

# Genetic Algorithm Based on a New Similarity for Probabilistic Transformation of Belief Functions †

Yilin Dong <sup>1</sup>, Lei Cao <sup>1,\*</sup> and Kezhu Zuo <sup>2</sup><sup>1</sup> Department of Artificial Intelligence, Shanghai Maritime University, Shanghai 201306, China<sup>2</sup> School of Cyber Science and Engineering, Southeast University, Nanjing 210096, China

\* Correspondence: lcao@shmtu.edu.cn

† This paper is an extended version of Dong, Y.; Li, X.; Dezert, J. A new probabilistic transformation based on evolutionary algorithm for decision making. In Proceedings of the 20th International Conference on Information Fusion (FUSION 2017), Xi'an, China, 10–13 July 2017; doi: 10.23919/ICIF.2017.8009821.

**Abstract:** Recent studies of alternative probabilistic transformation (PT) in Dempster–Shafer (DS) theory have mainly focused on investigating various schemes for assigning the mass of compound focal elements to each singleton in order to obtain a Bayesian belief function for decision-making problems. In the process of such a transformation, how to precisely evaluate the closeness between the original basic belief assignments (BBAs) and transformed BBAs is important. In this paper, a new aggregation measure is proposed by comprehensively considering the interval distance between BBAs and also the sequence inside the BBAs. Relying on this new measure, we propose a novel multi-objective evolutionary-based probabilistic transformation (MOEPT) thanks to global optimizing capabilities inspired by a genetic algorithm (GA). From the perspective of mathematical theory, convergence analysis of EPT is employed to prove the rationality of the GA used here. Finally, various scenarios in evidence reasoning are presented to evaluate the robustness of EPT.

**Keywords:** probabilistic transformation (PT); similarity measure; convergence analysis; belief functions (BFs)



**Citation:** Dong, Y.; Cao, L.; Zuo, K. Genetic Algorithm Based on a New Similarity for Probabilistic Transformation of Belief Functions. *Entropy* **2022**, *24*, 1680. <https://doi.org/10.3390/e24111680>

Academic Editor: Robert Niven

Received: 29 September 2022

Accepted: 15 November 2022

Published: 17 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. Background and Research Motivation

Since the pioneering work of Dempster and Shafer [1,2], known as Dempster–Shafer evidence theory (DST), belief functions are widely used in information fusion for decision making [3,4]. However, the computational complexity of reasoning with DST is one of the major points of criticism this formalism has to face. To overcome this difficulty, various approximating methods have been suggested that aim at reducing the number of focal elements in the frame of discernment (FoD) in order to maintain the tractability of computation. One common strategy is to simplify the FoD by removing or aggregating focal elements for approximating the original belief function [5]. Among these methods, probabilistic transformations (PTs) seem particularly desirable for reducing such computational complexity by means of assigning the mass of non-singleton elements to some singletons of the FoD [6,7]. The research on this probabilistic measure has received a lot of attention [8], and many efficient PTs have been proposed by scholars in recent years. Among them, a classical transformation, denoted as BetP [6], was usually adopted because it offered a compromise between the maximum of credibility (Bel) and the maximum of plausibility (Pl) for decision making. Unfortunately, BetP does not provide the highest probabilistic information content (PIC) [9], and Shenoy argued against BetP in his publication [10]. Sudano [11] also proposed series of alternatives and principles for them similar to BetP, which were called PrPI, PrBel and PrHyb. CuzzP [12], which was proposed by Cuzzolin in the framework of DST in 2009, showed its probabilistic transformation ability. Another novel transformation was proposed by Dezert and Smarandache in the

framework of Dezert–Smarandache theory (DSmT) which was called Dezert–Smarandache probability (DSmP) [9], and comprehensive comparisons were made in [9] to prove the advantages of DSmP with respect to other PTs.

### 1.2. Challenges

Although various techniques have been proposed to evaluate PTs, these methods have limitations. On the one hand, the PIC or Shannon’s entropy criterion was applied to evaluate PTs, with less uncertainty (*clarity*) for the BBAs obtained from PTs being preferred in order to make decisions easily. However, Han et al. [13] and Bucci in [14] illustrated the irrationality of the overemphasizing of the Shannon entropy or PIC. On the other hand, from the perspective of *fidelity*, to characterize the difference between the transformed BBAs and original BBAs, various distances were applied. The most famous distance was  $D_J$  [15], but it was not good enough to capture the difference between BBAs in some cases, and these can be seen in [16,17]. Thus, either the PIC or distance alone is not efficient and comprehensive enough to quantify all sorts of aspects of dissimilarity which inevitably need to be involved in PTs. To address incomprehensive evaluations, several two-dimensional measures [18] have been proposed in order to make sure that the results obtained by PTs are consistent in some manner with the original BBAs. Han in [19] proposed a 2D criteria which jointly uses distance and PIC measures to create a balance between fidelity and clarity. Liu [16] used a two-dimensional measure to effectively detect conflicts among the evidence. In [20], Liu proposed both a distance and a conflict coefficient based on probabilistic transformations  $BetP$  to characterize the dissimilarity, which are complementary in a certain sense. That aside, Deng [21] developed a new probability transformation method called the decision-based PT in belief functions theory. Following Deng’s idea [21], Zhao [22] proposed an importance-based PT to achieve the transformation of original basic belief assignments. Recently, Ma [23] integrated a fuzzy closeness and correlation coefficient to generate a new dissimilarity measure to characterize not only the difference between BBAs but also the divergence degree of the hypothesis that two BBAs support. By analyzing the mentioned existing methods, the relationship between the techniques of PTs and their corresponding evaluations are almost independent, except in [19]; that is to say, the methods of evaluation only assess the existing PTs instead of facilitating the development of novel PTs themselves.

### 1.3. Contributions

In this paper, we present a novel PT method based on a multi-objective algorithm (MOEPT) using reasonable and comprehensive two-dimensional criteria in order to capture the similarity in the process of a PT. This new method has some connections with the recent algorithm proposed in [19]. However, the main differences lie in the following aspects:

- The 2D criteria, PIC and Jousselme’s distance have pointed out its drawbacks in many references [17,24]. Thus, an efficient and different aggregation measure is proposed. Its novelty lies in considering the drawback of the past description of the distance between the evidence. In other words, up to now, most distances were defined according to the corresponding focal elements between two sources of evidence, and the sequence inside the assignments of the focal elements themselves was not considered. The sequence might also lead to dissimilarity, which is referred to as “self-conflict or self-contradiction” [25];
- More specific steps of evolutionary-based algorithms are given in detail. Aside from that, the convergence analysis of the MOEPT is illustrated to prove the rationality of using GAs. Moreover, some bugs are detected and fixed when using the MOEPT with traditional constraints;
- The specific application problem, target type tracking (TTT), is efficiently solved and discussed based on the proposed method with a novel simple constraint.

Compared with traditional PTs, global search replaces designing various assigning operators in classical PTs, and the evaluation criteria are embedded into an MOEPT to provide important guidance for the searching procedure. Specifically, masses of singletons

are randomly generated in an evolutionary-based framework, which needs to satisfy with the basic constraints for probability distributions in evidence reasoning. Additionally, an assessment factor is presented to assess the best individual in all populations by a special objective function (desirable evaluation criteria). The simulation results in 4D FoD test cases show that in these problems, the proposed MOEPT was able to outperform other PTs from the perspective of the 2D criteria. Moreover, we propose a simple constraint-handling strategy within the MOEPT that is well-suited for two-target type tracking (2-TTT) problems, which to some extent encourages the application of MOEPTs to more complex and real-world decision-making problems.

The remainder of this paper is structured as follows. In Section 2, we briefly summarize the basis of DST. The new aggregation measure is proposed in Section 3. In Section 4, multi-objective evolutionary algorithms (EAs) based on a two-dimensional objective function are proposed. In Section 5, several examples and comprehensive comparisons are carried out. A simple pattern recognition problem and also a target type tracking problem are presented and solved in detail at the end of this section. The conclusions are drawn in Section 6.

## 2. Basis of Belief Functions

In this section, we introduce the belief function terminology of DST and the notations used in the sequel to this paper.

### 2.1. DST Basis

In DST [2], the elements  $\theta_i$  ( $i = 1, \dots, N$ ) of the frame of discernment (FoD)  $\Theta \triangleq \{\theta_1, \dots, \theta_N\}$  must be mutually exhaustive and exclusive. The power set of the FoD is denoted as  $2^\Theta$ , and a basic belief assignment (BBA), also called a mass function, is defined by the mapping:  $2^\Theta \rightarrow [0, 1]$ , which satisfies  $m(\emptyset) = 0$  and

$$\sum_{A \subseteq 2^\Theta} m(A) = 1 \tag{1}$$

where  $m(A)$  is defined as the BBA of  $A$ . The element  $A$  is called a focal element of  $m(\cdot)$  if  $m(A) > 0$ . The belief and plausibility functions, which are in one-to-one mapping with the BBA  $m(\cdot)$ , are defined for all  $A \subseteq \Theta$  by

$$Bel(A) = \sum_{B \in 2^\Theta | B \subseteq A} m(B) \tag{2}$$

$$Pl(A) = 1 - Bel(\bar{A}) = \sum_{A, B \in 2^\Theta | A \cap B \neq \emptyset} m(B) \tag{3}$$

where  $\bar{A} \triangleq \Theta \setminus A$  is the complement to  $A$  in  $\Theta$ . The belief interval  $[Bel(A), Pl(A)]$  represents the uncertainty committed to  $A$ , and the bounds of this interval are usually interpreted as lower and upper bounds of the unknown (possibly subjective) probability of  $A$ .

In order to fuse  $n$  bodies Of evidence (BOEs), Dempster’s rule of combination is usually used in the DST framework. The combination of  $n$  distinct BOEs is achieved as follows:

$$m(A) = \begin{cases} 0, & \text{if } A = \emptyset \\ \frac{\sum_{\cap A_i = A} \prod_{1 \leq i \leq n} m_i(A_i)}{\sum_{\cap A_i \neq \emptyset} \prod_{1 \leq i \leq n} m_i(A_i)}, & \text{if } A \neq \emptyset \end{cases} \tag{4}$$

### 2.2. Classical Probabilistic Transformations

The efficiency of a probabilistic transformation (PT) in the field of decision making was analyzed in depth by Smets [6]. Various PTs have been proposed in the open literature such as BetP [6,26], CuzzP [12], DSmp [9], PrBP1 and PrBP2 [27], as well as Cobb and Shenoy’s normalization of plausibility [10]. The simple and classical transformation (BetP) is briefly recalled in this subsection.

**Pignistic Transformation (BetP)**

Smets in [6,26] first proposed pignistic (also called *betting*) probability to make decisions, which aims to transfer the mass of belief of each non-specific element onto the singletons. The classical pignistic probability is defined as  $BetP(\emptyset) = 0$ , and  $\forall A \in 2^\Theta \setminus \{\emptyset\}$ :

$$BetP(\theta_i) \triangleq \sum_{A \subseteq 2^\Theta, A \neq \emptyset} \frac{|\theta_i \cap A|}{|A|} \frac{m(A)}{1 - m(\emptyset)} \tag{5}$$

Because in Shafer’s framework  $m(\emptyset) = 0$ , Equation (5) can simply be rewritten for any singleton  $\theta_i \in \Theta$  as

$$BetP(\theta_i) = \sum_{B \in 2^\Theta, \theta_i \subseteq B} \frac{1}{|B|} m(B) = m(\theta_i) + \sum_{B \in 2^\Theta, \theta_i \subset B} \frac{1}{|B|} m(B) \tag{6}$$

**2.3. Distance Proposed by Han and Dezert  $d_{BI}^E$**

The Jousselme’s distance, which was widely denoted as  $D_J$  in [15], was applied in many recent works [19,28], but the particular choice of the  $D_J$  distance is not a very good choice because one knows that the  $D_J$  distance has bad behavior. This was clearly explained recently in [17,24]. Assuming that two independent BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  are defined on  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , for each focal element  $\theta_i \in \Theta$  ( $i = 1, 2, \dots, 2^n - 1$ ), belief intervals of  $\theta_i$  for  $m_1(\cdot)$  and  $m_2(\cdot)$  can be calculated, which are denoted by  $[Bel_1(\theta_i), Pl_1(\theta_i)]$  and  $[Bel_2(\theta_i), Pl_2(\theta_i)]$ , respectively. The strict distance between the interval numbers  $[a, b]$  and  $[c, d]$  is defined by [29]

$$d^I([a, b], [c, d]) = \sqrt{\left[\frac{a+b}{2} - \frac{c+d}{2}\right]^2 + \frac{1}{3}\left[\frac{b-a}{2} - \frac{d-c}{2}\right]^2} \tag{7}$$

Therefore, we can calculate the distance between  $BI_1(\theta_i) : [Bel_1(\theta_i), Pl_1(\theta_i)]$  and  $BI_2(\theta_i) : [Bel_2(\theta_i), Pl_2(\theta_i)]$  according to Equation (7). Thus, we can obtain a total of  $2^N - 1$  belief interval distance values for all  $\theta_i \in \Theta$ . Aside from that, the Euclidean family belief interval-based distance  $d_{BI}^E$  can be rewritten as

$$d_{BI}^E(m_1, m_2) = \sqrt{N_c \cdot \sum_{i=1}^{2^n-1} [d^I(BI_1(\theta_i), BI_2(\theta_i))]^2} \tag{8}$$

Here,  $N_c = 1/2^n - 1$  is the normalization factor. In this paper, we regard  $d_{BI}^E$  as one criterion for evaluating the degree of similarity (similarity representing the degree of difference between the original BBAs and the transformed ones in [30]) between the original BBAs and the transformed ones.

**3. New Evidence for Similarity Characterization**

As mentioned in previous section, those distances (i.e., Jousselme’s distance [28]) and other metrics such as the PIC [31] or entropy [27] were widely applied to measure the degree of “similarity or dissimilarity” between BBAs. However, only the corresponding focal elements (or the relevant focal element set) between two sources of evidence are described or characterized. This one-sided view does not consider the order of size of the assignment of each focal element in the evidence, which might lead to “self-conflict or self-contradiction”. To consider such “information” produced by the evidence itself, here, a new evidence similarity measure is defined between two evidential sources according to the order of size of the assignment. Prior to this, to give this new similarity measure, first we define the order correlation coefficient between two sets of data:

**Definition 1.** [32] Given two sets of data  $\{x_1, x_2, \dots, x_n\}$  and  $\{y_1, y_2, \dots, y_n\}$ , here,  $x_1, x_2, \dots, x_n$  and  $y_1, y_2, \dots, y_n$  are in an ascending order. After sorting, two sets of data  $x_{p_1}, x_{p_2}, \dots, x_{p_n}$  and

$y_{q_1}, y_{q_2}, \dots, y_{q_n}$  meet  $x_{p_1} \leq x_{p_2} \leq \dots \leq x_{p_n}$  and  $y_{q_1} \leq y_{q_2} \leq \dots \leq y_{q_n}$ , respectively, for each  $p_i$  and index their positions from  $q_1, q_2, \dots, q_n$ , assuming it is  $q_j$ ; that is,  $q_j = p_i$ . Note that  $j = f(i)$ , and the correlation coefficient is

$$\mu = \frac{\sum_{i=1}^n (i - j)^2}{\sum_{i=1}^n [n - (i - 1) - i]^2} \tag{9}$$

This satisfies  $0 \leq \mu \leq 1$ . When  $\mu = 0$ , the convergence of two sets of data is the largest. When  $\mu = 1$ , this is reversed.

### 3.1. The Consistency of Focal Elements between Two BOEs

**Definition 2.** [32] For any two sources of evidence (i.e.,  $S_1$  and  $S_2$ ),  $m_1(\cdot)$  and  $m_2(\cdot)$  are the basic belief assignments over the discernment framework  $\Theta$  of a size  $n$ . The number of focal elements and the focal elements of  $m_1(\cdot)$  and  $m_2(\cdot)$  can be different. We denote  $X_i$  and  $Y_i$  as the indexes of the focal elements whose masses are sorted by increasing order. The similarity function of the evidence to characterize the order of the size of the assignments over the subsets is as follows:

$$Sim_{seq}(m_1, m_2) = 1 - \frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n [n + 1 - 2i]^2} \tag{10}$$

As we all know, if there is a similarity function  $Sim(m_i, m_j)$ , which is the characterization of distance between any two evidence sources, then the following four basic conditions must be satisfied:

- Symmetry:  $\forall m_i(\cdot), m_j(\cdot), Sim(m_i, m_j) = Sim(m_j, m_i)$ ;
- Consistency:  $\forall m(\cdot), Sim(m, m) = 1$ ;
- Nonnegative:  $\forall m_i(\cdot), m_j(\cdot), 0 \leq Sim(m_i, m_j) \leq 1$ ;
- Triangle inequality:  $Sim(X, Y) + Sim(Y, Z) \geq Sim(X, Z)$ .

According to previous work [32], it is easy to find that  $Sim_{seq}$  satisfies symmetry, consistency and nonnegativity, but the last important condition is lost. Therefore, we prove the triangle inequality property of  $Sim_{seq}$  here:

**Proof.** Based on Equation (10), the triangle inequality can be rewritten as follows:  $1 - \frac{\sum_{i=1}^n (X_i - Z_i)^2}{\sum_{i=1}^n [n + 1 - 2i]^2} \leq 1 - \frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n [n + 1 - 2i]^2} + 1 - \frac{\sum_{i=1}^n (Y_i - Z_i)^2}{\sum_{i=1}^n [n + 1 - 2i]^2} \Rightarrow \sum_{i=1}^n [n + 1 - 2i]^2 \geq \frac{\sum_{i=1}^n (X_i - Y_i)^2 + \sum_{i=1}^n (Y_i - Z_i)^2 - \sum_{i=1}^n (X_i - Z_i)^2}{\sum_{i=1}^n [n + 1 - 2i]^2}$ .

Here,  $X, Y$  and  $Z$  denote vectors  $\Rightarrow \sum_{i=1}^n [n + 1 - 2i]^2 \geq (X - Y)^T(X - Y) + (Y - Z)^T(Y - Z) - (X - Z)^T(X - Z)$ .

According to the squared sum formula (SSF) (squared sum formula of a natural number:  $1^2 + 2^2 + 3^2 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}$ ; squared sum formula of an odd number:  $1^2 + 3^2 + 5^2 + \dots + (2n - 1)^2 = \frac{1}{3}n(4n^2 - 1)$ ; squared sum formula of an even number:  $2^2 + 4^2 + \dots + (2n)^2 = \frac{2}{3}n(n + 1)(2n + 1)$ ), we have

$$\begin{aligned} \sum_{i=1}^n [n + 1 - 2i]^2 &= \sum_{i=1}^n [(n + 1)^2 + 4i^2 - 4(n + 1)i] = \\ &= n(n + 1)^2 + 4 * \frac{1}{6}n(n + 1)(2n + 1) - 4(n + 1) \frac{n(n + 1)}{2} \\ &= \frac{2}{3}n(n + 1)(2n + 1) - n(n + 1)^2 \\ &= n(n + 1) \cdot [\frac{4}{3}n + \frac{2}{3} - n - 1] \\ &= \frac{1}{3} \cdot n \cdot (n^2 - 1) \end{aligned}$$

Because  $1 \leq X_i \leq n, 1 \leq Y_i \leq n, 1 \leq Z_i \leq n$ , we obtain the following:

$$(X - Y)^T(X - Y) + (Y - Z)^T(Y - Z) - (X - Z)^T(X - Z) \leq 1 + (k - 1)^2 + (k - 3)^2 + \dots$$

- When  $n - 1 = 2k \Rightarrow k = \frac{n-1}{2}, 1 + (k - 1)^2 + (k - 3)^2 + \dots = \frac{1}{6} \cdot k(k + 1)(k + 2)$ , and thus

$$\begin{aligned} & \sum_{i=1}^n [n + 1 - 2i]^2 - [(X - Y)^T(X - Y) + (Y - Z)^T(Y - Z) - (X - Z)^T(X - Z)] \\ & \geq \frac{1}{3} \cdot n \cdot (n^2 - 1) - \frac{1}{6} \cdot k(k + 1)(k + 2) \\ & = \frac{1}{3} \cdot n \cdot (n^2 - 1) - \frac{1}{6} \cdot \frac{n - 1}{2} \left(\frac{n - 1}{2} + 1\right) \left(\frac{n - 1}{2} + 2\right) \\ & = \frac{1}{3} \cdot n \cdot (n^2 - 1) - \frac{1}{6} \cdot \frac{n - 1}{2} \cdot \frac{n + 1}{2} \cdot \frac{n + 3}{2} \\ & = (n^2 - 1) \cdot \left(\frac{1}{4}n + \frac{1}{6}\right) \end{aligned}$$

Because  $n \geq 2$ , thus

$$\sum_{i=1}^n [n + 1 - 2i]^2 - [(X - Y)^T(X - Y) + (Y - Z)^T(Y - Z) - (X - Z)^T(X - Z)] \geq 0 \Rightarrow Sim_{seq} \text{ satisfy triangle inequality;}$$

- When  $n - 1 = 2k - 1 \Rightarrow k = \frac{n}{2}, 1 + (k - 1)^2 + (k - 3)^2 + \dots = \frac{1}{3} \cdot k \cdot (4k^2 - 1)$ , thus

$$\begin{aligned} & \sum_{i=1}^n [n + 1 - 2i]^2 - [(X - Y)^T(X - Y) + (Y - Z)^T(Y - Z) - (X - Z)^T(X - Z)] \\ & \geq \frac{1}{3} \cdot n \cdot (n^2 - 1) - \frac{1}{3} \cdot k \cdot (4k^2 - 1) \\ & = \frac{1}{3} \cdot n \cdot (n^2 - 1) - \frac{1}{3} \cdot \frac{n}{2} \cdot \left(4\left(\frac{n}{2}\right)^2 - 1\right) \\ & = \frac{1}{3} \cdot n \cdot (n^2 - 1) - \frac{n}{6} \cdot (n^2 - 1) \\ & = \frac{n}{6} \cdot (n^2 - 1) \end{aligned}$$

Because  $n \geq 2$ , thus  $\sum_{i=1}^n [n + 1 - 2i]^2 - [(X - Y)^T(X - Y) + (Y - Z)^T(Y - Z) - (X - Z)^T(X - Z)] \geq 0 \Rightarrow Sim_{seq}$  satisfies the triangle inequality.

□

**Definition 3.** [32] For any two sources of evidence (i.e.,  $S_1$  and  $S_2$ ),  $m_1(\cdot)$  and  $m_2(\cdot)$  are the basic belief assignments over  $n$  focal elements in the discernment framework  $\Theta$  (note that the BBA of each subproposition might be same). Assume that the  $s_1$  subpropositions' BBAs are the same in  $m_1(\cdot)$ , and the  $s_2$  subpropositions' BBAs are the same in  $m_2(\cdot)$ . Herein,  $X_i$  and  $Y_i$  are the serial numbers according to the order of the size of the subpropositions' BBAs, where the subscript  $i$  indicates the  $i$ th subproposition, due to the BBAs of some subpropositions being the same. For the evidence  $S_1$ , there might be  $s_1$  kinds of sorts. For  $S_2$ , there might be  $s_2$  kinds of sorts. Therefore, there are  $s_1 \times s_2$  kinds of sorts for  $S_1$  and  $S_2$ . The similarity measure functions are redefined in this case as follows:

$$Sim'_{seq}(m_1, m_2) = 1 - \frac{\sum_{t=1}^{s_1 s_2} \sum_{i=1}^n (X_i^t - Y_i^t)^2}{s_1 s_2 (\sum_{i=1}^n (n + 1 - 2i)^2)} \tag{11}$$

Similarly, it is easy to prove that  $Sim'_{seq}(m_X, m_Y)$  is a similarity measure function.

**Example 1. Bayesian BBAs:** Assuming two kinds of evidence  $m_1 = \{\theta_1, \theta_2, \theta_3\} = \{0, 0.1, 0.9\}$  and  $m_2 = \{\theta_1, \theta_2, \theta_3\} = \{0.9, 0.1, 0\}$ , then  $m_1$  and  $m_2$  are sorted from small to large so that

$X(m_1) = \{\theta_1, \theta_2, \theta_3\} = [1, 2, 3]$  and  $Y(m_2) = \{\theta_3, \theta_2, \theta_1\} = [3, 2, 1]$ , respectively. Thus, we can calculate the similarity measure based on Equation (10):

$$Sim_{seq}(m_1, m_2) = 1 - \frac{(1 - 3)^2 + (2 - 2)^2 + (3 - 1)^2}{(3 + 1 - 2 * 1)^2 + (3 + 1 - 2 * 2)^2 + (3 + 1 - 2 * 3)^2} = 0$$

According to  $Sim_{seq}(m_1, m_2)$ , we can find that  $m_1$  and  $m_2$  are completely different and lacking similarity.

### 3.2. The Inconsistency of the Focal Elements between Two BOEs

How do we calculate  $Sim_{seq}$  when the focal elements in BBAs are different? Let us consider the following example and put forward a different way compared with that in [32]:

**Example 2. General BBAs:** Assuming two types of evidence  $m_1 = \{\theta_1, \theta_2 \cup \theta_3, \theta_2 \cup \theta_4, \theta_1 \cup \theta_3 \cup \theta_4\} = \{0.3, 0.2, 0.2, 0.3\}$  and  $m_2 = \{\theta_1, \theta_1 \cup \theta_3, \theta_3, \theta_2 \cup \theta_3 \cup \theta_4, \theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4\} = \{0.4, 0.1, 0.1, 0.2, 0.2\}$ .

Borrowing ideas from Dezert ( $d_{BI}^E$ ) in [24], for each focal element  $\theta_i \in \Theta (i = 1, 2, \dots, 2^n - 1)$ , the belief intervals of  $\theta_i$  for  $m_1(\cdot)$  and  $m_2(\cdot)$  can be calculated, which are denoted by  $[Bel_1(\theta_i), Pl_1(\theta_i)]$  and  $[Bel_2(\theta_i), Pl_2(\theta_i)]$ , respectively. According to the theory of evidence, the width of such an interval as  $[Bel_1(\theta_i), Pl_1(\theta_i)]$  represents the degree of uncertainty for the corresponding focal element  $\theta_i$ . Therefore,  $X_i$  and  $Y_i$  in Equation (10) are obtained, which refer to the index of the width of the interval for each focal element, whose values are sorted in increasing order. The steps of this mechanism are illustrated as follows:

- Step 1:  $[Bel_1(\theta_1), Pl_1(\theta_1)] = [0.3, 0.6]$ ,  $[Bel_1(\theta_2), Pl_1(\theta_2)] = [0, 0.4]$ ,  $[Bel_1(\theta_3), Pl_1(\theta_3)] = [0, 0.5]$ ,  $[Bel_1(\theta_4), Pl_1(\theta_4)] = [0, 0.5]$ ,  $[Bel_2(\theta_1), Pl_2(\theta_1)] = [0.4, 0.7]$ ,  $[Bel_2(\theta_2), Pl_2(\theta_2)] = [0, 0.4]$ ,  $[Bel_2(\theta_3), Pl_2(\theta_3)] = [0.1, 0.6]$ ,  $[Bel_2(\theta_4), Pl_2(\theta_4)] = [0, 0.4]$ ;
- Step 2: The parameter  $\zeta$  denotes the width of the belief interval, where  $\zeta_1(\theta_1) = Pl_1(\theta_1) - Bel_1(\theta_1) = 0.3$ ,  $\zeta_1(\theta_2) = 0.4$ ,  $\zeta_1(\theta_3) = 0.5$ ,  $\zeta_1(\theta_4) = 0.5$ ,  $\zeta_2(\theta_1) = 0.3$ ,  $\zeta_2(\theta_2) = 0.4$ ,  $\zeta_2(\theta_3) = 0.5$  and  $\zeta_2(\theta_4) = 0.4$ ;
- Step 3:  $X_1$  and  $Y_1$  are the indexes of focal elements whose  $\zeta$  values are sorted in increasing order, where  $X_1 = \{1, 2, 3, 3\}$  and  $Y_1 = \{1, 2, 3, 2\}$ ;
- Step 4:  $Sim_{seq}$  is calculated based on Equation (10).

To consider the influence of the distance of the evidence, here, based on  $d_{BI}^E$  in Equation (8), we propose a new similarity measure which is presented as follows:

$$Cim(m_i, m_j) = w_1 \cdot d_{BI}^E(m_i, m_j) + w_2 \cdot F(Sim_{seq}(m_i, m_j)) \tag{12}$$

where  $w_1 = w_2 = 0.5$ . and  $F(\cdot)$  is the decreasing function within the interval  $[0, 1]$ , for which in this paper  $F(\cdot) = 1 - x^2$ . That aside, it is easy to prove that the improved measure function  $Cim(m_i, m_j)$  is still a similarity measure function. This is because the final metric between two similarity measure functions still meets the definition of a similarity measure function.

Additionally, to consider the normalization of Equation (12), it can be rewritten as follows:

$$C_{norm}(m_i, m_j) = w_1 * (d_{BI}^{E*}(m_i, m_j)) + w_2 * (F(Sim_{seq}^*(m_i, m_j))) \tag{13}$$

where  $d_{BI}^{E*}(\hat{m}) = \frac{d_{BI}^E(m_i, m_j) - \min(\mathbf{vector}_1)}{\max(\mathbf{vector}_1) - \min(\mathbf{vector}_1)}$ ,  $Sim_{seq}^*(m_i, m_j) = \frac{Sim_{seq}(m_i, m_j) - \min(\mathbf{vector}_2)}{\max(\mathbf{vector}_2) - \min(\mathbf{vector}_2)}$ ,  $\mathbf{vector}_1 = (D_1(m_i, m_1), D_2(m_i, m_2), \dots, D_j(m_i, m_j))$  and  $\mathbf{vector}_2 = (Sim_1(m_i, m_1), Sim_2(m_i, m_2), \dots, Sim_j(m_i, m_j))$ .

#### 4. Multi-Objective Evolutionary Algorithm Based on Two-Dimensional Criteria

In this section, we regard a PT as a general multi-objective problem consisting of two objectives also involved in a number of inequality and equality constraints. Then, a corresponding optimization model is proposed for selecting the *best* Bayesian BBA in the set of candidates.

##### 4.1. Multiple-Objective Evolutionary-Based Probabilistic Transformation

The idea to approximate any BBA into a Bayesian BBA (i.e., a subjective probability measure) using the minimization of the Shannon entropy under compatibility constraints was proposed recently by Han et al. in [13,19] using *on-the-shelf* optimization techniques. In this paper, we present in detail a new optimization method to achieve this PT based on a random evolutionary algorithm to acquire minimization of the new aggregation criteria, and this new comprehensive criteria represents different aspects of information in BBAs. For example, the *conflict coefficient* represents the degree of similarity in *conflict* between transformed BBAs and original BBAs (in other words, the more conflicts that exist between two BBAs, the less similarity they have). In addition,  $d_{B1}^E$  represents the interval distance between the original BBAs and the transformed ones.

Let us assume that the FoD of the original BBA  $m(\cdot)$  approximated by a Bayesian BBA is  $\Theta \triangleq \{\theta_1, \theta_2, \dots, \theta_N\}$ . The MOEPT method consists of the following steps, which are derived from GAs:

- Step 0 (setting parameters): Assume  $t_{\max}$  is the maximum number of iterations,  $n_{\max}$  is the population size in each iteration,  $P_s$  is the selection probability,  $P_c$  is the crossover probability, and  $P_m$  is the mutation probability.
- Step 1 (population generation and encoding mechanism): A set  $\mathbf{P}_t$  of  $j = 1, 2, \dots, n_{\max}$  random probability values  $P_t^j = \{P^j(\theta_1), \dots, P^j(\theta_N)\}$  is generated such that the constraints in Equations (14)–(16) for  $j = 1, 2, \dots, n_{\max}$  are satisfied in order to make each random set of probabilities  $P_t^j$  compatible with the original or target BBA  $m(\cdot)$  to approximate. (The lower (Bel) and upper (Pl) limits of each focal element are calculated using Equations (2) and (3) based on the value of  $m(\cdot)$ .) In other words, we have

$$P^j(\theta_i) \in [0, 1], \quad i = 1, 2, \dots, N \tag{14}$$

$$\sum_{i=1}^N P^j(\theta_i) = 1 \tag{15}$$

$$Bel(\theta_i) \leq P^j(\theta_i) \leq Pl(\theta_i), \quad i = 1, 2, \dots, N \tag{16}$$

- Step 2 (fitness assignment): For each probability set  $P_t^j$ , ( $j = 1, 2, \dots, n_{\max}$ ), we compute its fitness value  $F$  based on Equation (13). More precisely, one takes  $F(P_t^j) = C_{norm}(m(\cdot), P_t^j)$ .
- Step 3 (best approximation of  $m(\cdot)$ ): The best probability set  $P_t^{j_{\text{best}}}$  with a minimum fitness value is sought, and its associated index  $j_{\text{best}}$  is stored in *Best Individual* and *Index of Best Individual*.
- Step 4 (selection, crossover and mutation): The tournament selection, crossover and mutation operators drawn from the evolutionary theory framework [33] are implemented to create the associated offspring population  $\mathbf{P}'_t$  based on the parent population  $\mathbf{P}_t$ . If  $F(P_t^{j_{\text{best}}}) \leq F(P'_t^{j_{\text{best}}})$ , then the *Best Individual* remains unchanged; otherwise,  $\text{Best Individual} = P_t^{j_{\text{best}}}$ .
  - Crossover operator: The crossover operator is one of the most important operators in the genetic algorithm. The crossover operation is conducted for the selected pairs of individuals. The feasibility condition of each individual is described as follows. The value of each subsegment must be between 0 and 1, and the sum of the individuals should be 1. Although the initial population is

formed in a way that all individuals are feasible and correct, using the standard crossover operators leads to defective sub-segments, and a normalization procedure is needed for such a situation. Consider the following two individuals to be parents:  $X = (0.1, 0.2, 0.3, |0.4)$  and  $Y = (0.2, 0.2, 0.1, |0.5)$ . (Here, the vertical bar represents the intersection point with the crossover operator.) With the single-point classic crossover operator, the following offspring will be produced:  $X' = (0.1, 0.2, 0.3, 0.5)$  and  $Y' = (0.2, 0.2, 0.1, 0.4)$ , where  $\sum_{j=1}^4 X'_j$  is equal to 1.1, which is greater than 1, and  $\sum_{j=1}^4 Y'_j$  is equal to 0.9, which is less than 1. Therefore,  $X'$  and  $Y'$  have defective values for which a normalization factor is needed, which leads to the following:

$$X'' = \frac{X'}{\sum_{j=1}^4 X'_j} = (0.1/1.1, 0.2/1.1, 0.3/1.1, 0.5/1.1),$$

$$Y'' = \frac{Y'}{\sum_{j=1}^4 Y'_j} = (0.2/0.9, 0.2/0.9, 0.1/0.9, 0.4/0.9).$$

- Mutation operator: The mutation operator randomly alters the value of a sub-segment. After applying the mutation operator, normalization of the changed individuals is required. The normalization will be performed in a similar way to the crossover operator.
- Step 5 (stopping MOEPT): Steps 1–4 illustrate the  $t$ th iteration of the MOEPT method. If  $t \geq t_{\max}$ , then the MOEPT method is complete; otherwise, another iteration must be performed by taking  $t + 1 = t$  and going back to step 1.

The scheme of the MOEPT method is shown in Figure 1, and its pseudo-code is given in Algorithm 1.

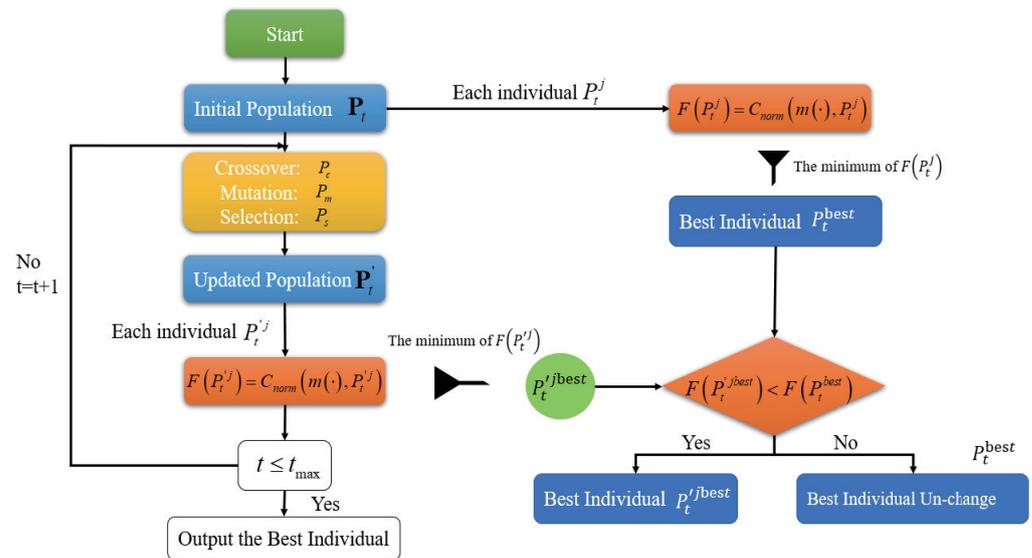


Figure 1. Scheme of MOEPT algorithm.

**Algorithm 1:** Multi-Objective Evolutionary-Based PT (MOEPT).

- 1: Define Stopping Criteria, ( $t \leq t_{\max}$ ); population Size  $n_{\max}$  for each iteration; crossover probability  $P_c$ , mutation probability  $P_m$  and selection probability  $P_s$ .
- 2: Generate an initial random population  $\mathbf{P}_t$  of consistent probabilities  $P_t^j$  with  $m(\cdot)$ .
- 3: For each individual  $P_t^j$  in  $\mathbf{P}_t$  do
- 4: Calculate Fitness  $F(P_t^j) = C_{norm}(m(\cdot), P_t^j)$  of  $P_t^j$
- 5: Store the best individual  $P_t^{j_{\text{best}}}$
- 6: End
- 7: Repeat:
- 8: Crossover: exchange parts of individuals with probability  $P_c$
- 9: Mutation: mutate the child individuals with probability  $P_m$
- 10: Selection: Select individuals based on fitness according to  $P_s$
- 11: After these three sub-steps, the updated population  $\mathbf{P}'_t$  is obtained
- 12: Calculate the fitness of individuals of  $\mathbf{P}'_t$ , and store the best individual  $P_t^{j_{\text{best}}}$
- 13: If  $F(P_t^{j_{\text{best}}}) \leq F(P_t^{j'_{\text{best}}})$
- 14: Best-Individual remains unchanged
- 15: else
- 16: Best-Individual =  $P_t^{j'_{\text{best}}}$
- 17: If  $t \geq t_{\max}$  then stops, otherwise  $t + 1 \rightarrow t$  and go back to line 7

**4.2. Convergence Analysis**

In order to mathematically prove the feasibility of an MOEPT, convergence analysis of our algorithm is given. First, we give a simplified description of the algorithm and also its symbolic representation for simplicity:

- Encoding mechanism: The size of the population is  $n_{\max}$ , the length of individual (chromosome) is  $N$ , and the initial population is  $\mathbf{P}_1$ ;
- Retain the best individual directly for the next generation;
- Randomly select the other non-optimal individuals in  $\mathbf{P}_t$  to cross over so as to form the intermediate population  $\mathbf{Y}_t$ ;
- The population  $\mathbf{Y}_t$  is mutated to form a population  $\mathbf{V}_t$ ;
- The better individuals in the population  $\mathbf{V}_t$  are selected as the new generation population  $\mathbf{P}_t$ .

Specifically, three operators (crossover operator, mutation operator and selection operator) can be described by the transition probability as follows:

- Crossover operator: For a single-point crossover, a new individual  $k$  is produced based on their parents: individuals  $i$  and  $j$ :

$$P_C^t(i \times j, k) = \begin{cases} |k|p_c/N, k \neq i, j \\ (1 - p_c) + |k|p_c/N, k = i \end{cases} \quad (17)$$

where  $|k|$  is the number of individuals  $k$ ,  $0 \leq p_c \leq 1$  is the crossover probability and  $a$  is the minimum probability for individuals  $|k|$ :

$$a = 1 - p_c + p_c/N. \quad (18)$$

- Mutation operator:

$$P_M^t(i, j) = p_m^{d(i,j)} (1 - p_m)^{N-d(i,j)} \tag{19}$$

where  $0 \leq p_m \leq 1$  is the mutation probability,  $d(i, j)$  is the Hamming distance between  $i$  and  $j$  and  $b$  is the minimum probability:

$$b = (1 - p_m)^N. \tag{20}$$

- Selection operator: An MOEPT uses the strategy of retaining the elite options, and the best individual is retained for the next generation which does not participate in the competition. Assume that  $m$  individuals are selected based on the following equations:

$$P_S^t(\mathbf{P}_t, P_t^j) = \frac{\sigma_n(F(P_t^j))}{\sum_{k=1}^{n_{max}} (F(P_t^k))}, j \in \mathbf{P}_t, n = 1, 2, \dots \tag{21}$$

where  $\sigma_n$  represents an increasing scale function. That aside, the probability of selecting the first individual in the next generation’s population is

$$P_S^*(\mathbf{P}_t, P_t^j) = \frac{|P_t^j|}{|B(\mathbf{P}_t)|}, P_t \in \mathbf{P}_t. \tag{22}$$

where  $|P_t^j|$  is the number of individuals  $P_t^j$  in  $\mathbf{P}_t$  and  $B(\mathbf{P}_t)$  is the cardinality of the optimal set of  $\mathbf{P}_t$ .

In order to facilitate the convergence analysis, the changing process of the fitness value  $F(P_t^j)$  is regarded as a Markov chain. If the MOEPT obtains the best individual  $P_t^{j_{best}}$  in generation  $t$ , we can denote this as  $\{\hat{F}(P_t)\} = P_t^{j_{best}}$ . Then, all the other populations in  $t + 1$  generations will also reach the best fitness value due to the elite strategy [34]. Therefore, the Markov chain  $\{\hat{F}(P_t)\}$  constitutes the lower martingale. According to the properties of the lower martingale and the convergence theorem of the lower martingale [35], convergence analysis of the MOEPT is converted into the convergence of  $\{\hat{F}(P_t)\}$ . The following three theorems are given, in which Theorem 4 is for proving that  $\{\hat{F}(P_t)\}$  satisfies the conditions of the martingale theorem, Theorem 5 proves the global convergence of the MOEPT and Theorem 6 constructs three conditions for the convergence of the lower martingale so that the optimal solution can almost be obtained everywhere.

**Theorem 1.** *The process of describing the values of the fitness functions in the MOEPT is a non-bounded martingale:*

$$E\{\hat{F}(\mathbf{P}_{t+1})/\mathbf{P}_t\} \geq \hat{F}(\mathbf{P}_t) \tag{23}$$

**Proof.** Because the algorithm retains the maximum fitness value of the previous generation for the next generation and does not participate in the genetic operation, the best individual mode is not destroyed, so the maximum fitness value of the next generation’s population will not be less than the maximum fitness value of the previous generation:

$$E\{\hat{F}(\mathbf{P}_{t+1})/\mathbf{P}_t\} \geq \hat{F}(\mathbf{P}_t) > 0 \tag{24}$$

□

**Theorem 2.** *The MOEPT converges to the global optimal solution based on the probability, which is mathematically expressed by the condition.*

**Proof.** When the population  $\mathbf{P}_t$  is updated to generation  $t$ , the minimum or best fitness is recorded as  $P_t^{j_{best}}$ , and the global optimal solution is noted as  $F^*$ , we assume that the MOEPT can converge to a global optimal solution at generation  $t$  such that

$$\{\hat{F}(\mathbf{P}_t)\} = F^* \tag{25}$$

□

Based on Theorem 4, we have

$$E\{\hat{F}(\mathbf{P}_{t+1})/\mathbf{P}_t\} = F^* \tag{26}$$

This is defined by the following conditional expectation:

$$\begin{aligned} E\{\hat{F}(\mathbf{P}_{t+1})/\mathbf{P}_t\} &= \\ &\sum_{i,j \in \mathbf{P}_t} P_C^t(i \times j, y) \sum_v P_M^t(y, v) \sum_k P_S^t(v, k) \hat{F}(k) \geq \\ &\sum_{i,j \in \mathbf{P}_t} P_C^t(i \times j, i) \sum_v P_M^t(y, v) \sum_k P_S^t(v, k) \hat{F}(k) \geq \\ &a \sum_v P_M^t(y, v) \sum_k P_S^t(v, k) \hat{F}(k) \geq \\ &a \sum_v P_M^t(y, y) \sum_k P_S^t(v, k) \hat{F}(k) \geq \\ &ab^m \left\{ \sum_{k \in B(\mathbf{P}_t)} [P_S^t(v, k) - P_S^{t*}(v, k)] \hat{F}(k) + \sum_{k \in B(\mathbf{P}_t)} P_S^{t*}(v, k) \hat{F}(k) \right\} \end{aligned}$$

When  $k \notin B(\mathbf{P}_t)$ ,  $P_S^{t*}(v, k) = 0$ , and when  $k \in B(\mathbf{P}_t)$ ,  $\hat{F}(k) = F^*$ ,  $E\{\hat{F}(\mathbf{P}_{t+1})/\mathbf{P}_t\}$  can be rewritten as

$$\begin{aligned} E\{\hat{F}(\mathbf{P}_{t+1})/\mathbf{P}_t\} &\geq \\ ab^m \sum_{k \in B(\mathbf{P}_t)} [P_S^t(v, k) \hat{F}(k) + F^*] &\geq ab^m F^*. \end{aligned}$$

Therefore, we obtain

$$ab^m F^* \leq F^*. \tag{27}$$

Because  $F^* > 0$ , one obtains

$$ab^m \leq 1. \tag{28}$$

Based on the above formula derivation, the MOEPT converges to the global optimal solution.

**Theorem 3.** When  $\forall n \geq 1$ , the following conditions are satisfied:

- $E[\hat{F}(\mathbf{P}_1)] < \infty, F^* < \infty$ ;
- $E[\hat{F}(\mathbf{P}_t)/\mathbf{P}_{t-1}] = \hat{F}(\mathbf{P}_{t-1}) + c_{t-1}F^*$ ;
- $c_t \in [0, 1], \lim_{t \rightarrow \infty} \sum_{k=0}^{t-1} c_k = 1 - \frac{\hat{F}(\mathbf{P}_1)}{F^*}$ .

Then, we have the random sequence  $\hat{F}(\mathbf{P}_t) \xrightarrow{a.s.} F^*$

Proof: By taking the mathematical expectation on both sides of condition (2), one has

$$\begin{aligned} E[\hat{F}(\mathbf{P}_t)] &= E[\hat{F}(\mathbf{P}_{t-1})] + c_{t-1}F^* = \\ E[\hat{F}(\mathbf{P}_{t-2})] &+ c_{t-1}F^* + c_{t-2}F^* = \dots = \\ E[\hat{F}(\mathbf{P}_1)] &+ F^* \sum_{k=0}^{t-1} c_k. \end{aligned}$$

According to conditions (1) and (3), we have

$$E[\hat{F}(\mathbf{P}_t)] < E[\hat{F}(\mathbf{P}_1)] + F^* < \infty$$

$$Sup_t E[\hat{F}(\mathbf{P}_t)] < sup_t E[\hat{F}(\mathbf{P}_1)] + sup_t F^* < \infty$$

Because  $\hat{F}(\mathbf{P}_t)$  is a non-bounded martingale, we have

$$\hat{F}(\mathbf{P}_t) \xrightarrow{a.s.} \hat{F}(\mathbf{P}_\infty) = \lim_{t \rightarrow \infty} \hat{F}(\mathbf{P}_t) \tag{29}$$

$$\lim_{t \rightarrow \infty} E[\hat{F}(\mathbf{P}_t)] =$$

$$\lim_{t \rightarrow \infty} E[\hat{F}(\mathbf{P}_1)] + F^* \lim_{t \rightarrow \infty} \sum_{k=0}^{t-1} c_k =$$

$$E[\hat{F}(\mathbf{P}_1)] + F^* (1 - \frac{\hat{F}(\mathbf{P}_1)}{F^*}) = F^*,$$

$$\hat{F}(\mathbf{P}_t) \xrightarrow{a.s.} F^*. \tag{30}$$

### 5. Simulation Results

According to the first step of the MOEPT, we initially set the related parameters as follows:  $t_{max} = 50, n_{max} = 1000, P_s = 0.3, P_c = 0.5$  and  $P_m = 0.1$ .

#### 5.1. Simple Examples

**Example 3.** Let us consider the frame  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$  and the corresponding BBA illustrated as follows:

$$m(\theta_1) = 0.16, m(\theta_2) = 0.14, m(\theta_3) = 0.01, m(\theta_4) = 0.02$$

$$m(\theta_1 \cup \theta_2) = 0.20, m(\theta_1 \cup \theta_3) = 0.09, m(\theta_1 \cup \theta_4) = 0.04$$

$$m(\theta_2 \cup \theta_3) = 0.04, m(\theta_2 \cup \theta_4) = 0.02, m(\theta_3 \cup \theta_4) = 0.01$$

$$m(\theta_1 \cup \theta_2 \cup \theta_3) = 0.10, m(\theta_1 \cup \theta_2 \cup \theta_4) = 0.03$$

$$m(\theta_1 \cup \theta_3 \cup \theta_4) = 0.03, m(\theta_2 \cup \theta_3 \cup \theta_4) = 0.03$$

$$m(\Theta) = 0.08$$

Based on the respective classical PTs, the original BBAs are transformed into their corresponding probabilities as illustrated in Table 1. Their corresponding  $C_{norm}$  values can be calculated using Equation (13), which is already listed in Table 1. Clearly, several interesting characteristics presented in Table 1 are worth mentioning: (1) MOEPT $_{d_{BI}^E + Sim_{seq}}$  had the minimum value from the perspective of the  $C_{norm}$  criteria, which consider both aspects of  $d_{BI}^E$  and  $Sim_{seq}$  rather than concentrating on a single aspect, and (2) compared with other PTs, especially MOEPT $_{D_J + Sim_{seq}}$  (here, to show the property of  $d_{BI}^E$ , we replaced  $d_{BI}^E$  with  $D_J$  in the MOEPT to make comparisons) our method performed better than the mentioned methods. However, in practice, the suitability of various PTs depends on a number of factors, including the designer’s choices; that is, from the perspective of sequence similarity,  $Sim_{seq}$  plays an important role in  $C_{norm}$ , but from the view of the whole distance, it involves transferring the principle role to  $d_{BI}^E$  or  $D_J$ . How does one quantify this role? Here, we depend on the parameters  $w_1$  and  $w_2$  in  $C_{norm}$  to distinguish our ideas from Han in [19], which initially set  $w_1$  and  $w_2$  as 0.5. Here, we discuss three different situations: (1).  $w_2$  is set to 0.8 so as to pay more attention to the similarity of the sequence, (2)  $w_1$  is set to 0.8 so as to focus more on the distance, and (3) considering both sequence similarity and distance,  $w_1 = w_2$  is set to 0.5, which is the same value used in [19], so that the similarity of both the sequence and distance are considered. This phenomenon, to some degree, reminds us of the importance of proper selections of weight in various applications when the MOEPT is applied. That aside, it is worth noting that  $C_{norm}$  turned out to be  $Sim_{seq}$  when  $w_1 = 0$ , and  $C_{norm}$  turned out to be  $d_{BI}^E$  when  $w_2 = 0$ .

**Table 1.** Results of different PTs in Example 3 ( $w_1 = w_2 = 0.5$ ).

	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$C_{norm}$
<b>CuzzP</b>	0.3860	0.3382	0.1607	0.1151	0.2800
<b>BetP</b>	0.3983	0.3433	0.1533	0.1050	0.2799
<b>DSmP<sub>0</sub></b>	0.5176	0.4051	0.0303	0.0470	0.1897
<b>DSmP<sub>0.001</sub></b>	0.5162	0.4043	0.0319	0.0477	0.1896
<b>PrBP1</b>	0.5419	0.3998	0.0243	0.0340	0.1918
<b>PrBP2</b>	0.5578	0.3842	0.0226	0.0353	0.1933
<b>MOEPT<sub>D<sub>J</sub>+Sim<sub>seq</sub></sub></b>	0.3980	0.3322	0.1156	0.1541	<b>0.1849</b>
<b>MOEPT<sub>d<sub>B1</sub><sup>E</sup>+Sim<sub>seq</sub></sub></b>	0.3985	0.3983	0.0623	0.1409	<b>0.0733</b>

**Example 4.** Let us consider another situation in the frame  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$  and the corresponding BBAs illustrated as follows:

$$\begin{aligned}
 m(\theta_1) &= 0.16, m(\theta_2) = 0.16, m(\theta_3) = 0.16, m(\theta_4) = 0.16 \\
 m(\theta_1 \cup \theta_2) &= 0.04, m(\theta_1 \cup \theta_3) = 0.04, m(\theta_1 \cup \theta_4) = 0.04 \\
 m(\theta_2 \cup \theta_3) &= 0.04, m(\theta_2 \cup \theta_4) = 0.04, m(\theta_3 \cup \theta_4) = 0.04 \\
 m(\theta_1 \cup \theta_2 \cup \theta_3) &= 0.03, m(\theta_1 \cup \theta_2 \cup \theta_4) = 0.03 \\
 m(\theta_1 \cup \theta_3 \cup \theta_4) &= 0.03, m(\theta_2 \cup \theta_3 \cup \theta_4) = 0.03
 \end{aligned}$$

In actuality, Example 4 is the extension of the case studied by Han in [13], which assumes a special scenario where no difference exists between  $m(\theta_1)$ ,  $m(\theta_2)$ ,  $m(\theta_3)$  and  $m(\theta_4)$  and where the traditional PTs become invalid and give unreasonable results, which can be seen in Table 2. The property of the original BBA where no difference exists between  $m(\theta_1)$ ,  $m(\theta_2)$ ,  $m(\theta_3)$  and  $m(\theta_4)$  was almost lost when classical PTs were applied. When a “sequence” is not considered in an MOEPT, which is denoted as  $MOEPT_{Distance}$ , the feature of equal mass in the original BBAs was also missing, as with other classical PTs. Fortunately, when information of the “sequence” was added into the objective function, the MOEPT performed better in keeping the original information, as expected.

**Table 2.** Results of different PTs in Example 4.

	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$C_{norm}$
<i>BetP</i>	0.3983	0.3433	0.1533	0.1050	0.3974
<i>DSmP<sub>0</sub></i>	0.2500	0.2500	0.2500	0.2500	0.5458
<i>PrBP1</i>	0.5419	0.3998	0.0243	0.0340	0.6368
<i>PrBP2</i>	0.5578	0.3842	0.0226	0.0353	0.6412
<i>MOEPT<sub>Distance</sub></i>	0.2500	0.1597	0.3578	0.2325	0.3415
<b>MOEPT<sub>D<sub>J</sub>+Sim<sub>seq</sub></sub></b>	0.2483	0.2485	0.2496	0.2536	0.1708
<b>MOEPT<sub>d<sub>B1</sub><sup>E</sup>+Sim<sub>seq</sub></sub></b>	0.2484	0.2488	0.2489	0.2539	<b>0.0450</b>

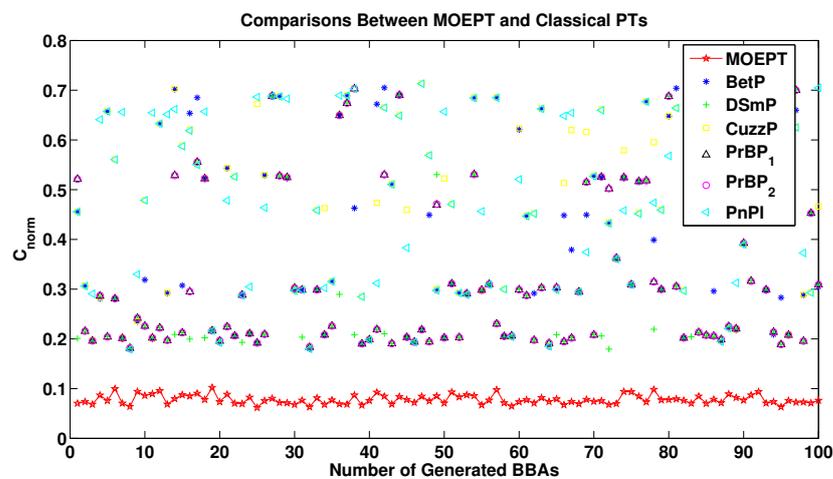
**Example 5.**  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$

To investigate the robustness of the MOEPT from a statistical point of view, in this example, we randomly generate BBAs and compare the MOEPT with classical PTs (BetP [6,26], CuzzP [12], DSmP [9], PrBP1 and PrBP2 [27]). The original BBAs for approximation are generated according to Algorithm 2 of [36].

**Algorithm 2:** Random generation of BBAs.

- 1: Input: Frame of Discernment  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$
- 2:  $N_{max}$ : Maximum number of focal element
- 3: Output: BBA-m
- 4: Generate  $K(\Theta)$ , which is the power set of  $\Theta$
- 5: Generate a random permutation of  $K(\Theta) \rightarrow R(\Theta)$
- 6: Generate an integer between 1 and  $N_{max} \rightarrow l$
- 7: For each First  $k$  elements of  $R(\Theta)$  do
- 8: Generate a value within  $[0, 1] \rightarrow m_i, i = 1, \dots, l$
- 9: End
- 10: Normalize the vector  $m = [m_1, m_2, \dots, m_l] \rightarrow m'$
- 11:  $m(\theta_i) = m'_i$

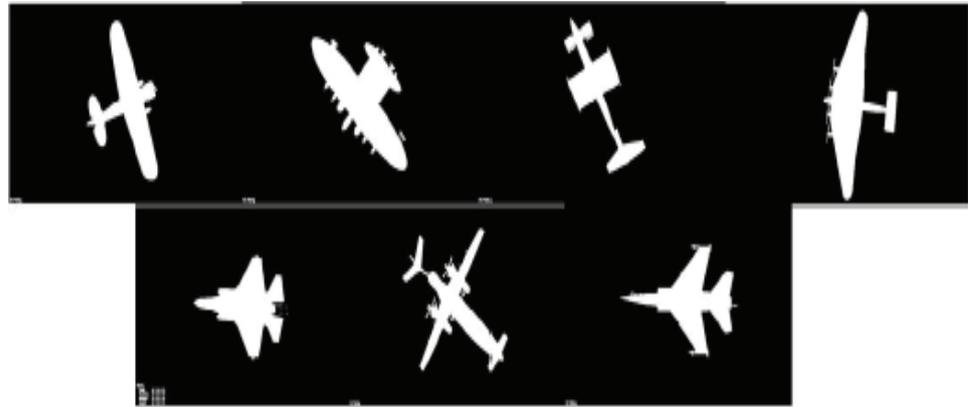
In our test, we set the cardinality of the FoD to 4 and fixed the number of focal elements to  $l = N_{max} = 15$ . We randomly generated  $L = 100$  BBAs. Six PT methods were tested, and  $C_{norm}$  was used to evaluate the quality of their corresponding results, shown in Figure 2. As we can see, and as was naturally expected, the MOEPT significantly outperformed the other methods based on the minimum  $C_{norm}$  criterion, which is absolutely normal because the method was developed for this aim.



**Figure 2.** Comparisons between MOEPT and state-of-the-art PTs.

### 5.2. Example of Pattern Classification Using the MOEPT

In this example, we used the evaluation of decision making under the evidence theory framework to indirectly evaluate the MOEPT. We considered seven classes of aircraft, which are illustrated in Figure 3, and the classifier used in this example was the probabilistic neural network (PNN). For each test example, the output of the classifier was represented by a BBA. The corresponding BBA for each test sample was generated according to Li's previous work [37].



**Figure 3.** The binary images of seven kinds of airplanes.

1. First, the image was preprocessed with binarization, and then multiple features were extracted, such as Hu moments, the normalized moment of inertia, affine invariant moments, discrete outline parameters and singular values. Secondly, five BBAs could be assigned to the evidence sources for each PNN. (Specifically, the transfer functions in five PNNs were set to a Gaussian function, the weighting function was set to the Euclidean distance; the input function was netprod, and the output function was compet.) Third, all five of these BBAs were fused by PCR6 [7] to form a single BBA  $m(\cdot)$ .

2. For the two classes  $t_1$  and  $t_2$  ( $t_1, t_2 \in 1, 2, 3, \dots, 7, t_1 \neq t_2$ ), with the top two values of  $m(i), i = 1, 2, 3, \dots, 7$ , the corresponding updated mass assignments were generated according to [38]:

$$m'(i) = m(i), \forall i = t_1, t_2 \quad (31)$$

The remaining mass was assigned to the total set  $\Theta$ :

$$m'(\Theta) = 1 - m'(t_1) - m'(t_2). \quad (32)$$

For example, for a test sample *target* – 1, we obtained the corresponding BBA from the PNNs, where  $m(1) = 0.7, m(2) = 0.05, m(3) = 0.2, m(4) = 0.01, m(5) = 0, m(6) = 0.02$  and  $m(7) = 0.02$ . The dominant class was class 1, and class 3 was in second place. The updated corresponding BBA was  $m'(1) = 0.7, m'(3) = 0.2$  and  $m'(2, 4, 5, 6, 7) = 0.1$ .

There were 100 samples for each class, with a total of 700 samples. For each class, 50 samples were randomly selected for training PNNs, and the remaining samples were used for testing. For the MOEPT, the decision result would be class  $t_{final}$  if

$$t_{final} = \operatorname{argmax}(MOEPT) \quad (33)$$

As we can see from Figure 4, the MOEPT performed well in this task of pattern classification.

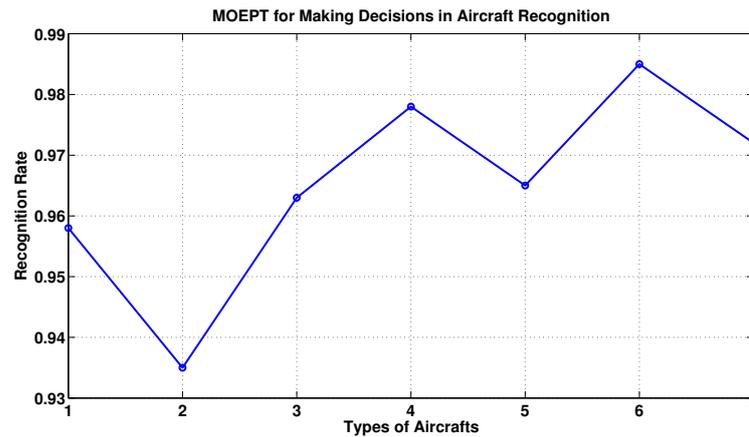


Figure 4. Recognition rate of MOEPT.

### 5.3. Example of Target Type Tracking Using the MOEPT

To further discuss the practicality of the proposed MOEPT, a target type tracking (TTT) problem in the area of decision making was used, which is briefly described below [39].

#### 5.3.1. Target Type Tracking Problem (TTT)

1. Consider  $\zeta = 1, 2, \dots, \zeta_{max}$  as the time index, and let there be  $N$  possible target types  $Tar_{\zeta} \in \Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$  in the surveillance area. For instance, in normal air target surveillance systems, the FoD could be  $\Theta = \{Fighter, Cargo\}$ ; that is,  $Tar_1 = \theta_1 \triangleq Fighter$ , and  $Tar_2 = \theta_2 \triangleq Cargo$ . Similarly, the FoD in a ground target surveillance system could be  $\Theta_{ground} = \{Tank, Truck, Car, Bus\}$ . In this paper, we just considered the air target surveillance systems to prove the practicability of EPT.

2. At every time  $\zeta$ , the true type of the target  $Tar(\zeta) \in \Theta$  was immediately observed by an attribute-sensor (here, we assumed a possible target probability).

3. A defined classifier was applied to process the attribute measurement of the sensor, which provided the probability  $Tar_d(\zeta)$  for the type of observed target at each instance  $\zeta$ .

4. The sensor was, in general, not totally reliable and was characterized by an  $N \times N$  confusion matrix:

$$\mathbf{M} = [M_{ij} = P(Tar_d = Tar_j | TrueType = Tar_i)] \tag{34}$$

where  $0 \leq i \leq N; 0 \leq j \leq N$ .

Here, we briefly summarize the main steps of the TTT using MOEPT:

1. Initialization: Determine the target type frame  $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$  and set the initial BBA  $m^{initial}(\theta_1 \cup \theta_2 \cup \dots \cup \theta_N) = 1$ , since there is no information about the first target type that will be observed;

2. Updating the BBA: An observed BBA  $m_{obs}(\cdot)$  on the types of unknown observed targets is defined from the current target type declaration and confusion matrix  $\mathbf{M}$ ;

3. Combination: We combine the current BBA  $m_{obs}(\cdot)$  with the initial BBA  $m^{initial}(\cdot)$  according to the PCR6 combination rule [7]:  $m_{PCR6}(\cdot) = m_{obs}(\cdot) \oplus m^{initial}(\cdot)$ ;

4. Approximation: Use  $MOEPT(\cdot)$  to approximate  $m_{PCR6}(\cdot)$  into a Bayesian BBA;

5. Decision making: Make a final decision about the type of the target at the current observation time based on the obtained Bayesian BBA;

6. Updating the BBA: Set  $m^{initial}(\cdot) = m_{PCR6}(\cdot)$ , and increase the time index  $\zeta = \zeta + 1$  before going back to step 2.

#### 5.3.2. Raw Dataset of TTT

We tested our MOEPT-based TTT on a very simple scenario for a 2D TTT problem, namely  $\Theta = \{Fighter, Cargo\}$ , for two types of classifiers. The matrix  $\mathbf{M}_1$  corresponds to

the confusion matrix of the good classifier, and  $\mathbf{M}_2$  corresponds to the confusion matrix of the poor classifier:

$$\mathbf{M}_1 = \begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{bmatrix}; \mathbf{M}_2 = \begin{bmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{bmatrix} \tag{35}$$

In our scenario, a true target type sequence over 120 scans was generated according to Figure 5. We can observe clearly from Figure 5 that Cargo (which is denoted as Type 2) appeared first in the sequence, and then the observation of the target type switched three times onto the Fighter type (Type 1) during different time durations (namely 20 s, 10 s and 5 s).

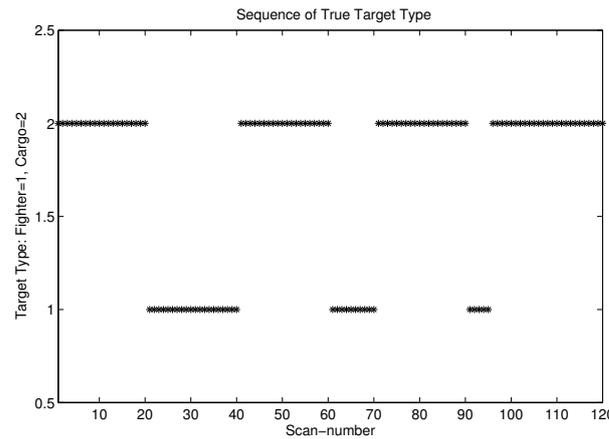


Figure 5. Raw sequence of true target type.

**A pathological case for TTT:** Our analysis showed that MOEPT can nevertheless be troublesome for tracking two target types, as proven in this particularly simple example (when  $0 \leq m(\theta_1 \cup \theta_2) \leq 0.1$ ). Let us consider the following BBA:

$$m_{target}(\cdot) = [\theta_1, \theta_2, \theta_1 \cup \theta_2] = [0, 1, 0]$$

According to the compatibility constraints in Equations (14)–(16), the population  $\mathbf{P}'_t$  was obtained from  $\mathbf{P}_t$  through a selection procedure. Next, an individual  $P'^j_t$  in  $\mathbf{P}'_t$ , which is denoted as  $P'^j_t = [m'(\theta_1), m'(\theta_2)]$ , was subjected to the initial constraint in Equations (1) and (36):

$$\begin{aligned} m'(\theta_1) &\geq (Bel(\theta_1) = m(\theta_1) = 0) \\ m'(\theta_1) &\leq (Pl(\theta_1) = m(\theta_1) + m(\theta_1 \cup \theta_2) = 0 + 0 = 0); \\ m'(\theta_2) &\geq (Bel(\theta_2) = m(\theta_2) = 1) \\ m'(\theta_2) &\leq (Pl(\theta_2) = m(\theta_2) + m(\theta_2 \cup \theta_1) = 1 + 0 = 1); \end{aligned} \tag{36}$$

From the above inequalities, one can see that only one probability measure,  $P^S_t = [m(\theta_1), m(\theta_2)] = [0, 1]$  (where the superscript index  $S$  means *single*), satisfied this constraint (the constraint was  $m(\theta_1) \in [Bel(\theta_1), Pl(\theta_1)] = [0, 0], m(\theta_2) \in [Bel(\theta_2), Pl(\theta_2)] = [1, 1]$ ). However because of the mechanism of MOEPT Equations (14)–(16), the  $P^j_t$  in population  $\mathbf{P}_t$ , which was randomly generated in the interval  $[Bel(\theta_i), Pl(\theta_i)], i = 1, 2, \dots, N$ , would be unable to generate enough candidates for evolutionary computation. (A sufficient number of candidate sets is a prerequisite for ensuring the global optimization performance of evolutionary algorithms.) That is why MOEPT becomes inefficient in this case, which occurs with a probability of  $1/n_{max}$ , where  $n_{max}$  denotes the size of the population  $\mathbf{P}_t$ . (In our simulation, we had  $n_{max} = 1000$ .) Unfortunately, in TTT decision-making problems, such a case cannot be avoided because it can really happen.

To circumvent this problem and make the MOEPT approach work in most circumstances, we needed modify the MOEPT method a bit to generate enough individuals for making the selection steps efficient when the bounds of the belief interval  $[Bel, Pl]$  took their minimum and maximum values ( $[0.9, 0.05, 0.05]$  and  $[0.05, 0.9, 0.05]$ , respectively). To achieve this, we proposed enlarging this particular interval through a parameter  $\lambda$  and maintaining the property of the original interval to some degree at the same time. More precisely, the modified belief interval, denoted as  $[Bel', Pl']$ , was heuristically computed by a simple thresholding technique as follows.

First, we assume that the original BBA we consider here for the FoD ( $\Theta = \{\theta_1, \theta_2\}$ ) is  $[\theta_1, \theta_2, \theta_1 \cup \theta_2] = [a, b, c]$ , with  $(a+b+c) = 1$  and  $0 \leq c \leq 0.1$ :

Step 1: Let  $m'(\theta_1 \cup \theta_2) = c + \lambda$ ;

Step 2: If  $a > b$ , then

$$m'(\theta_1) = a - \lambda; m'(\theta_2) = b; m'(\theta_1 \cup \theta_2) = c + \lambda; \tag{37}$$

Step 3: If  $a \leq b$ , then

$$m'(\theta_1) = a; m'(\theta_2) = b - \lambda; m'(\theta_1 \cup \theta_2) = c + \lambda; \tag{38}$$

Therefore, the values of  $[Bel'(\theta_1), Pl'(\theta_1)]$  and  $[Bel'(\theta_2), Pl'(\theta_2)]$  can be calculated based on Equations (37) and (38), which are presented as follows. When  $a > b$ , we have

$$\begin{cases} Pl'(\theta_1) = m(\theta_1) + m'(\theta_1 \cup \theta_2) = a - \lambda + c + \lambda = a + c; \\ Bel'(\theta_1) = 1 - Pl'(\bar{\theta}_1) = 1 - (b + c + \lambda) = a - \lambda. \end{cases} \tag{39}$$

$$\begin{cases} Pl'(\theta_2) = m(\theta_2) + m'(\theta_1 \cup \theta_2) = b + c + \lambda = b + c + \lambda; \\ Bel'(\theta_2) = 1 - Pl'(\bar{\theta}_2) \\ = 1 - (a - \lambda + c + \lambda) = 1 - (a + c) = b. \end{cases} \tag{40}$$

When  $a \leq b$ , we have

$$\begin{cases} Pl'(\theta_1) = m(\theta_1) + m'(\theta_1 \cup \theta_2) = a + c + \lambda; \\ Bel'(\theta_1) = 1 - Pl'(\bar{\theta}_1) \\ = 1 - (b - \lambda + c + \lambda) = 1 - (b + c) = a. \end{cases} \tag{41}$$

$$\begin{cases} Pl'(\theta_2) = m(\theta_2) + m'(\theta_1 \cup \theta_2) = b - \lambda + c + \lambda = b + c; \\ Bel'(\theta_2) = 1 - Pl'(\bar{\theta}_2) = 1 - (a + c + \lambda) = b - \lambda. \end{cases} \tag{42}$$

**Explanation:** Through step 1, one computes the total singleton mass one has in the entire BBA, and the threshold value of 0.9 allows one to evaluate if the percentage of the singleton mass is big enough or not. Here, we not only consider the unique extreme case  $m_{target}(\cdot) = [\theta_1, \theta_2, \theta_1 \cup \theta_2] = [0, 1, 0]$  but also other possible cases, such as  $m_{target}(\cdot) = [\theta_1, \theta_2, \theta_1 \cup \theta_2] = [0.0001, 0.9998, 0.0001]$ . Why do we consider the concept of this percentage? Actually, the higher the percentage of singleton mass, the smaller the interval for  $P_t^j$ ; in other words, the higher value of  $m(\theta_1 \cup \theta_2)$ , the bigger interval for  $P_t^j$ , which can be seen in Equation (36). Then, step 2 and step 3 give the method for calculating the updated upper bound of the belief interval  $[Bel', Pl']$ , and Equations (39)–(42) prove that the parameter  $\lambda$  determines the range of the interval. Next, we give two examples to show how the above method works.

**Pathological case one for TTT** (revisited with a modified MOEPT):

$$m_{target}(\cdot) = [\theta_1, \theta_2, \theta_1 \cup \theta_2] = [0.0001, 0.9998, 0.0001].$$

Here, the parameter  $\lambda$  is arbitrarily set to 0.4. (The value of the parameter  $\lambda$  can be chosen to be any value in  $[0, 1]$  by the designer for his or her own reasons to ensure the alternative interval effectively in the modified MOEPT.) Then, one computes in step 2 the modified plausibility bounds  $Bel'(\theta_1) = 0.0001$ ,  $Pl'(\theta_1) = 0.0001 + 0.0001 + \lambda = 0.4002$  and  $Bel'(\theta_2) = 0.9998 - 0.4 = 0.5998$ ,  $Pl'(\theta_2) = 0.9999$ . Therefore, we obtain  $[Bel'(\theta_1), Pl'(\theta_1)] = [0.0001, 0.4002]$  and  $[Bel'(\theta_2), Pl'(\theta_2)] = [0.5998, 0.9999]$ . The relationship between the original interval  $[Bel, Pl]$  and the updated interval  $[Bel', Pl']$  is illustrated in Figure 6.

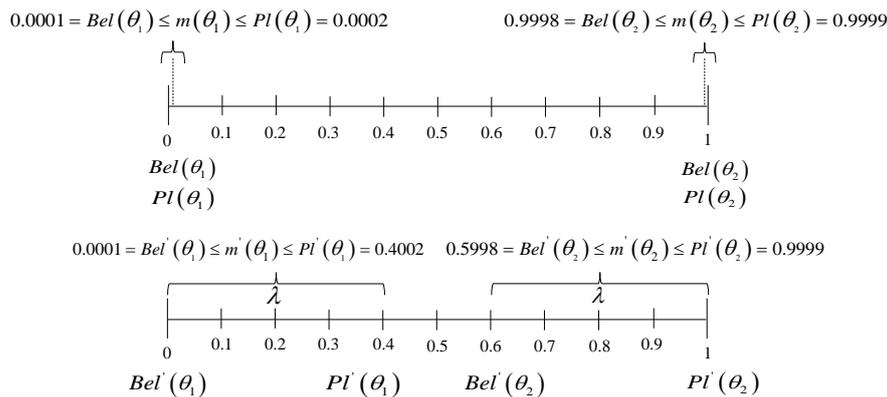


Figure 6. The principle of modified-constraint MOEPT ( $\lambda = 0.4$ ).

Consequently, any Bayesian BBA  $P_t^j = [m'(\theta_1), m'(\theta_2)]$  must be generated according the (modified) compatibility constraints:

$$m'(\theta_1) \in [Bel'(\theta_1), Pl'(\theta_1)] = [0.0001, 0.4002]$$

$$m'(\theta_2) \in [Bel'(\theta_2), Pl'(\theta_2)] = [0.5998, 0.9999]$$

**Pathological case two for TTT (revisited with a modified MOEPT):**

$$m_{target}(\cdot) = [\theta_1, \theta_2, \theta_1 \cup \theta_2] = [0.45, 0.48, 0.07].$$

Here, the parameter  $\lambda$  is set to 0.2. Then, any Bayesian BBA  $P_t^j = [m'(\theta_1), m'(\theta_2)]$  must be generated according the (modified) compatibility constraints:

$$m'(\theta_1) \in [Bel'(\theta_1), Pl'(\theta_1)] = [0.45, 0.72]$$

$$m'(\theta_2) \in [Bel'(\theta_2), Pl'(\theta_2)] = [0.28, 0.55]$$

In order to evaluate the influence of parameter  $\lambda$ , we reexamined all the pathological cases based on the following procedure:

1. The value of parameter  $\lambda$  was set to five possible values: 0, 0.1, 0.2, 0.3, 0.4 and 0.5;
2. We randomly generated an initial population  $P_t$  based on  $\lambda$ , which was also subject to the constraints in Equations (14)–(16).

With this simulation, we can observe in Figures 7 and 8 the impact of the  $\lambda$  value on the number of  $P_t^j$  in  $P_t$ . When we set  $\lambda = 0$  (in which the original MOEPT was applied), there existed no suitable  $P_t^j$  for case one, which demonstrates the necessity to circumvent the pathological case problem. Obviously, the number of  $P_t^j$  increased with the increase in the  $\lambda$  value, which efficiently proves the advantage of using the modified MOEPT approach to make the selection step of the evolutionary algorithm more efficient. One point we need to clarify is that the intervals (i.e.,  $[Bel'(\theta_1), Pl'(\theta_1)]$ ,  $[Bel'(\theta_2), Pl'(\theta_2)]$ ) induced

from the parameter  $\lambda$  above aim at guaranteeing enough of a number of  $P_t^j$  in  $P_t$  in the implementation of the MOEPT. Another point we also need to mention is that the number of  $P_t^j$  in  $P_t$  was not influenced by the weight. (Here, the weight equals  $w_2$  in Equation (13). And thus,  $w_1 = 1 - \text{Weight}$ , which to some degree guarantees the implementation of the MOEPT.)

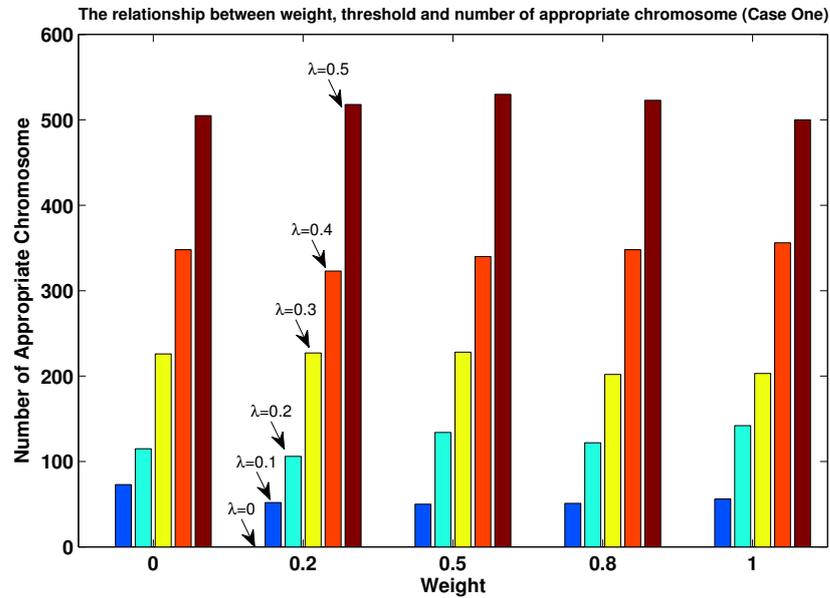


Figure 7. Impact of  $\lambda$  (x-axis) on individuals in  $P_t$  (y-axis) for case one.

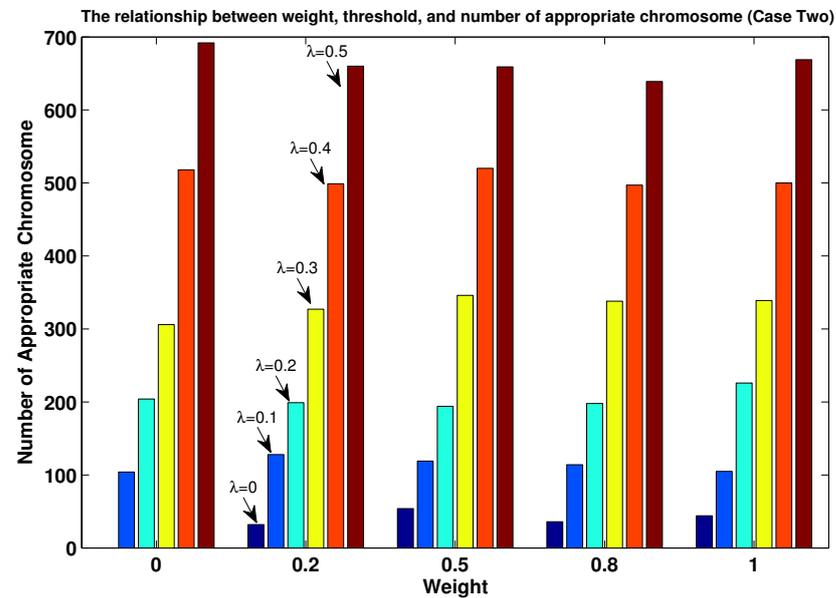


Figure 8. Impact of  $\lambda$  (x-axis) on individuals in  $P_t$  (y-axis) for case two.

### 5.3.3. Simulation Results of TTT Based on the Modified MOEPT

Our simulation consisted of 100 Monte Carlo runs, and we show in the sequel the averaged performances of the MOEPT. Figures 9 and 10 illustrate the Bayesian BBAs obtained by our new MOEPT method for solving the TTT problem using the PCR6 fusion rule. One can see that regardless of the good classifier  $M_1$  (recognition rate: 90.83 %) or poor classifier  $M_2$  (recognition rate: 80.83%) being used, the MOEPT was able to track properly the quick changes in target type.

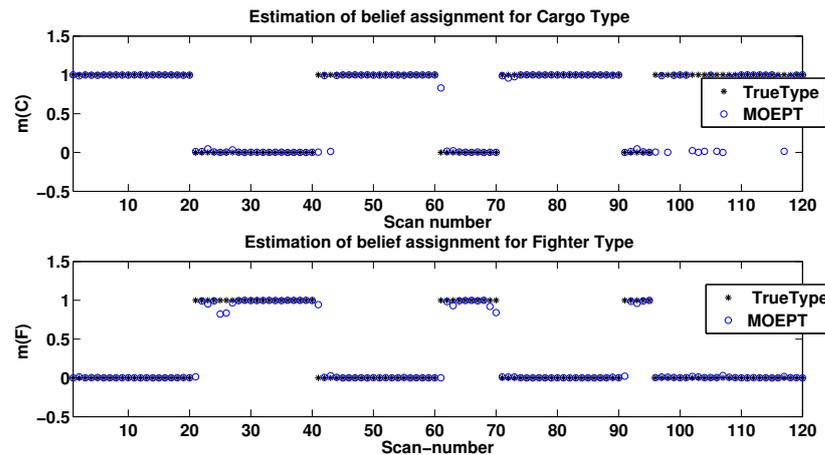


Figure 9. Results of MOEPT for Cargo and Fighter types using  $M_1$ .

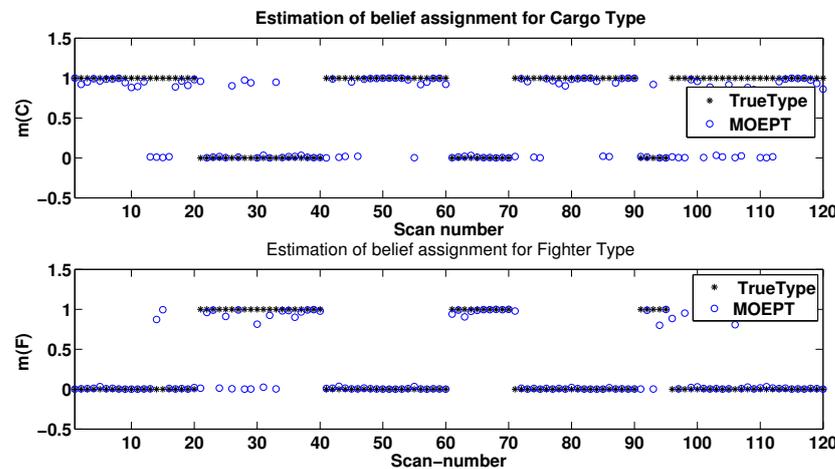


Figure 10. Results of MOEPT for Cargo and Fighter types using  $M_2$ .

### 6. Conclusions

A multi-objective evolutionary-based algorithm for probabilistic transformation (MOEPT) was proposed in this paper. It uses a genetic algorithm to obtain a Bayesian belief function and offer a comprehensive consideration concerning the closeness of distance between the original BBA and the Bayesian approximate one. In addition, a new aggregation measure was proposed in this paper to be combined into a more accurate “distance closeness” measure for MOEPT. More importantly, the convergence analysis of the MOEPT was given to prove the rationality of our proposed method. The effectiveness of the MOEPT was compared with respect to several probabilistic transformations proposed in the literature. Furthermore, the shortcomings of the original MOEPT version were clearly identified in two target type tracking problems, and they were solved thanks to modification of the belief interval constraints. As for future works, we would like to establish an adaptive scheme for the selection of weights in an MOEPT and make more comparisons between the performance of this MOEPT approach and other recently proposed evolutionary algorithms. We would also conduct more investigations to extend the MOEPT to DS<sub>m</sub>T using the DS<sub>m</sub> cardinal of elements. That aside, the current work in this paper is mainly for verifying the effectiveness of the algorithm through simulation examples from a theoretical perspective, and the feasibility of the proposed evolutionary-based PT will be verified through the practical and real problems in our future discussions.

**Author Contributions:** Conceptualization, Y.D. and L.C.; methodology, Y.D. and L.C.; software, Y.D. and L.C.; validation, Y.D. and L.C.; formal analysis, Y.D.; investigation, Y.D.; writing—original draft preparation, Y.D.; writing—review and editing, Y.D. and L.C.; visualization, Y.D. and K.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Natural Nature Science Fund, grant number 62103258.

**Institutional Review Board Statement:** Not applicable

**Informed Consent Statement:** Not applicable

**Data Availability Statement:** Not applicable.

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

## References

- Dempster, A.P. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Stat.* **1967**, *38*, 325–339. [\[CrossRef\]](#)
- Shafer, G. *A Mathematical Theory of Evidence*; Princeton University Press: Princeton, VA, USA, 1976.
- Ghaffari, M.; Ghadiri, N. Ambiguity-driven fuzzy C-means clustering: How to detect uncertainty clustered records. *Appl. Intell.* **2016**, *2*, 293–304. [\[CrossRef\]](#)
- Su, X.Y.; Mahadevan, S.; Han, W.H.; Deng, Y. Combining dependent bodies of evidence. *Appl. Intell.* **2015**, *3*, 634–644. [\[CrossRef\]](#)
- Dencœux, T. Inner and outer approximation of belief structures using a hierarchical clustering approach. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* **2001**, *9*, 437–460. [\[CrossRef\]](#)
- Smets, P. The Combination of Evidence in the Transferable Belief Model. *IEEE Trans. PAMI* **1990**, *5*, 29–39. [\[CrossRef\]](#)
- Smarandache, F.; Dezert, J. (Eds.) *Advances and Applications of DSMT for Information Fusion*; American Research Press: Rehoboth, NM, USA, 2015; pp. 1–4.
- Daniel, M. Probabilistic Transformations of Belief Functions. In Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU), Barcelona, Spain, 6–8 July 2005; pp. 539–551.
- Dezert, J.; Smarandache, F. A new probabilistic transformation of belief mass assignment. In Proceedings of the 11th International Conference on Information Fusion, Cologne, Germany, 30 June–3 July 2008; pp. 1–8.
- Cobb, B.R.; Shenoy, P.P. On the plausibility transformation method for translating belief function models to probability models. *Int. J. Approx. Reason.* **2006**, *41*, 314–330. [\[CrossRef\]](#)
- Sudano, J. Equivalence between belief theories and naive bayesian fusion for systems with independent evidential data: Part I, the theory. In Proceedings of the Fusion 2003, Cairns, Australia, 8–11 July 2003.
- Cuzzolin, F. The Intersection probability and its properties. In Proceedings of the ECSQARU Conference, Verona, Italy, 1–3 July 2009.
- Han, D.Q.; Dezert, J.; Han, C.; Yang, Y. Is entropy enough to evaluate the probability transformation approach of belief function? In Proceedings of the 13th Conference on Information Fusion, Edinburgh, UK, 26–29 July 2010; pp. 286–293.
- Bucci, D.J.; Acharya, S.; Pleskac, T.J.; Kam, T.J. Performance of probability transformations using simulated human opinions. In Proceedings of the 17th International Conference on Information Fusion, Salamanca, Spain, 7–10 July 2014; pp. 1–8.
- Yang, Y.; Han, D.Q. A new distance-based total uncertainty measure in the theory of belief function. *Knowl.-Based Syst.* **2016**, *94*, 114–123. [\[CrossRef\]](#)
- Liu, W.R. Analyzing the degree of conflict among belief functions. *Artif. Intell.* **2006**, *170*, 909–924. [\[CrossRef\]](#)
- Han, D.Q.; Dezert, J.; Yang, Y. New Distance Measures of Evidence based on Belief Intervals. In Proceedings of the International Conference on Belief Functions, Oxford, UK, 26–28 September 2014.
- Frikha, A. On the use of a multi-criteria approach for reliability estimation in belief function theory. *Inf. Fusion* **2014**, *18*, 20–32. [\[CrossRef\]](#)
- Han, D.Q.; Dezert, J.; Duan, Z.S. Evaluation of Probability Transformations of Belief Functions for Decision Making. *IEEE Trans. Syst. Man Cybern. Syst.* **2015**, *99*, 93–108. [\[CrossRef\]](#)
- Liu, Z.G.; Dezert, J.; Pan, Q.; Mercier, G. Combination of sources of evidence with different discounting factors based on a new dissimilarity measure. *Decis. Support Syst.* **2011**, *52*, 133–141. [\[CrossRef\]](#)
- Zhan, Deng, Z.; Wang, J.Y. A novel decision probability transformation method based on belief interval. *Knowl.-Based Syst.* **2020**, *208*, 427–450.
- Zhao, K.Y.; Chen, Z.Q.; Li, L.; Li, J.Y.; Sun, R.Z.; Yuan, G. DPT: An importance-based decision probability transformation method for uncertain belief in evidence theory. *Expert Syst. Appl.* **2022**, *11*, 120–133. [\[CrossRef\]](#)
- Ma, M.M.; An, J.Y. Combination of evidence with different weighting factors: A novel probabilistic-based dissimilarity measure approach. *J. Sens.* **2015**, *2015*, 509385. [\[CrossRef\]](#)
- Dezert, J.; Han, D.Q.; Tacnet, J.M.; Carladous, S.; Yang, Y. Decision-Making with Belief Interval Distance. In Proceedings of the International Conference on Belief Functions, Prague, Czech Republic, 21–23 September 2016.
- Han, D.Q.; Deng, Y.; Han, C.Z. Sequential weighted combination for unreliable evidence based on evidence variance. *Decis. Support Syst.* **2013**, *56*, 387–393. [\[CrossRef\]](#)

26. Smets, P. Decision making in the TBM: the necessity of the pignistic transformation. *Int. Approx. Reason.* **2005**, *38*, 133–147. [[CrossRef](#)]
27. Pan, W.; Hong, J.Y. New methods of transforming belief functions to pignistic probability functions in evidence theory. In Proceedings of the 2009 International Workshop on Intelligent Systems and Applications, Wuhan, China, 23–24 May 2009.
28. Jousselme, A.L.; Liu, C.S.; Grenier, D.; Bossé, E. Measuring Ambiguity in the Evidence Theory. *IEEE Trans. Syst. Man Cybern. Part Syst. Hum.* **2009**, *36*, 890–903. [[CrossRef](#)]
29. Irpino, A.; Verde, R. Dynamic clustering of interval data using a wasserstein-based distance. *Pattern Recognit. Lett.* **2008**, *29*, 1648–1658. [[CrossRef](#)]
30. Burger, T. Geometric views on conflicting mass functions: From distances to angles. *Int. J. Approx. Reason.* **2016**, *70*, 36–50. [[CrossRef](#)]
31. Sudano, J. The system probability information content (PIC) relationship to contributing components, combining independent multisource beliefs, hybrid and pedigree pignistic probabilities. In Proceedings of the 5th International Conference on Information Fusion. FUSION 2002.(IEEE Cat. No. 02EX5997), Annapolis, MA, USA, 8–11 July 2002; pp. 1277–1283.
32. Li, X.; Wang, F. A Clustering-Based Evidence Reasoning Method. *Int. J. Intell. Syst.* **2016**, *31*, 698–721. [[CrossRef](#)]
33. Srinivas, N.; Deb, K. Multiobjective function optimization using nondominated sorting genetic algorithms. *Evol. Comput.* **1995**, *2*, 221–248. [[CrossRef](#)]
34. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [[CrossRef](#)]
35. Zhu, Q.B. Analysis of Convergence of Ant Colony Optimization Algorithms. *Control Decis.* **2006**, *21*, 764–770. [[CrossRef](#)]
36. Jousselme, A.L.; Maupin, P. On some properties of distances in evidence theory. In Proceedings of the Workshop on Theory of Belief Functions, Brest, France, 31 March–2 April 2010; pp. 1–6.
37. Li, X.D.; Yang, W.D.; Dezert, J. An Airplane Image Target’s Multi-feature Fusion Recognition Method. *Acta Autom. Sin.* **2012**, *38*, 1299–1307. [[CrossRef](#)]
38. Bi, Y.; Bell, D.; Guan, J. Combining evidence from classifiers in text categorization. In Proceedings of the 8th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Wellington, New Zealand, 20–25 September 2004; pp. 521–528.
39. Dezert, J.; Tchamova, A.; Smarache, F.; Konstantinova, P. Target Type Tracking with PCR5 and Dempster’s rules: A Comparative Analysis. In Proceedings of the 2006 9th International Conference on Information Fusion, Florence, Italy, 10–13 July 2006.