Explainable Acceptance in Probabilistic Abstract Argumentation: Complexity and Approximation

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

[1] {g.alfano, greco, fparisi}@dimes.unical.it
Department of Informatics, Modeling, Electronics and System Engineering
University of Calabria, Italy

[2] Information Engineering and Computer Science Department
University of Trento, Italy

17th International Conference on Principles of Knowledge Representation and Reasoning

September 12-18, 2020

Rhodes, Greece

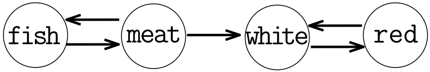
Argumentation in AI

- A general way for representing arguments and relationships between them
- It allows representing dialogues, making decisions, and handling inconsistency and uncertainty

Abstract Argumentation Framework (AF) [**Dung1995**]: arguments are abstract entities (no attention is paid to their internal structure) that may attack and/or be attacked by other arguments

Example (a simple AF)

John will have either `fish` or `meat`, and will drink either `white` wine or `red` wine. However, if he will have `meat`, then he will not drink `white` wine.



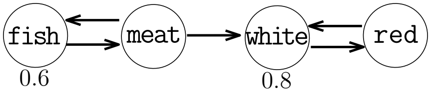
Probabilistic Abstract Argumentation Framework

- Arguments and attacks can be uncertain

Example (a simple PrAF)

There is some uncertainty:

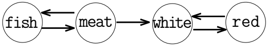
- about the fact that John will have `fish`
- about the fact that John will drink `white` wine



Argumentation Semantics for Deterministic AFs

In the deterministic setting, several semantics (such as *complete*, *preferred*, *stable*, *semi-stable*, and *grounded*) have been proposed to identify "reasonable" sets of arguments, called *extensions*.

Example (AF $\mathcal{A}_0$)



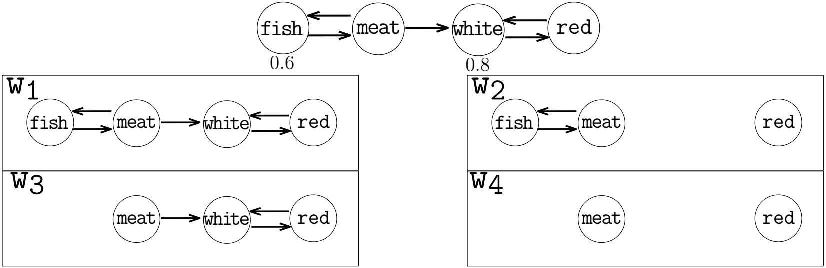
Semantic $\mathcal{S}$	Set of extensions of $\mathcal{A}_0$
complete ($\mathtt{co}$)	$\{\emptyset, \{\mathtt{fish}\}, \{\mathtt{red}\}, \{\mathtt{fish}, \mathtt{white}\}, \{\mathtt{fish}, \mathtt{red}\}, \{\mathtt{meat}, \mathtt{red}\}\}$
preferred ($\mathtt{pr}$)	$\{\{\mathtt{fish}, \mathtt{white}\}, \{\mathtt{fish}, \mathtt{red}\}, \{\mathtt{meat}, \mathtt{red}\}\}$
stable ($\mathtt{st}$)	$\{\{\mathtt{fish}, \mathtt{white}\}, \{\mathtt{fish}, \mathtt{red}\}, \{\mathtt{meat}, \mathtt{red}\}\}$
semi-stable ($\mathtt{sst}$)	$\{\{\mathtt{fish}, \mathtt{white}\}, \{\mathtt{fish}, \mathtt{red}\}, \{\mathtt{meat}, \mathtt{red}\}\}$
grounded ($\mathtt{gr}$)	$\{\emptyset\}$

- An argument g is credulously accepted w.r.t. $\mathcal{A}$ under semantics $\mathcal{S}$ iff it appear in at least an $\mathcal{S}$-extension of $\mathcal{A}$.

Argumentation Semantics for PrAFs

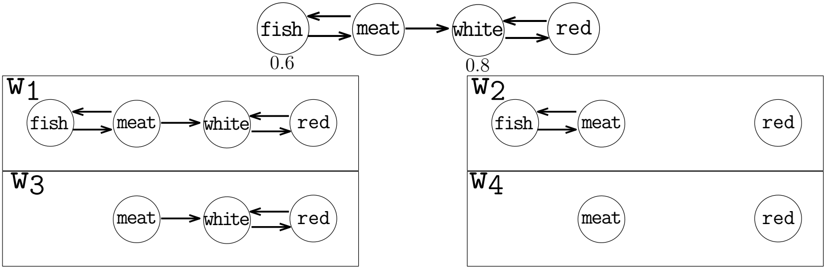
The meaning of a PrAF is given in terms of possible worlds that represents a probable (deterministic) scenario consisting of some subset of the arguments and defeats of the PrAF.

Example (Possible worlds of our PrAF)



Argumentation Semantics for PrAFs

Example (Possible worlds of our PrAF)



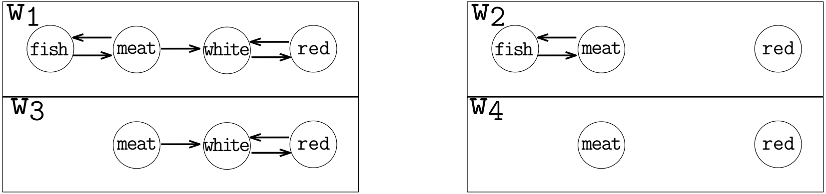
(Probabilistic credulous acceptance)

Given a PrAF Δ, an argument $g \in A$, the probability $PrCA^{\mathcal{S}}_{\Delta}(g)$ that g is credulously acceptable w.r.t $\mathcal{S}$ semantics is

$$PrCA^{\mathcal{S}}_{\Delta}(g) = \sum_{\substack{w \in pw(\Delta) \wedge \\ \exists E \in \mathcal{S}(w) \ s.t. \ g \in E}} I(w).$$

Argumentation Semantics for PrAFs

Example (Possible worlds of our PrAF)



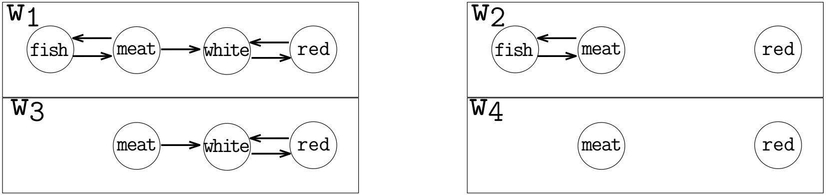
$$PrCA^{\mathcal{S}}_{\Delta}(g) = \sum_{\substack{w \in pw(\Delta) \wedge \\ \exists E \in \mathcal{S}(w) \, s.t. \, g \in E}} I(w).$$

w	$I(w)$	$E_1 = \{\mathtt{f}, \mathtt{w}\}$	$E_2 = \{\mathtt{f}, \mathtt{r}\}$	$E_3 = \{\mathtt{m}, \mathtt{r}\}$
w_1	0.48	✓	✓	✓
w_2	0.12		✓	✓
w_3	0.32			✓
w_4	0.08			✓

$PrCA^{\mathcal{S}}_{\Delta}(\mathtt{fish}) = I(w_1) + I(w_2) = 0.6$

Argumentation Semantics for PrAFs

Example (Possible worlds of our PrAF)



$$PrCA^{\mathcal{S}}_{\Delta}(g) = \sum_{\substack{w \in pw(\Delta) \wedge \\ \exists E \in \mathcal{S}(w) \text{ s.t. } g \in E}} I(w).$$

w	$I(w)$	$E_1 = \{\mathtt{f}, \mathtt{w}\}$	$E_2 = \{\mathtt{f}, \mathtt{r}\}$	$E_3 = \{\mathtt{m}, \mathtt{r}\}$
w_1	0.48	✓	✓	✓
w_2	0.12		✓	✓
w_3	0.32			✓
w_4	0.08			✓

$PrCA^{\mathcal{S}}_{\Delta}(\mathtt{meat}) = 1$

Non always reasonable

What we propose

A different approach called *Probabilistic Acceptance*

(Probabilistic Acceptance)

Given a PrAF $\Delta = \langle A, \Sigma, P \rangle$ and an argument $g \in A$, the probability $PrA_{\Delta}^{\mathcal{S}}(g)$ that g is acceptable w.r.t. semantics $\mathcal{S}$ is

$$PrA_{\Delta}^{\mathcal{S}}(g) = \sum_{\substack{w \in pw(\Delta) \wedge \\ E \in \mathcal{S}(w) \wedge g \in E}} I(w) \cdot Pr(E, w, \mathcal{S})$$

where $Pr(\cdot, w, \mathcal{S})$ is a PDF over the set $\mathcal{S}(w)$.

We show that a possible way to obtain $Pr(\cdot, w, \mathcal{S})$ is through ***explanations***, obtaining an instantiation of the above problem that we call *Explanation-based Probabilistic Acceptance* ($PrEA_{\Delta}^{\mathcal{S}}(g)$).

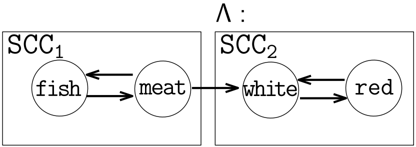
Example

In our example we have that :
$PrEA_{\Delta}^{\mathcal{S}}(\texttt{fish}) = 0.3$ $PrEA_{\Delta}^{\mathcal{S}}(\texttt{meat}) = 0.7$

Explanations: Intuitions and Example

- A sequence of necessary suggestions useful to construct a given extension.
- A sequence of choices (guided by ordering SCCs) to obtain the extension.

Example (Explanation for the extension $\{\mathtt{meat}, \mathtt{red}\}$)

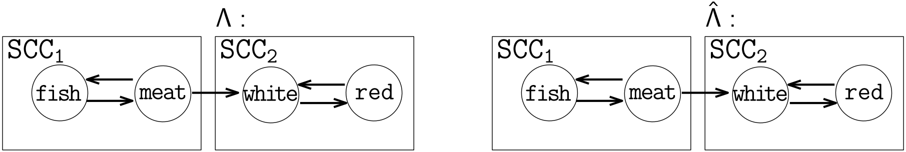


- For the stable extension $E = \{\mathtt{meat}, \mathtt{red}\}$ of Λ there is an explanation $X = \langle \mathtt{meat} \rangle$. **Why?**

Explanations: Intuitions and Example

- A sequence of necessary suggestions useful to construct a given extension.
- A sequence of choices (guided by ordering SCCs) to obtain the extension.

Example (Explanation for the extension $\{\mathtt{meat}, \mathtt{red}\}$)

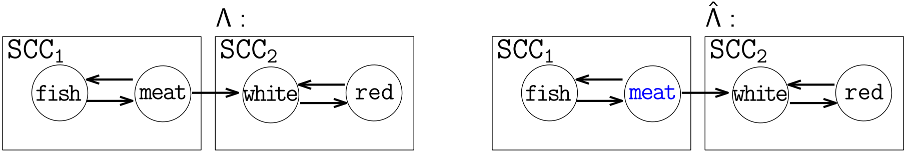


- For the stable extension $E = \{\mathtt{meat}, \mathtt{red}\}$ of Λ there is an explanation $X = \langle\mathtt{meat}\rangle$. **Why?**
- $\mathcal{GR}(\Lambda) = \{\emptyset\}$ does not help to determine any argument of the initial AF ($\hat{\Lambda} = \Lambda$).

Explanations: Intuitions and Example

- A sequence of necessary suggestions useful to construct a given extension.
- A sequence of choices (guided by ordering SCCs) to obtain the extension.

Example (Explanation for the extension $\{\texttt{meat}, \texttt{red}\}$)

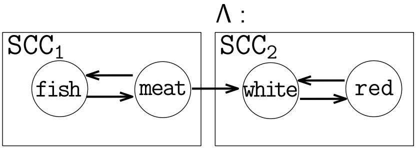


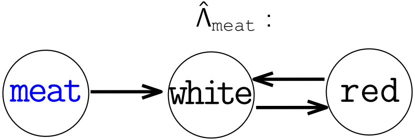
- meat can be chosen in the initial SCC of $\hat{\Lambda}$ w.r.t. E (which coincides with the initial SCC of $\hat{\Lambda}$).

Explanations: Intuitions and Example

- A sequence of necessary suggestions useful to construct a given extension.
- A sequence of choices (guided by ordering SCCs) to obtain the extension.

Example (Explanation for the extension $\{\texttt{meat}, \texttt{red}\}$)



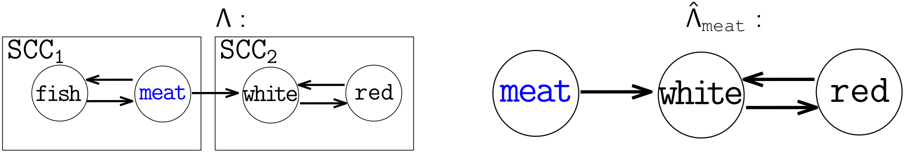


- meat can be chosen in the initial SCC of $\hat{\Lambda}$ w.r.t. E (which coincides with the initial SCC of $\hat{\Lambda}$).
- We look for an explanation for $\{\texttt{meat}, \texttt{red}\}$ w.r.t $\hat{\Lambda}_{\texttt{meat}}$.

Explanations: Intuitions and Example

- A sequence of necessary suggestions useful to construct a given extension.
- A sequence of choices (guided by ordering SCCs) to obtain the extension.

Example (Explanation for the extension $\{\texttt{meat}, \texttt{red}\}$)



- We look for an explanation for $\{\texttt{meat}, \texttt{red}\}$ w.r.t $\hat{\Lambda}_{\texttt{meat}}$.
- As $\mathcal{GR}(\hat{\Lambda}_{\texttt{meat}}) = \{\{\texttt{meat}, \texttt{red}\}\}$ we conclude that $X = \langle \texttt{meat} \rangle$ is an explanation for E.

Probabilities for explanations (1/2)

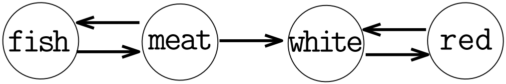
Since a given extension may have multiple explanations of different length, it is reasonable to assume that some explanations are preferred to others.

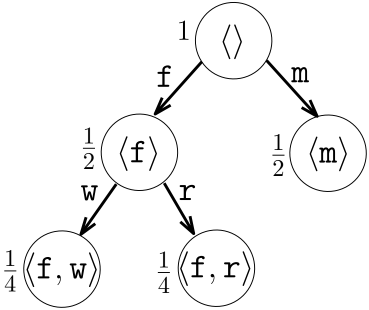
Probabilities for explanations (1/2)

To define probabilities of explanations, we use a *probabilistic trie.*

Example (Probabilistic Trie under preferred/stable/semi-stable semantics)

The AF Λ





$Pr(E_1 = \{\mathtt{fish}, \mathtt{white}\}, \Lambda, \mathcal{S}) = \frac{1}{4}$
$Pr(E_2 = \{\mathtt{fish}, \mathtt{red}\}, \Lambda, \mathcal{S}) = \frac{1}{4}$
$Pr(E_3 = \{\mathtt{meat}, \mathtt{red}\}, \Lambda, \mathcal{S}) = \frac{1}{2}$

$$Pr(\mathbf{S}, \Lambda, \mathcal{S}) = \sum_{X \in Exp_{\Lambda}^{\mathcal{S}}(\mathbf{S})} \pi(X)$$

Explanation-based Probabilistic Acceptance problem PrEA[$\mathcal{S}$]

(PrEA[$\mathcal{S}$] problem)

$$PrEA^{\mathcal{S}}_{\Delta}(g) = \sum_{\substack{w \in pw(\Delta) \wedge \\ E \in \mathcal{S}(w) \wedge g \in E}} I(w) \cdot Pr(E, w, \mathcal{S})$$

w	$I(w)$	$E_1 = \{\mathtt{f}, \mathtt{w}\}$ $Pr(E, w, _{\mathcal{ST}})$	$E_2 = \{\mathtt{f}, \mathtt{r}\}$ $Pr(E, w, _{\mathcal{ST}})$	$E_3 = \{\mathtt{m}, \mathtt{r}\}$ $Pr(E, w, _{\mathcal{ST}})$
w_1	0.48	1/4	1/4	1/2
w_2	0.12	0	1/2	1/2
w_3	0.32	0	0	1
w_4	0.08	0	0	1
		0.12	0.18	0.70

$$PrEA^{\mathcal{S}}_{\Delta}(\mathtt{fish}) = 0.12 + 0.18 = 0.3$$

$$PrEA^{\mathcal{S}}_{\Delta}(\mathtt{meat}) = 0.7$$

Exact and Approximate Complexity

(Theorem 1)

For $\mathcal{S} \in \{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$, PrA[$\mathcal{S}$] is $\mathsf{FP}^{\#\mathsf{P}}$-hard, even for acyclic PrAFs and for any chosen PDF.

It suggests that one would need to focus on approximations...

(Theorem 2)

Consider a semantics $\mathcal{S} \in \{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$. Unless NP $\subseteq$ BPP, there is no FPRAS for PrA[$\mathcal{S}$], even for acyclic PrAFs and for any chosen PDF.

(Theorem 3)

Let $\mathcal{S} \in \{\mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$. Unless NP $\subseteq$ BPP, there is no FPARAS for PrA[$\mathcal{S}$], for any chosen PDF.

Exact and Approximate Complexity

(Theorem 1)

For $\mathcal{S} \in \{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$, PrA[$\mathcal{S}$] is $\text{FP}^{\#\text{P}}$-hard, even for acyclic PrAFs and for any chosen PDF.

It suggests that one would need to focus on approximations...

(Theorem 2)

Consider a semantics $\mathcal{S} \in \{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$. Unless NP $\subseteq$ BPP, there is no FPRAS for PrA[$\mathcal{S}$], even for acyclic PrAFs and for any chosen PDF.

(Theorem 3)

Let $\mathcal{S} \in \{\mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$. Unless NP $\subseteq$ BPP, there is no FPARAS for PrA[$\mathcal{S}$], for any chosen PDF.

Approximate Complexity Results

It seems that all is dead! (Not properly so)
When $\mathcal{S} = \mathcal{GR}$ or when the input *PrAF* has no odd-length cycles, the use of explanations for devising a PDF over extensions allows us to construct an FPARAS.

	General PrAFs		PrAFs without odd cycles	
	FPRAS	FPARAS	FPRAS	FPARAS
$\mathcal{GR}$	$\times$	$\checkmark$	$\times$	$\checkmark$
$\mathcal{PR}$	$\times$	$\times$	$\times$	$\checkmark$
$\mathcal{ST}$	$\times$	$\times$	$\times$	$\checkmark$
$\mathcal{SST}$	$\times$	$\times$	$\times$	$\checkmark$

Devising an FPARAS

We report an FPARAS for the problem PrEA[$\mathcal{S}$], when either $\mathcal{S} = \mathcal{GR}$ or the input PrAF has no odd-length cycles.

Algorithm 1

Input: A PrAF $\Delta = \langle A, \Sigma, P \rangle$, a semantics $\mathcal{S}$, a goal argument $g \in A$, error parameter $\epsilon > 0$, and uncertainty parameter $0 < \delta < 1$.

Output: a random number p such that $PrEA_{\Delta}^{\mathcal{S}}(g) \in [p-\epsilon, p+\epsilon]$ with probability $1 - \delta$.

$n := \lceil \frac{1}{2\epsilon^2} \times \ln(\frac{2}{\delta}) \rceil$;
$c := 0$;
for $i \in \{1, \ldots, n\}$ **do**
  Choose $w \in pw(\Delta)$ with probability $I(w)$;
  Choose $E \in \mathcal{S}(w)$ with probability $Pr(E, w, \mathcal{S})$;
  if $g \in E$ **then**
    $c := c + 1$;
return $\frac{c}{n}$;

Devising an FPARAS

We report an FPARAS for the problem PrEA[$\mathcal{S}$], when either $\mathcal{S} = \mathcal{GR}$ or the input PrAF has no odd-length cycles.

Algorithm 1

Input: A PrAF $\Delta = \langle A, \Sigma, P \rangle$, a semantics $\mathcal{S}$, a goal argument $g \in A$, error parameter $\epsilon > 0$, and uncertainty parameter $0 < \delta < 1$.

Output: a random number p such that $PrEA_{\Delta}^{\mathcal{S}}(g) \in [p - \epsilon, p + \epsilon]$ with probability $1 - \delta$.

$n := \lceil \frac{1}{2\epsilon^2} \times \ln(\frac{2}{\delta}) \rceil$;
$c := 0$;
for $i \in \{1, \ldots, n\}$ **do**
    Choose $w \in pw(\Delta)$ with probability $I(w)$;
    Choose $E \in \mathcal{S}(w)$ with probability $Pr(E, w, \mathcal{S})$; $\longleftarrow$
    if $g \in E$ **then**
        $c := c + 1$;
return $\frac{c}{n}$;

Inapproximability for PrCA[$\mathcal{S}$]

Another issue for PrCA[$\mathcal{S}$] is...

(Theorem 6)

Consider a semantics $\mathcal{S} \in \{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$. Unless NP $\subseteq$ BPP, there is no FPRAS for PrCA[$\mathcal{S}$], even for acyclic PrAFs.

(Theorem 7)

Consider a semantics $\mathcal{S} \in \{\mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$. Unless *NP* $\subseteq$ *BPP*, there is no FPARAS for PrCA[$\mathcal{S}$], even for PrAFs without odd-length cycles.

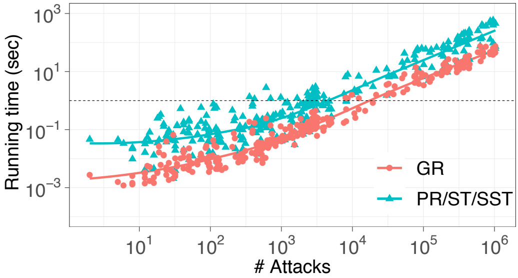
(Corollary 1)

The problem PrCA[$\mathcal{GR}$] admits an FPARAS.

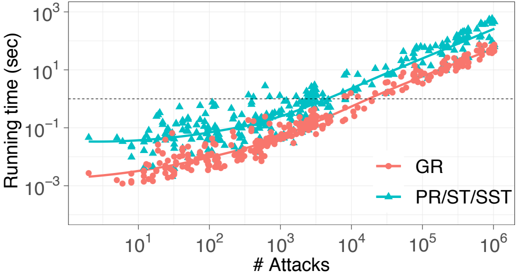
So, it seems to be all dead for PrCA[$\mathcal{S}$]...

Experimental Analysis

- Python prototype of Algorithm 1
- Generated PrAFs from AF benchmarks at ICCMA'19
- Results for 5 goals and $\epsilon = \delta = 5\%$



Experimental Analysis



(Results)

- Run time almost linearly on the number of attacks. It is lower for $\mathcal{GR}$ as only one extension exists (Alg.1 iterates once).
- Run time for the other semantics is not much higher (5.53) than that for $\mathcal{GR}$. Most PrAFs have a very large SCC containing 85% of the arguments on average, and thus the probabilistic trie of a world is not very deep.
- Alg.1 performs well ($<$ 1 sec) on large PrAFs (up to 10K attacks for almost 60% of PrAFs in the dataset).

Thank you!

... any ~~question~~ argument?