Agent-Based Systems

Michael Rovatsos
mrovatso@inf.ed.ac.uk

Lecture 13 – Argumentation in Multiagent Systems

Where are we?

Last time . . .

- Bargaining
- Alternating offers
- Negotiation decision functions
- Task-oriented domains
- Bargaining for resource allocation

Today . . .

- **Argumentation in Multiagent Systems**

Argumentation

- Agents may have mutually contradicting beliefs
  - I believe p; you believe $\neg p$
  - I believe p, $p \to q$; you believe $\neg q$
- How can agents reach agreements about **what to believe**?
- **Argumentation** provides principled techniques for deciding what to believe in the face of inconsistencies
- We achieve this by comparing arguments that can be compiled from the agents' beliefs
- **Arguments** usually present beliefs and describe reasonable justifications

Different modes of argument

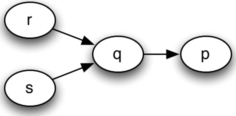
- At least four different modes of arguments can be identified between humans:

1. Logical mode (deductive, proof-like, concerned with making correct inferences)
2. Emotional mode (appeals to feelings, attitudes, etc.)
3. Visceral mode (physical, social aspects)
4. Kisceral mode (appeals to the intuitive, mystical or religious)

- Different types are used in different situations (e.g. logical mode (hopefully) in courts of law)

Abstract Argumentation

- We can decide what to believe while looking at arguments at the abstract level (Dung, 1995):
  - Disregarding their internal structures, e.g. arguments a, b, c, d
  - Focus on the **attack** relation, e.g. a attacks b or $a \rightarrow b$
  - Not concerned with the origin of arguments or the attack relation
- An **abstract argumentation system** $\mathcal{A} = \langle X, \rightarrow \rangle$ is defined by
  - a set of arguments X (just a collection of objects),
  - $\rightarrow \subseteq X \times X$ a binary **attack** relation on arguments
- Example: $\langle \{p, q, r, s\}, \{(r, q), (s, q), (q, p)\} \rangle$



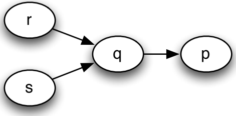
Arguments: p, q, r, s
Attacks: $r \rightarrow q$, $s \rightarrow q$, $q \rightarrow p$

- Which arguments can we consider to be rationally justified?

  There is no universal definition for acceptability

Terminology

- Lets consider some meaningful properties for rationally justified sets of arguments
- A set of arguments S is **conflict-free** if if there are no arguments a, b in S such that a attacks b, e.g.



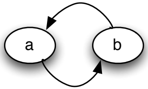
$$\emptyset, \{p\}, \{q\}, \{r\}, \{s\}, \{r, s\}, \{p, r\}, \{p, s\}, \{p, r, s\}$$

- An argument a is **acceptable** with respect to a set S of arguments iff for each argument a': if a' attacks a then a' is attacked by some argument in S
- A conflict-free set of arguments S is **admissible** iff each argument in S is acceptable w.r.t. S

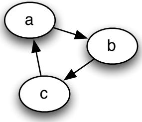
  e.g. $\emptyset, \{r\}, \{s\}, \{r, s\}, \{p, r\}, \{p, s\}, \{p, r, s\}$

Preferred Extensions

- **Preferred extensions** are maximal (w.r.t. set inclusion) admissible sets, e.g. $\{p, r, s\}$ is a preferred extension, but not $\emptyset$ or $\{p\}$
- Preferred extensions help determine which arguments should be accepted but are not always useful:



Preferred extensions are not necessarily unique e.g. $\{a\}$ and $\{b\}$ here



The only preferred extension may be the empty set

- An argument is **sceptically accepted** if it is a member of every preferred extension
- An argument is **credulously accepted** if it is a member of at least one preferred extension

Grounded Extensions (I)

- An alternative notion of acceptability is provided by the notion of **grounded** extension
- The (unique) grounded extension can be built incrementally:
  1. Arguments that are not attacked are "in"
  2. Delete from the graph every argument that is attacked by an argument that is in the grounded extension and go to Step 1
  - Iterate until there are no more changes to the argument graph
- The grounded extension
  - always exists and
  - is guaranteed to be unique, but
  - may be empty (if no arguments are free of attackers initially)

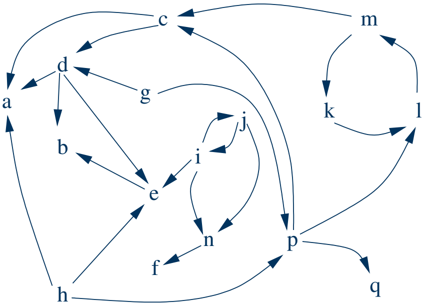
Grounded Extensions (II)

- The *characteristic function* of an argumentation system $\mathcal{A} = \langle X, \rightarrow \rangle$, is the function $\mathcal{F} : 2^X \rightarrow 2^X$, which is defined as follows:

$$\mathcal{F}(S) = \{a \mid a \textit{ is acceptable } \text{w.r.t. } S\}$$

- The grounded extension of an argumentation system is the least fixed point of the characteristic function $\mathcal{F}$
- Consider the sequence:
  - $\mathcal{F}^0 = \emptyset$,
  - $\mathcal{F}^{i+1} = \{a \in X \mid a \text{ is acceptable w.r.t. } \mathcal{F}^i\}$
  - $\cdots$ (until no arguments are added to the set)

Example



- Argument *h* has no attackers ➡ "in"
- Because of this, *a* is not acceptable ➡ "out"
- For same reason *p* is out
- *p* only attacker of *q*, thus *q* is ➡ "in"
- . . .

Deductive Argumentation Systems

- "Purest", most rational kind of argument: in classical logic, argument = sequence of inferences leading to a conclusion
- Write $\Gamma \vdash \varphi$ to denote that sequence of inference steps from premises Γ will allow us to establish proposition φ, where Γ is part of our overall knowledge base Δ

  Example: $\Gamma \vdash \mathit{mortal}(\mathit{Socrates})$ where

  $\Gamma = \{\mathit{human}(\mathit{Socrates}), \mathit{human}(X) \Rightarrow \mathit{mortal}(X)\}$
- A **deductive argument** is a pair $\langle \Gamma, \varphi \rangle$ with support Γ and conclusion φ where:
  i. $\Gamma \subset \Delta$, $\Gamma \vdash \varphi$
  ii. Γ is logically consistent
  iii. Γ is minimal (i.e. none of its subsets satisfies the above)
- Two important classes of arguments:
  - **Tautological arguments**: $\langle \Gamma, \varphi \rangle$ where $\Gamma = \emptyset$
  - **Non-trivial arguments**: $\langle \Gamma, \varphi \rangle$ where Γ is consistent

Example: Arguments

$$human(X) \Rightarrow mortal(X)$$
$$human(Hercules)$$
$$father(Heracles, Zeus)$$
$$father(Apollo, Zeus)$$
$$divine(X) \Rightarrow \neg mortal(X)$$
$$father(X, Zeus) \Rightarrow divine(X)$$
$$\neg(father(X, Zeus) \Rightarrow divine(X))$$

Examples of arguments:

$$Arg_1 = \langle \{human(Heracles), human(X) \Rightarrow mortal(X)\}, mortal(Heracles) \rangle$$
$$Arg_2 = \langle \{father(Heracles, Zeus), father(X, Zeus) \Rightarrow divine(X), divine(X) \Rightarrow \neg mortal(X)\}, \neg mortal(Heracles) \rangle$$
$$Arg_3 = \langle \{\neg(father(X, Zeus) \Rightarrow divine(X))\}, \neg(father(X, Zeus) \Rightarrow divine(X)) \rangle$$

The Attack Relation

The attack relation is defined as follows

- For any propositions φ and ψ, φ attacks ψ iff $\varphi \equiv \neg\psi$
- $\langle \Gamma_1, \varphi_1 \rangle$ **rebuts** $\langle \Gamma_2, \varphi_2 \rangle$ if φ_1 attacks φ_2
- $\langle \Gamma_1, \varphi_1 \rangle$ **undercuts** $\langle \Gamma_2, \varphi_2 \rangle$ if φ_1 attacks some $\psi \in \Gamma_2$
- $\langle \Gamma_1, \varphi_1 \rangle$ **attacks** $\langle \Gamma_2, \varphi_2 \rangle$ if it undercuts or rebuts it

Example:

$$
\begin{aligned}
Arg_1 &= \langle \{ human(Heracles), human(X) \Rightarrow mortal(X) \}, mortal(Heracles) \rangle \\
Arg_2 &= \langle \{ father(Heracles, Zeus), father(X, Zeus) \Rightarrow divine(X), \\
&\qquad divine(X) \Rightarrow \neg mortal(X) \}, \neg mortal(Heracles) \rangle \\
Arg_3 &= \langle \{ \neg(father(X, Zeus) \Rightarrow divine(X)) \}, \neg(father(X, Zeus) \Rightarrow divine(X)) \rangle
\end{aligned}
$$

- Arguments Arg_1 and Arg_2 are mutually rebutting
- Argument Arg_3 undercuts argument Arg_2

Argument Classes

We can identify five classes of argument type in order of increasing acceptability

- A1: The class of all arguments that can be constructed
- A2: The class of all non-trivial arguments that can be constructed
- A3: The class of all arguments that can be constructed with no rebutting arguments
- A4: The class of all arguments that can be constructed with no undercutting arguments
- A5: The class of all tautological arguments that can be constructed

Example: Argument Classes

$$
\begin{aligned}
Arg_1 =& \langle \{ human(Heracles), human(X) \Rightarrow mortal(X) \}, mortal(Heracles) \rangle \\
Arg_2 =& \langle \{ father(Heracles, Zeus), father(X, Zeus) \Rightarrow divine(X), \\
& \quad divine(X) \Rightarrow \neg mortal(X) \}, \neg mortal(Heracles) \rangle \\
Arg_3 =& \langle \{ \neg(father(X, Zeus) \Rightarrow divine(X)) \}, \neg(father(X, Zeus) \Rightarrow divine(X)) \rangle
\end{aligned}
$$

- Arg_1 and Arg_2 are mutually rebutting and thus in A2
- $\langle \emptyset, divine(Heracles) \vee \neg divine(Heracles) \rangle$ is in A5
- $\langle \{ father(apollo, Zeus), father(X, Zeus) \Rightarrow divine(X), divine(X) \Rightarrow \neg mortal(X) \}, \neg mortal(apollo) \rangle$ is in A4

Argumentation dialogue systems

- Agents engage in **dialogue** to convince other agents of some state of affairs
- Consider two agents 0 and 1 engaging in the following dialogue:
  - Agent 0 attempts to convince 1 of some argument
  - Agent 1 attempts to rebut or undercut it
  - Agent 0 in turn attempts to defeat 1's argument
  - And so on ...
- **Moves** $\langle Player, Arg \rangle$ are steps in such a dialogue, $Player \in \{0, 1\}$, $Arg \in \mathcal{A}(\Delta)$ (the set of all arguments constructed from Δ)
- A sequence $\langle m_0, \ldots m_k \rangle$ is a **dialogue history** if
  - $Player_{2i} = 0$, $Player_{2i+1} = 1$ for all $i \geq 0$
  - If $Player_i = Player_j$ and $i \neq j$, then $Arg_i \neq Arg_j$,
  - Arg_{i+1} defeats Arg_i for all $i \geq 0$
- A dialogue **ends** if no further moves are possible, the **winner** is $Player_k$

Types of dialogue

Typology due to Walton and Krabbe (1995):

Type	Initial situation	Main goal	Participants' aim
Persuasion	conflict of opinion	resolve the issue	persuade other
Negotiation	conflict of interest	make a deal	get best deal
Inquiry	general ignorance	growth of knowledge	find a proof
Deliberation	need for action	reach a decision	influence outcome
Information seeking	personal ignorance	spread knowledge	gain or pass on knowledge
Eristics	conflict/antagonism	reaching an accommodation	strike other party

Summary

- Argumentation
- Abstract argumentation systems
- Deductive argumentation systems
- Argumentation-based dialogue
- Next time: **Logics for Multiagent Systems**