Binary Aggregation by
Selection of the Most Representative Voter

Umberto Grandi

Department of Mathematics
University of Padova

14 May 2014

[Joint work with Ulle Endriss]
Selection of the Closest Opinion

\[
\text{argmin}_{\{o_i \mid i \in \mathcal{N}\}} \, d(o_i, o_1, \ldots, o_n)
\]
Selection of the Most Representative Voter

\[ \text{argmin} \left\{ o_i \mid i \in \mathcal{N} \right\} \]
\[ d(o_i, o_1, \ldots, o_n) \]
1. A general framework for aggregation problems:
   - Binary aggregation with integrity constraints
   - Preferences, judgments, multi-issue elections...
   - Generalised dictatorships

2. Selection of the most representative voter:
   - Average voter rule (AVR)
   - Majority voter rule (MVR)
   - Ranked voter rule (RVR)

3. Properties of most-representative-voter rules:
   - Approximation results
   - Computational complexity
   - Axiomatic properties
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 $B = (B_1, \ldots, B_n)$ to an element of the domain $\mathcal{D}$.

Wilson (1975), Dokow and Holzman (JET 2010), Grandi and Endriss (AIJ 2013)
A propositional language $\mathcal{L}$ to define the subset of rational ballots in $\{0, 1\}^I$:

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

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

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

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

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

**Example: Three agents with sensors**

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

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 \( C > A > B \)

\[ \text{Maj} \quad A > B > C > A !! \]

Condorcet Paradox (1785)

Preferences as binary ballots + integrity constraint

\[
\begin{array}{ccc}
A > B & B > C & A > C \\
\text{Agent 1} & 1 & 1 & 1 \\
\text{Agent 2} & 0 & 1 & 0 \\
\text{Agent 3} & 1 & 0 & 0 \\
\text{Maj} & 1 & 1 & 0 \\
\end{array}
\]
The Discursive Dilemma

<table>
<thead>
<tr>
<th>Agent</th>
<th>Propositions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>${\alpha, \beta, \alpha \land \beta}$</td>
</tr>
<tr>
<td>2</td>
<td>${-\alpha, \beta, -(\alpha \land \beta)}$</td>
</tr>
<tr>
<td>3</td>
<td>${\alpha, -\beta, -(\alpha \land \beta)}$</td>
</tr>
<tr>
<td>Maj</td>
<td>${\alpha, \beta, -(\alpha \land \beta)}$</td>
</tr>
</tbody>
</table>

Judgments as binary ballots + integrity constraint

$$IC = \neg(p_\alpha \land p_\beta \land p_{-(\alpha \land \beta)})$$

<table>
<thead>
<tr>
<th></th>
<th>$p_\alpha$</th>
<th>$p_\beta$</th>
<th>$p_{\alpha \land \beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Agent 2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Agent 3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maj</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Proposition - Majority rule

The majority rule does not generate a paradox with respect to IC if and only if IC is equivalent to a conjunction of clauses of size $\leq 2$ (i.e., 2-CNF)
Avoid paradoxes? Characterisation results and generalised dictatorship

Proposition - Majority rule

The majority rule does not generate a paradox with respect to IC if and only if IC is equivalent to a conjunction of clauses of size $\leq 2$ (i.e., 2-CNF).

How to avoid all paradoxes?

Proposition - Avoid all paradoxes

An aggregation procedure does not generate a paradox with respect to all IC if and only if it copies the ballot of a (possibly different) individual in every profile.

Grandi and Endriss, Lifting Integrity Constraints in Binary Aggregation. AIJ, 2013.
Distance-based rules in judgment aggregation

**Definition - Distance-based rule**

The DBR (aka Kemeny rule, Prototype) picks the consistent ballots minimising the sum of the Hamming distances to the individual ballots. $\Theta^p_2$-complete

**Definition - Slater rule**

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

**Definition - Ranked agenda**

The ranked-agenda rule picks the consistent ballots obtained by sequential majority following the order given by the strength of acceptance. $\Delta^p_2$-hard.

Lang and Slavkovijk, ECAI-2014.
Selection of the Most Representative Voter

Restrict the search space to $\text{Supp}(B) = \{B_1, \ldots, B_n\}$

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

$$\text{AVR}(B) = \arg\min_{B \in \text{Supp}(B)} \sum_{i \in N} H(B, B_i)$$

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:

$$\text{MVR}(B) = \arg\min_{B \in \text{Supp}(B)} \min \{H(B, B') \mid B' \in \text{Maj}(B)\}$$

The RVR is defined in a similar way...
An Example

The AVR, the MVR and the majority rule can give radically different results:

<table>
<thead>
<tr>
<th>Issue:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 voter:</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10 voters:</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10 voters:</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Maj: 0 0 0 0 0 0
MVR: 1 0 0 0 0 0
AVR: 0 1 1 0 0 0

Hamming distance of AVR from the profile: 53
Hamming distance of MVR from the profile: 70

Observation

\[ \mathcal{H}(\text{AVR}(B), B) \leq \mathcal{H}(\text{MVR}(B), B) \] where \( \mathcal{H}(B, B) = \sum_i H(B, B_i) \)
Can we compare the outcome of the ideal rule – DBR aka Kemeny – with that of our AVR, MVR, RVR?

**Definition - Approximation**

*F* is said to be an $\alpha$-approximation of $\text{DBR}^{IC}$ if for every profile $B$:

$$\mathcal{H}(F(B), B) \leq \alpha \cdot \mathcal{H}(\text{DBR}^{IC}(B), B)$$

Good approximation if $\alpha$ is a constant.
Preliminary facts about the distance-based procedure

The definition of the distance-based rule depends on the constraint:

$$\text{DBR}^{\text{IC}}(B) = \arg\min_{B \in \text{Mod}^{\text{IC}}} \sum_{i \in N} H(B, B_i)$$

In particular, $\text{DBR}^\top = \text{Maj}$ (the majority rule). With stronger constraints?

Lemma

If IC entails IC', then $H(\text{DBR}^{\text{IC}}(B), B) \geq H(\text{DBR}^{\text{IC}'}(B), B)$ for every profile $B \in \text{Mod}^{\text{IC}}$$^n$.

And a baseline result:

Proposition

Every rule based on the most representative voter is an $O(n)$-approximation of the DBR^{IC}.
Recall that $n$ is the number of individuals, $m$ is the number of issues.

**Theorem**

*The RVR is a $\Theta(n)$-approximation of Maj* (even if $m$ is bounded).

*Proof.* The upper bound is given by the result on the previous slide. The lower bound is obtained by showing a family of profiles where the result of the RVR is $n$-far from that of Maj.
Positive Results

**Theorem**

The AVR and the MVR are strict 2-approximations of the DBR

The AVR can go closer than 2 if \( m \) is bounded or the IC is restrictive:

**Theorem**

Let \( m \) be constant. Then the AVR is an \( \alpha \)-approximation of the DBR

**Theorem**

Let \( m \) be constant and let IC be a conjunction of \( k \) distinct literals. Then the AVR is an \( \alpha \)-approximation of the DBR

The MVR cannot do better, even with a bounded number of issues.
Recall that $m$ is the number of issues; $n$ is the number of voters.

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

Winner determination for the MVR is 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)

\[ F \text{ satisfies reinforcement if for any two profiles } B \text{ and } B' \text{ such that:} \]

- \( \text{Supp}(B) = \text{Supp}(B') \)
- \( F(B) \cap F(B') \neq \emptyset \)

we have that \( F(B \oplus B') = F(B) \cap F(B') \)

**Theorem**

*The AVR satisfies reinforcement, but the MVR does not.*
Conclusions

Characterisation results in binary aggregation with integrity constraints suggests novel simple procedures to be used in practice:

- **AVR**: voters minimizing average distance
- **MVR**: voters minimizing distance from majority
- **RVR**: voters chosen by sequential majority following strength order

Very good properties (AVR and MVR):

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

Not all definitions work well: RVR is a bad approximation.

Endriss and Grandi. Binary Aggregation by Selection of the Most Representative Voter. AAAI-2014