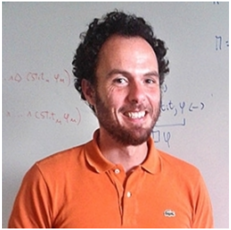
Empowering Data-driven AI by Argumentation


Leila Amgoud (CNRS-IRIT)


Emiliano Lorini (CNRS-IRIT)


Philippe Muller (UPS-IRIT)

Context

- Argumentation aims at increasing acceptability of claims by supporting them with arguments
- Argument is a set of *premises* intended to establish a *claim*

Premise 1 ⋮ Premise n	Generally, birds fly Tweety is a bird
Claim	Therefore, Tweety flies

Several Types of Claims

- Categorical arguments (X is a Y)
- Definitional arguments (X is a Y; the definition of Y is contested)
- Cause/Consequence arguments (X causes Y; Y is a consequence of X)
- Resemblance arguments (X is like Y)
- Evaluation arguments (X is good or bad; X is true or false)
- Proposal arguments (One should do X)
- . . .

$\Longrightarrow$ Several Arguments Schemes

Argumentation in AI

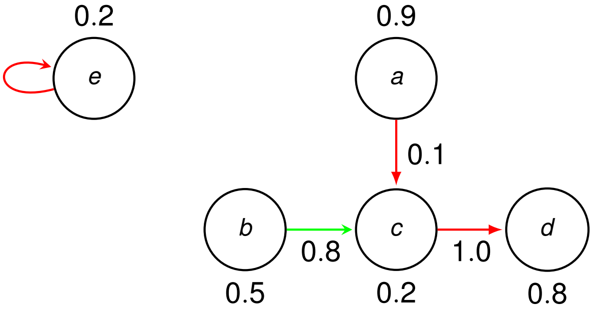
- **Practical applications**
  - Diagnosis in medical domain
  - Online dispute resolution (e.g., CyberSettle)
  - Online debate (e.g., DebateGraph, debate.org)
  - Committees
  - ...

- **Theoretical applications**
  - Reasoning with (inconsistent, defeasible) information
  - Decision making
  - Negotiation
  - Classification
  - ...

Argumentation Process

Given a **problem** (making a decision, classifying an object, ...)

- Construct **arguments**
- Identify their **basic strengths** + their **interactions**



Example of Arguments

Let Σ be a finite propositional knowledge base.

Definition

An argument is a pair $\langle \Psi, \psi \rangle$ such that

- $\Psi \subseteq \Sigma$
- Ψ is consistent
- $\Psi \vdash \psi$
- $\nexists \Psi'$ s.t. $\Psi' \subset \Psi$ and $\Psi' \vdash \psi$

$\Sigma = \{p \wedge q, \neg p \wedge t\}$

- $A = \langle \{p \wedge q\}, p \vee \neg t \rangle$
- $B = \langle \{p \wedge q\}, q \rangle$
- $C = \langle \{\neg p \wedge t\}, t \vee p \rangle$
- $\ldots$

Example of Attacks

Definition

(Ψ, ψ) attacks (Ψ', ψ') iff $\exists \phi \in \Psi'$ such that $\psi \vdash \neg \phi$.

$\Sigma = \{p \wedge q, \neg p \wedge t\}$

$$A = \langle \{p \wedge q\}, p \vee \neg t \rangle \textit{ attacks } C = \langle \{\neg p \wedge t\}, t \vee p \rangle$$

Argumentation Process

Given a **problem** (making a decision, classifying an object, . . .)

- Construct **arguments**
- Identify their **basic strengths** + their **interactions** ⇒ Graph
- **Analyse** the arguments ⇒ Semantics
- Conclude (the chosen option, the class of the object, . . .)

Semantics

Individual arguments

- Strength concerns the **quality** of argument's components (premises, link, conclusion)

  Characteristics: Uniqueness, Precise vs Vague

- Acceptability states whether an argument can be **accepted** so that its claim can safely be used for drawing conclusions, . . .

  Characteristics: Uniqueness, Binary (Accepted, Rejected)

Collections of arguments

- Coalitions Prevailing **viewpoints** expressed in an arg. graph

  Characteristics: Multiple sets

Three Families of Semantics

- A semantics is a function π that assigns to every $\mathbf{G} = \langle \mathcal{A}, w, \mathcal{R}, \pi \rangle$,
  - a *set* $\mathtt{Ext}^{\pi}_{\mathbf{G}} \in 2^{2^{\mathcal{A}}}$ *(Extension Semantics)*
  - a *weighting* $\mathtt{Deg}^{\pi}_{\mathbf{G}} : \mathcal{A} \to \mathcal{D}$ *(Weighting Semantics)*
  - a *preorder* $\succeq^{\pi}_{\mathbf{G}} \subseteq \mathcal{A} \times \mathcal{A}$ *(Ranking Semantics)*

  $\mathcal{D}$ is a totally ordered scale.

Weighting Semantics Ranking Semantics	Extension Semantics
Strength	Acceptability
	Coalitions

Weighting Semantics

Weighting semantics

Let $\mathbf{G} = \langle \mathcal{A}, w, \mathcal{R}, \pi \rangle$, $a \in \mathcal{A}$, $b_1, \ldots, b_n$ *its attackers.*

$$\texttt{Deg}(a) = f(w(a), g(h(\pi((b_1, a)), \texttt{Deg}(b_1)), \ldots, h(\pi((b_n, a)), \texttt{Deg}(b_n))))$$

- $h : [0,1] \times [0,1] \to [0,1]$
- $g : \bigcup_{n=0}^{+\infty} [0,1]^n \to [0, +\infty)$ *such that g is symmetric*
- $f : [0,1] \times \texttt{Range}(g) \to [0,1]$

Examples of Functions

$f_{comp}(x_1, x_2) = x_1(1 - x_2)$	$g_{sum}(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i$	$h_{prod}(x_1, x_2) = x_1 x_2$
$f_{exp}(x_1, x_2) = x_1 e^{-x_2}$	$g_{sum,\alpha}(x_1, \ldots, x_n) = (\sum_{i=1}^{n} (x_i)^{\alpha})^{\frac{1}{\alpha}}$	$h_{prod,\alpha}(x_1, x_2) = x_1^{\alpha} x_2, \ \alpha > 0$
$f_{frac}(x_1, x_2) = \frac{x_1}{1+x_2}$	$g_{max}(x_1, \ldots, x_n) = \max\{x_1, \ldots, x_n\}$	$h_{min}(x_1, x_2) = \min\{x_1, x_2\}$
$f_{min}(x_1, x_2) = \min\{x_1, 1 - x_2\}$	$g_{psum}(x_1, \ldots, x_n) = x_1 \oplus \cdots \oplus x_n$, where $x_1 \oplus x_2 = x_1 + x_2 - x_1 x_2$	$h_{Ham}(x_1, x_2) = \frac{x_1 x_2}{x_1 + x_2 - x_1 x_2}$; $h_{Ham}(x_1, x_2) = 0$ if $x_1 = x_2 = 0$

The choice of functions depends on axioms that need to be satisfied by a semantics

Example of a Weighting Semantics

h-Categorizer

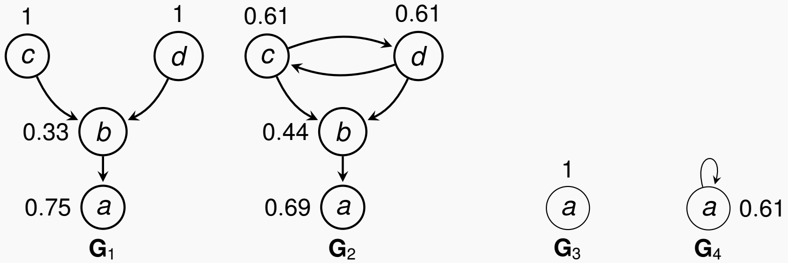
Let $\mathbf{G} = \langle \mathcal{A}, \mathcal{R} \rangle$ *and* $a \in \mathcal{A}$.

$$\texttt{Deg}^{\pi}_{\mathbf{G}}(a) = \frac{1}{1 + \sum\limits_{(b,a) \in \mathcal{R}} \texttt{Deg}^{\pi}_{\mathbf{G}}(b)}$$

$$\begin{cases} g_{sum}(x_1, \ldots, x_n) = \sum\limits_{i=1}^{n} x_i \\ \\ f_{frac}(x) = \frac{1}{1+x_2} \\ \\ \mathcal{D} = [0, 1] \end{cases}$$

Example of a Weighting Semantics

Examples



Objectives

- Argumentative counterparts of data-driven models
- Explanation theory and persuasion

Argumentative Counterparts of Data-driven Models

- Define argumentative view of existing data-driven models, namely NNs
- Improve predictions by incorporating arguments given by experts
- Reduce the need for large amounts of data. The new arguments introduce crucial domain knowledge,
- Improve search performance. The new arguments will constrain the combinatorial search among possible hypotheses

Explanation Theory and Persuasion

- Links between explanation and argument
- Explanation schemes
- Evaluation of explanations
- Persuasive explanation
- Which explanation to present to users and under which format ?

Planned PhD / post doc proposals

- 2 PhD thesis (one for each part of the project)
- 1 (or 2) postdoc on the first part

Interactions with other chairs

- Joao Marques-Silva
- Louise Travé-Massouyès
- Jean-Michel Loubes