Multi-Agent Systems

Albert-Ludwigs-Universität Freiburg

Bernhard Nebel, Felix Lindner, and Thorsten Engesser

Summer Term 2017

Course outline

Outline

- Modeling agents exchanging arguments
  - Argumentation frameworks
  - Semantics
  - Algorithms

Dispute I

- A: My government cannot negotiate with your government because your government does not even recognize my government.
- B: Your government does not recognize my government either.
- A: But your government is a terrorist government.
- Which arguments should be accepted?

Dispute II

- A: Ralph goes fishing, because it is sunday.
- B: Ralph does not go fishing, because it is Mother's day, so he visits his parents.
- C: Ralph cannot visit his parents, because it is a leap year, so they are on vacation.

- Which arguments should be accepted?

⇒Content does not seem to matter but structure does!

Core idea

- A statement is believable if it can be argued successfully against attacking arguments.
- Whether or not a rational agent believes in a statement depends on whether or not the argument supporting this statement can be successfully defended against the counterarguments.

Abstract argumentation framework

We can decide what to believe while looking at arguments at the abstract level (Dung, 1995):

- Disregarding internal structures of arguments
- Focus on the attack relation between arguments $(a, b, c, d, \dots)$: a **attacks** b or $a \rightsquigarrow b$
- Not concerned with the origin of arguments or the attack relation

Abstract argumentation framework

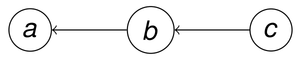
An argumentation framework is a pair $\mathcal{AF} = (Arg, \rightsquigarrow)$ where Arg is a set of arguments and $\rightsquigarrow \subseteq Arg \times Arg$. We say that $a \in Arg$ attacks $b \in Arg$ iff $(a, b) \in \rightsquigarrow$.

Example: Argumentation framework

- Remember:
  - A: Ralph goes fishing, because it is sunday.
  - B: Ralph does not go fishing, because it is Mother's day, so he visits his parents.
  - C: Ralph cannot visit his parents, because it is a leap year, so they are on vacation.

- Representation as an argumentation framework:

$$\mathcal{AF} = \langle \{a, b, c\}, \{(b, a), (c, b)\} \rangle,$$



- Argument-based semantics get as input an argumentation framework and output zero or more sets of acceptable arguments.

Argument labellings

Definition: Labelling

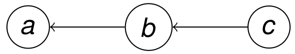
Let $\mathcal{AF} = (Arg, \rightsquigarrow)$ be an argumentation framework. A labelling of $\mathcal{AF}$ is a total function $\mathcal{L}ab : Arg \to \{in, out, undec\}$. The set of all labellings will be denoted by $\mathcal{L}(\mathcal{AF})$.

- $in(\mathcal{L}ab) = \{a \mid \mathcal{L}ab(a) = \mathbf{in}\}$
- $out(\mathcal{L}ab) = \{a \mid \mathcal{L}ab(a) = \mathbf{out}\}$
- $undec(\mathcal{L}ab) = \{a \mid \mathcal{L}ab(a) = \mathbf{undec}\}$

- To refer to a labelling $\mathcal{L}ab$ we will also write $\langle in(\mathcal{L}ab), out(\mathcal{L}ab), undec(\mathcal{L}ab)\rangle$

Application to initial example

$$\mathcal{AF} = \langle\{a,b,c\},\{(b,a),(c,b)\}\rangle,$$



$\mathcal{L}(\mathcal{AF}) = \{\langle\emptyset,\emptyset,\{a,b,c\}\rangle, \langle\emptyset,\{a\},\{b,c\}\rangle \ldots\}$

- How to identify the appropriate labellings?
- E.g., we do not want to accept both a and b, thus if $\mathcal{L}ab(a) = \mathbf{in}$ then $\mathcal{L}ab(b) \neq \mathbf{in}$.

Definition: Admissible labelling

Definition

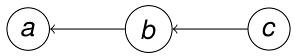
Let $\mathcal{L}ab$ be a labelling of argumentation framework $\mathcal{AF}$. An **in**-labelled argument is said to be legally in iff all its attackers are labelled **out**. An **out**-labelled argument is said to be legally out iff it has at least one attacker that is labelled **in**.

Definition

Let $\mathcal{AF}$ be an argumentation framework. An admissible labelling is a labelling where each **in**-labelled argument is legally **in** and each **out**-labelled argument is legally **out**.

Application to initial example

$$\mathcal{AF} = \langle\{a,b,c\},\{(b,a),(c,b)\}\rangle,$$



Admissible labellings

- $\langle\emptyset,\emptyset,\{a,b,c\}\rangle$
- $\langle\{a,c\},\{b\},\emptyset\rangle$

Argumentation semantics

Definition

Given an argumentation framework $\mathcal{AF} = (Arg, \rightsquigarrow)$, a labelling semantics S associates with $\mathcal{AF}$ a subset of $\mathcal{L}(\mathcal{AF})$, denoted as $\mathcal{L}_S(\mathcal{AF})$.

Complete semantics

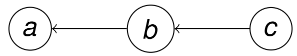
Definition

Let $\mathcal{AF} = (Arg, \rightsquigarrow)$ be an argumentation framework and $\mathcal{L}ab : Arg \rightarrow \{in, out, undec\}$ be a total function. We say that $\mathcal{L}ab$ is a complete labelling iff it satisfies the following:

$\forall a \in Arg : (\mathcal{L}ab(a) = \textbf{out} \leftrightarrow \exists b \in Arg : (b \rightsquigarrow a \wedge \mathcal{L}ab(b) = \textbf{in}))$

$\forall a \in Arg : (\mathcal{L}ab(a) = \textbf{in} \leftrightarrow \forall b \in Arg : (b \rightsquigarrow a \rightarrow \mathcal{L}ab(b) = \textbf{out}))$

$$\mathcal{AF} = \langle \{a,b,c\}, \{(b,a),(c,b)\} \rangle,$$



Complete labellings:

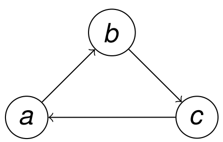
1. $\langle \{a,c\}, \{b\}, \emptyset \rangle$

Why not $\langle \emptyset, \emptyset, \{a,b,c\} \rangle$?

Bert, Ernie, and Elmo

- A: Bert says that Ernie is unreliable, therefore everything that Ernie says cannot be relied on.
- B: Ernie says that Elmo is unreliable, therefore everything that Elmo says cannot be relied on.
- C: Elmo says that Bert is unreliable, therefore everything that Bert says cannot be relied on.

$$\mathcal{AF} = \langle \{a,b,c\}, \{(a,b),(b,c),(c,a)\} \rangle,$$



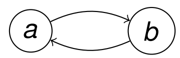
Complete labellings:

1. $\mathcal{L}ab_1 : \langle \emptyset, \emptyset, \{a,b,c\} \rangle$

Nixon Diamond

- A: Nixon is a pacifist, because he is a quaker.
- B: Nixon is not a pacifist, because he is a republican.

$$\mathcal{AF} = \langle \{a,b\}, \{(a,b),(b,a)\} \rangle,$$



Complete labellings:

1. $\mathcal{L}ab_1 : \langle \emptyset, \emptyset, \{a,b\} \rangle$
2. $\mathcal{L}ab_2 : \langle \{a\}, \{b\}, \emptyset \rangle$
3. $\mathcal{L}ab_3 : \langle \{b\}, \{a\}, \emptyset \rangle$

$\Rightarrow$Three resonable positions a rational agent can take.

Grounded semantics

Definition

Let $\mathcal{AF}$ be an argumentation framework. The grounded labelling of $\mathcal{AF}$ is a complete labelling $\mathcal{L}ab$ where $in(\mathcal{L}ab)$ is minimal w.r.t. set inclusion.

- Grounded semantics picks the complete labelling with minimal **in**, minimal **out**, and maximal **undec**.
- Intuitively, the arguments in **in** are those that must be accepted by every rational agent.
- These arguments are in the **in** set of every complete labelling.
- The grounded labelling coincedes with the intersection of all complete labellings.
- The grounded labelling is unique.

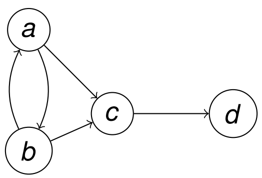
Preferred semantics

Definition

Let $\mathcal{AF}$ be an argumentation framework. The preferred labelling of $\mathcal{AF}$ is a complete labelling $\mathcal{L}ab$ where $in(\mathcal{L}ab)$ is maximal w.r.t. set inclusion.

- Preferred semantics picks the complete labelling with maximal **in**, maximal **out**, and minimal **undec**.
- For every argumentation framework at least one preferred labelling exists.

Preferred semantics: Instance



- Only ground labelling: $\langle \emptyset, \emptyset, \{a,b,c,d\} \rangle$
- Preferred labelling: $\langle \{a,d\}, \{b,c\}, \emptyset \rangle$, $\langle b,d\}, \{a,c\}, \emptyset \rangle$
- Here: Ground labelling is not among the preferred labellings and non of the preferred labellings is the ground labelling. Also, it is not the case that the ground labelling coincedes with the intersection of all preferred labellings.

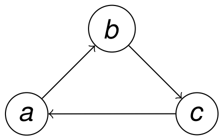
Stable semantics

Definition

Let $\mathcal{L}ab$ be a labelling of an argumentation framework $\mathcal{AF}$. $\mathcal{L}ab$ is a stable labelling of $\mathcal{AF}$ iff it is a complete labelling with **undec**$(\mathcal{L}ab) = \emptyset$.

- Stable semantics decides for every argument if it is **in** or **out**.
- As it minimizes **undec** it maximizes **in** and **out**. Thus, every stable labelling is a preferred labelling.
- But not vice versa: Whereas a preferred labelling always exists, the existence of a stable labelling is not guaranteed.

Applicability of stable semantics



Complete labellings:

1. $\mathcal{L}ab_1 : \langle \emptyset, \emptyset, \{a, b, c\} \rangle$

$\Rightarrow \mathcal{L}ab_1$ is complete, ground, preferred, but not stable.

Reasoning tasks

- Given an argument A and and argumentation framework $\mathcal{AF}$, is A in the **in** set of $\mathcal{AF}$'s ground labelling?
  - Skeptical acceptance of A: Corresponds to the question if A is **in** in all complete labellings.
- Given an argument A and and argumentation framework $\mathcal{AF}$, is A in the **in** set of some of $\mathcal{AF}$'s preferred labellings?
  - Credulous acceptance of A

Partial labelling

Definition

A partial labelling is a partial function $\mathcal{L}ab : Args \rightarrow \{\textbf{in}, \textbf{out}\}$ such that

- if $\mathcal{L}ab(A) = \textbf{in}$ then for each attacker B $\mathcal{L}ab(B) = \textbf{out}$
- if $\mathcal{L}ab(A) = \textbf{out}$ then for some attacker B $\mathcal{L}ab(B) = \textbf{in}$

- Partial labellings are admissible labellings
- A partial labelling $\mathcal{L}ab$ can be extended to a complete labelling $\mathcal{L}ab' \supseteq \mathcal{L}ab$
- For each complete labelling $\mathcal{L}ab'$ there exists a partial labelling $\mathcal{L}ab \subseteq \mathcal{L}ab'$ (just remove the **undec** labels)

Grounded labelling: Algorithm

Definition

$extendin(\mathcal{L}ab) = \mathcal{L}ab \cup \{(A, \textbf{in}) \mid \forall B [B \rightsquigarrow A \rightarrow \mathcal{L}ab(B) = \textbf{out}]\}$
$extendout(\mathcal{L}ab) = \mathcal{L}ab \cup \{(A, \textbf{out}) \mid \exists B [B \rightsquigarrow A \wedge \mathcal{L}ab(B) = \textbf{in}]\}$
$extendinout(\mathcal{L}ab) = extendin(\mathcal{L}ab) \circ extendout(\mathcal{L}ab)$

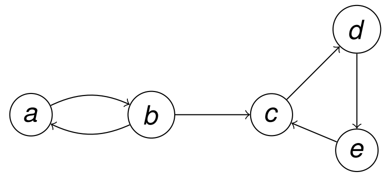
- If $\mathcal{L}ab$ is a partial labelling, then $extendin(\mathcal{L}ab)$, $extendout(\mathcal{L}ab)$, $extendinout(\mathcal{L}ab)$ are partial labellings.

```
function GROUNDLABELLING(AF)
    L ← ∅
    repeat
        L_old ← L
        L ← extendin(L)
        L ← extendout(L)
    until L = L_old
    return L ∪ {(A, undec) | (A, in) ∉ L and (A, out) ∉ L}
end function
```

From Socrates's dialog to preferred labelling

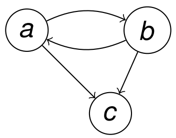
- Idea: Take the other's opinion and then derive a contradiction:
  - Proponent (M) makes a statement (A)
  - Opponent (S) derives from A more statements M will be committed to
  - S aims at letting M commit himself to a contradiction
- Dialog game
  - M starts and claims the existence of a reasonable position (admissible labelling) in which a particular argument is accepted (labelled **in**).
  - S confronts M with the consequences of M's own position, and asks M to resolve these consequences.
  - S wins if she leads M to a contradiction.
- If M wins then his argument is in the **in** set of an admissible labelling, and thus in the **in** of a preferred labelling.

Example Dialog



- M: in(D) *I have an admissible labelling in which D is in*
- S: out(C) *But then in your labelling C is out. Why?*
- M: in(B) *Because B is in*
- S: out(A) *But then A must be out. Why?*
- M: in(B) *Because B is in.*

Example Dialog



- M: in(C) *I have an admissible labelling in which C is in*
- S: out(A) *But then in your labelling A is out. Why?*
- M: in(B) *Because B is in*
- S: out(B) *But B must be out!*

Admissible discussion I

Definition

Let $\mathcal{AF} = (Arg, \rightsquigarrow)$ be an argumentation framework. An admissible discussion is a sequence of moves $[\Delta_1, \ldots, \Delta_n] (n \geq 0)$ such that:

- each move $\Delta_i (1 \leq i \leq n)$ where i is odd is called M-move and is of the form $in(A)$
- each move $\Delta_i (1 \leq i \leq n)$ where i is even is called S-move and is of the form $out(A)$
- for each S-move $\Delta_i = out(A) (2 \leq i \leq n)$ there exists an M-move $\Delta_j = in(B) (j < i)$ such that A attacks B
- for each M-move $\Delta_i = in(A) (3 \leq i \leq n)$ it holds that Δ_{i-1} is of the form $out(B)$, where A attacks B
- there exist no two S-moves $\Delta_i = \Delta_j$ with $i \neq j$

Admissible discussion II

Definition

An admissible discussion $[\Delta_1, \ldots, \Delta_n]$ is said to be finished iff

1. There exists no Δ_{n+1} such that $[\Delta 1, \ldots, \Delta_n, \Delta_{n+1}]$ is an admissible discussion, or there exists a M-move and a S-move containing the same argument
2. No subsequence of the discussion is finished.

Definition

A finished admissible discussion is won by player S if there exist a M-move and a S-move containing the same argument. Otherwise, it is won by the player making the last move.

Theorem

Theorem

Let g be an admissible discussion won by M and let $\mathcal{L}ab : Ar \to \{\mathbf{in}, \mathbf{out}, \mathbf{undec}\}$ be a function defined as follows.
For every argument $B \in Ar$:

- $\mathcal{L}ab(B) = \mathbf{in}$ if B was labeled in during g
- $\mathcal{L}ab(B) = \mathbf{out}$ if B was labeled out during g
- $\mathcal{L}ab(B) = \mathbf{undec}$ otherwise

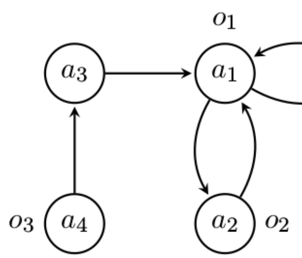
Then $\mathcal{L}ab$ is an admissible labelling.

- Thus, if there is a winning game for M defending A then A is in the **in** set of some preferred labelling (add **undec** arguments to **in** as long as possible).

Applications of argumentation frameworks

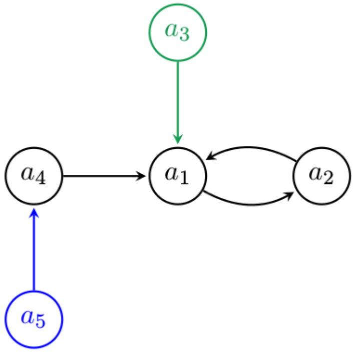
- Can be used to decide what to do next.
- Can be used to find perfect matchings [3]
  - Arg: The couples
  - $(m_1, w_1) \rightsquigarrow (m_2, w_2)$ iff
    - $m_1 = m_2$ and m_1 prefers w_2 to w_1, or
    - $w_1 = w_2$ and w_1 prefers m_1 to m_2
- Ressource allocation
  - Arg: Pairs $(agent, task)$
  - $(agent_i, task_i) \rightsquigarrow (agent_j, task_j)$ iff one of:
    - $(agent_i, task_i)$ is preferred to $(agent_j, task_j)$
    - $(agent_i, task_i)$ exclused $(agent_j, task_j)$
    - Agent is unable to do $task_i$ (then self attack of $(agent_i, task_i)$)
- Can be used to compute the set of arguments an agent should utter / keep for itself (Persuation).

Argument-Based Decisions



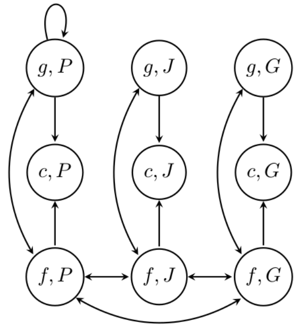
Source: [4]

Argument-Based Debates



Source: [4]

Argument-Based Allocation



Souce: [4]

Extensions of abstract argumentation systems

- In abstract argumentation systems all arguments are equally strong—relaxation
  $\rightsquigarrow$ Preference-based argumentation systems (e.g., Amgoud et al. 1998f) model preference (weights) of arguments.
- Acceptability of arguments can depend on the target audience (e.g., newspaper vs. scientific article)
  $\rightsquigarrow$ Value-based argumentation systems (Bench-Capon et. al, 2003ff)
- Arguments in abstract argumentation systems do not have an internal (logical) structure
  $\rightsquigarrow$ Deductive argumentation systems

Course outline

Literature

- M. Caminada, A gentle introduction to argumentation semantics. Technical report, University of Luxembourg, Summer 2008.
- M. Caminada, W. Dvorak, S. Vesic, Preferred semantics as socratic discussion. Journal of Logic and Computation, 2014.
- P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence 77, pp. 321-357, 1995.
- J.-G. Mailly, Dynamics of Argumentation Frameworks, PhD Thesis, 2015.