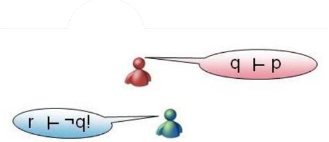
Explaining rational decision making by arguing

Francesca Toni

Workshop on Decision Making, Toulouse, 2017

Department of Computing, Imperial College London, UK
CLArg (Computational Logic and Argumentation) Group



Argumentation in AI

Non-Monotonic Reasoning (NMR)

from late 1980s (e.g. Lin, Shoham, Dung, Kowalski, Kakas, Toni):
$\Rightarrow$ abstract (and bipolar) argumentation, ABA

Defeasible Reasoning as studied in philosophy

from late 1980s (e.g. Pollock, Nute):
$\Rightarrow$ DeLP, ASPIC, ASPIC+

Resolving inconsistencies (paraconsistent reasoning)

from mid 1990s (e.g. Cayrol, Amgoud, Hunter):
$\Rightarrow$ logic-based argumentation

Decision making

from early 1990s (e.g. Fox, Krause, Ambler):
$\Rightarrow$ Amgoud and Prade (2009), . . .

Outline

- Argumentative approaches to “explained” decision-making:
  - descriptive, rational/socially optimal, privacy preserving
- Essential background on argumentation
  - abstract, bipolar, value-based, assumption-based

Main references

- L. Carstens, X. Fan, Y. Gao, F. Toni: An Overview of Argumentation Frameworks for Decision Support. GKR 2015
- M. Aurisicchio, P. Baroni, D. Pellegrini, F. Toni: Comparing and Integrating Argumentation-Based with Matrix-Based Decision Support in Arg&Dec. TAFA 2015
- Y. Gao, F. Toni, H. Wang, F. Xu: Argumentation-Based Multi-Agent Decision Making with Privacy Preserved. AAMAS 2016

Collaborative MAS decisions vs Abstract Argumentation

- socially optimal and privacy preserving distributed constraint satisfaction
- explanations via related admissibility in abstract argumentation

Abstract Argumentation (AA) – [Dung 1995]

An AA framework is a pair $\langle Args, attacks \rangle$ where

- *Args* is a set (the *arguments*)
- $attacks \subseteq Args \times Args$ is a binary relation over *Args*

Example (AA framework represented as a directed graph)

α: I love Toulouse because it is nice and small

β: Small? with 500k people? γ: It is small wrt London!

$$\alpha \longleftarrow \beta \longleftarrow \gamma$$

Semantics, e.g. $A \subseteq Args$ is

- **conflict-free** (c-f) iff it does not attack itself
- **admissible** iff it is c-f and attacks each attacking argument

Example

$\{\beta\}$ is conflict-free, $\{\gamma\}$, $\{\alpha, \gamma\}$ are admissible

Related admissible sets of arguments in AA [Fan&Toni 2015]

$A \subseteq Args$ is *related admissible* iff

$\exists a \in A$: A is admissible & A **r-defends** a (a is *a topic* of A), where

- $a \in Args$ *r-defends* $b \in Args$ iff
  $a = b$ or
  $\exists c \in Args$ s.t. a attacks c and c attacks b or
  $\exists c \in Args$ s.t. a r-defends c and c r-defends b
- $A \subseteq Args$ *r-defends* $a \in Args$ iff for each $b \in A$: b r-defends a

$A \subseteq Args$ is an *explanation* of $a \in Args$ iff

A is related admissible and a is a topic of A

Example

$$\omega \qquad\qquad \alpha \longleftarrow \beta \longleftarrow \gamma$$

$\{\alpha, \gamma\}$ is an explanation of α
$\{\alpha, \gamma, \omega\}$ is admissible but not an explanation of α

Privacy preserving decisions in collaborative MAS

Problems requiring information sharing, conflict resolution and privacy preservation.

Example (Variant of the battle of the sexes)

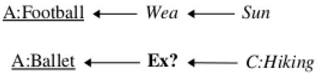
Alice (A): I definitely prefer ballet. **But will Bob's ex-wife be there**? Caroline (C) said that she will be hiking. . . . Bob (B): I definitely prefer football. **Does Alice like football?** She surely enjoys sports, as she enjoys tennis. Caroline (C) posted on Facebook that she is in the ballet hall with her mother. . . .

Solutions = *strategy profiles* which are:

- *feasible*: all actions are 'doable' according to all agents (e.g. attending ballet is not doable for A if B's ex-wife is there too)
- *acceptable*: all constraints are met (e.g. A and B want to be together)
- *socially optimal*: no other solution is "better" for any agent
- *secure*: **private information** is not (in)directly disclosed

"Battle of the sexes" example

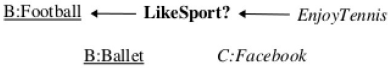
Alice's AA (internal) framework:



A:Football ← *Wea* ← *Sun*

A:Ballet ← **Ex?** ← *C:Hiking*

Bob's AA (internal) framework:



B:Football ← **LikeSport?** ← *EnjoyTennis*

B:Ballet

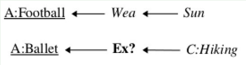
C:Facebook

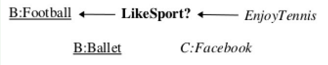
- several types of arguments: private practical, **private epistemic**, *disclosable epistemic*
- several restrictions over attacks: practical arguments are c-f, practical arguments do not attack epistemic ones, . . .
- there may be attacks across (between disclosable arguments), e.g. *C: Facebook* attacks *C: Hiking*

Solving collaborative MAS by arguing

- distributed constraints satisfaction algorithm (with backtracking), incorporating
- variant of TPI-dispute to exchange “compact reasons” drawn from explanations (guaranteed to be disclosable!)

Example





A: C says she will be hiking with your ex-wife today. . .
({*C: Hiking*,A:Ballet} is the only explanation for A:Ballet)

B: But she has just posted on Facebook that they are at the ballet now.

A: I see. Shall we go and watch football?

B: if I’m not mistaken, you enjoy watching sport, right?
({*B: EnjoyTennis*,B:Football} is the only explanation for B:Football)

Collaborative MAS decisions vs Value-Based Argumentation

- Reinforcement Learning agents - converging to optimal policy
- actions are supported by arguments, which promote values; preferences over values

Value-based Argumentation (VbA) [Bench-Capon 2003]

Example

- Consider the AA framework $a \leftrightarrow b$ where

  a: Let's have dinner at home today
  b: Let's have dinner in a restaurant today

- $\{a\}$ and $\{b\}$ are both admissible

VbA uses **preferences over values promoted by arguments**

Example ($a \leftrightarrow b$)

- Consider **values**

  $v1$: Money-saving, where a **promotes** $v1$
  $v2$: Time-saving, where b **promotes** $v2$

- if $v1 > v2$ then $a \rightarrow b$: $\{a\}$ is admissible, $\{b\}$ is not
- if $v2 > v1$ then $a \leftarrow b$: $\{b\}$ is admissible, $\{a\}$ is not

VbA for Cooperative Multi-Agent Decisions (CMAD)

Decisions = actions:

- "Internal conflicts": each agent may have multiple alternative actions to take, but can only choose one at a time
- "External conflicts": multiple agents may want to perform the same action, but this action can/should be performed by one agent only

RoboCup

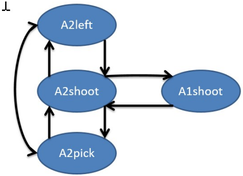
Exit	Ag2 (gold)	Wumpus	Ag1	

Multi-agent wumpus world

Example of VbA for CMAD

Exit	Ag2 (gold)	Wumpus	Ag1	

⇒



- **A1shoot**: Ag1 should do *shoot_left* because there is a Wumpus next to Ag1, on its left
- **A2left**: Ag2 should do *go_left* because the exit is on its left
- **A2pick**: Ag2 should do *pickup* because gold is in its square.

Vsafe: agents' safety

Vmoney: money-making

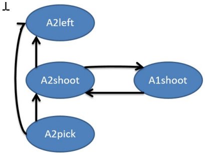
Vexit: exit wumpus world

A1shoot and **A2shoot** promote *Vsafe*

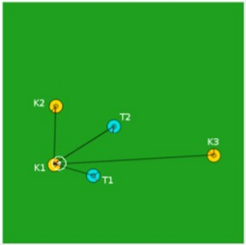
A2pick promotes *Vmoney*

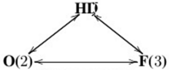
A2left promotes *Vexit*

$Vmoney > Vsafe > Vexit \Rightarrow$

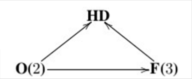


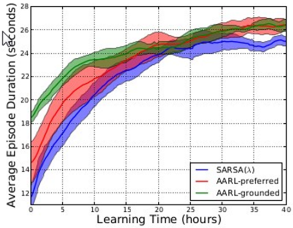
VbA+Reinforcement Learning for RoboCup [Gao&Toni 2014]





(a) Keepaway game





(a) 2-Keepaway

Decision matrices vs Bipolar Argumentation

- matrices: selection criteria for decisions/concept variants
- debates in Bipolar Argumentation (attack and support) over selection criteria and decisions

Bipolar Argumentation (BA) [Cayrol&Lagasquie-Schiex 2005], ...

An BA framework is a triple $\langle Args, attacks, supports \rangle$ where

- $\langle Args, attacks \rangle$ is an AA framework
- $supports \subseteq Args \times Args$ is a binary relation over $Args$

Example (BA framework represented as a directed graph)

γ: Toulouse is small wrt London! δ: London has over 10M people

$$\alpha \longleftarrow \beta \longleftarrow \gamma \Longleftarrow \delta$$

Semantics, e.g.

- $A \subseteq Args$ is **admissible** iff ...
- the (dialectical) **strength** of $a \in Args$ is ...

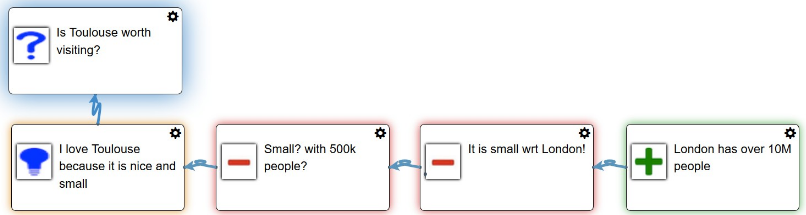
Example

$\{\alpha, \gamma, \delta\}$ is "admissible", $\{\beta\}$ is not
α has strength 0.4375, β has strength 0.125 (within [0,1])

QuAD (Quantitative Argumentation Debates) for Bipolar Argumentation

Arg&Dec (`www.arganddec.com`)

$$\alpha \longleftarrow \beta \longleftarrow \gamma \Longleftarrow \delta$$



Is Touloue worth visiting?

I love Toulouse because it is nice and small

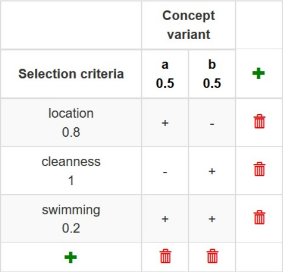
Small? with 500k people?

It is small wrt London!

London has over 10M people

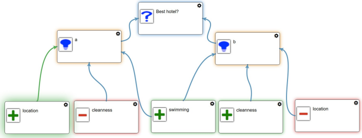
QuAD and DF-QuAD methods for determining “strength”

Arg&Dec for decision-making

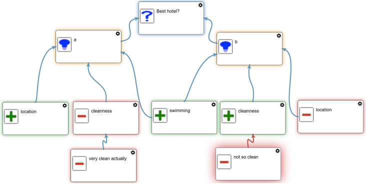


	Concept variant		
Selection criteria	a 0.5	b 0.5	✚
location 0.8	+	-	🗑
cleanness 1	-	+	🗑
swimming 0.2	+	+	🗑
✚	🗑	🗑	

b "better than" a



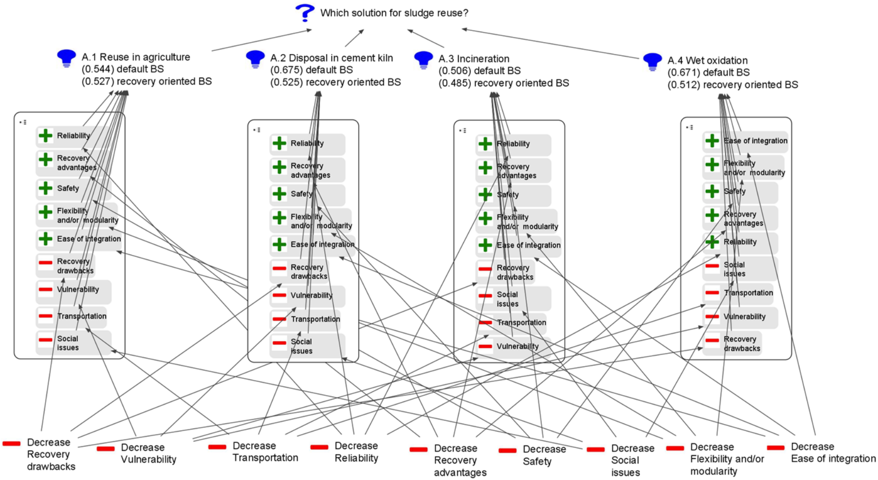
b stronger than a



a stronger than b

a "better than" b

BA/QuAD: applications



Which solution for sludge reuse?

A.1 Reuse in agriculture
(0.544) default BS
(0.527) recovery oriented BS

- + Reliability
- + Recovery advantages
- + Safety
- + Flexibility and/or modularity
- + Ease of integration
- − Recovery drawbacks
- − Vulnerability
- − Transportation
- − Social issues

A.2 Disposal in cement kiln
(0.675) default BS
(0.525) recovery oriented BS

- + Reliability
- + Recovery advantages
- + Safety
- + Flexibility and/or modularity
- + Ease of integration
- − Recovery drawbacks
- − Vulnerability
- − Transportation
- − Social issues

A.3 Incineration
(0.506) default BS
(0.485) recovery oriented BS

- + Reliability
- + Recovery advantages
- + Safety
- + Flexibility and/or modularity
- + Ease of integration
- − Recovery drawbacks
- − Social issues
- − Transportation
- − Vulnerability

A.4 Wet oxidation
(0.671) default BS
(0.512) recovery oriented BS

- + Ease of integration
- + Flexibility and/or modularity
- + Safety
- + Recovery advantages
- + Reliability
- − Social issues
- − Transportation
- − Vulnerability
- − Recovery drawbacks

− Decrease Recovery drawbacks
− Decrease Vulnerability
− Decrease Transportation
− Decrease Reliability
− Decrease Recovery advantages
− Decrease Safety
− Decrease Social issues
− Decrease Flexibility and/or modularity
− Decrease Ease of integration

Optimal decisions vs Assumption-based Argumentation

- decisions (have attributes that) fulfil goals, (possibly) preferences over goals, various notions of optimal decisions
- structured argumentation, debate trees as explanations

Assumption-based Argumentation (ABA) [Bondarenko et al 1997]

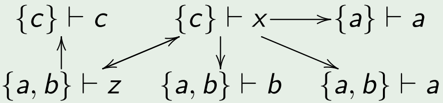
- a form of *structured* argumentation:
  - arguments are constructed from *rules*, and supported by *assumptions*
  - attacks are on the assumptions supporting arguments, by arguments for *contraries* of these assumptions

Example (Flat ABA frameworks give AA frameworks)

An ABA framework with

- rules $\mathcal{R} = \{x \leftarrow c, \quad z \leftarrow b, \quad a \leftarrow b\}$,
- assumptions $\mathcal{A} = \{a, b, c\}$,
- contraries $\overline{a} = x, \overline{b} = y, \overline{c} = z$

gives the AA framework:



$\{c\} \vdash c$

$\{c\} \vdash x$

$\{a\} \vdash a$

$\{a, b\} \vdash z$

$\{a, b\} \vdash b$

$\{a, b\} \vdash a$

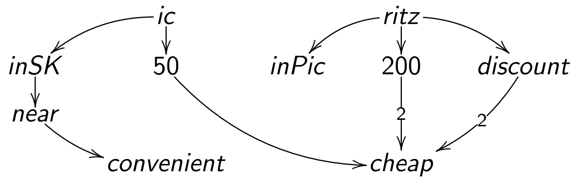
ABA for Multi-Criteria Decision Making

- from decision frameworks to (flat) ABA frameworks: "optimal decisions" form admissible sets of arguments
- "dispute trees" explain (optimality of) decisions:
  1. each node of a dispute tree $\mathcal{T}$ is labelled by some $\chi \in Args$ and is by the *proponent* 🟢 or the *opponent* 🔴
  2. for each 🟢 node *n*, labelled by some $\beta \in Args$, and for every $(\gamma, \beta) \in attacks$ there is a 🔴 child of *n* labelled by γ
  3. for each 🔴 node *n*, labelled by some $\beta \in Args$, there is *exactly* one child of *n* which is by 🟢 and labelled by some γ such that $(\gamma, \beta) \in attacks$
  4. there are no other nodes in $\mathcal{T}$

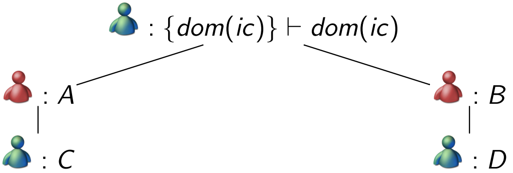
The set of all 🟢 arguments in *admissible dispute trees* (where no argument labels both 🟢 and 🔴 nodes) is admissible.

Example: ABA for decision graphs and "dominant" decisions

- decision graph:



- dominant decision: ic (meets all goals: convenient and cheap)
- ABA dispute tree:



: $\{dom(ic)\} \vdash dom(ic)$

: A

: B

: C

: D

$A = \{notMet(ic, convenient)\} \vdash notDom(ic)$

$B = \{notMet(ic, cheap)\} \vdash notDom(ic)$

$C = \{\ldots\} \vdash met(ic, convenient)$ $D = \{\ldots\} \vdash met(ic, cheap)$

Summary

- AA and VbA for cooperative MAS decisions
- BA and QuAD for matrix-based decisions
- ABA for multi-attribute decisions

rational, explainable decisions, supported by tools for computational argumentation

AA-CBR

Case-based Reasoning (CBR):

- Given *past cases* (S, o) (S features, $o \in \{+, -\}$ outcome)

  e.g. $(\{ensuite, wireless\}, +)$, $(\{small\}, -)$
- a default outcome $d \in \{+, -\}$

  e.g. $d =+$
- Determine the outcome of new case (with features) N

  e.g. $N =\{ensuite, small\}$

CBR by mapping onto AA:

- Arguments: past cases, $(N, ?)$, $(\emptyset, d)$

  e.g. $(\{ensuite, wireless\}, +)$, $(\{small\}, -)$, $(\{ensuite, small\}, ?)$, $(\emptyset, +)$
- Attack by $\neq$outcome&specificity&coincision/irrelevance:

  e.g. $(\{small\}, -)$ attacks $(\emptyset, +)$, $(\{ensuite, small\}, ?)$ attacks $(\{ensuite, wireless\}, +)$
- outcome of N is d ($\overline{d}$) if $(\emptyset, d)$ is (not) in grounded extension

  e.g. the outcome for $N =\{ensuite, small\}$ is $-$
- dispute trees as explanations of outcomes