Explaining Argument Acceptance in ADFs

Tjitze Rienstra [1] Jesse Heyninck [2] Gabriele Kern-Isberner [3] Kenneth Skiba [4]
Matthias Thimm [4]

[1]Maastricht University

[2]Open Universiteit Heerlen

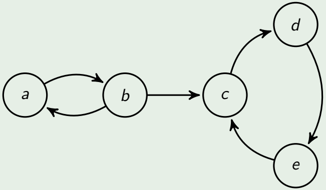
[3]Technische Universtät Dortmund

[4]FernUniversität Hagen

September 12, 2022

Abstract Argumentation - Basics

Example

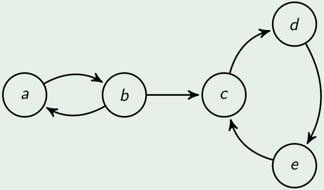


$$L_1 = \{a = \mathtt{I}, b = \mathtt{O}, c = \mathtt{U}, d = \mathtt{U}, e = \mathtt{U}\}$$
$$L_2 = \{a = \mathtt{O}, b = \mathtt{I}, c = \mathtt{O}, d = \mathtt{I}, e = \mathtt{O}\}$$
$$L_3 = \{a = \mathtt{U}, b = \mathtt{U}, c = \mathtt{U}, d = \mathtt{U}, e = \mathtt{U}\}$$

Dialogical Proof

Example

"Preferred Semantics as Socratic Discussion" (Caminada, Dvorak, Vesic, JLC 2016).



- PRO: I(d)
  "I have an admissible labelling in which d is labelled I."
- OPP: O(c)
  "But then in your labelling it must also be the case that d's attacker c is labelled O. Based on which grounds?"
- PRO: I(b)
  "c is labelled O because b is labelled I."
- OPP: O(a)
  "But then in your labelling it must also be the case that b's attacker a is labelled O. Based on which grounds?"
- PRO: I(b)
  "a is labelled O because b is labelled I."

There exists a dialogue where PRO wins iff the initial argument is credulously accepted under the preferred semantics.

Abstract Dialectical Frameworks

(Brewka, Woltran, KR 2010)

- An ADF is a tuple $D = (At, L, C)$ where
  - At is a finite set of atoms (also referred to as *arguments*)
  - $L \subseteq At \times At$ is a set of *links*
  - $C = \{\phi_x\}_{x \in At}$ is a set of *acceptance conditions* such that an atom $y \in At$ appears in ϕ_x only if $(y, x) \in L$.
- An *interpretation* of At is a function $v : At \to \{\mathbf{T}, \mathbf{F}, \mathbf{U}\}$.
- A *semantics* σ maps every ADF $D = (At, L, C)$ to a set $\sigma(D)$ of interpretations of At.

Abstract Dialectical Frameworks

- $\leq_i$ is the reflexive closure of the s.p.o. $<_i$ over $\{\mathbf{T}, \mathbf{F}, \mathbf{U}\}$ defined by $\mathbf{U} <_i \mathbf{T}$ and $\mathbf{U} <_i \mathbf{F}$.
- $v \leq_i u$ iff $v(x) \leq_i u(x)$ for all $x \in At$.
- $\sqcap$ is the meet of the complete meet-semilattice $(\{\mathbf{T}, \mathbf{F}, \mathbf{U}\}, \leq_i)$.
- $v \sqcap u$ is defined by $(v \sqcap u)(x) = v(x) \sqcap u(x)$ for all $x \in At$.
- $[v]_2$ denotes the set of two-valued interpretations u s.t. $v \leq_i u$.
- Given an ADF $D = (At, \{\phi_x\}_{x \in At})$ and three-valued-interpretation v, $\Gamma_D(v)$ denotes the three-valued interpretation defined by

$$\Gamma_D(v)(x) = \sqcap\{u(\phi_x) | u \in [v]_2\}.$$

- A three-valued interpretation v of an ADF D is:
  - admissible iff $v \leq_i \Gamma_D(v)$
  - preferred iff v is $\leq_i$-maximal admissible

Prime Implicants

- A *literal* is an atom or its negation.
- A *term* is a consistent conjunction of literals with $\top$ denoting the empty conjunction.
- We sometimes equate a term τ with the set of literals it contains.
- For instance, $\tau \setminus \tau'$ denotes the term τ with every literal that appears in τ' removed.
- An *implicant* of a formula ϕ is a term τ such that $\tau \models \phi$.
- A *prime implicant* of ϕ is an implicant τ of ϕ such that there exists no implicant τ' of ϕ such that $\tau' \subset \tau$.

Admissibility in terms of Prime Implicants

Lemma

An interpretation v of an ADF $D = (At, C)$ is admissible if and only if, for all $x \in At$:

(1) *If $v(x) = \boldsymbol{T}$ then v satisfies a prime implicant of ϕ_x.*

(2) *If $v(x) = \boldsymbol{F}$ then v satisfies a prime implicant of $\neg\phi_x$.*

Note: satisfaction defined as in Kleene's three-valued logic.

Example

Example

Let D be an ADF with argument a. Three examples of acceptance conditions for a:

1. $\phi_a = \neg b \wedge \neg c$. Then ϕ_a has one prime implicant, namely ϕ_a itself. Accepting a therefore requires both b and c to be rejected. The prime implicants of $\neg\phi_a$ are b and c. Rejecting a therefore requires b or c to be accepted.
2. $\phi_a = b \wedge (c \vee d)$. Then the prime implicants of ϕ_a are $b \wedge c$ and $b \wedge d$. Accepting a thus requires b to be accepted as well as c or d. The prime implicants of $\neg\phi_a$ are $\neg b$ and $\neg c \wedge \neg d$. Rejecting a thus requires b to be rejected or both c and d to be rejected.
3. $\phi_a = b \vee \neg b$. Then the only prime implicant of ϕ_a is $\top$. We can therefore accept a regardless of the status of b. Rejecting a is impossible since $\neg\phi_a$ does not have a prime implicant.

Preferred Dialogues for ADFs

Definition

Let $D = (At, L, C)$ be an ADF and $x \in At$ be an argument. A *preferred dialogue for* x is a sequence $(p_1, o_1, \ldots, p_n, o_n)$ of formulas (with p_i called the i-th *proponent move* and o_i the i-th *opponent move*) such that:

1. $p_1 = x$
2. For $i \geq 1$, $o_i = \Theta_D(p_i \setminus p_1 \cup \ldots \cup p_{i-1})$, where $\Theta_D(\phi)$ denotes the result of replacing every atom in ϕ with its acceptance condition.
3. For $i > 1$, p_i is a prime implicant of o_{i-1}
4. $p_1 \wedge \ldots \wedge p_n$ is satisfiable.

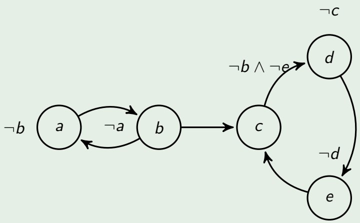
The dialogue $(p_1, o_1, \ldots, p_n, o_n)$ is *successful* if $o_n \equiv \top$.

Theorem

Let D be an ADF and x an argument. There exists a successful preferred dialogue for x iff D has a preferred interpretation v such that $v(x) = \boldsymbol{T}$.

Preferred Dialogues for ADFs

Example

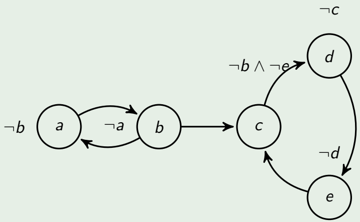


Preferred interpretations:

$$v_1 = \{a = \mathbf{T}, b = \mathbf{F}, c = \mathbf{U}, d = \mathbf{U}, e = \mathbf{U}\}$$
$$v_2 = \{a = \mathbf{F}, b = \mathbf{T}, c = \mathbf{F}, d = \mathbf{T}, e = \mathbf{F}\}$$

Preferred Dialogues for ADFs

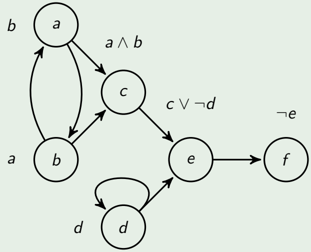
Example



p_1	o_1	p_2	o_2	p_3	o_3	p_4	o_4	p_5	o_5
d	$\neg c$	$\neg c$	$b \vee e$	b	$\neg a$	$\neg a$	b	b	$\top$

Preferred Dialogues for ADFs

Example

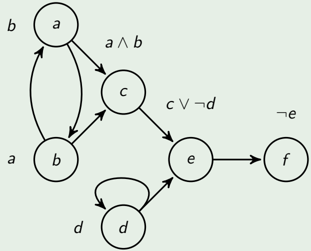


Preferred interpretations:

$$v_1 = \{a = \mathbf{T}, b = \mathbf{T}, c = \mathbf{T}, d = \mathbf{T}, e = \mathbf{T}, f = \mathbf{F}\}$$
$$v_2 = \{a = \mathbf{F}, b = \mathbf{F}, c = \mathbf{F}, d = \mathbf{T}, e = \mathbf{F}, f = \mathbf{T}\}$$

Preferred Dialogues for ADFs

Example



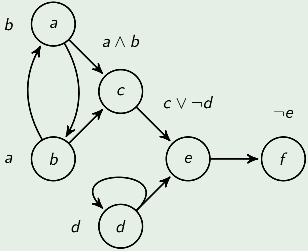
Successful dialogical proofs for f:

p_1	o_1	p_2	o_2	p_3	o_3	p_4	o_4	p_5	o_5	p_6	o_6
f	$\neg e$	$\neg e$	$\neg(c \vee \neg d)$	$\neg c \wedge d$	$\neg(a \wedge b) \wedge d$	$\neg a \wedge d$	$\neg b$	$\neg b$	$\neg a$	$\neg a$	$\top$

p_1	o_1	p_2	o_2	p_3	o_3	p_4	o_4	p_5	o_5	p_6	o_6
f	$\neg e$	$\neg e$	$\neg(c \vee \neg d)$	$\neg c \wedge d$	$\neg(a \wedge b) \wedge d$	$\neg b \wedge d$	$\neg a$	$\neg a$	$\neg b$	$\neg b$	$\top$

Preferred Dialogues for ADFs

Example



Succesful dialogical proofs for e:

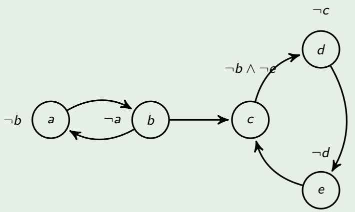
p_1	o_1	p_2	o_2	p_3	o_4
e	$c \vee \neg d$	$\neg d$	$\neg d$	$\neg d$	$\top$

p_1	o_1	p_2	o_2	p_3	o_3	p_4	o_4
e	$c \vee \neg d$	c	$a \wedge b$	$a \wedge b$	$b \wedge a$	$b \wedge a$	$\top$

Related Work

Preferred Dialogues for Abstract Argumentation

Recall the preferred game for abstract argumentation (Caminada, Dvorak, Vesic, 2016).



- PRO: I(d)
  "I have an admissible labelling in which d is labelled I."
- OPP: O(c)
  "But then in your labelling it must also be the case that d's attacker c is labelled O. Based on which grounds?"
- PRO: I(b)
  "c is labelled O because b is labelled I."
- OPP: O(a)
  "But then in your labelling it must also be the case that b's attacker a is labelled O. Based on which grounds?"
- PRO: I(b)
  "a is labelled O because b is labelled I."

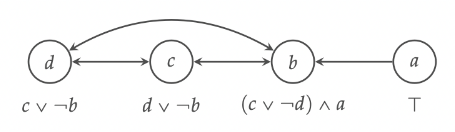
We can translate this into:

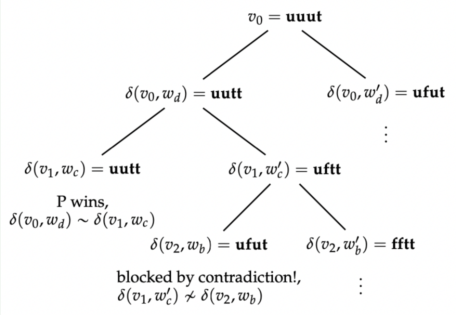
PRO	OPP	PRO	OPP	PRO
d	$\neg c$	b	$\neg a$	b

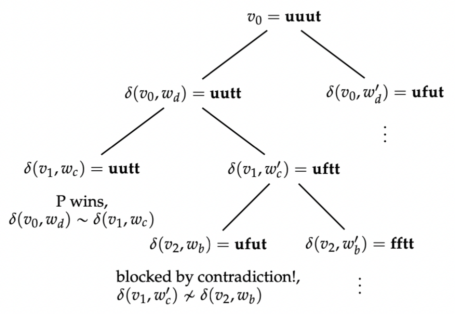
Compare with:

p_1	o_1	p_2	o_2	p_3	o_3	p_4	o_4	p_5	o_5
d	$\neg c$	$\neg c$	$b \vee e$	b	$\neg a$	$\neg a$	b	b	$\top$

Discussion Games for Preferred Semantics of Abstract Dialectical Frameworks (Zafarghandi, Verbrugge, Verheij. Ecsqaru 2019).







p_1	o_1	p_2	o_2	p_3	o_3
d	$c \vee \neg b$	c	$d \vee \neg b$	d	$\top$

p_1	o_1	p_2	o_2	p_3	o_3	p_4	o_4
d	$c \vee \neg b$	c	$d \vee \neg b$	$\neg b$	$(\neg c \wedge d) \vee \neg a$	$\neg c \wedge d$	$\dots$

p_1	o_1	p_2	o_2	p_3	o_3	p_4	o_4
d	$c \vee \neg b$	c	$d \vee \neg b$	$\neg b$	$(\neg c \wedge d) \vee \neg a$	$\neg a$	$\bot$

p_1	o_1	p_2	o_2	p_3	o_3	p_4	o_4
d	$c \vee \neg b$	$\neg b$	$(\neg c \wedge d) \vee \neg a$	$\neg c \wedge d$	$\neg d \wedge b$	$\neg d \wedge b$	$\dots$

Conclusion

Conclusion

Take-away points:

- Dialogical proof theory for credulous acceptance under preferred semantics of ADFs
- Also captures dialogical proofs for AFs
- Also applies to other special cases of ADFs.

Future work

- Other semantics
- Complexity

Discussion

- To what extent does a dialogical proof *explain* argument acceptance?