Multiagent Systems (BE4M36MAS)

(Computational) Social Choice

Branislav Bošanský

Artificial Intelligence Center,
Department of Computer Science,
Faculty of Electrical Engineering,
Czech Technical University in Prague

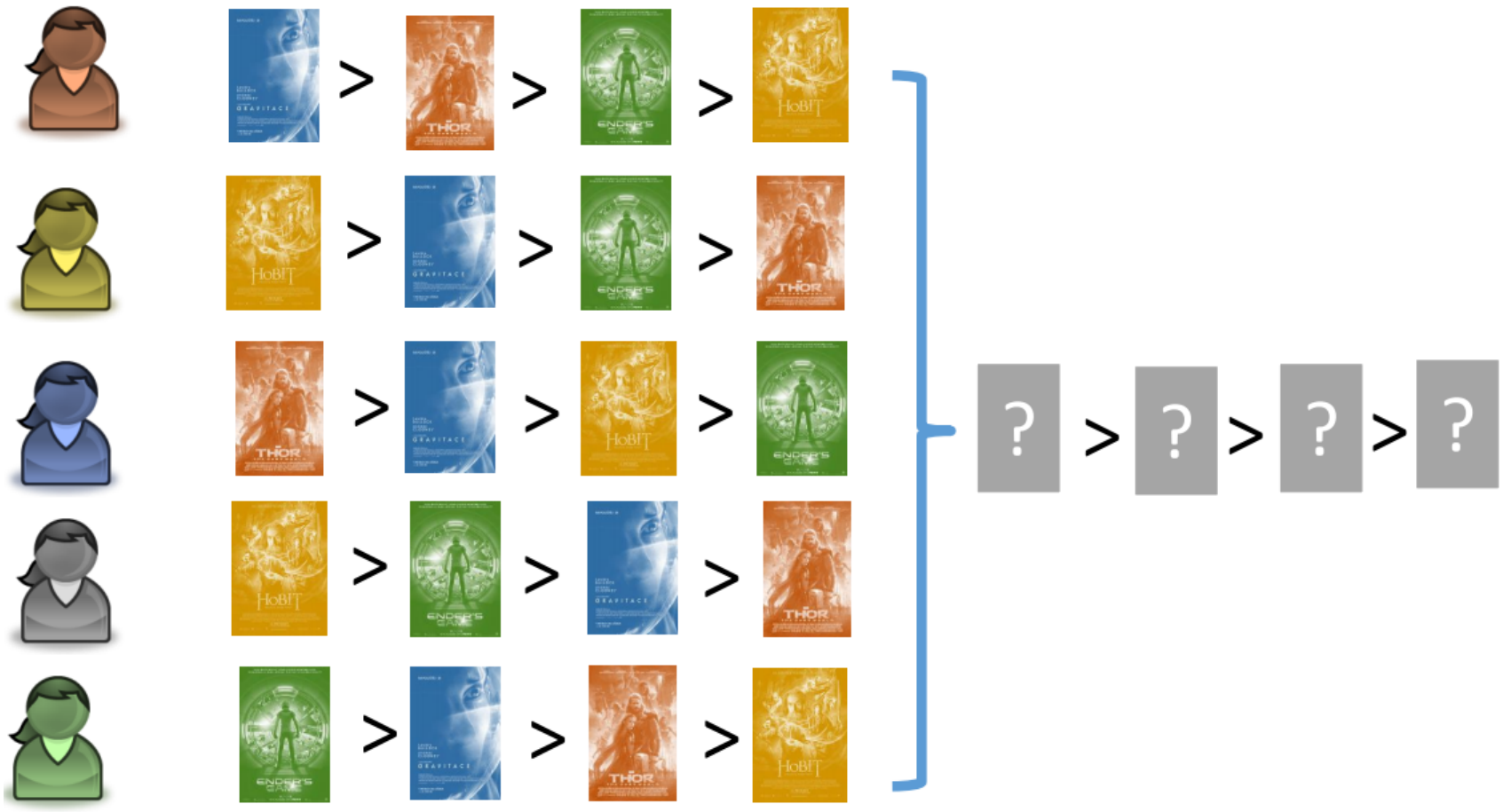
branislav.bosansky@agents.fel.cvut.cz

December 10, 2019

Previously ... on multi-agent systems.

1. Agent Architectures
2. Non-cooperative Game Theory
3. Distributed Constraint Satisfaction/Optimization
4. Cooperative/Coalitional Game Theory

Social Choice and Motivational Example



Social Choice and Applications

- Elections

- Joint plans for cooperating agents
- Resource allocation
- Recommendation and reputation systems
- Preference aggregation
- Human computation (crowdsourcing)

- Webpage ranking and meta-search engines

Social Welfare Function

Consider:

- a finite set $N = \{1, \ldots, n\}$ of at least two agents (a.k.a individuals or voters),
- a finite universe U of at least two alternatives (candidates),
- each agent i has preferences over the alternatives in U, which are represented by a transitive and complete preference relation $\succeq_i$
- the set of all preference relations over the universal set of alternatives U is denoted as $\mathcal{R}(U)$.
- the set of preference profiles, associating one preference relation with each individual agent is then given by $\mathcal{R}(U)^n$

Definition

A *social welfare function (SWF)* is a function $f : \mathcal{R}(U)^n \to \mathcal{R}(U)$.

Social Welfare Function - properties

SWF must satisfy two basic properties:

- transitivity: $a \succeq_i b \succeq_i c$ implies $a \succeq_i c$
- completeness: for any pair of alternatives $a, b \in U$ either $a \succeq_i b$, or $b \succeq_i a$ (or both if indifference is allowed, $a \sim_i b$)

Different SWFs satisfy different additional properties

Definition (Pareto optimality)

A social welfare function f is *Pareto optimal* if $a \succ_i b$ for all $i \in N$ implies that $a \succ_f b$.

Social Welfare Function - properties

An SWF satisfies *independence of irrelevant alternatives (IIA)* if the social preference between any pair of alternatives only depends on the individual preferences restricted to these two alternatives.

Definition (IIA)

Let R and R' be two preference profiles and a and b be two alternatives such that $R|_{\{a,b\}} = R'|_{\{a,b\}}$ i.e., the pairwise comparisons between a and b are identical in both profiles. Then, IIA requires that a and b are also ranked identically in $\succ_f |_{\{a,b\}} = \succ'_f |_{\{a,b\}}$.

Independence of Irrelevant Alternatives Example

Consider a plurality voting system (the candidate that is voted as the top one by a majority wins) of 7 voters.

There are 2 alternatives – (A, B):

- 3 voters rank $A \succ B$
- 4 voters rank $B \succ A$

B should win based under the plurality voting rule

A new alternative C is introduced – (A, B, C):

- 3 voters rank $A \succ B \succ C$
- 2 voters rank $B \succ A \succ C$
- 2 voters rank $C \succ B \succ A$

A now wins under the plurality voting rule

Introducing an "irrelevant" alternative affects the outcome between A and B.

Social Welfare Function - properties

Definition (Non-dictatorship)

An SWF is *non-dictatorial* if there is no agent who can dictate a strict ranking no matter which preferences the other agents have. Formally, an SWF is non-dictatorial if there is no agent i such that for all preference profiles $\mathcal{R}$ and alternatives a, b, $a \succ_i b$ implies that $a \succ_f b$.

There is no agent who can dictate a strict ranking no matter which preferences the other agents have.

Arrow's Theorem

Theorem (Arrow, 1951)

There exists no SWF that simultaneously satisfies IIA, Pareto-optimality, and non-dictatorship whenever $|U| \geq 3$.

Negative result: At least one of the desired properties has to be omitted or relaxed in order obtain a positive result.

If $|U| = 2$, IIA is trivially satisfied by any SWF and reasonable SWFs (e.g. the majority rule) also satisfy remaining conditions.

Social Choice Functions

In many applications, a full social preference relation is not needed; rather, we just wish to identify the socially most desirable alternatives.

Definition (Social Choice Function)

A *social choice function (SCF)* is a function $f : \mathcal{R}(U)^n \times \mathcal{F}(U) \to \mathcal{F}(U)$ such that $f(R, A) \subseteq A$ for all R and A.

where $\mathcal{F}(U)$ is the set of all non-empty subsets of U.

Social Choice Functions

Arrows Theorem can be reformulated for SCFs by appropriately redefining Pareto-optimality, IIA, and non-dictatorship and introducing a new property called the *weak axiom of revealed preference*.

- *Pareto optimality:* $a \notin f(R, A)$ if there exists some $b \in A$ such that $b \succ_i a$ for all $i \in N$
- *Non-dictatorship:* an SCF is non-dictatorial iff there is no agent i such that for all preference profiles R and alternatives a such that $a \succ_i b$ for all $b \in A \setminus \{a\}$ implies $a \in f(R, A)$
- *IIA:* an SCF satisfies IIA iff $f(R, A) = f(R', A)$ if $R|_A = R'|_A$

Weak Axiom of Revealed Preference

Definition (Weak Axiom of Revealed Preference (WARP))

An SCF f satisfies WARP iff for all feasible sets A and B and preference profiles R:

$$\text{if } B \subseteq A \text{ and } f(R, A) \cap B \neq \emptyset \text{ then } f(R, A) \cap B = f(R, B).$$

WARP requires that the choice set of B consists precisely of those alternatives in B that are also chosen in A, whenever this set is non-empty.

Theorem (Arrow)

There exists no social choice function that simultaneously satisfies IIA, Pareto optimality, non-dictatorship, and WARP whenever $|U| \geq 3$.

Voting Rule

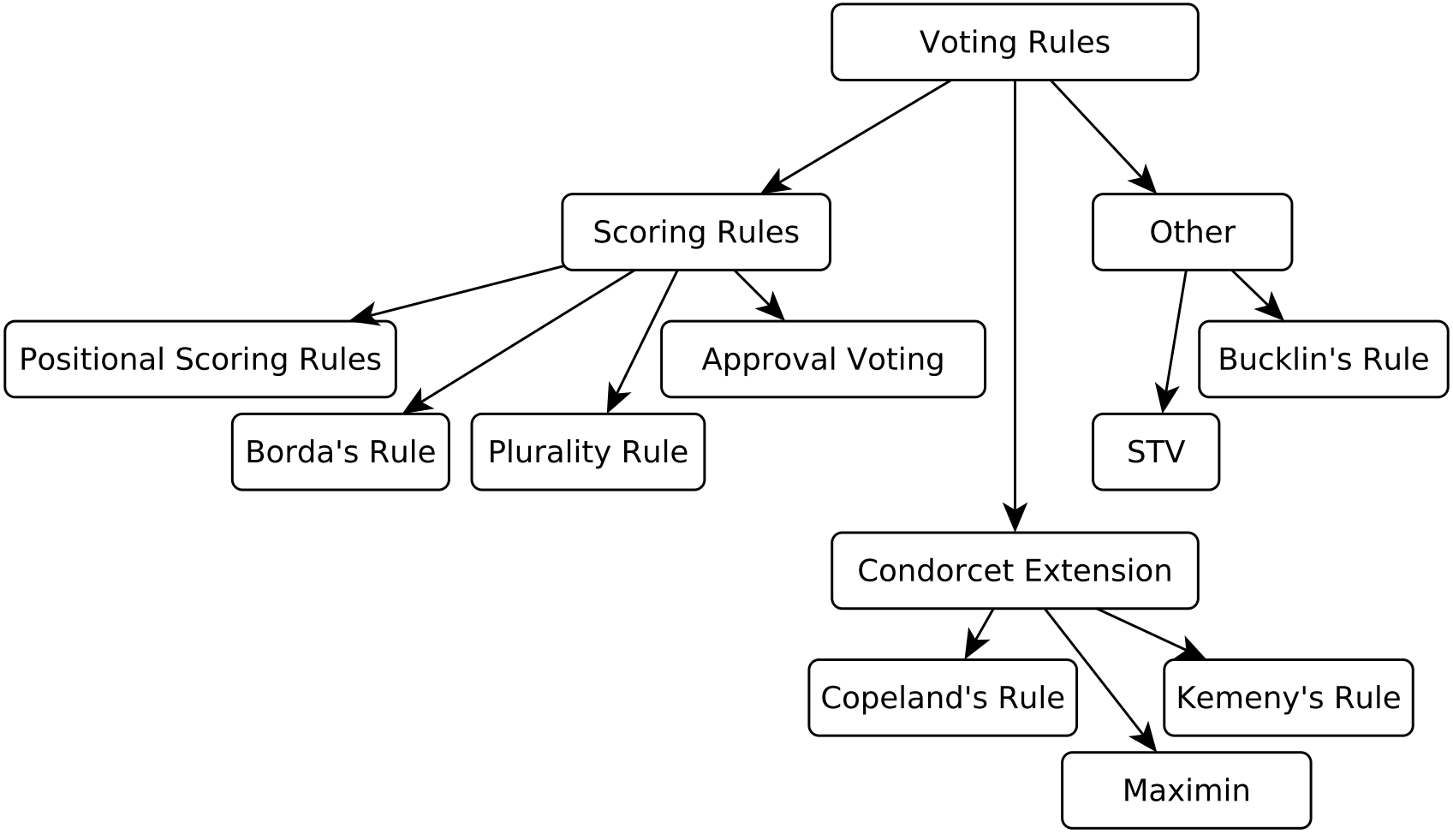
Voting rules are specific social choice functions

Definition (Voting Rule)

A voting rule is a function $f : \mathcal{R}(U)^n \to \mathcal{F}(U)$.

A voting rule f is *resolute* if $|f(R)| = 1$ for all preference profiles R.

List of Voting Rules



Voting Rules

Positional Scoring Rules Assuming $m = |U|$ alternatives, we define a score vector $s = (s_1, \ldots, s_m) \in \mathbb{R}^m$ such that $s_1 \geq \ldots \geq s_m$ and $s_1 > s_m$. Each time an alternative is ranked i-th by some voter, it gets a particular score s_i.

The scores of each alternative are summed and the alternative with the highest cumulative score is selected.

- **Borda's rule** the score vector is $s = (m-1, m-2, \ldots, 0)$
- **Plurality rule** the score vector is $s = (1, 0, \ldots, 0)$
- **Anti-plurality rule / Approval voting** the score vector is $s = (1, 1, \ldots, 1, 0)$ (or a subset of alternatives).

Condorcet Winner and Extension

An alternative a is a **Condorcet winner** if, when compared with every other candidate, is preferred by more voters (or is the winner in every pairwise comparison).

- Condorcet winner is unique but does not always exist
- Condorcet extension: a voting rule that selects Condorcet winner whenever it exists.

Question

Do the positional scoring rules satisfy the Condorcet extension?

Condorcet Winner and Extension

Question

Do the positional scoring rules satisfy the Condorcet extension?

No.

B is the Condorcet winner, but plurality rule selects D

- $2 : A \succ B \succ D \succ C$
- $2 : C \succ B \succ A \succ D$
- $3 : D \succ B \succ C \succ A$

B is the Condorcet winner, but approval and Borda's rule select A

- $2 : A \succ B \succ C \succ D$
- $2 : C \succ B \succ A \succ D$
- $2 : D \succ B \succ A \succ C$
- $1 : A \succ C \succ D \succ B$

Condorcet Winner and Extension

Rules that satisfy Condorcet extension:

- **Copelands rule**: an alternative gets a point for every pairwise majority win, and some fixed number of points between 0 and 1 (say, 1/2) for every pairwise tie. The winners are the alternatives with the highest number of points.
- **Maximin rule**: evaluate every alternative by its worst pairwise defeat by another alternative; the winners are those who lose by the lowest margin in their worst pairwise defeats. (If there are any alternatives that have no pairwise defeats, then they win.)

Kemeny's Rule

$$\arg\max_{\succ} \sum_{i \in N} |\succ \cap \succ_i|$$

i.e. all strict rankings that agree with as many pairwise preferences as possible.

Maximum likelihood interpretation: agents provide noisy estimates of a "correct" ranking

Computation is NP-hard for 4 or more voters.

Other Voting Rules

Single transferable vote (STV): looks for the alternatives that are ranked in first place the least often, removes them from all voters' ballots, and repeats. The alternatives removed in the last round win.

Pairwise elimination: pairs the candidates according to some ordering (agenda), the loser of the pairwise election drops out; repeat.

Strategic Manipulation

So far, we assumed that the true preferences of all voters are known.

Voters may be better off by misrepresenting their preferences.

- 1 voter ranks $A \succ B \succ C \succ D$
- 2 voters rank $A \succ C \succ B \succ D$
- 2 voters rank $B \succ D \succ C \succ A$
- 2 voters rank $C \succ B \succ D \succ A$

Plurality winner A ... but B can be the winner if the last two voters vote for B instead of C.

Borda's winner B ... but C wins if the voters in the second row, who prefer C to B move B to the bottom.

Strategic Manipulation

Definition

A resolute voting rule f is *manipulable* by voter i if there exist preference profiles R and R' such that $R_j = R'_j$ for all $j \neq i$ and $f(R) \succ_i f(R')$. A voting rule is *strategyproof* if it is not manipulable.

Negative Aspects of Strategic Manipulation

Inefficient: Energy and resources are wasted on manipulative activities.

Unfair: Manipulative skills are not spread evenly across the population.

Erratic: Predictions or theoretical statements about election outcomes become extremely difficult.

Question

Are there any voting methods which are non-manipulable, in the sense that voters can never benefit from misrepresenting preferences?

The Gibbard-Satterthwaite Impossibility Theorem

Theorem (Gibbard-Satterthwaite)

Every non-imposing, strategyproof, resolute voting rule is dictatorial when $|U| \geq 3$.

A voting rule is **non-imposing** if its image contains all singletons of $\mathcal{F}(U)$, i.e., every single alternative is returned for some preference profile.

Computational Hardness of Manipulation

Gibbard-Satterthwaite tells us that manipulation is possible in principle but does not give any indication of how to misrepresent preferences.

There are voting rules that are prone to manipulation in principle, but where manipulation is computationally complex (e.g. Single Transferable Vote rule is NP-hard to manipulate).

Problem: NP-hardness is a worst-case measure.

Recent negative result (Isaksson et al., 2010): Essentially, for every efficiently computable, neutral voting rule, a manipulable preference profile with a corresponding manipulation can easily be found.