DRAFT

full paper published in:

Lecture Notes in Computer Science, Volume 5219, Berlin/Heidelberg, 2008, pp. 291-304

Proceedings of the 27th International Conference on Computer Safety, Reliability and Security SAFECOMP 2008, Newcastle, UK, 2008, pp 291-304

Paper published in the journal and presented at the conference

Expert Assessment of Arguments: a Method and Its Experimental Evaluation

Lukasz Cyra[1], Janusz Górski[1]

[1] Gdansk University of Technology, Department of Software Engineering
Narutowicza 11/12, 80-952 Gdansk, Poland
lukasz.cyra@eti.pg.gda.pl, jango@pg.gda.pl

Abstract. Argument structures are commonly used to develop and present cases for safety, security and other properties. Such argument structures tend to grow excessively. To deal with this problem, appropriate methods of their assessment are required. Two objectives are of particular interest: (1) systematic and explicit assessment of the compelling power of an argument, and (2) communication of the result of such an assessment to relevant recipients. The paper gives details of a new method which deals with both problems. We explain how to issue assessments and how they can be aggregated depending on the types of inference used in arguments. The method is fully implemented in a software tool. Its application is illustrated by examples. The paper also includes the results of experiments carried out to validate and calibrate the method.

Keywords: Argument assessment, Dempster-Shafer model, Argument structures, Safety Case, Trust Case, Assurance Case.

1 Introduction

Arguments are commonly used in ‘cases’ (safety cases [13, 17, 18], assurance cases [2], trust cases [9, 10], conformity cases [4, 5], etc.) to justify various qualities of objects (like safety, security, privacy, conformity with standards and so on). Recently, there is a growing interest in these subjects, which leads to the development of relevant methodologies and finding new application areas for argument structures [6, 7].

The idea which lies behind the development of argument structures is to make expert judgment explicit in order to redirect the dependence on judgment to issues on which we can trust this judgment [22]. In this way, it is possible to analyze the argument structure and take a position on it.

However, argument structures tend to grow excessively, becoming too complex to be analyzed by non-experts. Therefore appropriate methods of assessment of argument structures are required. Two objectives are of particular interest: (1) assessment of the compelling power of an argument structure, and (2) communication of the result of such an assessment to relevant recipients.

The paper gives details of such an appraisal method. Although, this approach is developed in connection with the Trust-IT methodology [9, 10], which focuses on trust cases, it is general enough to be applied to other kinds of cases. The appraisal mechanism described in the paper is based on the Dempster-Shafer model [20, 21]. It provides for issuing assessments and their aggregation depending on the types of inference used in arguments. The mechanism has been already implemented in the TCT (Trust Case Toolbox) software tool [14]. The paper includes examples of its application which have been borrowed from the trust case developed for a real system [19]. The paper also reports on the results of some experiments carried out to validate and calibrate the appraisal method.

The presented method extends and modifies the approach to arguments appraisal proposed in [12], which had problems with adapting itself to the types of arguments occurring in trust cases. It proposes linguistic appraisal scales which are then represented in terms of Dempster-Shaffer belief functions and can be mapped onto the Josang's opinion triangle [15, 16]. Some of the aggregation rules (not described in detail in the paper) are following Yager's modification of Dempster's rule of combination [20]. A general discussion of the role of confidence in dependability cases can be found in [3].

2 Representing Arguments

The proposed approaches to argument representation in 'cases' [1, 9, 17] are influenced by the Toulmin's argument model [23]. In our approach (the Trust-IT methodology) we adopt this model in a fairly straightforward way as shown in Fig. 1.

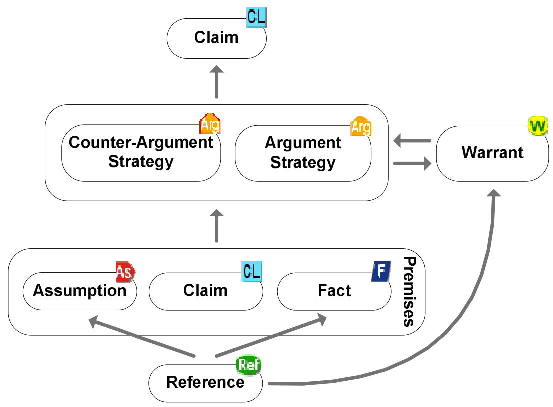

Fig. 1. Trust-IT argument model.

The presented structure includes a conclusion to be justified represented as a *claim* (denoted). The claim is supported by an *argument strategy* (denoted), which contains a basic idea how to support the conclusion. In the case of *counter-argument strategies* (denoted) it includes the idea of rebuttal of the claim. The argument strategy is related to a *warrant* (denoted) which justifies the inference from *premises* to the conclusion. This justification may require additional, more specific arguments, which is shown by the arrow leading from the argument to the warrant node.

A premise can be of three different types: it can be an *assumption* (denoted), in which case the premise is accepted without further justification; it can be a more specific claim which is justified further; or it can represent a *fact* (denoted) which is obviously true or, otherwise, is supported by some evidence. Evidence is provided in external (to the trust case) documents, which are pointed at by nodes of type *reference* (denoted). In the case of an assumption, the referenced document can contain explanation of the context in which the assumption is made.

As claims and warrants can be demonstrated using other (sub-)claims the argumentation structure can grow recursively. The icons labeling different nodes shown in Fig.1 and implemented in the supporting tool (TCT [14]) to denote elements of argument structure will be used in the subsequent examples presented in this paper.

An example argument following the model introduced in Fig. 1 is presented in Fig. 2. The example refers to the PIPS system, the system delivering to its users health and lifestyle related personalized services [19]. The top claim of the structure postulates validity of information supplied to PIPS. It is demonstrated by considering different channels through which information related to a patient's state is supplied to the system. This leads to four premises which are used by the argument. Three of them: *'Validity of information from PIPS-enabled devices'*, *'Validity of information from questionnaires'* and *'Validity of product codes'* are claims and are supported by more detailed arguments. The fourth one, *'Truthfulness of the information provided by a patient'*, is an assumption and is not further analyzed.

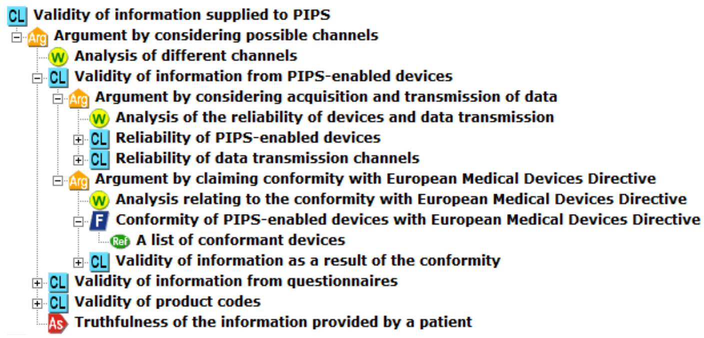

Fig. 2. An example argument coming from a trust case for the e-health system PIPS.

To appraise the compelling power of such an argument we have introduced an appraisal method which is presented in the following part of the paper. The method consists of two steps:

Step 1 – appraisal of warrants and premises

1.1. Estimate the 'strength' of warrants (these, which do not have their supporting arguments) occurring in the argument. (This assessment is based on the assessment of the evidence linked to the warrant but also the common knowledge and the logical bases for the inference.)

1.2. Estimate the 'strength' of the facts and assumptions occurring in the argument. (This appraisal is mostly based on the assessment of the evidence linked to the premises by the reference nodes.)

Referring to the example shown in Fig. 2, the appraisal of the *'Analysis of different channels'* warrant would take into account if the validity of information received from the devices, questionnaires and by reading product codes with the additional assumption that patients are not cheating intentionally are sufficient to conclude the validity of information supplied to the system.

The appraisal of the premises would assess the acceptability of the assumption that patients are not cheating intentionally (note that this is context dependent and the result would depend on the knowledge about the system and its environment). The appraisal of the *'Conformity of the PIPS-enabled devices with European Medical Devices Directive'* fact would take into account the evidence linked to this fact by the corresponding reference node.

Implementation of **Step 1** requires that we have an appropriate scale to express the appraisals of warrants, facts and assumptions. This should be complemented by appropriate guidelines supporting the assessor.

Step 2 – automatic aggregation of the partial appraisals

2.1. Starting from the leaves of the argumentation tree, aggregate the appraisals of the premises and warrants to obtain the appraisal of the conclusions.

2.2. Repeat the process until the top conclusion has been reached.

Referring to the example from Fig. 2, this step would result in the appraisal of the top claim taking as an input the appraisals of warrants, facts and assumptions occurring in the argumentation and recursively applying the aggregation rules.

Implementation of **Step 2** requires that the appropriate aggregation rules were defined covering all relevant types of warrants occurring in the arguments.

3 Appraisal Mechanism

To support experts during the appraisal process we have introduced two linguistic scales, the *Decision scale* and *Confidence scale*. The former provides for expressing the attitude towards acceptance or rejection of the assessed element. It distinguishes four decision values: *acceptable*, *tolerable*, *opposable* and *rejectable*. The latter scale provides for expressing the confidence in this decision. It distinguishes six levels of confidence: *for sure*, *with very high confidence*, *with high confidence*, *with low confidence*, *with very low confidence* and *lack of confidence*.

The scales can be combined together which results in twenty four values of the *Assessment scale* as shown in Fig. 3. The elements of the scale, which are represented as small circles, have intuitively understandable linguistic values. For instance, the element represented as the white circle reads: '*with very low confidence tolerable'*.

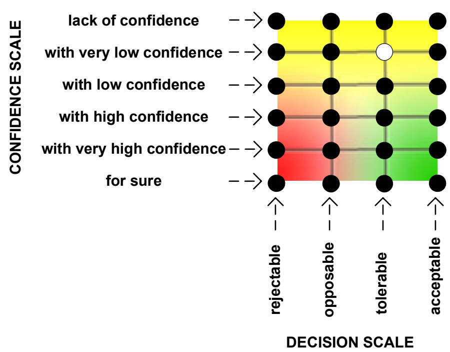

Fig. 3. The *Assessment scale* as a product of *Confidence scale* and *Decision scale*.

The semantics of the scales can be formalized using Dempster-Shafer's belief and plausibility functions [20, 21]. If *s* is a statement, then

- $Bel(s) \in [0,1]$ is the *belief* function representing the amount of belief that directly supports *s*,
- $Pl(s) \in [0,1]$ is a *plausibility* function representing the upper bound on the belief in *s* that can be gained by adding new evidence.

We can formally represent *confidence* as:

$$Conf(s) = Bel(s) + 1 - Pl(s)\,,\ Conf(s) \in [0,1] \tag{1}$$

and map the interval [0,1] onto linguistic values from the *Confidence scale* in such a way that *lack of confidence* = 0, and *for sure*=1.

Decision scale distinguishes four levels to express the ratio between belief (acceptance of a statement) and the overall confidence in the statement (without distinguishing if we want it to be accepted or rejected).
Using Dempster-Shafer's functions we can formally represent the decision concerning *s* as:

$$Dec(s) = \begin{cases} Bel(s)/(Bel(s)+1-Pl(s)) & Bel(s)+1-Pl(s) \neq 0 \\ 1 & Bel(s)+1-Pl(s) = 0 \end{cases}, \; Dec(s) \in [0,1] \;. \qquad \textbf{(2)}$$

The interval [0,1] is mapped onto linguistic values of the *Decision scale* in such a way that '*rejectable'*=0 and *'acceptable'*=1.

We can observe that the difference between stating that something is acceptable or rejectable is significant if enough evidence supporting such assessment is available (which corresponds to e.g. *'for sure'* or *'with very high confidence'* assessments), however, there is no difference if no evidence is present (which corresponds to the *'lack of confidence'* assessment). To address this particular aspect, the *Assessment scale* can be represented as a triangle shown in Fig. 4 and called Josang's opinion triangle [15, 16].

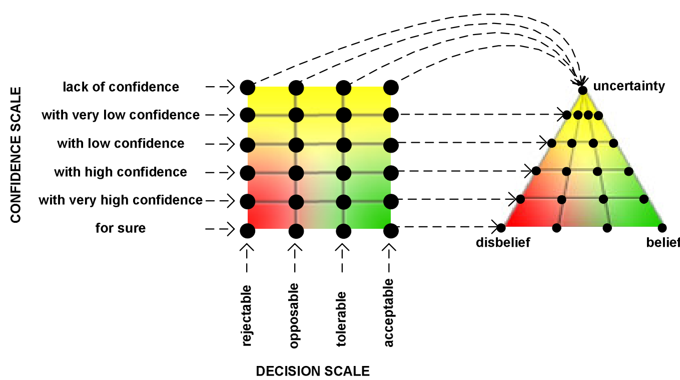

Fig. 4. Mapping *Assessment scale* on Josang's opinion triangle.

In the Josang's opinion triangle *'lack of confidence'* is mapped onto uncertainty. The other vertices of the triangle represent the total disbelief (equivalent to '*for sure rejectable'*) and total belief (equivalent to '*for sure acceptable'*).

3.1 Appraisal Procedure

The *Assessment scale* is applied to express opinions and the level of confidence in these opinions in relation to assumptions, facts and warrants which are not supported by a (counter-)argument strategy.

The assessment of a single node of the structure proceeds as follows:

1) If no evidence for or against the statement representing the node is available the *'lack of confidence'* assessment is issued and the procedure is broken.
2) In the other case, the ratio between the evidence supporting the acceptance and rejection of the statement is assessed and an appropriate value from the *Decision scale* is chosen.
3) Then, it is assessed how much evidence could additionally be provided to be sure about the decision chosen in step 2. This amount of missing evidence drives the selection from the *Confidence scale*.
4) The final assessment from the *Assessment scale* is obtained by combining the two partial assessments from steps 2 and 3.

Fig. 5 presents a fragment of the user interface for issuing assessments of the TCT tool [14]. The user can drag a small white marker over the opinion triangle shown on the left hand side. Then, the linguistic values corresponding to the current position of the marker are displayed in the Confidence level: and Decision: windows. It is also possible to directly choose an appropriate linguistic assessment. Additionally, the current levels of belief, disbelief and uncertainty are displayed as horizontal bars just above the opinion triangle.

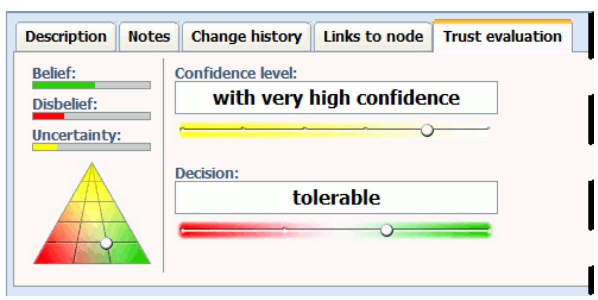

Fig. 5. User interface for issuing assessments of the TCT tool.

3.2 Appraisal Examples

As an example let us consider fact F stating that *'PIPS-enabled devices are conformant with European Medical Devices Directive'* (see Fig. 2).

Let us assume that:

(1) There is some evidence relevant to F.

(2) All the available evidence supports F which results in choosing the *'acceptable'* value from the *Decision scale*.

(3) The evidence is almost complete, however, the certification process has not been performed yet – which leads to the *'with very high confidence'* assessment.

Consequently, the final appraisal of F is: *'with very high confidence acceptable'*.

Let us now consider F in a different situation:

(1) There is some evidence relevant to F.

(2) The evidence demonstrates that most of the requirements of the directive are met, however, one significant requirement is not yet fulfilled; this results in the *'opposable'* assessment.

(3) The evidence is substantial, however, not complete which gives the *'with high confidence'* assessment.

Consequently, the final appraisal of F is: *'with high confidence opposable'*.

As another example let us take warrant W (*'Analysis of different channels'*) from Fig. 2. The warrant identifies different types of channels providing information to the system and explains that if they provide valid information, the information available to the system is also valid.

Let us assume that:

(1) An inventory of types of channels exists.

(2) It identifies four major types of channels represented in the argument and in addition, some other less important ones, which were not considered in the argument; this leads to the *'tolerable'* assessment of the warrant.

(3) The inventory resulted from a formalized procedure of review of system design; this leads to the *'for sure'* assessment.

Consequently, the final appraisal of W is: *'for sure tolerable'*.

As yet another example let us take assumption A (*'Truthfulness of the information provided by a patient'*) from Fig. 2, which states that patients will not intentionally input false data into the system.

Let us assume that:

(1) There are bases to assess the assumption as there is some information about in what situations and what kind of data patients input into the system.

(2) The assessor tends to accept the assumption although sees some situations where it does not necessarily hold; the decision is to assess it as *'tolerable'*.

(3) The assessor has no doubts that she/he sees the whole scope of relevant situations; this leads to the *'for sure'* assessment.

Consequently, the final appraisal of A is: *'for sure tolerable'*.

4 Aggregation Rules

Aggregation rules define how the appraisals of the premises and the appraisal of the warrant are used to calculate the appraisal of the conclusion. We will briefly discuss four basic argument types and the corresponding aggregation rules which we have identified and illustrate their application by examples. An interested reader is referred to [8] for more detail and the formal definition of the aggregation rules. The examples show the results of application of the aggregation mechanisms as they are presently implemented in the TCT tool [14].

C-argument (*Complementary argument*) is such where the premises provide complementary support for the conclusion. In the case of *C-argument* not only the assessments of the premises and the warrant but also the weight associated with each premise is taken into account. The final assessment of the conclusion is a sort of weighed mean value of the contribution of all the premises.

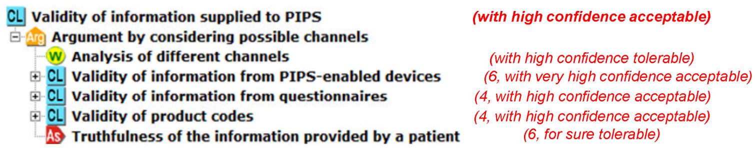

Fig. 6. Assessment of the conclusion of *C-argument*.

Let us consider a *C-argument* shown in Fig. 6. The assessments of the warrant and premises (together with the associated weights) are shown on the right hand side in italic. The resulting assessment of the conclusion is *'with high confidence acceptable'* (printed in bold). Note that despite the fact that one of the premises is *'tolerable'* the conclusion is *'acceptable'*. This results form the fact that the other premises ranked *'acceptable'* outweighed in this case. Additionally, it can be seen that the confidence in the conclusion is slightly lower than it could be expected while looking at the assessments of the premises. This results from the fact that there were some doubts concerning the strength of the inference rule, reflected in the assessment of the warrant.

Let us consider the example from Fig. 6 but with the assessment of the assumption modified to *(6, lack of confidence)*. In such a case the assessment of the conclusion would be *(with low confidence acceptable)* which results form the fact that the other premises are fairly high assessed and there is relatively high assessment of the warrant. Nevertheless, the assessment of the conclusion would be lower than in the example shown in Fig. 6.

Another type of argument, called *A-argument* (*Alternative argument*) is encountered in situations where we have two or more independent justifications of the common conclusion. In *A-arguments* the confidence in the assessments coming from different argument strategies is reinforced if the assessments agree, or it is decreased if they contradict each other.

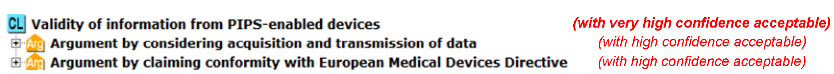

Fig. 7. Assessment of the conclusion of *A-argument*.

Let us illustrate the situation with the example shown in Fig. 7. Both arguments support the conclusion providing high confidence. In this case the resultant assessment of the conclusion is *'with very high confidence acceptable'*.

In the case the arguments contradict each other, the effect is opposite. If one of the arguments in Fig. 7 would support rejection of the conclusion and another recommend acceptance, there would be no confidence in the conclusion at all and the *'lack of confidence'* assessment would result.

In *NSC-arguments* (*Necessary and Sufficient Condition list argument)*, negative assessments are strongly reinforced. In such arguments the acceptance of all premises leads to the acceptance of the conclusion, whereas rejection of a single premise leads to the rebuttal of the conclusion. An example of such an argument is shown in Fig. 8. Each premise is a necessary condition for the conclusion. Therefore, if even one of them is rejected, the conclusion cannot be accepted. Consequently, low assessments of the premises leads to a rapid drop in assessment of the conclusion.

Fig. 8. Assessment of the conclusion of *NSC-argument*.

In *SC-arguments* (*Sufficient Condition list argument*) acceptance of the premises leads to the acceptance of the conclusion similar to *NSC-arguments*. An example of such an argument is presented in Fig. 9. The difference to *NSC-argument* is that in this case rejection of a single premise leads to the rejection of the whole inference. For instance, the lack of conformity with EMD Directive does not result in invalid information received from the devices. The only reasonable conclusion is that in such a case we do not know anything new concerning the validity of this information.

Fig. 9. Assessment of the conclusion of *SC-argument*.

Detailed analysis of the inference types encountered in argument structures is presented in [8] and the aggregation rules for different types of algorithms have been proposed. Each warrant occurring in the argumentation has its type explicitly identified and the corresponding aggregation rule assigned, which is done by the developer of the argument structure. Additionally, for *C-arguments* it is necessary to assign weights to the premises, which indicate the influence of a given premise on the conclusion.

5 Assessment Scenarios

Arguments can be used in different contexts and their appraisal serves two basic purposes:

1. to assess the compelling power of the argumentation,
2. to construct a simple and understandable message communicating the strength of the argument to the receivers who do not have capacity or resources to study and assess the argument themselves.

In the ‘standard’ scenario of (safety, assurance, trust) cases the argument aims to justify some distinguished property of a considered object in its application context. Such justification is often complex and difficult to understand without sufficient expertise and resources. In such situations the appraisal mechanism can be used by experts to record and accumulate their opinions about the argument. The mechanism provides full traceability to the elementary assessments and the way they were combined into the final one. Different experts’ opinions can be compared and, if necessary, the resultant assessment can be easily computed. Such opinions can then be communicated to the managers and other decision-makers to support their decisions concerning suitability of the considered object for the expected purpose (for instance, granting a license for using the object in its target context). The results of the appraisal can be also communicated to broader public which can promote trust and the feeling of safety/security.

In [11] a collaborative development process for trust cases was proposed. Its extension, taking into account the proposed appraisal mechanism which supports the assessment phase in the process is illustrated in Fig. 10. The assessment phase provides, as part of its feedback, the report on structural errors in the trust case which is a positive side effect of the application of the appraisal mechanism. It provides also an assessed trust case which can be presented to viewers and used as a base for the next development cycle.

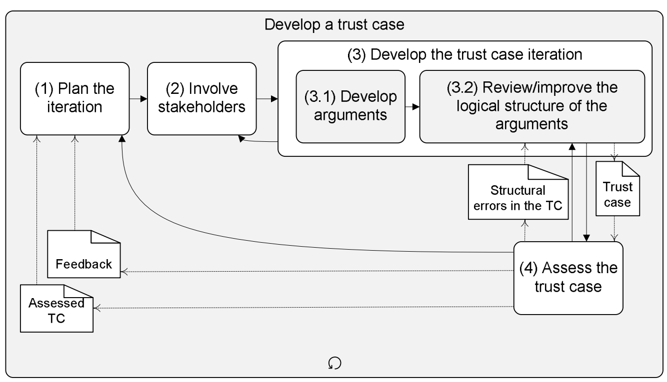

Fig. 10. Trust Case development process.

Another application scenario, to which the appraisal mechanism can bring a significant value, is application of standards. Conformity assessment is a process which ends with a binary assessment (acceptance or rejection of the claim of conformity). In this process, fulfillment of numerous requirements is checked applying the same binary decision scale. This is mainly because of the lack of practical and usable mechanisms supporting more fine differentiation of the fulfillment level of various requirements of the standard. In every assessment project, some of the requirements are fulfilled better than others and are supported by stronger evidence. To exploit this fact, more sophisticated methods of evidence and justification appraisal are needed, which would make it possible to perform appraisal in parallel to the conformity achievement process to get feedback on how the project is approaching full satisfaction of the conformity criteria. This aspect is of particular interest if self-declaration is taken into account (first party conformity assessment). Application of the appraisal mechanisms could help to make self-assessment more objective and provide useful feedback during the conformity achievement process.

To this end *Standards Conformity Framework (SCF)* [4, 5] has been proposed which supports demonstration of conformity with standards. The aim of SCF is to develop and maintain a document which justifies the claim of conformity. Its component is a conformity case template which is a skeleton of argumentation about conformity, derived from a given standard. The template has gaps to be filled in during the process of conformity demonstration. The gaps are supplemented with project specific evidence which proves that an examined object fulfils the requirements of the standard. It results in a complete argument which is called a *Conformity Case*. Such a conformity case can be easily assessed using the appraisal mechanism proposed in this paper. The whole process of development of templates of conformity cases, their conversion to complete arguments and later appraisal of those arguments is fully supported by the TCT tool.

6 Experimental Evaluation

The linguistic scales for the appraisal of arguments have been chosen deliberately to support assessors by offering them a (not too large) set of intuitively understood values. However, in order to perform the calculations defined in the aggregation rules, it was necessary to represent the linguistic values as numbers from the [0,1] interval and this mapping could have a significant impact on the computed results. There was no evidence, that the most obvious, even distribution of the linguistic values over the [0,1] interval is the most proper one. And in fact, the experiments showed the opposite.

We decided to find the mapping between linguistic scales and the [0,1] interval experimentally by calibrating the aggregation rules relating them to the expert assessments of conclusions of the selected set of arguments.

A group of 31 students of the last year of a computer science university course took part in the experiment. The students were divided into three groups, each of which was supposed to apply one of the aggregation rules: *A-rule, NSC-rule* and *C-rule*. *SC-argument* type was dropped because of its similarity to *NSC-argument* type. Each student was provided with five simple trust cases composed of a claim, an argument strategy, a warrant and premises (in the case of C-rule and NSC-rule) or a claim with a few argument strategies (in the case of A-rule).

The experiment participants were asked to assess the warrant and, in the case of C-rule to assign weights to the premises. Then, assuming the pre-defined assessments of each premise (in the case of C-rule and NSC-rule) or the assessments assigned to each of the argument strategies (in the case of A-rule) the participants were asked to give their assessment of the conclusion using the Assessment Scale. They were supposed to repeat this step for 10 different sets of initial assessments of the premises (chosen randomly) for each trust case. That makes the total of 50 assessments of the conclusions issued by each participant. To check for consistency, each participant was additionally asked to repeat his/her assessments for 10 randomly selected situations.

Some students were excluded from the experiment for formal reasons or because their assessments apparently were not reasonable (for instance, they declared high confidence in acceptance of a conclusion in a situation where the premises were with high confidence rejectable). Finally, 8 questionnaires related to A-argument type, 6 questionnaires related to NSC-argument type and 10 questionnaires related to C-argument type were used in the following analysis.

In the experiment the consistency of the students' answers was measured (i.e. the average change in a repeated assessment of the same item) and the accuracy of assessments obtained by application of the aggregation rules (measured as the average distance between the students' answers and the results obtained by application of the aggregation rules). The results are presented in table 1. The numbers in the table are normalized, which means that 1 represents the distance between two adjacent positions on the linguistic scale.

The data shows that the accuracy of the results obtained by application of the aggregation rules is similar to the consistency of the participants' answers (i.e. the corresponding metrics for each aggregation rule do not differ significantly). This is the maximum of what could have been achieved regarding the data set used to calibrate the aggregation rules. The data show that using the (calibrated) aggregation rules we can expect to obtain the results which are fairly close to the results which would be obtained while engaging humans in the aggregation process.

Table 1. Results of experiments.

		Aggregation rule		
		A-rule	***NSC-rule***	***C-rule***
Consistency of students' assessments	***Confidence scale***	1,03	0,94	0,84
	Decision scale	0,64	0,62	0,87
Accuracy of assessments obtained by application of aggregation rules	***Confidence scale***	1,06	1,10	0,90
	Decision scale	0,80	0,78	0,66

Further calibration requires more data which we plan to collect in the subsequent experiments.

7 Conclusion

This article introduced an innovative method of argument structures appraisal. The method provides for gathering expert opinions about the inferences used in the argumentation and the value of the supporting evidence. It can be applied to assess the compelling power of arguments used in different contexts. In particular, it can be used with respect to the arguments contained in different cases, like safety cases, security cases, assurance cases or trust cases. It can also be used to support standards conformity processes. The method has been fully implemented in the TCT tool which supports full-scale application of our Trust-IT methodology. Some experimental validation of the method has been already performed and further experiments are under preparation.

Acknowledgement

This work was partially supported by the project ANGEL - 'Advanced Networked embedded platform as a Gateway to Enhance quality of Life' (IST project 2005-IST-5-033506-STP) and by the project PIPS – 'Personalized Information Platform for health and life Services' (Contract No. 507019 IST2.3.1.11 e-Health) within the European Commission 6th Framework Programme.

Contributions by Michal Nawrot and Michal Witkowicz in the development of the appraisal mechanism in TCT tool are to be acknowledged.

References

1. Bishop, P., Bloomfield, R.: A Methodology for Safety Case Development, Industrial Perspectives of Safety-critical Systems: Proceedings of the Sixth Safety-critical Systems Symposium, Birmingham (1998)
2. Bloomfield, R., Guerra, S., Masera, M., Miller, A., Sami Saydjari, O.: Assurance Cases for Security, A report from a Workshop on Assurance Cases for Security, Washington, USA (2005)
3. Bloomfield, R. E., Littlewood, B., Wright, D.: Confidence: Its Role in Dependability Cases for Risk Assessment, 37th Annual IEEE/IFIP International Conference Dependable Systems and Networks, 338--346 (2007)
4. Cyra, L., Gorski, J.: Supporting Compliance with Safety Standards by Trust Case Templates, Proc. of ESREL 2007, Volume 2, Norway, 1367--1374 (2007)
5. Cyra, L., Gorski, J.: Standard Compliance Framework for Effective Requirements Communication, Polish Journal of Environmental Studies, Volume 16 no. 5B, 312--316 (2007)
6. Cyra, L., Gorski, J.: Extending GQM by Argument Structures, 2nd IFIP Central and East European Conference on Software Engineering Techniques CEE-SET, 1--16 (2007)
7. Cyra, L. Gorski, J.: Using Argument Structures to Create a Measurement Plan, Polish Journal of Environmental Studies, Volume 16 no. 5B, 230--234 (2007)
8. Cyra, L., Gorski, J.: Supporting Expert Assessment of Argument Structures in Trust Cases, Ninth International Probabilistic Safety Assessment and Management Conference PSAM 9, Hong Kong, China, 1--9 (2008)
9. Gorski, J., Jarzebowicz, A., Leszczyna, R., Miler, J., Olszewski, M.: Trust Case: Justifying Trust in IT Solution, Elsevier, Reliability Engineering and System Safety, Volume 89, 33--47 (2005)
10. Gorski, J.: Trust Case – a Case for Trustworthiness of IT Infrastructures, Cyberspace Security and Defence: Research Issues, NATO ARW Series, Springer-Verlag, 125--142 (2005)
11. Gorski, J.: Collaborative Approach to Trustworthiness of IT Infrastructures, Proc. of IEEE International Conference on Technologies for Homeland Security and Safety TEHOSS 2005, 137--142 (2005)
12. Gorski, J., Zagorski, M.: Reasoning about Trust in IT Infrastructures, Proc. of ESREL 2005, 689-- 695 (2005)
13. Greenwell, W., Strunk, E., Knight, J.: Failure Analysis and the Safety-Case Lifecycle, Human Error, Safety and Systems Development 2004, 163--176 (2004)

14. Information Assurance Group: TCT User Manual, Gdansk University of Technology, http://kio.eti.pg.gda.pl/trust_case/download/TCTEditor_Users_Manual.pdf (2007)
15. Josang, A., Grandison, T.: Conditional Inference in Subjective Logic, Proc. of the 6th International Conference on Information Fusion, Cairns, 471--478 (2003)
16. Josang, A., Pope, S., Daniel, M.: Conditional Deduction Under Uncertainty. Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2005). Barcelona, Spain, 824--835 (2005)
17. Kelly, T.: Arguing Safety – A Systematic Approach to Managing Safety Cases. PhD Thesis, University of York, UK (1998)
18. Kelly, T., McDermid J.: A Systematic Approach to Safety Case Maintenance, Proc. Of SAFECOMP, 271--284 (1999)
19. PIPS Project website, http://www.pips.eu.org
20. Sentez, K., Ferson, S.: Combination of Evidence in Dempster-Shafer Theory, SANDIA National Laboratories (2002)
21. Shafer G.: Mathematical Theory of Evidence, Princetown University Press (1976)
22. Strigini, L.: Formalism and Judgement in Assurance Cases, Workshop on Assurance Cases: Best Practices, Possible Obstacles, and Future Opportunities, Proc. of DSN 2004, Florence, Italy (2004)
23. Toulmin, S.: The Uses of Argument, Cambridge University Press, Cambridge (1969)