Confidence Assessment Framework for Safety Arguments

Rui Wang, Jérémie Guiochet, Gilles Motet

▶ **To cite this version:**

Rui Wang, Jérémie Guiochet, Gilles Motet. Confidence Assessment Framework for Safety Arguments. International Conference on Computer Safety, Reliability, and Security (SAFECOMP), Trento, Italy, Sep 2017, Trento, Italy. 14p. hal-01533221

HAL Id: hal-01533221
https://hal.science/hal-01533221v1

Submitted on 6 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Confidence Assessment Framework for Safety Arguments

Rui Wang, Jérémie Guiochet, and Gilles Motet

LAAS-CNRS, Université de Toulouse, CNRS, INSA,UPS, Toulouse, France

Abstract. Confidence in safety critical systems is often justified by safety arguments. The excessive complexity of systems nowadays introduces more uncertainties for the arguments reviewing. This paper proposes a framework to support the argumentation assessment based on experts' decision and confidence in the decision for the lowest level claims of the arguments. Expert opinion is extracted and converted in a quantitative model based on Dempster-Shafer theory. Several types of argument and associated formulas are proposed. A preliminary validation of this framework is realized through a survey for safety experts.

Keywords: safety argument, confidence assessment, belief function theory

1 Introduction

Safety case is an important representative of structured arguments adopted for critical systems. It is used to formally present that a system is free from unacceptable risks. This justification often demonstrates the compliance of the system with safety regulation and includes a great amount of convincing evidence in parallel. Both developers of critical systems and regulation bodies have to spend time on evaluating such argumentation in order to produce trustful systems or make a justified decision for certification. Many works have been done to help this evaluation process. 1) Building a clear safety argument with a graphical representation of safety arguments ([15, 6]); 2) Adding confidence arguments to justify the confidence in safety arguments ([2, 12]); 3) Assessing the confidence in arguments with quantitative methods ([7, 8, 11]).

This paper focuses on the third perspective. A confidence assessment framework with specific steps based on Dempster-Shafer theory is proposed to facilitate the argumentation assessment process. It requires the safety experts' opinions only on the lowest level claims of safety arguments. Then, the proposed framework aggregates these opinions in a quantitative way to deduce the decision and the confidence in this decision for the top goal of the safety argument. We made a first experiment through a survey among safety experts and a preliminary validation of this framework is obtained.

This paper is organized as follows. In Section 2, the background on GSN and belief function theory is provided. In Section 3, the overview of the proposed

safety argument assessment framework is given. Two argument types with quantitative confidence aggregation models are introduced. Afterwards, in Section 4, expert data collection is realized through a survey. We present the survey implementation; then its responses are analyzed. In Section 5 the related works are introduced. Finally, the contributions of our approach are summarized and future works are highlighted in Section 6.

2 Background

2.1 Safety Argumentation

Structuring an argument to convince regulation bodies is a main challenge for critical systems. Many approaches, such as safety case [16, 4], assurance case [13], trust case [6], and dependability case [5], provide concepts and notations for taking up this challenge.

Safety cases, a popular form of safety argumentation, could be defined as [4] "a documented body of evidence that provides a convincing and valid argument that a system is adequately safe for a given application in a given environment". A graphical argumentation notation, named as Goal Structuring Notation (GSN), has been developed [15] to represent the different elements of an assurance case and their relationships with individual notations. GSN allows the representation of the supporting evidence, objectives to be achieved, safety argument, context, etc. An example of GSN is given in Figure 1, which is derived from the Hazard Avoidance Pattern [16]. The five main elements of GSN presented in this figure are: *goal* (e.g., G1): the claim about the system; *solution* (e.g., Sn1): the reference to evidence item(s); *strategy* (e.g., S1): the nature of inference that exists between a goal and its supporting sub-goal(s); *context* (e.g., C1): a reference to contextual information, or a statement.

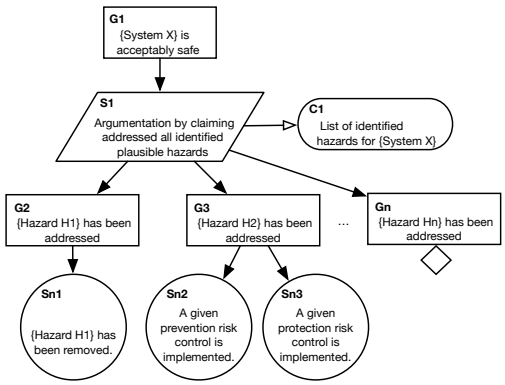


Fig. 1. GSN example adapted from Hazard Avoidance Pattern [16]

2.2 Dempster-Shafer Theory

Among uncertainty theories (such as probabilistic approaches, possibility theory, fuzzy set, etc.), Dempster-Shafer (D-S) Theory or evidence theory, was developed by Arthur Dempster and Glenn Shafer successively [18]. This theory offers a powerful tool to model human belief in evidence from different sources, and an explicit modeling of epistemic uncertainties, which is not the case in other theories. As presented later, we propose to use the D-S Theory as it allows uncertainty, imprecision or ignorance, i.e., "we know that we don't know" to be explicitly expressed.

Let X be a variable taking values in a finite set Ω representing a *frame of discernment*. Ω is composed of all the possible situations of interest. In this paper, we consider only binary frame of discernment, i.e. $\Omega_X = \{\bar{X}, X\}$. For instance, if X would be the state of a bulb, then $\Omega = \{on, off\}$. The *mass function* on Ω (m^Ω) is the mapping of the power set of Ω on the closed interval [0,1] that is, $2^\Omega \rightarrow [0, 1]$. The mass $m^\Omega(P)$ reflects the degree of belief committed to the hypothesis that the truth lies in P. The sum of the masses of all elements in the power set is equal to one. For instance, we can have the following assignment of belief: $m_1(\{on\}) = 0.5$, $m_1(\{off\}) = 0.3$, $m_1(\{on, off\}) = 0.2$. Note that $m_1(\{on, off\})$ does not represent the belief that the bulb might be in {on} or {off} state, but the degree of belief in the statement "we don't know".

More generally, an opinion about a statement X is assessed with 3 masses: *belief* ($bel_X = m(X)$), *disbelief* ($disb_X = m(\bar{X})$), and the *uncertainty* ($uncer_X = m(\Omega)$). This leads to $m(X)+m(\bar{X})+m(\Omega) = 1$ (*belief + disbelief + uncertainty = 1*). Thus we have:

$$\begin{cases} bel_X = m(X) \text{ represents the belief} \\ disb_X = m(\bar{X}) \text{ represents the disbelief} \\ uncer_X = m(\Omega) = 1 - bel_X - disb_X \text{ represents the uncertainty} \end{cases} \quad (1)$$

where bel_X, $disb_X$ and $uncer_X \in [0, 1]$.

3 Safety Argument Assessment Framework

In this section, we introduce an assessment framework for safety arguments, which allows 1) experts to provide their opinions on the lowest level claims of a structured safety argument based on available evidence (e.g. test reports, verification reports, etc.); and 2) to aggregate these opinions hierarchically until we obtain the opinion of the top claim of the argument. The opinion aggregation adopts a quantitative assessment method of argument confidence proposed in our previous works [19]. A new formula to calculate the degree of disbelief and uncertainty is provided in this paper.

3.1 Framework Overview

The proposed assessment framework of safety argument is summarized in Figure 2 with an argument showing that *Goal B and Goal C support Goal A*. This schema also illustrates the three main steps:

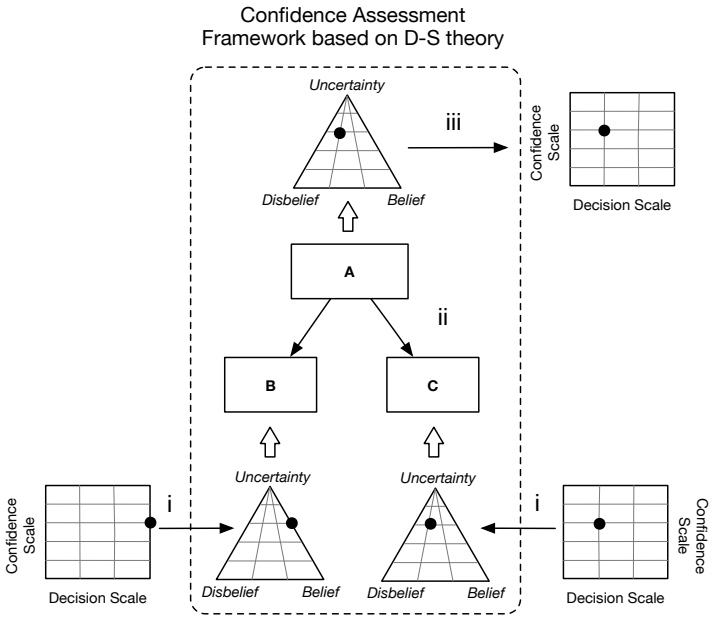


Fig. 2. Schema of the assessment framework for safety argument

i Transforming safety experts' opinions of a goal of a structured argument into a 3-tuple *(bel,uncer,disb)* representing *Belief*, *Uncertainty* and *Disbelief* in this goal. In Figure 2, the process for B and C starts from using a scaled evaluation matrix and then an uncertainty triangle named Jøsang triangle [14]. We refer to the transformation expression between decision/confidence and belief functions proposed in paper [7]. The experts' opinion is presented in two dimensions: *decision* and *confidence in this decision*. Instead of the original proposal with belief function and plausibility function, we convert the opinion directly into (bel,uncer,disb), to make explicit the formal concepts.

ii Aggregating all the 3-tuple estimation of lower-level claims into the upper-level claim. This step is based on the confidence assessment method derived from the D-S Theory [19]. As shown in Figure 2, the *Belief*, *Uncertainty* and *Disbelief* of B and C are aggregated to produce the three values of A. This aggregation requires some basic information on the argument, such as: the argument types, the weights of B and C, etc.

iii The last step presents the inverse process of Step 1, which aims to generate the opinion on A, i.e. the *decision* on A and the *confidence in this decision*.

This approach is detailed in Section 3.2 and 3.3. In Section 3.4, an application to an argument example helps for a better understanding of this assessment framework for safety argument.

3.2 Argument Types and Assessment

Like most of structured arguments, a GSN argument has a tree structure, which is composed of a top goal and branches of sub-goals. As described in [12], the assessment of the confidence in the top goal may be based on the estimation of the *trustworthiness* of each sub-goal and the *appropriateness* of the sub-goals in regard to the top goal. Thus, we propose two assessment parameters corresponding to these two aspects. Figure 3 shows an example of a simple argument: goal A is supported by two sub-goals B and C; the assessment parameters are annotated on this GSN argument and interpreted in the following way:

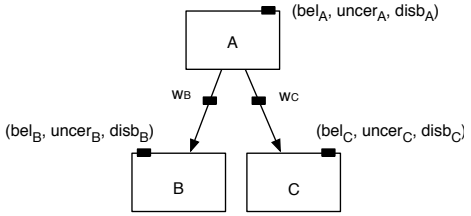


Fig. 3. An example of simple argument annotated with assessment parameters

- A 3-tuple *(bel,uncer,disb)*, such as $(bel_A, uncer_A, disb_A)$ (see the definition in Equation (1)) represents the *trustworthiness* of a claim. They do not only assess the confidence in the claims, but also allow our degree of distrust and uncertainty in them to be explicitly expressed.
- w_B and w_C are the *disjoint contributing weights* of B and C, w_B, $w_C \in [0, 1]$ and $w_B + w_C \leq 1$. A *disjoint contributing weight* means the degree that B or C can independently contribute to the trustworthiness in A. It refers to the *appropriateness* of sub-goals.

In order to propagate the trustworthiness estimation from B and C nodes to A, we propose two argument types:

- Dependent argument (D-Arg): When the contribution (to the trustworthiness in A) of a sub-goal B depends on the trustworthiness in another sub-goal C, the argument $B{+}C{\rightarrow}A$ is called *dependent argument*. For example, an argument is *"B: Test process is correct"* and *"C: Test results are correct"* support *"A: System is acceptably safe"*. The contribution of C to the trustworthiness in A depends on the trustworthiness in B.
- Redundant argument (R-Arg): When sub-goals belonging to the same top goal have a certain degree of overlapping to contribute to the trustworthiness in the top goal, the argument is called *redundant argument*. For example, an argument is *"B: Formal verification is passed"* and *"C: Test is conclusive"* support *"A: System is acceptably safe"*. B and C are two different techniques to assure the system safety. Either of them can support A in certain degree without depending on the other one.

Based on these two argument types, we proposed in [19] the aggregation formulas to integrate the trustworthiness of sub-goals. The related formulas are presented as Equations (II) and (IV) in Table 1 for the dependent and redundant arguments, respectively. In this calculation, we introduce a factor to represent the degree of dependency and redundancy among sub-goals. For an argument such as *B and C support A*, this factor, representing the *degree of correspondence* of sub-goals B and C, is expressed as: $c_A = 1 - w_B - w_C$, where $c_A \in [0, 1]$. While c_A varies between 0 and 1, the two formulas (II) and (IV) lead to several special cases of argument types, which are also described in Table 1. In particular:

- Fully dependent argument (FD-Arg): For a *dependent argument*, when $c_A = 1$, i.e. $w_B = w_C = 0$, the argument is a *fully dependent argument*. B have a total interdependence on C. One sub-goals cannot contribute to the trustworthiness in A without the other one.
- Fully redundant argument (FR-Arg): For a *redundant argument*, when $c_A = 1$, i.e. $w_B = w_C = 0$, the argument is *fully redundant argument*. Either of B and C can contribute to the full trustworthiness in the top goal.
- Disparate argument (I-Arg): When the *correspondence* between B and C c_A decreases (i.e. w_B, w_A increase) to $c_A = 0$ (i.e. $w_B + w_C = 1$), the aggregation formulas of the dependent and redundant arguments become the same formulas ((III) in Table 1). B and C contribute independently to only one part of the trustworthiness in the top goal.

Except the above three special argument types, other arguments are either *partial dependent argument (PD-Arg)* or *partial redundant argument (PR-Arg)*.

Table 1. Comparison of two different argument types

Arg. types		c_A	w_B,w_C	Aggregation formula	
D-Arg	FD-Arg	1	0	$bel_A = bel_B bel_C$ $disb_A = disb_B + disb_C - disb_B disb_C$	(I)
D-Arg	PD-Arg	$\downarrow$	$\downarrow$	$bel_A = bel_B w_B + bel_C w_C + bel_B bel_C c_A$ $disb_A = disb_B w_B + disb_C w_C + (disb_B + disb_C - disb_B disb_C) c_A$	(II)
R-Arg	I-Arg	0	$\sum = 1$	$bel_A = bel_B w_B + bel_C w_C$ $disb_A = disb_C w_B + disb_C w_C$	(III)
R-Arg	PR-Arg	$\uparrow$	$\uparrow$	$bel_A = bel_B w_B + bel_C w_C + [1 - (1 - bel_B)(1 - bel_C)] c_A$ $disb_A = disb_B w_B + disb_C w_C + disb_B disb_C c_A$	(IV)
R-Arg	FR-Arg	1	0	$bel_A = 1 - (1 - bel_B)(1 - bel_C)$ $disb_A = disb_B disb_C$	(V)

Due to the limited space of this paper, the parameter formalization and the development process of assessment aggregation formulas are not presented. For more details and a general assessment model for N-sub-goal arguments, please refer to paper [19]. In Table 1, we directly provide the formulas to calculate $(bel_A, uncer_A, disb_A)$ according to the argument types.

3.3 Expert Judgment Extraction

While assessing an argument, a safety expert has to evaluate all the elements of this argument, i.e. statement, evidence, context, etc. In Figure 4 a), a goal G1: *"Low-level requirements coverage is achieved"* is to be assessed. It is supported by the evidence S1: *"Low-level requirement coverage verification reports"*, which records the coverage verification of low-level requirements based on the contexts C1: *"Complete low-level requirements"* and C2: *"Structural coverage analysis (statement coverage, branch coverage, etc) reports"*. We adopt an evaluation matrix as proposed by [7] to assess G1 by two criteria: the *decision* on the goal and the *confidence in the decision (dec, conf)*. In Figure 4 b), there are 4 levels for decision scale from *"rejectable"* to *"acceptable"* and 6 levels for Confidence Scale from *"lack of confidence"* to *"for sure"*. We assume that, in both scales, the levels are evenly and linearly distributed. A solid dot represents the evaluation of this goal by an expert. Here, the expert accepts this goal with very high confidence. The decision *"acceptable"* indicates that the expert believes that all the low-level requirements were actually covered. Moreover, the *"very high confidence"* comes from relatively high coverage rate and thorough explanation of discrepancies in evidence S1.

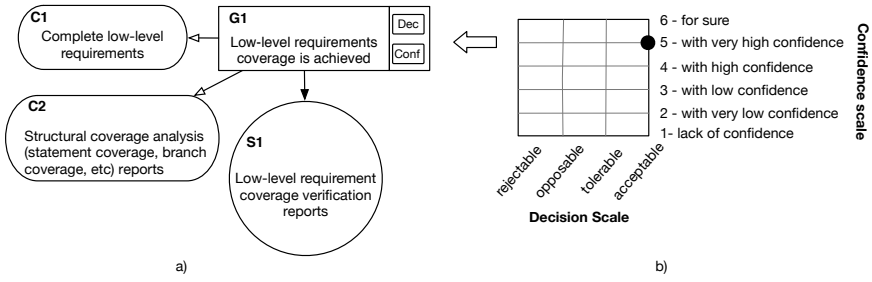


Fig. 4. An evaluation matrix for safety argument

In order to further assess the upper-level goals, we need to aggregate the expert's evaluation of sub-goals. As mentioned in the Section 3.1, the evaluation of the experts *(dec, conf)* will be transformed to *belief*, *uncertainty* and *disbelief*. In fact, this step is used to formalize the evaluation as a mass function in order to take advantage of the D-S Theory to combine uncertain information. This uncertainty theory offers a powerful tool to explicitly model and process information with uncertainty. We adopt the definition of *decision* and *confidence in the decision* of any claim A based on belief functions proposed in [7] to fit the input of the aggregation model (refer to Table 1). The modified definition is presented in Equations (2) and (3).

$$conf_A = bel_A + disb_A \tag{2}$$

$$\begin{cases} dec_A = bel_A/(bel_A + disb_A), & bel_A + disb_A \neq 0 \\ dec_A = 0, & bel_A + disb_A = 0 \end{cases} \tag{3}$$

Due to the constrains of mass function of D-S Theory, we can deduce that $conf_A$, $dec_A \in [0, 1]$.

Once, the $(dec_A, conf_A)$ is obtained based on $(bel_A, uncer_A, disb_A)$, their values may not be exactly one of the values of 4 decision levels and 6 confidence levels. If so, these numbers should be rounded to find the nearest levels.

Furthermore, the inverse functions from $(dec, conf)_A$ to $(bel_A, uncer_A, disb_A)$ are given in the Equation 4.

$$\begin{cases} bel_A = conf_A * dec_A \\ disb_A = conf_A * (1 - dec_A) \\ uncer_A = 1 - b_A - d_A \end{cases} \tag{4}$$

3.4 Application Example

In this subsection, we use a fragment of GSN argument shown in Figure 5 as an example to present the use of the proposed safety case assessment framework. In this GSN model, it is assumed that *"G1: system is acceptably safe"* (claim A), if *"G2: Low-level requirements coverage is achieved "* (sub-goal B) and *"G3: High-level requirements coverage is achieved "* (sub-goal C) are fulfilled. The confidence in A is based on the assessment of sub-goals B and C. To illustrate the calculus, we provide the arbitrary values to assess B as *"opposable"* (weak reject) with *"very low confidence"* and C as *"acceptable"* with *"very high confidence"*. The low-level requirements coverage is verified through the structural coverage analysis based on functional testing; the high-level requirements coverage is also based on function testing. B and C are linked to each other, but they also cover two different aspects. Thus, they are considered as partial dependent arguments. We arbitrarily choose the values $c_A = 0.5$ and equal disjoint contributing weights $w_B = w_C = (1 - c_A)/2 = 0.25$. A possible approach is presented in Section 4 to extract the information about argument types and weights with the help of a survey.

Here follows the three-step process of the framework proposed in Section 3.1 to realize the assessment of confidence in A.

- Transforming the evaluation $(dec, conf)$ of B and C to $(bel, uncer, disb)$ using Equation 4. $(bel_B, uncer_B, disb_B) = (0.066, 0.8, 0.134)$, $(bel_C, uncer_C, disb_C) = (0.8, 0.2, 0.0)$
- Aggregating the estimations of B and C with the aggregation formula of dependent argument (refer to (II) in Table 1). $(bel_A, uncer_A, disb_A) = (0.243, 0.657, 0.101)$
- Calculating the decision on A and the confidence in the decision $(dec_A, conf_A) = (0.707, 0.343)$. The level of decision and confidence in this decision are selected by the nearest value of the results. Thus, goal A is *"tolerable", "with low confidence"*.

In brief, the framework can be regarded as a function f: $(dec_A, conf_A) = f[(dec_B, conf_B), (dec_C, conf_C)]$, where inputs are the evaluation of sub-goals B and C, the output is the assessment of the top goal A. More generally, this framework can be applied for a safety argument with multiple sub-goals and more hierarchical levels, thanks to the general version of aggregation formulas.

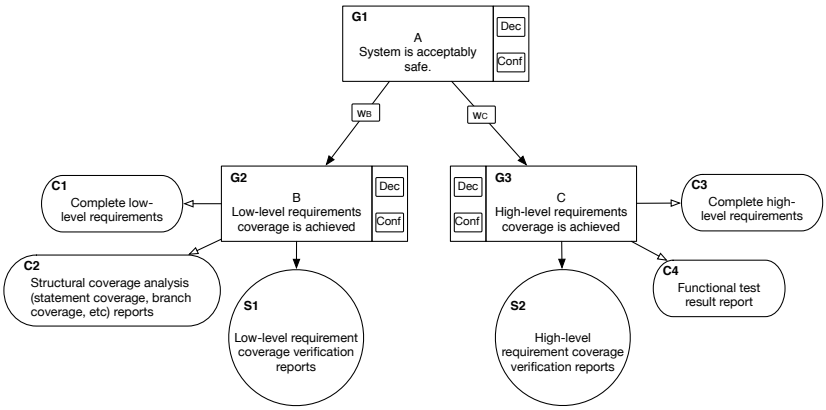


Fig. 5. A safety argument example to be estimated

4 A Survey for Expert Data Collection

To study the argument properties, such as the argument types and the sub-goal weights, we carried out a survey among experts in system safety domain.

4.1 Implementation of the Survey

In the questionnaire, four argument fragments are provided. These arguments includes Arg1 represented in Figure 5 and Arg2-Arg4 in Figures 6. They have the same form with an identical top goal A and two sub-goals B and C. For each argument, two pairs of estimation results of B and C (corresponding to Q1 and Q2 in Table 2) are initial information given to the respondents. Then, they are asked to make a decision on the top goals, that is, choosing an appropriate decision level among *rejectable (rej), opposable (opp), tolerable (tol), acceptable (acc)*; and select their confidence level in this decision from *1-lack of confidence* to *6-for sure*. For better understanding of the assessment process, an introduction of the evaluation matrix is given at the beginning of the questionnaire; and explanations and assumptions of the 4 arguments are also provided. Furthermore, an extra question follows each argument asking respondents for their understanding degree of the argument. The degrees are "to great extent", "somewhat", "very little" and "not at all". An online version of this questionnaire is available [1].

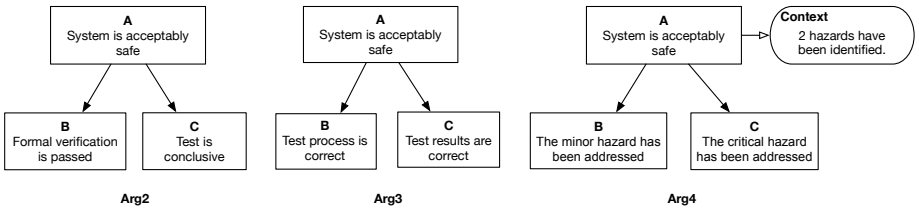


Fig. 6. Argument fragments questioned in the survey

35 experts answered this questionnaire, including: 18 system safety engineers, safety managers and other engineers of critical system fields, and 17 researchers and PhD candidates working in the system dependability domain. Due to no significant difference in the analysis results between the respondent's profiles, their answers are processed together.

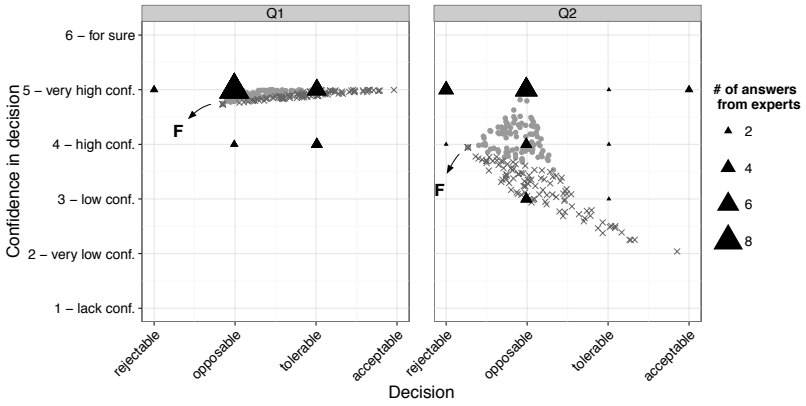
4.2 Result Analysis

The case study aims to analyze the properties of the argument examples from the questionnaire responses, that is, to estimate the sub-goal weights and argument types implicitly considered by the experts . The collected data (*expert data*) are compared with the data calculated based on the assessment formulas introduced in Section 3 (*theoretical data*).

In Figure 7, the theoretical data of dependent argument is shown as a cloud of dots derived from random trials of possible values of w_B and w_C. Note that the triangles will be explained later. Different shapes of clouds are due to the two pairs of inputs of B and C for questions Q1(Figure 7 a)) and Q2 (Figure 7 b)). According to the process of the assessment framework, we calculate the values of $(dec_A, conf_A)$ from $(dec_B, conf_B)$ and $(dec_C, conf_C)$. Then we plot the values in the evaluation matrix. The solid dots represent the values with the constraint that $w_B > w_C$; whereas the crosses represent the values of $w_B \leq w_C$. In the figures, the "F" letters represent the output of a special case of dependent argument: *fully dependent argument*.

Then, in order to extract the consensus of experts, we filtered the data using the confidence intervals. Also, if the respondents chose "not at all" for the understanding of one argument, the answers for the corresponding two questions were removed. The expert data are presented with triangles in the evaluation matrix (Figure 7). The size of the triangle indicates the number of respondents giving the same opinion.

Finally, the expert data are compared with the theoretical data clouds. Taking question Q1 of the argument Arg1 for example, we consider this argument as dependent argument, since there is some dependency between the two sub-goals. Hence, in Figure 7, the expert data are compared with framework output of dependent argument. Two large dots are matched with the distribution of dependent graphs. We assume that the argument type can be validated by the



a) b)

Fig. 7. Experts estimations of Argument 1 and theoretical data of dependent argument

degree of overlapping of these two sets of data. Thus, the percentage of the answers positioned in the cloud (matched answers) is calculated (see Table 2). Large percentages of overlapping for both of figures a) and b) confirm that the Arg1 is a dependent argument.

Furthermore, some weight information of B and C can be deduced. Looking at the biggest triangle in Figure 7 a), it shows that 8 experts have given the *opposable* decision with *very high confidence* (denoted as *opp-5*). Compared with the theoretical data cloud, these 8 answers indicate that the argument can be either fully dependent argument with $w_B = w_C = 0$ or partial dependent argument with $w_B > w_C$.

Table 2. Validation of safety argument assessment approach

Arg.	Ques.	Initial values	Expert answer examples	Expected arg. types	Validation		
					Validated arg. types	Answer in cloud (%)	Weight info.
Arg1	Q1	B: opp-5 C: acc-5	A: opp-5, tol-5	PD-Arg.	PD-Arg.	65.0%	$w_B \geq w_C$
	Q2	B: opp-5 C: acc-2	A: rej-5, opp-3,4,5			63.6%	-
Arg2	Q1	B: opp-5 C: acc-5	A: opp-5, tol-5	PR-Arg.	PR-Arg.	72.2%	$w_B \geq w_C$
	Q2	B: opp-5 C: acc-2	A: opp-4,5			62.5%	$w_B > w_C$
Arg3	Q1	B: opp-5 C: acc-5	A: opp-5, tol-5	PD-Arg.	PD-Arg.	62.5%	$w_B \geq w_C$
	Q2	B: opp-5 C: acc-2	A: rej-5, opp-5			58.3%	$w_B > w_C$
Arg4	Q1	B: opp-5 C: acc-5	A: tol-4,5	I-Arg.	PD-Arg.	57.7%	$w_B = w_C$
	Q2	B: opp-5 C: acc-2	A: opp-3, opp-5			77.8%	-

In Table 2, the analysis of the expert answers for 4 arguments is summed up. Some representative examples of the expert answers are given in this table. Comparing the "expected argument types" with "validated argument types", Arg4 is considered as "partial dependent argument" rather than the "disparate

argument". The percentages of the answers in the cloud are calculated for all the argument examples. These results show that experts have a certain degree of consensus on the type of the arguments based on our approach. Moreover, the experts' preference of weights for B and C are listed in the last column. "-" means that there is no clear opinion on the preference of weights.

A large percentage of the consensus answers matches the model output of the assessment framework proposed in this paper, which is a first validation of the framework. Furthermore, based on the above analysis of the survey data, we deduce the properties of the 4 argument examples including argument types and the disjoint contributing weights.

5 Related Work

Confidence assessment of safety case has been mainly addressed with two perspectives. The first one focuses on the identification of "defeaters" of an argument, and the construction of an additional argument dedicated to confidence [3, 12]. Such approaches are mainly qualitative. A second trend is the development of quantitative approaches of confidence in argument. Indeed, excessive growth of argument leads it to make analysis for estimating confidence too complex; then quantitative tools might help analysts to estimate the confidence. To refer to some of them, we can cite [8], based on Bayesian Network, and [7, 11] based on belief function theory. As presented in [10], many approaches are studied for quantitative assessment of safety argument confidence. In this last paper the authors study the flaws and counterarguments for each approaches, and conclude that whereas quantitative approaches for confidence assessment are of high interest, no method is currently fully applicable. Moreover, we argue that these quantitative approaches lack of practicability between assurance case and confidence assessment, or do not provide clear interpretation of confidence calculation parameters. Our framework over comes this flaw.

Compared to our approach, the paper [17] provides an expert judgment extraction of confidence and a propagation calculation based on belief theory in order to build a confidence case as proposed in [12]. Nevertheless they do not address inference type when aggregating information. They also do not study how the confidence level could be used by the analysts to make a decision regarding the safety case.

In [2], the authors mainly introduce four argument types and formulas to combine confidence regarding these types. They also use belief theory for calculation, and the result is provided with belief, disbelief and uncertainty estimation for each evidence of the safety case. Even if some types of argument are comparable with our proposal (e.g., their "Alternative" is near our "Redundant"), they do not provide any justification of the combining formulas, with a relative low intuitive interpretation of the parameters (which is a main drawback for potential users). Moreover, once calculation is performed, the results do not provide any justification for a decision regarding the acceptability of the safety case.

As already mentioned, we reuse a part of the approach presented in [7]. In this paper, the authors introduce a way to convert a decision on the acceptability of a statement in a safety case and its confidence, into belief theory parameters. We also use similar steps in our approach, from expert judgment extraction to calculation of a decision and its confidence in the top statement. Compared to our work, they did not use GSN for safety case modeling as we proposed; but the main difference is in the argument types and associated formulas. Indeed, they extended the work from [9] to propose 6 types of arguments. We found them too complex for an intuitive identification in a real safety case. Moreover, according to each of these types, several parameters are difficult to determine and interpret. Our objective is really to provide an efficient and pragmatic approach for analysts; thus we actually only propose 2 types of argument, and a direct application to GSN safety cases.

6 Conclusion

In this paper, an assessment framework has been put forward to support the safety argument assessment process. This 3-step framework only requires the evaluation results of the lowest-level claims; then it aggregates them to estimate the confidence in the top claim. The quantitative aggregation approach based on Dempster-Shafer theory was proposed in our previous work [19]. An evaluation matrix for extracting experts opinion is adopted [7] with the scales of decision and confidence in the decision. We define two main argument types: dependent and redundant arguments. By changing the weights of sub-goals, we also proposed to refine these types using the same formula. Meanwhile a possible approach to estimate argument properties is introduced. A survey was carried out to make a preliminary validation of our framework. We focus in this survey on validating the types of arguments, their aggregation models and the expert judgment extraction. 35 safety engineers and researchers in system dependable domain participated to this survey. We compared the questionnaire results with the theoretical output calculated by applying our assessment framework. This first experiment shows that our aggregation models are consistent with expert judgments. The framework makes practical the theoretical model in terms of the extraction of experts' opinion on the trustworthiness of sub-goals. However, while assessing an argument, this framework still requires the expert to determine the argument types and weights of sub-goals. A method to identify the argument types and weights for a given safety argument will be our future work.

References

1. Questionnaire for safety argument assessment research, 1 2017. https://goo.gl/forms/V3vMnl59cTWA6Lws2.
2. Anaheed Ayoub, Jian Chang, Oleg Sokolsky, and Insup Lee. Assessing the overall sufficiency of safety arguments. In *21st Safety-Critical Systems Symposium (SSS'13)*, pages 127–144, 2013.

3. Anaheed Ayoub, BaekGyu Kim, Insup Lee, and Oleg Sokolsky. A systematic approach to justifying sufficient confidence in software safety arguments. In *Computer Safety, Reliability, and Security*, pages 305–316. Springer, 2012.
4. Peter Bishop and Robin Bloomfield. A methodology for safety case development. In *Industrial Perspectives of Safety-Critical Systems*, pages 194–203. Springer, 1998.
5. Robin Bloomfield, Bev Littlewood, and David Wright. Confidence: its role in dependability cases for risk assessment. In *Dependable Systems and Networks, 2007. DSN'07. 37th Annual IEEE/IFIP International Conference on*, pages 338–346. IEEE, 2007.
6. Lukasz Cyra and Janusz Gorski. Supporting compliance with security standards by trust case templates. In *Dependability of Computer Systems, 2007. DepCoS-RELCOMEX'07. 2nd International Conference on*, pages 91–98. IEEE, 2007.
7. Lukasz Cyra and Janusz Gorski. Support for argument structures review and assessment. *Reliability Engineering & System Safety*, 96(1):26–37, 2011.
8. Ewen Denney, Ganesh Pai, and Ibrahim Habli. Towards measurement of confidence in safety cases. In *Empirical Software Engineering and Measurement (ESEM), 2011 International Symposium on*, pages 380–383. IEEE, 2011.
9. Trudy Govier. *A practical study of argument*. Wadsworth, Cengage Learning, 2013.
10. Patrick J. Graydon and C. Michael Holloway. An investigation of proposed techniques for quantifying confidence in assurance arguments. *Safety Science*, 92:53 – 65, 2017.
11. Jérémie Guiochet, Quynh Anh Do Hoang, and Mohamed Kaaniche. A model for safety case confidence assessment. In *Computer Safety, Reliability, and Security (SAFECOMP)*, pages 313–327. Springer, 2015.
12. Richard Hawkins, Tim Kelly, John Knight, and Patrick Graydon. A new approach to creating clear safety arguments. In *Advances in systems safety*, pages 3–23. Springer, 2011.
13. ISO/IEC 15026-2. Systems and software engineering - systems and software assurance - part 2: Assurance case, 2011. International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC).
14. Audun Jøsang. A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(03):279–311, 2001.
15. Tim Kelly. *Arguing Safety - A Systematic Approach to Safety Case Management*. PhD thesis, Department of Computer Science, University of York, 1998.
16. Tim Kelly and John McDermid. Safety case construction and reuse using patterns. In *Computer Safety, Reliability, and Security (SAFECOMP)*, pages 55–69. Springer, 1997.
17. S. Nair, N. Walkinshaw, T. Kelly, and J. L. de la Vara. An evidential reasoning approach for assessing confidence in safety evidence. In *2015 IEEE 26th International Symposium on Software Reliability Engineering (ISSRE)*, pages 541–552, Nov 2015.
18. Glenn Shafer. *A mathematical theory of evidence*, volume 1. Princeton university press Princeton, 1976.
19. Rui Wang, Jérémie Guiochet, Gilles Motet, and Walter Schön. DS theory for argument confidence assessment. In *International Conference on Belief Functions*, pages 190–200. Springer, 2016.