Quantifying confidence of safety cases with belief functions

Yassir Idmessaoud, Didier Dubois, Jérémie Guiochet

Quantifying confidence of safety cases with belief functions

Yassir Idmessaoud[1], Didier Dubois[2], and Jérémie Guiochet[1]

[1] LAAS-CNRS, University of Toulouse, France
[2] IRIT, University of Toulouse, France

Abstract. Structured safety argument based on graphical representations such as GSN (Goal Structuring Notation) are used to justify the certification of critical systems. However, such approaches do not deal with uncertainties that might affect the merits of arguments. In the recent past, some authors proposed to model the confidence in such arguments using Dempster-Shafer theory. It enables us to determine the confidence degree in conclusions for some basic GSN patterns. In this paper, we refine this approach and improve the elicitation method for expert opinions used in previous papers.

Keywords: Safety cases · Goal Structuring Notation (GSN) · Dempster–Shafer theory (DST) · Belief elicitation · Confidence assessment.

1 Introduction

GSN (Goal Structuring Notation) is a graphical formalism used to represent argument structures (assurance cases, dependability cases, etc). Originally, GSN structures do not include a representation of uncertainty in the arguments. Several independent works, e.g., [2, 13] proposed to augment this approach to argumentation with confidence assessment methods. They design numerical confidence propagation models for some GSN patterns. However, in [2, 15], the data collection method, that enable these mathematical models to be fed with initial confidence values, allowing the computation of the overall confidence in the system, needs improvement. The previous elicitation methods also present some technical defects. In this paper, after reviewing previous work in Section 2, we introduce an extensive confidence propagation method in Section 3, starting with a brief description of argument types. In Section 4, we present an improved expert opinion elicitation method. Finally, in section 5, we illustrate our approach on a small example.

2 Related work

The issue of confidence assessment in argument structures has been addressed in multiple works. Safety cases, which represent one type of such structures, aim at proving the safety of a system by producing several pieces of documented evidence. In this paper, we focus on safety cases modeled by the so-called Goal Structuring Notation (GSN) defined in [9]. Safety argumentation is a core activity in safety critical systems development. Such an argumentation can be carried out using structured notations. It decomposes the safety requirements of the system into elementary pieces as presented in Figure 1, called *goals*. *Strategies* are the components that justify this decomposition. Each goal is supported by one or multiple pieces of evidence called *Solutions*. The example of GSN in Figure 1, is a classical pattern of an argument that treats the hazards existing in a system, and listed in the context box. However, many other patterns exist. This type of representation does not consider the uncertainty that may pervade each premise or the support relation between solutions and goals. Moreover, it is important to note that GSN models are non formal, and no explicit formal logical relations is expressed between elements.

In [6], we compared some works that deal with confidence assessment in GSN and give some recommendation to improve these methods. For instance, we showed why it is more adequate to use implication instead of equivalence (used in [13–15]) to represent argument types. We also discussed why Dempster rule of combination is more suitable for combining evidence, in our case, than other methods used in [2] for instance. Building such a confidence model relies

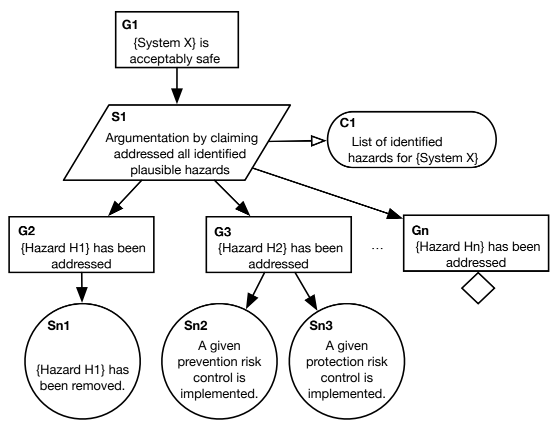

Fig. 1. GSN example adapted from Hazard Avoidance Pattern [10]

on input values, usually provided by experts in qualitative form, and transformed into quantitative values. Such an activity could be called "Expert opinion elicitation". This method is more often used with probabilistic models. For instance, in [3], authors used an expert elicitation procedure in a risk assessment approach in fault trees. However, it can also be used in evidence theory. Ben Yaghlane et al. [16], generate belief functions from a preference relation between events provided by experts. In relation to our framework, few authors augmented their confidence assessment method by such a data elicitation procedure in order to provide quantitative values for their models. Only some authors such as [2, 15] used an elicitation method that transforms expert opinions given in the form of qualitative values, into quantitative ones.

Uncertainty propagation can be addressed by standard existing belief function software based on results in [11] (e.g., the belief function machine implemented in MatLab), but the GSNs we study have a particular tree-like structure that enable an explicit symbolic calculation of the belief function on the conclusion space. The explicit formulas, that can be obtained from approaches such as the one we propose, make the calculation more efficient and we can predict the effect of changing selected inputs, thus better explaining the obtained results, and validating the approach.

3 Argument types

In this section, we define the argument types used in our confidence assessment method. Remember that an argument type represents the relationship between premises and a conclusion. For instance, in Figure 1, (G_2) to (G_n) are the premises of (G_1) and they are all necessary to support it. This logical relation could be assimilated to the strategy component in a GSN (e.g., S1). When adopting a logical viewpoint [6], we then speak of a rule. Unlike the types of arguments proposed in [14], which use the equivalence connective to model such rules, we decided to break down this equivalence into two implications. Each implication brings a single piece of information about the conclusion, given premises. For instance, in the case of a single premise (P) supporting one conclusion (C), $P \Rightarrow C$ (that we call *direct rule*) can only infer the acceptance of the conclusion if the premise is true. On the other hand, the rule $\neg P \Rightarrow \neg C$ (that we call *reverse rule*) can only infer the rejection of the conclusion when the premise is false. We believe that, when assessing uncertainty of the relationship between P and C, this separate handling makes the resulting models more accurate and easy to assess. We also decided to associate a simple support mass function to each rule to avoid dependence in the confidence between premises. Below are uncertainty propagation formulas for various argument types; all calculation details are available in [7].

D-Arg (Disjunctive): In this situation, each premise can support alone the whole conclusion. We formally define this argument by : $\wedge_{i=1}^{n}(p_i \Rightarrow C)$. Rules that infer the rejection of the conclusion ($\neg C$) can be deduced from this argument type by reversing this rule to obtain : $(\wedge_{i=1}^{n}\neg p_i) \Rightarrow \neg C$. To get formulas (1) and (2), we first assign to each rule a simple support mass function (resp. m_{dir}^{i} for $p_i \Rightarrow C$ and m_{rev} for the reverse rule). We also assign one mass function m_p^i to each premise p_i. This function uses three masses on p_i, $\neg p_i$ and the tautology ($\top$) summing to one. Then, we combine, using DS rule of combination, all masses on premises together ($m_p = m_p^1 \oplus m_p^2 \oplus ... \oplus m_p^n$) and masses on rules together ($m_r = [m_{dir}^1 \oplus m_{dir}^2 \oplus ... \oplus m_{dir}^n] \oplus m_{rev}$). Finally, we combine the resulting masses on the rules and premises ($m = m_p \oplus m_r$). We obtain degrees of belief and disbelief in C:

$$Bel_C(C) = 1 - \prod_{i=1}^{n}[1 - Bel_p^i(p_i)Bel_{\Rightarrow}^i(p_i \Rightarrow C)] \tag{1}$$

$$Disb_C(C) = Bel_{\Leftarrow}(\wedge_{i=1}^{n}[\neg p_i] \Rightarrow \neg C)\prod_{i=1}^{n} Disb_p^i(p_i). \tag{2}$$

We can notice that (1) expresses a "*Multivalued Disjunction*". To have maximal belief in the conclusion, it is enough that the degree of belief in one single premise equals 1 (assuming that the mass on the direct rule also equals 1). Formula (2), in contrast, expresses a ''Multivalued Conjunction". To have a maximal disbelief in the conclusion, all the disbelief degrees on premises should be equal to 1, supposing that the mass on the reverse rule equals 1 too. We can also notice that, when $Bel_C(C)$ is maximal, $Disb_C(C)$ is minimal.

C-Arg (Conjunctive): This argument type describes the situation when two premises or more are jointly needed to support a conclusion. Following the same reasoning as in the previous type, we define it formally by two rules : $(\wedge_{i=1}^{n} p_i) \Rightarrow C$ and its reverse $\wedge_{i=1}^{n}(\neg p_i \Rightarrow \neg C)$. Following the same calculation in the disjunctive type, we get the formulas below :

$$Bel_C(C) = Bel_{\Rightarrow}([\wedge_{i=1}^{n} p_i] \Rightarrow C)\prod_{i=1}^{n} Bel_p(p_i) \tag{3}$$

$$Disb_C(C) = 1 - \prod_{i=1}^{n}[1 - Disb_p^i(p_i)Bel_{\Leftarrow}^i(\neg p_i \Rightarrow \neg C)] \tag{4}$$

We can notice that, in contrast with formulas obtained for D-Arg, (3) and (4) respectively express a "*Multivalued Conjunction*" and "*Multivalued Disjunction*".

H-Arg (Hybrid): This argument describes the case when it is difficult to choose between the conjunctive or disjunctive types. Each premise supports the conclusion to some extent, and the conjunction of the premises does it to a larger extent. We obtain degrees of belief and disbelief in C:

$$\begin{aligned} Bel_C(C) =& Bel_{\Rightarrow}([\wedge_{i=1}^{n} p_i] \Rightarrow C) \times \prod_{i=1}^{n} Bel_p^i(p_i)[1 - Bel_{\Rightarrow}^i(p_i \Rightarrow C)] \\ &+ \{1 - \prod_{i=1}^{n}[1 - Bel_p^i(p_i)Bel_{\Rightarrow}^i(p_i \Rightarrow C)]\}. \end{aligned} \tag{5}$$

$$\begin{aligned} Disb_C(C) =& Bel_{\Leftarrow}([\wedge_{i=1}^{n} \neg p_i] \Rightarrow \neg C) \times \prod_{i=1}^{n} Disb_p^i(p_i)[1 - Bel_{\Leftarrow}^i(\neg p_i \Rightarrow \neg C)] \\ &+ \{1 - \prod_{i=1}^{n}[1 - Disb_p^i(p_i)Bel_{\Leftarrow}^i(\neg p_i \Rightarrow \neg C)]\}. \end{aligned} \tag{6}$$

We can notice from (5) and (6) that these formulas subsume those in conjunctive and disjunctive types. On the one hand, if masses on $p_i \Rightarrow C$ are zero, it becomes the formula of the conjunctive type. On the other hand, if the mass on $[\wedge_{i=1}^{n} p_i] \Rightarrow C$ is zero, we get the disjunctive type formula.

This argument provides a general framework that allow as to calculate belief and disbelief values in different situations. D-Arg and C-Arg represent extreme cases where the value of some rules is null.

Note that moving away from these extreme cases may lead to encounter situations of conflict. A contradiction may appear when we have opposite opinion about two premises along corresponding direct and reverse rules. Formally, it always takes the form of a combination of four items of the form: $\{p_i, p_i \Rightarrow C, \neg p_j, \neg p_j \Rightarrow \neg C\}$. The sum $(Bel_C(C) + Disb_C(C))$ is then greater than 1. This may indicate something wrong in the GSN or in the way the experts replied questions, or yet on the reported experiments. Equation (7) represents the conflict calculation formula :

$$m(\bot) = \sum_{i=1, j\neq i}^{n} [Bel_p^i(p_i) Bel_{\Rightarrow}^i(p_i \Rightarrow C) \times Disb_p^j(p_j) Bel_{\Leftarrow}^j(\neg p_j \Rightarrow \neg C)] \quad (7)$$

To address this issue, we choose to subtract the mass of the conflict $m(\bot)$ from $Bel_C(C)$ and $Disb_C(C)$ in (5) and (6), and get contradiction-free degrees $bel(C) = Bel_C(C) - m(\bot)$ and $disb_C(C) = Disb_C(C) - m(\bot)$. We choose not to normalize the results (dividing by $1 - m(\bot)$) as proposed in the usual DS rule of combination because this operation will eliminate the conflict and proportionally increase the contradiction-free degrees of beliefs and disbelief $bel_C(C)$ and $disb_C(C)$ in a misleading way in the case of strong conflict. On the other hand, keeping $m(\bot)$ and subtracting it from $Bel_C(C)$ and $Disb_C(C)$ will legitimately increase uncertainty (i.e., $bel_C(C) + disb_C(C)$ is small) and show that the system is not that safe because of the presence of a conflict.

4 Expert opinion elicitation

In the previous section, we defined three argument types and proposed analytical formulas to calculate the belief and disbelief degrees in conclusions. However, using these models requires the presence of two types of information: belief degrees in premises (e.g., Bel_p^i) and the belief degrees for rules (e.g., $Bel_{\Rightarrow}^i$). In this section, we are first going to see how we can transform an expert opinion about a premise into belief, uncertainty and disbelief degrees. Then, in the second part, we provide some hints on how we can identify masses on rules.

4.1 Elicitation of belief and disbelief on premises

In order to directly obtain belief and disbelief degrees in a premise p from an expert, authors in [2] consider asking two pieces of information : one, called decision index $Dec(p)$, describes which side the expert leans towards, acceptance or rejection of p; the other, called confidence $Conf(p)$, reflects the amount of information an expert possesses that can justify his opinion. Namely $Dec(p) = 1$ (resp., 0) indicates the certainty that p is true (resp false), while $Conf(p) = 1$ (resp., 0) indicates the expert has full (resp. no) information supporting the choice of $Dec(p)$. This is represented on Fig. 2.

The problem is then to define the belief and disbelief degrees in a proposition p in terms of $Dec(p)$ and $Conf(p)$. In [2], it is proposed to let $Bel(p) = Dec(p) \cdot Conf(p)$ and $Disb(p) = (1 - Dec(p)) \cdot Conf(p)$, which implies a natural result:

$$Conf(p) = Bel(p) + Disb(p). \tag{8}$$

However, it also implies that $Dec(p) = \frac{Bel(p)}{Bel(p)+Disb(p)}$. Note that this formula, presents a discontinuity in case of no information ($Bel(p) + Disb(p) = 0$). The expression is then completed by assuming $Dec(p) = 1$ [2] or 0 [15], in this case, which sounds arbitrary.

It is more convincing to use the *Pignistic* transform [12] that turns a mass function m on a set Ω (the frame of discernment) into a probability, changing the focal sets into uniform distributions. When $\Omega = \{p, \neg p\}$ has two possible states, $Dec(p)$ is the midpoint between belief and plausibility of p, which reads :

$$Dec(p) = \frac{1 + Bel(p) - Disb(p)}{2} \tag{9}$$

Note that when $Bel(p) = Disb(p) = 0$, we get $Dec(p) = 1/2$.

Some authors suggest to define $Dec(p)$ by renormalising the pair $(Pl(p), Pl(\neg p)$, where $Pl(p) = 1 - Bel(p)$, dividing them by $Pl(p) + Pl(\neg p)$ (plausibility transformation method [1]). This method is in agreement with Dempster rule of combination. However, we do not get the midpoint between belief and plausibility, which is intuitively surprizing, and in case of more than 3 elements in the frame, such a transformation may give probability values outside the range $[Bel, Pl]$[5].

Using equations (8) and (9) and the knowledge of $Dec(p)$ and $Conf(p)$, we can calculate belief and disbelief values : $Bel(p) = \frac{Conf(p)-1}{2} + Dec(p)$, $Disb(p) = \frac{Conf(p)+1}{2} - Dec(p)$. Viewing $Bel(p)$ as a lower probability, the pignistic transformation computes the center of gravity of the convex set of probabilities $\{P : P \geq Bel\}$.

However, the pignistic transform also presents one issue for the elicitation procedure. Some values of the pair $(Dec, Conf)$ provided by the expert may lead to negative values of belief $Bel(p)$ or disbelief degrees $Disb(p)$, which makes no sense. This is because there are constraints relating $Conf(p)$ and $Dec(p)$: (8) and (9) imply $1 - Conf(p) \leq \min(2Dec(p), 2(1 - Dec(p))$, which is known as Josang triangle [8]. To fix this problem, we can express the range of $Dec(p)$ for a given confidence level as:

$$\frac{1 - Conf(p)}{2} \leq Dec(p) \leq \frac{1 + Conf(p)}{2} \tag{10}$$

For instance, a strong decision (full acceptance or rejection) should only be made when we have a very high level of confidence, since when $Conf(p) = 1$,

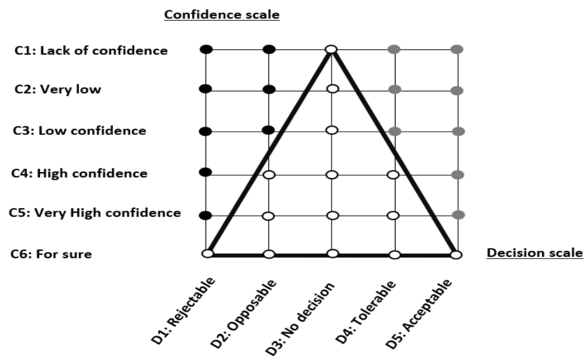

Fig. 2. Expert opinion extraction matrix

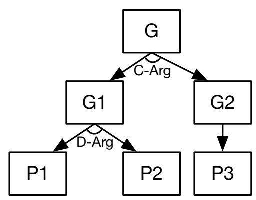

Fig. 3. GSN toy example

$Dec(p)$ is not constrained and ranges on [0, 1]. In contrast, under ignorance $Conf(p) = 0$ enforces $Dec(p) = 1/2$. So, when the pair $(Dec(p), Conf(p))$ is situated outside the triangle, and $Dec(p) < \frac{1-Conf(p)}{2}$ (rejection: black dots on Fig. 2), we set $Dec(p) = \frac{1-Conf(p)}{2}$. On the other hand, when $Dec(p) > \frac{1+Conf(p)}{2}$ (acceptance: grey dots on Fig. 2), we set $Dec(p) = \frac{1+Conf(p)}{2}$. Choosing scales for $(Conf, Dec)$ and translating such pairs into numerical degrees is not trivial, we thus make the equidistance assumption for simplicity and to be comparable to previous works.

4.2 Determination of belief weights for rules

Now, consider the mass functions for rules. Unlike belief and disbelief degrees in premises, belief degrees in rules are more difficult to obtain directly from an expert. Remember that a rule is representing a support relation between a conclusion and its premises. As a first approach, Wang et al. [14] proposed to exploit this relation so as to identify these masses. To this end, they propose certain values of the pair $(Dec, Conf)$ on premises, as inputs for the model, and ask the expert his opinion about the conclusion using the matrix of Figure 2. Then, for each type of argument, they use a non-linear least square method to estimate the values of parameters (belief in rules).

A second approach, which is under study, aims to determine these masses and the argument type (C, D or H-Arg) through a series of questions, assuming clear-cut knowledge for premises ($Bel = 1$ or 0 and $Disb = 1$ or 0). For example, *If decision and confidence on a premise are maximal (acceptable for sure), what is your decision and confidence on the conclusion ?* When belief or disbelief in premises are maximal, the mass on the conclusion is the mass of the rule. For instance, if in (3) we let $\forall i, Bel(p_i) = 1$, then $Bel_C(C) = Bel([\wedge_{i=1}^{n} p_i] \Rightarrow C)$ is obtained from the expert.

5 Toy example

We apply our approach in this section to a simple GSN example presented in Figure 3, including two types of arguments. It presents a top goal (G) supported by two sub-goals (G_1) and (G_2). (G_1) is supported by two premises (P_1) and (P_2). On the other hand, (G_2) is supported by a single premise (P_3). For simplicity, we chose a *C-Arg* for the argument type used to calculate (G) and a *D-Arg* in the calculation of (G_1).

In order to see how the variation of belief and disbelief degrees in the premises affect the conclusion, we show six different configurations in table 1, where we set for each premise (P_1, P_2 and P_3) a qualitative pair (decision, confidence) (and the corresponding pair (Bel, Disb)), and calculate the conclusion (G) by means of formulas (1), (2), (3) and (4). We also set the values of the masses on the rules to 1. As a result, the values of belief and disbelief in the conclusion will depend only on the masses on the premises.

Table 1. Qualitative (decision, confidence) and quantitative (belief, disbelief) pairs for the example (see Fig. 2 for the meaning of symbols)

	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}
P_1	(R;C_6)	(A;C_5)	(A;C_5)	(T;C_5)	(A;C_5)	(T;C_6)
	(0 ; 1)	(0.8 ; 0)	(0.8 ; 0)	(0.65 ; 0.15)	(0.8 ; 0)	(0.75 ; 0.25)
P_2	(R;C_5)	(A;C_6)	(R;C_5)	(T;C_6)	(A;C_6)	(T;C_1)
	(0 ; 0.8)	(1 ; 0)	(0 ; 0.8)	(0.75 ; 0.25)	(1 ; 0)	(0 ; 0)
P_3	(R;C_6)	(A;C_6)	(A;C_6)	(O;C_6)	(A;C_2)	(T;C_6)
	(0 ; 1)	(1 ; 0)	(1 ; 0)	(0.25 ; 0.75)	(0.2 ; 0)	(0.75 ; 0.25)
G	(R;C_6)	(A;C_6)	(A;C_5)	(O;C_6)	(A;C_2)	(T;C_4)
	(0 ; 1)	(1 ; 0)	(0.8 ; 0)	(0.23 ; 0.76)	(0.2 ; 0)	(0.56 ; 0)

We can notice, on Table 1, that when we have either three rejectable or three acceptable premises with high levels of confidence (1^{st} and 2^{nd} columns), the models maintain the same decision with the same high level of confidence. On the other hand, when we have divergent opinions on the premises, either by opposite decisions (3^{rd} and 4^{th} columns) or opposite confidence levels (5^{th} and 6^{th} columns), the results will depend on the nature of the argument involved.

In the 3^{rd} and 6^{th} column, decision levels (resp. acceptable and tolerable) were maintained because the divergence is located in a D-Arg. Due to its disjunctive nature, this argument favors the propagation of the premises that maximally support the conclusion. However, confidence levels were slightly decreased because of a C-Arg, which cumulates the uncertainty present in each premise and propagates it to the conclusion. In the 4^{th} and 5^{th} columns the divergence is located in a C-Arg. Unlike D-Arg, this argument favors the propagation of the premises that support the conclusion with the least strength. Thus, we end up with a mildly negative ("opposable") decision level in the 4^{th} column and a very low level of confidence in the 5^{th} column.

6 Conclusion

In this article, we propose a method for confidence assessment in GSN. It covers both the definition of argument types (belief propagation formulas) and data transformation (from elicited qualitative data to belief and disbelief pairs). We also illustrate this approach on a toy example. First results show that it was possible to improve previous work on uncertainty propagation and elicitation issues. We still need to conduct a full experiment for assessing beliefs in rules. We will investigate the expert questionnaire. We also want to propose an approach for automatic rule type identification. In the long range, we also plan to do away with the qualitative to quantitative transformation that contains some arbitrariness, by developing the purely qualitative approach to information fusion outlined in [4], and compare it to the quantitative one.

References

1. Cobb, B.R., Shenoy, P.P.: On the plausibility transformation method for translating belief function models to probability models. Int. J. Approx. Reason. **41**(3), 314–330 (2006)
2. Cyra, L., Górski, J.: Support for argument structures review and assessment. Reliability Engineering & System Safety **96**(1), 26–37 (2011)
3. De Persis, C., Bosque, J.L., Huertas, I., Wilson, S.P.: Quantitative system risk assessment from incomplete data with belief networks and pairwise comparison elicitation. arXiv preprint arXiv:1904.03012 (2019)
4. Dubois, D., Faux, F., Prade, H., Rico, A.: A possibilistic counterpart to shafer evidence theory. In: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), New Orleans, LA, USA, June 23-26. pp. 1–6. IEEE (2019)
5. Dubois, D., Prade, H.: Practical methods for constructing possibility distributions. Int. J. Intell. Syst. **31**(3), 215–239 (2016)
6. Idmessaoud, Y., Dubois, D., Guiochet, J.: Belief functions for safety arguments confidence estimation: A comparative study. In: International Conference on Scalable Uncertainty Management. pp. 141–155. Springer (2020)
7. Idmessaoud, Y., Guiochet, J., Dubois, D.: Calculation of aggregation formulas for GSN argument types using belief functions. Tech. rep., LAAS-CNRS (Apr 2021), https://hal.laas.fr/hal-03210201

8. Jøsang, A.: Subjective logic. Springer (2016)
9. Kelly, T.: Arguing Safety – A Systematic Approach to Safety Case Management. Ph.D. thesis, Department of Computer Science, University of York, UK (1998)
10. Kelly, T.P., McDermid, J.A.: Safety case construction and reuse using patterns. In: International Conference on Computer Safety, Reliability, and Security (Safecomp) 97, pp. 55–69. Springer (1997)
11. Shenoy, P.P., Shafer, G.: Axioms for probability and belief-function propagation. In: Shachter, R.D., Levitt, T.S., Kanal, L.N., Lemmer, J.F. (eds.) Uncertainty in Artificial Intelligence, Machine Intelligence and Pattern Recognition, vol. 9, pp. 169–198. North-Holland (1990)
12. Smets, P.: Decision making in the tbm: the necessity of the pignistic transformation. International Journal of Approximate Reasoning **38**, 133–147 (2005)
13. Wang, R., Guiochet, J., Motet, G., Schön, W.: D-S Theory for Argument Confidence Assessment. In: 4th International Conference on Belief Functions (BELIEF 2016). pp. 190–200. Prague, Czech Republic (Sep 2016)
14. Wang, R., Guiochet, J., Motet, G., Schön, W.: Modelling Confidence in Railway Safety Case. Safety Science **110 part B**, 286–299 (Dec 2018)
15. Wang, R., Guiochet, J., Motet, G., Schön, W.: Safety Case Confidence Propagation Based on Dempster-Shafer theory. International Journal of Approximate Reasoning **107**, 46–64 (Apr 2019)
16. Yaghlane, A.B., Denœux, T., Mellouli, K.: Elicitation of expert opinions for constructing belief functions. In: Uncertainty and intelligent information systems, pp. 75–89. World Scientific (2008)