Coalition Logic: links to current research

Natasha Alechina   Brian Logan

Utrecht University

n.a.alechina@uu.nl   b.s.logan@uu.nl

In this recording:

- properties of coalition logic (axiom system)
- combining CL with epistemic operators
- complexity of reasoning
- use of CL for formalising the notion of responsibility

Complete axiom system

- a system of axioms and inference rules is *complete* for some logic if any valid (universally true) formula in the logic is derivable from the axioms using the rules
- not all logics have a complete axiomatisation
- coalition logic does

Complete axiom system for CL

PL complete set of axioms for propositional logic

Safety $[\mathbb{A}\mathrm{gt}]\top$

Liveness $\neg[A]\bot$

$\mathbb{A}\mathrm{gt}$-maximality $\neg[\emptyset]\varphi \to [\mathbb{A}\mathrm{gt}]\neg\varphi$

Superadditivity $[A_1]\varphi \wedge [A_2]\psi \to [A_1 \cup A_2](\varphi \wedge \psi)$ for any disjoint $A_1, A_2 \subseteq \mathbb{A}\mathrm{gt}$

inference rules

Modus Ponens from φ and $\varphi \to \psi$ derive ψ

Monotonicity if $\varphi \to \psi$ is a theorem (you already proved it), then $[A]\varphi \to [A]\psi$ is also a theorem.

Alternative semantics for CL

- it is often easier to work with so-called effectivity semantics for CL (instead of CGS or action semantics)
- an effectivity model $\mathcal{E} = (St, E, \mathcal{V})$ where $E : St \to 2^{\mathbb{A}\mathrm{gt}} \to 2^{2^{St}}$ is an *effectivity function*
- intuitively, in a state q, a coalition A is *effective for*, or has an action guaranteeing the outcome to be in a certain set of states
- $\mathcal{E}, q \models [A]\varphi$ iff $\{q' \mid \mathcal{E}, q' \models \varphi\} \in E(q)(A)$
- for example, in Prisoner's Dilemma, $E(q_0)(\{1\}) = \{\{q_1, q_2\}, \{q_3, q_4\}, \{q_1, q_2, q_3\}, \{q_1, q_2, q_4\}, \{q_1, q_3, q_4\}, \{q_2, q_3, q_4\}, \{q_1, q_2, q_3, q_4\}$ plus all the same sets with q_0 added $\}$
- an effectivity function corresponds to a concurrent game structure iff it has certain properties (is 'truly playable')

Truly playable effectivity function

Outcome monotonicity $X \in E(q, A)$ and $X \subseteq Y$ implies $Y \in E(q, A)$

Safety $E(q, A) \neq \emptyset$ ($\mathbb{A}\mathrm{gt}$ should be able to enforce *something*
(fluffed this in the recording)

Liveness $\emptyset \notin E(q, A)$

Superadditivity If $A_1 \cap A_2 = \emptyset$, $X \in E(q, A_1)$ and $Y \in E(q, A_2)$, then $X \cap Y \in E(q, A_1 \cup A_2)$

$\mathbb{A}\mathrm{gt}$-maximality $\overline{X} \notin E(q, \emptyset)$ implies $X \in E(q, \mathbb{A}\mathrm{gt})$

Determinacy If $X \in E(q, \mathbb{A}\mathrm{gt})$, then $\{x\} \in E(q, \mathbb{A}\mathrm{gt})$ for some $x \in X$

Adding epistemic operators to CL

- we can add $\sim_i$ relations between states in a CGS
- epistemic operators are defined as usual

Interaction between knowledge and ability for effectivity functions

- agents have the same ability in two states that they cannot distinguish:
- $q \sim_i q'$ implies $X \in E(q, \{i\})$ iff $X \in E(q', \{i\})$
- corresponding axiom: $[\{i\}]\varphi \rightarrow K_i[\{i\}]\varphi$

Complexity of the satisfiability problem

- Satisfiability: given a formula φ of CL, is it satisfiable?
- satisfiability for CL is in PSPACE (Pauly 2001)
- satisfiability for CL + K + D and interaction axioms: also PSPACE
- see Agotnes and Alechina: Coalition logic with individual, distributed and common knowledge, Journal of Logic and Computation, Volume 29, Issue 7, 2019, 1041-1069
- satisfiability for CL + C + interaction axiom: still open?

Responsibility and blameworthiness

- CL started in Philosophy
- used for reasoning about games
- another relevant problem for AI: what does it mean that a group of agents is responsible (or to blame) for a state of affairs φ
- CL or CL-like logic can be used to make this notion precise

Responsibility and blameworthiness

- classic approach: a group A is to blame for φ if:
  - φ is the case now
  - and in the initial state:
  - A could have prevented φ: $[A]\neg\varphi$
  - A knew this: $E_A[A]\neg\varphi$ (or $D_A[A]\neg\varphi$, $C_A[A]\neg\varphi$)
  - A is a minimal such group: for no $B \subset A$, $[B]\neg\varphi$.
- for more, see: Naumov, Tao: An epistemic logic of blameworthiness. *Artificial Intelligence* 283: article 103269 (2020)
- added in additional materials for week 5