Binary Aggregation by Selection of the Most Representative Voter

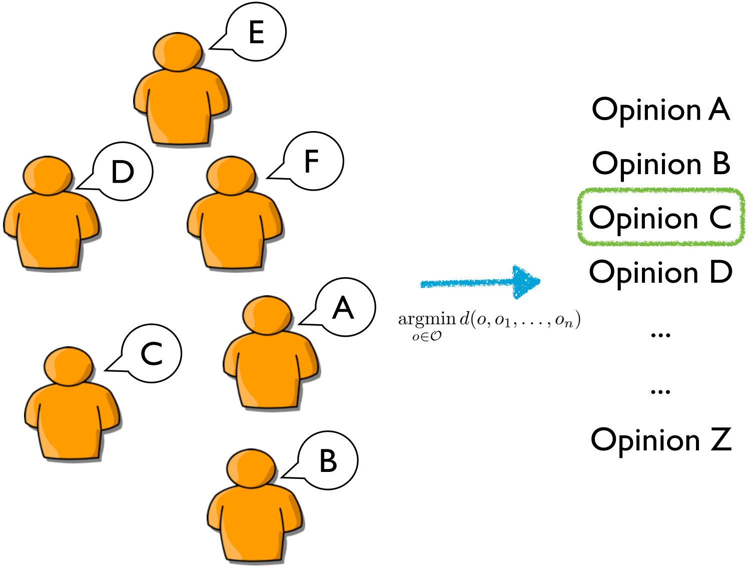
Umberto Grandi

Department of Mathematics
University of Padova

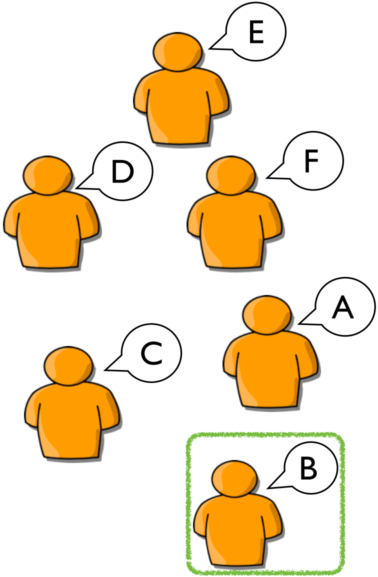
4 August 2013

[Joint work with Ulle Endriss]

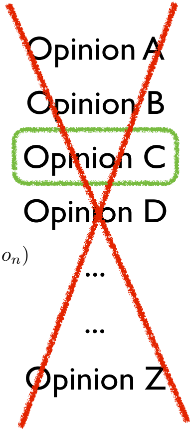
Selection of the Closest Opinion



Selection of the Most Representative Voter



$$\underset{\{o_i \mid i \in \mathcal{N}\}}{\operatorname{argmin}}\ d(o_i, o_1, \ldots, o_n)$$



Outline

1. A general framework for aggregation problems:
   - Binary aggregation with integrity constraints
   - Preferences, judgments, multi-issue elections...
2. Distance-based rules:
   - Paradoxical profiles
   - Kemeny rule
   - Slater rule
3. Selection of the most representative voter:
   - Low computational complexity
   - Axiomatic properties
   - Approximation results

Binary Aggregation

Ingredients:

- A finite set $\mathcal{N}$ of individuals
- A finite set $\mathcal{I} = \{1, \ldots, m\}$ of issues
- A boolean combinatorial domain: $\mathcal{D} = \{0, 1\}^{\mathcal{I}}$

Definition

An aggregation procedure is a function $F : \mathcal{D}^{\mathcal{N}} \to \mathcal{D}$ mapping each profile of ballots $\boldsymbol{B} = (B_1, \ldots, B_n)$ to an element of the domain $\mathcal{D}$.

Dokow and Holzman (JET 2010), Grandi and Endriss (AIJ-2013)

Binary Aggregation

Ingredients:

- A finite set $\mathcal{N}$ of individuals
- A finite set $\mathcal{I} = \{1, \ldots, m\}$ of issues
- A boolean combinatorial domain: $\mathcal{D} = \{0,1\}^{\mathcal{I}}$

Definition

An aggregation procedure is a function $F : \mathcal{D}^{\mathcal{N}} \to \mathcal{D}$ mapping each profile of ballots $\boldsymbol{B} = (B_1, \ldots, B_n)$ to an element of the domain $\mathcal{D}$.

Example: Three agents with sensors

- $\mathcal{N} = \{1,2,3\}$
- $\mathcal{I} = \{T_1, T_2, A\}$
- Individuals submit ballots in $\mathcal{D} = \{0,1\}^3$

$B_1 = (0,1,0)$ the first individual think the action should not be performed.

Integrity Constraints

A propositional language $\mathcal{L}$ to define the subset of rational ballots in $\{0,1\}^{\mathcal{I}}$:

- One propositional symbol for every issue: $PS = \{p_1, \ldots, p_m\}$
- $\mathcal{L}_{PS}$ closed under connectives $\wedge$, $\vee$,$\neg$, $\rightarrow$ the set of atoms PS

Given an integrity constraint $\text{IC} \in \mathcal{L}_{PS}$, a rational ballot is $B \in \text{Mod}(\text{IC})$

Integrity Constraints

A propositional language $\mathcal{L}$ to define the subset of rational ballots in $\{0,1\}^{\mathcal{I}}$:

- One propositional symbol for every issue: $PS = \{p_1, \ldots, p_m\}$
- $\mathcal{L}_{PS}$ closed under connectives $\wedge$, $\vee$,$\neg$, $\rightarrow$ the set of atoms PS

Given an integrity constraint IC $\in \mathcal{L}_{PS}$, a rational ballot is $B \in \mathrm{Mod}(\mathrm{IC})$

Example: Three agents with sensors

Perform action A if both parameters exceed the thresholds.
Propositional constraint: $\mathrm{IC} = (p_{T_1} \wedge p_{T_2}) \rightarrow p_A$

Individual 1 submits $B_1 = (1,1,1)$: B_1 satisfies IC ✓
Individual 2 submits $B_2 = (0,1,0)$: $B_2 \models \mathrm{IC}$ ✓
Individual 3 submits $B_3 = (1,0,0)$: $B_3 \models \mathrm{IC}$ ✓

Majority aggregation outputs $(1,1,0)$: IC not satisfied.

Preference Aggregation as Binary Aggregation

Agent 1	$A > B > C$
Agent 2	$B > C > A$
Agent 3	$C > A > B$
Maj	$A > B > C > A$!!

Condorcet Paradox (1785)

Preferences as binary ballots + integrity constraint

	$A > B$	$B > C$	$A > C$
Agent 1	1	1	1
Agent 2	0	1	0
Agent 3	1	0	0
Maj	1	1	0

Judgment Aggregation as Binary Aggregation

Agent 1	$\{\alpha, \beta, \alpha \wedge \beta\}$
Agent 2	$\{\neg\alpha, \beta, \neg(\alpha \wedge \beta)\}$
Agent 3	$\{\alpha, \neg\beta, \neg(\alpha \wedge \beta)\}$
Maj	$\{\alpha, \beta, \neg(\alpha \wedge \beta)\}$

Doctrinal Paradox

Kornhauser and Sager (1986)

Judgments as binary ballots + integrity constraint

$$\text{IC} = \neg(p_\alpha \wedge p_\beta \wedge p_{\neg(\alpha\wedge\beta)})$$

	p_α	p_β	$p_{\alpha\wedge\beta}$
Agent 1	1	1	1
Agent 2	0	1	0
Agent 3	1	0	0
Maj	1	1	0

Distance-Based Rules

How to avoid paradoxes?

- Only consider outcomes that respect the integrity constraint.
- Pick the one closest to the individual inputs!

Depending on the distance chosen there may be several distance-based rules:

Definition

The DBR (aka Kemeny rule, Prototype) picks the consistent ballot minimising the sum of the Hamming distances to the individual ballots. Θ_2^p-complete!!

Definition

The Slater rule (aka Endpoint) picks the consistent ballot minimising the Hamming distance to the outcome of the majority rule. NP-hard (at least)!!

The Hamming distance H between an individual input and the outcome is the number of issues on which they differ.

Selection of the Most Representative Voter

Basic idea:
Restrict the search space to $\textsc{Supp}(\boldsymbol{B}) = \{B_1, \ldots, B_n\}$

Definition

The average-voter rule is the aggregation rule that selects those individual ballots that minimise the Hamming distance to the profile:

$$\mathrm{AVR}(\boldsymbol{B}) = \operatorname*{argmin}_{B \in \textsc{Supp}(\boldsymbol{B})} \sum_{i \in \mathcal{N}} H(B, B_i)$$

Definition

The majority-voter rule is the aggregation rule that selects those individual ballots that minimise the Hamming distance to one of the majority outcomes:

$$\mathrm{MVR}(\boldsymbol{B}) = \operatorname*{argmin}_{B \in \textsc{Supp}(\boldsymbol{B})} \min\{H(B, B') \mid B' \in \mathrm{Maj}(\boldsymbol{B})\}$$

An Example

The AVR and the MVR can give radically different results:

Issue:	**1**	**2**	**3**	**4**	**5**
1 voter:	0	1	1	1	1
2 voters:	1	0	0	0	0
10 voters:	0	1	1	0	0
10 voters:	0	0	0	1	1
Maj:	0	0	0	0	0
MVR:	1	0	0	0	0
AVR:	0	1	1	0	0
AVR:	0	0	0	1	1

Hamming distance of AVR from the profile: 48
Hamming distance of MVR from the profile: 65

Computational Complexity

Recall that m is the number of issues; n is the number of voters.

Winner determination for the AVR is in $O(mn \log n)$

- compute the vector of sums in $O(mn)$
- compute the difference between each ballot (multiplied by n) to the vector of sums in $O(mn \log n)$ [O(log n) because of integers up to n]

Winner determination for the MVR is in $O(mn)$

- compute the majority vector in $O(mn)$
- compare each ballot to the majority vector in $O(mn)$

Conclusion? Both rules are easy to compute (MVR is easier)

Axiomatic Properties

Rules based on the most representative voter satisfy interesting properties:

- No paradoxes ever, whatever the IC (no other rule has this property)
- Unanimity guaranteed (obvious)
- Neutrality guaranteed (less obvious)

F satisfies reinforcement if for any two profiles $\boldsymbol{B}$ and $\boldsymbol{B}'$ such that:

- $\textsc{Supp}(\boldsymbol{B}) = \textsc{Supp}(\boldsymbol{B}')$
- $F(\boldsymbol{B}) \cap F(\boldsymbol{B}') \neq \emptyset$

we have that $F(\boldsymbol{B} \oplus \boldsymbol{B}') = F(\boldsymbol{B}) \cap F(\boldsymbol{B}')$

If two groups independently agree that a certain outcome is best, we would expect them to uphold this choice when choosing together.

Theorem

The AVR satisfies reinforcement, but the MVR does not.

Approximation Results

F is said to be an α-approximation of F' if for every profile $\boldsymbol{B}$:

$$d(F(\boldsymbol{B}), \boldsymbol{B}) \leqslant \alpha \cdot d(F'(\boldsymbol{B}), \boldsymbol{B})$$

Good in case F' is intractable (like distance-based rules) and α is a constant.

Theorem

Both the AVR and the MVR are 2-approximations of the DBR (for any IC).

Very likely that α decreases if we increase the logical complexity of IC.

This result extends to preference aggregation: cf. approximations of the Kemeny rule (better approximation ratio but rather "technical" rules) by Dwork et al. 2001, Kenyon-Mathieu and Schudy, 2007.

Conclusions

Binary aggregation with integrity constraints is a general framework for the study of aggregation problems such as preference and judgment aggregation: We have seen that it suggests novel simple procedures to be used in practice!

Restrict minimisation of a distance-based rule to the individual ballots received:

- Outcome will never be paradoxical
- Very low complexity
- Social-choice theoretic properties (not independence!)
- 2-approximation of distance-based rule (aka Kemeny)

Future work:

- Tideman's ranked-pairs rule as a distance-based rule
- Better approximation results on restricted domains