Ontology Merging as Social Choice

Daniele Porello
(ILLC, University of Amsterdam)

Ontologies and Lexical Semantics

30 September and 1 October 2011

This talk is based on a joint work with Ulle Endriss, *CLIMA 2011*.

Overview

- Part I: Social choice theory: Judgment aggregation.
  - Modelling aggregation of individual judgments.
  - Impossibility results.
  - Characterization results.
- Part II: Ontology aggregation
  - Description Logics and Ontologies,
  - Modelling ontology aggregation;
  - Methods from SCT and JA;
  - Voting procedures on ontologies;
  - Balancing efficiency and fairness;
  - Conclusion and future work.

Part I: Social choice theory and Judgment Aggregation

A discursive dilemma

- (Kornhauser and Sager, 1986) discussed the following decisional dilemma that actually emerged in the deliberative practice of the American supreme court (they had to decide whether a trial had to be revised):
- Suppose p: "the confession was forced"; $p \rightarrow q$: "if the confession was forced, then the trial has to be revised", q: "the trial has to be revised". Judges vote as follows:

	p	$p \rightarrow q$	q
Agent 1:	Yes	Yes	Yes
Agent 2:	No	Yes	No
Agent 3:	Yes	No	No
Majority:	Yes	Yes	No

- Each judge holds a set of proposition that is *consistent*, but the collective judgment set derived using the majority rule is not.

Is it just a bad example? Does it happen for any voting procedure?

SCT: Judgment aggregation, the framework[1]

An **agenda** is a finite nonempty set $\Phi \subseteq \mathcal{L}_{PS}$. We call a judgment set $J \subseteq \Phi$:

- **complete** if $\alpha \in J$ or $\sim\alpha \in J$ for all $\alpha \in \Phi$;
- **complement-free** if for all $\alpha \in \Phi$ it is not the case that both α and its complement are in J;
- **consistent** if there exists an assignment that makes all formulas in J true.

Denote with $J(\Phi)$ the set of all complete consistent subsets of Φ. Given a set $N = \{1, \ldots, n\}$ of $n \geqslant 3$ *individuals* (or *agents*), denote with $\mathbf{J} = (J_1, \ldots, J_n)$ a *profile* of judgment sets, one for each individual.

An **aggregation procedure** for agenda Φ and a set of n individuals is a function $F : J(\Phi)^n \to \mathcal{P}(\Phi)$.

[1] This presentation is based on Endriss, Grandi and Porello, AAMAS 2010, a rephrase of List and Pettit 2002.

Condition on the aggregation: fairness and rationality

Fairness conditions:

- **Unanimity** (U): If $\varphi \in J_i$ for all i then $\varphi \in F(\mathbf{J})$.
- **Anonymity** (A): For any profile $\mathbf{J}$ and any permutation $\sigma : N \to N$ we have $F(J_1, \ldots, J_n) = F(J_{\sigma(1)}, \ldots, J_{\sigma(n)})$.
- **Neutrality** (N): For any φ, ψ in the agenda Φ and profile $\mathbf{J} \in J(\Phi)$, if for all i we have that $\varphi \in J_i \Leftrightarrow \psi \in J_i$, then $\varphi \in F(\mathbf{J}) \Leftrightarrow \psi \in F(\mathbf{J})$.
- **Independence** (I): For any φ in the agenda Φ and profiles $\mathbf{J}$ and $\mathbf{J}'$ in $J(\Phi)$, if $\varphi \in J_i \Leftrightarrow \varphi \in J'_i$ for all i, then $\varphi \in F(\mathbf{J}) \Leftrightarrow \varphi \in F(\mathbf{J}')$.
- **Systematicity** (S): For any φ, ψ in the agenda Φ and profiles $\mathbf{J}$ and $\mathbf{J}'$ in $J(\Phi)$, if $\varphi \in J_i \Leftrightarrow \psi \in J'_i$ for all i, then $\varphi \in F(\mathbf{J}) \Leftrightarrow \psi \in F(\mathbf{J}')$.
- **Monotonicity** $(\mathrm{M}^{\mathrm{I}})$: For any φ in the agenda Φ and profiles $\mathbf{J} = (J_1, \ldots, J_i, \ldots, J_n)$ and $\mathbf{J}' = (J_1, \ldots, J'_i, \ldots, J_n)$ in $J(\Phi)$, if $\varphi \notin J_i$ and $\varphi \in J'_i$, then $\varphi \in F(\mathbf{J}) \Rightarrow \varphi \in F(\mathbf{J}')$.

Rationality conditions:

- F is **rational** iff $F(\mathbf{J})$ is complete and consistent for every $\mathbf{J} \in J(\Phi)$.

Impossibility result

- List and Pettit (*Economics and Philosophy*, 2002) prove the following result:

Theorem (List and Pettit, 2002)

(On sufficiently complex agendas) there is no rational aggregation procedure satisfying (A) and (S).

In particular the majority rule is not rational. We can characterize the majority rule as follows:

Theorem (EGP, AAMAS 2010)

If the number of individuals is odd, an aggregation procedure F satisfies (A), (S) and (M) and *weak rationality* (completeness and complement-free, i.e. not consistency) if and only if F is the majority rule.

Safe agendas

We can see now when an agenda is sufficiently complex:

Median property (Mp)

We say that an agenda Φ satisfies the **median property** (MP), if every non trivially inconsistent subset of Φ has itself an inconsistent subset of size 2.

E.g. $\{\neg p, \underbrace{p, p \rightarrow q, \neg q}, \neg q, \neg(p \rightarrow q)\}$

Theorem (List and Puppe, 2009, EGP, 2010)

An agenda Φ is safe for the majority rule if and only if Φ satisfies the MP.

Part II: Ontology merging

Motivations

- We want to apply the analysis of the aggregation procedures provided by SCT to the problem of aggregating several ontologies.
- Different individual ontologies provide different and possibly contrasting information and we ask which ontology better represent the *group* information.
- Social choice theory shows that the notion of *group information* strongly depends on the aggregation procedure we use.
- Fairness conditions here can be interpreted as constraints of impartiality on agents or on propositions. E.g. anonymity assumes that we do not know which individual ontology is more reliable.

Description logics

- The language of $\mathcal{ALC}$ is based on an alphabet consisting of *atomic concepts*, *role names*, and *object names*.
- The set of *concept descriptions* is generated as follows (where A represents atomic concepts and R role names):

$$C \quad ::= \quad A \mid \neg C \mid C \sqcap C \mid C \sqcup C \mid \forall R.C \mid \exists R.C$$

- A *TBox* is a finite set of formulas of the form $A \sqsubseteq C$ and $A \equiv C$ (where A is an atomic concept and C a concept description).
- An *ABox* is a finite set of formulas of the form $A(a)$ and $R(a, b)$.
- The semantics of $\mathcal{ALC}$ is given by *interpretations* that map each object name to an element of its domain, each atomic concept to a subset of the domain, and each role name to a binary relation on the domain.
- A set of (TBox and ABox) formulas is *satisfiable* if there exists an interpretation in which they are all true.

Ontology Aggregation

- Let us fix a finite set $\mathcal{L}$ of $\mathcal{ALC}$ formulas over this alphabet that includes all the possible ABox formulas.
- We call $\mathcal{L}$ the *agenda* and any set $O \subseteq \mathcal{L}$ an *ontology*. O can be divided into a TBox O^T and an ABox O^A. Let $\mathrm{On}(\mathcal{L})$ be the set of all those ontologies that are *satisfiable*. The *closure* of a set of formulas $\Phi \subseteq \mathcal{L}$ is given by $\mathrm{Cl}(\Phi) := \{\varphi \in \mathcal{L} \mid \Phi \models \varphi\}$
- Let $\mathcal{N} = \{1, \ldots, n\}$ be a finite set of *agents*. Each agent $i \in \mathcal{N}$ provides a satisfiable ontology $O_i \in \mathrm{On}(\mathcal{L})$.
- An *ontology profile* $\mathbf{O} = (O_1, \ldots, O_n) \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$ is a vector of such ontologies, one for each agent. We write $\mathcal{N}_\varphi^{\mathbf{O}} := \{i \in \mathcal{N} \mid \varphi \in O_i\}$ for the set of agents including φ in their ontology under profile $\mathbf{O}$.
- An *ontology aggregator* is a function $F : \mathrm{On}(\mathcal{L})^{\mathcal{N}} \to 2^{\mathcal{L}}$ mapping any profile of satisfiable ontologies to an ontology.
- E.g. an ontology aggregator is F with $F(\mathbf{O}) := O_1 \cup \cdots \cup O_n$, which returns the union of the individuals ontologies. The ontology obtained may not be satisfiable.

An Example

- The *majority rule* (accept a formula if and only if a majority of the agents do) can lead to unsatisfiable outcomes, as we can easily simulate the *discursive dilemmas* from *Judgment Aggregation*.
- Suppose three agents share a common TBox with two formulas:

$$C_3 \equiv C_1 \sqcap C_2 \qquad C_4 \sqsubseteq \neg C_3$$

- Furthermore, suppose the three ABoxes are as follows:

	$C_1(a)$	$C_2(a)$	$C_3(a)$	$C_4(a)$
Agent 1	yes	yes	yes	no
Agent 2	yes	no	no	yes
Agent 3	no	yes	no	yes
Majority	yes	yes	no	yes

- Individual ontologies are satisfiable but the collective one is not.
- E.g. The original example of discursive dilemma can be viewed as a case in which O^T expresses the legal doctrine (Kornhauser and Sager, 1986).

Basic features

- We restrict to "coarse" merging: the ontology to be constructed will be a list of some of the formulas included in the individual ontologies. We do not deal with "fine" merging, where we might also want to construct entirely new formulas from those provided by the individuals.
- *Open vs Closed World Assumption*: we need to adapt the standard JA framework by dropping *completeness* (a agent accepts A or accepts $\neg A$): here it would entail that an agent cannot express her lack of knowledge concerning the application of both A and $\neg A$ to a certain object.
- *Syntactic vs Semantic axioms (explicit vs implicit knowledge)*. We define "syntactic" axioms, they relate to the formulas that occur *explicitly* in the ontologies of individual agents or in the collective ontology. We will contrast this with "semantic" axioms that make reference to the formulas that can be inferred *implicitly* from those ontologies.

Syntactic Axioms (standard)

The usual social choice theory and judgment aggregation axioms can be restated as follows:

- **Unanimity:** F is called *unanimous* if $O_1 \cap \cdots \cap O_n \subseteq F(\mathbf{O})$ for every profile $\mathbf{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.
- **Anonymity:** F is called *anonymous* if for any profile $\mathbf{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$ and any permutation $\pi : \mathcal{N} \to \mathcal{N}$ we have that $F(O_1, \ldots, O_n) = F(O_{\pi(1)}, \ldots, O_{\pi(n)})$.
- **Independence:** F is called *independent* if for any $\varphi \in \mathcal{L}$ and profiles $\mathbf{O}, \mathbf{O}' \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$, we have that $\varphi \in O_i \Leftrightarrow \varphi \in O_i'$ for all $i \in \mathcal{N}$ implies $\varphi \in F(\mathbf{O}) \Leftrightarrow \varphi \in F(\mathbf{O}')$.
- **Monotonicity:** F is called *monotonic* if for any $i \in \mathcal{N}$, $\varphi \in \mathcal{L}$, and $\mathbf{O}, \mathbf{O}' \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$ with $O_j = O_j'$ for all $j \neq i$, we have that $\varphi \in O_i' \setminus O_i$ and $\varphi \in F(\mathbf{O})$ imply $\varphi \in F(\mathbf{O}')$.

Syntactic Axioms (specific)

We introduce the following specific axioms we might want for our ontology aggregators:

- **Groundedness:** F is called *grounded* if $F(\mathbf{O}) \subseteq O_1 \cup \cdots \cup O_n$ for every profile $\mathbf{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.
- **Exhaustiveness:** F is called *exhaustive* if there exists no satisfiable set $\Phi \subseteq O_1 \cup \cdots \cup O_n$ with $F(\mathbf{O}) \subset \Phi$ for any profile $\mathbf{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.
- **Group Closure:** F is called *group-closed* if there exists no set $\Phi \subseteq O_1 \cup \cdots \cup O_n$ with $F(\mathbf{O}) \models \Phi$ and $F(\mathbf{O}) \subset \Phi$ for any profile $\mathbf{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.

Neutrality

An important axiom is *neutrality*, which, intuitively, requires all formulas to be treated symmetrically. In fact, there are a number of possible interpretations of this notion, including these:

- **Neutrality:** F is called *neutral* if for any $\varphi, \psi \in \mathcal{L}$ and $\mathbf{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$ we have that $\varphi \in O_i \Leftrightarrow \psi \in O_i$ for all $i \in \mathcal{N}$ implies $\varphi \in F(\mathbf{O}) \Leftrightarrow \psi \in F(\mathbf{O})$.
- **Acceptance-Rejection Neutrality:** F is called *acceptance-rejection neutral* if for any $\varphi \in \mathcal{L}$ and $\mathbf{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$ we have that $\varphi \in O_i \Leftrightarrow \psi \notin O_i$ for all $i \in \mathcal{N}$ implies $\varphi \in F(\mathbf{O}) \Leftrightarrow \psi \notin F(\mathbf{O})$.

In the case of JA without completeness, acceptance-rejection neutrality is usually assumed. Our objection to this axiom is stated as follows:

Proposition

Any ontology aggregator that satisfies acceptance-rejection neutrality violates exhaustiveness.

Semantic Axioms

We propose the following definition of semantic axioms that model a form of implicit knowledge.

- **Semantic Unanimity:** F is called *semantically unanimous* if $\mathrm{Cl}(O_1) \cap \cdots \cap \mathrm{Cl}(O_n) \subseteq \mathrm{Cl}(F(\mathbf{O}))$ for every profile $\mathbf{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.
- **Semantic Groundedness:** F is called *semantically grounded* if $\mathrm{Cl}(F(\mathbf{O})) \subseteq \mathrm{Cl}(O_1) \cup \cdots \cup \mathrm{Cl}(O_n)$ for every $\mathbf{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.
- **Semantic Exhaustiveness:** F is called *semantically exhaustive* if there exists no satisfiable set $\Phi \subseteq \mathrm{Cl}(O_1) \cup \cdots \cup \mathrm{Cl}(O_n)$ with $\mathrm{Cl}(F(\mathbf{O})) \subset \Phi$ for any $\mathbf{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.

Implicit vs explicit knowledge

- For most aggregators, syntactic and semantic axioms do not entail each other. E.g. The majority rule is syntactically unanimous but not semantically unanimous.
- Intuitively, semantic unanimity is the (much) stronger property. This intuition can be confirmed for the following aggregators:

Proposition

Any satisfiable and exhaustive ontology aggregator that is semantically unanimous is unanimous.

- Analogous results hold for the other axioms.
- In the next slides, we shall discuss some concrete voting procedures.

Majority rule

- Let $\mathrm{M}(\mathbf{O}) = \{\varphi \in \mathcal{L} \mid |\mathcal{N}_\varphi^{\mathbf{O}}| > \frac{n}{2}\}$ for all $\mathbf{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.
- In JA, the majority rule provides consistent outcomes on agendas that satisfy *the median property* (List and Puppe, 2009). We can refine this result as follows.
- We say that $\mathcal{L}$ satisfies the $\mathcal{T}$*-median property* if and only if for every set of ABox formulas $X \subseteq \mathcal{L}^A$ such that $\mathcal{T} \cup X$ is unsatisfiable there exists a set $Y \subseteq X$ with cardinality at most 2 such such $\mathcal{T} \cup Y$ is also unsatisfiable.
- We obtain the following characterisation:

Proposition

The majority rule will return a satisfiable ontology for any profile with a common TBox $\mathcal{T}$ if and only if the agenda $\mathcal{L}$ satisfies the $\mathcal{T}$-median property.

Quota rules

- We can generalise the idea underlying the majority rule and accept a formula for the collective ontology whenever the number of agents who do so meet a certain quota. This gives rise to the family of quota rules:

Quota rules

Let $q \in [0, 1]$. The quota rule with quota q is the ontology aggregator F_q with $F_q(\mathbf{O}) = \{\varphi \in \mathcal{L} \mid |\mathcal{N}_\varphi^{\mathbf{O}}| \geqslant q \cdot n\}$ for all $\mathbf{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$.

- The majority rule violates semantic unanimity. In fact, any quota rule does, unless we lower the quota so far as to obtain the trivial union aggregator:

Proposition

A quota rule with quota q for n agents is semantically unanimous if and only if $q \leqslant \frac{1}{n}$.

Support-based procedures

- We order the formulas in terms of the number of agents supporting them. We introduce a *priority rule* $\gg$ mapping each profile $\mathbf{O}$ to a strict linear order $\gg_{\mathbf{O}}$ on $\mathcal{L}$ such that $\varphi \gg_{\mathbf{O}} \psi$ implies $|\mathcal{N}_{\varphi}^{\mathbf{O}}| \geqslant |\mathcal{N}_{\psi}^{\mathbf{O}}|$ for all $\varphi, \psi \in \mathcal{L}$.
- *Support-based procedures*:

Given a priority rule $\gg$, the support-based procedure with $\gg$ is the ontology aggregator $\mathrm{SBP}_{\gg}$ mapping any profile $\mathbf{O} \in \mathrm{On}(\mathcal{L})^{\mathcal{N}}$ to $\mathrm{SBP}_{\gg}(\mathbf{O}) := \Phi$ for the unique set $\Phi \subseteq \mathcal{L}$ for which $\varphi \in \Phi$ if and only if

(*i*) $\mathcal{N}_{\varphi}^{\mathbf{O}} \neq \emptyset$ and

(*ii*) $\{\psi \in \Phi \mid \psi \gg_{\mathbf{O}} \varphi\} \cup \{\varphi\}$ is satisfiable.

- The SBP clearly satisfies the axioms of anonymity, monotonicity, groundedness (due to condition (*i*)), and exhaustiveness (due to condition (*ii*)). Neutrality is violated by virtue of having to fix a priority rule $\gg$

Two-stages procedures

- We may give priority to the terminology or to the assertions.
- From JA, we have the *premise-based procedure*: individuals vote on the premises by majority and then draw the conclusions, and the *conclusion-based procedure*: individuals draw their own conclusions and then votes on them by majority.
- E.g. assertion-based procedures stress the information coming from the ABox:

An (irresolute) assertion-based procedure maps each profile $\mathbf{O}$ to the set of ontologies obtained as follows:

1. Choose an aggregator F_A restricted to ABox formulas, and let $F_A(\mathbf{O})$ be the outcome.
2. Then the TBox is defined as follows:

$$F_T(\mathbf{O}) = \text{argmin}_{O \in \text{On}(\mathcal{L})} \sum_{i \in \mathcal{N}} d(F_A(\mathbf{O}) \cup O_i^T, O)$$

(d is a distance)

Conclusions and future work

- We presented a model inspired by social choice theory and judgment aggregation to define aggregation of individual ontologies for the case of the coarse merging.
- Different aggregation procedures define different notion of *group information* and the axiomatic analysis spells out the properties of such notions.
- We presented our analysis distinguishing between implicit and explicit knowledge by the distinction between semantic and syntactic axioms.
- We introduced and discussed voting rules and properties of aggregators with the aim of balancing the satisfiability and fairness.

Future works include:

- Using social choice theory methods for fine merging; e.g. distance-based procedures can potentially be adapted to deal with *concept merging*.
- Ontologies suggest a rich notion of agent, they allow for representing the preferences an agent might have over a set of alternatives together with her criteria for choosing. In this sense, our approach to ontology aggregation can lead to structured aggregation problems.