Artificial Intelligence Review (2024) 57:239
https://doi.org/10.1007/s10462-024-10833-z

Evidence generalization-based discounting method: assigning unreliable information to partial ignorance

Qiying Hu[1,2] · Qianli Zhou[1] · Zhen Li[3] · Yong Deng[1,4] · Kang Hao Cheong[5,6]

Accepted: 24 June 2024 / Published online: 9 August 2024

Abstract
Conflict management is an important topic when dealing with unreliable sources information fusion in Dempster–Shafer theory. Discounting unreliable bodies of evidence has proven to be effective to decrease conflict. Based on the generalization of belief functions, a new generalization-based discounting method is proposed. When resolving conflicts with the same degree, our method can realize less information loss in comparison with other discounting methods. By simulating the process of resolving conflicts of randomly generated bodies of evidence, using entropy measurements and binary conflict as evaluation index, we show our method's rationality and superiority. Finally, the hyperparameters of the conflict metrics are generated and generalization-based discounting is applied to classify real-world datasets. The improved classification performance further illustrates the usefulness of the method.

Keywords Discounting · Generalization of belief function · Information fusion · Dempster–Shafer theory · Conflict management

1 Introduction

Objects Dempster–Shafer theory (DST) (Dempster 2008; Shafer 1976), which can model ignorance, is widely used in combining uncertain information. DST extends the event space in probability theory to the power set space. Basic probability assignment (BPA) is used to model uncertain information. Dempster proposes Dempster rule of combination (DRC), which can fuse multi-source BPAs. DRC combines multi-source information without requiring prior knowledge and is therefore a flexible tool for uncertainty information processing (Hanyu et al. 2022; Zhao et al. 2023). DST is widely used in pattern classification (Liu et al. 2023b), pattern clustering (Zhang et al. 2023), information fusion (Huang et al. 2023; Huang 2023), decision making (Hanyu et al. 2023; Zhao and Cheong 2023), and reasoning (Pan and Gao 2023; Chang et al. 2022). Liu et al. (2024) introduce iterative evidential reasoning technique to solve the problem of real-time multi-mode fault diagnosis problem and obtained excellent diagnostic performance. Some extended forms of DST have also been proposed. Yang and Xu (2013) propose a new rule for combining multi-source BPAs, i.e., the Evidential Reasoning rule (ER rule). ER rule combines BPAs using

---

Extended author information available on the last page of the article

DRC after preprocessing, which not only satisfies the exchange law and union law, but also keeps the specificity of BPA unchanged in the preprocessing process. In practice, the information sources used for fusion are usually dependent and unreliable, which is inconsistent with the assumption of DRC. However, the reliability of the source can be quantified, and evidence discounting is widely applied to correct for unreliable information sources. Discounting method usually consists of two parts: the calculation of the discounting factor and the reassignment of the unreliable belief. Discounting factor usually reflects the unreliability of the discounted BPA. Shafer (1976) proposes to reassign unreliable belief to the total ignorance, and this method results in significant information loss.

Motivation How to discount the unreliable part of BPA with relatively small information loss has high research value. We propose a new method and give two motivations for such an assignment. The first motivation stems from improvement in evidence combination methods. Dubois and Prade (1988) propose to distribute the conflicting part to union sets rather than the total ignorance. The distribution of discounted masses can also take this idea, we choose to distribute discounted masses to partial ignorances and total ignorance instead of total ignorance. The second motivation stems from one operation: ballooning extension (Delmotte and Smets 2004). Inspired by this space-extending operation, we propose to distribute the discounted mass to sets containing the focal element.

Contribution A new generalization-based discounting method is proposed in this paper. The discounting factor of the BPA is calculated by considering its contribution to global conflict and local conflict, and this method does not require prior knowledge. The discounted masses are reassigned to partial ignorances and total ignorance. We then introduce the concept of the degree of conservatism to represent the proportion of the discounted masses that are distributed to the total ignorance. After operating the conflict resolution process by using SDM, contextual discounting, and generalization-based discounting for randomly generated BPAs, it can be found that generalization-based discounting method enables similar conflict management by losing less information. Finally, the proposed method is applied to classify real-world datasets. The experiment illustrates that reassigning the unreliable parts of BPA through generalization-based discounting can effectively improve the classification accuracy.

The structure of our paper is as follows:

- Some concepts about DST is introduced in Sect. 3.
- In Sect. 4 we introduce how to calculate the discounting factor and how to operate the generalization-based discounting method.
- In Sect. 5 we show the advantages of the generalization-based discounting method.
- In Sect. 6 we conclude this paper and look ahead to our work.

2 Related work

As for the calculation of discounting factor, Schubert (2011) proposes a method using the BPA's contribution to the global conflict. Yang et al. (2013) analogize Schubert's idea and propose the degree of disagreement. Klein and Colot (2010), Murphy (2000) propose to discount a BPA by the Jousselme distance (Jousselme and Maupin 2012) between the discounted BPA and the average of all BPAs. Martin et al. (2008) propose to discount a BPA by average Jousselme distance between the discounted BPA and other BPAs, in which the contribution of the discounted BPA to the local conflict is considered. Huynh (2009) proposes to discount a BPA based on its uncertainty degree and the uncertainty degree is measured through the Shannon entropy of its corresponding pignistic probability functions. Zhou et al. (2022a) propose a multi-criteria assessment method that integrates the reliability, relevance metrics of BPA. Liu et al. (2023a) propose to calculate the discounting factor based on game theory and composite metrics. What's more, the discounting factor can be context-dependent and learnt based on the training data (Pichon et al. 2012). In the proposed discounting method, the discounting factor is calculated by considering its contribution to the global and local conflict.

Some scholars also propose different methods for discounted belief redistribution. Smarandache et al. (2010) propose importance discounting. The value of discounted masses is determined by an importance factor and the discounted masses are distributed to the empty set. Denoeux et al. (2019) propose contextual discounting, which refers to the reliability of each singleton to redistribute unreliable belief to partial ignorances. Yang and Xu (2013), Du and Zhong (2021) propose ER rule, which redistribute the unreliable part to the power set. ER rule preprocesses a BPA by its weight, while weight is determined by BPA's importance factor and reliability. Zhou and Deng (2022a) propose a negation-based discounting method, which provides less commitment.

Although discounting method has shown its power in conflict management, the inevitable disadvantage of discounting is the significant information loss. Based on this, we consider to assign unreliable belief to partial and total ignorances. By considering the information volume of the BPA, we give a more reasonable redistribution method. This method can calculate the redistribution proportion to the global ignorance and partial ignorances. Finally, we apply the proposed method to classify real-world datasets to show its usefulness.

3 Preliminaries

3.1 Dempster–Shafer theory

DST is originated from the decision-making of expert systems. It mainly uses two forms: BPA and belief functions to express information.

Definition 3.1 (*BPA*) Dempster (2008) Given an n-element FoD Θ, the elements in FoD compose a closed space, its power set is:

$$2^{\Theta} = \{\{\emptyset\}, \{\theta_1\}, \cdots, \{\theta_n\}, \{\theta_1, \theta_2\}, \cdots, \{\theta_1, \cdots, \theta_n\}\}, \tag{1}$$

BPA m satisfies

$$m(\emptyset) = 0, \ \ m(A) \in [0, 1], \ \ \sum_{A \in 2^{\Theta}} m(A) = 1, \tag{2}$$

BPA is also called mass function or a piece of evidence. The subsets A of Θ which satisfy $m(A) > 0$ are called focal elements.

Definition 3.2 (*Belief functions*) Shafer (1976) Given an n-element FoD Θ with BPA m, the belief (Bel) function, plausibility (Pl) function and commonality (q) function of focal element A are defined as

$$Bel(A) = 1 - Pl(\overline{A}) = \sum_{\emptyset \neq B \subseteq A} m(B), \tag{3}$$

$$Pl(A) = 1 - Bel(\overline{A}) = \sum_{B \cap A \neq \emptyset} m(B), \tag{4}$$

$$q(A) = \sum_{A \subseteq B} m(B), \tag{5}$$

where $\bar{A}$ means the complement of A in FoD. $[Bel(A), Pl(A)]$ (belief interval) is a powerful tool to indicate the uncertainty degree of propositional A.

$Bel(A)$ represents the supporting degree to A, $Pl(A)$ represents the non-negation degree to A, and $q(A)$ represents the supporting degree to the total focal elements of A.

When we receive information from multiple independent sources, DRC is provided to fuse multiple pieces of evidence to express more information.

Definition 3.3 (*DRC*) Given an n-element FoD Θ with L BPAs $m_1, m_2, ..., m_L$, DRC, denoted as $m_{1\oplus ... \oplus L} = m_1 \oplus ... \oplus m_L$ is defined as

$$m(A) = \begin{cases} \frac{1}{1-K} \sum_{\cap_{i=1}^{L} A_i = A, A_i \subseteq \Theta} \prod_{i=1}^{L} m_i(A_i), A \neq \emptyset \\ 0, A = \emptyset \end{cases}, \tag{6}$$

where $K = \sum_{\cap_{i=1}^{L} A_i = \emptyset, A_i \subseteq \Theta} \prod_{i=1}^{L} m_i(A_i)$.

K is called the conflict coefficient, which reflects global conflict between all BPAs.

Jousselme et al. (2001) propose Jousselme distance to measure the difference between two BPAs.

Definition 3.4 (*Jousselme distance*) Given an n-element FoD Θ with 2 BPAs m_1 and m_2, the Jousselme distance is defined as

$$d_{BPA}(m_1, m_2) = \sqrt{\frac{1}{2}(\vec{m}_1 - \vec{m}_2)\underline{D}(\vec{m}_1 - \vec{m}_2)^T}, \tag{7}$$

where $\vec{m}_1$ and $\vec{m}_2$ are two 2^N-dimensional vectors with coordinates $m(A)$. $\underline{D}$ is a $2^N \times 2^N$ matrix whose elements are $\underline{D}(A, B) = \frac{|A \cap B|}{|A \cup B|}, A, B \subseteq \Theta$.

Using only K in Eq. (6) to measure conflict among BPAs yields many counterintuitive results. Liu et al. (2023a) propose a new binary conflict by introducing Jousselme distance as an alternative metric.

Definition 3.5 (*Binary conflict*) Given an n-element FoD Θ with L BPAs $m_1, m_2, ..., m_L$, the binary conflict $B_{1...L}$ between L BPAs is defined as

$$B_{1...L} = \frac{K_{1...L} + d_{BPA_{1...L}}^{\max}}{2} \exp\left(-\left|\frac{K_{1...L} + d_{BPA_{1...L}}^{\max}}{2} - 1\right|\right), \tag{8}$$

where $d^{\max}_{BPA_{1\ldots L}} = \max(d_{BPA}(m_i, m_j))(1 \le i,j \le L)$, $K_{1\ldots L}$ represents the conflict coefficient between $m_1, m_2, \ldots, m_L$.

Since the actual obtained evidence is not completely reliable, Shafer (1976) proposes the discounting operation to modify the unreliable body of evidence.

Definition 3.6 (*SDM*) Given an n-element FoD Θ with BPA m, SDM is defined as

$$m^{\alpha}(A) = \begin{cases} \alpha + (1-\alpha)m(\Theta) & A = \Theta \\ (1-\alpha)m(A) & A \neq \Theta \end{cases}, \tag{9}$$

α is called the discounting factor, which represents the unreliability of m. The discounted m is represented as m^{α}.

3.2 Pignistic probability transform

Due to the compound elements in BPA, we are unable to make a direct decision through the BPA. Smets et al. propose Pignistic probability transform for BPA, which assigns the masses of compound elements to singletons.

Definition 3.7 (*Pignistic transform*) Given an n-element identification FoD Θ with BPA m, the Pignistic transform of m is defined as

$$BetP_m(A) = \sum_{\substack{B \in 2^{\Theta} \\ B \neq \emptyset}} \frac{|A \cap B|}{|B|} \frac{m(B)}{1 - m(\emptyset)}, \tag{10}$$

where $|\ B\ |$ represents the cardinality of set B.

The Pignistic transform (PT) of the singletons is called the Pignistic probability transform for m. As for the decision result, the singleton with the largest probability distribution is chosen. What's more, probabilistic information content (PIC) is a useful tool for calculating the amount of PT's information.

Definition 3.8 (*PIC*) Given an n-element FoD Θ with BPA m, the PT of m is represented as P, the PIC value of P is defined as

$$\mathrm{PIC}(P) = \frac{1}{\log_2 n} \sum_{A \in \Theta} P(A) \log_2(P(A)) + 1. \tag{11}$$

It can be seen that amount of information available for decision making is greater when the value of PIC is larger. When PIC is equal to 0, it is almost impossible to decide correctly because all singletons have the same probability distribution; when PIC is equal to 1, the probability is concentrated in one singleton and we can make a correct judgment.

3.3 Ballooning extension

Definition 3.9 (*Ballooning extension*) Delmotte and Smets (2004) Let Θ be a FoD, Θ' be a subset of Θ. Suppose mass function $m^{\Theta'}$ defined on Θ' and we need mass function on Θ. This transform operation given by ballooning extension is denoted as $m^{\Theta' \Uparrow \Theta}$, and its values are

$$m^{\Theta' \Uparrow \Theta}(A) = \begin{cases} 0 & \text{if } A = B \cup \overline{\Theta'}, B \subseteq \Theta' \\ m^{\Theta'}(B) & \text{otherwise} \end{cases}. \tag{12}$$

The ballooning extension operation is useful when the received beliefs are built on a limited frame.

3.4 Uncertainty measurements

Many scholars have proposed uncertainty measurements for BPA (Zhou et al. 2021; Balakrishnan et al. 2022).

Definition 3.10 (*Diversity*) Zhou et al. (2022b) Given an n-element FoD Θ with BPA m, the diversity of m is defined as

$$\mathrm{Div}(m) = \frac{\sum_{A \subseteq \Theta; |A|>1} m(A)(|A|-1)}{n-1}, \tag{13}$$

where $\mid A \mid$ represents the cardinality of set A.

This function can be seen as a non-specificity measurement of BPA's uncertainty for it only depends on the non-singletons' belief, and the range value is $[0, 1]$. When $\mathrm{Div}(m) = 0$, the belief of m is all in singletons (Bayesian mass function); when $\mathrm{Div}(m) = 1$, the belief of m is all in the total ignorance (vacuous mass function).

Definition 3.11 (*Fractal-based belief entropy*) Zhou and Deng (2022b) Given an n-element FoD with BPA m, the fractal-based belief (FB) entropy of m is defined as

$$E_{\mathrm{FB}}(m) = -\sum_{A \subseteq \Theta} \sum_{A \subseteq B} \frac{m(B)}{2^{|B|}-1} \log \sum_{A \subseteq B} \frac{m(B)}{2^{|B|}-1}, \tag{14}$$

FB entropy satisfies additivity, monotonicity and probabilistic consistency, which makes it a powerful tool for measuring the uncertainty degree of BPA.

4 Generalization-based discounting method

4.1 Motivation

There are two main motivations for this paper, which have already been mentioned in Sect. 1. In this part we will give a more detailed explanation of motivation 2. Suppose FoD $\Theta = \{\theta_1, \theta_2, \theta_3\}$. Θ', as a subset of Θ, is $\{\theta_2, \theta_3\}$. We extend the Θ' space to Θ space using ballooning extension, the extended sets are $\{\theta_1, \theta_2\}, \{\theta_1, \theta_3\}, \{\theta_1, \theta_2, \theta_3\}$. Suppose FoD $\Theta = \{\theta_1, \theta_2, \theta_3\}$. Θ'', as a subset of Θ, is $\{\theta_3\}$. We extend the Θ'' space to Θ space using ballooning extension, the extended sets are $\{\theta_1, \theta_2, \theta_3\}$. The first situation is to add $\{\theta_1\}$ to $\{\theta_2, \theta_3\}$ space and the second situation is to add $\{\theta_1, \theta_2\}$ to $\{\theta_3\}$ space. Suppose we have a BPA m under a 3-element FoD Θ with $\{\theta_1\}$ and $\{\theta_1, \theta_2\}$ as it's focal elements, α is the discounting factor for m. By borrowing idea from ballooning extension, we consider using the following operation to add the discounted masses back into m, that is, $\alpha m(\{\theta_1\})$ to sets $\{\theta_1, \theta_2\}, \{\theta_1, \theta_3\}, \{\theta_1, \theta_2, \theta_3\}$ while $\alpha m(\{\theta_1, \theta_2\})$ to set $\{\theta_1, \theta_2, \theta_3\}$. The distributed sets are the same as the extended sets described above. It can be found that this operation can also be seen as doing ignorance to the discounted masses, but is more conservative than SDM, for we do not completely distribute the discounted masses to the total ignorance.

4.2 Calculation of discounting factor

In this section, we will give the calculation procedure of the discounting factor of BPA by considering its contribution to the global conflict and the local conflict.

***Step* 1** In this step we calculate the reliability of the BPA by considering its contribution to the global conflict. Suppose we have a BPA set $M = \{m_1, m_2, ..., m_L\}$ defined on the same FoD, the global conflict contribution of $m_i (i = 1, 2, ..., L)$ is defined as (Schubert 2011)

$$GC_{\mathrm{M}}\left(m_i\right) = \frac{K_{\mathrm{M}} - K_{\mathrm{M_i}}}{1 - K_{\mathrm{M_i}}}, \tag{15}$$

where K_{M} is the conflict coefficient of BPA set M, $K_{\mathrm{M_i}}$ is the conflict coefficient of BPA set M_i, M_i is obtained by discarding m_i in the BPA set M. When $GC_{\mathrm{M}}\left(m_i\right)$ is 0, $K_{\mathrm{M}} = K_{\mathrm{M_i}}$, it represents that BPA m_i has no influence on the global conflict; when $0 \le GC_{\mathrm{M}}\left(m_i\right) \le 1$, it represents that BPA m_i have some influence on the global conflict; when $GC_{\mathrm{M}}\left(m_i\right)$ is 1, it represents that BPA m_i is the main contribution to the global conflict. We often multiply $GC_{\mathrm{M}}(m_i)$ by gain factor $\epsilon (0 \le \epsilon \le 1)$ as the unreliability of m_i, and the reliability of m_i is calculated as:

$$Re_{\mathrm{M}}^{GC}(m_i) = 1 - \epsilon GC_{\mathrm{M}}\left(m_i\right). \tag{16}$$

Noting that the larger the selected ϵ, the smaller the reliability $Re_{\mathrm{M}}^{GC}(m_i)$ calculated. We illustrate the function $f(x;\epsilon) = 1 - \epsilon x (0 \le x \le 1, 0 \le \epsilon \le 1)$ for $\epsilon = 1$, $\epsilon = \frac{2}{3}$, $\epsilon = \frac{1}{2}$, $\epsilon = \frac{1}{3}$, and $\epsilon = 0$ in Fig. 1.

***Step* 2** In this step we calculate the reliability of the BPA by considering its contribution to the local conflict. Suppose we have a BPA set $M = \{m_1, m_2, ..., m_L\}$ defined on the same FoD, the average local conflict contribution of m_i is defined as (Martin et al. 2008)

$$LC_{\mathrm{M}}\left(m_i\right) = \frac{1}{L-1} \sum_{j=1, i \ne j}^{L} d_{BPA}(m_i, m_j), \tag{17}$$

where $d_{BPA}(m_i, m_j)$ is the Jousselme distance between BPA m_i and m_j, and $0 \le LC_{\mathrm{M}}(m_i) \le 1$. The larger the $LC_{\mathrm{M}}\left(m_i\right)$, the higher the mean local conflict with other evidence and the less reliable the corresponding m_i. In this case, the reliability of m_i is calculated as:

$$Re_{\mathrm{M}}^{LC}(m_i) = (1 - LC_{\mathrm{M}}\left(m_i\right)^{\lambda})^{\frac{1}{\lambda}}, \lambda > 0. \tag{18}$$

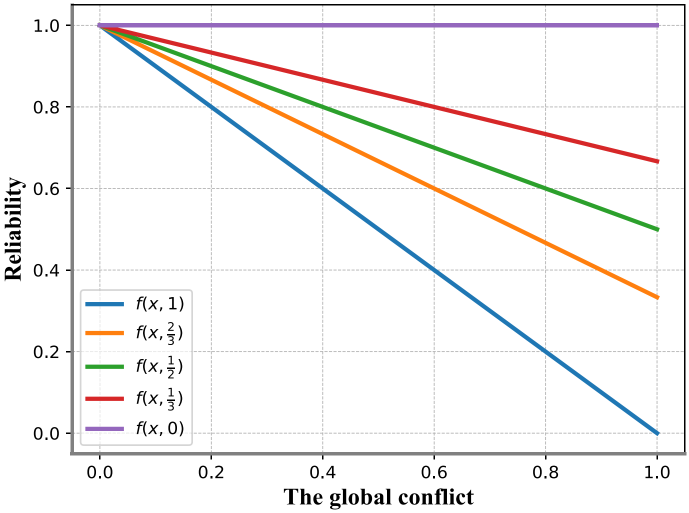


Fig. 1 Reliability of one evidence according to the global conflict

Noting that the larger the selected λ, the larger the reliability $Re_{\mathrm{M}}^{LC}(m_i)$ calculated. We can easily prove the function $f(x;\lambda) = (1 - x^{\lambda})^{\frac{1}{\lambda}} (0 \leq x \leq 1, \lambda > 0)$ is increasing with λ. For any $\lambda_1, \lambda_2 (\lambda_1 > \lambda_2 > 0)$:

$$(1 - x^{\lambda_1})^{\frac{1}{\lambda_1}} > (1 - x^{\lambda_1})^{\frac{1}{\lambda_2}} > (1 - x^{\lambda_2})^{\frac{1}{\lambda_2}}. \tag{19}$$

We illustrate the function $f(x;\lambda)$ for $\lambda = 3$, $\lambda = 2$, $\lambda = \frac{3}{2}$, $\lambda = 1$, $\lambda = \frac{2}{3}$, $\lambda = \frac{1}{2}$, and $\lambda = \frac{1}{3}$ in Fig. 2.

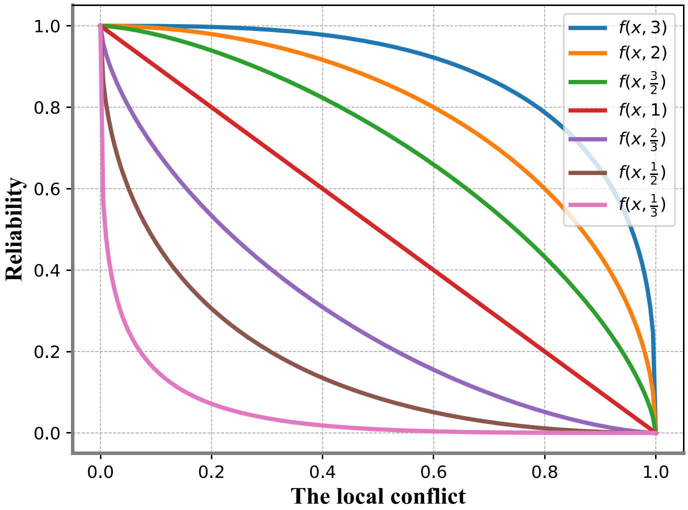


Fig. 2 Reliability of one evidence according to the local conflict

***Step* 3** In this step we calculate the discounting factor of the BPA. Taking into account both BPA's impact on global and local conflict, the reliability of BPA $m_i (i = 1, 2, ..., L)$ is calculated as:

$$Re_{\mathrm{M}}(m_i) = Re^{GC}_{\mathrm{M}}(m_i) \times Re^{LC}_{\mathrm{M}}(m_i). \tag{20}$$

The vector of discounting factors $\alpha = [\alpha_1, \alpha_2, ..., \alpha_L]$ of all BPAs $m_i (i = 1, 2, ..., L)$ is defined as:

$$\begin{cases} \alpha = [\alpha_1, \alpha_2, ..., \alpha_L] \\ \alpha_i = 1 - Re_{\mathrm{M}}(m_i), i = 1, 2, ..., L \end{cases}. \tag{21}$$

It can be found that when the conflict of the BPA gets larger, the lower the reliability calculated, and the larger the corresponding discounting factor. Meanwhile, controlling the value of ϵ or λ can adjust the effect of global or local conflict on the value of discounting factor. As shown in Figs. 1 and 2, it can be found that when we fix global or local conflict, a higher ϵ or lower λ value results in lower reliability value and a higher discounting factor.

4.3 Generalization-based discounting method

Supposing there is one BPA m defined on the FoD $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$, the elements in m are not arranged sequentially but in binary order, the i-th element of the column vector $\mathbf{a} = [a_i]$ is the set which the element at the value of 1 in the binary expression of $i - 1$. The elements in column vector **a** correspond one by one to the elements in the power set 2^{Θ}. Consider element a_{12}, the binary representation of 12 is 1100, so $a_{12} = \{\theta_3, \theta_4\}$. Table 1 shows what are the vectors for $\Theta = \{\theta_1, \theta_2, \theta_3\}$. **BfrM** (Zhou and Deng 2023) is a $2^n \times 2^n$ matrix and the values of it are $BfrM(i,j) = 1$ iff $a_j \subseteq a_i$ and 0 otherwise.

Suppose we have a BPA set $M = \{m_1, m_2, ..., m_L\}$ defined on the FoD $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$, we perform generalization-based discounting method on the BPA set M. There are $2^n \times 2^n$ generalization matrices $P_1, P_2, ..., P_L$ for each BPA respectively. The element of row k_1, column k_2 of matrix P_i is calculated as follows $(k_1, k_2 = 1, 2, ..., 2^n)$:

Table 1 Order of the elements in vector **a**

Position	$\{\theta_3, \theta_2, \theta_1\}$	Θ
1	000	$\{\emptyset\}$
2	001	$\{\theta_1\}$
3	010	$\{\theta_2\}$
4	011	$\{\theta_1, \theta_2\}$
5	100	$\{\theta_3\}$
6	101	$\{\theta_1, \theta_3\}$
7	110	$\{\theta_2, \theta_3\}$
8	111	$\{\theta_1, \theta_2, \theta_3\}$

$$P_i(k_1,k_2)=\begin{cases}1-\alpha_i & k_1=k_2, k_1\neq 2^n, k_2\neq 2^n\\ \alpha_i w_i & k_1=2^n, sum_{k_2}>2\\ 1 & k_1=k_2=2^n\\ \frac{\alpha_i(1-w_i)}{sum_{k_2}-2} & a_{k_2}\subseteq a_{k_1}, k_1\neq 2^n, k_1\neq k_2, sum_{k_2}>2\\ \alpha_i & k_1=2^n, sum_{k_2}=2\\ 0 & \text{otherwise}\end{cases}, \tag{22}$$

where α_i is the discounting factor of m_i. Calculation of sum_{k_2} is as follows:

$$sum_{k_2}=\sum_{j=1}^{2^n} BfrM(j,k_2). \tag{23}$$

The generalization-based discounting operation is as follows:

$$m_i^{\alpha_i}(\mathbf{a})=P_i\times m_i(\mathbf{a}). \tag{24}$$

When $\Theta=\{\theta_1,\theta_2,\theta_3\}$, the discounted BPA $m_i^{\alpha_i}$ is as follows:

$$\begin{aligned}
m_i^{\alpha_i}(\emptyset) =&(1-\alpha_i)m_i(\emptyset)\\
m_i^{\alpha_i}\left(\{\theta_1\}\right) =&(1-\alpha_i)m_i(\{\theta_1\})+\frac{\alpha_i(1-w_i)}{6}m_i(\emptyset)\\
m_i^{\alpha_i}\left(\{\theta_2\}\right) =&(1-\alpha_i)m_i(\{\theta_2\})+\frac{\alpha_i(1-w_i)}{6}m_i(\emptyset)\\
m_i^{\alpha_i}\left(\{\theta_1,\theta_2\}\right) =&(1-\alpha_i)m_i(\{\theta_1,\theta_2\}))\\
&+\frac{\alpha_i(1-w_i)}{2}(m_i(\{\theta_1\})+m_i(\{\theta_2\}))+\frac{\alpha_i(1-w_i)}{6}m_i(\emptyset)\\
m_i^{\alpha_i}\left(\{\theta_3\}\right) =&(1-\alpha_i)m_i(\{\theta_3\})+\frac{\alpha_i(1-w_i)}{6}m_i(\emptyset)\\
m_i^{\alpha_i}\left(\{\theta_1,\theta_3\}\right) =&(1-\alpha_i)m_i(\{\theta_1,\theta_3\})+\frac{\alpha_i(1-w_i)}{2}(m_i(\{\theta_1\})+m_i(\{\theta_3\}))\\
&+\frac{\alpha_i(1-w_i)}{6}m_i(\emptyset)\\
m_i^{\alpha_i}\left(\{\theta_2,\theta_3\}\right) =&(1-\alpha_i)m_i(\{\theta_2,\theta_3\})+\frac{\alpha_i(1-w_i)}{2}(m_i(\{\theta_2\})\\
&+m_i(\{\theta_3\}))+\frac{\alpha_i(1-w_i)}{6}m_i(\emptyset)\\
m_i^{\alpha_i}\left(\{\theta_1,\theta_2,\theta_3\}\right) =&m_i(\{\theta_1,\theta_2,\theta_3\})\\
&+\alpha_i(m_i(\{\theta_1,\theta_2\})+m_i(\{\theta_1,\theta_3\})+m_i(\{\theta_2,\theta_3\}))\\
&+\alpha_i w_i(m_i(\emptyset)+m_i(\{\theta_1\})+m_i(\{\theta_2\})+m_i(\{\theta_3\}))
\end{aligned}. \tag{25}$$

In Table 2, we show the corresponding generalization matrix P_i. It can be found that the discounted masses for partial ignorances are evenly distributed. We fully believe that there exists more reasonable distribution methods, but we have not found it yet, so there is still a lot of room for generalization-based discounting method. The larger the value of w_i, the more the discounted masses are distributed to the total ignorance, resulting in more effective conflict management effect but more information loss. We therefore call w_i the degree

Table 2 Generalization matrix P_i when $\Theta = \{\theta_1, \theta_2, \theta_3\}$, we use $\cdot$ to stand for zero value

	$\{\varnothing\}$	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_1, \theta_2\}$	$\{\theta_3\}$	$\{\theta_1, \theta_3\}$	$\{\theta_2, \theta_3\}$	$\{\theta_1, \theta_2, \theta_3\}$
$\{\varnothing\}$	$1-\alpha_i$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
$\{\theta_1\}$	$\frac{\alpha_i(1-w_i)}{6}$	$1-\alpha_i$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
$\{\theta_2\}$	$\frac{\alpha_i(1-w_i)}{6}$	$\cdot$	$1-\alpha_i$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
$\{\theta_1, \theta_2\}$	$\frac{\alpha_i(1-w_i)}{6}$	$\frac{\alpha_i(1-w_i)}{2}$	$\frac{\alpha_i(1-w_i)}{2}$	$1-\alpha_i$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
$\{\theta_3\}$	$\frac{\alpha_i(1-w_i)}{6}$	$\cdot$	$\cdot$	$\cdot$	$1-\alpha_i$	$\cdot$	$\cdot$	$\cdot$
$\{\theta_1, \theta_3\}$	$\frac{\alpha_i(1-w_i)}{6}$	$\frac{\alpha_i(1-w_i)}{2}$	$\cdot$	$\cdot$	$\frac{\alpha_i(1-w_i)}{2}$	$1-\alpha_i$	$\cdot$	$\cdot$
$\{\theta_2, \theta_3\}$	$\frac{\alpha_i(1-w_i)}{6}$	$\cdot$	$\frac{\alpha_i(1-w_i)}{2}$	$\cdot$	$\frac{\alpha_i(1-w_i)}{2}$	$\cdot$	$1-\alpha_i$	$\cdot$
$\{\theta_1, \theta_2, \theta_3\}$	$\alpha_i w_i$	$\alpha_i w_i$	$\alpha_i w_i$	α_i	$\alpha_i w_i$	α_i	α_i	1

of conservatism of BPA m_i and its range value is [0,1]. When the value of w_i is 1, the generalization-based discounting method degenerates to SDM and it is essential to choose a reasonable value of w_i. It can be found that when diversity function of the BPA is small, the more the belief is concentrated in singletons or in compound sets with small number of elements. The information in this case tends to be more clearly directed with a small degree of uncertainty, the BPA would lose more information if discounted masses are ignored completely, so we choose a small value of w_i. On the contrary, when the diversity function of BPA is large, the more the belief is concentrated in compound sets with more elements or in the total ignorance, the BPA is more uncertain than before. In this case we choose a large value of w_i, for the loss of information will be small even if the discounted massed are completely ignored. Based on this, we think that the relationship between w_i and the diversity function of BPA is a linear positive correlation. The diversity function is numerically normalized, so we directly take the diversity function of BPA as the value of w_i, that is, $w_i = \mathrm{Div}(m_i)$.

In Algorithm 1 we describe the sequential discount form of the generalization-based discounting method. We use m_i^d to represent BPA m_i after d discounts, α_i^d to represent discounting factor for m_i^d, P_i^d to represent generalization matrix for m_i^d. By operating sequential discount, the conflict can be reduced to any level that we are satisfied with.

Table 3 BPA set M_1 generated by Algorithm 2, · stands for zero value

	Arrangement order of elements							
	$\{\varnothing\}$	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_1,\theta_2\}$	$\{\theta_3\}$	$\{\theta_1,\theta_3\}$	$\{\theta_2,\theta_3\}$	$\{\theta_1,\theta_2,\theta_3\}$
m_1	·	·	0.4192	0.1073	0.2190	0.2545	·	·
m_2	·	·	·	·	·	0.2392	·	0.7608
m_3	·	1.0000	·	·	·	·	·	·
m_4	·	0.0550	0.2605	0.0195	0.1535	0.1262	0.1610	0.2243

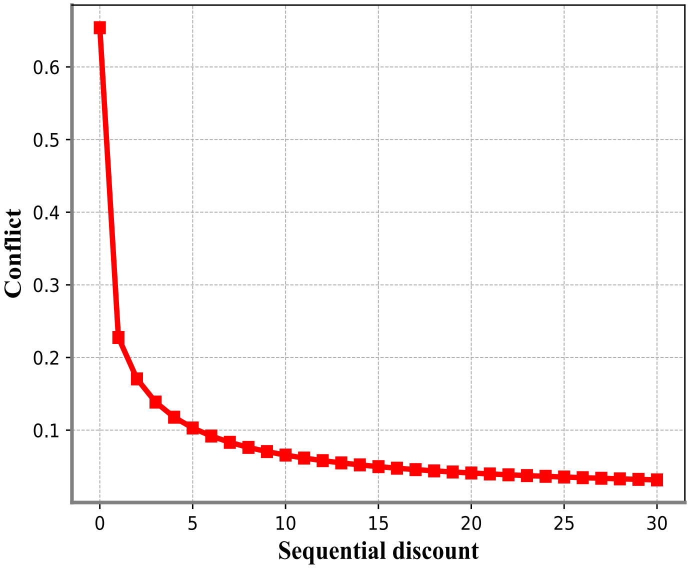


Fig. 3 Conflict of generalization-based discounting method in the sequential discounting process

Algorithm 1 Algorithm for sequential discount using generalization matrix.

```
Input: $m_i(A)(i = 1,2,...,L)$, $A \subseteq \Theta$; $B_1$ (the maximum binary conflict allowed); $d = -1$;
Output: $m_i^{d+1}(A)(i = 1,2,...,L)$;
while $B > B_1$ do
    $d = d + 1$;
    Calculate discounting factor $\alpha_i^d$ for each BPA using Eq.(21);
    Calculate the degree of conservatism $w_i^d$ using Eq.(13);
    Calculate generalization matrix $P_i^d$ using Eq.(22);
    Calculate $m_i^{d+1}(A)(i = 1,2,...,L)$, $A \subseteq \Theta$ using Eq.(24);
    Calculate the binary conflict $B$ using Eq.(8);
end while
```

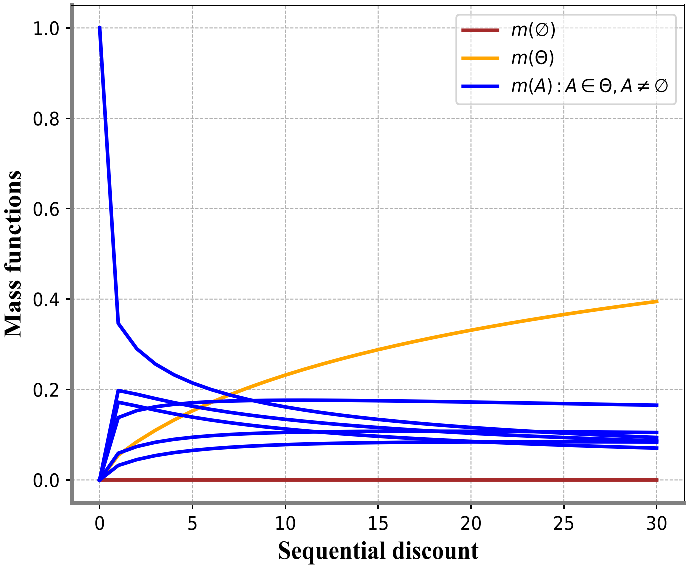


Fig. 4 Combined results in the sequential discounting process. Orange line represents the mass of Θ. Brown line represents the mass of ∅. (Color figure online)

4.4 Numerical example and discussion

In Algorithm 2 (Yang et al. 2013) we describe how to generate BPAs randomly. Suppose the FoD is $\Theta = \{\theta_1, \theta_2, \theta_3\}$, we randomly generate BPA set $M_1 = \{m_1, m_2, m_3, m_4\}$ defined on Θ as shown in Table 3. Calculation of discounting factors is using Eq. (21) and we let gain factor ε be 0.5 and index λ be 2. Sequential discount is performed on M_1 through Algorithm 1, the binary conflict and the combined results of discounted BPAs are shown in Figs. 3 and 4, respectively. It can be found that the supporting degree of preferred propositions become stable after about 30 sequential discounts. This indicates generalization-based sequential discount does not make decision making difficult, on the contrary, the obtained results are more reliable due to the decreasing binary conflict.

Algorithm 2 Algorithm for random generation of normalized BPAs

```
Input: $\Theta$: FoD; $N$: The number of elements in power set $2^\Theta$; $M$: The number of BPAs generated;
Output: $M$ BPAs $(\tilde{m}_1, ..., \tilde{m}_M)$;
Generate set $P(\Theta)$ by randomly arrange the elements in $2^\Theta$;
Randomly generate an integer $k$, the range value of $k$ is $[2, N]$;
The first $k$ elements of $P(\Theta)$ are randomly given values within $[0, 1]$;
Set the value assigned to the empty set to 0.
Normalize the column vector $m_1 \to \tilde{m}_1$ as the first BPA;
Repeat the above operations for $M - 1$ times to generate $M$ BPAs;
```

5 Comparisons of the original discounting methods and the proposed generalization-based discounting method

5.1 Discussion with existing discounting methods

Discount unreliable part of BPA consist of two main parts: the determination of the discounting factor and the reassignment of unreliable belief. Discounting factor can be determined, or partial information on discounting factor can be provided, if prior knowledge is available. Typical work is the meta-knowledge proposed by Pichon et al. (2012). Meta-knowledge proposes to discount by taking into account the truthfulness and relevance of the source information. These metrics can be obtained from real-world data. However, in many cases, due to the lack of prior knowledge, the discounting factor can only be obtained by analyzing the combined BPAs. There are two types of methods commonly used, one is to measure the conflict between the evidence with certain conflict metrics, and then express the discounting factor with a certain function (Schubert 2011; Martin et al. 2008); the other method is to obtain the discounting factor based on the uncertainty of the BPA. Different methods for assigning unreliable belief have also been proposed, including Shafer's classical discounting method, contextual discounting (Mercier et al. 2008), and weighted belief assignment in ER rule (Yang et al. 2013). Shafer proposes to reassign belief to the full set; ER rule reassigns belief to the power set; and contextual discounting uses more detailed information in different contexts to reassign belief to the partial and total ignorances. In generalization-based discounting, we use global and local conflict to measure the reliability of BPA, and diversity metrics to determine the assigning weights. In Table 4, we compare in detail the generalization-based discounting and the other discounting methods in aspects of prior knowledge requirement, discounting factor determination, and belief assignment. Compared with contextual discounting and ER rule, the generalization-based discounting has a wider application range (no prior knowledge required). Compared with Schubert, Zhou et al.'s, and Liu et al.'s discounting method, the generalization-based discounting adopts a more conservative belief assignment way, which allows for less information loss when resolving conflict with the same degree.

5.2 Numerical comparison with SDM

Table 4 Comparison results of the generalization-based discounting with the existing typical discounting method

Method	Prior knowledge	Discounting factor	Belief assignment
Schubert (2011)	No	Conflict	Full set
Zhou et al. (2022a)	No	Conflict Relevance	Full set
Liu et al. (2023a)	No	Conflict Uncertainty	Full set
Contextual discounting (Denoeux et al. 2019)	Yes	Meta-knowledge	Partial and total ignorances
ER rule (Yang et al. 2023)	Yes	Weight Reliability	Power set
Generalization-based discounting	No	Conflict	Partial and total ignorances

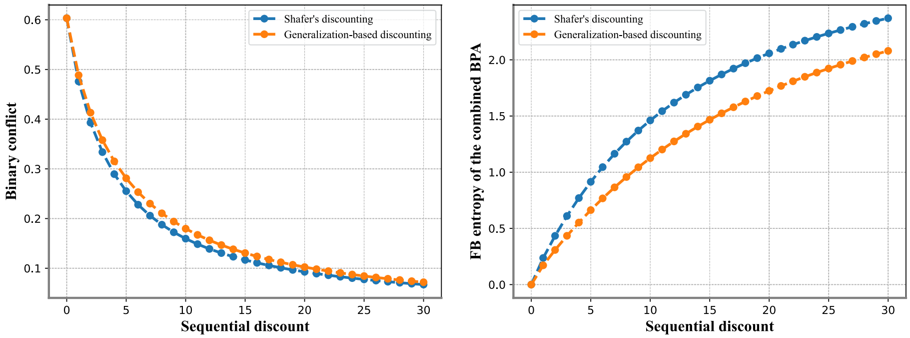
Table 5 BPA set M_2 generated by Algorithm 2, · stands for zero value

	Arrangement order of elements							
	$\{\varnothing\}$	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_1,\theta_2\}$	$\{\theta_3\}$	$\{\theta_1,\theta_3\}$	$\{\theta_2,\theta_3\}$	$\{\theta_1,\theta_2,\theta_3\}$
m_1	·	·	·	1.0000	·	·	·	·
m_2	·	0.0798	0.2887	0.2400	0.1251	0.0459	0.0093	0.2112
m_3	·	·	0.3652	·	·	0.4616	0.1731	·
m_4	·	·	0.4131	·	·	·	·	0.5869
m_5	·	0.0239	0.2353	0.2496	0.0893	0.0355	0.1747	0.1917
m_6	·	·	0.0646	0.2815	·	0.2330	0.4166	0.0043
m_7	·	·	0.4517	·	·	·	0.5483	·
m_8	·	0.1068	0.1424	0.0498	0.0726	0.2279	0.1955	0.2051
m_9	·	·	·	1.0000	·	·	·	·

Information is continuously lost along with the discounting process, which can be seen as a necessary drawback if we want to operate conflict management. FB entropy, as a modified version of Deng entropy, has been proven as a powerful tool for measuring the uncertainty of the BPA (Zhou and Deng 2022c). What's more, we prefer to see the most supporting proposition's uncertainty degree as small as possible, so we choose to observe the belief interval of the most supporting proposition. The PIC values of the transformed probability distribution are also compared. It is more conductive to make reasonable decision when PIC values are larger.

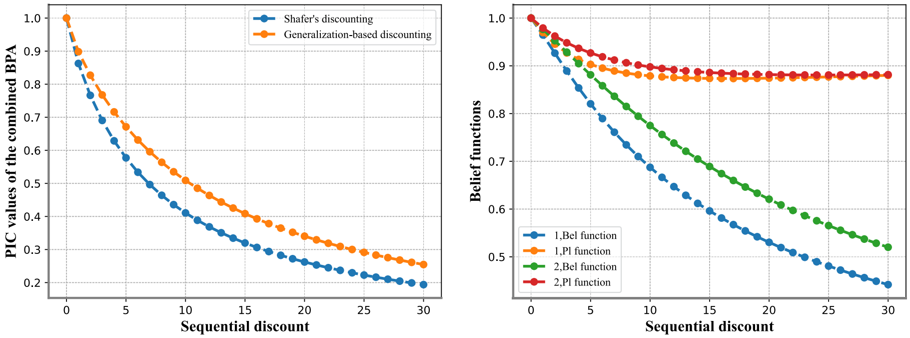
Suppose the FoD is $\Theta = \{\theta_1, \theta_2, \theta_3\}$, we randomly generate BPA set $M_2 = \{m_1, m_2, ..., m_9\}$ defined on Θ as shown in Table 5. Calculation of discounting factors is using Eq. (21) and we let gain factor ε and index λ be 0.1 and 2, respectively. We perform sequential discount on BPA set M_2 using SDM and generalization-based discounting method. In Fig. 5, it can be found that generalization-based discounting method has achieved almost the same conflict management effect as SDM. However, the FB entropy of the generalization-based discounting method is significantly lower than that of SDM; the PIC value is significantly higher; the difference between the *Pl* function and the *Bel* function is significantly smaller. This illustrates that generalization-based discounting method loses less information than SDM in the sequential discount process. Through further analysis, it can be found that although the generalization-based discounting method uses more number of discounts when both methods reduce the conflict to a small degree, the information metrics are all better than SDM. The comparison results are shown in Table 6.

For computational complexity, suppose we have M BPAs to be fused and FoD $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$. For each BPA to be discounted, SDM only needs to recalculate the belief of each focal element, so the time complexity is $O(M2^n)$. While generalization-based discounting needs to redistribute the belief by the matrix calculation through Eq. (24), so the time complexity is $O(M2^{2n})$. Thus the computational complexity of the generalization-based discounting method is about 2^n times that of SDM. When n is small, the added computational complexity is almost negligible; when n is large, the proposed method may introduce more computational complexity than SDM.



(a) Binary conflict of SDM and our method.

(b) FB entropy of SDM and our method.



(c) PIC values of SDM and our method.

(d) Belief functions of SDM and our method. 1 represents SDM and 2 represents our method.

Fig. 5 Binary conflict, FB entropy, PIC values and belief functions of SDM and our method in the sequential discounting process

Table 6 The number of sequential discounts used and the information metrics of the combined BPAs when a similar conflict management effect is reached

	Number of discounts	Conflict	FB entropy	PIC	Interval
SDM	27	0.0730	2.2937	0.2103	0.4139
Generalization-based discounting	30	**0.0719**	**2.0796**	**0.2547**	**0.3615**

In Table, bold stands for the best-performing metrics(lower conflict,lower FB entropy,higher PICvalue,smaller belief interval)

5.3 Numerical comparison with contextual discounting

Contextual discounting exploits more fine-grained reliability to correct evidence and has a wide range of applications in areas like pattern recognition or decision making. In our hypothesis, we don't know accurately whether the information source is reliable or not so we use $1-\alpha$ to represent the degree of reliability. As SDM operated by a discounting factor

Table 7 Generalization matrix for m using contextual discounting, · stands for zero value

	$\{\emptyset\}$	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_1,\theta_2\}$	$\{\theta_3\}$	$\{\theta_1,\theta_3\}$	$\{\theta_2,\theta_3\}$	$\{\theta_1,\theta_2,\theta_3\}$
$\{\emptyset\}$	0.216	·	·	·	·	·	·	·
$\{\theta_1\}$	0.144	0.360	·	·	·	·	·	·
$\{\theta_2\}$	0.144	·	0.360	·	·	·	·	·
$\{\theta_1,\theta_2\}$	0.096	0.240	0.240	0.600	·	·	·	·
$\{\theta_3\}$	0.144	·	·	·	0.360	·	·	·
$\{\theta_1,\theta_3\}$	0.096	0.240	·	·	0.240	0.600	·	·
$\{\theta_2,\theta_3\}$	0.096	·	0.240	·	0.240	·	0.600	·
$\{\theta_1,\theta_2,\theta_3\}$	0.064	0.160	0.160	0.400	0.160	0.400	0.400	1.000

Table 8 Generalization matrix for m using generalization-based discounting, · stands for zero value

	$\{\emptyset\}$	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_1,\theta_2\}$	$\{\theta_3\}$	$\{\theta_1,\theta_3\}$	$\{\theta_2,\theta_3\}$	$\{\theta_1,\theta_2,\theta_3\}$
$\{\emptyset\}$	0.6000	·	·	·	·	·	·	·
$\{\theta_1\}$	0.0667	0.6000	·	·	·	·	·	·
$\{\theta_2\}$	0.0667	·	0.6000	·	·	·	·	·
$\{\theta_1,\theta_2\}$	0.0667	0.2000	0.2000	0.6000	·	·	·	·
$\{\theta_3\}$	0.0667	·	·	·	0.6000	·	·	·
$\{\theta_1,\theta_3\}$	0.0667	0.2000	·	·	0.2000	0.6000	·	·
$\{\theta_2,\theta_3\}$	0.0667	·	0.2000	·	0.2000	·	0.6000	·
$\{\theta_1,\theta_2,\theta_3\}$	·	·	·	0.4000	·	0.4000	0.4000	1.0000

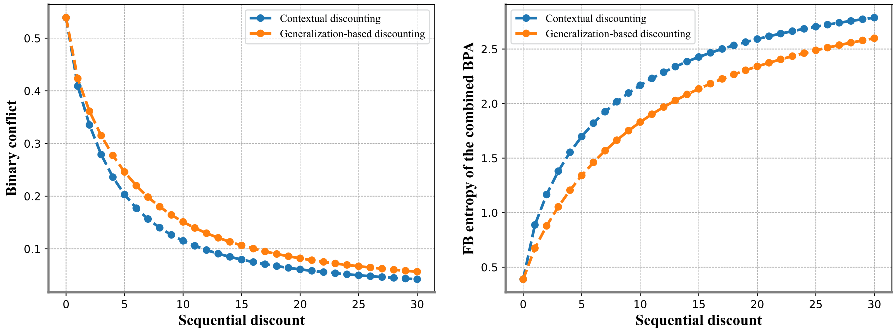
Table 9 Discounted BPAs through contextual discounting and generalization-based discounting, · stands for zero value

	$\{\emptyset\}$	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_1,\theta_2\}$	$\{\theta_3\}$	$\{\theta_1,\theta_3\}$	$\{\theta_2,\theta_3\}$	$\{\theta_1,\theta_2,\theta_3\}$
Contextual discounting	·	0.18	0.18	0.24	·	0.12	0.12	0.16
Generalization-based discounting	·	0.30	0.30	0.20	·	0.10	0.10	·

α, the contextual discounting is operated by a vector $(\alpha_1, \alpha_2, ..., \alpha_n)$ for an n-element FoD Θ ($\alpha_1, ..., \alpha_n$ represent the unreliabilities of the corresponding n singletons). Suppose FoD $\Theta = \{\theta_1, \theta_2, \theta_3\}$. $\{\theta_1\}, \{\theta_2\}, \{\theta_3\}$ correspond to helicopter, rocket, and airplane in targets. Suppose one sensor provides BPA m: $m(\{\theta_1\}) = 0.5$ and $m(\{\theta_2\}) = 0.5$, α_1 represents the unreliability of helicopter is 0.4 and the other sources are equally unreliable. In Tables 7 and 8, we show the generalization matrix for contextual discounting and generalization-based discounting method, respectively. In Table 9, we show the discounted results. We can find that helicopter, rocket, airplane are all not completely reliable identification target, so the masses are transferred to $\{\theta_1, \theta_2\}$, $\{\theta_1, \theta_3\}$, $\{\theta_2, \theta_3\}$ and $\{\theta_1, \theta_2, \theta_3\}$ through both two methods, which indicates both two methods can increase the reliability of the BPA. Further we calculate the PIC values of the two discounted BPAs, the value of generalization-based discounting method is 0.1363 while the value of contextual discounting method is 0.0587.

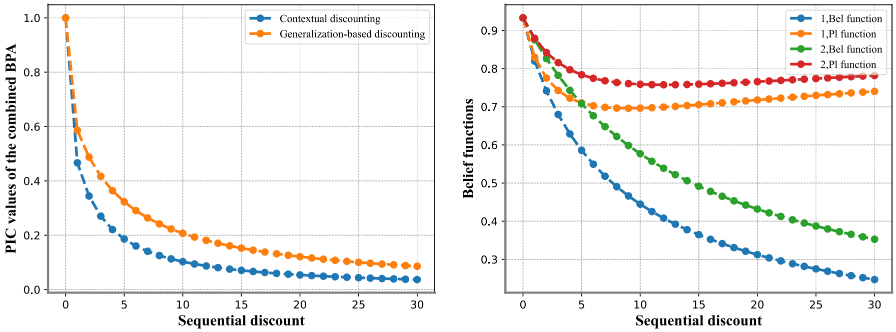
Table 10 BPA set M_3 generated by Algorithm 2, · stands for zero value

	Arrangement order of elements							
	$\{\varnothing\}$	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_1,\theta_2\}$	$\{\theta_3\}$	$\{\theta_1,\theta_3\}$	$\{\theta_2,\theta_3\}$	$\{\theta_1,\theta_2,\theta_3\}$
m_1	·	0.0544	0.1904	0.1664	0.1459	0.1007	0.0438	0.2983
m_2	·	·	0.2158	0.4343	·	·	0.3499	·
m_3	·	·	0.2045	0.1715	·	0.6240	·	·
m_4	·	0.1414	0.1450	0.0980	0.1311	0.1665	0.2174	0.1006
m_5	·	0.0512	0.6887	·	·	0.2601	·	·
m_6	·	·	·	0.5995	0.2321	0.1684	·	·
m_7	·	·	0.4090	0.0605	·	·	·	0.5305
m_8	·	0.0235	0.3573	·	0.2740	·	0.3452	·



(a) Binary conflict of contextual discounting and our method.

(b) FB entropy of contextual discounting and our method.



(c) PIC values of contextual discounting and our method.

(d) Belief functions of contextual discounting and our method. 1 represents contextual discounting and 2 represents our method.

Fig. 6 Binary conflict, FB entropy, PIC values and belief functions of contextual discounting and our method in the sequential discounting process

The discounted BPA obtained by the generalization-based discounting method contains more information for making decision.

On this basis we guess that the generalization-based discounting method can improve the reliability of the BPA while containing more information for making decision. We have done more relative experiments to confirm this and we give one of the numerical examples here. BPA set $M_3 = \{m_1, m_2, ..., m_8\}$ is randomly generated as shown in Table 10 defined on FoD $\Theta = \{\theta_1, \theta_2, \theta_3\}$. Calculation of discounting factors is using Eq. (21) and we let gain factor ε be 0.5 and index λ be 2. Since the proposed method cannot assess the unreliability degree of each singleton, we treat them as equally unreliable. In Fig. 6, it can be found that both methods have a conflict management effect, while the effect of contextual discounting is a little better. However, the FB entropy of the generalization-based discounting method is significantly lower than that of contextual discounting; the PIC value is significantly higher; the difference between the *Pl* function and the *Bel* function is significantly smaller. This illustrates that generalization-based discounting method loses less information than contextual discounting in the sequential discounting process. Through further analysis, it can found that although the generalization-based discounting method uses more number of discounts when both methods reduce the conflict to the same degree, the information metrics are all better than contextual discounting, and the detailed comparison results are shown in Table 11.

5.4 Comparison analysis with SDM and contextual discounting

When the calculation method of the discounting factor is fixed: the given two examples illustrate that the generalization-based discounting method requires more sequential discounting numbers to achieve the same conflict management effect. However, the information loss of the generalization-based discounting method is significantly smaller during sequential discounting process. The reason is that when comparing with SDM, the proposed method allows the discounted masses to be more uncertain but not completely ignored; when comparing with contextual discounting, the proposed method commits back less belief to partial ignorances and total ignorance, so the cost of information loss is smaller than the compared two methods. Besides, due to the introduction of the degree of conservatism, the distribution of the discounted masses is more rational. When the proposed method achieves the same conflict management effects as the other two methods, the information metrics are all significantly better. Finally, SDM has the lowest computational

Table 11 The number of sequential discounts used and the information metrics of the combined BPAs when a similar conflict management effect is reached

	Number of discounts	Conflict	FB entropy	PIC	Interval
Contextual discounting	6	0.2030	1.6980	0.1862	0.1243
Generalization-based discounting	8	**0.1983**	**1.5678**	**0.2638**	**0.1202**

In Table, bold stands for the best-performing metrics(lower conflict,lower FB entropy,higher PICvalue,smaller belief interval) complexity, and the computational complexity between the proposed method and contextual discounting is close.

5.5 Application to real-world dataset

In this section we first give a multi-attribute classification model based on DST; after that we give a learning-based method to generate the parameters ϵ and λ based on Genetic Algorithm (Mirjalili and Mirjalili 2019); and finally we conduct experiments on real-world datasets, and the results illustrate that the generalization-based discounting method can effectively improve the classification accuracy.

5.5.1 Multi-attribute classification model based on DST

Since the distinctions of attributes between different categories in multi-attribute classification problems are frequently arbitrary or overlapping, incorrect classification results often occur. Dempster–Shafer theory (DST), as an multi-source information fusion tool, can effectively combine the complementary knowledge from different information sources to improve classification performance. Xu et al. (2013), Liu et al. (2020) propose a multi-attribute classification model based on Gaussian distribution to generate BPA. Assume that the original data set, containing n classes and K attributes, is denoted as the FoD $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$. The classification of data sets is achieved through the following four steps.

1. The training set for the initial data set is made up of l samples drawn at random from each class, and the testing set is made up of the remaining l' samples.
2. Calculate the mean value μ_{ij} and standard deviation σ_{ij} of all training samples in the $i(i = 1, 2, ..., n)$-th class on the $j(j = 1, 2, ..., K)$-th attribute.
3. Based on the obtained μ_{ij} and σ_{ij}, construct the Gaussian grade of membership function $f(x;\mu_{ij};\sigma_{ij}^2)$ for class i on the j-th attribute:

$$f\left(x;\mu_{ij},\sigma_{ij}^2\right) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp\left(-\frac{\left(x-\mu_{ij}\right)^2}{2\sigma_{ij}^2}\right). \tag{26}$$

   According to Eq. (26), we can obtain the membership functions of K attributes to the n classes.
4. For the $j(j = 1, 2, ..., K)$-th attribute on the $y(y = 1, 2, ...l')$-th test sample, normalize the n membership values based on $f_{ij}^y = \frac{f_{ij}^y}{\sum_i f_{ij}^y}$, where f_{ij}^y denotes the membership value of

Table 12 The calculated mean values and standard deviations of four attributes in three classes

	Mean values				Standard deviations			
	SL	SW	PL	PW	SL	SW	PL	PW
Setosa	5.0375	3.4525	1.4600	0.2350	0.3621	0.3609	0.1722	0.0975
Versicolor	6.0100	2.7800	4.3175	1.3500	0.5232	0.3330	0.4511	0.2075
Virginica	6.6225	2.9600	5.6075	1.9900	0.6841	0.3365	0.5876	0.2725

the i-th class that provide by the j-th attribute. Rank the n normalized membership values in decreasing order: $\widetilde{f}_{1j}^{y}, \widetilde{f}_{2j}^{y}, ..., \widetilde{f}_{nj}^{y}$. Their corresponding classes are $\theta'_1, \theta'_2, ..., \theta'_n$, the BPA m_j^y is generated as follows:

$$\begin{array}{l} m_j^y\left(\left\{\theta'_1\right\}\right) = \widetilde{f}_{1j}^{y} \\ m_j^y\left(\left\{\theta'_1, \theta'_2\right\}\right) = \widetilde{f}_{2j}^{y} \\ \dots \\ m_j^y\left(\left\{\theta'_1, \theta'_2, \dots, \theta'_n\right\}\right) = m(\Theta) = \widetilde{f}_{nj}^{y} \end{array}. \tag{27}$$

Take Iris data set for example, there are three classes, namely Setosa, Versicolor, and Virginica. Each class has 50 samples and 4 attributes: sepal length (SL), sepal width (SW), petal length (PL), and petal width (PW). Firstly, we randomly select 30 out of 50 samples from each class as training samples and the rest as testing samples. Secondly, we calculate the mean values and standard deviations of the four attributes of different classes in training samples, and the results are shown in Table 12. Thirdly, 12 Gaussian membership functions can be obtained through Eq. (26). Finally, for a given testing sample, we can generate four BPAs based on the values of its four attributes. Suppose the SL attribute of a testing sample is 5, the membership value of belonging to Setosa, Versicolor, and Virginica can be calculated according to the three corresponding Gaussian membership functions, respectively. As shown in Fig. 7, the membership that the testing sample belongs to class Setosa, Versicolor, and Virginica is 1.0959, 0.1183, 0.0350, respectively. Rank these three membership values in decreasing order, we can obtain class Setosa, Versicolor, and Virginica corresponds to $\theta'_1, \theta'_2, \theta'_3$, respectively. Then we normalize the membership values, the obtained BPA m is shown in Eq. (28).

$$\begin{array}{l} m(\{\theta'_1\}) = m(\{Setosa\}) = 0.8773 \\ m(\{\theta'_1, \theta'_2\}) = m(\{Setosa, Versicolor\}) = 0.0947 \\ m(\{\theta'_1, \theta'_2, \theta'_3\}) = m(\{Setosa, Versicolor, Virginica\}) = 0.0280 \end{array}. \tag{28}$$

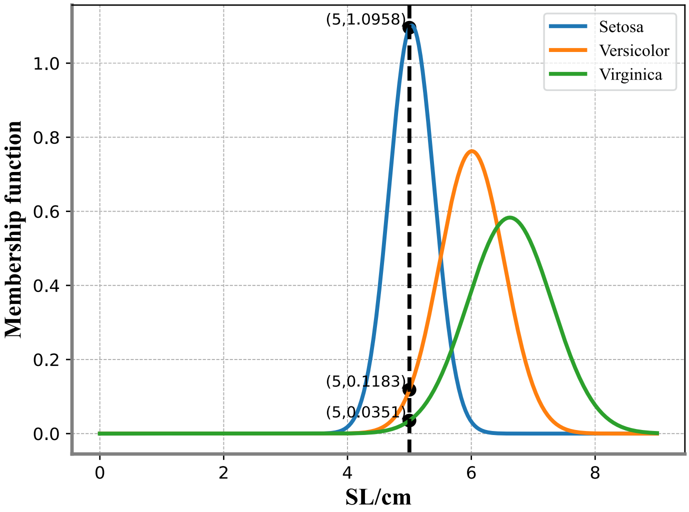
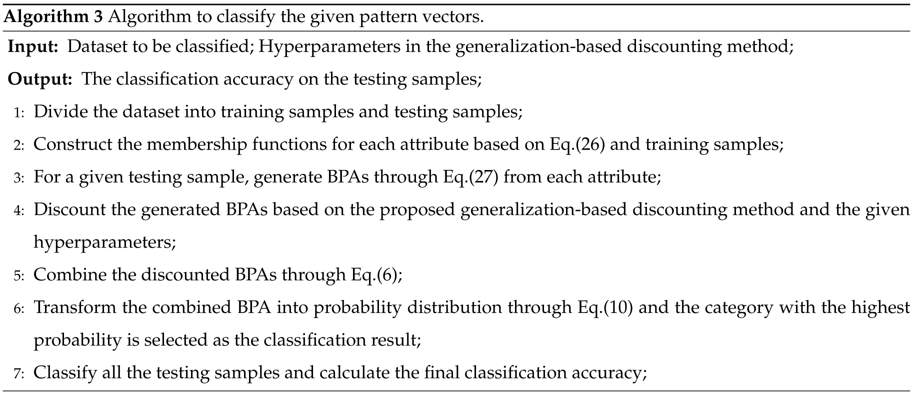


Fig. 7 Membership functions and three intersections for attribute SL

5. By discounting the generated K BPAs $(m_1^y, m_2^y, \ldots, m_K^y)$ through the proposed generalization-based discounting method, combine the discounted BPAs through Eq. (6), and transform the combined BPA to probability distribution through Eq. (10), we can obtain the predicted result of the y-th test sample (the class corresponding to the largest probability distribution is seen as the predicted result);
6. Repeat the above processes $l' - 1$ times to obtain the predicted results for all testing samples. The whole procedure is summarized in Algorithm 3.

Algorithm 3 Algorithm to classify the given pattern vectors



```
Algorithm 3 Algorithm to classify the given pattern vectors.
Input: Dataset to be classified; Hyperparameters in the generalization-based discounting method;
Output: The classification accuracy on the testing samples;
Divide the dataset into training samples and testing samples;
Construct the membership functions for each attribute based on Eq.(26) and training samples;
For a given testing sample, generate BPAs through Eq.(27) from each attribute;
Discount the generated BPAs based on the proposed generalization-based discounting method and the given
  hyperparameters;
Combine the discounted BPAs through Eq.(6);
Transform the combined BPA into probability distribution through Eq.(10) and the category with the highest
  probability is selected as the classification result;
Classify all the testing samples and calculate the final classification accuracy;
```

5.5.2 Parameter generation

In Algorithm 4, we describe how to generate parameters ϵ and λ through training data and Genetic Algorithm.

Table 13 The average classification accuracy with and without discounting on the Haberman and Iris datasets

Dataset	Sample	Class	Attribute	Accuracy (no discount)	Accuracy (discount)
Haberman	306	2	3	61.80	**64.04**
Iris	150	3	4	94.58	**94.77**

In Table, bold stands for higher classification accuracy performance compared with no discount case

Algorithm 4 Algorithm for generating parameters ϵ and λ

```
Input: $d = 0$(the number of interation); $D$(the maximum number of interation); $\epsilon_{\max}$(the upper bound of $\epsilon$); $\lambda_{\max}$(the upper bound of $\lambda$); Crossover probability $P_1$; Mutation Probability $P_2$;
Output: the selected parameter;
while $d \leq D$ do
    $d = d + 1$;
    Generate random population for $\epsilon$ and $\lambda$, the range of $\epsilon$ and $\lambda$ is $[0, \epsilon_{\max}]$ and $[0, \lambda_{\max}]$, respectively;
    Calculate the classification accuracy of the training data base on the population and the proposed classification model in Section 5.5.1;
    Save the best solution in current population;
    Select two parent chromosomes from a population according to their fitness (the better fitness, the bigger chance to be selected);
    With a crossover probability $P_1$ cross over the parents to form a new offspring (children). If no crossover was performed, offspring is an exact copy of parents;
    With a mutation probability $P_2$ mutate new offspring at each locus (position in chromosome);
    Place new offspring in a new population;
    Use new generated population for a further run of algorithm;
end while
Return the best solution in the loop procedure;
```

5.5.3 Classification performance on UCI datasets

We apply the classification model in Sect. 5.5.1 on Haberman and Iris dataset, where we randomly select 60% of the data as training samples and the rest as test samples. The average classification accuracies without and with generalization-based discounting method in the case of randomly dividing the dataset 200 times are given in Table 13. The maximum values of ϵ and λ are set to 1 and 20, respectively, the crossover probability is 0.8, the mutation probability is 0.1, the population size is 50, and the maximum number of iterations is 5. After the process of parameter learning and discounting the unreliable parts, a classification accuracy increase can be found.

6 Conclusion

In this paper, a new generalization-based discounting method is proposed. The proposed method is inspired by ballooning extension, and the discounted masses are distributed to partial ignorances and total ignorance. The value of discounted masses is determined by the discounting factor, which is calculated based on the BPA's global and local conflict contribution. We then introduce the concept of the degree of conservatism to represent the proportion of the discounted masses that are distributed to the total ignorance. For the purpose of losing less useful information, we choose the diversity metrics. The discounted masses for partial ignorance(s) are evenly distributed. Based on the above mentioned, we can calculate the corresponding generalization matrices for BPAs to perform discounting operations. Verified by numerical examples, generalization-based discounting method loses much less information when achieving similar conflict management effects compared to SDM and contextual discounting. We then apply the generalization-based discounting to classify real-world datasets and give a hyperparameters generation algorithm. There exists some improvements in classification accuracy compared to the no discount case. In future research, we can extend this work from the following aspects: (1) the determination for the degree of conservatism is only in terms of uncertainty degree, we can consider using other properties of BPA; (2) for the discounted mass distributed to the partial ignorances, we can explore more reasonable methods instead of an even distribution; (3) the proposed discounting method can be used to correct the unreliable part of BPA and can be applied in more DST-based machine learning system to improve its robustness and reliability.

Acknowledgements The work is partially supported by the National Natural Science Foundation of China (Grant No. 62373078). We truly appreciate the associate editor and the reviewers for their useful and careful feedback that improved this paper.

Author Contributions Qiying Hu: writing-original draft, conceptualization, methodology, software, validation. Qianli Zhou: writing-original draft, methodology, validation. Zhen Li: validation, supervision. Yong Deng: validation, resources, writing-review & editing, supervision, funding acquisition. Kang Hao Cheong: validation, writing-review & editing.

Declarations

Competing Interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Balakrishnan N, Buono F, Longobardi M (2022) A unified formulation of entropy and its application. Physica A. https://doi.org/10.1016/j.physa.2022.127214

Chang L, Zhang L, Fu C, Chen Y-W (2022) Transparent digital twin for output control using belief rule base. IEEE Trans Cybern 52(10):10364–10378. https://doi.org/10.1109/TCYB.2021.3063285

Delmotte F, Smets P (2004) Target identification based on the transferable belief model interpretation of Dempster-Shafer model. IEEE Trans Syst Man Cybern A 34(4):457–471

Dempster AP (2008) Upper and lower probabilities induced by a multivalued mapping. In: Classic works of the Dempster-Shafer theory of belief functions. Springer, pp 57–72

Denoeux T, Kanjanatarakul O, Sriboonchitta S (2019) A new evidential k-nearest neighbor rule based on contextual discounting with partially supervised learning. Int J Approx Reason 113:287–302

Du Y-W, Zhong J-J (2021) Generalized combination rule for evidential reasoning approach and Dempster-Shafer theory of evidence. Inf Sci 547:1201–1232

Dubois D, Prade H (1988) Representation and combination of uncertainty with belief functions and possibility measures. Comput Intell 4(3):244–264

Hanyu E, Cui Y, Pedrycz W, Fayek AR, Li Z, Li J (2022) Design of fuzzy rule-based models with fuzzy relational factorization. Expert Syst Appl 206:117904

Hanyu E, Cui Y, Pedrycz W, Li Z (2023) Design of distributed rule-based models in the presence of large data. IEEE Trans Fuzzy Syst 31(7):2479–2486. https://doi.org/10.1109/TFUZZ.2022.3226250

Huang L (2023) Combination of information in labeled and unlabeled data via evidence theory. IEEE Trans Artif Intell. https://doi.org/10.1109/TAI.2023.3316194

Huang L, Zhao W, Liu Y, Yang D, Liew AW-C, You Y (2023) An evidential multi-target domain adaptation method based on weighted fusion for cross-domain pattern classification. IEEE Trans Neural Netw Learn Syst. https://doi.org/10.1109/TNNLS.2023.3275759

Huynh V-N (2009) Discounting and combination scheme in evidence theory for dealing with conflict in information fusion. In: International conference on modeling decisions for artificial intelligence. Springer, pp 217–230

Jousselme A-L, Maupin P (2012) Distances in evidence theory: comprehensive survey and generalizations. Int J Approx Reason 53(2):118–145

Jousselme A-L, Grenier D, Bossé É (2001) A new distance between two bodies of evidence. Inf Fusion 2(2):91–101

Klein J, Colot O (2010) Automatic discounting rate computation using a dissent criterion. In: Workshop on the theory of belief functions. pp 1–6

Liu Z-G, Liu Y, Dezert J, Cuzzolin F (2020) Evidence combination based on credal belief redistribution for pattern classification. IEEE Trans Fuzzy Syst 28(4):618–631. https://doi.org/10.1109/TFUZZ.2019.2911915

Liu X, Liu S, Xiang J, Sun R (2023a) A conflict evidence fusion method based on the composite discount factor and the game theory. Inf Fusion 94:1–16

Liu Z-G, Fu Y-M, Pan Q, Zhang Z-W (2023b) Orientational distribution learning with hierarchical spatial attention for open set recognition. IEEE Trans Pattern Anal Mach Intell 45(7):8757–8772. https://doi.org/10.1109/TPAMI.2022.3227913

Liu Z, Li C, He X (2024) Evidential ensemble preference-guided learning approach for real-time multimode fault diagnosis. IEEE Trans Ind Inf 20(4):5495–5504. https://doi.org/10.1109/TII.2023.3332112

Martin A, Jousselme A-L, Osswald C (2008) Conflict measure for the discounting operation on belief functions. In: (2008) 11th international conference on information fusion. IEEE, pp 1–8

Mercier D, Quost B, Denœux T (2008) Refined modeling of sensor reliability in the belief function framework using contextual discounting. Inf Fusion 9(2):246–258

Mirjalili S, Mirjalili S (2019) Genetic algorithm, evolutionary algorithms and neural networks: theory and applications. pp 43–55

Murphy CK (2000) Combining belief functions when evidence conflicts. Decis Support Syst 29(1):1–9

Pan L, Gao X (2023) Evidential Markov decision-making model based on belief entropy to predict interference effects. Inf Sci 633:10–26

Pichon F, Dubois D, Denoeux T (2012) Relevance and truthfulness in information correction and fusion. Int J Approx Reason 53(2):159–175

Schubert J (2011) Conflict management in Dempster-Shafer theory using the degree of falsity. Int J Approx Reason 52(3):449–460

Shafer G (1976) A mathematical theory of evidence, vol 42. Princeton University Press, Princeton

Smarandache F, Dezert J, Tacnet J-M (2010) Fusion of sources of evidence with different importances and reliabilities. In: 2010 13th international conference on information fusion. IEEE, pp 1–8

Xu P, Deng Y, Su X, Mahadevan S (2013) A new method to determine basic probability assignment from training data. Knowl Based Syst 46:69–80

Yang J-B, Xu D-L (2013) Evidential reasoning rule for evidence combination. Artif Intell 205:1–29

Yang Y, Han D, Han C (2013) Discounted combination of unreliable evidence using degree of disagreement. Int J Approx Reason 54(8):1197–1216
Yang J-B, Xu D-L, Xu X, Fu C (2023) Likelihood analysis of imperfect data. IEEE Trans Syst Man Cybern Syst 53(8):5046–5057. https://doi.org/10.1109/TSMC.2023.3262585
Zhang Z-W, Liu Z-G, Martin A, Zhou K (2023) BSC: belief shift clustering. IEEE Trans Syst Man Cybern Syst 53(3):1748–1760. https://doi.org/10.1109/TSMC.2022.3205365
Zhao J, Cheong KH (2023) Obfuscating community structure in complex network with evolutionary divide-and-conquer strategy. IEEE Trans Evol Comput 27(6):1926–1940. https://doi.org/10.1109/TEVC.2023.3242051
Zhao J, Wang Z, Cao J, Cheong KH (2023) A self-adaptive evolutionary deception framework for community structure. IEEE Trans Syst Man Cybern Syst 53(8):4954–4967. https://doi.org/10.1109/TSMC.2023.3240765
Zhou Q, Deng Y (2022a) NPNT: non-oscillating process negation transformation of mass functions and a negation-based discounting method in information fusion. Eng Appl Artif Intell 116:105381
Zhou Q, Deng Y (2022b) Fractal-based belief entropy. Inf Sci 587:265–282
Zhou Q, Deng Y (2022c) Higher order information volume of mass function. Inf Sci 586:501–513. https://doi.org/10.1016/j.ins.2021.12.005
Zhou Q, Deng Y (2023) Generating Sierpinski gasket from matrix calculus in Dempster-Shafer theory. Chaos Solitons Fractals 166:112962
Zhou M, Zhu S-S, Chen Y-W, Wu J, Herrera-Viedma E (2021) A generalized belief entropy with nonspecificity and structural conflict. IEEE Trans Syst Man Cybern Syst 52(9):5532–5545
Zhou L, Cui H, Mi X, Zhang J, Kang B (2022a) A novel conflict management considering the optimal discounting weights using the BWM method in Dempster-Shafer evidence theory. Inf Sci 612:536–552
Zhou Q, Bossé É, Deng Y (2022b) Modeling belief propensity degree: measures of evenness and diversity of belief functions. IEEE Trans Syst Man Cybern Syst. https://doi.org/10.1109/TSMC.2022.3219996

Authors and Affiliations

Qiying Hu[1,2] · Qianli Zhou[1] · Zhen Li[3] · Yong Deng[1,4] · Kang Hao Cheong[5,6]

✉ Qianli Zhou
zhouqianli@std.uestc.edu.cn

✉ Yong Deng
dengentropy@uestc.edu.cn

✉ Kang Hao Cheong
kanghao.cheong@ntu.edu.sg

Qiying Hu
huqiyinguestc@hotmail.com

Zhen Li
zhen.li@pku.edu.cn

[1] Institute of Fundamental and Frontier Science, University of Electronic Science and Technology of China, Chengdu 610054, China

[2] School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

[3] China Mobile Information Technology Center, Beijing 100029, China

[4] School of Medicine, Vanderbilt University, Nashville 37240, USA

[5] Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371, Singapore

[6] College of Computing and Data Science, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore