On Explainable Acceptance in Probabilistic and Incomplete Abstract Argumentation Frameworks

(Discussion Paper)

Gianvincenzo Alfano[1,†], Marco Calautti[2,†], Sergio Greco[1,†], Francesco Parisi[1,†] and Irina Trubitsyna[1,*,†]

[1]*Department of Informatics, Modeling, Electronics and System Engineering (DIMES), University of Calabria, Rende, Italy*
[2]*Department of Computer Science (DI), University of Milan, Milan, Italy*

Abstract

Dung's Argumentation Framework (AF) has been extended in several directions, including the possibility of representing uncertainty about the existence of arguments and attacks. In this regard, two main proposals have been introduced in the literature: Probabilistic Argumentation Framework (PrAF) and Incomplete Argumentation Framework (iAF). PrAF is an extension of AF with probability theory, thus representing quantified uncertainty. In contrast, iAF represents unquantified uncertainty, that is it can be seen as a special case where we only know that some elements (arguments or attacks) are uncertain. We discuss the problem of computing the probability that a given argument is accepted in PrAF, which is based on the concept of probabilistic explanation for any given (probabilistic) extension [1]. Our approach can be extended to iAF, as it can be viewed as a special case of PrAF where uncertain elements have associated a probability equal to 1/2.

Keywords

Formal Argumentation Theory, Explainable AI, Probabilistic Argumentation Framework.

1. Introduction

The abstract Argumentation Framework (AF) is a simple, yet powerful formalism for modeling disputes between two or more agents [2]. An AF consists of a set of *arguments* and a binary *attack* relation over the set of arguments that specifies the interactions between arguments: intuitively, if argument a attacks argument b, then b is acceptable only if a is not. Hence, arguments are abstract entities whose role is entirely determined by the interactions specified by the attack relation.

Recently, there has been an increasing interest in extending argumentation frameworks to manage uncertain information. This has been carried out by either considering quantified uncertainty about the existence of arguments and attacks, thus combining formal argumentation with probability theory, or considering unquantified uncertainty by explicitly denoting the elements (arguments and attacks) which are uncertain. In fact, Probabilistic Argumentation [3] can be viewed as part of the several proposals that have been made in the last decades for extending reasoning tasks in AI frameworks with probabilities. These include for instance Probabilistic SAT (PSAT) [4], Probabilistic Logic [5], Probabilistic Logic Programming [6], and Probabilistic Databases [7].

One of the most popular approaches based on probability theory for modeling the uncertainty is the so called *constellations* approach [8, 9, 10, 11, 12], where alternative scenarios, called *possible*

8th Workshop on Advances in Argumentation in Artificial Intelligence, co-located with the 23rd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2024), 25-28 November, 2024, Bolzano, Italy

*Corresponding author.
† These authors contributed equally.
g.alfano@dimes.unical.it (G. Alfano); marco.calautti@unimi.it (M. Calautti); greco@dimes.unical.it (S. Greco); fparisi@dimes.unical.it (F. Parisi); i.trubitsyna@dimes.unical.it (I. Trubitsyna)
https://gianvincenzoalfano.net/ (G. Alfano); https://www.unimi.it/it/ugov/person/marco-scalautti (M. Calautti); https://people.dimes.unical.it/sergiogreco/ (S. Greco); http://wwwinfo.deis.unical.it/~parisi/ (F. Parisi); https://sites.google.com/dimes.unical.it/itrubitsyna/home (I. Trubitsyna)
0000-0002-7280-4759 (G. Alfano); 0000-0003-0921-4040 (M. Calautti); 0000-0003-2966-3484 (S. Greco); 0000-0001-9977-1355 (F. Parisi); 0000-0002-9031-0672 (I. Trubitsyna)

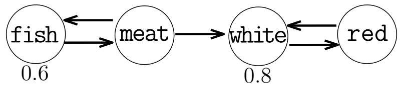

Figure 1: Probabilistic argumentation framework Δ of Example 1.

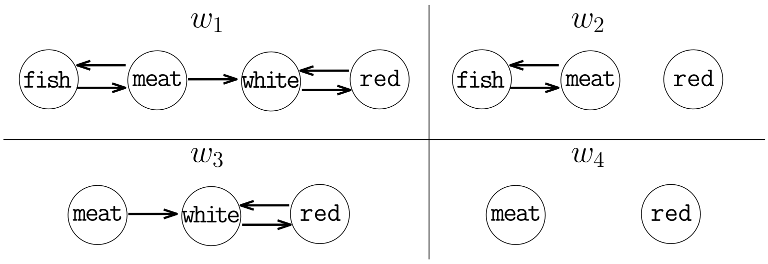

Figure 2: Possible worlds of the probabilistic argumentation framework Δ of Example 1.

worlds, are associated with probabilities. In particular, in a *Probabilistic Argumentation Framework* (PrAF) [12, 13, 14, 15, 16, 17, 18] a probability distribution function (PDF) on the set of possible worlds is entailed by the probabilities that are associated with arguments and attacks.

Example 1. Consider a PrAF $\Delta = \langle\{\texttt{fish}, \texttt{meat}, \texttt{white}, \texttt{red}\}, \{(\texttt{fish}, \texttt{meat}), (\texttt{meat}, \texttt{fish}), (\texttt{meat}, \texttt{white}), (\texttt{white}, \texttt{red}), (\texttt{red}, \texttt{white})\}, \{\texttt{fish/0.6}, \texttt{white/ 0.8}\}\rangle$, whose corresponding graph is shown in Figure 1, where nodes and edges represent arguments and attacks, respectively, and probabilities different from 1 are specified nearby them. For the sake of brevity, we do not specify the probabilities of certain elements in Δ (all the elements different from `fish` and `white` have probability 1). Intuitively, Δ describes what a person is going to have for lunch as follows. They will have either `fish` or `meat`, and will drink either `white` wine or `red` wine. However, if they will have `meat`, then they will not drink `white` wine. Furthermore, the probability that `fish` is available is 0.6, whereas the probability that `white` wine is available is 0.8. □

Intuitively, PrAF is a combination of two powerful approaches to reasoning and decision making: probabilistic reasoning and abstract argumentation. Probabilities are assigned to arguments and attacks to indicate their degree of uncertainty. One of the benefits of probabilistic abstract argumentation is its ability to handle quantified uncertainty in the analysis. In fact, PrAF can help to model and analyze situations where there is uncertainty by capturing both the relationships between arguments and the uncertainty degrees of arguments and attacks.

Several argumentation semantics—e.g. *grounded* (gr), *complete* (co), *preferred* (pr), *stable* (st), and *semi-stable* (sst)—have been defined for AFs, leading to the characterization of *σ-extensions*, which intuitively consist of the sets of arguments that can be collectively accepted under semantics σ. Consider for instance the deterministic version of the PrAF in Example 1, obtained by assuming that all arguments are certain (i.e. they have probability 1). Considering the preferred semantics, the pr-extensions are $E_1 = \{\texttt{fish}, \texttt{white}\}$, $E_2 = \{\texttt{fish}, \texttt{red}\}$, and $E_3 = \{\texttt{meat}, \texttt{red}\}$.

The semantics of a PrAF is given by considering all possible worlds (i.e. AFs) obtained by removing consistent subsets of the probabilistic elements. Here, for consistent subset we mean any subset of probabilistic elements (arguments and attacks) whose deletion from the initial framework results in an AF (for instance we cannot delete an argument without also deleting the attacks towards or from that argument). Every possible world has associated a probability value derived from the probabilities of the elements that have been kept or removed. Moreover, every possible world admits a set of σ-extensions. The probability of a possible world w is computed by multiplying the probabilities of the elements occurring in w and the complement to 1 of the probabilities of the elements not occurring in w.

Example 2. Continuing with Example 1, the possible worlds of Δ are shown in Figure 2. The probability of a possible world w_i is obtained by multiplying the probabilities $P(a)$ of each argument a occurring in w_i and the probabilities $(1 - P(b))$ of every argument b not occurring in w_i. Since $P(\texttt{fish}) = 0.6$, $P(\texttt{white}) = 0.8$, and $P(\texttt{meet}) = P(\texttt{red}) = 1$, the probabilities of w_1, w_2, w_3, and w_4 are $0.6 \cdot 1 \cdot 0, 8 \cdot 1 = 0.48$, $0.6 \cdot 1 \cdot 0.2 \cdot 1 = 0.12$, $0.4 \cdot 1 \cdot 0.8 \cdot 1 = 0.32$, and $0.4 \cdot 1 \cdot 0.2 \cdot 1 = 0.08$. Since w_1 coincides with the deterministic version of Δ, its pr-extensions are E_1, E_2, and E_3 given earlier. The pr-extensions of w_2 are E_2 and E_3, while w_3 and w_4 admit only E_3 as their preferred extension. $\square$

2. Explanation-based Probabilistic Acceptance

Interesting problems recently investigated in the context of probabilistic argumentation are *probabilistic credulous acceptance* (PrCA) and *probabilistic skeptical acceptance* (PrSA) [19, 15]. In particular, given a PrAF Δ whose set of arguments is A, a goal argument $g \in A$ and a semantics σ, PrCA is the problem of computing the probability $PrCA^{\sigma}_{\Delta}(g)$ that the goal g is credulously accepted, that is, there is a possible world w of Δ such that g belongs to a σ-extension of w. Moreover, PrSA is the problem of computing the probability $PrSA^{\sigma}_{\Delta}(g)$ that the goal g is skeptically accepted, that is, g is credulously accepted and belongs to all σ-extensions of w.

However, the answer to these problems does not reflect our intuition of probability that a goal argument is accepted under a given semantics. For instance, considering the PrAF Δ of Figure 1, the probability that meat is credulously accepted under preferred semantics is 1, whereas the probability that meat is skeptically accepted under preferred semantics is 0.4. However, the fact that $PrCA^{\texttt{pr}}_{\Delta}(\texttt{meat}) = 1$ does not mean that the person in our example will surely have meat in any scenario (i.e. possible world). In fact, even if meat belongs to at least one preferred extension of every world of Δ, we expect that the probability of acceptance of meat should be lower than 1. Indeed, in any possible world, the presence of multiple extensions is an additional source of uncertainty that should be taken into account.

To better grasp the issue behind the probability of credulous acceptance, consider the following AF (where all elements are certain): $\Lambda = \langle\{\texttt{fish}, \texttt{meat}\}, \{(\texttt{fish}, \texttt{meat}), (\texttt{meat}, \texttt{fish})\}\rangle$ saying that fish and meat are mutually exclusive. Again, the probability that a person will have meat is 1, under probabilistic credulous acceptance, when considering the preferred semantics, whereas we believe that the expected answer should be 0.5. Moreover, if we consider AF w_1 of Example 2 (that can be obtained from Λ by adding arguments white and red and attacks $(\texttt{white}, \texttt{red})$, $(\texttt{red}, \texttt{white})$ and $(\texttt{meat}, \texttt{white})$) we expect that the probability of having meat does not change.

With the aim of providing more intuitive answers for probabilistic acceptance, a new problem called *Probabilistic Acceptance* (denoted as PrA, or PrA$[\sigma]$ when considering a given semantics σ) has beed investigated [1, 20], i.e. given a PrAF Δ and a goal argument g, compute the probability that g is accepted under semantics $\sigma \in \{\texttt{gr}, \texttt{co}, \texttt{pr}, \texttt{st}, \texttt{sst}\}$. In this framework, acceptance still relies on σ-extensions but, differently from credulous acceptance, we get rid of the assumption that no uncertainty exists at the level of the extensions of a world (i.e. AF). In more detail, PrA$[\sigma]$ implicitly assumes that a PDF over the set of σ-extensions of any AF (and thus of any possible world of PrAF Δ) is defined. Thus, a concrete instance of PrA is obtained after defining such a PDF. This can be carried out by exploiting the concept of *explanation* for an extension.

In general, in abstract argumentation an explanation for an extension E can be viewed as a (possibly minimal) subset $S \subseteq E$ such that, by assuming that the elements in S are acceptable, it turns out that all elements in $E \setminus S$ are "univocally" determined as acceptable (w.r.t. the underlying semantics). For instance, considering AF w_1 of Example 2, for the preferred extension $E = \{\texttt{meat}, \texttt{red}\}$, the set $S_1 = \{\texttt{meat}\}$ is an explanation for E, whereas the set $S_2 = \{\texttt{red}\}$ is not. In our perspective, explanations are sequences of "choices" to be made to justify how an extension is obtained and they provide a tool to assign probabilities to extensions. Integrating explanations in argumentation systems is important for enhancing the argumentation and persuasion capabilities of software agents [21, 22, 23]. For this reasons, several researchers have explored how to deal with explanations in formal argumentation [24, 25, 26].

An instantiation of PrA$[\sigma]$ where the PDF over the set of σ-extensions of a world relies on the concept of *explanation* is called *Explanation-based Probabilistic Acceptance* problem, and denoted by PrEA (and $\mathsf{PrEA}[\sigma]$ for a specific semantics σ). Intuitively, an explanation for an σ-extension E is a *sequence of arguments* occurring in E that "justify" E. Every explanation is associated with a probability entailed by the possible choices that can be made when building it. These choices must be consistent with an ordering entailed by the strongly connected components of the given AF, and they are used to guide the construction of an extension. The sum of the probabilities of the explanations for an extension E gives the probability of E. Thus, we still assign to each possible world w of Δ a probability as in the standard way, but in addition propose to distinguish among extensions of a given world w by associating with them a probability based on explanations.

Example 3. Continuing with Example 1, take for instance the possible world w_1 having probability 0.48. As shown in Example 2, w_1 has three $\mathtt{pr}$-extensions, namely E_1, E_2 and E_3. As shown in [1], in this case, for each extension there is only one explanation. In particular, $X_1 = \langle \mathtt{fish}, \mathtt{white} \rangle$ is the explanation for E_1. The intuition of explanation X_1 is that, considering that the AF consists of two strongly connected components, we first choose `fish` (with probability $1/2$ as we can only choose between `fish` and `meat`) in the first component and determine that `meat` cannot belong to the extension; then we choose `white` (with probability $1/2$ as we can only choose between `white` and `red`) in the second component, obtaining that X_1 has probability $1/2 \cdot 1/2 = 1/4$. Analogously, $X_2 = \langle \mathtt{fish}, \mathtt{red} \rangle$ is the only explanation for E_2 with probability $1/2 \cdot 1/2 = 1/4$. Considering explanation $X_3 = \langle \mathtt{meat} \rangle$ for extension E_3, we have that we first choose `meat` with probability $1/2$ as it belongs to the first component, and we can only choose between `fish` and `meat`. Next, since we determine that `fish` and `white` cannot belong to the extension, whereas `red` does, the probability of X_3 turns out to be $1/2$. Since the probabilities of X_1, X_2 and X_3 are $1/4, 1/4$ and $1/2$, respectively, the probabilities associated with E_1, E_2 and E_3 in the world w_1 are $1/4, 1/4$ and $1/2$, respectively. Moreover, since E_1 is not an extension of any other possible world, the probability of E_1 in Δ is $1/4 \cdot 0.48 = 0.12$. It turns out that the answer to $\mathsf{PrEA}[\mathtt{pr}]$ for `meat` is 0.70, while that for `fish` is 0.30. □

The definition of *Explanation-based Probabilistic Acceptance* has been also carried out to another argumentation framework extending AF that has received an increasing attention in the last years and is tightly related to PrAF, that is, to incomplete AF (iAF) [27, 28]. This follows from the fact that iAF can be viewed as a special case of PrAF where uncertain elements have associated a probability equal to 1/2.

Acknowledgements

We acknowledge the support from PNRR MUR project PE0000013-FAIR and PE0000014-SERICS, project Tech4You ECS0000009, and MUR project PRIN 2022 EPICA (CUP H53D2300 3660006).

References

[1] G. Alfano, M. Calautti, S. Greco, F. Parisi, I. Trubitsyna, Explainable acceptance in probabilistic and incomplete abstract argumentation frameworks, Artif. Intell. 323 (2023) 103967.
[2] P. M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artif. Intell. 77 (1995) 321–358.
[3] A. Hunter, S. Polberg, N. Potyka, T. Rienstra, M. Thimm, Probabilistic argumentation: A survey, Handbook of Formal Argumentation 2 (2021) 397–441.
[4] G. F. Georgakopoulos, D. J. Kavvadias, C. H. Papadimitriou, Probabilistic satisfiability, J. Complex. 4 (1988) 1–11.
[5] N. J. Nilsson, Probabilistic logic revisited, Artif. Intell. 59 (1993) 39–42.
[6] F. Riguzzi, T. Swift, A survey of probabilistic logic programming, in: M. Kifer, Y. A. Liu (Eds.), Declarative Logic Programming: Theory, Systems, and Applications, ACM / Morgan & Claypool, 2018, pp. 185–228.

[7] D. Suciu, D. Olteanu, C. Ré, C. Koch, Probabilistic Databases, Synthesis Lectures on Data Management, Morgan & Claypool Publishers, 2011.
[8] P. M. Dung, P. M. Thang, Towards (probabilistic) argumentation for jury-based dispute resolution, in: Proc. of Int. Conf. on Computational Models of Argument (COMMA), 2010, pp. 171–182.
[9] T. Rienstra, Towards a probabilistic Dung-style argumentation system, in: Proc. of Int. Conf. on Agreement Technologies (AT), 2012, pp. 138–152.
[10] D. Doder, S. Woltran, Probabilistic argumentation frameworks-A logical approach, in: Proc. of Int. Conf. on Scalable Uncertainty Management (SUM), 2014, pp. 134–147.
[11] A. Hunter, Some foundations for probabilistic abstract argumentation, in: Proc. of Int. Conf. on Computational Models of Argument (COMMA), 2012, pp. 117–128.
[12] H. Li, N. Oren, T. J. Norman, Probabilistic argumentation frameworks, in: Proc. of Int. Workshop on Theorie and Applications of Formal Argumentation (TAFA), 2011, pp. 1–16.
[13] B. Fazzinga, S. Flesca, F. Parisi, On the complexity of probabilistic abstract argumentation frameworks, ACM Trans. on Comput. Log. 16 (2015) 22:1–22:39.
[14] B. Fazzinga, S. Flesca, F. Parisi, On efficiently estimating the probability of extensions in abstract argumentation frameworks, Int. J. Approx. Reason. 69 (2016) 106–132.
[15] B. Fazzinga, S. Flesca, F. Furfaro, Complexity of fundamental problems in probabilistic abstract argumentation: Beyond independence, Artif. Intell. 268 (2019) 1–29.
[16] N. Potyka, A polynomial-time fragment of epistemic probabilistic argumentation, Int. J. Approx. Reason. 115 (2019) 265–289.
[17] R. Riveret, N. Oren, G. Sartor, A probabilistic deontic argumentation framework, Int. J. Approx. Reason. 126 (2020) 249–271.
[18] P. Dondio, Toward a computational analysis of probabilistic argumentation frameworks, Cybern. Syst. 45 (2014) 254–278.
[19] B. Fazzinga, S. Flesca, F. Furfaro, Credulous and skeptical acceptability in probabilistic abstract argumentation: complexity results, Intelligenza Artificiale 12 (2018) 181–191.
[20] G. Alfano, M. Calautti, S. Greco, F. Parisi, I. Trubitsyna, Explainable acceptance in probabilistic abstract argumentation: Complexity and approximation, in: Proc. of the 17th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR), 2020, pp. 33–43.
[21] B. Moulin, H. Irandoust, M. Bélanger, G. Desbordes, Explanation and argumentation capabilities: Towards the creation of more persuasive agents, Artif. Intell. Rev. 17 (2002) 169–222.
[22] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, Artif. Intell. 267 (2019) 1–38.
[23] X. Fan, F. Toni, On computing explanations in argumentation, in: Proc. of AAAI Conf. on Artificial Intelligence, 2015, pp. 1496–1502.
[24] M. Ulbricht, J. P. Wallner, Strong explanations in abstract argumentation, in: Proc. of Thirty-Fifth AAAI Conf. on Artificial Intelligence, 2021, pp. 6496–6504.
[25] G. Brewka, M. Ulbricht, Strong explanations for nonmonotonic reasoning, in: Description Logic, Theory Combination, and All That - Essays Dedicated to Franz Baader on the Occasion of His 60th Birthday, volume 11560, 2019, pp. 135–146.
[26] Z. G. Saribatur, J. P. Wallner, S. Woltran, Explaining non-acceptability in abstract argumentation, in: Proc. of the 24th Eur. Conf. on Artificial Intelligence (ECAI), volume 325, 2020, pp. 881–888.
[27] D. Baumeister, D. Neugebauer, J. Rothe, H. Schadrack, Verification in incomplete argumentation frameworks, Artif. Intell. 264 (2018) 1–26.
[28] G. Alfano, S. Greco, F. Parisi, I. Trubitsyna, Incomplete argumentation frameworks: Properties and complexity, in: Proc. of AAAI Conf. on Artificial Intelligence, 2022, pp. 5451–5460.