \tag{10}$$

$$\mathrm{PraPl}(\theta_i) = \mathrm{Bel}(\{\theta_i\}) + \frac{1 - \sum_j \mathrm{Bel}(\{\theta_j\})}{\sum_j \mathrm{Pl}(\{\theta_j\})} \cdot \mathrm{Pl}(\{\theta_i\}) \tag{11}$$

$$\mathrm{PrHyb}(\theta_i) = \mathrm{PraPl}(\theta_i) \cdot \sum_{Y \in 2^\Theta, \theta_i \in Y} \frac{m(Y)}{\sum_{\cup_j \theta_j = Y} \mathrm{PraPl}(\theta_j)} \tag{12}$$

$$\mathrm{PrScP}(\theta_i) = \sum\nolimits_{\theta_i \in Y} \left(\frac{\mathrm{PrScP}(\theta_i)}{\sum_j \mathrm{PrScP}(\theta_j)} \right) \cdot m(Y). \tag{13}$$

Note that the iterative PrScP should be initiated by some other transformation.

3) Cobb–Shenoy's Normalization of Plausibility: This PT is defined as the normalized plausibility function of singletons [8]

$$\mathrm{PnPl}(\theta_i) = \frac{\mathrm{Pl}(\{\theta_i\})}{\sum_j \mathrm{Pl}(\{\theta_j\})}. \tag{14}$$

4) Cuzzolin's Intersection Probability: From a geometric interpretation of evidence theory, an intersection probability measure [12] was proposed using the proportional repartition of the total nonspecific mass (total non-specific mass (TNSM) $= \sum_{A \in 2^\Theta, |A|>1} m(A)$) for each contribution of the nonspecific masses involved

$$\mathrm{CuzzP}(\theta_i) = m(\{\theta_i\}) + \frac{\mathrm{Pl}(\{\theta_i\}) - m(\{\theta_i\})}{\sum_j (\mathrm{Pl}(\{\theta_j\}) - m(\{\theta_j\}))} \cdot \mathrm{TNSM}. \tag{15}$$

5) DSmP: The $\mathrm{DSmP}_\varepsilon(\theta_i)$ [9] can be directly obtained by

$$\mathrm{DSmP}_\varepsilon(\theta_i) = m(\{\theta_i\}) + (m(\{\theta_i\}) + \varepsilon) \times \left(\sum_{\substack{X \in 2^\Theta \\ \theta_i \subset X, |X| \geq 2}} \frac{m(X)}{\sum_{\substack{Y \in 2^\Theta \\ Y \subset X, |Y|=1}} m(Y) + \varepsilon \cdot |X|} \right). \tag{16}$$

In DSmP, both mass assignments and cardinality of focal elements are used in the proportional redistribution. DSmP makes an improvement compared with Sudano's, Cuzzolin's, and BetP, in that DSmP makes a more judicious redistribution of the ignorance masses to the singletons involved. DSmP works for both DST and DSmT.

6) HDSmP: It is a hierarchical version of DSmP (see [13] for details). When the mass for the focal elements with the same cardinality are zero, HDSmP_0 cannot be computed due to its hierarchical nature. Therefore, the parameter ϵ is necessary to improve the applicability of HDSmP [13].

7) PrBP1: The proportional transformation hypothesis used in PrBP1 [11] assumes that the masses are distributed proportionally to the product of $\mathrm{Bel}(\theta_i)$ and $\mathrm{Pl}(\theta_i)$ among each singleton element of $\theta_i \in Y$ with $Y \subseteq \Theta$

$$\mathrm{PrBP1}(\theta_i) = \sum_{Y,\theta_i \in Y} \left(\frac{\mathrm{Bel}(\theta_i)\mathrm{Pl}(\theta_i)}{\sum_{j,\theta_j \in Y} \mathrm{Bel}(\theta_j)\mathrm{Pl}(\theta_j)} \right) \cdot m(Y). \quad (17)$$

8) PrBP2: It [11] assumes that the masses are distributed proportionally to some given parameters $s_i = \mathrm{Bel}(\theta_i)/1 - \mathrm{Pl}(\theta_i)$ or $s_i = \mathrm{Pl}(\theta_i)/1 - \mathrm{Bel}(\theta_i)$

$$\mathrm{PrBP2}(\theta_i) = \sum_{Y,\theta_i \in Y} \left(\frac{s_i}{\sum_{j,\theta_j \in Y} s_j} \right) \cdot m(Y). \quad (18)$$

A PT outputs a Bayesian BBA (having only singleton focal elements) corresponding to a given (non-Bayesian) BBA. That is why the PT is also called the Bayesian transformation. A Bayesian BBA is not a probability measure, but if $m(.)$ is a Bayesian BBA, then its corresponding Bel(.) and Pl(.) coincide with a probability measure, i.e., $\mathrm{Bel}(.) = \mathrm{Pl}(.) = P(.)$. Due to the tradition, it is still called the "PT" in this paper.

V. Questioning of Traditional Evaluation of Probability Transformation

The evaluation of different PTs is important for analysis and their applications. It is also important for the design of new transformations. In this section, we will provide some comments on traditional evaluation approaches for PTs.

A. Traditional Evaluation Approaches for Probability Transformation

Qualitative evaluation approaches were proposed. In [13], three desired properties of a PT are introduced.

1) p Consistency: A PT is p-consistent (probability consistent) if $\mathrm{PT}(m) = m$ for any Bayesian BBA m.

2) ULB Consistency: A PT is upper and lower bounds (ULB)-consistent (upper and lower bounds consistent) if its resulting transformed probability $P = \mathrm{PT}(m)$ satisfies $\mathrm{Bel}(X) \le P(X) \le \mathrm{Pl}(X)$.

3) Combination Consistency: It means that we will obtain the same result either, if we combine two BBAs m_1 and m_2 using the combination rule first and perform PT, thereafter, or perform PT to both input BBAs m_1 and m_2 first and combine them thereafter. It is defined through commutation property of combination rule and PT. It is difficult to be satisfied, and PnPl [8] is the only one known to the authors that can satisfy it when using Dempster's rule of combination.

There also exist some quantitative metrics measuring the strength of a critical decision based on a probability measure.

a) Normalized Shannon entropy: Suppose that $P(\theta)$ is a PMF, where $\theta \in \Theta$, $|\Theta| = N$. An evaluation criterion for the PMF transformed is as follows [11]:

$$E_H = \frac{-\sum_{\theta \in \Theta} P(\theta)\log_2(P(\theta))}{\log_2 N} \quad (19)$$

i.e., the ratio of Shannon entropy [23] and the maximum Shannon entropy. Clearly E_H is normalized. The larger (smaller) E_H gets, the larger (smaller) the degree of uncertainty gets. When $E_H = 0$, one singleton proposition will have probability 1 and the others will have zero probabilities. Therefore, the agent or system can make decision without error if the probability $P(\cdot)$ corresponds to the real probability of the events. When $E_H = 1$, it is unlikely to make a correct decision, because $P(\theta)$ are equal, for all $\theta \in \Theta$, i.e., one has a uniform PMF.

b) Probabilistic information content: The PIC criterion [5] is the dual of the normalized Shannon entropy. The PIC value of a PMF obtained from a PT indicates the total knowledge to make a correct decision

$$\mathrm{PIC}(P) = 1 + \frac{1}{\log_2 N} \cdot \sum_{\theta \in \Theta} P(\theta)\log_2(P(\theta)). \quad (20)$$

Obviously, $\mathrm{PIC} = 1 - E_H$. A PIC value of zero indicates that the knowledge to make a correct decision is not informative enough (all propositions have equal probabilities, i.e., one has the maximal entropy).

Less uncertainty means that the corresponding PT result is more helpful in making a decision. According to such an idea, the PT should attempt to enlarge the belief differences among all the propositions and thus to achieve a clearer decision result. Is this rational? Is uncertainty degree always judicious at all to evaluate a PT for decision-making purpose? If this is true, a PT approach based on direct uncertainty minimization should be the best choice. Is that true? In the next section, we examine the legitimacy of using uncertainty degree as a criterion to evaluate a PT.

B. Probability Transformation Based on Uncertainty Minimization

As mentioned above, the "best" PT can be obtained by directly minimizing E_H (or equivalently maximizing PIC) as follows:

$$\min_{\{P(\theta)|\theta\in\Theta\}} \left\{ -\sum_{\theta \in \Theta} P(\theta)\log_2(P(\theta)) \right\}$$
$$\text{s.t.} \begin{cases} \mathrm{Bel}(B) \le \sum_{\theta \in B} P(\theta) \le \mathrm{Pl}(B) \\ 0 \le P(\theta) \le 1 \forall \theta \in \Theta \\ \sum_{\theta \in \Theta} P(\theta) = 1 \end{cases} \quad (21)$$

where the objective function is the Shannon entropy and the constraints are the ULB consistency and the property of probability. Given a belief function, the solution of (21) is guaranteed to have the least uncertainty and is thus seemingly more preferable in decision making. This so called the "best" transformation is denoted by $\mathrm{Un}_{\min}$.

Clearly, the problem of finding a minimum-entropy PMF does not have a unique solution in general. We use the quasi-Newton method followed by a global optimization algorithm [24] to solve (21) to alleviate the effect of the local extremum problem. Other optimization algorithms [25], [26] can also be used, e.g., genetic algorithm and particle swarm optimization.

TABLE I
PT RESULTS FOR EXAMPLE 1

	θ_1	θ_2	θ_3	θ_4	PIC
PnPl	0.3614	0.3168	0.1931	0.1287	0.0526
CuzzP	0.3860	0.3382	0.1607	0.1151	0.0790
BetP	0.3983	0.3433	0.1533	0.1050	0.0926
PraPl	0.4021	0.3523	0.1394	0.1062	0.1007
PrPl	0.4544	0.3609	0.1176	0.0671	0.1638
PrHyb	0.4749	0.3749	0.0904	0.0598	0.2014
PrBel	0.5176	0.4051	0.0303	0.0470	0.3100
DSmP_0	0.5176	0.4051	0.0303	0.0470	0.3100
$\text{DSmP}_{0.001}$	0.5162	0.4043	0.0319	0.0477	0.3058
HDSmP_0	0.5293	0.3960	0.0310	0.0437	0.3161
$\text{HDSmP}_{0.001}$	0.5258	0.3943	0.0344	0.0455	0.3064
PrScP	0.5420	0.3870	0.0324	0.0386	0.3247
PrBP1	0.5419	0.3998	0.0243	0.0340	0.3480
PrBP2	0.5578	0.3842	0.0226	0.0353	0.3529
Un_{min}	0.7300	0.2300	0.0100	0.0300	0.4813

TABLE II
PT RESULTS FOR EXAMPLE 2

	θ_1	θ_2	θ_3	θ_4	PIC
PrBel	N/A due to 0 value of singletons				
PrScP	N/A due to 0 value of singletons				
PrBP1	N/A due to 0 value of singletons				
DSmP_0	N/A due to 0 value of singletons				
HDSmP_0	N/A due to 0 value of singletons				
PnPl	0.4348	0.2609	0.1643	0.1401	0.0733
CuzzP	0.4498	0.2540	0.1599	0.1364	0.0822
PraPl	0.4630	0.2478	0.1561	0.1331	0.0907
BetP	0.4600	0.2550	0.1533	0.1317	0.0910
PrPl	0.6161	0.2160	0.0960	0.0719	0.2471
PrBP2	0.6255	0.2109	0.0936	0.0700	0.2572
PrHyb	0.6368	0.2047	0.0909	0.0677	0.2698
$\text{DSmP}_{0.001}$	0.8820	0.0486	0.0400	0.0294	0.6464
$\text{HDSmP}_{0.001}$	0.8646	0.0604	0.0454	0.0296	0.6106
Un_{min}	0.9000	0.0900	0.0000	0.0100	0.7420

C. Analysis of Probability Transformation Based on Uncertainty Minimization

To compare different PTs, the following two examples drawn from [6] and [11] are considered, where PIC is used for evaluation.

Example 1: For FOD $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$, the corresponding BBA is as follows:

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

Based on the PTs defined in (8)–(18) and (21), respectively, the BBA can be transformed into different probabilities as illustrated in Table I. Their corresponding PICs can be calculated using (20), which are also listed in Table I. The Un_{min} provides the maximum PIC as expected.

Example 2: For FOD $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$, the corresponding BBA is as follows:

$$\begin{aligned} m(\{\theta_1\}) &= 0.05, m(\{\theta_2\}) = 0.00, m(\{\theta_3\}) = 0.00 \\ m(\{\theta_4\}) &= 0.00, m(\{\theta_1, \theta_2\}) = 0.39 \\ m(\{\theta_1, \theta_3\}) &= 0.19, m(\{\theta_1, \theta_4\}) = 0.18 \\ m(\{\theta_2, \theta_3\}) &= 0.04, m(\{\theta_2, \theta_4\}) = 0.02 \\ m(\{\theta_3, \theta_4\}) &= 0.01, m(\{\theta_1, \theta_2, \theta_3\}) = 0.04 \\ m(\{\theta_1, \theta_2, \theta_4\}) &= 0.02, m(\{\theta_1, \theta_3, \theta_4\}) = 0.03 \\ m(\{\theta_2, \theta_3, \theta_4\}) &= 0.03, m(\Theta) = 0.00. \end{aligned}$$

By using the PTs defined in (8)–(18) and (21), respectively, we can transform the BBA into different probabilities as illustrated in Table II. Their corresponding PICs can be calculated using (20), which are also listed in Table II. In this example, the masses for some singletons are zero, so some PTs cannot be applied as shown in Table II, where N/A means "not available." The notation DSmP_0, $\text{DSmP}_{0.001}$, HDSmP_0, and $\text{HDSmP}_{0.001}$ mean that the values of the parameter ε in DSmP and HDSmP transformations are chosen to 0 and 0.001.

From the experimental results in Tables I and II, the PMF obtained from the proposed Un_{min} approach has the least uncertainty (and thus the greatest PIC) when compared with the others. That is, the difference among all propositions of the existing PT approaches can be further enlarged, which is seemingly helpful for more consolidated and clearer decision making.

Remark: However, there exist serious deficiencies associated with Un_{min}, as shown in the following example.

Example 3: The BBA defined on FOD $\Theta = \{\theta_1, \theta_2\}$ is $m(\{\theta_1\}) = 0.3,\ m(\{\theta_2\}) = 0.1,\ m(\{\theta_1, \theta_2\}) = 0.6.$

The experimental results of different approaches are listed in Table III.

Is the PT based on PIC maximization (i.e., entropy minimization) rational?

In this simple example, all of the mass 0.6 committed to $\{\theta_1, \theta_2\}$ is only redistributed to the singleton $\{\theta_1\}$ when the Un_{min} transformation is used in order to achieve the maximum PIC.

It can also be shown that for the Un_{min}, the mass assignments $m(\{\theta_1, \theta_2\}) > 0$ is always completely redistributed to $\{\theta_1\}$ as long as $m(\{\theta_1\}) > m(\{\theta_2\})$ in order to achieve the maximum PIC.

This is also true in the situations where the difference between masses of singletons is very small as demonstrated by the following BBA defined on FOD $\Theta = \{\theta_1, \theta_2\}$:

$$\begin{aligned} m(\{\theta_1\}) &= 0.1000001, m(\{\theta_2\}) = 0.1 \\ m(\{\theta_1, \theta_2\}) &= 0.7999999. \end{aligned}$$

Such a case shows that $m(\{\theta_1\})$ is almost the same as $m(\{\theta_2\})$ and there is no specific reason to obtain a very high

TABLE III
PT RESULTS FOR EXAMPLE 3

	θ_1	θ_2	PIC
PnPl	0.5625	0.4375	0.0113
CuzzP	0.6000	0.4000	0.0291
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	0.7500	0.2500	0.1887
$\text{DSmP}_{0.001}$	0.7493	0.2507	0.1875
HDSmP_0	0.7500	0.2500	0.1887
$\text{HDSmP}_{0.001}$	0.7485	0.2515	0.1864
PrBel	0.7500	0.2500	0.1887
PrScP	0.7500	0.2500	0.1887
PrBP1	0.7765	0.2235	0.2334
PrBP2	0.8400	0.1600	0.3657
Un_{min}	0.9000	0.1000	0.5310

probability for θ_1 and a small one for θ_2. Therefore, the decision based on the result from Un_{min} transformation appears to be very risky or dogmatic. In some applications, a decision has to be made and we cannot avoid to make one (good or bad). However, when the time to make a decision is not too limited or rejection decision is permitted, it is better to collect more observations (information) or to make a rejection rather than to take high risk to make an erroneous decision. So the criterion of uncertainty minimization, which can bring such risky results, is not always judicious to evaluate a PT for decision-making purpose. Furthermore, when we use Un_{min}, there are also some other problems. See the next example for details.

Example 4: The BBA defined on the FOD $\Theta = \{\theta_1, \theta_2, \theta_3\}$ is

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

Using Un_{min}, we can obtain six different PMFs yielding the same minimal entropy, which are listed as follows:

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

It is clear that the problem of finding a PMF with the minimal entropy does not have a unique solution in general. Then how to choose a unique one? In this example, different admissible PMF yields different decision result. This is a serious problem for decision making.

From all the above examples, we conclude that the maximization of PIC (or minimization of Shannon entropy) is not satisfactory for evaluation. Therefore, more comprehensive evaluation methods are needed, which is an open and challenging problem.

VI. New Bi-Criteria Solution for Probability Transformation Evaluation

To design a single criterion for the comprehensive evaluation of PTs is always difficult. Jointly using multiple criteria is one option meaning that besides entropy or PIC, another measure, which describes some other aspects of a PT, may be incorporated for evaluation.

The level of PIC characterizes the clarity of a given PMF. Indeed, higher PIC (lower entropy) means that the PMF tends to concentrate on a specific hypothesis of the FOD, which makes the decision easier for the decider. Also, the PMF is transformed from a given BBA representing the original information source. If the obtained PMF is in some sense closer to the original BBA, it will be preferred. This is because such a PMF has high degree of fidelity to the original BBA (i.e., with less loss of information in the transformation). The clarity and fidelity can always be balanced. So by adding one more criterion representing the degree of fidelity to the evaluation measure, it will be more comprehensive. How to characterize the closeness between the obtained PMF and original BBA? The answer is to use the distance (or dissimilarity) of evidence [15]. In summary, we attempt to propose bi-criteria evaluation approaches by jointly using the distance of evidence and PIC.

By treating the PMF obtained from some PT as a special BBA (not strict), a distance of evidence can be used to describe the dissimilarity between the PMF and original BBA. We use the distance of evidence together with PIC (or entropy) as the elements of a two-tuple

$$\langle \text{PIC}(P), d_J(P, m)\rangle \quad \text{or} \quad \langle \text{Entropy}(P), d_J(P, m)\rangle \tag{22}$$

where P is the transformed probability (i.e., PMF) and m is the original BBA.

A two-tuple can provide more comprehensive information; however, how to jointly use them to evaluate a PT? Larger PIC represents greater clarity and smaller d_J represents greater fidelity. Over-emphasizing on any single criterion is not preferred. As aforementioned, if we want to choose a better PT approach, there should be a tradeoff between the two criteria. Here, we propose two approaches to jointly use the two criteria. The first one is directly based on the arithmetic operations over PIC and d_J. The second one first sorts PTs to be evaluated according to PIC and d_J, respectively, to obtain two corresponding ranks. And then, the two ranks are fused through some rank fusion rule to obtain a comprehensive rank for evaluation.

A. Comprehensive Evaluation With Value-Based Fusion

Suppose that there are N PMFs denoted by $P_1, P_2, \ldots, P_N$, which are obtained from N different PTs. Their entropies are $En_1, En_2, \ldots, En_N$ and their d_J's (between each PMF and the original BBA) are $d_1, d_2, \ldots, d_N$. To jointly use these two criteria, we calculate the comprehensive scores for different PTs as follows:

$$C_{\text{joint}}(P_i) = \alpha \cdot \text{Ent}'(i) + (1 - \alpha) \cdot d'(i) \tag{23}$$

where $i = 1, \ldots, N$ and α denotes a weight representing the degree of preference on entropy. Distances (and entropies) usually take different values depending on the PT. Therefore, to be consistent, we need to first normalize them by their ranges as

$$\begin{cases} \mathrm{Ent}'(i) = \dfrac{\mathrm{Ent}(i) - \min(\mathrm{Ent})}{\max(\mathrm{Ent}) - \min(\mathrm{Ent})} \\ d'(i) = \dfrac{d(i) - \min(d)}{\max(d) - \min(d)} \end{cases} \tag{24}$$

where the vector $d = [d(1), d(2), \ldots, d(N)]$ and the vector $\mathrm{Ent} = [\mathrm{Ent}(1), \mathrm{Ent}(2), \ldots, \mathrm{Ent}(N)]$ are the vectors of distances and entropies corresponding to the PTs $P_1, P_2, \ldots, P_N$.

Smaller entropy (larger PIC, i.e., bigger clarity) and smaller distance (bigger fidelity) are desired. Then, by sorting the values of $C_{\text{joint}}(P_i)$ in ascending order, we can obtain the rank as

$$\Lambda_C = (r_C(P_1), r_C(P_2), \ldots, r_C(P_N)). \tag{25}$$

The PTs with the best rank, i.e., those having the smallest rank value, are preferred.

B. Comprehensive Evaluation With Rank-Based Fusion

Let us consider N PMFs denoted by $P_1, P_2, \ldots, P_N$, which are obtained from N different PTs. A comprehensive rank can then be obtained from the rank-based fusion (rank fusion for short) [27], [28] implemented in the following steps.

Step 1: Obtain the PIC-based rank. Sort the PMFs in descending order according to their PICs (this is because higher PIC value is desired). Then the rank of all the PMFs is

$$\Lambda_{\text{PIC}} = (r_{\text{PIC}}(P_1), r_{\text{PIC}}(P_2), \ldots, r_{\text{PIC}}(P_N)). \tag{26}$$

Step 2: Obtain the distance-based rank. Sort the PMFs in ascending order according to the Jousselme distances (d_J) (this is because smaller d_J is desired), then the rank for all the PMFs can be obtained as

$$\Lambda_d = (r_d(P_1), r_d(P_2), \ldots, r_d(P_N)). \tag{27}$$

Step 3: Obtain the global rank by a rank fusion. To find the joint (or comprehensive) rank of Λ_{PIC} and Λ_d, a rank fusion is applied as

$$\Lambda_f = f(\Lambda_{\text{PIC}}, \Lambda_d) \tag{28}$$

where f is a rank fusion rule and Λ_f is

$$\Lambda_f = \big(r_j(P_1), r_j(P_2), \ldots, r_j(P_N)\big). \tag{29}$$

The PTs with the best rank, i.e., those having the smallest rank, are preferred.

The selection of rank fusion rule is crucial, which is discussed below.

1) Min Rule:

$$r_j(P_k) = \min(r_{\text{PIC}}(P_k), r_d(P_k)) \forall k = 1, \ldots, N. \tag{30}$$

2) Max Rule:

$$r_j(P_k) = \max(r_{\text{PIC}}(P_k), r_d(P_k)) \forall k = 1, \ldots, N. \tag{31}$$

3) Arithmetic Averaging Rule:

$$r_j(P_k) = w_1 \cdot r_{\text{PIC}}(P_k) + w_2 \cdot r_d(P_k) \forall k = 1, \ldots, N \tag{32}$$

where w_1 and w_2 are weights for the two different ranks to be fused.

4) Optimization Rule: The optimization-based rank fusion rule is

$$\Lambda^* = \arg\min_{\Lambda} \frac{1}{L} \sum_{j=1}^{L} d_r\big(\Lambda, \Lambda_j\big) \tag{33}$$

where $\Lambda_1, \ldots, \Lambda_L$ are the L different ranks to fuse and $d_r(\cdot, \cdot)$ is a distance between two ranks that will be presented in the sequel.

Here, we have two ranks Λ_{PIC} and Λ_d. The above equation could be rewritten as

$$\Lambda_f = \arg\min_{\Lambda} \frac{1}{2}[d_r(\Lambda, \Lambda_{\text{PIC}}) + d_r(\Lambda, \Lambda_d)] \tag{34}$$

$d_r(\cdot, \cdot)$ could be any rank distance including the footrule distance [29], Kendall distance [30], [31], and Spearman distance [32] as introduced below.

Suppose that Λ_1 and Λ_2 are two ranks. Let $X = \{x_1, x_2, \ldots, x_N\}$ be a set of items to be ranked. $\Lambda_j(i)$ is the rank associated with the item x_i, where $j = 1, 2$ and $i = 1, 2, \ldots, N$.

5) Footrule Distance:

$$F(\Lambda_1, \Lambda_2) = \sum_{i=1}^{N} |\Lambda_1(i) - \Lambda_2(i)|. \tag{35}$$

6) Kendall Distance:

$$K(\Lambda_1, \Lambda_2) = \sum_{\{i,j\} \in P} K^*_{i,j}(\Lambda_1, \Lambda_2) \tag{36}$$

where P is the set of the unordered pairs of distinct items in X and

$$K^*_{i,j}(\Lambda_1, \Lambda_2) = \begin{cases} 0, & \text{if } x_i \text{ and } x_j \text{ are in the same order in } \Lambda_1 \text{ and } \Lambda_2 \\ 1, & \text{if } x_i \text{ and } x_j \text{ are in the reverse order in } \Lambda_1 \text{ and } \Lambda_2. \end{cases}$$

7) Spearman Distance:

$$\rho(\Lambda_1, \Lambda_2) = 1 - \frac{6 \cdot \sum_{i=1}^{N} (\Lambda_1(i) - \Lambda_2(i))^2}{N(N^2 - 1)}. \tag{37}$$

Clearly, $\rho \in [-1, 1]$. $\rho = 1$ means a total positive correlation between the ranks and $\rho = -1$ means a total negative one.

Here, Λ_1 and Λ_2 could be Λ_{PIC} and Λ_d, respectively.

C. Illustrative Example of Rank Fusion

Suppose that $X = \{x_1, x_2, x_3, x_4\}$ corresponds to a set of possible choices in a decision-making problem. $\Lambda_1 = [1, 2, 3, 4]$ and $\Lambda_2 = [1, 3, 4, 2]$ are two ranks provided by two experts for X.

When using different rank fusion rules, the results (denoted by Λ_f) are as follows.

TABLE IV
EVALUATIONS OF PT RESULTS OF BBA IN EXAMPLE 1 USING DIFFERENT CRITERIA

	PIC	d_J	Λ_{PIC}	Λ_d	Λ_{Value}	Λ_{min}	Λ_{max}	Λ_{ave}	Λ_{opt}
PnPl	0.0526	0.2504	15	4	15	7	4	15	15
CuzzP	0.0790	0.24651	14	3	14	5	12	14	14
BetP	0.0926	0.2462	13	1	13	1	10	1	1
PraPl	0.1007	0.24647	12	2	12	3	8	1	2
PrPl	0.1638	0.2524	11	5	11	9	6	15	13
PrHyb	0.2014	0.2589	10	6	10	11	5	5	7
PrBel	0.3100	0.28084	6	9	6	11	2	3	3
DSmP_0	0.3100	0.28084	6	9	6	11	2	3	3
$\text{DSmP}_{0.001}$	0.3058	0.2801	9	7	8	14	2	5	5
HDSmP_0	0.3161	0.2827	5	11	5	9	6	5	8
$\text{HDSmP}_{0.001}$	0.3064	0.28082	8	8	9	15	1	5	6
PrScP	0.3247	0.2853	4	12	4	7	8	5	9
PrBP1	0.3480	0.2887	3	13	1	5	10	5	12
PrBP2	0.3529	0.2917	2	14	2	3	12	5	10
Un_{min}	0.4813	0.3676	1	15	3	1	4	5	10

1) Min Rule:

$$\begin{aligned}\Lambda_f &= [\min(1,1), \min(2,3), \min(3,4), \min(4,2)]\\ &= [1,2,3,2].\end{aligned}$$

This rule has a tie.

2) Max Rule:

$$\begin{aligned}\Lambda_f &= [\max(1,1), \max(2,3), \max(3,4), \max(4,2)]\\ &= [1,3,4,4].\end{aligned}$$

This rule also has a tie.

3) Arithmetic Averaging Rule:

$$\begin{aligned}\Lambda_f &= [0.5\cdot 1 + 0.5\cdot 1, 0.5\cdot 3 + 0.5\cdot 2, 0.5\cdot 3\\ &\quad + 0.5\cdot 4, 0.5\cdot 4 + 0.5\cdot 2]\\ &= [1, 2.5, 3.5, 3].\end{aligned}$$

As we can see that both weights are equal to 0.5.

We encounter the noninteger rank value. This does not matter. What we care is only the relative value of a rank. Therefore, we obtain the final result as [1, 2, 4, 3], which is the rank obtained by ordering [1, 2.5, 3.5, 3] in ascending order.

4) Optimization Rule: Here, we use footrule distance. Suppose that $\Lambda_f = [r_f(1), r_f(2), r_f(3), r_f(4)]$. We try to find a Λ_f which minimizes

$$\begin{aligned}F(\Lambda_f, \Lambda_1) + F(\Lambda_f, \Lambda_2) &= \left|r_f(1) - 1\right| + \left|r_f(2) - 2\right|\\ &\quad + \left|r_f(3) - 3\right| + \left|r_f(4) - 4\right|\\ &\quad + \left|r_f(1) - 1\right| + \left|r_f(2) - 3\right|\\ &\quad + \left|r_f(3) - 4\right| + \left|r_f(4) - 2\right|.\end{aligned}$$

By using the optimization rank fusion rule in (33), one gets $\Lambda_f^* = [1, 2, 4, 3]$.

VII. EXPERIMENTS FOR THE BI-CRITERIA EVALUATION APPROACH

In this section, we examine the previous Examples 1–3 using the new bi-criteria evaluation approach.

A. Example 1 Revisited

Table IV shows the evaluation results of different PTs (the initial PMF for PrScP used here is BetP), their distances and PICs, their ranks obtained using two criteria (Λ_{PIC} and Λ_d), and the joint rank using value-based fusion (Λ_{Value}). Table IV also provides the evaluation results using rank fusion, where Λ_{min} denotes the fused rank using min rule; Λ_{max} denotes the fused rank using max rule; Λ_{ave} denotes the fused rank using arithmetic averaging rule; and Λ_{opt} denotes the fused rank using optimization. The weight for value-based fusion is $\alpha = 0.5$ while the weights for arithmetic averaging rank fusion are $w_1 = w_2 = 0.5$. In optimization-based rank fusion, the distance used is the Spearman distance in (37) due to its quadric form, which is mathematically more tractable for optimization. The comparisons among evaluation results of different criteria are also shown in Fig. 1. Note that in all experiments here, smaller value of rank represents higher rank.

In Table IV, there exist cases of tie. Our strategy for the tie is as follows. When a tie happens, the alternatives in the tie will be assigned the same rank. The rank of the closest following the alternative of the tie will be increased by the number of alternatives in the tie. For example, in Table IV, the PICs of PrBel and DSmP_0 are the same, so their ranks are both 6. The PIC value of $\text{HDSmP}_{0.001}$ is the closest following the alternative, so its rank becomes 8. That is, there is no rank 7 here because rank 6 appeared twice.

From Table IV, although the PMF obtained from Un_{min} has the maximum PIC (thus it seems to be the best choice), it also has the maximum d_J. Therefore, it is not the best choice according to d_J. The joint evaluation results show that Un_{min} is not preferred. The bi-criteria evaluation appears more natural and helpful than PIC alone.

Also, PnPl has the minimum d_J. Thus, according to Λ_d, it is the best. But PnPl has the lowest PIC which is not good for making a clear or solid decision. From this angle, it is the worst. As we can see, the evaluation based on PIC or d_J alone is not satisfactory. In Table IV, PnPl has obtained the worst

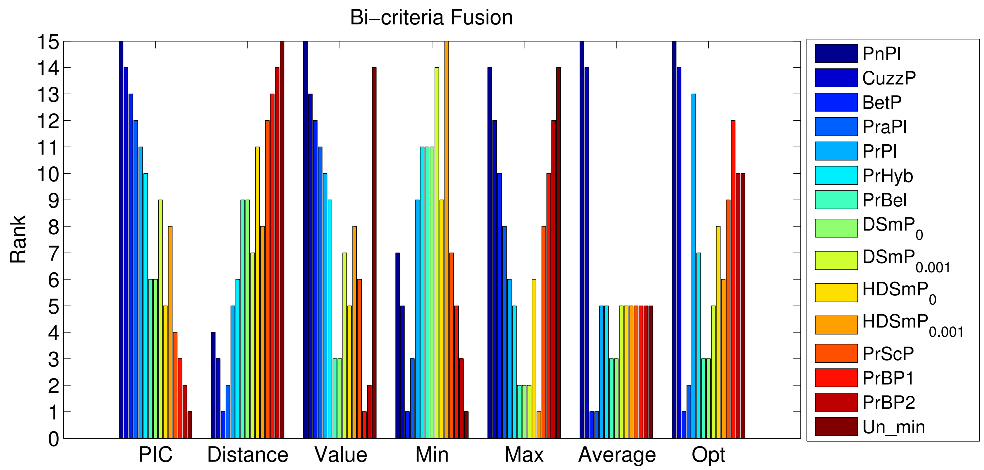


Fig. 1. Evaluation results for Example 1 using different criteria.

TABLE V
EVALUATIONS OF PT RESULTS OF BBA IN EXAMPLE 2 USING DIFFERENT CRITERIA

	PIC	d_J	Λ_{PIC}	Λ_d	Λ_{Value}	Λ_{min}	Λ_{max}	Λ_{ave}	Λ_{opt}
PrBel	N/A due to 0 value of singletons								
DSmP_0	N/A due to 0 value of singletons								
HDSmP_0	N/A due to 0 value of singletons								
PrBP1	N/A due to 0 value of singletons								
PnPl	0.0733	0.3128	11	4	9	7	10	11	11
CuzzP	0.0822	0.3123	10	3	7	5	8	10	10
BetP	0.0910	0.3121	8	1	5	1	4	1	1
PraPl	0.907	0.3122	9	2	6	3	6	2	2
PrPl	0.2471	0.3373	7	5	1	9	2	3	3
PrHyb	0.2698	0.3440	5	7	3	9	2	3	3
$\mathrm{DSmP}_{0.001}$	0.6464	0.4684	2	10	11	3	8	3	3
$\mathrm{HDSmP}_{0.001}$	0.6094	0.4572	3	9	10	5	6	3	3
PrScP	0.5987	0.4309	4	8	4	7	4	3	3
PrBP2	0.2572	0.3402	6	6	2	11	1	3	3
Un_{min}	0.7421	0.4764	1	11	8	1	10	3	3

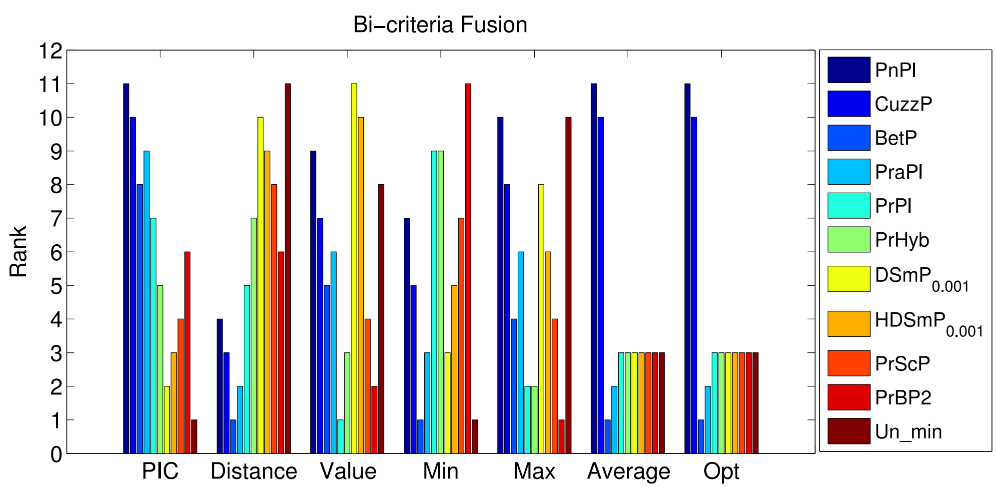


Fig. 2. Evaluation results for Example 2 using different criteria.

score based on our proposed bi-criteria evaluation approach. This new evaluation approach can assure a "good" score for both elements of the two-tuple and meanwhile it can also counteract the exaggeration of a single factor.

It can be seen from the experimental results listed in Table IV and Fig. 1, although there are some differences among different evaluation approaches, BetP, PraPl, DSmP, and HDSmP all perform pretty well in this example,

TABLE VI
EVALUATION OF PT RESULTS OF BBA IN EXAMPLE 3 USING DIFFERENT CRITERIA

	PIC	d_J	Λ_{PIC}	Λ_d	Λ_{Value}	Λ_{min}	Λ_{max}	Λ_{ave}	Λ_{opt}
PnPl	0.0113	0.3023	15	3	15	5	14	15	15
CuzzP	0.0290	0.3000	13	1	11	1	10	1	1
BetP	0.0290	0.3000	13	1	11	1	10	1	1
PraPl	0.0553	0.3023	11	3	2	5	8	1	3
PrPl	0.0553	0.3023	11	3	2	5	8	1	3
PrHyb	0.0984	0.3111	10	6	1	13	3	9	14
PraBel	0.1887	0.3354	5	10	7	9	3	5	5
DSmP_0	0.1887	0.3354	5	10	7	9	3	5	5
$\text{DSmP}_{0.001}$	0.1875	0.3351	8	8	5	15	1	9	9
HDSmP_0	0.1887	0.3354	5	10	7	9	3	5	5
$\text{HDSmP}_{0.001}$	0.0553	0.3023	9	7	4	14	2	9	10
PrScP	0.1887	0.3354	5	10	7	9	3	5	5
PrBP1	0.2334	0.3481	3	13	10	5	10	9	11
PrBP2	0.3657	0.3842	2	14	13	4	13	9	12
Un_{min}	0.5310	0.4243	1	15	14	1	14	9	12

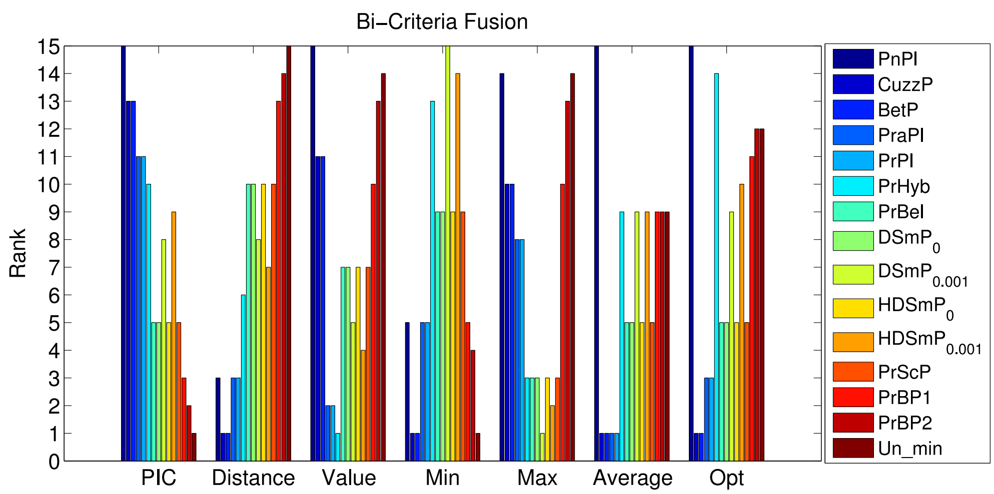


Fig. 3. Evaluation results for Example 3 using different criteria.

which make a good balance between PIC and d_J, i.e., the clarity and fidelity.

B. Example 2 Revisited

The parameters are the same as in revisited Example 1. The results are shown in Table V and Fig. 2.

In this experiment, we can see that four PTs (PrBel, PrBP1, HDSmP_0, and DSmP_0) cannot provide results due to the zero value for singletons. This shows that they are not robust and have more requirements for the original BBA. DSmP and HDSmP have the parameter which can counteract this negative effect.

From the comparisons among the remaining 11 PTs in Table V, we see that although there are some differences among different evaluation approaches, those exaggerated transformations, e.g., PnPl and Un_{min}, over-emphasizing only one criterion, do not perform that well. The bi-criteria evaluation appears more natural and helpful than using PIC alone.

C. Example 3 Revisited

The results are shown in Table VI and Fig. 3.

In this experiment, those transformations (PrPl, PrHyb, PraPl, $\text{DSmP}_{0.001}$, and $\text{HDSmP}_{0.001}$) making a good balance between clarity and fidelity always perform well using different evaluation approaches.

The above results show that DSmP and HDSmP can always generate a probability measure with less uncertainty. At the same time, this is not too risky, i.e., they can achieve a better tradeoff between PIC and risk in decision making.

We prefer the evaluation approaches using rank fusion when compared with that using value-based fusion. This is because in the evaluation approaches using rank fusion, the values are not that important. The rank fusion is not sensitive to the ranges of different criteria. In this sense, it is relatively more robust. Although in the evaluation approaches using value-based fusion, we added the step of normalization to counteract the sensitivity to the value ranges, it cannot be avoided but suppressed to some extent.

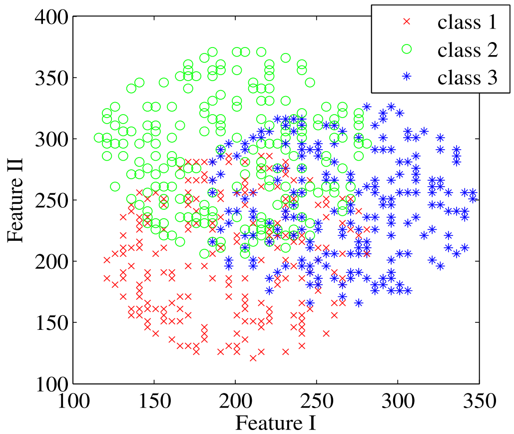


Fig. 4. Samples for classification.

Among all the evaluation approaches using rank fusion, we prefer the arithmetic averaging and the optimization based ones. This is because they are moderate, i.e., neither too pessimistic (like min rule) nor too optimistic (like max rule).

VIII. Application-Oriented Evaluation of Probability Transformations

The performance evaluation approaches for decision-making applications such as classification, are available. Therefore, we can use the evaluation of decision making under the evidence theory framework to indirectly evaluate PTs, which is illustrated in Example 5.

Example 5: In this example, a pattern classification application is considered. We consider only three classes of samples in this example, which are illustrated in Fig. 4. The 2-D-dataset is artificially generated. Samples of each class are uniformly distributed around three different centers. Abscissa and ordinate in Fig. 5 represent two feature dimensions of each sample.

The classifier used in this example is the K-nearest neighbor (K-NN) [33]. For each test sample, the output of the classifier is represented by a BBA. The corresponding BBA for each test sample is generated as follows.

1) The class space is $C = \{1, 2, 3\}$. For a test sample, find its K-NNs. In the K-NNs, calculate the ratio of the samples belonging to each class as follows:

$$P(i) = \frac{k(i)}{\sum_{j=1}^{3} k(j)} \tag{38}$$

where $P(i)$ represents the ratio of class i and $k(i)$ represents the number of samples belonging to class i in the K-NNs, $i = 1, 2, 3$. Obviously, $K = \sum_{j=1}^{3} k(j)$.

2) For the two classes s and t ($s, t \in 1, 2, 3, s \neq t$) with the top two values of $k(i), i = 1, 2, 3$, the corresponding mass assignments are generated according to [34]

$$m(\{i\}) = P(i), \forall i = s, t \tag{39}$$

The remaining mass is assigned to the total set Θ

$$m(\Theta) = 1 - m(\{s\}) - m(\{t\}). \tag{40}$$

For example, for a test sample x_q, among its seven nearest neighbors, four belong to class 1, two belong to class 2, and one belongs to class 3. The class distribution is then $P(1) = 4/7, P(2) = 2/7,$ and $P(3) = 1/7$. The dominant class is classes 1 and 2 are at the second place. The corresponding BBA is $m(\{1\}) = 4/7$, $m(\{2\}) = 2/7$, and $m(\{1, 2, 3\}) = 1/7$.

There are 200 samples for each one class with a total of 600 samples. In each experiment cycle, the samples are randomly selected from each class with 100 samples for training (300 training samples in total) and the remaining samples are used for testing (300 test samples in total).

For a PT, the decision result will be class i_1 if

$$i_1 = \arg\max_{j} \mathrm{PT}(j), i_2 = \arg\max_{j, j \neq i_1} \mathrm{PT}(j) \tag{41}$$

$$\mathrm{PT}(i_1) - \mathrm{PT}(i_2) \geq \tau \tag{42}$$

where τ is the threshold for decision making. If (42) is not satisfied, i_1 will be rejected. The threshold τ is selected from

$$\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 0.8, 0.9, 1.0\}.$$

For each of these tested thresholds, the above experiment procedure is repeated 100 times to calculate the average performance of different PTs.[2] Based on the simulation results, the average rejection rates (average over all threshold values) of different PTs are shown in Fig. 5.

As we can see from Fig. 5, the rejection rate of $\mathrm{Un_{min}}$ is the minimum one. This is because $\mathrm{Un_{min}}$ emphasizes on the clarity. Therefore, the difference between different alternatives is relatively large. We propose to use "rejection–error" curve as shown in Fig. 6 to evaluate PTs.

The abscissas and ordinates of the points on rejection–error curves are, respectively, the average rejection rate values and the average error rate values for different PTs at each threshold τ. For any PT at each threshold value, the average rejection and average error rates are the mean of the repeated 100 times simulations. The deviations of the rejection rate and error rate of each PT at different thresholds are listed in Tables VII and VIII, respectively.

From the results in Fig. 6, it can be seen that given the same rejection rate, the classification error rate of $\mathrm{Un_{min}}$ is always the highest. The decision results based on $\mathrm{Un_{min}}$ are the worst. Although $\mathrm{Un_{min}}$ has the least uncertainty degree and the minimum rejection rate, it is not the winner. The smaller rejection rate is at the price of higher classification error rate. The rejection–error curves can be used as a comprehensive and indirect application-oriented evaluation approach for PTs.

The performance of other PTs (except for $\mathrm{Un_{min}}$) is similar when using application-oriented performance evaluation (rejection–error curves).

IX. Desired Properties of Probability Transformations

In this section, we discuss some desired properties of PTs.

[2]The PTs $\mathrm{HDSmP_0}$, PrBP1, and PrBP2 are not included in our simulation because they cannot be computed when zero masses occur.

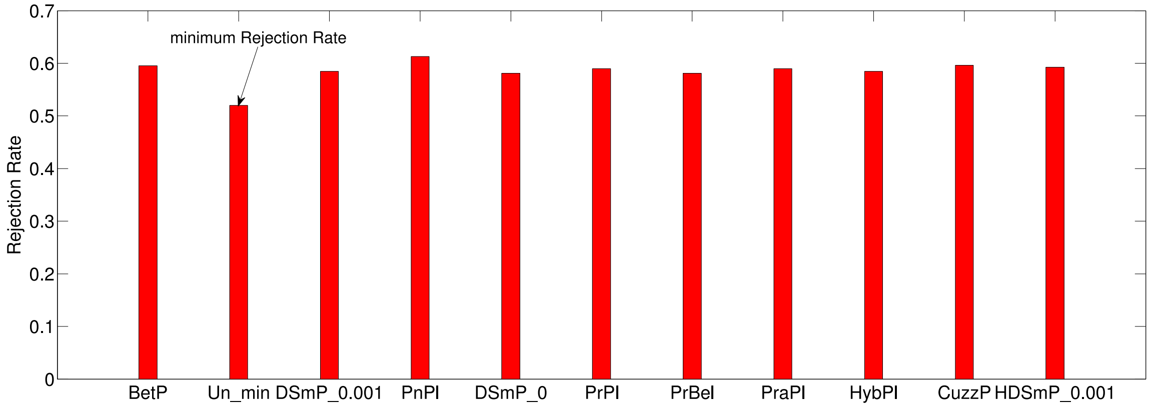


Fig. 5. Comparison of average rejection rate over all thresholds among different PTs.

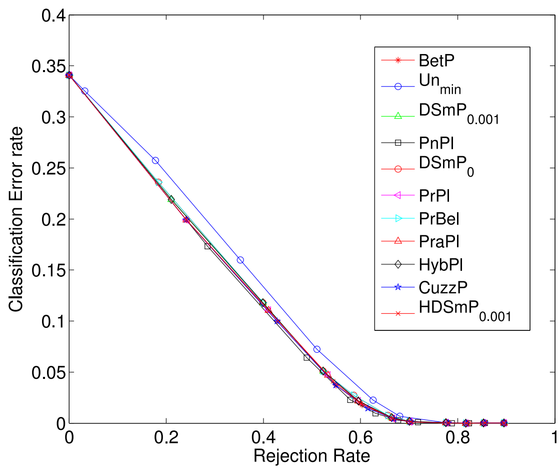


Fig. 6. Rejection-error curves comparison of different PTs.

A. Order Preservation

It is preferred that a PT can maintain some order before and after the transformation, e.g., the order of the uncertainty degree. Given N BBAs: $m_1, \ldots, m_N$, we can obtain the rank or order of their uncertainty degree according to some uncertainty measure in evidence theory as introduced in the sequel

$$\mathrm{Rank}_m = [r_m(1), \ldots, r_m(N)].$$

After applying the PTs, the rank of $P_1, \ldots, P_N$ in terms of uncertainty degree in probability theory[3] is

$$\mathrm{Rank}_P = [r_P(1), \ldots, r_P(N)].$$

It is very hard to completely maintain the order after the transformation. However, the degree of accordance or consistency between Rank_m and Rank_P, i.e., the degree of order preserving, can be used to evaluate different PTs. Such a degree of order preservation can be defined using the distance between ranks as introduced before. Less degree of order preserving represents more twist or loss of information in the procedure of PTs.

[3]Here, we only consider Shannon's entropy as the ranking criterion for convenience.

When calculating the degree of order preservation, we need to specify an uncertainty measure of belief functions. Some uncertainty measures in evidence theory are as follows.

1) Nonspecificity: It [35] is defined as

$$N(m) = \sum_{A \subseteq \Theta} m(A) \log_2 |A|. \tag{43}$$

2) Confusion: It [36] is defined using the BBA m and the Bel in the spirit of entropy as

$$\mathrm{Conf}(m) = -\sum_{A \in \Theta} m(A) \log_2(\mathrm{Bel}(A)). \tag{44}$$

3) Dissonance: It [37] is defined using the BBA m and the Pl in the spirit of entropy as

$$\mathrm{Diss}(m) = -\sum_{A \in \Theta} m(A) \log_2(\mathrm{Pl}(A)). \tag{45}$$

4) Aggregate Uncertainty Measure: Let Bel be a belief measure on the FOD Θ. The aggregate uncertainty (AU) [38] associated with Bel is measured by

$$\mathrm{AU}(\mathrm{Bel}) = \max_{\mathcal{P}_{\mathrm{Bel}}} \left[-\sum_{\theta \in \Theta} p_\theta \log_2 p_\theta \right] \tag{46}$$

where the maximum is taken over all probability distributions that are consistent with the given belief function. $\mathcal{P}_{\mathrm{Bel}}$ consists of all probability distributions $\langle p_\theta | \theta \in \Theta \rangle$ satisfying

$$\begin{cases} p_\theta \in [0, 1] \forall \theta \in \Theta \\ \sum_{\theta \in \Theta} p_\theta = 1 \\ \mathrm{Bel}(A) \leq \sum_{\theta \in A} p_\theta \leq 1 - \mathrm{Bel}(\bar{A}) \ \forall A \subseteq \Theta. \end{cases} \tag{47}$$

AU is an aggregated total uncertainty measure.

AU satisfies all the requirements of uncertainty measure [39] including probability consistency, set consistency, value range, sub-additivity, and additivity for the joint BBA in Cartesian space. AU also has the drawbacks [3] of high computing complexity, high insensitivity to the changes of evidence, etc.

5) Ambiguity Measure: Let m be a BBA defined over the FOD $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$. Ambiguity measure (AM) [40] is defined as

$$\mathrm{AM}(m) = -\sum_{\theta \in \Theta} \mathrm{BetP}_m(\theta) \log_2(\mathrm{BetP}_m(\theta)) \tag{48}$$

TABLE VII
DEVIATIONS OF THE REJECTION RATE OF EACH PT AT DIFFERENT τ

τ	BetP	Un_{min}	$\mathrm{DSmP}_{0.001}$	PnPl	DSmP_0	PrPl	PrBel	PraPl	HybPl	CuzzP	$\mathrm{HDSmP}_{0.001}$
0	0	0	0	0	0	0	0	0	0	0	0
0.1	0.0406	0.0130	0.0382	0.0434	0.0360	0.0406	0.0360	0.0406	0.0382	0.0406	0.0406
0.2	0.0401	0.0332	0.0421	0.0356	0.0421	0.0412	0.0421	0.0412	0.0421	0.0401	0.0412
0.3	0.0374	0.0496	0.0418	0.0312	0.0418	0.0411	0.0418	0.0411	0.0418	0.0374	0.0387
0.4	0.0284	0.0453	0.0309	0.0238	0.0330	0.0309	0.0330	0.0309	0.0309	0.0276	0.0298
0.5	0.0265	0.0320	0.0282	0.0267	0.0284	0.0274	0.0284	0.0274	0.0282	0.0265	0.0274
0.6	0.0209	0.0247	0.0214	0.0199	0.0214	0.0209	0.0214	0.0209	0.0214	0.0209	0.0209
0.7	0.0208	0.0213	0.0213	0.0206	0.0213	0.0213	0.0213	0.0213	0.0213	0.0208	0.0208
0.8	0.0185	0.0192	0.0192	0.0185	0.0192	0.0192	0.0192	0.0192	0.0192	0.0195	0.0195
0.9	0.0165	0.0167	0.0167	0.0165	0.0167	0.0167	0.0167	0.0167	0.0167	0.0165	0.0165
1.0	0.0131	0.0131	0.0131	0.0131	0.0131	0.0131	0.0131	0.0131	0.0131	0.0131	0.0131

TABLE VIII
DEVIATIONS OF THE ERROR RATE OF EACH PT AT DIFFERENT τ

τ	BetP	Un_{min}	$\mathrm{DSmP}_{0.001}$	PnPl	DSmP_0	PrPl	PrBel	PraPl	HybPl	CuzzP	$\mathrm{HDSmP}_{0.001}$
0	0.0204	0.0204	0.0204	0.0204	0.0204	0.0204	0.0204	0.0204	0.0204	0.0204	0.0199
0.1	0.0318	0.0182	0.0308	0.0319	0.0300	0.0318	0.0300	0.0318	0.0308	0.0318	0.0318
0.2	0.0283	0.0238	0.0322	0.0233	0.0322	0.0311	0.0322	0.0311	0.0322	0.0283	0.0311
0.3	0.0191	0.0268	0.0228	0.0127	0.0228	0.0214	0.0228	0.0214	0.0228	0.0191	0.0197
0.4	0.0099	0.0236	0.0118	0.0071	0.0144	0.0118	0.0144	0.0118	0.0118	0.0094	0.0108
0.5	0.0059	0.0140	0.0080	0.0057	0.0085	0.0070	0.0085	0.0070	0.0080	0.0059	0.0070
0.6	0.0036	0.0075	0.0039	0.0027	0.0039	0.0039	0.0039	0.0039	0.0039	0.0036	0.0036
0.7	0.0005	0.0008	0.0008	0.0005	0.0008	0.0008	0.0008	0.0008	0.0008	0.0005	0.0005
0.8	3e-17	0.0005	0.0005	3e-17	0.0005	0.0005	0.0005	0.0005	0.0005	3e-17	3e-17
0.9	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17
1.0	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17

where $\mathrm{BetP}_m(\theta) = \sum_{\theta \in B, B \subseteq \Theta} m(B)/|B|$ is the pignistic probability. Jousselme *et al.* [40] declared that the AM satisfies the requirements of uncertainty measure and at the same time it overcomes the defects of AU, but in fact AM does not satisfy the sub-additivity [41]. Moreover, in [39], AM has been proved to be logically nonmonotonic under some conditions.

6) Contradiction Measure: The contradiction measure [42] is defined as

$$\mathrm{CM}(m) = \sqrt{\frac{2n}{n-1}} \cdot \sum_{X \in \mathcal{X}} m(X) \cdot d_J(m, m_X) \tag{49}$$

where $\mathcal{X}$ denotes the set of all focal elements of m and d_J is Jousselme's distance. There exists contradiction measure (CM) $\in [0, 1]$.

B. Simulation of Degree of Order Preservation

Our simulation consists of the following steps.

Step 1: Randomly generate ten BBAs and calculate the degree of uncertainty in each BBA to generate Rank_m.

Step 2: Apply the transformation using N types of PTs. We can obtain various Rank_p^i, $i = 1, \ldots, N$.

Step 3: Calculate the distance $(d_p(i))$ between Rank_m and each Rank_p^i, $i = 1, \ldots, N$.

TABLE IX
ALGORITHM 1. RANDOM GENERATION OF BBA

Input: Θ: Frame of discernment;
N_{max}: Maximum number of focal elements
Output: **Output**: m: BBA
Generate $\mathcal{P}(\Theta)$, which is the power set of Θ;
Generate a random permutation of $\mathcal{P}(\Theta) \rightarrow \mathcal{R}(\Theta)$;
Generate an integer between 1 and $N_{max} \rightarrow l$;
FOReach First k elements of $\mathcal{R}(\Theta)$ do
Generate a value within $[0, 1] \rightarrow m_i$, $i = 1, \ldots, l$;
END
Normalize the vector $m = [m_1, \ldots, m_l] \rightarrow m'$;
$m(A_i) = m_i'$;

Step 4: Repeat step 1–3 a 100 times. Calculate the average distance and the corresponding standard deviation as follows:

$$d_{\mathrm{pm}}(i) = \frac{1}{100} \sum_{j=1}^{100} d_p^j(i) \tag{50}$$

$$d_{p-\mathrm{std}}(i) = \sqrt{\frac{1}{100-1} \sum_{j=1}^{100} \left(d_p^j(i) - d_{\mathrm{pm}}(i)\right)^2}. \tag{51}$$

The PTs with smaller $d_{\mathrm{pm}}(i)$ and $d_{p-\mathrm{std}}(i)$ are preferred.

TABLE X
EVALUATION OF DEGREE OF ORDER CHANGE (AVERAGE VALUE)

Prob Trans	PnPl	CuzzP	BetP	PraPl	PrPl	PrHyb	$DSmP_{0.001}$	$HDSmP_{0.001}$	Un_{min}
Nonspecificity	0.4140	0.3958	0.4100	0.3640	0.4411	0.4240	0.4760	0.4476	0.4671
Contradiction	0.3567	0.3756	0.3618	0.4111	0.3320	0.3502	0.3551	0.3767	0.5022
AM	0.1282	0.0476	0	0.0989	0.0622	0.0589	0.2571	0.2284	0.4560
AU	0.0416	0.0304	0.0311	0.0324	0.0489	0.0507	0.1151	0.1082	0.2029
Confusion	0.5896	0.6009	0.6004	0.6084	0.6064	0.6071	0.6213	0.6076	0.5302
Dissonance	0.4882	0.4998	0.4836	0.5389	0.4511	0.4696	0.4453	0.4673	0.5247

TABLE XI
EVALUATION OF DEGREE OF ORDER CHANGE (STANDARD DEVIATION)

Prob Trans	PnPl	CuzzP	BetP	PraPl	PrPl	PrHyb	$DSmP_{0.001}$	$HDSmP_{0.001}$	Un_{min}
Nonspecificity	0.1278	0.1183	0.1207	0.1177	0.1237	0.1177	0.1335	0.1263	0.1271
Contradiction	0.1289	0.1193	0.1170	0.1249	0.1176	0.1127	0.1317	0.1302	0.1387
AM	0.0711	0.0378	0	0.0567	0.0417	0.0372	0.1127	0.0915	0.1119
AU	0.0545	0.0404	0.0400	0.0432	0.0542	0.0579	0.1033	0.0945	0.1282
Confusion	0.1297	0.1206	0.1192	0.1200	0.1187	0.1188	0.1175	0.1282	0.1131
Dissonance	0.1380	0.1309	0.1308	0.1294	0.1317	0.1296	0.1409	0.1296	0.1309

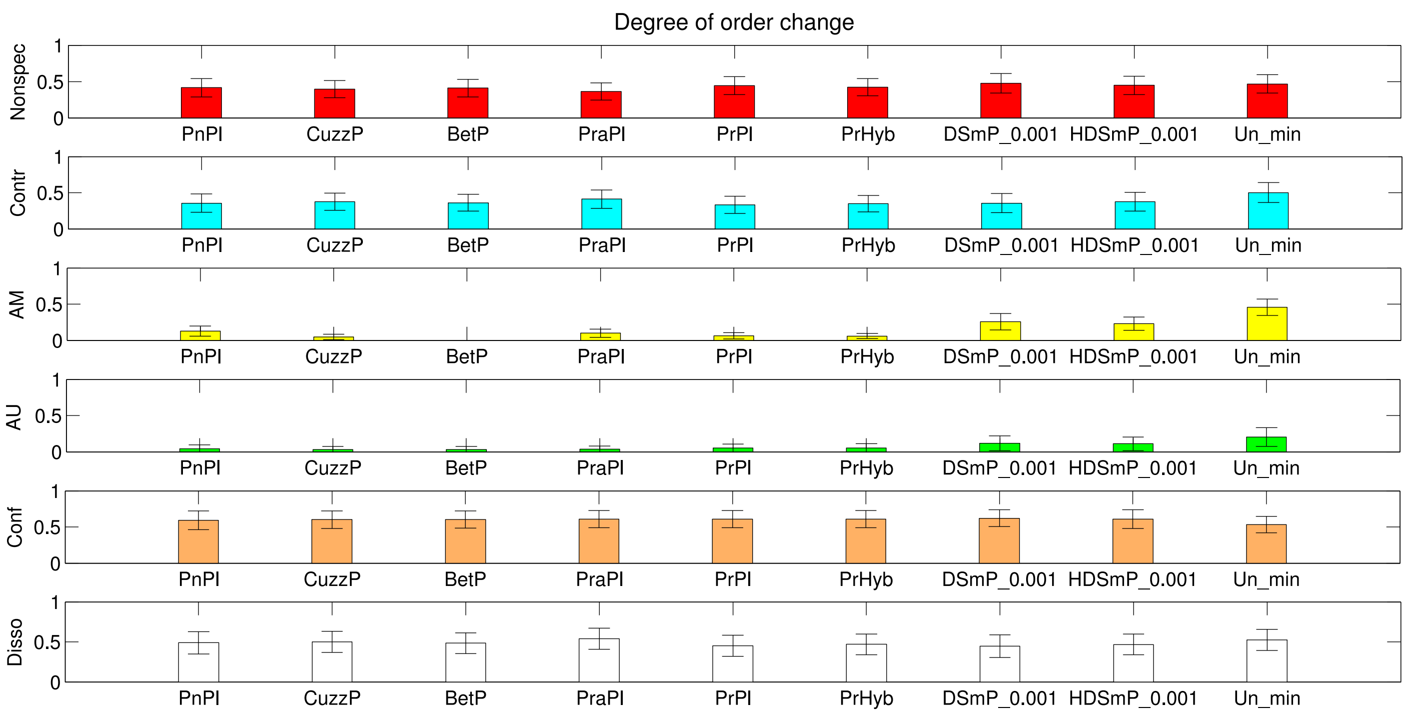


Fig. 7. Evaluation in terms of the degree of order change.

In step 1, BBAs are generated using Algorithm 1 [17] in Table IX. All six types of uncertainty measures introduced above are used in the simulations. The distance between Rank_m and Rank_p used is the Kendall distance in (36). The evaluation results are shown in Tables X and XI and Fig. 7.

From Tables X and XI and Fig. 7, Un_{min} provides the worst (with the lowest degree of order preservation). Although all the degree of order change based on different uncertainty measures are all listed in Tables X and XI and Fig. 7, we prefer to use the ones based on AU. The reasons are as follows. First, in all the uncertainty measures used here, only AU is a strict total uncertainty measure, which can describe both discord and nonspecificity in a body of evidence. Second, as we can see from the subfigure using AM, BetP's degree of order change is zero. However, this does not make sense, because AM is defined based on BetP. Therefore, it is partial (or nonneutral) when BetP is also included for evaluation. AU is relative more appropriate to be used here. According to the subfigure based on AU, Un_{min} is the worst. HDSmP and DSmP are also not that good in terms of order preservation. Other PTs perform similar to each other in terms of order preservation.

Even AU is still not absolutely impartial (or neutral), because AU is designed also based on some PT (maximization of entropy). If such an entropy maximization-based PT is also included for evaluation, it will be partial (or nonneutral). So some new uncertainty measure for BBA not related to PT is needed for an impartial evaluation. We think that the CM in (49) is a good attempt. It is an uncertainty measure which is not based on a PT, although its strictness (satisfying the requirements of the uncertainty measure) still deserves further research.

X. Conclusion

In this paper, we focus on the evaluations of PTs of a belief function. The existing transformations are briefly reviewed and compared. Our experimental results and analysis show that PIC criterion alone is insufficient to truly measure the quality of a PT. A compromise between fidelity and clarity is achieved by the joint use of PIC and the distance of evidence. We have also proposed an application-oriented evaluation approach for PTs. Furthermore, we have evaluated PTs by their robustness to preserve the uncertainty order of the original BBAs. The simulation results show that our proposed evaluation approaches are able to make rational comparison of different PTs. Future work includes the development of general and direct measures of uncertainty of a BBA, which do not depend on the choice of PTs. This is important for the property of uncertainty order preservation.

Note that the evaluations for the issues in evidence theory (like the PTs, the evidence combination, the determination of BBA, etc.) lack solid theoretical foundation so far. In the future, we will attempt to propose more rational and useful criteria for PTs and try to establish more theoretically sound evaluation approaches for PTs, which are important and challenging problems.

Acknowledgment

The authors would like to thank the reviewers for providing very helpful comments and suggestions.

References

[1] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Ann. Math. Stat.*, vol. 38, no. 4, pp. 325–339, 1967.

[2] G. Shafer, *A Mathematical Theory of Evidence*, vol. 1. Princeton, NJ, USA: Princeton Univ. Press, 1976.

[3] P. Smets and R. Kennes, "The transferable belief model," *Artif. Intell.*, vol. 66, no. 2, pp. 191–234, 1994.

[4] F. Smarandache and J. Dezert, Eds., *Applications and Advances of DSmT for Information Fusion*, vol. 3. Rehoboth, DE, USA: Amer. Res. Press, 2009. [Online]. Available: http://www.gallup.unm.edu/~smarandache/DSmT-book3.pdf

[5] J. J. Sudano, "Pignistic probability transforms for mixes of low- and high-probability events," in *Proc. 4th Int. Conf. Inf. Fusion*, Montreal, QC, Canada, Aug. 2001, pp. 23–27.

[6] J. J. Sudano, "The system probability information content (PIC) relationship to contributing components, combining independent multi-source beliefs, hybrid and pedigree pignistic probabilities," in *Proc. 5th Int. Conf. Inf. Fusion*, Annapolis, MD, USA, Jul. 2002, pp. 1277–1283.

[7] J. J. Sudano, "Yet another paradigm illustrating evidence fusion (YAPIEF)," in *Proc. 9th Int. Conf. Inf. Fusion*, Florence, Italy, Jul. 2006, pp. 1–7.

[8] B. R. Cobb and P. P. Shenoy, "On the plausibility transformation method for translating belief function models to probability models," *Int. J. Approx. Reason.*, vol. 41, no. 3, pp. 314–330, 2006.

[9] J. Dezert and F. Smarandache, "A new probabilistic transformation of belief mass assignment," in *Proc. 11th Int. Conf. Inf. Fusion*, Cologne, Germany, Jun./Jul. 2008, pp. 1–8.

[10] J. J. Sudano, "Belief fusion, pignistic probabilities, and information content in fusing tracking attributes," in *Proc. IEEE Radar Conf.*, Philadelphia, PA, USA, Apr. 2004, pp. 218–224.

[11] W. Pan and H. Yang, "New methods of transforming belief functions to pignistic probability functions in evidence theory," in *Proc. Int. Workshop Intell. Syst. Appl.*, Wuhan, China, 2009, pp. 1–5.

[12] F. Cuzzolin, "On the properties of the intersection probability," in *Proc. 10th Eur. Conf. Symbol. Quant. Approach. Reason. Uncertainty (ECSQARU)*, Verona, Italy, Jul. 2007, pp. 287–298.

[13] J. Dezert, D. Han, Z. Liu, and J.-M. Tacnet, "Hierarchical proportional redistribution for bba approximation," in *Belief Functions: Theory and Applications.* Advances in Intelligent and Soft Computing, vol. 164, T. Denoeux and M.-H. Masson, Eds. Berlin, Germany: Springer, 2012, pp. 275–283.

[14] D. Han, J. Dezert, C. Han, and Y. Yang, "Is entropy enough to evaluate the probability transformation approach of belief function?" in *Proc. 13th Conf. Inf. Fusion*, Edinburgh, U.K., Jul. 2010, pp. 1–7.

[15] A.-L. Jousselme, D. Grenier, and É. Bossé, "A new distance between two bodies of evidence," *Inf. Fusion*, vol. 2, no. 2, pp. 91–101, 2001.

[16] M. Daniel, "Probabilistic transformations of belief functions," in *Proc. 8th Eur. Conf. Symbol. Quant. Approach. Reason. Uncertainty (ECSQARU)*, Barcelona, Spain, Jul. 2005, pp. 539–551.

[17] A.-L. Jousselme and P. Maupin, "Distances in evidence theory: Comprehensive survey and generalizations," *Int. J. Approx. Reason.*, vol. 53, no. 2, pp. 118–145, 2012.

[18] D. Han, Y. Deng, C. Han, and Y. Yang, "Some notes on betting commitment distance in evidence theory," *Sci. China Inf. Sci.*, vol. 55, no. 3, pp. 558–565, 2012.

[19] M. Bouchard, A.-L. Jousselme, and P.-E. Doré, "A proof for the positive definiteness of the Jaccard index matrix," *Int. J. Approx. Reason.*, vol. 54, no. 5, pp. 615–626, 2013.

[20] R. R. Yager, "Decision making under Dempster–Shafer uncertainties," *Int. J. Gen. Syst.*, vol. 20, no. 3, pp. 233–245, 1992.

[21] W. Lv, "Decision-making rules based on belief interval with D-S evidence theory," in *Fuzzy Information and Engineering*. Advances in Soft Computing, vol. 40, B. Cao, Ed. Berlin, Germany: Springer, 2007, pp. 619–627.

[22] P. Smets, "Decision making in the TBM: The necessity of the pignistic transformation," *Int. J. Approx. Reason.*, vol. 38, no. 2, pp. 133–147, 2005.

[23] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[24] R. Horst, P. M. Pardalos, and N. V. Thoai, *Introduction to Global Optimization*, 2nd ed. Boston, MA, USA: Kluwer Academic, 2000.

[25] D. T. Pham and D. Karaboga, *Intelligent Optimisation Techniques*. London, U.K.: Springer, 2000.

[26] J. Kennedy, J. F. Kennedy, and R. C. Eberhart, *Swarm Intelligence*. San Francisco, CA, USA: Morgan Kaufmann, 2001.

[27] M. E. Renda and U. Straccia, "Web metasearch: Rank vs. score based rank aggregation methods," in *Proc. 18th ACM Symp. Appl. Comput.*, Melbourne, FL, USA, Mar. 2003, pp. 841–846.

[28] A. Argentini, "Ranking aggregation based on belief function theory," Ph.D. dissertation, Dept. Inf. Eng. Comput. Sci., Univ. Trento, Trento, Italy, 2012.

[29] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar, "Rank aggregation methods for the Web," in *Proc. 10th Int. Conf. World Wide Web*, Hong Kong, China, May 2001, pp. 613–622.

[30] M. G. Kendall, *Rank Correlation Methods*. London, U.K.: Griffen, 1948.

[31] R. Fagin, R. Kumar, and D. Sivakumar, "Comparing top k lists," in *Proc. 14th Annu. ACM-SIAM Symp. Discrete Algorithms*, Baltimore, MD, USA, 2003, pp. 28–36.

[32] J. L. Myers, A. Well, and R. F. Lorch, *Research Design and Statistical Analysis*, 3rd ed. New York, NY, USA: Routledge, 2001.

[33] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, NY, USA: Wiley, 2012.

[34] Y. Bi, D. Bell, and J. Guan, "Combining evidence from classifiers in text categorization," in *Proc. 8th Int. Conf. Knowl.-Based Intell. Inf. Eng. Syst.*, Wellington, New Zealand, Sep. 2004, pp. 521–528.

[35] D. Dubois and H. Prade, "A note on measures of specificity for fuzzy sets," *Int. J. Gen. Syst.*, vol. 10, no. 4, pp. 279–283, 1985.

[36] U. Höhle, "Entropy with respect to plausibility measures," in *Proc. 12th IEEE Int. Symp. Multiple-Valued Logic*, Paris, France, May 1982, pp. 167–169.

[37] R. R. Yager, "Entropy and specificity in a mathematical theory of evidence," *Int. J. Gen. Syst.*, vol. 9, no. 4, pp. 249–260, 1983.

[38] D. Harmanec and G. J. Klir, "Measuring total uncertainty in Dempster–Shafer theory: A novel approach," *Int. J. Gen. Syst.*, vol. 22, no. 4, pp. 405–419, 1994.

[39] J. Abellán and A. Masegosa, "Requirements for total uncertainty measures in Dempster–Shafer theory of evidence," *Int. J. Gen. Syst.*, vol. 37, no. 6, pp. 733–747, 2008.

[40] A.-L. Jousselme, C. Liu, D. Grenier, and É. Bossé, "Measuring ambiguity in the evidence theory," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 36, no. 5, pp. 890–903, Sep. 2006.

[41] G. J. Klir and H. W. Lewis, "Remarks on 'measuring ambiguity in the evidence theory,"' *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 38, no. 4, pp. 995–999, Jul. 2008.

[42] F. Smarandache, D. Han, and A. Martin, "Comparative study of contradiction measures in the theory of belief functions," in *Proc. 15th Int. Conf. Inf. Fusion*, Singapore, Jul. 2012, pp. 271–277.

Deqiang Han (S'08–M'09) was born in Xi'an, China, in 1980. He received the bachelor's degree in communication and control engineering and the master's and Ph.D. degrees in control science and engineering from Xi'an Jiaotong University, Xi'an, in 2001, 2004, and 2008, respectively.

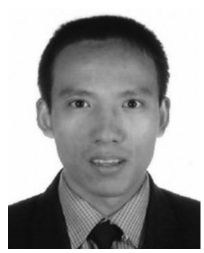
He is currently an Associate Professor with Xi'an Jiaotong University. His current research interests include evidence theory, information fusion, and pattern classification.

Dr. Han is an International Society of Information Fusion (ISIF) Member and Technical Program Committee Member for the 1st–6th Chinese Conference on Information Fusion in 2009–2014.

Jean Dezert was born in l'Haÿ-les-Roses, France, in 1962. He received the D.E.E. degree from the École Française de Radioélectricité Électronique et Informatique, Paris, France, the D.E.A. degree from University Paris VII (Jussieu), Paris, and the Ph.D. degree from University Paris XI, Orsay, France, in 1985, 1986, and 1990, all in automatic control and signal processing.

Since 1993, he has been a Senior Research Scientist with the Information and Fusion Systems Research Unit, Information and Modeling and Processing Department, ONERA. His current research interests include autonomous navigation, estimation theory, stochastic systems theory, and its applications to multisensor-multitarget tracking, information fusion, and plausible reasoning.

Dr. Dezert has served as a Local Arrangements Organizer for the 2000 3rd International Conference on Information Fusion (Fusion) in Paris, a Secretary for ISIF in 2001, and an Executive Vice-President for ISIF in 2004. He has been involved in the Technical Program Committee of Fusion 2001–2014 International Conferences. He was a Board Member of the International Society of Information Fusion. He gave several invited plenary talks and seminars on information fusion in Europe, America, and China during last year.



Zhansheng Duan (S'07–M'11) was born in Shaanxi, China, in 1977. He received the B.S. and Ph.D. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 1999 and 2005, respectively, and the Ph.D. degree in electrical engineering from the University of New Orleans, New Orleans, LA, USA, in 2010.

In 2010, he was an Assistant Professor/Research with the Department of Computer Science, University of New Orleans, for four months, then, he was with the Center for Information Engineering Science Research, Xi'an Jiaotong University, where he is currently an Associate Professor. His current research interests include Bayesian learning, estimation and detection theory, target tracking, information fusion, and performance evaluation. He has coauthored the book entitled *Multisource Information Fusion* (Tsinghua Univ. Publishing House, 2006).

Dr. Duan is a member of the International Society of Information Fusion and the Honor Society of Eta Kappa Nu.