SHADED: Shapley Value-based Deceptive Evidence Detection in Belief Functions

Haifei Zhang[0000−0003−4488−1631]

Université de technologie de Compiègne, UMR-CNRS 7253 Heudiasyc, France
haifei.zhang@hds.utc.fr

Abstract. Deceptive evidence detection is an important issue in the theory of belief functions, which can be used to solve the problem of conflicts among evidence and to assess the credibility of evidence sources. In this paper, we first define strong and weak deceptive evidence. Then, we propose a deceptive evidence detection approach that directly investigates the process of Dempster's combination rule and decision-making based on the pignistic transformation. It can distinguish between strong and weak deceptive evidence and assess the importance of each piece of evidence. Several numerical examples are used to illustrate the effectiveness and efficiency of our proposed approach.

Keywords: Dempster-Shafer theory · Deceptive evidence detection · Conflict management · Shapley value · Information fusion.

1 Introduction

Dempster-Shafer theory (DST) [2,11], also known as theory of belief functions or theory of evidence, is a mathematical framework designed for combining and reasoning with uncertain information from multiple sources and has been applied in various fields, such as classification [3,4,20,21] and fault diagnosis [16,17].

In DST, combining evidence from different sources is a key issue, which provides the basis for subsequent reasoning. There are several available combination methods, among which Dempster's combination rule (DCR) is a fundamental one [2]. However, when confronted with high-conflicting evidence, using DCR may yield counter-intuitive fusion results, partly because it allocates conflicts totally to the empty set, and partly because of the unreliability of evidence sources. To address this issue, many previous works have been presented. The first strategy attempts to modify and improve the evidence combination rules by reallocating the conflict [5,7,13,18], while the second one focuses on evaluating the reliability of pieces of evidence and discounting them before the combination process [8,19].

Rather than considering all evidence as in the approaches described above, recent studies have started to investigate whether a piece of evidence should be discarded in the fusion process [6]: if a piece of evidence is determined as deceptive, then it should not be combined. In reality, deceptive evidence can arise for a variety of reasons, which may be deliberate deception (e.g., from human sources) or natural causes (e.g., failure of sensors). It will be dangerous if we trust any conclusion derived from the combination of different pieces of evidence without knowing if there are deceptive ones. Therefore, detecting deceptive evidence is crucial, not only to address the problem of conflict among evidence but also to assess the credibility of evidence sources.

As simple and obvious deceptive evidence, the "negation or the complement of true evidence" has been presented and investigated by several researchers [9,15]. However, in real-world applications, deceptive evidence is more complicated than negation. For this issue, Schubert proposed to detect deceptions using the falsity degree of each piece of evidence based on conflict and entropy, as well as their combination [10]. Schubert's approach is limited to considering only the interaction between each piece of evidence and the remaining evidence. Subsequently, Zhou et al. [22] introduced a deception detection method by calculating the Shapley value of each piece of evidence based on the distances between each pair of two subsets of evidence. Cui et al. [1] proposed a method of belief gravitational clustering for this problem, using a similarity measure based on the entropy of evidence and the conflict coefficient between two pieces of evidence. A notable limitation of these methods is their inability to connect the criteria for assessing evidence credibility with Dempster's combination rule.

Recently, Kang and Zhao [6] provided the definition of deceptive evidence and proposed to detect deceptive evidence using reinforcement learning with off-policy Q-learning: rewarding the reduction of uncertainty and the production of reasonable decisions. This approach does not assess the credibility of the evidence, nor does it distinguish between strong deceptive evidence (which can alter the decision) and weak deceptive evidence (which does not change the decision but increases the uncertainty of the decision). In addition, it may be ineffective in identifying deceptive evidence that is difficult to detect.

Considering all the issues mentioned above, we propose a new approach based on Shapley values associated with the uncertainty of combined results for deceptive evidence detection, which is designed for the following objectives:

1. to directly investigate the process of Dempster's combination rule and decision-making based on the pignistic transformation;
2. to distinguish between strong and weak deceptive evidence;
3. to provide an importance assessment for each piece of evidence, which can be further used to conduct a weighted averaging of credible evidence.

The rest of this paper is structured as follows. Section 2 briefly reviews DST and Shapley value. Our proposed method is presented in Section 3 and tested with several examples in Section 4. Finally, a conclusion is drawn in Section 5.

2 Preliminaries

2.1 Theory of belief functions

Let $\mathbf{\Omega} = \{\omega_1, \ldots, \omega_K\}$, $K \geq 2$, be the *frame of discernment*. Given a piece of evidence, the information is represented by a *mass function*, which is a mapping $m : 2^{\mathbf{\Omega}} \rightarrow [0, 1]$, such that $m(\emptyset) = 0$ and $\sum_{A \subseteq \mathbf{\Omega}} m(A) = 1$.

For any subset $A \subseteq \mathbf{\Omega}$, the uncertainty of the proposition that the true state lies in A can be quantified by the degrees of *belief* $Bel(A)$ and *plausibility* $Pl(A)$:

$$Bel(A) = \sum_{B \subseteq A} m(B) \text{ and } Pl(A) = \sum_{B \cap A \neq \emptyset} m(B). \tag{1}$$

$Bel(A)$ and $Pl(A)$ measure the support (belief) and compatibility (plausibility), respectively, associated with the proposition that the truth lies in A.

Given two independent mass functions m_1 and m_2 defined on the same problem, they can be combined via *Dempster's combination rule* (DCR) [2]:

$$m(A) = (m_1 \oplus m_2)(A) = \frac{1}{1-\mathcal{K}} \sum_{B \cap C = A} m_1(B) m_2(C), \quad \forall A \subseteq \mathbf{\Omega}, \tag{2}$$

where $\mathcal{K}$ is the degree of conflict between the two mass functions, defined as:

$$\mathcal{K} = \sum_{B \cap C = \emptyset} m_1(B) m_2(C). \tag{3}$$

Dempster's combination rule is also called the *orthogonal sum* of m_1 and m_2, and requires that the mass functions to be combined are independent and their conflict is smaller than one. This operation is commutative and associative, which makes it possible to sequentially combine a series of evidence in any order.

The combined mass function m can be transformed into a probability distribution through the *pignistic transformation* [14]:

$$BetP(\omega_k, m) = \sum_{A \subseteq \mathbf{\Omega}, \omega_k \in A} \frac{m(A)}{|A|}, \quad \forall \omega_k \in \mathbf{\Omega}, \tag{4}$$

in which the mass of focal sets is equally assigned to their elements. Then, the state of nature that maximizes the pignistic probability is taken as the decision.

2.2 Shapley values

The *Shapley value* is a concept from cooperative game theory, which offers a fair way to distribute the total gains to the players based on their individual contributions to the collective outcome of a coalitional game [12]. Formally, a coalitional game is defined as $(\mathbf{N}, v)$ where $\mathbf{N}$ is a set of n players and payoff function $v : 2^{\mathbf{N}} \to \mathbb{R}$ maps the subsets of players (coalitions) to a real number. The Shapley value (marginal contribution) of player i is defined as:

$$\varphi_i = \sum_{S \subseteq \mathbf{N} \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!} \left[v(S \cup \{i\}) - v(S) \right]. \tag{5}$$

3 SHADED: SHApley value-based Deceptive Evidence Detection

3.1 Definition of deceptive evidence

Following the definition of deceptive evidence in [6], we refine it into strong deceptive evidence and weak deceptive evidence. This refinement allows us to better understand deceptive evidence, as well as to better manage them.

Definition 1 (Strong deceptive evidence). *For a given set of credible evidence and a certain fusion system (i.e., Dempster's combination rule), strong deceptive evidence is a piece of or a group of evidence that can alter the decision.*

Definition 2 (Weak deceptive evidence). *For a given set of credible evidence and a certain fusion system (i.e., Dempster's combination rule), weak deceptive evidence is a piece of or a group of evidence that can not alter the decision but can increase the uncertainty associated with the decision.*

In the next section, we will delve into a Shapley value-based approach designed to identify deceptive evidence, as well as to distinguish between strong and weak deceptive evidence based on the above definitions.

3.2 Detection of deceptive evidence

In this paper, for a given collection of n mass functions $\mathbf{M} = \{m1, m2, \ldots, m_n\}$, we consider that the combined decision is the state of nature with the highest pignistic probability associated with the combined mass function via DCR:

$$cd = \arg\max_{\omega \in \mathbf{\Omega}} BetP(\omega, \overset{n}{\underset{i=1}{\oplus}}[m_i]). \tag{6}$$

However, the combined decision may be unreasonable because of the conflict in evidence. Therefore, the first step of our method is to determine the reasonable decision by conflict resolution. We apply Murphy's approach [8], which can maintain the majority opinion, and attain the convergence of the certainty on the decision. It combines the n identical averaged mass functions using DCR:

$$rd = \arg\max_{\omega \in \mathbf{\Omega}} BetP(\omega, \overset{n}{\underset{i=1}{\oplus}}[m]), \text{ where } m = \frac{1}{n}\sum_{i=1}^{n} m_i. \tag{7}$$

The second step is to calculate the Shapley value of each mass function. For a coalition (set) of mass functions S, the payoff function v is defined as the pignistic probability associated with the reasonable decision. Thus the Shapley value of the mass function $m_i, i = 1, \ldots, n$, can be calculated by

$$\varphi_i = \sum_{S \subseteq \mathbf{M} \setminus \{m_i\}} \frac{|S|!(n-|S|-1)!}{n!} \left[BetP(rd, \underset{m_j \in S'}{\oplus}[m_j]) - BetP(rd, \underset{m_j \in S}{\oplus}[m_j]) \right], \tag{8}$$

where $S' = S \cup \{m_i\}$. Moreover, we set $BetP(rd, \emptyset) = 1/K$ with K the number of states of nature, thus $\varphi_0 = 1/K$. Here, φ_0 represents the utility (payoff) of the evidence combination (coalitional game) when there is no evidence provided or only a piece of vacuous evidence is provided, which is equivalent to the case of making a decision with total ignorance.

Based on the above setting and the properties of Shapley value, we can distinctly draw the following propositions:

1. *efficiency*: the pignistic probability associated with the combined mass function and the reasonable decision is equal to the sum of the Shapley values of the given mass functions, i.e., $BetP(rd, \overset{n}{\underset{i=1}{\oplus}}[m_i]) = \sum_{i=0}^{n} \varphi_i$;
2. *symmetry*: two identical mass functions have the same Shapley value, i.e., if $m_i = m_j$, then $\varphi_i = \varphi_j$;
3. *dummy*: the Shapley value of a vacuous mass function is zero because it brings no contribution to any coalition in the evidence combination process, i.e., if $m_i(\mathbf{\Omega}) = 1$, then $\varphi_i = 0$.

The third step involves determining the nature of each piece of evidence. The set of credible evidence (having positive or zero contributions to the reasonable decision), denoted as CE, includes all evidence with non-negative Shapley values: $CE = \{m_i \mid \varphi_i \geq 0, m_i \in \mathbf{M}\}$. Consequently, the set of deceptive evidence (DE) comprises all evidence having negative contributions (with negative Shapley values) to the reasonable decision, i.e., $DE = \mathbf{M} \setminus CE$. Next, we differentiate between strong and weak deceptive evidence. If the decision derived from combining all evidence matches the reasonable decision, it implies the absence of strong deceptive evidence (SDE), rendering all deceptive evidence as weak (WDE), i.e., $WDE = DE$ and $SDE = \emptyset$. Otherwise, there must be strong deceptive evidence and it necessitates examining all subsets of DE by combining them with CE to determine the piece of or the group of strong deceptive evidence. In this case, $WDE = DE \setminus SDE$.

The Alg. 1 summarises the three steps described above[1]. The Shapley values of the credible evidence can be used to calculate the weights for credible evidence:

$$\mathrm{w}_j = \frac{\varphi_j}{\sum\limits_{m_i \in CE} \varphi_i}, \ \forall\ m_j \in CE. \tag{9}$$

Then, the final fusion result is

$$p(\omega_k) = BetP(\omega_k, \overset{|CE|}{\underset{i=1}{\oplus}}[m]), \text{ where } m = \sum_{m_j \in CE} \mathrm{w}_j \times m_j. \tag{10}$$

4 Experiments

In this section, we provide several numerical examples to illustrate the effectiveness and efficiency of our proposed deceptive evidence detection method.

[1] Our code is available on GitHub: https://github.com/Haifei-ZHANG/shaded.

Algorithm 1: SHApley value-based Deceptive Evidence Detection

Input: A set of n mass functions $\mathbf{M} = \{m_1, \ldots, m_n\}$
Output: Set of credible evidence, strong and weak evidence
Calculate the combined decision cd via Eq. (6)
Calculate the reasonable decision rd via Eq. (7)
for $i = 1, \ldots, n$ **do**
  Calculate Shapley value φ_i via Eq. (8)
$CE = \{m_i \mid \varphi_i \geq 0, m_i \in \mathbf{M}\}$ // set of credible evidence
$DE = \mathbf{M} \setminus CE$ // set of deceptive evidence
if $cd = rd$ **then**
  $WDE = DE$ // set of weak deceptive evidence
  $SDE = \emptyset$. // set of strong deceptive evidence
else
  $SDE = \{\}$
  for *all* $S \subseteq DE$ **do**
    $S' = CE \cup S$
    if $rd \neq \arg\max_{\omega \in \mathbf{\Omega}} BetP(\omega, \bigoplus_{m_j \in S'} [m_j])$ **then**
      $SDE = SDE \cup S$
  $WDE = DE \setminus SDE$
Return (CE, SDE, WDE)

4.1 Effectiveness comparison

The first example is drawn from the article of Cui et al. [1] for the fault diagnosis. Five sensors can provide evidence about three kinds of fault $\mathbf{\Omega} = \{\omega_1, \omega_2, \omega_3\}$. The set of mass functions and their Shapley values are listed in Table 1.

Table 1. Mass functions and Shapley values of the first example.

	$m(\{\omega_1\})$	$m(\{\omega_2\})$	$m(\{\omega_3\})$	$m(\{\omega_1, \omega_3\})$	φ_i
m_1	0.41	0.29	0.30	0.00	0.014
m_2	0.00	0.90	0.10	0.00	-0.743
m_3	0.58	0.07	0.00	0.35	0.135
m_4	0.55	0.10	0.00	0.35	0.126
m_5	0.60	0.10	0.00	0.30	0.136

For the first example, the pignistic transformation from the result of Dempster's combination rule finds the combined decision is ω_3, while the reasonable decision is ω_1, which means there must be strong deceptive evidence. According to the calculated Shapley values, we can determine that m_2 is strong deceptive evidence, which is the same as the detection made by the method of Cui et al. After deleting deceptive evidence and re-conducting combination, we find in Table 2 that our approach can reach a higher pignistic probability for the reasonable decision ω_1 than other methods, i.e., the decision is more certain.

Table 2. Fusion results of different methods for the first example.

	Method	$p(\omega_1)$	$p(\omega_2)$	$p(\omega_3)$	Reasonable
Consider all mass functions	DCR [2]	0.0000	0.1422	0.8578	No
	Murphy [8]	0.9636	0.0210	0.0154	Yes
Delete $m2$	Cui et al. [1]	0.9684	0.0007	0.0309	Yes
	Ours	**0.9893**	**0.0001**	**0.0106**	**Yes**

The second example is used to compare with the reinforcement learning method [6]. The set of mass functions and their Shapley values are listed in Table 3.

Table 3. Mass functions and Shapley values of the second example.

	$m(\{\omega_1\})$	$m(\{\omega_2\})$	$m(\{\omega_3\})$	$m(\mathbf{\Omega})$	φ_i
m_1	0.30	0.60	0.00	0.10	-0.023
m_2	0.70	0.00	0.00	0.30	0.206
m_3	0.65	0.15	0.00	0.20	0.170
m_4	0.75	0.00	0.05	0.20	0.238
m_5	0.05	0.45	0.50	0.00	-0.368
m_6	0.05	0.50	0.45	0.00	-0.376

For the second example, the decision derived by Dempster's combination rule and the pignistic transformation is ω_2, while the reasonable decision is ω_1. According to the obtained Shapley values, our method can detect that m_1 is a piece of weak deceptive evidence (it can not flip the decision but reduce the certainty of the decision) and $\{m_5, m_6\}$ are strong deceptive evidence. However, Kang and Zhao's method failed to detect m_1 as a piece of deceptive evidence. Table 4 shows that our approach can make a reasonable decision more certain.

Table 4. Fusion results of different methods for the second example.

	Method	$p(\omega_1)$	$p(\omega_2)$	$p(\omega_3)$	Reasonable
Consider all mass functions	DCR [2]	0.1814	0.7428	0.0758	No
	Murphy [8]	0.8230	0.1555	0.0215	Yes
Delete $\{m_5, m6\}$	Kang et Zhao [6]	0.9566	0.0413	0.0021	Yes
Delete $\{m_1, m_5, m6\}$	Ours	**0.9762**	**0.0147**	**0.0091**	**Yes**

4.2 Efficiency comparison

In this experiment, we compared our method with Kang and Zhao's method in terms of computational efficiency. For Kang and Zhao's method, the default parameter setting in [6] was used. The experiments were conducted on a Windows 11 system, powered by an Intel i5-12450H CPU. All used mass functions, of which there are two pieces of deceptive evidence, are defined on $\mathbf{\Omega} = \{\omega_1, \omega_2, \omega_3\}$ with one to four focal sets. Fig. 1 shows that our method is much more efficient, regardless of the number of mass functions.

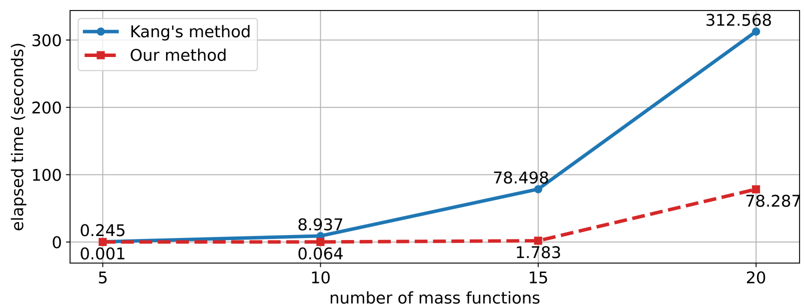

Fig. 1. Time required to detect deception as a function of the number of mass functions.

5 Conclusion

In this paper, we have defined strong and weak deceptive evidence in the information fusion of belief functions and proposed a new Shapley value-based deceptive evidence detection method, which provides more insights directly linked with the combination rule and the decision-making strategy. This method can distinguish between strong and weak deceptive evidence and assess the credibility of the evidence, which can be further used to conduct a weighted averaging process. Some numerical examples have demonstrated that our method is outperforming state-of-the-art methods in terms of detection effectiveness and efficiency.

In future works, we intend to optimize the efficiency of the proposed algorithm (especially with larger frames of discernment and a larger number of mass functions) and apply it in real-world applications. Moreover, the determination of the reasonable decision for a given set of mass functions is also an interesting open issue to investigate.

References

1. Cui, H., Zhang, H., Chang, Y., Kang, B.: Bgc: Belief gravitational clustering approach and its application in the counter-deception of belief functions. Engineering Applications of Artificial Intelligence **123**, 106235 (2023)

2. Dempster, A.P.: Upper and Lower Probabilities Induced by a Multivalued Mapping. The Annals of Mathematical Statistics **38**, 325–339 (1967)
3. Denoeux, T.: A k-nearest neighbor classification rule based on dempster-shafer theory. IEEE transactions on systems, man, and cybernetics **25**(5), 804–813 (1995)
4. Denoeux, T.: A neural network classifier based on dempster-shafer theory. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans **30**(2), 131–150 (2000)
5. Dubois, D., Prade, H.: Representation and combination of uncertainty with belief functions and possibility measures. Computational intelligence **4**(3), 244–264 (1988)
6. Kang, B., Zhao, C.: Deceptive evidence detection in information fusion of belief functions based on reinforcement learning. Information Fusion **103**, 102102 (2024)
7. Lefevre, E., Colot, O., Vannoorenberghe, P.: Belief function combination and conflict management. Information fusion **3**(2), 149–162 (2002)
8. Murphy, C.K.: Combining belief functions when evidence conflicts. Decision support systems **29**(1), 1–9 (2000)
9. Pichon, F., Dubois, D., Denoeux, T.: Relevance and truthfulness in information correction and fusion. International Journal of Approximate Reasoning **53**(2), 159–175 (2012)
10. Schubert, J.: Counter-deception in information fusion. International Journal of Approximate Reasoning **91**, 152–159 (2017)
11. Shafer, G.: A mathematical theory of evidence. Princeton University Press (1976)
12. Shapley, L.S.: Notes on the n-person game—ii: The value of an n-person game (1951)
13. Smets, P.: The combination of evidence in the transferable belief model. IEEE Transactions on pattern analysis and machine intelligence **12**(5), 447–458 (1990)
14. Smets, P.: Decision making in the tbm: the necessity of the pignistic transformation. International journal of approximate reasoning **38**(2), 133–147 (2005)
15. Smets, P.: Managing deceitful reports with the transferable belief model. In: 2005 7th International Conference on Information Fusion. vol. 2, pp. 7–pp. IEEE (2005)
16. Xiao, F., Cao, Z., Jolfaei, A.: A novel conflict measurement in decision-making and its application in fault diagnosis. IEEE Transactions on Fuzzy Systems **29**(1), 186–197 (2020)
17. Xu, Y., Li, Y., Wang, Y., Zhong, D., Zhang, G.: Improved few-shot learning method for transformer fault diagnosis based on approximation space and belief functions. Expert Systems with Applications **167**, 114105 (2021)
18. Yager, R.R.: On the dempster-shafer framework and new combination rules. Information sciences **41**(2), 93–137 (1987)
19. Yong, D., WenKang, S., ZhenFu, Z., Qi, L.: Combining belief functions based on distance of evidence. Decision support systems **38**(3), 489–493 (2004)
20. Zhang, H., Quost, B., Masson, M.H.: Cautious decision-making for tree ensembles. In: European Conference on Symbolic and Quantitative Approaches with Uncertainty. pp. 3–14. Springer (2023)
21. Zhang, H., Quost, B., Masson, M.H.: Cautious weighted random forests. Expert Systems with Applications **213**, 118883 (2023)
22. Zhou, L., Cui, H., Huang, C., Kang, B., Zhang, J.: Counter deception in belief functions using shapley value methodology. International Journal of Fuzzy Systems **24**(1), 340–354 (2022)