Judgment Aggregation

Ulle Endriss
Institute for Logic, Language and Computation
University of Amsterdam

Example

Suppose three robots are in charge of climate control for this building. They need to make judgments on p (the temperature is below 17°C), q (we should switch on the heating), and the "policy" $p \to q$.

	p	$p \to q$	q
Robot 1:	Yes	Yes	Yes
Robot 2:	No	Yes	No
Robot 3:	Yes	No	No

Exercise: *Should we switch on the heating?*

Outline

This will be an introduction to the theory of *judgment aggregation*.

- Reminder: what you need to know about *logic* to understand this
- The *paradox* of judgment aggregation: a second example
- Main question: Is there a *reasonable method* of aggregation?

Reminder: Consistency

A set of propositional formulas is said to be *consistent* (or *satisfiable*) if we can assign truth values to the propositional variables occurring within the set so that all the formulas in the set become true.

Exercise: *Which of the following sets are consistent?*

- $\{\neg q,\ p \rightarrow q,\ p\}$
- $\{p \vee q,\ \neg p \vee \neg q,\ r\}$
- $\{p,\ q,\ \neg(p \wedge q)\}$

Example

A defendant is accused of a breach of contract . . .

Legal doctrine stipulates that you are *guilty* *if and only if* it is the case that the agreement was *binding* (p) *and* has *not* been *honoured* ($\neg q$).

	p	q	$p \wedge \neg q$
Judge 1:	Yes	No	Yes
Judge 2:	Yes	Yes	No
Judge 3:	No	No	No

Exercise: *Should we pronounce the defendant guilty?*

The Paradox of Judgment Aggregation

Once again our two examples:

	p	$p \to q$	q
Robot 1:	Yes	Yes	Yes
Robot 2:	No	Yes	No
Robot 3:	Yes	No	No

	p	q	$p \wedge \neg q$
Judge 1:	Yes	No	Yes
Judge 2:	Yes	Yes	No
Judge 3:	No	No	No

Why do we call this a *paradox*? Two explanations:

- Premise-driven rule and conclusion-driven rule disagree
- Majority rule produces judgment set that is not consistent

Formal Framework

An *agenda* Φ is a set of propositional formulas (and their negations).

Example: $\Phi = \{p,\ \neg p,\ p \rightarrow q,\ \neg(p \rightarrow q),\ q,\ \neg q\}$

A *judgment set* J for the agenda Φ is a subset of Φ. We call J:

- *complete* if $\varphi \in J$ or $\neg\varphi \in J$ for all formulas $\varphi, \neg\varphi \in \Phi$
- *consistent* if J has a satisfying truth assignment

Now n individual *agents* each express judgments on the formulas in Φ, producing a *profile* $\boldsymbol{J} = (J_1, \ldots, J_n)$ of complete and consistent sets.

Example: $\boldsymbol{J} = (\{p, p \rightarrow q, q\},\ \{\neg p, p \rightarrow q, \neg q\},\ \{p, \neg(p \rightarrow q), \neg q\})$

An *aggregation rule* F for an agenda Φ and a group of n agents is a function mapping every given profile of complete and consistent sets to a single collective judgment set.

Example: Majority Rule

Suppose three agents express judgments on the formulas in the agenda $\Phi = \{p, \neg p, q, \neg q, p \vee q, \neg(p \vee q)\}$.

For simplicity, we only show the positive formulas in our tables:

	p	q	$p \vee q$	
Agent 1:	Yes	No	Yes	$J_1 = \{p, \neg q, p \vee q\}$
Agent 2:	Yes	Yes	Yes	$J_2 = \{p, q, p \vee q\}$
Agent 3:	No	No	No	$J_3 = \{\neg p, \neg q, \neg(p \vee q)\}$

The (strict) *majority rule* F_{maj} takes a (complete and consistent) profile and returns the set of formulas accepted by $> \frac{n}{2}$ agents.

In our example: $F_{\text{maj}}(\boldsymbol{J}) = \{p, \neg q, p \vee q\}$ [complete and consistent!]

Other Rules

Instead of using the *majority rule*, we could also use:

- *Premise-driven rule:* use majority voting on literals and infer other formulas from the literals accepted
- *Quota-based rules:* e.g., accept a formula if $\geqslant \frac{2}{3}$ of the agents do

There are many more options. *So how do you choose?*

The Axiomatic Method

What makes for a "good" aggregation rule F? The following so-called *axioms* all express intuitively appealing properties:

- *Anonymity:* Treat all individual agents symmetrically!
- *Neutrality:* Treat all formulas symmetrically!
- *Independence:* To decide whether to accept formula φ, you should only have to consider which individual agents accept φ!

Observe that the *majority rule* satisfies all of these axioms . . .

. . . but so do various other rules!

Exercise: *Can you think of some examples?*

Impossibility Theorem

We have seen that the majority rule does *not* preserve *consistency*.
Is there another "reasonable" rule that does not have this problem?

Surprise: No! (at least not for certain agendas)

Theorem 1 (List and Pettit, 2002) *No judgment aggregation rule for $\geqslant 2$ agents and an agenda Φ with $\{p, q, p \wedge q\} \subseteq \Phi$ that satisfies the axioms of anonymity, neutrality, and independence will always return a collective judgment set that is complete and consistent.*

Remark: Also true for other agendas (such as all those we saw today).

C. List and P. Pettit. Aggregating Sets of Judgments: An Impossibility Result. *Economics and Philosophy*, 18(1):89–110, 2002.

Proof

First, understand the impact of our three axioms:

- *Independence:* acceptance of φ only depends on *who* accepts φ.
- Add *anonymity:* it only depends on *how many* agents accept φ.
- Add *neutrality:* must use *same* acceptance criterion for all formulas.

We now prove the theorem for *odd* n (it's even easier for even n).

Let $N_\varphi^{\boldsymbol{J}}$ be the set of agents who accept formula φ in profile $\boldsymbol{J}$.

Consider a profile $\boldsymbol{J}$ where $\frac{n-1}{2}$ agents accept p and q; one accepts p but not q; one accepts q but not p; and $\frac{n-3}{2}$ accept neither p nor q. That is: $|N_p^{\boldsymbol{J}}| = |N_q^{\boldsymbol{J}}| = |N_{\neg(p \wedge q)}^{\boldsymbol{J}}| = \frac{n+1}{2}$. Then:

- Accepting all three formulas contradicts consistency.
- But if we accept none, completeness forces us to accept their complements, which also contradicts consistency.

So it is indeed *impossible* to satisfy all of our requirements. ✓

Related Research at the ILLC

- Finding aggregation rules that *maximise the chance* of returning the *"correct"* judgment (think of agents as domain experts)
- Analysing under what circumstances an agent might derive an advantage from *strategically misrepresenting* her judgments
- Understanding simpler aggregation scenarios by *embedding* them into judgment aggregation (example: *participatory budgeting*)
- Designing JA-inspired methods for *crowdsourcing* of linguistic judgments, to support research in *computational linguistics*

Z. Terzopoulou and U. Endriss. Optimal Truth-Tracking Rules for the Aggregation of Incomplete Judgments. SAGT-2019.

S. Botan and U. Endriss. Majority-Strategyproofness in JA. AAMAS-2020.

S. Rey, U. Endriss, and R. de Haan. Designing Participatory Budgeting Mechanisms Grounded in Judgment Aggregation. KR-2020.

C. Qing, U. Endriss, R. Fernández, and J. Kruger. Empirical Analysis of Aggregation Methods for Collective Annotation. COLING-2014.

Last Slide

This has been an introduction to *judgment aggregation*. We saw:

- Formal framework for aggregating views on complex matters
- Applicable to many diverse settings (thus: important)
- Modelling *coherent* judgments as *consistent* sets of formulas
- *Paradox:* majority view of coherent judges may be incoherent
- Thus: need to carefully analyse the problem (*axiomatic method*)
- *Impossibility:* no "reasonable" rule can always be coherent
- Active research topic at the ILLC