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

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

Rationality constraints: *completeness*, *consistency*.

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

Requirements for JAF: *universal domain*, *anonymity*, *independence*.

Impossibility Result

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

C. List, P. Pettit (2002), "Aggregating Sets of Judgments: an Impossibility Result", in: *Economics and Philosophy*, 18: 89-110.

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.

G. Pigozzi (2006), "Belief Merging and the Discursive Dilemma: an Argument-Based Account to Paradoxes of Judgment Aggregation", in: *Synthese*, 152(2): 285-298.

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. Douven, J. W. Romeijn (2006), "The Discursive Dilemma as a Lottery Paradox", in: *Proceedings of the 1st International Workshop on Computational Social Choice (COMSOC-2006)*, ILLC University of Amsterdam: 164-177.

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.

I. Levi (1967), *Gambling with Truth. An Essay on Induction and the Aims of Science*, MIT Press: Cambridge.

Decision-Making Under Uncertainty

	Probability					
	$P(v_1)$	...	$P(v_i)$	...	$P(v_m)$	
	Possible states of the world					
Actions	v_1	...	v_i	...	v_m	**Expected utility**
A_1	$u(A_1, v_1)$	...	$u(A_1, v_i)$	...	$u(A_1, v_m)$	$EU(A_1)$
...	...	...	...	...	...	...
A_j	$u(A_j, v_1)$	...	$u(A_j, v_i)$	...	$u(A_j, v_m)$	$EU(A_j)$
...	...	...	...	...	...	...
A_n	$u(A_n, v_1)$	...	$u(A_n, v_i)$	...	$u(A_n, v_m)$	$EU(A_n)$

Maximization of expected utility:

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

Actions

Actions are the acts of acceptance of *possible collective judgments*.

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

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

$$\mathcal{CJ} = \{\{p, q, r\}, \{\neg p, q, \neg r\}, \{p, \neg q, \neg r\}, \{\neg p, \neg q, \neg r\}, \{\neg p, \neg r\}, \{\neg q, \neg r\}, \{\neg r\}, \emptyset\}$$

Possible States of the World

$\mathcal{M}_\Phi = \{v_1, ..., v_l\}$ is the set of all *possible states of the world* with respect to Φ, where each v_j is a unique *truth valuation* for the formulas from Φ.

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

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

$$
\begin{array}{ll}
v_1: & v_1(p) = 1,\ v_1(q) = 1,\ v_1(r) = 1 \\
v_2: & v_2(p) = 0,\ v_2(q) = 1,\ v_2(r) = 0 \\
v_3: & v_3(p) = 1,\ v_3(q) = 0,\ v_3(r) = 0 \\
v_4: & v_4(p) = 0,\ v_4(q) = 0,\ v_4(r) = 0
\end{array}
$$

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 $v \models \Phi_i$:

$$P(v|\Phi_i) = \frac{P(\Phi_i|v)P(v)}{\sum_j P(\Phi|v_j)P(v_j)}$$

Example ($\mathcal{M}_\Phi = \{v_1, v_2, v_3, v_4\}$, $r = 0.7$)

$$P(v_1) = 0.25 \quad P(v_2) = 0.25 \quad P(v_3) = 0.25 \quad P(v_4) = 0.25$$

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 $v \models \Phi_i$:

$$P(v|\Phi_i) = \frac{P(\Phi_i|v)P(v)}{\sum_j P(\Phi|v_j)P(v_j)}$$

Example ($\mathcal{M}_\Phi = \{v_1, v_2, v_3, v_4\}$, $r = 0.7$)

$$P(v_1) = 0.44 \quad P(v_2) = 0.19 \quad P(v_3) = 0.19 \quad P(v_4) = 0.19$$

$$v_1 \models \Phi_1$$

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 $v \models \Phi_i$:

$$P(v|\Phi_i) = \frac{P(\Phi_i|v)P(v)}{\sum_j P(\Phi|v_j)P(v_j)}$$

Example ($\mathcal{M}_\Phi = \{v_1, v_2, v_3, v_4\}$, $r = 0.7$)

$$P(v_1) = 0.64 \quad P(v_2) = 0.12 \quad P(v_3) = 0.12 \quad P(v_4) = 0.12$$

$$v_1 \models \Phi_1, \qquad v_1 \models \Phi_2$$

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 $v \models \Phi_i$:

$$P(v|\Phi_i) = \frac{P(\Phi_i|v)P(v)}{\sum_j P(\Phi|v_j)P(v_j)}$$

Example ($\mathcal{M}_\Phi = \{v_1, v_2, v_3, v_4\}$, $r = 0.7$)

$$P(v_1) = 0.56 \quad P(v_2) = 0.24 \quad P(v_3) = 0.10 \quad P(v_4) = 0.10$$

$$v_1 \models \Phi_1, \quad v_1 \models \Phi_2, \quad v_2 \models \Phi_3$$

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*).

$$u(\Psi, v_i) \quad \sim \quad u_\varepsilon(\Psi, v_i) + u_s(\Psi)$$

$u_\varepsilon(\Psi, v_i)$ — *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):

$$\text{cont}(\Psi) = \frac{|v_i \in \mathcal{M}_\Phi : v_i \not\models \Psi|}{|\mathcal{M}_\Phi|}$$

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

$$\text{cont}(\{p, q, r\}) = 0.75 \qquad \text{cont}(\{\neg r\}) = 0.25$$

Measure of truth:

$$\mathrm{T}(\Psi, v_i) = \begin{cases} 1 \text{ iff } v_i \models \Psi \\ 0 \text{ iff } v_i \not\models \Psi \end{cases}$$

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

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:

$$u(\Psi, v_i) = \beta \underbrace{(\alpha\ \mathrm{cont}(\Psi) + (1-\alpha)\ \mathrm{T}(\Psi, v_i))}_{u_\varepsilon(\Psi, v_i)} + (1-\beta) \underbrace{\mathrm{SA}(\Psi)}_{u_s(\Psi)}$$

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

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

$$\mathrm{JAF}(\{\Phi_i\}_{i\in\mathcal{A}}) = \bigcap \Psi$$

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

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

$$u(\Psi, v_i) = \beta \underbrace{(\alpha\ \mathsf{cont}(\Psi) + (1-\alpha)\ \mathrm{T}(\Psi, v_i))}_{} + (1-\beta)\ \underbrace{\mathrm{SA}(\Psi)}_{}$$
$$= \beta\ u_{\varepsilon}(\Psi, v_i) + (1-\beta)\ u_s(\Psi)$$

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