Aggregating Opinions in Abstract Argumentation

Richard Booth
Cardiff University, UK

1 Introduction and Background

The problem of *judgment aggregation* (JA), i.e., the problem of aggregating the opinions of a group of agents over a set of logically inter-connected propositions, has received quite a lot of attention recently from researchers in multi-agent systems, philosophy and economics. In this talk we report recent and ongoing work on a version of this problem in the setting of abstract argumentation.

We assume a fixed set $Ag = \{1, \ldots, n\}$ of agents, who are evaluating arguments of a given argumentation framework (AF) $\mathcal{A}$. Each evaluation takes the form of an $\mathcal{A}$-*labelling*, i.e., an assignment of one of the labels `in` (denoting *accepted*), `out` (*rejected*) or `undec` (*undecided*) to each argument of $\mathcal{A}$. Not all possible $\mathcal{A}$-labellings are *feasible*. The feasible labellings are the ones that conform to the particular argumentation *semantics* under consideration. This could be, for example, the *complete*, *stable* or *preferred* semantics, etc. We assume each agent submits a feasible $\mathcal{A}$-labelling. In this talk we will just use complete semantics, because that has been the focus of our work until now.

Given an AF $\mathcal{A} = \langle Args, \rightharpoonup \rangle$, a complete $\mathcal{A}$-labelling is a function $L : Args \to \{\texttt{in}, \texttt{out}, \texttt{undec}\}$ that satisfies, for all $a, b \in Args$:

- $L(a) = \texttt{in}$ iff $L(b) = \texttt{out}$ for all $b \in Args$ s.t. $b \rightharpoonup a$.
- $L(a) = \texttt{out}$ iff $L(b) = \texttt{in}$ for some $b \in Args$ s.t. $b \rightharpoonup a$.

We will sometimes also talk about *admissible* $\mathcal{A}$-labellings, i.e, labellings that are required to satisfy only the left-to-right directions of the above two conditions.

An $\mathcal{A}$-profile $\mathbf{L} = (L_1, \ldots, L_n)$ is a sequence of complete $\mathcal{A}$-labellings, one for each agent. Our aim is to define, in a principled manner, a concrete *aggregation method* for the agents.

Definition 1 *A* (resolute) aggregation method *is a function F that assigns, to every AF $\mathcal{A}$ and every $\mathcal{A}$-profile $\mathbf{L}$, an $\mathcal{A}$-labelling $F_{\mathcal{A}}(\mathbf{L})$.*

$F_{\mathcal{A}}(\mathbf{L})$ represents the $\mathcal{A}$-labelling of the group, given each agent i submits L_i.

2 Postulates for Aggregation Methods

What properties do we want our aggregation methods to satisfy? A most basic requirement is that we want the output to be feasible[1]:

Collective Completeness $F_{\mathcal{A}}(\mathbf{L})$ is a complete $\mathcal{A}$-labelling.

Compatibility enforces a level of consistency of the output with each of the agents' individual labellings. Here, for any label `x`, $\neg$`x` denotes `in` if `x` = `out`, `out` if `x` = `in`, and `undec` otherwise.

Compatibility For all $i \in Ag$ and $a \in Args_{\mathcal{A}}$ we have $[F_{\mathcal{A}}(\mathbf{L})](a) = \neg L_i(a)$ implies $[F_{\mathcal{A}}(\mathbf{L})](a) = \texttt{undec}$.

A weakening of *Compatibility* is the following.

`in`/`out`**-Plurality** If `x` is the `in`/`out`-loser in $(L_i(a))_{i \in Ag}$ then $[F_{\mathcal{A}}(\mathbf{L})](a) \neq \texttt{x}$

Here, the `in`/`out`-loser (resp. winner) of a given tuple of labels is that label among $\{\texttt{in}, \texttt{out}\}$ that appears the fewer (resp. more) number of times.

The next property says that the collective labelling of a given argument a is independent of whichever other arguments might be present or absent in the given AF. Here $\mathbf{L}[A]$ denotes the restriction of a given profile $\mathbf{L}$ to just the arguments in A for any $A \subseteq Args_{\mathcal{A}}$.

AF-Independence If $\mathbf{L}_1$ and $\mathbf{L}_2$ are profiles over $\mathcal{A}_1$ and $\mathcal{A}_2$ respectively and $a \in Args_{\mathcal{A}_1} \cap Args_{\mathcal{A}_2}$ then $\mathbf{L}_1[a] = \mathbf{L}_2[a]$ implies $[F_{\mathcal{A}_1}(\mathbf{L}_1)](a) = [F_{\mathcal{A}_2}(\mathbf{L}_2)](a)$.

AF-Independence turns out to be incompatible with *Collective Completeness* in the presence of a couple more mild requirements [Booth *et al.*, 2014]. A weaker version, that is arguably more appropriate anyway in an argumentation setting, is:

Directionality Suppose $\mathcal{A} \subseteq_{\mathrm{f}} \mathcal{A}'$ and suppose $Args_{\mathcal{A}}$ is unattacked in $\mathcal{A}'$. Then for any $\mathcal{A}'$-profile $\mathbf{L}$ and $a \in Args_{\mathcal{A}}$ we have $[F_{\mathcal{A}'}(\mathbf{L})](a) = [F_{\mathcal{A}}(\mathbf{L}[Args_{\mathcal{A}}])](a)$.

3 Families of Aggregation Methods

3.1 Interval Methods

In *interval aggregation methods* [Booth *et al.*, 2014], the collective label for $a \in Args_{\mathcal{A}}$ is taken to be the `in`/`out`-winner `x` in $\mathbf{L}[a]$, *provided that* the victory of `x` over $\neg$`x` is "sufficiently decisive". Otherwise we just take `undec`. Formally, let $Int_n = \{(k, l) \mid k < l,\ k, l \in \{0, 1, \ldots, n\}\}$. Let

[1] These and other postulates are discussed in [Booth *et al.*, 2014].

$Y \subseteq Int_n$ be such that $(0, n) \in Y$. Then we define aggregation method F^Y by setting, for each $\mathcal{A}$, $\mathcal{A}$-labelling profile $\mathbf{L}$ and $a \in Args_{\mathcal{A}}$:

$$[F^Y_{\mathcal{A}}(\mathbf{L})](a) = \begin{cases} \mathtt{x} & \text{if } \mathtt{x} \in \{\mathtt{in}, \mathtt{out}\} \text{ and} \\ & \quad (|V^{\mathbf{L}}_{a:\neg\mathtt{x}}|, |V^{\mathbf{L}}_{a:\mathtt{x}}|) \in Y \\ \mathtt{undec} & \text{otherwise} \end{cases}$$

Here, $V^{\mathbf{L}}_{a:\mathtt{x}}$ denotes the set of agents who labelled argument a with x. If Y satisfies [$(k, l) \in Y$ implies $(s, t) \in Y$ whenever $s \leqslant k$ and $l \leqslant t$] then F^Y is a *widening interval method*. If Y satisfies [$(k, l) \in Y$ implies $k = 0$] then F^Y is *zero-based*. As special cases of both of these classes we have the *sceptical initial* aggregation method, for which $Y = \{(0, n)\}$, and the *credulous initial* aggregation method, for which $Y = \{(0, l) \mid l > 0\}$ [Caminada and Pigozzi, 2011].

Interval methods have been axiomatised in [Booth *et al.*, 2014]. Their characteristic postulates include *AF-Independence* and in/out-*Plurality*. Also the widening and zero-based interval methods have been characterised. An interval method is zero-based iff it satisfies *Compatibility*.

3.2 DAUC Interval Methods

Interval methods don't satisfy *Collective Completeness*. To remedy this we use the *down-admissible* and *up-complete* procedures to *repair* $F^Y(\mathbf{L})$ and make it complete. The down-admissible labelling $\downharpoonleft L$ of an arbitrary labelling L is obtained by relabelling every illegally in- or out-labelled argument with undec. The resulting labelling will be admissible but not necessarily complete. The up-complete labelling $\upharpoonleft L$ of a given admissible labelling can be obtained by relabelling every illegally undec-labelled argument with in or out as appropriate. (See [Caminada and Pigozzi, 2011] for details.)

The *DAUC version* of an interval method F^Y is obtained by taking as output, for any given AF $\mathcal{A}$ and $\mathcal{A}$-profile $\mathbf{L}$, $\upharpoonleft\downharpoonleft F^Y_{\mathcal{A}}(\mathbf{L})$. *Collective Completeness* is guaranteed by construction. DAUC interval methods don't satisfy *AF-Independence* but they satisfy instead *Directionality*. Surprisingly, in/out-*Plurality* holds only if F^Y is zero-based [Booth *et al.*, 2014]. In [Caminada and Booth, 2016] we outline *discussion-based procedures* for determining whether an argument should be labelled in by $\upharpoonleft\downharpoonleft F^Y_{\mathcal{A}}(\mathbf{L})$ for the special cases in which F^Y is the sceptical initial and credulous initial methods.

3.3 Partial Resolution-based Methods

Another way to use a family of intervals Y to define an aggregation method is the *partial resolution* approach [Baroni *et al.*, 2011]. Let's say $a \in Args_{\mathcal{A}}$ is Y*-supported in* $\mathbf{L}$ if $(|V^{\mathbf{L}}_{a:\mathtt{out}}|, |V^{\mathbf{L}}_{a:\mathtt{in}}|) \in Y$. Then modify $\mathcal{A}$ by resolving any mutual attack in favour of Y-supported arguments, i.e., if $a \rightleftharpoons b$, remove one of these attacks if it goes from an un-Y-supported argument to a Y-supported one. Then $G^Y_{\mathcal{A}}(\mathbf{L})$ is defined by taking the *grounded* labelling of this modified AF. The resulting method G^Y satisfies *Collective Completeness*. Furthermore if F^Y is widening then F^Y and G^Y give the same results when restricting to the class of *symmetric* AFs (i.e., AFs for which every attack is mutual).

3.4 Distance-based Approaches

Another approach to aggregation is to base it on some notion of *distance* $d(L_1, L_2)$ between labellings. For instance, given such a distance, we could take the result of aggregation to be that complete labelling that is *closest* to the group. Of course it may well turn out that we end up with several closest labellings, which means the output will be a *set* of labellings rather than a single one. We also need to define an appropriate distance measure. One possibility here is the *issue-based distance* [Booth *et al.*, 2012]. Roughly, the arguments in an AF can be partitioned into a set of *issues*, i.e., arguments whose labels are either always the same or always opposite (in vs out) in every complete labelling. The distance $d(L_1, L_2)$ is then determined by taking one representative a from each issue and summing $\mathit{diff}(L_1(a), L_2(a))$ over all these representatives, where diff is some measure of difference between *labels* (e.g., $\mathit{diff}(\mathtt{in}, \mathtt{out}) = 2$, $\mathit{diff}(\mathtt{in}, \mathtt{undec}) = \mathit{diff}(\mathtt{out}, \mathtt{undec}) = 1$, etc.). By summing just over these representatives rather than over all arguments we avoid *double counting*, i.e., we disregard some differences in the labelling that might already be necessary consequences of some other differences. An initial study of the resulting aggregation methods can be found in [Podlaszewski, 2015].

4 Further and Ongoing Work

We are in the process of expanding the results not just to AFs but also to *abstract dialectical frameworks (ADFs)*. Indeed DAUC interval methods have been extended to ADFs in [Booth, 2015]. Secondly, it remains to explore the precise relation with JA: does our setting subsume the JA one, or the other way around? Finally we'd like to fully characterise the families of aggregation methods described here.

References

[Baroni *et al.*, 2011] Pietro Baroni, Paul E Dunne, and Massimiliano Giacomin. On the resolution-based family of abstract argumentation semantics and its grounded instance. *Artificial Intelligence*, 175(3):791–813, 2011.

[Booth *et al.*, 2012] Richard Booth, Martin Caminada, Mikołaj Podlaszewski, and Iyad Rahwan. Quantifying disagreement in argument-based reasoning. In *Proc. AAMAS 2012*, 2012.

[Booth *et al.*, 2014] Richard Booth, Edmond Awad, and Iyad Rahwan. Interval methods for judgment aggregation in argumentation. In *Proc. KR 2014*, 2014.

[Booth, 2015] Richard Booth. Judgment aggregation in abstract dialectical frameworks. In *Advances in Knowledge Representation, Logic Programming, and Abstract Argumentation*, pages 296–308. Springer, 2015.

[Caminada and Booth, 2016] Martin Caminada and Richard Booth. A dialectical approach for argument-based judgment aggregation. In *Proc. COMMA 2016*, 2016.

[Caminada and Pigozzi, 2011] Martin Caminada and Gabriella Pigozzi. On judgment aggregation in abstract argumentation. *Autonomous Agents and Multi-Agent Systems*, 22(1):64–102, 2011.

[Podlaszewski, 2015] Mikołaj Podlaszewski. *Poles Apart: Navigating the Space of Opinions in Argumentation*. PhD thesis, University of Luxembourg, Luxembourg, 2015.