Binary Aggregation by Selection of the Most Representative Voter

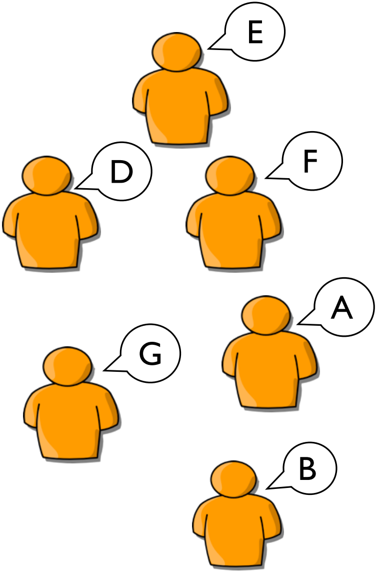
Umberto Grandi

Department of Mathematics
University of Padova

21 November 2013

[Joint work with Ulle Endriss]

Selection of the Closest Opinion



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

Opinion A

Opinion B

Opinion C

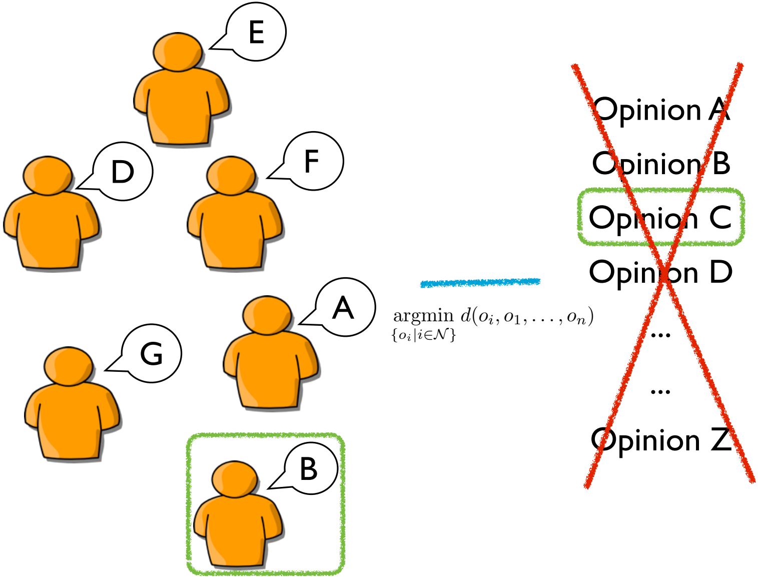
Opinion D

...

...

Opinion Z

Selection of the Most Representative Voter



Outline

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

Everything Starts From Paradoxical Situations

Suppose three agents in a multi-agent system need to decide whether to perform a collective decision A. The decision is performed if two parameters T_1 and T_2 exceed a given threshold. Consider the following situation:

	T_1	T_2	A
Agent 1	Yes	Yes	Yes
Agent 2	No	Yes	No
Agent 3	Yes	No	No
Majority	Yes	Yes	No

Should the agents perform action A or not?

- A majority of agents think the first parameter exceeds the threshold.
- A majority of agents think the second parameter exceeds the threshold.
- But: a majority of agents think action A should not be performed!!

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

Wilson (1975), 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: $PS = \{p_1, \dots, p_m\}$
- $\mathcal{L}_{PS}$ 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: IC $= (p_{T_1} \wedge p_{T_2}) \rightarrow p_A$

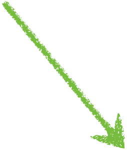
Individual 1 submits $B_1 = (1,1,1)$:	B_1 satisfies IC ✓
Individual 2 submits $B_2 = (0,1,0)$:	$B_2 \models$ IC ✓
Individual 3 submits $B_3 = (1,0,0)$:	$B_3 \models$ IC ✓

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

The Doctrinal Paradox

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

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

	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

Collective Rationality

Definition

F is collectively rational (CR) for IC $\in \mathcal{L}_{PS}$ if for all profiles $\boldsymbol{B}$ such that $\boldsymbol{B}_i \models$ IC for all $i \in N$ then $F(\boldsymbol{B}) \models$ IC.

F lifts the rationality assumption given by IC
from the individual to the collective level.

If you want to know more about collective rationality:

Grandi and Endriss, Lifting Integrity Constraints in Binary Aggregation. AIJ, 2013.

Avoid paradoxes? Generalised Dictatorship

A generalised dictatorship copies the ballot of a (possibly different) individual (aka local dictatorships, positional dictatorships, rolling dictatorships):

Proposition

F is collectively rational with respect to all IC in $\mathcal{L}_{PS}$ if and only if F is a generalised dictatorship

This class includes:

- Classical dictatorships $F(B_1, \ldots, B_n) = B_i$ for $i \in \mathcal{N}$
- Rules based on the selection of the most representative voter

Distance-Based Rules

Another route has been taken in the literature to avoid paradoxes:

- Given an integrity constraint, consider $\mathrm{Mod}(\mathsf{IC})$
- Define a distance between a profile and a ballot
- Pick the ballot in $\mathrm{Mod}(\mathsf{IC})$ which is closest to the individual inputs!

Examples of distance-based rules have been defined in preference aggregation and more recently imported in the literature on judgment aggregation

Konieczny and Pino Pérez, Merging information under constraints... JLC, 2002.
Pigozzi. Belief merging and the discursive dilemma... Synthese, 2006
Lang Et Al. Judgment Aggregation Rules based on Minimisation. TARK-2011.

Kemeny and Slater Rule

In preference aggregation we have a number of individuals submitting preferences in the form of a linear order.

Definition

The Kemeny rule picks the linear orders which minimises the sum of the Kendall τ distance to the individual preferences. Θ_2^p*-complete*

Definition

The Slater rule picks the linear orders minimising the Kendall τ distance to the outcome of the majority rule. NP-hard (at least)

The Kendall τ distance counts the number of pairwise disagreements between two linear orders.

Hemaspaandra et Al. The complexity of Kemeny elections. Theoretical Computer Science, 2005

Brandt et Al. The Computational Complexity of Choice Sets. Mathematical Logic Quarterly, 2009.

DBR and Slater Rule

The Kemeny and Slater rule comes under different names in binary aggregation:

Definition

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

Definition

The Slater rule (aka Endpoint) picks the consistent ballots 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:

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

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:

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

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

Observation

$\mathcal{H}(\mathrm{AVR}(\boldsymbol{B}), \boldsymbol{B}) \leqslant \mathcal{H}(\mathrm{MVR}(\boldsymbol{B}), \boldsymbol{B})$ *where* $\mathcal{H}(B, \boldsymbol{B}) = \sum_i H(B, B_i)$

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 paradox 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 the best, we would expect them to uphold this choice when choosing together.

Theorem

The AVR satisfies reinforcement, but the MVR does not.

Approximation Results

Can we compare the outcome of AVR and MVR with that of Kemeny rule?

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.

Short and Long Proof

A short direct proof can be obtained using the triangular inequality:

$$
\begin{aligned}
\mathcal{H}(\mathrm{AVR}(\boldsymbol{B}), \boldsymbol{B}) &= \sum_{i=1}^{n} H(\mathrm{AVR}(\boldsymbol{B}), B_i) \leqslant \sum_{i=1}^{n} \frac{1}{n} \cdot \sum_{k=1}^{n} H(B_k, B_i) \\
&\leqslant \sum_{i=1}^{n} \frac{1}{n} \cdot \sum_{k=1}^{n} [H(B_k, \mathrm{DBR}(\boldsymbol{B})) + H(\mathrm{DBR}(\boldsymbol{B}, B_i)] \\
&= \frac{1}{n} \cdot [n \cdot \mathcal{H}(\mathrm{DBR}(\boldsymbol{B}), \boldsymbol{B}) + n \cdot \mathcal{H}(\mathrm{DBR}(\boldsymbol{B}), \boldsymbol{B}))] \\
&= 2 \cdot \mathcal{H}(\mathrm{DBR}(\boldsymbol{B}), \boldsymbol{B})
\end{aligned}
$$

But longer proofs lead us to more precise bounds:

Theorem

If $n > m$, then the AVR and the MVR are α-approximations of the $\mathrm{DBR}^{\mathrm{IC}}$ with $\alpha = 2 \cdot \frac{m-1}{m}$ for any integrity constraint IC.

Preference Aggregation (future work)

Known approximation results for the Kemeny rule in preference aggregation:

- 2-approximation by Dwork et Al. 2001
- $\frac{11}{7}$-approximation by Ailon et Al. 2008
- PTAS by Kenyon-Mathieu and Schudy, 2007

Better approximation ratio but rather "technical" rules!

Question 1

Can we obtain smaller approximation bounds for AVR and MVR if we restrict to the domain of preferences?

Question 2

To the best of our knowledge AVR and MVR have not been studied as preference aggregation rules. Do they have interesting additional properties?

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!

Rules based on the selection of the most representative voter:

- 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