Typicality, Conditionals and a Probabilistic Semantics for Gradual Argumentation

Mario Alviano[1] Laura Giordano[2] Daniele Theseider Dupré[2]

1 Università della Calabria, Italy

2 Università del Piemonte Orientale Italy

CILC 2024, Roma, June 25-28

Aims of the talk

Argumentation is one of the major formalisms used for explainability.

- We propose a general approach to define a *many-valued preferential interpretation* of *gradual argumentation semantics*.
- *Conditional reasoning over arguments* and boolean combination of arguments through the *verification of graded (strict or defeasible) implications* over a preferential interpretation.
- We also discuss a *probabilistic interpretation for gradual argumentation*, which builds on the many-valued preferential semantics.

The approach

Given an *argumentation graph G* and a *gradual semantics S*, satisfying *weak conditions on the domain* of argument interpretation, we consider:

- a *many-valued propositional logic with typicality*, where arguments play the role of propositional variables (inspired to PTL and DLs with typicality)
- *graded conditionals* of the form $\mathbf{T}(\alpha) \rightarrow \beta \geq l$, meaning that "normally argument α implies argument β with degree at least l" (with α and β *boolean combination of arguments*):

$$\mathbf{T}(granted\_loan) \rightarrow high\_salary \wedge young \geq 0.7$$

- Build a multi-preferential interpretation of a graph G under a semantic S, I_G^S
- Verification of conditional properties over I_G^S by *model checking*

Domain of argument interpretation and argumentation graphs: some assumptions

- We let the *domain of argument interpretation* be a set $\mathcal{D}$, equipped with a *preorder relation* $\leq$ [Baroni et al. 2019]
- Let a *(weighted) argumentation graph* be a tuple:

$$G = \langle \mathcal{A}, \mathcal{R}, \sigma_0, \pi \rangle$$

  - $\mathcal{A}$ is a set of *arguments*,
  - $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ a set of *edges*,
  - $\sigma_0 : \mathcal{A} \rightarrow \mathcal{D}$ assigns a *base score* of arguments,
  - $\pi : \mathcal{R} \rightarrow \mathbb{R}$ is a *weight function* assigning a positive or negative weight to edges.

  A pair $(B, A) \in \mathcal{R}$ is regarded as a *support* of argument B to argument A when the weight $\pi(B, A)$ is positive and as an *attack* of argument B to A when $\pi(B, A)$ is negative.

Labellings and gradual semantics

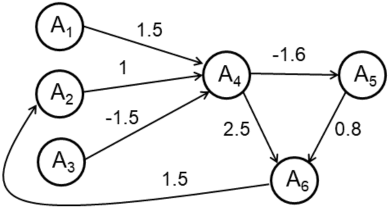
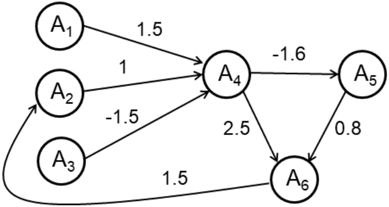


Figure: Example weighted argumentation graph *G* where the base score is not represented

- A *labelling* σ *of G over* $\mathcal{D}$ is a function $\sigma : \mathcal{A} \to \mathcal{D}$, which assigns to each argument an *acceptability degree (or a strength) in* $\mathcal{D}$.
- A *gradual semantics S* for an argumentation graph *G* identifies *a set* Σ^S *of labellings* of the graph *G* over a domain of argument valuation $\mathcal{D}$.

Example



- φ-coherent semantics [NMR 2022];
- $\mathcal{D}$ equal to $\mathcal{C}_n = \{0, \frac{1}{n}, \ldots, \frac{n-1}{n}, 1\}$.
- With $n = 5$, the graph G has *36 φ-coherent labellings*, while, for $n = 9$, G has *100 φ-coherent labellings*.
- For instance, $\sigma = (0, 4/5, 3/5, 2/5, 2/5, 3/5)$ (meaning that $\sigma(A_1) = 0$, $\sigma(A_2) = 4/5$, and so on) is a labelling for $n = 5$.

A many valued logic (of arguments)

- Given an argumentation graph $G = \langle \mathcal{A}, \mathcal{R}, \sigma_0, \pi \rangle$, we introduce a *propositional language* $\mathcal{L}$, whose set of propositional variables *Prop is the set of arguments* $\mathcal{A}$.
- Language $\mathcal{L}$ contains the boolean connectives $\wedge$, $\vee$, $\neg$ and $\rightarrow$, and that formulas are defined inductively, as usual.
- $\mathcal{D}$ is the *truth degree set*.
- We let $\otimes$, $\oplus$, $\rhd$ and $\ominus$ be the *truth degree functions* in $\mathcal{D}$ for the connectives $\wedge$, $\vee$, $\neg$ and $\rightarrow$ (respectively).
  - E.g., when $\mathcal{D}$ is $[0, 1]$ or $\mathcal{C}_n$, $\otimes$, $\oplus$, $\rhd$ and $\ominus$ can be a t-norm, s-norm, implication function, and negation function in some system of many-valued logic.

Labellings as many-valued valuations

- A labelling $\sigma : \mathcal{A} \to \mathcal{D}$ of graph G, assigning to each argument $A_i \in \mathcal{A}$ a truth degree in $\mathcal{D}$, as a *many-valued valuation*.
- σ is extended to all propositional formulas of $\mathcal{L}$:
  $\sigma(\alpha \wedge \beta) = \sigma(\alpha) \otimes \sigma(\beta)$ $\quad$ $\sigma(\alpha \vee \beta) = \sigma(\alpha) \oplus \sigma(\beta)$
  $\sigma(\alpha \to \beta) = \sigma(\alpha) \rhd \sigma(\beta)$ $\quad$ $\sigma(\neg \alpha) = \ominus \sigma(\alpha)$
- A labelling σ uniquely assigns a truth degree to any *boolean combination of arguments*.
- We assume that the false argument $\bot$ and the true argument $\top$ are formulas of L and that $\sigma(\bot) = 0_{\mathcal{D}}$ and $\sigma(\top) = 1_{\mathcal{D}}$, for all labellings σ.

Preferences over labellings in Σ

- Given a set of labellings Σ, we define a *preference relation* $<_{A_i}$ on Σ, for each argument $A_i \in \mathcal{A}$:

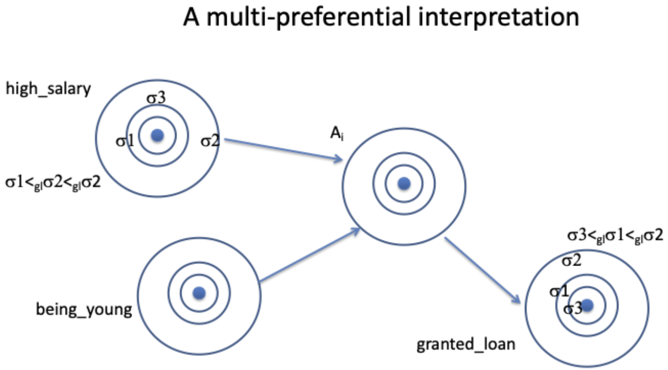
$$\sigma <^{\Sigma}_{A_i} \sigma' \text{ iff } \sigma'(A_i) < \sigma(A_i), \text{ for } \sigma, \sigma' \in \Sigma$$

  σ *is* more plausible *than* σ' as a situation for argument A_i to holds.
- The preference relation $<^{\Sigma}_{A_i}$ is a *strict partial order relation* on Σ. We write $<_{A_i}$. We restrict to sets of labellings such that $<_{A_i}$ and $<_{\neg A_i}$ are *well-founded*.
- Similarly, *for boolean combinations of arguments* α:

$$\sigma <_{\alpha} \sigma' \text{ iff } \sigma'(\alpha) < \sigma(\alpha).$$

- For example, $\sigma = (1, 4/5, 0, 1, 1/5, 1)$ is preferred to all other labellings with respect to $<_{A_6}$, being the only one with $\sigma(A_6) = 1$.

Preferences with respect to arguments



A many-valued logic with typicality

- Given an argumentation graph G, a gradual semantics S with domain of argument valuation $\mathcal{D}$, and the *set of labellings Σ^S of G* wrt S, we let the *preferential interpretation of G wrt S* to be the pair $I_G^S = (\mathcal{D}, \Sigma^S, \{<_\alpha\})$.
- Language $\mathcal{L}^{\mathbf{T}}$ is obtained by extending $\mathcal{L}$ with a unary *typicality operator* **T**. Intuitively, "a sentence of the form $\mathbf{T}(\alpha)$ is understood to refer to the *typical situations in which* α *holds*" [Booth et al., 2019]
- The typicality operator allows the formulation of *conditional implications* (or *defeasible implications*) of the form $\mathbf{T}(\alpha) \rightarrow \beta$, "normally, if α then β"
- As in PTL also *general implications* $\alpha \rightarrow \beta$, where α and β may contain **T**

A many-valued logic with typicality

- Given a preferential interpretation $I = (\mathcal{D}, \Sigma)$, and a labelling $\sigma \in \Sigma$, the valuation of a propositional formula $\mathbf{T}(\alpha)$ in σ is defined as follows:

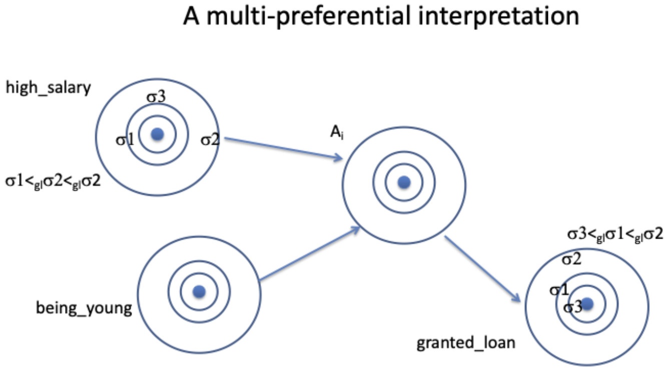
$$\sigma(\mathbf{T}(\alpha)) = \begin{cases} \sigma(\alpha) & \text{if there is no } \sigma' \text{ such that } \sigma' <_\alpha \sigma \\ 0_\mathcal{D} & \text{otherwise} \end{cases} \tag{1}$$

- When $\sigma(\mathbf{T}(A)) > 0_\mathcal{D}$, σ is *a labelling maximizing the acceptability of argument A*, among all the labellings in I.

Example

Under Gödel logic with standard involutive negation with $n = 5$, the boolean combination of arguments $A_1 \wedge A_2 \wedge \neg A_3$ has 4 maximally preferred labellings, with $\sigma(A_1 \wedge A_2 \wedge \neg A_3) = 4/5$. For such labellings, $\sigma(\mathbf{T}(A_1 \wedge A_2 \wedge \neg A_3)) = 4/5$, while equal to 0 for all other labellings.

Labellings and gradual semantics



We may check, for instance:

$$\mathbf{T}(\mathit{granted\_loan}) \rightarrow \mathit{high\_salary} \wedge \mathit{being\_young} \geq 0.7$$

Graded implications

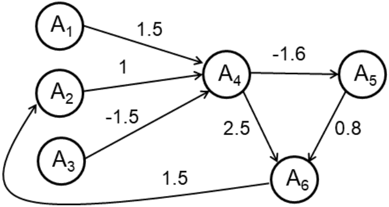
- Given a preferential interpretation $I = (\mathcal{D}, \Sigma)$, we can now define the satisfiability in I of a *graded implication*, having form $\alpha \rightarrow \beta \geq l$ or $\alpha \rightarrow \beta \leq u$, with l and u in $\mathcal{D}$ and α and β boolean combination of arguments.
- the *truth degree of an implication* $\alpha \rightarrow \beta$ *wrt.* I is defined as:

$$(\alpha \rightarrow \beta)^I = inf_{\sigma \in \Sigma}(\sigma(\alpha) \rhd \sigma(\beta)).$$

- *I satisfies a graded implication* $\alpha \rightarrow \beta \geq t$ (written $I \models \alpha \rightarrow \beta \geq t$) iff $(\alpha \rightarrow \beta)^I \geq t$;

  I satisfies a graded implication $\alpha \rightarrow \beta \leq u$ (written $I \models \alpha \rightarrow \beta \leq u$) iff $(\alpha \rightarrow \beta)^I \leq u$.

Graded implications:example



- The following graded conditionals are among the ones satisfied in the preferential interpretation $I = (\mathcal{C}_5, \Sigma, \{<_\alpha\})$, under the φ-coherent semantics:
  $\mathbf{T}(A_1 \wedge A_2 \wedge \neg A_3) \rightarrow A_6 \geq 1$
  (with 4 preferred labellings);
  $\mathbf{T}(A_1 \wedge A_2) \rightarrow A_6 \geq 4/5$ (12 preferred labellings);
  $\mathbf{T}(A_6) \rightarrow A_1 \wedge A_2 \geq 4/5$ (1 preferred labelling).

Properties

Given an interpretation $I^S = (\mathcal{S}, \Sigma^S)$, associated with an argumentation semantics S of a graph G:

- Under the choice of combination functions as in Gödel logic, interpretation $I^S = (\mathcal{S}, \Sigma^S)$ satisfies the *KLM postulates* of a preferential consequence relation, suitably reformulated:
  $\alpha \mid\!\sim \beta$ *is interpreted as* $\mathbf{T}(\alpha) \rightarrow \beta \geq 1$
  $\models A \rightarrow B$ *is interpreted as* $\alpha \rightarrow \beta \geq 1$
- For a *finite* interpretation $I^S = (\mathcal{S}, \Sigma^S)$, *satisfiability* of a graded conditional $\mathbf{T}(\alpha) \rightarrow \beta \geq k$ in I^S can be decided in *polynomial time* in the product of the size of the interpretation and the size of the formula.

Towards a probabilistic semantics of gradual argumentation

- The fuzzy interpretation of arguments also suggests a probabilistic semantics of gradual argumentation, based on Zadeh's *probability of fuzzy events* [Zadeh1968].
- An approach previously considered for SOMs [JLC2022].
- Consider the set Σ^S of labellings of G in a gradual semantics S, with *domain of argument valuation in* $[0,1]$, and a suitable (continuous) t-norm [Montes et al, 2013].
- Assuming a discrete probability distribution $p : \Sigma^S \rightarrow [0,1]$ over a set Σ^S one can define the *probability of a boolean combination of arguments* α as:

$$P(\alpha) = \sum_{\sigma \in \Sigma^S} \sigma(\alpha)\, p(\sigma)$$

When the labellings are two-valued ($\sigma(\alpha)$ is 0 or 1), this definition relates to the probability of a boolean term α by Hunter and Thimm [2020].

Towards a probabilistic semantics of gradual argumentation

- We let the conditional probability of A given B, where A and B are boolean combinations of arguments, to be

$$P(A \mid B) = P(A \wedge B)/P(B)$$

(provided $P(B) > 0$).

- As observed by Dubois and Prade [1993], this generalizes both conditional probability and the fuzzy inclusion index advocated by Kosko [1992].
- Let us extend the language $L^{\mathbf{T}}$ by introducing a new proposition $\{\sigma\}$, for each $\sigma \in \Sigma$ (with $\sigma(\{\sigma\}) = 1$ and $\sigma'(\{\sigma\}) = 0$, for any $\sigma' \neq \sigma$). Then

$$P(A|\{\sigma\}) = \sigma(A)$$

which can be regarded as a *subjective probability* (i.e., the degree of belief we put into A when we are in a state represented by labelling σ).

Towards a probabilistic semantics of gradual argumentation

- The notion of probability P defined satisfies Kolmogorov's axioms for any P_Z-compatible t-norm, with associated t-conorm, and the negation function $\ominus x = 1 - x$ [Montes2013].
- But, there are properties of classical probability which do not hold (depending on the choice of t-norm), as a consequence of the fact that not all classical logic equivalences hold in a fuzzy logic.

Conclusions and Related work

- We have proposed an approach for defeasible reasoning over argumentation graphs.
- As a case of study, for the φ-coherent semantics in the finite valued case, the approach has been implemented through an *ASP encoding* [ASPOCP 2023]
- FW: the use of this formalism for explanation

Related work

- Weydert [2013] has proposed one of the first approaches for combining abstract argumentation with a conditional semantics. He has proposed the *JZ-evaluation semantics*.
- The correspondence between *Abstract Dialectical Frameworks* (Brewka2013) and Conditional Logics have been studied by Heyninck, Kern-Isberner and Thimm [FLAIRS2020].
- In the work by Skiba and Thimm [2022] *Ordinal Conditional Functions (OCFs)* are interpreted and formalized for Abstract Argumentation, by developing a framework that allows to rank sets of arguments wrt. their plausibility. They propose an OCF inspired by System Z ranking function.
- Thimm's probabilistic semantics for AF [ECAI 2012]
- Epistemic graphs [Hunter, Pollberg, Thimm 2021] allow epistemic constraints involving statements about probabilities (we have not considered them so far).

Thank you!!!!!