

Abstract Argumentation Frameworks with Fallible Evidence

Kenneth SKIBA ^{a,1}, Matthias THIMM ^a, Andrea COHEN ^b,
Sebastian GOTTIFREDI ^b, and Alejandro J. GARCÍA ^b

^a *University of Koblenz-Landau, Germany*

^b *Universidad Nacional del Sur, CONICET, Argentina*

Abstract. We consider a generalisation of abstract argumentation frameworks where arguments need to be backed by pieces of evidence in order to be actually present in the argumentation framework. These pieces of evidence come with an associated cost for retrieval and may not be available at any given time. We model an information-seeking agent in this scenario that aims at deciding whether a certain argument is acceptable while minimising the total evidence retrieval cost. We investigate the computational complexity of decision variants of this optimisation problem and find that, depending on the underlying classical argumentation semantics, complexity rises one level in the polynomial hierarchy compared to the classical case.

Keywords. abstract argumentation, computational complexity, evidence retrieval cost

1. Introduction

Computational models of argumentation [1] aim at modelling rational decision-making through the representation of arguments and their relationships. In particular, abstract argumentation frameworks [6] provide a simple representation formalism of such situations by focusing on the representation of arguments and a conflict relation between arguments through modeling this setting as a directed graph. Here, arguments are identified by vertices and an *attack* from one argument to another is represented as a directed edge. This simple model already provides an interesting object of study, see [2] for an overview. Several extensions of this model have been investigated as well, such as considering an additional support relation [5], recursive interactions [9], attacks by sets of arguments [10], and others.

In this paper, we investigate yet another extension of abstract argumentation. In many real-life application scenarios for argumentation, arguments are not standalone entities but rely on pieces of *evidence* in order to be *active* in an argumentation [4, 12]. Consider an online discussion forum and a discussion about the

¹Corresponding Author: Kenneth Skiba, University of Koblenz-Landau, Germany; E-mail: kennethskiba@uni-koblenz.de.

spread of the COVID-19 virus. A specific argument in this context could be “The COVID-19 virus is a serious danger because (1) its incubation phase is quite long and (2) infected people can easily infect other people during this phase. Moreover, although (3) the mortality rate is quite small, (4) the impact on public health systems and the elderly can be severe”. In order for this argument to be believable at all, facts (1), (2), (3), and (4) need to be backed by some evidence. For example, the author of that argument can link its argument to some articles from the World Health Organisation (WHO) or other resources of authority, also providing concrete numbers to the imprecise facts of the argument. If the recipient of the argument wishes to assess the validity of the argument, she can visit the linked pieces of evidence and verify the claims. However, this verification step involves time and effort, so pieces of evidence usually come with an associated cost (such as time) that need to be spent in order to verify the argument. Moreover, retrieval of pieces of evidence may fail, because a web server may be down or the article is no longer available.

Here, we model scenarios such as the one outlined above by extending abstract argumentation frameworks through the addition of pieces of evidence, a function associating arguments with pieces of evidence and a function determining their associated cost, generalising the formalisation of [4]. More precisely, the contributions of this paper are as follows:

1. We present Abstract Argumentation Frameworks with Fallible Evidence (AAFE) as an extension to Dung’s abstract argumentation frameworks that take evidence for arguments into account (Section 3).
2. We investigate the computational complexity of a certain optimisation problem within our new setting (Section 4).

We also provide necessary preliminaries on abstract argumentation in Section 2, discuss related works in Section 5, and conclude in Section 6.

2. Abstract Argumentation

Following [6], an (*abstract*) *argumentation framework* AF is a pair $(\mathcal{A}, \mathcal{R})$, where $\mathcal{A}$ is a finite set of arguments and $\mathcal{R}$ is a set of attacks between arguments, i. e. $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$. An argument a is said to *attack* b if $(a, b) \in \mathcal{R}$. We call an argument a *acceptable with respect to a set* $S \subseteq \mathcal{A}$ if for each $b \in \mathcal{A}$ with $(b, a) \in \mathcal{R}$, there is an argument $c \in S$ with $(c, b) \in \mathcal{R}$. An argumentation framework $(\mathcal{A}, \mathcal{R})$ can be illustrated by a directed graph with vertex set $\mathcal{A}$ and edge set $\mathcal{R}$.

For an argumentation framework $AF = (\mathcal{A}, \mathcal{R})$ and a set $\mathcal{A}' \subseteq \mathcal{A}$ we define the *projection* $AF_{\mathcal{A}'}$ of AF onto $\mathcal{A}'$ via $AF_{\mathcal{A}'} = (\mathcal{A}', \mathcal{R} \cap (\mathcal{A}' \times \mathcal{A}'))$.

Semantics are given to argumentation frameworks by means of *extensions*, i. e., sets of mutually acceptable arguments. A set $S \subseteq \mathcal{A}$ is *conflict-free* (CF) if there are no arguments a and b in S such that $(a, b) \in \mathcal{R}$. We call a conflict-free set S *admissible* (AD) if every argument $a \in S$ is acceptable with respect to S .

Definition 1. Let $(\mathcal{A}, \mathcal{R})$ be an argumentation framework and $S \subseteq \mathcal{A}$.

- S is a *complete extension* (CP) if it is admissible and contains every argument that is acceptable with respect to S .

- S is the *grounded extension* (GR) if it is the minimal complete extension (wrt. set inclusion).
- S is a *preferred extension* (PR) if it is a maximal complete extension (wrt. set inclusion).
- S is a *stable extension* (ST) if it is conflict-free and for each $b \in \mathcal{A} \setminus S$, there is at least one argument $a \in S$ such that $(a, b) \in \mathcal{R}$.

Note that the grounded extension is uniquely defined and stable extensions may not exist [6]. Given an argumentation framework $(\mathcal{A}, \mathcal{R})$ and an argument $a \in \mathcal{A}$, we say that a is *credulously* (*skeptically*) accepted under the semantics σ if it is contained in at least one σ -extension (all σ -extensions) of $(\mathcal{A}, \mathcal{R})$, respectively.

3. Abstract Argumentation with Fallible Evidence

We now define abstract argumentation frameworks with fallible evidence (abbreviated AAFE) as a generalisation of abstract argumentation frameworks. In this generalisation, each argument is associated with a set of evidence, which models observations needed to be present in order to make the argument “active”. Each evidence comes with an associated cost that needs to be paid in order to attempt to retrieve the evidence. This cost must also be paid if it turns out that the evidence is not available.

Example 1. Consider the following arguments exchanged by doctors trying to diagnose a patient:

- a:* If the patient shows the set S_1 of signs and symptoms, there are indications that the patient has the disease D .
- b:* If the patient also shows the set S_2 of signs and symptoms, he could have disease D' instead. We can perform a high sensitivity test T_1 for D' . If the result is positive, then we have reasons to diagnose the patient with D' .
- c:* If T_1 is positive, a high specificity test T_2 for D' has to be performed. If T_2 's result is negative, then we can refrain from diagnosing the patient with D' .
- d:* We can also run a high sensitivity test T_3 for D . If we get a negative result, then we can refrain from diagnosing the patient with D .

The arguments are based on different sets of signs and symptoms and/or the results of different tests being performed over the patient. Then, these observations are the pieces of evidence the arguments are based on and, furthermore, they come with an associated cost. On the one hand, the doctors have to spend some time in order to identify the sets S_1 and S_2 . On the other hand, the cost for performing each test has to be paid (both time and money), regardless of whether the result is as expected (as specified by the corresponding piece of evidence) or not; in cases where the tests' outcome are not as expected, we can consider that the corresponding pieces of evidence are unavailable and cannot be retrieved.

We define *abstract argumentation frameworks with fallible evidence* as follows.

Definition 2. An *abstract argumentation framework with fallible evidence* (AAFE) F is a tuple $F = (\mathcal{A}, \mathcal{R}, \mathcal{E}, \delta, \mu)$ where $\mathcal{A}$ is a set of arguments, $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ is the attack relation, $\mathcal{E}$ is a set of evidence, $\delta : \mathcal{A} \rightarrow 2^{\mathcal{E}}$ assigns to each argument a set of evidence, and $\mu : \mathcal{E} \rightarrow \mathbb{N}$ is the evidence cost function.

The function μ is extended to sets of evidence $E \subseteq \mathcal{E}$ via $\mu(E) = \sum_{e \in E} \mu(e)$. Furthermore, by abusing notation, for a set of evidence $E \subseteq \mathcal{E}$, we write $\delta^{-1}(E) = \{a \in \mathcal{A} \mid \delta(a) \subseteq E\}$.

Given a concrete set of evidence $E \subseteq \mathcal{E}$, an AAFE $F = (\mathcal{A}, \mathcal{R}, \mathcal{E}, \delta, \mu)$ is instantiated to an abstract argumentation framework $F_E = (\mathcal{A}_E, \mathcal{R}_E)$ by projecting on the arguments activated by the set of evidence, i.e. $F_E = (\mathcal{A}, \mathcal{R})_{\delta^{-1}(E)}$. We assume that not all pieces of evidence in $\mathcal{E}$ can actually be retrieved, let $\hat{\mathcal{E}} \subseteq \mathcal{E}$ denote the set of evidence which is actually available. We now consider an agent Ag that has to inquire whether a certain argument $a_{\text{query}} \in \mathcal{A}$ is credulously or skeptically accepted in $F_{\hat{\mathcal{E}}}$ (i.e. in the framework obtained by projecting on the set of active arguments whose evidence is available) with respect to a semantics σ while $\hat{\mathcal{E}}$ is not known to Ag .

We denote by $\Delta_{\sigma}^{\circ}(F, a) \in \{y, n, na\}$ (yes, no, not active) the acceptance status of a in F with respect to the semantics σ and the problem $\circ \in \{\text{cred}, \text{skep}\}$; the first two labels correspond to active arguments belonging to $\delta^{-1}(\hat{\mathcal{E}})$, whereas the latter corresponds to arguments for which the acceptance status cannot be determined because they are inactive.

The agent can ask for every piece of evidence $e \in \mathcal{E}$ by paying its cost $\mu(e)$ and, if $e \in \hat{\mathcal{E}}$, then he also retrieves e and activates the corresponding arguments. Of course, Ag should be economic and only ask for as little evidence as required (set $E_{\text{Ag}} \subseteq \mathcal{E}$) in order to make sure that for the final set of evidence E that he actually collected (i.e. $E = E_{\text{Ag}} \cap \hat{\mathcal{E}}$) we have that both frameworks $F_{\hat{\mathcal{E}}}$ and F_E yield the same answer regarding a_{query} and he paid as little cost as possible. Formally, given $F = (\mathcal{A}, \mathcal{R}, \mathcal{E}, \delta, \mu)$ and an argument $a_{\text{query}} \in \mathcal{A}$, Ag must solve the following optimisation problem:

Minimise $\mu(E_{\text{Ag}})$ such that for every argument $a \in \mathcal{A} \setminus \delta^{-1}(E_{\text{Ag}} \cap \hat{\mathcal{E}})$ with

$$\Delta_{\sigma}^{\circ}((\mathcal{A}, \mathcal{R})_{\delta^{-1}(E_{\text{Ag}} \cap \hat{\mathcal{E}})}, a_{\text{query}}) \neq \Delta_{\sigma}^{\circ}((\mathcal{A}, \mathcal{R})_{\delta^{-1}(E_{\text{Ag}} \cap \hat{\mathcal{E}}) \cup \{a\}}, a_{\text{query}})$$

we have $E_{\text{Ag}} \cap \delta(a) \cap (\mathcal{E} \setminus \hat{\mathcal{E}}) \neq \emptyset$.

In other words, Ag seeks to determine the cheapest set of evidence E_{Ag} such that the acceptance status of a_{query} does not change. For that Ag made sure that for every argument a that cannot be constructed from this set of evidence (specifically, from its subset of available evidence) and would change the acceptance status of a_{query} , he asks for at least one piece of evidence of a which cannot be retrieved because it is not available. We denote the optimal solution of the above problem as $\text{OPT}_{F, \hat{\mathcal{E}}}^{\circ, \sigma}(a_{\text{query}})$.

Example 2. Consider the argumentation framework with fallible evidence $F = (\mathcal{A}, \mathcal{R}, \mathcal{E}, \delta, \mu)$ depicted in Figure 1 and defined via

$$\mathcal{A} = \{a, b, c, d, e, f\}$$

$$\mathcal{R} = \{(a, b), (b, c), (d, c), (e, c), (e, d), (f, e)\}$$

$$\begin{aligned}
\mathcal{E} &= \{e_1, e_2, e_3, e_4\} \\
\delta(a) &= \{e_1\} & \delta(b) &= \{e_1, e_2\} & \delta(c) &= \{e_2\} \\
\delta(d) &= \{e_3\} & \delta(e) &= \{e_3, e_4\} & \delta(f) &= \{e_3, e_4\} \\
\mu(e_1) &= 2 & \mu(e_2) &= 6 & \mu(e_3) &= 3 & \mu(e_4) &= 9
\end{aligned}$$

For example, we have that argument b can only be considered if the pieces of evidence e_1 and e_2 are available and these have a cost of 2 and 6, respectively. In Figure 1 we use dotted edges to indicate which piece of evidence is needed for which argument.

Let us consider grounded semantics (under credulous reasoning) and observe that in the complete framework where all pieces of evidence can be retrieved, we have that argument c is not accepted, i.e., $\Delta_{\text{GR}}^{\text{cred}}(F_{\mathcal{E}}, c) = n$. Let us now assume that $\hat{\mathcal{E}} = \{e_1, e_2\}$, i.e., it is not possible to retrieve e_3 and e_4 ; then we have $\Delta_{\text{GR}}^{\text{cred}}(F_{\hat{\mathcal{E}}}, c) = y$. The question now is, which pieces of evidence must **Ag** attempt to retrieve in order to come to the same conclusion as $F_{\hat{\mathcal{E}}}$ (recall that $\hat{\mathcal{E}}$ is not known to the agent)? Let us consider some scenarios:

- **Ag** can attempt to retrieve the entire set of evidence $E_1 = \mathcal{E} = \{e_1, e_2, e_3, e_4\}$. He would then retrieve e_1 and e_2 and learn that e_3 and e_4 are unavailable. Thus, **Ag** knows that $F_{E_1 \cap \hat{\mathcal{E}}} = F_{\hat{\mathcal{E}}}$ and therefore trivially $\Delta_{\text{GR}}^{\text{cred}}(F_{E_1 \cap \hat{\mathcal{E}}}, c) = \Delta_{\text{GR}}^{\text{cred}}(F_{\hat{\mathcal{E}}}, c)$. The cost for his attempts of retrieval then is $\mu(E_1) = 2 + 6 + 3 + 9 = 20$.
- **Ag** can retrieve $E_2 = \{e_1, e_2\}$ to obtain $F_{E_2 \cap \hat{\mathcal{E}}} = F_{E_1 \cap \hat{\mathcal{E}}}$ with cost only $\mu(E_2) = 2 + 6 = 8$. However, as **Ag** does not know whether any of the other pieces of evidence e_3 and e_4 are actually unavailable, he cannot be sure that $\Delta_{\text{GR}}^{\text{cred}}(F_{E_2 \cap \hat{\mathcal{E}}}, c) = \Delta_{\text{GR}}^{\text{cred}}(F_{\hat{\mathcal{E}}}, c)$. For example, e_3 could be available, changing the acceptability status of c due to the presence of d .
- **Ag** can attempt to retrieve $E_3 = \{e_1, e_2, e_3\}$ to again obtain the same framework $F_{E_3 \cap \hat{\mathcal{E}}} = F_{E_1 \cap \hat{\mathcal{E}}}$ with cost $\mu(E_3) = 2 + 6 + 3 = 11$. As he learns that e_3 is not available, he can be sure that $\Delta_{\text{GR}}^{\text{cred}}(F_{E_3 \cap \hat{\mathcal{E}}}, c) = \Delta_{\text{GR}}^{\text{cred}}(F_{\hat{\mathcal{E}}}, c)$, independently of whether e_4 is available. Obviously searching for E_3 is better for **Ag** than searching for E_1 due to the lower cost.
- **Ag** can further minimise the cost while still being sure that the answer remains the same. In fact, **Ag** can attempt to retrieve $E_4 = \{e_2, e_3\}$ with cost $\mu(E_4) = 6 + 3 = 9$. As only e_2 can be retrieved from E_4 , the framework $F_{E_4 \cap \hat{\mathcal{E}}}$ consists only of the argument c and we have $\Delta_{\text{GR}}^{\text{cred}}(F_{E_4 \cap \hat{\mathcal{E}}}, c) = y$. However, if e_1 would be searched for and retrieved, the result stays the same as both the attacker b of c and its defender a could be constructed.

In conclusion, the best strategy for **Ag** is to attempt to retrieve $E_4 = \{e_2, e_3\}$ with cost $\mu(E_4) = 9$ in order to learn that c is credulously accepted under the grounded semantics given the available evidence. So we have $\text{OPT}_{F, \hat{\mathcal{E}}}^{\text{cred, GR}}(c) = 9$.

Example 3. The situation introduced in Example 1 can be modelled with the AAFE $F = (\mathcal{A}, \mathcal{R}, \mathcal{E}, \delta, \mu)$, where

σ	σ -CRED-UAAFE	σ -SKEP-UAAFE
GR	NP-c	NP-c
AD	Π_2^P -c	trivial
CP	Π_2^P -c	NP-c
ST	Π_2^P -c	Σ_2^P -c

Table 1. Summary of complexity results where NP-c stands for “NP-complete”.

5. Related works

The idea of using evidence to determine active arguments is not new in the computational argumentation community. In [3] one of the first approaches that uses evidence for this purpose was presented. There, the authors present a structured argumentation system based on DeLP [8] where arguments are pre-compiled (i. e. built beforehand) and then the available evidence is used to activate some of these pre-compilations. The work in [12] further generalises this idea of using evidence to activate pre-existing argument structures in the context of abstract argumentation. Like in the AAFEs, [12] characterises a set of evidence as a set of pieces of information, and establishes which pieces of evidence are required to activate an argument. However, differently from us, that work mostly focuses on providing a formal characterisation of the dynamics of the elements of the framework.

More closely related to our work is [4], where the problem being tackled there motivated our research. In [4] the authors use a simplified version of the framework presented in [12] but extended to consider (like in this paper) that the evidence associated with the arguments has to be retrieved, and such retrieval comes at a cost. Similarly to us, they aim to minimise the evidence retrieval cost incurred for determining the acceptance status of an argument. However, they do not provide a formal characterisation of the task as an optimisation problem nor study its complexity. Instead, they present an algorithm adopting a heuristic-based pruning technique for the construction of argumentation trees. Also, differently from us, they focus on a single semantics which is derived from the one adopted by DeLP and is also similar to the grounded semantics.

6. Summary and Conclusion

We introduced abstract argumentation frameworks with fallible evidence as a generalisation of Dung’s abstract argumentation framework that models arguments being backed by pieces of evidence, which in turn may be available or not and have an associated cost when being attempted to be retrieved. As the main contribution, we formulated an optimisation problem characterising reasoning with AAFEs wrt. admissible, complete, grounded and stable semantics, studied its computational complexity, and showed that complexity rises one level in the polynomial hierarchy compared to the classical case.

Our findings can be particularly used to justify the choice of [4] to develop a heuristic algorithm for solving problems very similar to ours. As finding the optimal choice of which pieces of evidence to attempt to retrieve is intractable

even for grounded semantics, there is little hope of finding a polynomial-time algorithm; thus, heuristic approaches akin to [4] could be developed for AAFEs with the aim of decreasing the evidence retrieval cost. Part of current work is to establish the computational complexity of our problem wrt. preferred semantics, which we conjecture to be on the third level of the polynomial hierarchy.

Acknowledgments

The work reported here has been partially supported by the Deutsche Forschungsgemeinschaft (grant KE 1413/11-1) and by Universidad Nacional del Sur (grant 24/N046).

References

- [1] K. Atkinson, P. Baroni, M. Giacomin, A. Hunter, H. Prakken, C. Reed, G. R. Simari, M. Thimm, and S. Villata. Toward artificial argumentation. *AI Magazine*, 38(3):25–36, October 2017.
- [2] P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, editors. *Handbook of Formal Argumentation*. College Publications, 2018.
- [3] M. Capobianco, C. I. Chesñevar, and G. R. Simari. Argumentation and the dynamics of warranted beliefs in changing environments. *Autonomous Agents and Multi-Agent Systems*, 11(2):127–151, 2005.
- [4] A. Cohen, S. Gottifredi, and A. J. García. A heuristic pruning technique for dialectical trees on argumentation-based query-answering systems. In *Proceedings of FQAS’19*, pages 101–113, 2019.
- [5] A. Cohen, S. Gottifredi, A. J. García, and G. R. Simari. A survey of different approaches to support in argumentation systems. *Knowledge Engineering Review*, 29(5):513–550, 2014.
- [6] P. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n -person games. *Artificial Intelligence*, 77(2):321–357, 1995.
- [7] W. Dvořák and P. E. Dunne. Computational problems in formal argumentation and their complexity. In *Handbook of Formal Argumentation*. College Publications, 2018.
- [8] A. J. García and G. R. Simari. Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming*, 4(1-2):95–138, 2004.
- [9] S. Gottifredi, A. Cohen, A. J. Garcia, and G. R. Simari. Characterizing acceptability semantics of argumentation frameworks with recursive attack and support relations. *Artificial Intelligence*, 262:336–368, 2018.
- [10] S. Nielsen and S. Parsons. A generalization of Dung’s abstract framework for argumentation: Arguing with sets of attacking arguments. In *Proceedings of ArgMAS’06*, pages 54–73, Berlin, Heidelberg, 2007. Springer.
- [11] C. Papadimitriou. *Computational Complexity*. Addison-Wesley, 1994.
- [12] N. D. Rotstein, M. O. Moguillansky, A. J. García, and G. R. Simari. A dynamic argumentation framework. In *Proceedings of COMMA’10*, pages 427–438, 2010.