Assumption-based argumentation

Francesca Toni

Department of Computing
Imperial College London

Outline

- Argumentation
- Computational models of argumentation:
  - Abstract argumentation
  - Assumption-based argumentation
- Dispute derivations and CaSAPI system
- Applications
- Limitations

work with Bondarenko, Dung, Gaertner, Mancarella, Kowalski:

AIJ97, AIJ06,COMMA06, AIJ07,ArgNMR07,ArgMAS07

Argumentation

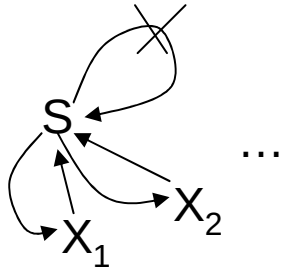
- It allows the evaluation of "*possible conclusions*" by considering reasons for and against (the conclusions and their support)
- It is useful to resolve conflicts (within or across "agents")
- It helps understanding several problems:
  - in Philosophy, Logic, Law, Artificial Intelligence, Computer Science, etc
  - e.g. dispute resolution, decision-making, negotiation, security, bioinformatics etc.
- Many recent argumentation events in AI:
  - *Comput. Models of Nat.Lang. Argument* 01-07
  - *NMR: Arg., Dialogue, and Dec. Making* 02-04-06
  - *ArgMAS* 04-07, *COMMA* 06, *ArgNMR* 07, etc

Plethora of applications

- Decision-making/practical+epistemic reasoning
  - Alternative decisions
  - Contradictory beliefs
  - Preferences
- Dispute resolution (legal setting)
  - Alternative outcomes
  - Incoherent information
- *Security (validating requirements, firewall policy verification)*
- *Bio-informatics (interpretation of results by protein structure prediction tools)*

Computational argumentation

- Needed to support applications!
- Abstract argumentation (Dung AIJ95):
  - Given framework: *(arguments,attack)*
  - A subset *S* of *arguments* is
    - *Admissible* iff *S* does not *attack S* and *S attacks* each X that *attacks* S
    - *Preferred* iff *S* is maximally admissible
    - *Grounded* iff *S* is minimal such that it contains every *a* such that *S* attacks every *X* that attacks *a*
    - *Ideal* iff *S* is admissible and contained in each preferred set
    - …

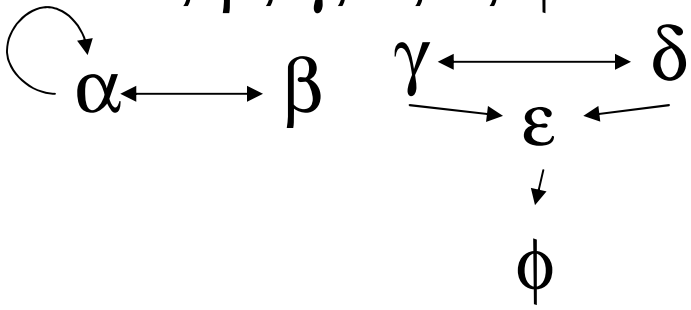


- Several computational mechanisms

Abstract argumentation: different semantics

- Framework:
  - arguments: α; β; γ; δ
  - attack: 
    - {} grounded
    - {β, δ} and {β,γ} (admissible and) preferred
    - {β} sceptical (all) preferred and ideal

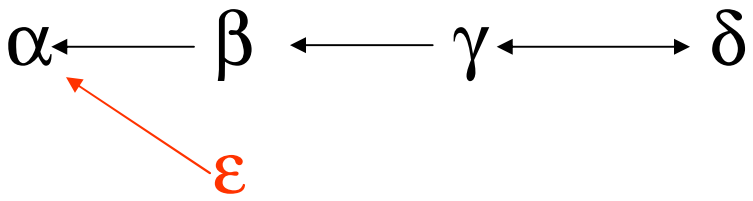
Abstract argumentation: different semantics

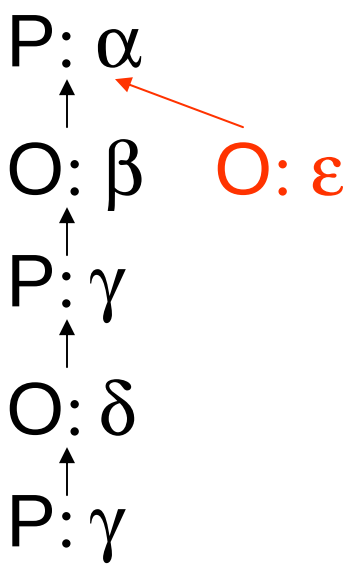
- Ideal semantics is more sceptical than the sceptical preferred semantics*:*
  - arguments: α; β; γ; δ; ε; ϕ
  - attacks: 

  - $\{\beta, \delta, \phi\}$ and $\{\beta, \gamma, \phi\}$ preferred
  - $\{\beta, \phi\}$ sceptical preferred
  - $\{\beta\}$ ideal

Abstract argumentation: computation

- Framework:
  - arguments: $\alpha; \beta; \gamma; \delta ; \varepsilon$
  - attack:





~~{α,γ} is admissible!~~

Pros and cons of abstract argumentation

- Many instances/applications: non-monotonic reasoning, games etc
- Intuitive semantics/computation: game/dispute, “last word wins”
- A lot of work to identify arguments and attacks
- Overlapping between arguments ignored

Pros and cons of abstract argumentation

- A lot of work to identify arguments and attacks:
  - α: *"John is guilty because he was seen with the murder weapon by a reliable witness …."*
  - β: *"but the witness is not reliable because …."*
- Overlapping between arguments ignored
  - α: *"A because B and C …."*
  - β: *"but not B because D…."*
  - γ: *"but not D because C and E …."*

Assumption-based argumentation

In assumption-based argumentation frameworks:

- *arguments* defined in terms of:
  - a **deductive system** **(rules)**
    - e.g. laws/regulations, policy rules, argumentation schemas
  - a set of candidate **assumptions**
    - e.g. uncertain/unsupported beliefs, decisions, “names” of rules
- *attacks* defined in terms of:
  - a notion of **contrary** of assumptions
    - e.g. Negation, alternative decisions, exceptions to rules

Assumption-based argumentation formally

- An assumption-based argumentation framework is $(\mathcal{L},\mathcal{R},A,\bar{\ })$ where
  - $(\mathcal{L},\mathcal{R})$ is a deductive system
    - $\mathcal{L}$ set of sentences
    - $\mathcal{R}$ set of inference rules $s \leftarrow s_1,\ldots,s_n$
  - $A \subseteq \mathcal{L}$ is a set of candidate assumptions
  - $\bar{a}$ is the contrary of assumption a

Assumption-based argumentation formally (cntd)

- *arguments* are *tight* deduction supported by sets of assumptions
- an argument α *attacks* another argument β if the conclusion of α is the contrary of one of the assumptions supporting β

(undermining attacks –

but rebuttal attacks can be obtained too)

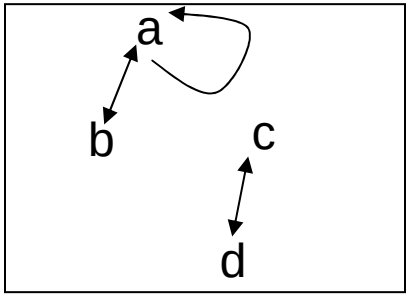
Example

$(\mathcal{L},\mathcal{R},A,\bar{\ })$:

- $\mathcal{L}$={a, b,c,d, ¬a, ¬b, ¬c, ¬d }
- $\mathcal{R}$={¬a ←a; ¬a ← b; ¬b ← a; ¬c ← d; ¬d ← c}
- A={a,b,c,d}
- $\bar{a}$= ¬a; $\bar{b}$= ¬b; $\bar{c}$= ¬c; $\bar{d}$= ¬d

{a} ├ ¬a attacks itself
{b} ├ ¬a attacks {a} ├ ¬b, etc

{a} attacks itself
{b} attacks {a}, etc



All arguments supported by subsets of {b,d} and {b,c} are preferred

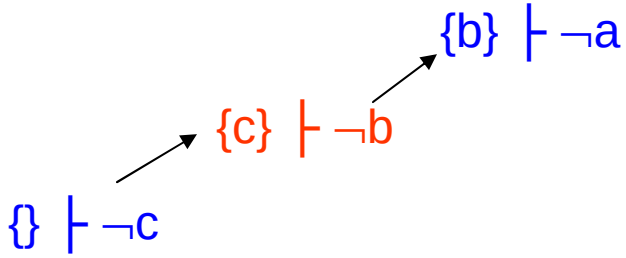
{b,d} and {b,c} preferred

Assumption-based argumentation: computation

- (various kinds of) dispute derivations:
  - Dispute between proponent and opponent
  - Construction of arguments/identification of attacks
  - Outcomes:
    - the initial claim is “acceptable” (e.g. preferred/admissible) or not
    - Assumptions supporting the arguments by the proponent (if acceptable)
    - Assumptions supporting arguments by the opponent and chosen by the proponent to be counter-attacked (if acceptable)

Dispute derivations: example

- $\mathcal{R}$={¬a ← b; ¬b ← c; ¬c}
- ā= ¬a; $\bar{b}$= ¬b; $\bar{c}$= ¬c



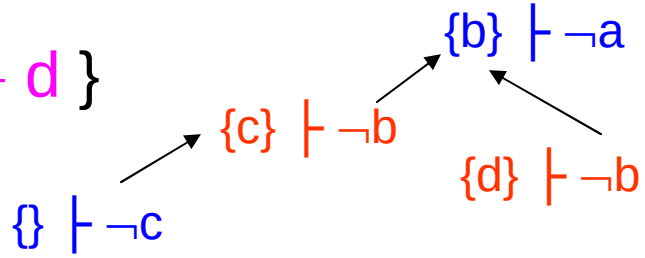
Is ¬a "acceptable"?

Proponent	Opponent	Assumptions supporting Proponent	Culprits chosen in Opponent
{¬a}	{}	{}	{}
{b}	{}	{b}	{}
{}	{{¬b}}	{b}	{}
{}	{{c}}	{b}	{}
{¬c}	{}	{b}	{c}
{}	{}	{b}	{c}

Yes!

Dispute derivations: example (cntd)

- $\mathcal{R}$={¬a ← b; ¬b ← c; ¬c; ¬b ← d }
- ā= ¬a; $\bar{b}$= ¬b; $\bar{c}$= ¬c; $\bar{d}$= ¬d;



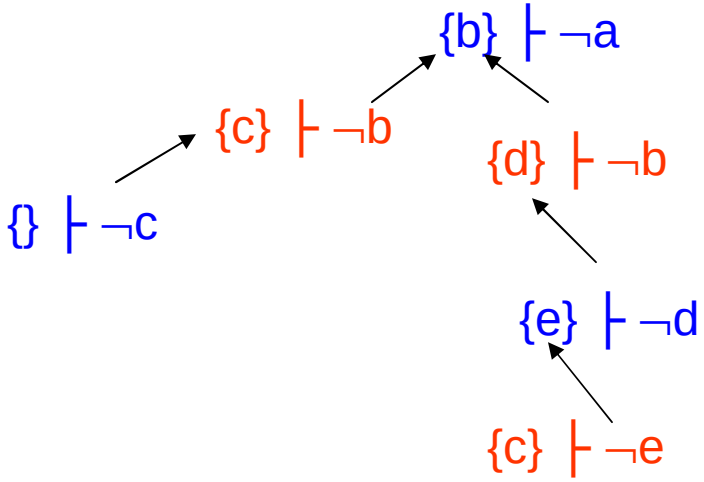
Is ¬a “acceptable”?

Proponent	Opponent	Assumptions supporting Proponent	Culprits chosen in Opponent
{¬a}	{}	{}	{}
{b}	{}	{b}	{}
{}	{{¬b}}	{b}	{}
{}	{{c}, {d}}	{b}	{}
{¬c}	{{d}}	{b}	{c}
{¬d}	{}	{b}	{c}

No!

Dispute derivations: example (cntd)

Overlapping between arguments *not* ignored:



Filtering of culprits by culprits

Various other forms of filtering to exploit overlapping between arguments

Assumption-based argumentation: implementation

- CaSAPI (Credulous and Sceptical Argumentation: Prolog Implementation)

  http://www.doc.ic.ac.uk/~dg00/casapi.html

  http://casapi.sourceforge.net/ (soon)

- 3 kinds of dispute derivations, verbose or silent output, 1 or all answers
- Version 2 (assumptions)
- Version 3 (assumptions and arguments)

demo

Outline revisited

- Argumentation
- Computational models of argumentation:
  - Abstract argumentation
  - Assumption-based argumentation
- Dispute derivations and CaSAPI system
- Applications
- Limitations

Application: dispute resolution

- Rules $\mathcal{R}$: payment ←goodJob

  badJob ←tooLateJob

  badJob ←incompleteJob

  incompleteJob ←delivered, not(accordingToSpec)

  accordingToSpec ←reqA, reqB

  delivered

  not(reqB)

- Assumptions A: goodJob, not(accordingToSpec), reqA, reqB

- Contraries: contrary(goodJob)=badJob;
  contrary(not(accordingToSpec)=accordingToSpec
  contrary(reqA)=not(reqA); contrary(reqB)=not(reqB)

Dispute derivations: payment? NO

Application: decision making

Example from [Bench-Capon,Prakken 06]

- Possible decisions: *prison; fine; (community) service*
- Beliefs:

  punish ← prison
  punish ← fine
  punish ← service
  deter ← prison
  deter ← fine
  ¬ deter ← service
  rehabilitate ← service
  ¬ rehabilitate ← prison
  protect ← prison
- Which decision,
  given goals: punish, deter, rehabilitate (in some order)?

Application: decision making (cntd)

- A = {*prison; fine; service, a,b,c,d*}
- contrary(*prison*) = {*fine, service*} contrary(*fine*) ={*prison; service*}
  contrary(*service*)= {*prison; fine*}
  contrary(*a*)= $\neg$ *deter* contrary(*b*)=*deter*
  contrary(*c*)= $\neg$ *rehabilitate* contrary(*d*)= *rehabilitate*
- $\mathcal{R}$ consists of
  punish ← *prison*
  punish ← *fine*
  punish ← *service*
  deter ← *prison, a*
  deter ← *fine, a*
  $\neg$ *deter* ← *service, b*
  rehabilitate ← *service, c*
  $\neg$ *rehabilitate* ← *prison, d*
  protect ← *prison*
- Dispute derivations: given goals *punish&deter*, decision *prison*

Application: dynamic preferences

Example from [Prakken, Sartor 96]

- Defeasible rules:
  - r1(X): X 's exterior may not be modified if X is a protected building.
  - r2(X): X's exterior may be modified if X needs restructuring.
  - r3(X,Y): R1(X) > R2(Y) if R1(X) concerns artistic bldgs and R2(Y) concerns town planning
  - t(X,Y): R1(X) > R2(Y) if R1(X) is later than R2(Y)
- Strict rules:
  - r1(X) concerns artistic buildings; r2(X) concerns town planning
  - r2(X) is later than r1(X); r3(X,Y) is later than t(X,Y)
  - villa is a protected building and needs restructuring
- What should be believed about modifying villa's exterior?

Application: dynamic preferences (cntd)

In assumption-based argumentation
(see [Kowalski, Toni 96] – but other translations possible)

- Rules $\mathcal{R}$:

  villa 's exterior may not be modified ← not defeated(r1(villa))
  villa's exterior may be modified ← not defeated(r2(villa))
  defeated(r1(villa)) ← not defeated(t(villa,villa)), not defeated(r2(villa))
  defeated(r2(villa)) ← not defeated(r3(villa,villa)),
  not defeated(r1(villa))
  defeated(t(villa,villa)) ← not defeated(t(villa,villa)),
  not defeated(r3(villa,villa))

- Assumptions A are *not defeated(X))*, with contrary *defeated(X).*

Dispute derivations: *villa's exterior may not be modified*

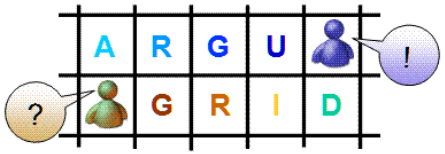
Limitations

- Gathering of information ($\mathcal{R}$, A, etc)?
- Automatic compilation into assumption-based argumentation frameworks (preferences, decisions, argumentation schemas)?
- Presentation of outcomes?
- Multi-party argumentation?
- Scalability ?
- Real-life applications?

Conclusions

- Argumentation tool for various Dung-style (credulous and sceptical) semantics
- Deployed to support decision making in agents and conflict-resolution amongst agents :
  - Morge&Mancarella, ArgMAS07 : MARGO
  - Gaertner&Toni, ArgMAS07 : BDI-like agents
- Ongoing:
  - Interfaces (preferences, decisions)
  - Graph-based visualisation of abstract arguments + attacks
  - Further applications (Security? Bioinformatics?)

Acknowledgements



EU-funded, IST Framework VI

www.argugrid.eu

Royal Academy of Engineering/Leverhulme Foundation Senior Research Fellowship (Francesca Toni)