Judgment Aggregation
as Maximization of Social and Epistemic Utility

Szymon Klarman

Institute for Logic, Language and Computation
University of Amsterdam
sklarman@science.uva.nl

ComSoC-2008, Liverpool
Problem of Judgment Aggregation

Let $\Phi$ be an agenda, such that for every $\varphi \in \Phi$ there is also $\neg \varphi \in \Phi$, and $A = \{1, \ldots, n\}$ be a set of agents.

An individual judgment of agent $i$ with respect to $\Phi$ is a subset $\Phi_i \subseteq \Phi$ of those propositions from $\Phi$ that $i$ accepts. The collection $\{\Phi_i\}_{i \in A}$ is the profile of individual judgments with respect to $\Phi$. A collective judgment with respect to $\Phi$ is a subset $\Psi \subseteq \Phi$.

Rationality constraints: completeness, consistency.

A judgment aggregation function is a function that assigns a single collective judgment $\Psi$ to every profile $\{\Phi_i\}_{i \in A}$ of individual judgments from the domain.

Requirements for JAF: universal domain, anonymity, independence.
Imppossibility Result

The *propositionwise majority voting* rule entails the *discursive dilemma*.


Escape routes:

- Relaxing *completeness*: no obvious choice for the propositions to be removed from the judgement.
- Relaxing *independence*: doctrinal paradox
  - Conclusion-driven procedure,
  - Premise-driven procedure,
  - Argument-driven procedure.

Inspiration

There is a similar problem known as the *lottery paradox* that has been discussed in the philosophy of science.

The lottery paradox concerns the problem of *acceptance of logically connected propositions in science* on the basis of the support provided by empirical evidence. Propositionwise acceptance based on *high probability* leads to inconsistency.


I. Levi suggested that *acceptance* can be seen as a special case of *decision making* and thus analyzed in a *decision-theoretic framework*. He showed also how the lottery paradox can be tackled in this framework.

### Decision-Making Under Uncertainty

<table>
<thead>
<tr>
<th>Probability</th>
<th>$P(v_1)$</th>
<th>...</th>
<th>$P(v_i)$</th>
<th>...</th>
<th>$P(v_m)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Possible states of the world</td>
<td>$v_1$</td>
<td>...</td>
<td>$v_i$</td>
<td>...</td>
<td>$v_m$</td>
</tr>
<tr>
<td>Actions</td>
<td>$A_1$</td>
<td></td>
<td>$u(A_1, v_1)$</td>
<td>...</td>
<td>$u(A_1, v_i)$</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>$A_j$</td>
<td></td>
<td>$u(A_j, v_1)$</td>
<td>...</td>
<td>$u(A_j, v_i)$</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>$A_n$</td>
<td></td>
<td>$u(A_n, v_1)$</td>
<td>...</td>
<td>$u(A_n, v_i)$</td>
</tr>
<tr>
<td>Expected utility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Maximization of expected utility:**

Choose $A$ that maximizes $EU(A) = \sum_{i \in [1, m]} P(v_i) u(A, v_i)$. 

Szymon Klarman
Actions are the acts of acceptance of possible collective judgments.

The set of possible collective judgments $\mathcal{CJ} = \{\Psi_1, \ldots, \Psi_m\}$ typically contains judgments that are consistent, though not necessarily complete.

Example ($\Phi = \{p, \neg p, q, \neg q, r, \neg r\}$, where $r \equiv p \land q$)

\[
\mathcal{CJ} = \{\{p, q, r\}, \{\neg p, q, \neg r\}, \{p, \neg q, \neg r\}, \{\neg p, \neg q, \neg r\}, \\
\{\neg p, \neg r\}, \{\neg q, \neg r\}, \{\neg r\}, \emptyset\}
\]
Possible States of the World

\[ \mathcal{M}_\Phi = \{v_1, \ldots, v_l\} \] is the set of all \textit{possible states of the world} with respect to \( \Phi \), where each \( v_j \) is a unique \textit{truth valuation} for the formulas from \( \Phi \).

Example \((\Phi = \{p, \neg p, q, \neg q, r, \neg r\}, \text{ where } r \equiv p \land q)\)

\[ \mathcal{M}_\Phi = \{v_1, v_2, v_3, v_4\}, \text{ such that:} \]

\[
\begin{align*}
    v_1 : & \quad v_1(p) = 1, \quad v_1(q) = 1, \quad v_1(r) = 1, \\
    v_2 : & \quad v_2(p) = 0, \quad v_2(q) = 1, \quad v_2(r) = 0, \\
    v_3 : & \quad v_3(p) = 1, \quad v_3(q) = 0, \quad v_3(r) = 0, \\
    v_4 : & \quad v_4(p) = 0, \quad v_4(q) = 0, \quad v_4(r) = 0
\end{align*}
\]
Probability

Given the *degree of reliability* of agents \(0.5 < r < 1\) and the profile of individual judgments we can derive the *probability distribution* over \(\mathcal{M}_\Phi\) using the Bayesian Update Rule.

The degree of reliability represents *the likelihood* that an agent *correctly identifies the true state*.

A single update for \(\nu \models \Phi_i\):

\[
P(\nu | \Phi_i) = \frac{P(\Phi_i | \nu)P(\nu)}{\sum_j P(\Phi | \nu_j)P(\nu_j)}
\]

**Example** \((\mathcal{M}_\Phi = \{\nu_1, \nu_2, \nu_3, \nu_4\}, \ r = 0.7)\)

\[
\begin{align*}
P(\nu_1) &= 0.25 & P(\nu_2) &= 0.25 & P(\nu_3) &= 0.25 & P(\nu_4) &= 0.25
\end{align*}
\]
Probability

Given the *degree of reliability* of agents \(0.5 < r < 1\) and the profile of individual judgments we can derive the *probability distribution* over \(M_\Phi\) using the Bayesian Update Rule.

The degree of reliability represents *the likelihood* that an agent *correctly identifies the true state*.

A single update for \(\nu \models \Phi_i\):

\[
P(\nu|\Phi_i) = \frac{P(\Phi_i|\nu)P(\nu)}{\sum_j P(\Phi|\nu_j)P(\nu_j)}
\]

**Example** \((M_\Phi = \{\nu_1, \nu_2, \nu_3, \nu_4\}, r = 0.7)\)

\[
P(\nu_1) = 0.44 \quad P(\nu_2) = 0.19 \quad P(\nu_3) = 0.19 \quad P(\nu_4) = 0.19
\]

\(\nu_1 \models \Phi_1\)
Given the *degree of reliability* of agents (0.5 < r < 1) and the profile of individual judgments we can derive the *probability distribution* over $\mathcal{M}_\Phi$ using the Bayesian Update Rule.

The degree of reliability represents *the likelihood* that an agent *correctly identifies the true state*.

A single update for $\nu \models \Phi_i$:

$$P(\nu | \Phi_i) = \frac{P(\Phi_i | \nu)P(\nu)}{\sum_j P(\Phi | \nu_j)P(\nu_j)}$$

**Example** ($\mathcal{M}_\Phi = \{\nu_1, \nu_2, \nu_3, \nu_4\}$, $r = 0.7$)

$$P(\nu_1) = 0.64 \quad P(\nu_2) = 0.12 \quad P(\nu_3) = 0.12 \quad P(\nu_4) = 0.12$$

$\nu_1 \models \Phi_1$, $\nu_1 \models \Phi_2$
Probability

Given the degree of reliability of agents \((0.5 < r < 1)\) and the profile of individual judgments we can derive the probability distribution over \(\mathcal{M}_\Phi\) using the Bayesian Update Rule.

The degree of reliability represents the likelihood that an agent correctly identifies the true state.

A single update for \(\nu \models \Phi_i\):

\[
P(\nu | \Phi_i) = \frac{P(\Phi_i | \nu)P(\nu)}{\sum_j P(\Phi | \nu_j)P(\nu_j)}
\]

Example (\(\mathcal{M}_\Phi = \{\nu_1, \nu_2, \nu_3, \nu_4\}, r = 0.7\))

<table>
<thead>
<tr>
<th>(\nu)</th>
<th>(P(\nu))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\nu_1)</td>
<td>0.56</td>
</tr>
<tr>
<td>(\nu_2)</td>
<td>0.24</td>
</tr>
<tr>
<td>(\nu_3)</td>
<td>0.10</td>
</tr>
<tr>
<td>(\nu_4)</td>
<td>0.10</td>
</tr>
</tbody>
</table>

\(\nu_1 \models \Phi_1, \quad \nu_1 \models \Phi_2, \quad \nu_2 \models \Phi_3\)
Utility Function

The collective judgment selected by a group is expected to fairly reflect opinions of the group’s members (social goal) as well as to have good epistemic properties, i.e. to be based on a rational cognitive act (epistemic goals).

\[ u(\Psi, v_i) \sim u_\varepsilon(\Psi, v_i) + u_s(\Psi) \]

- \( u_\varepsilon(\Psi, v_i) \) — epistemic utility — adopted from the cognitive decision model of I. Levi. Involves a trade-off between epistemic goals.
- \( u_s(\Psi) \) — social utility — a distance measure of the judgment from the majoritarian choice.
Epistemic Goals

*Epistemically good* judgments are ones that convey a large amount of *information* about the world and are very likely to be *true*.

**Measure of information content** (completeness):

\[
\text{cont}(\Psi) = \frac{|\{v_i \in M_\Phi : v_i \not\models \Psi\}|}{|M_\Phi|}
\]

Example \((\Phi = \{p, \neg p, q, \neg q, r, \neg r\}, \text{ where } r \equiv p \land q)\)

\[
\text{cont}(\{p, q, r\}) = 0.75 \quad \text{cont}(\{\neg r\}) = 0.25
\]

**Measure of truth**:

\[
T(\Psi, v_i) = \begin{cases} 
1 & \text{iff } v_i \models \Psi \\
0 & \text{iff } v_i \not\models \Psi
\end{cases}
\]
Social Goal

The social value of a collective judgment depends on how well the judgment responds to individual opinions of agents, i.e. to what extent agents individually agree on it.

Measure of social agreement:

- for any $\varphi \in \Phi$: $SA(\varphi) = \frac{|A_\varphi|}{|A|}$,
- for any $\Psi_i \in \mathcal{CJ}$: $SA(\Psi_i) = \frac{1}{|\Psi_i|} \sum_{\varphi \in \Psi_i} SA(\varphi)$,

The measure expresses what proportion of propositions from a judgment is on average accepted by an agent (normalized Hamming distance).
Acceptance Rule

The total utility of accepting a collective judgment:

\[
u(\Psi, v_i) = \beta \left( \alpha \, \text{cont}(\Psi) + (1 - \alpha) \, T(\Psi, v_i) \right) + (1 - \beta) \, \text{SA}(\Psi)
\]

\[
= \beta \, u_\varepsilon(\Psi, v_i) + (1 - \beta) \, u_s(\Psi)
\]

Coefficient \( \beta \in [0, 1] \) should reflect the 'compromise' preference of the group between the epistemic and social goals; coefficient \( \alpha \in [0, 1] \) — between information content and truth.

(Provisional) tie-breaking rule:
In case of a tie accept the common information contained in the selected collective judgments.

The utilitarian judgment aggregation function

\[
\text{JAF}(\{\Phi_i\}_{i \in \mathcal{A}}) = \bigcap \Psi
\]

such that \( \Psi \in \arg \max_{\Psi \in \mathcal{C} \mathcal{J}} \sum_{v_i \in \mathcal{M}_\Phi} P(v_i)u(\Psi, v_i) \)
Conclusions

The utilitarian model of judgment aggregation:

- brings together perspectives of social choice theory and epistemology,
- relaxes independence and completeness requirements in a justified and controlled manner (the discursive dilemma resolved!),
- is predominantly a tool for theoretical analysis of judgment aggregation procedures, amenable to various extensions and revisions.

However:

- unless trimmed it is hardly feasible as a practical aggregation method,
- some ingredients of the model are debatable (the tie-breaking rule, probabilities...).
Conclusions

\[ u(\Psi, v_i) = \beta \left( \alpha \text{cont}(\Psi) + (1 - \alpha) T(\Psi, v_i) \right) + (1 - \beta) \text{SA}(\Psi) \]

\[ = \beta \underbrace{u_\varepsilon(\Psi, v_i)}_{\text{propositionwise majority voting}} + (1 - \beta) \underbrace{u_s(\Psi)}_{\text{argument-based aggregation}} \]

- \( \beta = 0, \mathcal{CJ} = \text{all complete judgments: propositionwise majority voting} \),
- \( \beta = 0, \mathcal{CJ} = \text{complete and consistent judgments: argument-based aggregation} \) (Pigozzi, 2006),
- \( \alpha = 1: \text{completeness vs. responsiveness trade-off} \),
- \( \beta = 1: \text{cognitive decision model} \) (Levi, 1967).