Two Forms of Explanations in Computational Assumption-based Argumentation

(Extended Abstract)

Xiuyi Fan, Siyuan Liu, Huiguo Zhang, Chunyan Miao
Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly (LILY)
Nanyang Technological University, Singapore

Cyril Leung
Electrical and Computer Engineering
The University of British Columbia, Canada

ABSTRACT

Computational Assumption-based Argumentation (CABA) has been introduced to model argumentation with numerical data processing. To realize the "explanation power" of CABA, we study two forms of argumentative explanations, *argument explanations* and *CU explanations* representing *diagnosis* and *repair*, resp.

Keywords

Argumentation, Explanation

1. INTRODUCTION

Assumption-based Argumentation (ABA) [8] is a form of structured argumentation with applications in many areas [6]. However, when used as a modeling tool, ABA has limited ability to directly model systems involving numerical calculation. For instance, in ABA based decision making work, e.g. [3, 4], the relations between decision candidates and agent goals need to be "pre-compiled" into binary predicates rather than analyzed from data. The lack of numerical calculation is a major hindrance to ABA applications requiring intensive data processing.

The Computational Assumption-based Argumentation (CABA) framework [2], an ABA extension, introduced Computation Units (CUs) [5] to capture computation that is difficult to represent with standard ABA. A unique advantage of CABA is that, while supporting numerical calculation, it enhances the "explanation power" of argumentation by connecting results obtained from numerical calculation to high-level arguments. We study two forms of CABA explanations, *argument explanation (arg-explanation)* and *CU-explanation*, for non-acceptable arguments. We leverage on the established relation between CABA and Abstract Argumentation (AA) [1] for our work. For a non-acceptable argument A, its arg-explanation gives a form of *diagnosis*, identifying attacking arguments that cannot be defended. Its CU-explanation represents a form of *repair*, identifying "fixes" that would render A acceptable.

2. EXPLANATION IN CABA

We introduce CABA explanations with a version of the Multiple Attribute Decision Making problem presented in [9]. *Good College* is admitting students. To evaluate candidates, four attributes are considered: $Exam_1$, $Exam_2$, *Interview* and *Extracurricular Activity (EA)*. $Exam_1$ and $Exam_2$ are scores ranging from 0 to 100; *Interview* is a rank from E to A; *EA* is a binary value, (*Yes* /*No*). The selection criterion is specified with two conditions C1 and C2, such that: (**C1**) The average score of $Exam_1$ and $Exam_2$ is greater than 90, or *EA* is *Yes*; and (**C2**) the *Interview* rank is A. A student is admitted iff both C1 and C2 are met.

Table 1: Student Candidate Admission Data.

Student	$Exam_1$	$Exam_2$	*Interview*	*EA*
$s1$	92	89	A	*No*
$s2$	93	85	A	*No*

Table 1 presents the attributes of two candidates, $s1$ and $s2$. Here, we can see that for student $s1$, his average exam score is $(92 + 89)/2 = 90.5$, hence meeting condition C1; his interview rank is A, meeting condition C2; therefore $s1$ should be admitted. For $s2$, his average exam score is $(93+85)/2 = 89$ and he has not performed any extracurricular activity, thus failing to meet C1; although $s2$ has an A for his interview, $s2$ cannot be admitted. Here, we need to compute the average scores of $Exam_1$ and $Exam_2$ and test if the average is greater than 90. We pack this computation into a CU, $u_{90} = \langle \mathsf{T}_{90}, \mathsf{C}_{90}, \mathsf{E}_{90} \rangle$, in which:

- $\mathsf{T}_{90} \subseteq \mathbb{Z} \times \mathbb{Z}$ are the two exam scores;
- $\mathsf{C}_{90}(x, y) = (x + y)/2$;
- $\mathsf{E}_{90} = \top$ if $\mathsf{C}_{90} > 90$ and $\mathsf{E}_{90} = \bot$ otherwise.

Similarly, we pack the checks for *Interview* and *EA* into CUs u_{int} and u_{ea}, resp, as follows.

$u_{int} = \langle \mathsf{T}_{int}, \mathsf{C}_{int}, \mathsf{E}_{int} \rangle$ in which:

- $\mathsf{T}_{int} = \{A, B, C, D, E\}$;
- $\mathsf{C}_{int}(x) = x$;
- $\mathsf{E}_{int} = \top$ if $\mathsf{C}_{int} = A$ and $\mathsf{E}_{int} = \bot$ otherwise.

$u_{ea} = \langle \mathsf{T}_{ea}, \mathsf{C}_{ea}, \mathsf{E}_{ea} \rangle$ in which:

- $\mathsf{T}_{ea} = \{\text{Yes, No}\}$;
- $\mathsf{C}_{ea}(x) = x$;
- $\mathsf{E}_{ea} = \top$ if $\mathsf{C}_{ea} =$ Yes and $\mathsf{E}_{ea} = \bot$ otherwise.

We use the following framework to model the admission problem.

- $\mathcal{U}$ is the following CUs:
  $u_{90}(s1)$ $u_{ea}(s1)$ $u_{int}(s1)$
  $u_{90}(s2)$ $u_{ea}(s2)$ $u_{int}(s2)$
- $\mathcal{L}$ is the following sentences:

C1(s1)	C2(s1)	Ave>90(s1)	EA(s1)
notC1(s1)	notC2(s1)	Adm(s1)	INT(s1)
C1(s2)	C2(s2)	Ave>90(s2)	EA(s2)
notC1(s2)	notC2(s2)	Adm(s2)	INT(s2)

Appears in: *Proc. of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017), S. Das, E. Durfee, K. Larson, M. Winikoff (eds.), May 8–12, 2017, São Paulo, Brazil.*

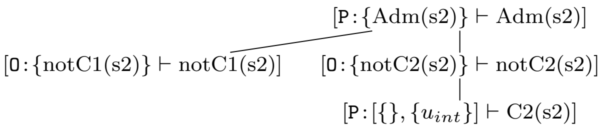

Figure 1: An illustration of the non-admissible dispute tree for {Adm(s2)} $\vdash$ Adm(s2). This dispute tree is not admissible as it contains opponent leaf nodes $[\mathtt{O}:\{\mathrm{notC1(s2)}\} \vdash \mathrm{notC1(s2)}]$ that is not counter-attacked by any proponent argument.

- $\mathcal{R}$ is the following rules:
  C1(s1) ← Ave>90(s1) Ave>90(s1) ← $u_{90}(s1)$
  C1(s1) ← EA(s1) EA(s1) ← $u_{ea}(s1)$
  C2(s1) ← INT(s1) INT(s1) ← $u_{int}(s1)$
  C1(s2) ← Ave>90(s2) Ave>90(s2) ← $u_{90}(s2)$
  C1(s2) ← EA(s2) EA(s2) ← $u_{ea}(s2)$
  C2(s2) ← INT(s2) INT(s2) ← $u_{int}(s2)$
- $\mathcal{A}$ is the following assumptions:
  Adm(s1), C1(s1), C2(s1), Adm(s2), C1(s2), C2(s2)
- $\mathcal{C}$ is the following contrary mappings:
  $\mathcal{C}$(Adm(s1)) = {notC1(s1), notC2(s1)},
  $\mathcal{C}$(notC1(s1)) = {C1(s1)}, $\mathcal{C}$(notC2(s1)) = {C2(s1)}.
  $\mathcal{C}$(Adm(s2)) = {notC1(s2), notC2(s2)},
  $\mathcal{C}$(notC1(s2)) = {C1(s2)}, $\mathcal{C}$(notC2(s2)) = {C2(s2)}.

Here, u_{90}, u_{ea} and u_{int} model the requirement for exam scores, extracurricular activity and interview, resp. The rules in $\mathcal{R}$ describe conditions for meeting the two conditions C1 and C2. Assumptions in $\mathcal{A}$ and contraries in $\mathcal{C}$ model the selection criterion that a student is admitted iff both C1 and C2 are met. We see that the set of arguments $\{\{Adm(s1)\} \vdash Adm(s1)$, $[\{\}, \{u_{90}(s1)\}] \vdash C1(s1)$, $[\{\}, \{u_{int}(s1)\}] \vdash C2(s1)\}$ is admissible and the argument $\{Adm(s2)\} \vdash Adm(s2)$ is not.

We realize the "explaning power" of CABA with *argument explanation* and *computation unit explanation*, for diagnosis and repair [7], resp. For a non-admissible argument A, its non-admissibility can be attributed to either its own non-applicability or the existence of attackers which cannot be defended. We formalize *argument explanation (arg-explanation)* for non-acceptable CABA arguments as follows.

Definition 1. *Given a CABA framework* F *with corresponding AA framework* $\langle \mathtt{A}^{\mathrm{F}}, \mathtt{R}^{\mathrm{F}} \rangle$*, let A be a non-admissible CABA argument in* F*. Then, if there exists some $As \subseteq \mathtt{A}^{\mathrm{F}}$, such that:*

1. *A is admissible in $\langle \mathtt{A}^{\mathrm{F}}, \mathtt{R}^{\mathrm{F}} \rangle \setminus As$, and*
2. *there is no $As' \subset As$ such that A is admissible in $\langle \mathtt{A}^{\mathrm{F}}, \mathtt{R}^{\mathrm{F}} \rangle \setminus As'$, then As is an* arg-explanation *of A.*

Otherwise, $\{A\}$ is the arg-explanation *of A.*

Figure 1 shows a non-admissible dispute tree for $\{\mathrm{Adm(s2)}\} \vdash \mathrm{Adm(s2)}$. We can see that, if $\{\mathrm{notC1(s2)}\} \vdash \mathrm{notC1(s2)}$ is removed from the tree, then $\{\mathrm{Adm(s2)}\} \vdash \mathrm{Adm(s2)}$ would be admissible. Thus, $\{\{\mathrm{notC1(s2)}\} \vdash \mathrm{notC1(s2)}\}$ is an arg-explanation for $\{\mathrm{Adm(s2)}\} \vdash \mathrm{Adm(s2)}$. We interpret this as:

> *An explanation for not admitting student s2 is that s2 does not meet requirement C1.*

Arg-explanations are focused on identifying attacks that cannot be addressed. Thus, they can be viewed as a form of diagnosis, i.e., the identification of the cause of non-admissibility. In the context of CABA, since only applicable arguments can defend others, the "lack of support" can be traced to the unsuccessfulness of certain CUs in the sense that, if these CUs were successful, then the argument in question would be acceptable. Changing these CUs can be viewed as a means for "repair". Formally:

Definition 2. *Given a CABA framework* $\mathrm{F} = \langle \mathcal{U}, \mathcal{L}, \mathcal{R}, \mathcal{A}, \mathcal{C} \rangle$*, let $\mathbb{U}$ be the set of all CUs, $f : \mathbb{U} \mapsto \mathbb{U}$ an injective function such that for all $u \in \mathbb{U}$, $f(u)$ is a successful CU, then the* repaired framework of *F* wrt some $\Gamma \subseteq \mathcal{U}$ *is* $\mathrm{F}^+(\Gamma) = \langle \mathcal{U}', \mathcal{L}, \mathcal{R}', \mathcal{A}, \mathcal{C} \rangle$ *in which*

- $\mathcal{U}' = \{f(u) | u \in \Gamma\} \cup \{u | u \in \mathcal{U} \setminus \Gamma\}$ *and*
- $\mathcal{R}' = \{s \leftarrow f(u), \_, \ldots | s \leftarrow u, \_, \ldots \in \mathcal{R}$ *and* $u \in \Gamma\} \cup \{\rho | \rho \in \mathcal{R}$ *such that there is no CU* $u \in \Gamma$ *in the body of* $\rho\}$.[1]

The intuition of Definition 2 is that given a CABA framework F and some CUs Γ in F, the repaired framework $F^+(\Gamma)$ is another CABA framework with all CUs in Γ made successful and rules $\mathcal{R}$ updated to $\mathcal{R}'$ with new CUs replacing the ones in Γ. $\mathcal{L}, \mathcal{A}$ and $\mathcal{C}$ remain unchanged. We assume that for all $u \in \mathcal{U}$, $f(u) \notin \mathcal{U}$.

Definition 3. *Given* $\mathrm{F} = \langle \mathcal{U}, \mathcal{L}, \mathcal{R}, \mathcal{A}, \mathcal{C} \rangle$ *with some non-admissible argument A in* F*, $\Gamma \subseteq \mathcal{U}$ is a CU-explanation for A iff the following conditions hold: (1) A is admissible in* $\mathrm{F}^+(\Gamma)$*, (2) there is no $\Gamma' \subset \Gamma$ such that A is admissible in* $\mathrm{F}^+(\Gamma')$*, (3) there is an admissible dispute tree $\mathcal{T}$ for A in* $\mathrm{F}^+(\Gamma)$ *such that for each $u \in \Gamma$, $f(u)$ is not in any argument in the culprit [8] of $\mathcal{T}$.*

Conditions 1 and 2 in Definition 3 specify that a CU-explanation needs to be minimum (wrt $\subseteq$). Condition 3 specifies that all CUs in Γ must be in arguments defending A in the repaired framework.

In our example, student $s2$ is not admitted as he does not meet condition C2, represented by the argument $A = \{\mathrm{notC1(s2)}\} \vdash \mathrm{notC1(s2)}$. There are two (non-applicable) arguments, $B = [\{\}, \{u_{90}(s2)\}] \vdash C1(s2)$, and $C = [\{\}, \{u_{ea}(s2)\}] \vdash C1(s2)$ attacking A. If u_{90} or u_{ea} were successful, then B or C would be applicable and A would be attacked. Hence, $\{Adm(s2)\} \vdash Adm(s2)$ would be admissible. Thus, $\{u_{90}\}$ and $\{u_{ea}\}$ are two CU-explanations for $\{Adm(s2)\} \vdash Adm(s2)$. We read this as:

> *To admit s2, his average score needs to exceed 90;* and
> *To admit s2, he needs to do some extracurricular activity.*

The following two propositions describe relations between arg-explanations and CU-explanations. If an applicable argument has a CU-explanation, then it has an arg-explanation not including itself.

Proposition 1. *Given a CABA framework* F*, let A be an applicable non-admissible argument in* F*; if Γ is a CU-explanation for A, then A also has an arg-explanation As' such that $A \notin As'$.*

Non-successful CUs are not always in a CU-explanation.

Proposition 2. *There exists a CABA framework* F *with non-admissible A in* F *such that As is an arg-explanation for A; there exists some non-applicable argument $[\_, \Gamma] \vdash \_$ in* F *attacking some argument in As; for some non-successful $u \in \Gamma$, u is not in any CU-explanation for A.*

3. CONCLUSION

In this paper, we studied two forms of explanations for non-admissible arguments in CABA, *argument explanation* and *CU explanation*, to realize diagnosis and repair, resp. In the future, we plan to explore justifications for acceptable arguments. We also plan to investigate CABA's use in other applications and its properties related to existing argumentation frameworks.

4. ACKNOWLEDGEMENTS

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its IDM Futures Funding Initiative.

[1]The symbol _ denotes an anonymous variable, as in Prolog.

REFERENCES

[1] P. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *AIJ*, 77(2):321–357, 1995.

[2] X. Fan, S. Liu, H. Zhang, C. Leung, and C. Miao. Explained activity recognition with computational assumption-based argumentation. In *Proc. ECAI*, 2016.

[3] X. Fan and F. Toni. Decision making with assumption-based argumentation. In *Proc. TAFA*, pages 127–142. Springer, 2013.

[4] X. Fan, F. Toni, A. Mocanu, and M. Williams. Multi-agent decision making with assumption-based argumentation. In *Proc. AAMAS*, pages 533–540, 2014.

[5] X. Fan, H. Zhang, C. Leung, and C. Miao. A first step towards explained activity recognition with computational abstract argumentation. In *Proc. IEEE MFI*, 2016.

[6] S. Modgil, F. Toni, F. Bex, I. Bratko, C. Chesñevar, W. Dvořák, M. Falappa, X. Fan, S. Gaggl, A. García, M. González, T. Gordon, J. Leite, M. Možina, C. Reed, G. Simari, S. Szeider, P. Torroni, and S. Woltran. The added value of argumentation. In *Agreement Technologies*, volume 8, pages 357–403. Springer, 2013.

[7] Y. Peng and J. Reggia. *Abductive Inference Models for Diagnostic Problem Solving*. Springer, 1990.

[8] F. Toni. A tutorial on assumption-based argumentation. *Argument & Computation, Special Issue: Tutorials on Structured Argumentation*, 5(1):89–117, 2014.

[9] K. P. Yoon and C.-L. Hwang. *Multiple Attribute Decision Making: An Introduction*. Sage Publications Inc, March 1995.