Ontology Merging as Social Choice

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

[joint work with Daniele Porello]

Ontology Merging as Social Choice

The Semantic Web might provide access to *several ontologies* describing the same domain (hopefully using the *same vocabulary*).

Ontology merging is the problem of amalgamating this kind of distributed information into a single ontology.

A new perspective: think of this as an *aggregation* problem.

Talk Outline

The talk will focus on technical issues that arise when we adapt the framework of *judgment aggregation* to this new application. The main difference is that we will adopt an *open world assumption*.

What next?

- Introductory example
- Definition of the formal framework
- Discussion of axioms + some simple results
- Discussion of possible aggregation rules

The Doctrinal Paradox

Three agents agree on the following concept definition (TBox):

$$C_3 \equiv C_1 \sqcap C_2$$

But they have diverging opinions about which concepts the object a is an instance of (ABox):

	$C_1(a)$	$C_2(a)$	$C_3(a)$
Agent 1	Yes	Yes	Yes
Agent 2	Yes	No	No
Agent 3	No	Yes	No
Majority	Yes	Yes	No

Paradox: even though each individual ontology is satisfiable, the ontology obtained by applying the majority rule is not.

Formal Framework

An *agenda* Φ is finite set of formulas (in some logic). Unlike for standard JA, we don't require Φ to be closed under complementation.

An *ontology* is a set $O \subseteq \Phi$. The set of *satisfiable* ontologies is $\mathrm{On}(\Phi)$.

Let $\mathcal{N} = \{1, \ldots, n\}$ be a finite set of *agents*. If each of them provides a satisfiable ontology, we get a *profile* $\boldsymbol{O} = (O_1, \ldots, O_n) \in \mathrm{On}(\Phi)^{\mathcal{N}}$.

An *ontology aggregator* is a function $F : \mathrm{On}(\Phi)^{\mathcal{N}} \to 2^{\Phi}$ mapping any such profile of satisfiable ontologies to an ontology.

<u>We will</u>

- look for aggregators that ensure the *satisfiability* of outcomes;
- *not* talk about *complete* outcomes (makes no sense here); and
- *not* talk about *deductively closed* outcomes either (not attractive for ontology engineering: information overload).

Basic Axioms

Axioms are used to describe desirable properties of aggregators.

A couple of standard axioms we'd surely want also here:

- F is *anonymous* if $F(O_1, \ldots, O_n) = F(O_{\pi(1)}, \ldots, O_{\pi(n)})$ for any profile $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$ and any permutation $\pi : \mathcal{N} \to \mathcal{N}$.
- F is *unanimous* if $O_1 \cap \cdots \cap O_n \subseteq F(\boldsymbol{O})$ for any $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$.

Neutrality

The standard neutrality axiom from JA:

- F is *neutral* if for any $\varphi, \psi \in \Phi$ and $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$ we have $\varphi \in O_i \Leftrightarrow \psi \in O_i$ for all $i \in \mathcal{N}$ implies $\varphi \in F(\boldsymbol{O}) \Leftrightarrow \psi \in F(\boldsymbol{O})$.

Do we want it? Maybe (but also questionable in standard JA).

In standard JA, the next axiom would come for free from neutrality + completeness + complement-freeness:

- F is *$\star$-neutral* if for any $\varphi \in \Phi$ and $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$ we have $\varphi \in O_i \Leftrightarrow \psi \notin O_i$ for all $i \in \mathcal{N}$ implies $\varphi \in F(\boldsymbol{O}) \Leftrightarrow \psi \notin F(\boldsymbol{O})$.

This we don't want: it suggests that not including a formula into your ontology is like including its negation (closed world assumption).

Groundedness

If we perform "coarse" ontology aggregation, then we'd want this:

- F is *grounded* if $F(\boldsymbol{O}) \subseteq O_1 \cup \cdots \cup O_n$ for any $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$.

This axiom is not needed in standard JA: groundedness wrt. φ is implied by unanimity wrt. $\neg\varphi$.

Exhaustiveness

If we take all information provided by the agents as potentially useful (as long as it doesn't contradict other information we want to accept), then we should adopt this axiom:

- F is *exhaustive* if there exists no satisfiable set $\Delta \subseteq O_1 \cup \cdots \cup O_n$ with $F(\boldsymbol{O}) \subset \Delta$ for any profile $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$.

Exhaustiveness vs. $\star$-Neutrality

Another reason why we don't want to adopt $\star$-neutrality:

Proposition 1 *Any ontology aggregator that satisfies $\star$-neutrality violates exhaustiveness.*

Recall the meaning of these axioms:

- $\star$-neutral: treat acceptance symmetrically to rejection
- exhaustive: only reject a proposed φ if it yields a contradiction

Proof: Let $\Phi = \{p, q\}$. If each agent accepts exactly one of p and q, i.e., $p \in O_i \Leftrightarrow q \notin O_i$ for all $i \in \mathcal{N}$, then *$\star$-neutrality* forces accepting exactly one of p and q, but *exhaustiveness* forces accepting both. ✓

"Semantic" Axioms

It not only matters what information is *explicitly included* in an ontology, but also what information can be *inferred* from it.

This suggests formulating "semantic" variants of known axioms, e.g.:

- F is *unanimous* if $O_1 \cap \cdots \cap O_n \subseteq F(\boldsymbol{O})$ for any $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$.
- Define $\mathrm{Cl}(\Delta) := \{\varphi \in \Phi \mid \Delta \models \varphi\}$. F is *semantically unanimous* if $\mathrm{Cl}(O_1) \cap \cdots \cap \mathrm{Cl}(O_n) \subseteq \mathrm{Cl}(F(\boldsymbol{O}))$ for any $\boldsymbol{O} \in \mathrm{On}(\Phi)^{\mathcal{N}}$.

For standard JA this distinction is not relevant, because judgment sets are already assumed to be deductively closed.

Comparing the Unanimity Axioms

Neither of them is stronger than the other:

- U $\not\Rightarrow$ SU: Suppose three agents accept TBox $\{C_1 \equiv C_2, C_2 \equiv C_3\}$ and the ABox of agent i is $\{C_i(a)\}$. Then the (*unanimous*!) majority rule produces an *empty* ABox, in violation of *semantic unanimity*, which requires returning an ABox that entails $C_1(a)$.
- SU $\not\Rightarrow$ U: The rule mapping any profile to $\{C \equiv D \sqcap \neg D,\ C(a)\}$ is (vacuously) semantically unanimous but not unanimous.

But for "well-behaved" aggregators we get the expected entailment:

Proposition 2 *Any satisfiable and exhaustive ontology aggregator that is semantically unanimous is also unanimous.*

Proof: Routine.

An Impossibility

The *union aggregator* is defined via $F_u(\boldsymbol{O}) := O_1 \cup \cdots \cup O_n$.
It is neither satisfiable nor otherwise exciting. But:

Proposition 3 *The only aggregator that is anonymous, neutral, independent and semantically unanimous is the union aggregator.*

Proof: By *anonymity* and *independence*, collective acceptance of φ only depends on the *cardinality* of the coalition accepting φ.
By *neutrality*, the acceptable cardinalities do not depend on φ.

Suppose F does *not* accept a formula when exactly $k > 0$ agents do.

Construct profile in which agents $i, \ldots, (i{+}k{-}1 \mod n)$ accept $C_i(a)$ and all accept $C_1 \equiv C_2, \ldots, C_{n-1} \equiv C_n$. But by *semantic unanimity*, F should accept at least one $C_i(a)$. $\leadsto$ Contradiction. ✓

Remark: The proof assumes F is defined for any agenda; this might open up the way for (more positive) agenda characterisation results.

Aggregation Rules

For the final part of the talk, let us consider a few pragmatic ways of dealing with the aggregation problem:

- Quota-based rules (simple; only high quota ensures satisfiability)
- Support-based rules
- Distance-based rules
- Two-stage rules

Support-based Rules

(Same idea also in Marija's dissertation.)

Given profile $\boldsymbol{O}$, fix an order $\gg$ on the agenda Φ such that $\varphi \gg \psi$ only when $\#\{i \mid \varphi \in O_i\} \geqslant \#\{i \mid \psi \in O_i\}$.

Now define an aggregator $F_{\gg}$ that accepts φ iff:

- $\{i \mid \varphi \in O_i\} \neq \emptyset$ and
- $\{\psi \in F_{\gg}(\boldsymbol{O}) \mid \psi \gg \varphi\} \cup \{\varphi\}$ is satisfiable

Anonymous, unanimous, grounded, exhaustive, monotonic, satisfiable.

Possible variations:

- Could also take union over all possible $\gg$'s: irresolute, neutral.
- Could also define $\gg$ differently, e.g., in terms of $\{i \mid O_i \models \varphi\}$ or only respecting subset-relation on coalitions rather than cardinality.

Distance-based Rules

The usual idea, with a minor tweak, because we don't want our "distance" to be *symmetric:* dropping a proposed formula is much worse than adding one you had not thought of yourself.

$$F_d(\boldsymbol{O}) \quad = \quad \mathrm{argmin}_{O \in \mathrm{On}(\Phi)} \sum_{i \in \mathcal{N}} d(O_i, O),$$

$$\text{where } d(X, Y) := \#(X \setminus Y)$$

Two-stage Rules

Division into assertional and terminological knowledge could be exploited during aggregation, giving precedence to one.

(Note: well-defined, unlike premise/conclusion division in JA.)

For instance:

(1) Aggregate ABoxes using the majority rule or some other quota rule (satisfiability is guaranteed if concepts are atomic).

(2) Use support-based rule or distance-based rule for the TBoxes, taking the collective ABox as fixed.

Last Slide

Suggested *ontology merging* as a possible application for JA.

Technical *differences* from the point of view of JA:

- not accepting a formula is different from accepting its negation
- agents might not provide judgments on parts of the agenda at all
- natural partition of agenda is possible (ABox vs. TBox)

These differences suggest some new *directions:*

- new axioms: groundedness and exhaustiveness
- variants of axioms referring to deductive closures of judgment sets
- aggregation rules may take TBox/ABox division into account

Full details are in our CLIMA-2011 paper.

D. Porello and U. Endriss. Ontology Merging as Social Choice. *Proc. 12th Internat. Workshop on Computational Logic in Multiagent Systems*, Springer, 2011.