ESSENCE 2014: Argumentation-Based Models of Agent Reasoning and Communication

Sanjay Modgil

Department of Informatics, King's College London

Outline

Logic, Argumentation and Reasoning

- Dung's Theory of Argumentation
- The Added Value of Argumentation
- Rationality Postulates for Logic-based Argumentation

Argumentation Based Dialogue

- Argument Game Proof Theories
- Generalisation to Dialogue
- Applications

Dung's Abstract Argumentation Theory and non-monotonic reasoning

Dung's Abstract Argumentation Theory *

 A Dung argumentation framework AF is a directed graph (Args,Att)
 Where the nodes Args denote arguments and Att is a conflict based binary attack relation between arguments

$\Box \quad \text{Given a logic } \mathcal{L} \text{ define :}$

- 1) What constitutes an argument
- 2) What constitutes an attack between two arguments
- 3) Given a set of wff Δ in \mathcal{L} construct all the arguments and relate them by the attacks in an AF (i.e., *instantiate* an AF)

* P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and *n*-person games. *Artificial Intelligence*, 77:321–357, 1995

Logic Programming Instantiation of a Dung Argumentation Framework

- ∆ = {q :- p, not s ; s :- not g ; g :- m p ;
 - m }
- Given a set of wff in some logic *L* define :
 1) What constitutes an argument
 - r) what constitutes an argument
 - X = [q :- p, not s ; p]
 - Y = [s :- not g]
 - Z = [g :- m ; m]

q is the **claim** of argument X s is the **claim** of argument Y g is the **claim** of argument Z

Logic Programming Instantiation of a Dung Argumentation Framework

- ∆ = {q :- p, not s ; s :- not g ; g :- m p ;
 - m }
- Given a set of wff in some logic *L* define :
 What constitutes an attack

X = [q :- p, not s; p] and Y = [s :- not g] and Z = [g :- m; m]

$$(Y,X) \in Att$$

Logic Programming Instantiation of a Dung Argumentation Framework

- Δ = {q :- p, not s ; s :- not g ; g :- m p ; m }
- Given a set of wff in some logic *L* define :
 What constitutes an attack

X = [q :- p, not s; p] and Y = [s :- not g] and Z = [g :- m; m]

 $(Y,X) \in Att$, $(Z,Y) \in Att$

Logic Programming Instantiation of a Dung Argumentation Framework

(Args,Att) =

$Z \longrightarrow Y \longrightarrow X$

Argument Evaluation

$\square \quad \mathsf{AF} = (\mathcal{A} rgs, \mathcal{A} tt)$



□ What are the justified / rejected / undecided arguments ?

Evaluation based on intuitive notion of reinstatement / defence



Evaluation based on intuitive notion of reinstatement / defence



Z defends/reinstates X (X is *acceptable* w.r.t. S)

Evaluation based on intuitive notion of reinstatement / defence



- Z defends/reinstates X (X is acceptable w.r.t. S)
- □ If S is conflict free (contains no two arguments that attack), and all arguments in S are acceptable w.r.t. S, then S is admissible

Evaluation based on intuitive notion of reinstatement / defence



The set S of arguments is admissible since it is conflict free and all its contained arguments are acceptable (defended against attacks)

Dung semantics

Let S be admissible:

- S is a *complete* extension iff every argument acceptable w.r.t. S is in S
- S is the grounded extension iff it is the smallest complete extension
- S is a *preferred* extension iff it is a maximal *complete* extension

Other semantics defined in the literature * e.g.

- Stable semantics (maximal conflict free set attacking all arguments outside it)
- Semi-stable semantics
- Ideal Semantics
- Semantics defined for the sake of getting more publications

* P. Baroni and M.Giacomin. Semantics of Abstract Argument Systems. *Argumentation in Artificial Intelligence* (eds. I.Rahwan and G.Simari) 25-45, Springer, 2009

Is \varnothing admissible ? Is \varnothing complete ?

Is {A} admissible ? Is {A} complete ?

Is {A,D} admissible ? Is {A,D} complete ?

 $A \longrightarrow C \longrightarrow D$

What are the grounded (smallest complete) and preferred (largest complete) extensions?

Is \varnothing admissible ?	Yes
Is \varnothing complete ?	No
Is {A} admissible ?	Yes
Is {A} complete ?	No
Is {A,D} admissible ?	Yes
Is {A,D} complete ?	Yes

 $A \longrightarrow C \longrightarrow D$

What are the grounded (smallest complete) and preferred (largest complete) extensions?

Is \varnothing admissible ? Is \varnothing complete ?

Is {A,C} admissible ?

Is {A} admissible ? Is {A} complete ?

Is {C} admissible ? Is {C} complete ?

 $A \longleftrightarrow C$

What are the grounded and preferred extensions?

Is \varnothing admissible ? Yes Is \varnothing complete ? Yes

Is {A,C} admissible ? No

Is {A} admissible ? Yes Is {A} complete ? Yes

Is {C} admissible ? Yes Is {C} complete ? Yes

 $A \longleftrightarrow C$

What are the grounded and preferred extensions?



What are the preferred extensions ? What is the grounded extension ?



What are the preferred extensions ? What is the grounded extension ?





 \varnothing is grounded extension

Labelling Approach to Evaluating Extensions *

Given an AF = (Args, Att)

 $\cdot X \in Args$ is IN iff (Y,X) $\in Att \rightarrow Y$ is OUT

- $X \in Args$ is OUT iff $\exists (Y,X) \in Att$ such that Y is IN
- $X \in Args$ is UNDEC iff $\exists (Y,X) \in Att$ such that Y is UNDEC and $\neg \exists (Y,X) \in Att$

Att such that Y is IN

Each framework can have many legal labellings Legal labelling *minimising* IN is grounded Legal labelling *maximising* IN is preferred Legal labelling UNDEC = \emptyset is stable

* S. Modgil and M.Caminada. Proof Theories and Algorithms for Abstract Argument Frameworks. *Argumentation in Artificial Intelligence* (eds. I.Rahwan and G.Simari) 105-132, Springer, 2009

Example 3

Is Ø admissible ? Is {A} admissible ? Is {A,B} admissible ?



What are the grounded, preferred and stable extensions?

Properties of Extensions

- Many properties of extensions have been studied, e.g.:
- Each AF has a single grounded extension that is the intersection of all complete extensions
- Each stable extension is preferred, but not vice versa
- If X is acceptable w.r.t. an admissible extension E, then E ∪ X is admissible (Fundamental Lemma)

The Justified Arguments of a Framework

X is sceptically justified under semantics E if X is in all E extensions

X is *credulously justified* under semantics E if X is in *at least one* E extension



{A,D} and {B,D} are preferred extensions

➔ D is sceptically justified under preferred

Ø is grounded extension

 no argument is sceptically or credulously justified under grounded semantics

Argumentation-based Nonmonotonic inference relation

- $\Box \quad \text{Abstract} (\mathcal{A} rgs, \mathcal{A} tt) \text{ defined by set of wff } \Delta \text{ in logic } \mathcal{L}$
- $\Box \quad \Delta \mid_{\widetilde{AF}} \alpha$ iff α is the claim of a sceptically justified argument in $\mathcal{A}rgs$
- Logic programming, default logic, auto-epistemic logic, defeasible logic,
 ... all shown to conform to Dung's semantics
- e.g.

 $\Delta \mid_{LP} \alpha$ under well founded semantics iff $\Delta \mid_{AF} \alpha$ under grounded semantics

Argumentation-based characterisation of non-monotonic inference in logic programming

• $\Delta \mid_{\mathcal{AF}} \alpha$ under well founded semantics iff $\Delta \mid_{\mathcal{AF}} \alpha$ under grounded semantics

 $X = [q :- p, not s ; p] \leftarrow Y = [s :- not g] \leftarrow Z = [g :- m ; m]$

Grounded extension is {X , Z} and so $\Delta \mid_{\widetilde{_{AF}}} q$, g

corresponding to $\Delta |_{P}$ q, g

Argumentation-based Nonmonotonic inference relation

- $\Box \quad \text{Abstract} (\mathcal{A} rgs, \mathcal{A} tt) \text{ defined by set of wff } \Delta \text{ in logic } \mathcal{L}$
- $\Box \quad \Delta \mid_{\widetilde{AF}} \alpha$ iff α is the claim of a sceptically justified argument in $\mathcal{A}rgs$
- Logic programming, default logic, auto-epistemic logic, defeasible logic, ... all shown to conform to Dung's semantics :
 Dialectical Semantics alternative to model theoretic semantics :
 True to the extent that all attempts to prove otherwise fail

Argumentation-based Nonmonotonic inference relation

- $\Box \quad \text{Abstract} (\mathcal{A} rgs, \mathcal{A} tt) \text{ defined by set of wff } \Delta \text{ in logic } \mathcal{L}$
- $\Box \quad \Delta \mid_{\widetilde{AF}} \alpha$ iff α is the claim of a sceptically justified argument in Args
- Define arguments and attacks from a possibly inconsistent set Δ of wff in a monotonic logic *L*.
 Yields non-monotonic inference relation Δ |_{AF} α thus resolving inconsistencies in underlying Δ

Classical Logic-based Argumentation *

- Define arguments and attacks from a possibly inconsistent set ∆ of propositional classical wff
- A *CIArg* argument is a pair (Γ , α) such that $\Gamma \subseteq \Delta$, and

1) Γ ⊢_{CL} α
 2) Γ is consistent
 3) No proper subset of Γ entails α

- (Γ, α) attacks (Σ, β) if $\alpha = \neg \delta$ for some $\delta \in \Sigma$

* P. Besnard and A.Hunter. *Elements of Argumentation*, MIT Press, 2008

□ Framework defined by $\Delta = (p, q, p \rightarrow \neg q)$





preferred extension 1



preferred extension 2



preferred extension 3



• 3 preferred/stable extensions corresponding to three max consistent subsets of { p , q , p \rightarrow \neg q } !

• No argument is in every extension (i.e., no argument is sceptically justified)

Argumentation-based Nonmonotonic inference relation

- $\Box \quad \text{Abstract} (\mathcal{A} rgs, \mathcal{A} tt) \text{ defined by set of wff } \Delta \text{ in logic } \mathcal{L}$
- $\Box \quad \Delta \mid_{\widetilde{AF}} \alpha \text{ iff } \alpha \text{ is the claim of a sceptically justified argument in Args}$



• 3 preferred/stable extensions corresponding to three max consistent subsets of { p , q , p \rightarrow \neg q } !

- So AF inference relation does not arbitrate between conflicting conclusions !
- When instantiating with monotonic logic what does argumentation do for you unless you have some way of arbitrating between conflicts ?
- One solution is to use preferences over arguments to arbitrate
- Partial ordering ≤ (preference relation) over arguments

 $(Args, Att, \leq)$

- If X attacks Y and Y strictly preferred to X (X < Y) then X cannot be moved as a successful attack on Y
- So based on ≤ and Att remove unsuccessful attacks and evaluate extensions and justified arguments

* L Amgoud, C Cayrol. <u>A reasoning model based on the production of acceptable arguments</u>. Annals of Mathematics and Artificial Intelligence 34 (1-3), 197-215



 $\{q, p \rightarrow \neg q\}: \neg p < \{p\}: p$ $\{p, p \rightarrow \neg q\}: \neg q < \{q\}: q$



 $\{q, p \rightarrow \neg q\}: \neg p < \{p\}: p$ $\{p, p \rightarrow \neg q\}: \neg q < \{q\}: q$



 $\{p, p \rightarrow \neg q\}$: $\neg q \leq \{q\}$: q

Single preferred extension = { $\{p\} : p, \{q\} : q, \{p,q\} : p \land q \}$

Argumentation-based Nonmonotonic inference relation

- $\Box \quad \Delta \mid_{AF} \alpha$ iff α is the claim of a sceptically justified argument in Args
- $\Delta \mid_{\mathcal{IP}} \alpha$ under well founded semantics iff $\Delta \mid_{\mathcal{AF}} \alpha$ under grounded semantics
 - Is there a non-monotonic inference relation that corresponds to the argumentation inference relation defined by classical logic argumentation with preferences ?

Argumentation based characterisation of Brewka's Non-monotonic **Preferred Subtheories** *

Totally ordered set of propositional classical wff inducing a stratification:

T1	p , q
T2	¬ p , s, ¬s
Tn	

- Start with maximal consistent subset of T1 then maximally consistently extend with formulae in T2 then ... all the way to Tn
- □ → many preferred subtheories consequences of formulae in intersection are non-monotonic inferences (_{ps})

* G. Brewka. Preferred subtheories: an extended logical framework for default reasoning. In *Proc. 11th International Joint Conference on Artificial intelligence*, 1043–1048, 1989.

Argumentation based characterisation of Brewka's Non-monotonic **Preferred Subtheories** *

Totally ordered set of propositional classical wff inducing a stratification:

T1	p , q
T2	¬ p , s, ¬s
Tn	

- Start with maximal consistent subset of T1 then maximally consistently extend with formulae in T2 then ... all the way to Tn
- □ → many preferred subtheories consequences of formulae in intersection are non-monotonic inferences (ps)
- $\Box \quad \text{E.g., } \{p,q,s\} \text{ and } \{p,q,\neg s\} \qquad |\sim_{ps} = Cn(p,q)$

Argumentation based characterisation of Brewka's Preferred Subtheories

- Build classical logic arguments from {T1, ..., Tn }
- X < Y if there is premise in X that is strictly ordered below all premises in Y according to total ordering</p>

 $\{q, p \rightarrow \neg q\}: \neg p < \{p\}: p \qquad \{p, p \rightarrow \neg q\}: \neg q < \{q\}: q$

Evaluate justified arguments under preferred semantics using argument preference ordering to determine successful attacks

* S. Modgil, H. Prakken. A General Account of Argumentation and Preferences. Artificial



Argumentation based characterisation of Brewka's Preferred Subtheories

- Build classical logic arguments from {T₁, ..., T_n }
- X < Y if there is premise in X that is strictly ordered below all premises in Y according to total ordering</p>

Evaluate justified arguments under preferred semantics using argument preference ordering

$$\Box$$
 We * show that $\vdash_{ps} = \vdash_{AF}$

* S. Modgil, H. Prakken. A General Account of Argumentation and Preferences. *Artificial Intelligence* 195(0), 361 - 397, 2013.



More on Abstract Argumentation

- So far we have seen how argumentation can define an inference relation over set of instantiating formulae
- Other works in which argumentation used for decision making (e.g., L. Amgoud, H. Prade. Using arguments for making and explaining decisions. In: Artificial Intelligence (AIJ). V.173, pp. 413-436, 2009)
 - arguments for beliefs (epistemic) and decision options (practical) and evaluation makes use of decision principles
- Extensions of abstract argumentation, e.g.,

 values associated with arguments and ordering over values used to arbitrate amongst arguments (TJM Bench-Capon. *Persuasion in practical argument using valuebased argumentation frameworks*. Journal of Logic and Computation 13 (3), 429-448)

- AFs extended with arguments that attack attacks, so integrating argumentation-based reasoning about preferences (S.Modgil. *Reasoning about preferences in Argumentation Frameworks*. In: Artificial Intelligence (AIJ). V.173, 9-10, 2009.)

The Added Value of Argumentation

Abstract Argumentation



E.g. Logic Programming, Default Logic ...

E.g. Preferred Subtheories

So what accounts for the popularity of argumentation. Who cares and why ?

Basis for defining procedures for *distributed* non-monotonic reasoning based on simple, intuitive principle of reinstatement



Basis for defining procedures for *distributed* non-monotonic reasoning based on simple, intuitive principle of reinstatement



Basis for defining procedures for *distributed* non-monotonic reasoning based on simple, intuitive principle of reinstatement



Basis for defining procedures for distributed non-monotonic reasoning based on simple, intuitive principle of reinstatement



□ Argument Game proof theories → basis for dialogues in which agents exchange arguments to persuade, deliberate over a course of action, negotiate … Evaluation of exchanged arguments decides dialogue outcome (2nd part of tutorial)

- Reinstatement principle intuitive and familiar to human modes of reasoning a debate
- Argumentation based characterisations of computational reasoning understandable and accessible to human reasoning ¹
- Abstractions that accommodate computational and human reasoning can provide bridging role so that ²:
 - Computational reasoning augments human reasoning
 - Human reasoning augments computational reasoning
 - Integrating human & computational reasoning (ontologies/crowd sourcing)

1. H. Mercier and D. Sperber. Why do humans reason? arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2):57–747, 2011.

2. S. Modgil, F. Toni et.al. *The Added Value of Argumentation*. Book chapter in: <u>Agreement Technologies</u>. Springer Verlag, 2013.

- Reinstatement principle intuitive and familiar to human modes of reasoning and debate
- Argumentation based characterisations of computational reasoning understandable and accessible to human reasoning ¹
- Abstractions that accommodate computational and human reasoning can provide bridging role so that ²:

Computational reasoning informs human reasoning (normative/rational)

- Human reasoning augments computational reasoning
- Integrating human & computational reasoning (ontologies/crowd sourcing)

1. H. Mercier and D. Sperber. Why do humans reason? arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2):57–747, 2011.

2. S. Modgil, F. Toni et.al. *The Added Value of Argumentation*. Book chapter in: <u>Agreement Technologies</u>. Springer Verlag, 2013.

Rationality Postulates

Rationality postulates *

- Given an (Args,Att, ≤) defined by set of wff ∆ in logic L what properties would we rationally expect to hold of arguments contained in a complete extension E (remember grounded = smallest complete and preferred = maximal complete)?
- **Consistency** : the claims of arguments in *E* are mutually consistent
- Sub-argument Closure : If X is an argument E then every subargument of X is in E (e.g., [p] is a sub-argument of [q :- p, not s ; p])
- **Closure under Deductive (Strict) Inference** : If $β_1... β_n$ are claims of arguments in E, and If $β_1... β_n$ deductively entail γ then there is an argument in E with claim γ

1. M. Caminada and L.Amgoud. <u>On the evaluation of argumentation formalisms.</u> Artificial Intelligence 171(5-6):286-310 (2007)

Rationality postulates

The ASPIC and ASPIC+ framework

_

- M. Caminada and L.Amgoud. <u>On the evaluation of argumentation formalisms.</u> Artificial Intelligence, 171(5-6):286-310 (2007)
- S. Modgil, H. Prakken. <u>A General Account of Argumentation and Preferences</u>. In: Artificial Intelligence, 195(0), 361 - 397, 2013.

specify conditions under which different logical instantiations (including instantiations by logics with *defeasible inference* rules) of Dung's framework, and different preference relations, satisfy rationality postulates

ASPIC+ Example



 $b \Rightarrow f$ is a defeasible inference rule – 'birds usually fly'

 $p \rightarrow b$ is a strict inference rule – 'penguins are without exception birds'

Attacks can only be directed at the conclusions of defeasible inference rules, and not at the conclusions of strict inference rules



Attacks can only be directed at the conclusions of defeasible inference rules, and not at the conclusions of strict inference rules



Attacks can only be directed at the conclusions of defeasible inference rules, and not at the conclusions of strict inference rules



Suppose we allowed attacks on the conclusions of strict inference rules



Suppose we allowed attacks on the conclusions of strict inference rules



Can anyone see why consistency would be violated ?



There is a preferred extension containing arguments with mutually inconsistent conclusions





Things can go wrong with preferences. Suppose :



Attacks do not succeed



Preferred extension now contains arguments with inconsistent conclusions



But if preference relation satisfies property of being *reasonable* then all three strict preferences are not possible (for otherwise it would result in a cycle in <)


ASPIC+ Example Conditions for satisfaction of postulates



$$\{p,q\}:p \land q \not\leq \{p \rightarrow \neg q\}:p \rightarrow \neg q$$

lf

then

ASPIC+ Example Conditions for satisfaction of postulates



More on ASPIC+

I've glossed over many details, but take home message is:

ASPIC+ is a general framework that allows for a broad range of possible logical instantiations, and provides guidelines for your choice of inference rules, how you define attacks, how you define preferences etc, so that you can be sure that your logical instantiation of Dung frameworks with preferences satisfies rationality postulates.

More papers

S. Modgil, H. Prakken. *The ASPIC+ framework for structured argumentation: a tutorial.* In: **Argument and Computation**, (in press) 2014.

S. Modgil, H. Prakken. <u>A General Account of Argumentation and Preferences</u>. In: Artificial Intelligence (AIJ) . 195(0), 361 - 397, 2013.

H. Prakken. <u>An abstract framework for argumentation with structured arguments</u>. In: **Argument and Computation**, 1(2):93–124, 2010.

Summary

- Dung's abstract theory of argumentation and example logical instantiations (with preferences)
- Correspondences between non-monotonic inference relation of instantiating logic and inference relation defined by claims of justified arguments
- The added value of argumentation generalisation to dialogue (distributed reasoning) and familiar principles in everyday reasoning and debate
- Rationality postulates and ASPIC+

Outline

Logic, Argumentation and Reasoning

- Dung's Theory of Argumentation
- The Added Value of Argumentation
- Rationality Postulates for Logic-based Argumentation

□ Argumentation Based Dialogue

- Argument Game Proof Theories
- Generalisation to Dialogue
- Applications