Heuristics in Argumentation: A Game-Theorical Investigation

Régis Riveret, University of Bologna,
Henry Prakken, Utrecht University and University of Groningen,
Antonino Rotolo, University of Bologna,
Giovanni Sartor, European University Institute.

May 30, 2008

A four-layered view on argumentation

H. Prakken's four-layered view on argumentation:

1. The logical layer defines how single arguments can be built.

A four-layered view on argumentation

H. Prakken's four-layered view on argumentation:

1. The logical layer defines how single arguments can be built.
2. The dialectical layer, focuses on conflicting arguments and defines dialectical status of arguments.

A four-layered view on argumentation

H. Prakken's four-layered view on argumentation:

1. The logical layer defines how single arguments can be built.
2. The dialectical layer, focuses on conflicting arguments and defines dialectical status of arguments.
3. The procedural layer regulates the conduct of argumentative dialogues.

A four-layered view on argumentation

H. Prakken's four-layered view on argumentation:

1. The logical layer defines how single arguments can be built.
2. The dialectical layer, focuses on conflicting arguments and defines dialectical status of arguments.
3. The procedural layer regulates the conduct of argumentative dialogues.
4. The heuristic layer deals with the strategies in dialogues.

A four-layered view on argumentation

H. Prakken's four-layered view on argumentation:

1. The logical layer defines how single arguments can be built.
2. The dialectical layer, focuses on conflicting arguments and defines dialectical status of arguments.
3. The procedural layer regulates the conduct of argumentative dialogues.
4. The heuristic layer deals with the strategies in dialogues.

We are interested here in the heuristic layer.

Game-theorical heuristics

- Observation: an arguer makes moves by taking into account moves of the other player.

Game-theorical heuristics

- Observation: an arguer makes moves by taking into account moves of the other player.
- Problem: how to determine optimal strategies in a dialogue games for argumentation?

Game-theorical heuristics

- Observation: an arguer makes moves by taking into account moves of the other player.
- Problem: how to determine optimal strategies in a dialogue games for argumentation?
- Solution: we propose the use of game-theorical tools.

Adjudication debates

We focus on ‘adjudication debates’:

1. Two parties argue on a claim,
2. A neutral party decides whether to accept the statements stated during the debate.

Preferences over strategies

Preferences over strategies

1. Moves have costs and benefits.

Preferences over strategies

1. Moves have costs and benefits.
2. Opposing arguers make estimates how likely it is that the premises of their arguments will be accepted by the adjudicator.

Assumptions on the logic

1. Arguments have a finite nonempty set of premises and one conclusion.
2. There is a binary relation of defeat between arguments.

Assumptions on the game protocol

1. An *argument game* is played by two players *Pro* and *Opp*.

Assumptions on the game protocol

1. An *argument game* is played by two players *Pro* and *Opp*.
2. A *move* is a withdrawal or is an argument that defeats an argument previously moved by the other party (except the first move).

Assumptions on the game protocol

1. An *argument game* is played by two players *Pro* and *Opp*.
2. A *move* is a withdrawal or is an argument that defeats an argument previously moved by the other party (except the first move).
3. Player *Pro* does not repeat moves.

Assumptions on the game protocol

1. An *argument game* is played by two players *Pro* and *Opp*.
2. A *move* is a withdrawal or is an argument that defeats an argument previously moved by the other party (except the first move).
3. Player *Pro* does not repeat moves.
4. Each *turn* of an argument game consists of a withdrawal or a sequence of maximum *m* arguments. The first turn consists of a single argument or a withdrawal (i.e. no debate takes place).

Assumptions on the game protocol

1. An *argument game* is played by two players *Pro* and *Opp*.
2. A *move* is a withdrawal or is an argument that defeats an argument previously moved by the other party (except the first move).
3. Player *Pro* does not repeat moves.
4. Each *turn* of an argument game consists of a withdrawal or a sequence of maximum *m* arguments. The first turn consists of a single argument or a withdrawal (i.e. no debate takes place).
5. The turn shifts after a player has made 1 or at maximum *m* moves in a row and indicates explicitly that she has ended her turn.

Assumptions on the game protocol

1. An *argument game* is played by two players *Pro* and *Opp*.
2. A *move* is a withdrawal or is an argument that defeats an argument previously moved by the other party (except the first move).
3. Player *Pro* does not repeat moves.
4. Each *turn* of an argument game consists of a withdrawal or a sequence of maximum *m* arguments. The first turn consists of a single argument or a withdrawal (i.e. no debate takes place).
5. The turn shifts after a player has made 1 or at maximum *m* moves in a row and indicates explicitly that she has ended her turn.
6. Each argument move other than the first one defeats its target argument.

Assumptions on argument games

1. A game *terminates* if a player withdraws. If the set of arguments is finite then each game terminates, since the proponent may not repeat arguments.

Assumptions on argument games

1. A game *terminates* if a player withdraws. If the set of arguments is finite then each game terminates, since the proponent may not repeat arguments.
2. Each game induces a *reply tree*, which consists of the argument moves as nodes and their target relations as links.

Assumptions on argument games

1. A game *terminates* if a player withdraws. If the set of arguments is finite then each game terminates, since the proponent may not repeat arguments.
2. Each game induces a *reply tree*, which consists of the argument moves as nodes and their target relations as links.
3. Reply trees can be labeled as follows: a node is *in* iff all its children are *out*; and a node is *out* iff it has a child that is *in*.

Assumptions on argument games

1. A game *terminates* if a player withdraws. If the set of arguments is finite then each game terminates, since the proponent may not repeat arguments.
2. Each game induces a *reply tree*, which consists of the argument moves as nodes and their target relations as links.
3. Reply trees can be labeled as follows: a node is *in* iff all its children are *out*; and a node is *out* iff it has a child that is *in*.
4. An argument move in a reply tree *favours Pro* if the argument move is *in*; otherwise it favours *Opp*.

Assumptions on argument games

1. A game *terminates* if a player withdraws. If the set of arguments is finite then each game terminates, since the proponent may not repeat arguments.
2. Each game induces a *reply tree*, which consists of the argument moves as nodes and their target relations as links.
3. Reply trees can be labeled as follows: a node is *in* iff all its children are *out*; and a node is *out* iff it has a child that is *in*.
4. An argument move in a reply tree *favours Pro* if the argument move is *in*; otherwise it favours *Opp*.
5. A game is *won* by a player if at termination the initial move favours the player.

Four structures

1. A defeat graph in which the nodes are arguments and the links are defeat relations; which is a declarative representation of a set of available arguments with their defeat relations.

Four structures

1. A defeat graph in which the nodes are arguments and the links are defeat relations; which is a declarative representation of a set of available arguments with their defeat relations.
2. A reply tree of a single-move argument game in which the nodes are arguments and the links are reply links.

Four structures

1. A defeat graph in which the nodes are arguments and the links are defeat relations; which is a declarative representation of a set of available arguments with their defeat relations.
2. A reply tree of a single-move argument game in which the nodes are arguments and the links are reply links.
3. A multi-move argument game which is a sequence of turns by two players *Pro* and *Opp*. Each turn consists of zero or more arguments;

Four structures

1. A defeat graph in which the nodes are arguments and the links are defeat relations; which is a declarative representation of a set of available arguments with their defeat relations.
2. A reply tree of a single-move argument game in which the nodes are arguments and the links are reply links.
3. A multi-move argument game which is a sequence of turns by two players *Pro* and *Opp*. Each turn consists of zero or more arguments;
4. A game tree of all possible turn games in which the nodes are turns and the links express their temporal order in a game.

Example

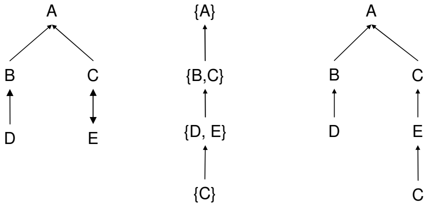


Figure: In the middle, a single terminated argument game based on the defeat graph on the left, and its reply graph on the right.

Game-theorical assumptions

1. Arguers can plan moves whenever she has to move: we model dialogues as *extensive games*.

Game-theorical assumptions

1. Arguers can plan moves whenever she has to move: we model dialogues as *extensive games*.
2. Arguers are perfectly informed about the arguments previoulsy advanced by the other arguer: *extensive games with perfect information*.

Game-theorical assumptions

1. Arguers can plan moves whenever she has to move: we model dialogues as *extensive games*.
2. Arguers are perfectly informed about the arguments previoulsy advanced by the other arguer: *extensive games with perfect information*.
3. The set of all arguments and their defeat relations is given in advance, is finite, stays fixed during a game and is known by both players between the games: *extensive games with perfect and complete information*.

Dialogue games as extensive games

An extensive game is composed of:

1. Players: opponent and proponent.

Dialogue games as extensive games

An extensive game is composed of:

1. Players: opponent and proponent.
2. Histories: sequences of turns.

Dialogue games as extensive games

An extensive game is composed of:

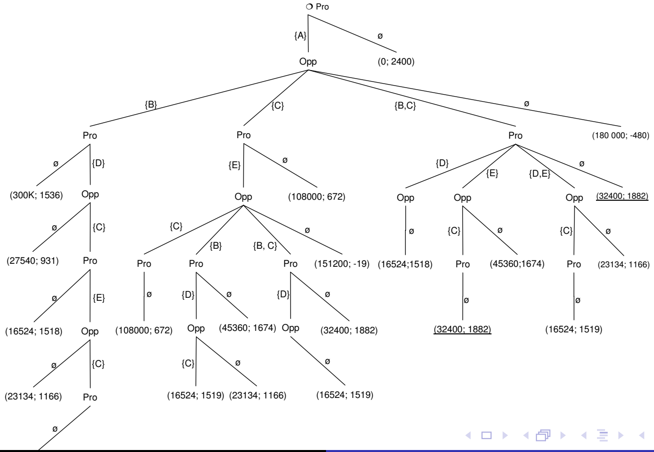
1. Players: opponent and proponent.
2. Histories: sequences of turns.
3. A player turn function: the arguer turn function.

Dialogue games as extensive games

An extensive game is composed of:

1. Players: opponent and proponent.
2. Histories: sequences of turns.
3. A player turn function: the arguer turn function.
4. A preference relation for each player over terminated histories.

Dialogue games as extensive games



Strategies

The strategy of an arguer is the specification of the sequences of arguments chosen by the arguer for every history after which it is her turn to move.

Definition

A strategy of arguer $i \in N$ in an extensive argumentation game with perfect information $\langle N, H, P, (\succeq_i) \rangle$ is a function that assigns a move $M(h)$ to each nonterminal history $h \in H - Z$ for which $P(h) = i$.

Equilibrium

- In strategic games, it is usual to consider Nash equilibrium: no player has anything to gain by changing her strategy.

Equilibrium

- In strategic games, it is usual to consider Nash equilibrium: no player has anything to gain by changing her strategy.
- In extensive game, we consider *subgame perfect equilbrium*: a subgame perfect equilibrium is a Nash equilibrium of every subgame of the original game.

Equilibrium

Definition

A subgame perfect equilibrium of an extensive argumentation game with perfect information $\Gamma = \langle N, H, P, (\succeq_i) \rangle$ is a strategy profile s^* such that for every nonterminal history $h \in H - Z$ for which $P(h) = i$, $i \in \{Opp, Pro\}$, we have:

$$Out_h(s^*_{Pro}|_h, s^*_{Opp}|_h) \succeq_{Opp|h} Out_h(s^*_{Pro}|_h, s_{Opp})$$

$$Out_h(s^*_{Pro}|_h, s^*_{Opp}|_h) \succeq_{Pro|h} Out_h(s_{Pro}, s^*_{Opp}|_h)$$

for every s_{Pro} and s_{Opp} in the subgame $\Gamma(h)$.

Backwards induction

- The subgame perfect equilbrium can be compiled by using standard backwards induction.

Backwards induction

- The subgame perfect equilbrium can be compiled by using standard backwards induction.
- Backward induction: start at a player's final decision nodes to see what a player will do there, and then reasons backwards to tell which action is best for the other player.

Preferences specifications

The preference relation is defined by means of an utility function $EU_i : Out(s) \rightarrow \mathbb{R}$ such that:

$$Out(s) \succeq_i Out(s') \text{ if and only if } EU_i(Out(s)) \geq EU_i(Out(s')).$$

Expected utility

The utility function is specified in terms of expected utility.

$$EU(X) = \sum_{i=1}^{n} Pr(o_i).u(o_i)$$

where $o_1, \ldots, o_n$ are the possible (and mutually exclusive) outcomes of X.

Outcomes of a game

The game-theorical outcome $Out(s)$ of a strategy profile s is a terminal history, i.e. the dialogue resulting from s. For each terminated game associated to a strategy profile s, we have two mutually exclusive utility outcomes: an arguer can win or lose In other words, the initial argument is succcessful or not.

$$\begin{aligned} EU_i(Out(s)) = {} & Pr(Succ(A, Out(s))) \times u_i(Succ(A, Out(s))) \\ & + Pr(\neg Succ(A), Out(s)) \times u_i(\neg Succ(A, Out(s))) \end{aligned} \tag{1}$$

- $Pr(Succ(A, Out(s)))$: the probability of success of the initial argument A w.r.t. the dialogue $Out(s)$
- $u_i(Succ(A, Out(s)))$ is the utility value of the success of A w.r.t. the dialogue $Out(s)$.

Probability of success of an argument

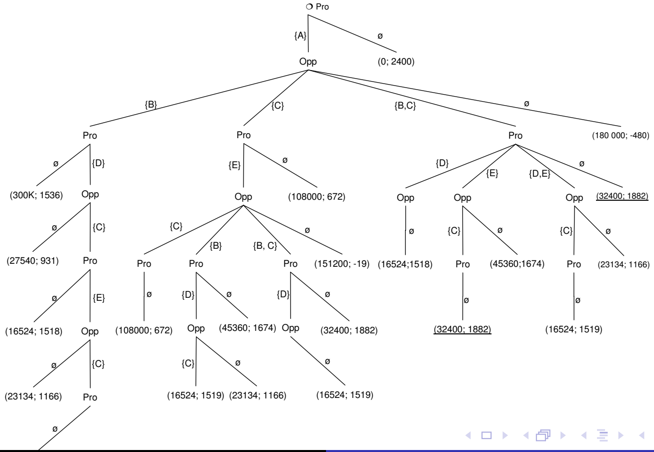
The probability of success of an argument is intended to mean the probability that the argument is accepted as justified given a knowledge base of which the statements are assigned a probability of acceptance by the adjudicator.

Utility values

The utility values $u_i(Succ(A, Out(s)))$ and $u_i(\neg Succ(A, Out(s)))$ incorporate costs and benefits of moves.
We distinguish:

1. Fixed costs/benefits capture costs/benefits independent of the success of the player (e.g. trial expenses).
2. Costs/benefits of moves dependant upon success.

Dialogue games as extensive games



Conclusion

- An interpretation of a dialectical setting in game-theorical terms.
- A specification of preferences over outcomes has been provided in terms of expected utility combining the probability of success of arguments, costs and benefits of arguments.