A target-oriented discussion framework to support collective decision making*

Jordi Ganzer-Ripoll[1], Maite Lopez-Sanchez[1], Juan Antonio Rodriguez-Aguilar[2]

[1] Departament de Matemàtiques i Informàtica, Facultat de Matemàtiques i Informàtica, Universitat de Barcelona, Spain
[2] IIIA, Artificial Intelligence Research Institute
CSIC, Spanish National Research Council
jordi891@gmail.com, maite_lopez@ub.edu, jar@iiia.csic.es

Abstract. Argumentative debates are a powerful tool for resolving conflicts and reaching agreements in open environments such as on-line communities. Here we introduce an argumentation framework to structure argumentative debates. Our framework represents the arguments issued by the participants involved in a debate, the (attack and defence) relationships between them, as well as participants' opinions on them. Furthermore, we tackle the problem of computing a collective decision from participants' opinions. With this aim, we design an aggregation function to ensure that participants reach a coherent collective decision.

1 Introduction

As argued in [10,11], argumentative debates are a powerful tool for reaching agreements in open environments such as on-line communities. Nowadays, this is particularly true in our society due to the increasing interest and deployment of e-participation and e-governance ICT-systems that involve citizens in governance [14]. Not surprisingly some European cities are opening their policy making to citizens (e.g., Reykjavík [2], Barcelona [1]). Moreover, the need for argumentative debates has also been deemed as necessary for open innovation systems [12]. On-line debates are usually organised as threads of arguments and counter-arguments that users issue to convince others so that debates eventually converge to agreements. Users are allowed to express their opinions on arguments by rating them (e.g., [11]). There are two main issues in the management of large-scale on-line debates. First, as highlighted by [10] and [11], there is simply too much noise when many individuals participate in a discussion, and hence there is the need for *structuring* it to keep the focus. Second, the opinions on arguments issued by users must be aggregated to achieve a collective decision about the topic under discussion [4]. In this paper we try to make headway on these two issues.

Recently, argumentation has become one of the key approaches to rational interaction in artificial intelligence [5,13]. Here, we propose to follow an argumentation-based approach that allows agents to issue arguments in favour of or against a *topic* under discussion as well as about other agents' arguments. Furthermore, we will consider that agents express their opinions about each other's arguments and the topic itself.

* Funded by Collectiveware TIN2015-66863-C2-1-R (MINECO/FEDER) and 2014 SGR 118.

Within our multi-agent framework, we face the following collective decision problem: *given a set of agents, each with an individual opinion about a given set of arguments related to a topic, how can agents reach a collective decision on the topic under discussion?* To solve this problem, we propose: 1) A novel multi-agent argumentation framework, the so-called *target-oriented discussion framework*, to support discussions about the acceptance of a target proposal; and 2) A social choice function that aggregates agents' opinions to infer the overall opinion about the topic under discussion. Our aggregation function is based on combining opinions and exploiting dependencies between arguments to produce an aggregated opinion. Moreover, and most importantly, our aggregation function guarantees the resulting aggregated opinion to be *coherent*, naely free of contradictions.

1.1 Example

Next, we introduce a simple example to illustrate some of the presented concepts along the paper.

Example 1 (Flatmates' discussion) *Consider three flatmates (Alan, Bart, and Cathy) discussing about norm (N): "Flatmates take fixed turns for dishwashing at 10 p.m." and issuing the following arguments: a_1 ="10 p.m. is too late and should be changed"; a_2 ="Schedule is too rigid"; and a_3 ="Fair distribution". Notice that: arguments a_1 and a_2 attack N whereas a_3 defends it; and a_1 is in favour of a_2. Once all arguments and their relations are clear, flatmates express their opinions by accepting, rejecting (or not opining about) each argument : (1) Alan (Ag_1) gets up early 4 days per week, and so (as first row in Table 1 shows) he rejects norm N and accepts arguments a_1 and a_2. Nevertheless, he acknowledges and accepts argument a_3. (2) Bart (Ag_2) has spare time at night and is clearly pro norm N. Second row in Table 1 shows he accepts N and a_3, and rejects a_1 and a_2. Finally, (3) Cathy (Ag_3) is keen on routines so she rejects a_2 and accepts N, a_1, and a_3 (see third row in Table 1).*

		Arguments			
		N	a_1	a_2	a_3
Agents	Ag 1	✗	✓	✓	✓
	Ag 2	✓	✗	✗	✓
	Ag 3	✓	✓	✗	✓

Table 1. Flatmates' opinions in the discussion on the dish-washing norm.

Therefore, the question that arises is how to aggregate all these opinions so that a consensus is reached over the acceptance (or not) of this dish-washing norm.

2 The target-oriented discussion framework

The purpose of this section is to formally capture all the core elements of our argumentation framework.

2.1 Formalising our argumentation framework

Our purpose is to provide an argumentation framework that allows one to capture both attack and defence relationships between arguments, as done in bipolar argumentation frameworks [8,3].[3] The motivation for including defence relationships is based on recent studies in large-scale argumentation frameworks involving humans (e.g., [12,11]). There, humans naturally handle both attack and defence relationships between arguments. Our notion of *discussion framework* aims at offering such expressiveness.

Definition 1 *A discussion framework is a triple $DF = \langle \mathcal{A}, \mapsto, \Vdash \rangle$, where $\mathcal{A}$ is a finite set of arguments, and $\mapsto \subseteq \mathcal{A} \times \mathcal{A}$ and $\Vdash \subseteq \mathcal{A} \times \mathcal{A}$ stand for attack and defence relationships that are disjoint, namely $\mapsto \cap \Vdash = \emptyset$. We say that an argument $b \in A$ attacks another argument $a \in A$ iff $b \mapsto a$, and that b defends a iff $b \Vdash a$.*

A discussion framework can be depicted as a graph whose nodes stand for arguments and whose edges represent either attack or defence relationships between arguments. Figure 1 shows our graphical representation of attack and defence relationships.

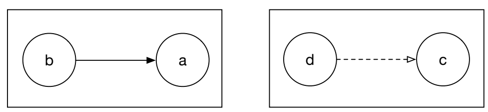

Fig. 1. Representation of an attack relationship $b \mapsto a$ and a defence relationship $d \Vdash c$.

.

Definition 2 *Let $DF = \langle \mathcal{A}, \mapsto, \Vdash \rangle$ be a discussion framework and $a \in A$ one of its arguments. We say that an argument $b \in \mathcal{A}$ is a descendant of a if there is a finite subset of arguments $\{c_1, \cdots, c_r\} \subseteq \mathcal{A}$ such that $b = c_1$, $c_1 R_1 c_2$, $\cdots$, $c_{r-1} R_{r-1} c_r$, $c_r = a$ and $R_i \in \{\mapsto, \Vdash\}$ for all $1 \leq i < r$.*

Definition 3 *A target-oriented discussion framework $TODF = \langle \mathcal{A}, \mapsto, \Vdash, \tau \rangle$ is a discussion framework satisfying the following properties: (i) for every argument $a \in \mathcal{A}$, a is not a descendant of itself; and (ii) there is an argument $\tau \in \mathcal{A}$, called the target, such that for all $a \in \mathcal{A} \setminus \{\tau\}$, a is a descendant of τ.*

Observation 1 *From the previous definitions we infer some properties that help us further characterise a target-oriented discussion framework: Attack and defence relations are irreflexive and non-reciprocal. Moreover, the target neither attacks nor defends any other argument. This distinguishes the special role of the target as the center of discussion to which attacks and supports are directly or indirectly pointed.*

Proposition 4 *Let $TODF = \langle \mathcal{A}, \mapsto, \Vdash, \tau \rangle$ be a target-oriented discussion framework and $E = \mapsto \cup \Vdash$. The graph associated with a TODF, $G = \langle \mathcal{A}, E \rangle$, is a directed acyclic graph, where $\mathcal{A}$ is the set of nodes and E the edge relationship.*

[3] Nevertheless, there are notable differences, e.g., bipolar argumentation frameworks do not consider labellings (different opinions on arguments), nor their aggregation.

Proof. *Straightforward from definition 2 and observation 1.*

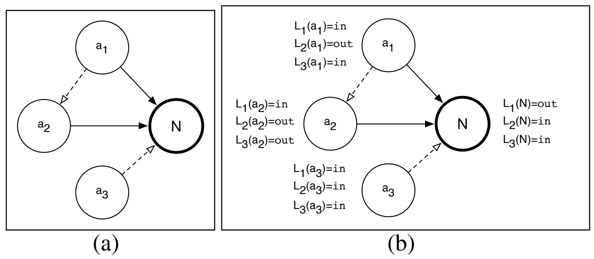

Fig. 2. Flatmates example: (a) TODF's associated graph; (b) $TODF$ together with labellings.

Example 2 (Flatmates' example formalization) *Figure 2(a) depicts the flatmates' target-oriented discussion framework. The nodes in the graph represent the set of arguments* $\mathcal{A} = \{N, a_1, a_2, a_3\}$ *in the example of previous section, where* N *is the dish-washing norm, and* a_1, a_2, a_3 *are the rest of arguments. Thus,* N*, the norm under discussion, is taken to be* τ *in our* $TODF$*. As to edges, they represent both the attack and defence relationships:* $a_1 \mapsto N$*,* $a_2 \mapsto N$ *and* $a_1 \Vdash a_2$*,* $a_3 \Vdash N$ *respectively.*

2.2 Argument labellings

Agents encode their opinions about arguments through *labellings* [6,7]. An agent expresses its support to an argument by labelling it as *in*, rejects it with *out* labels, and abstains from deciding whether to accept it or reject it by labelling it as *undec*. This *undec* label also stands for the absence of an opinion.

Definition 5 (Argument labelling) *Let* $TODF = \langle \mathcal{A}, \mapsto, \Vdash, \tau \rangle$ *be a target-oriented discussion framework. An argument labelling for* $TODF$ *is a function* $L : \mathcal{A} \longrightarrow \{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}$ *that maps each argument of* $\mathcal{A}$ *to one out of the following labels:* `in` *(accepted),* `out` *(rejected), or* `undec` *(undecidable).*

We note as $Ag = \{ag_1, \ldots, ag_n\}$ the set of agents taking part in a $TODF$, and as L_i the labelling encoding the opinion of agent $ag_i \in Ag$. We will put together the opinions of all the agents participating in an argumentation as follows.

Definition 6 (Labelling profile) *Let* $L_1, \ldots, L_n$ *be argument labellings of the agents in* Ag*, where* L_i *is the argument labelling of agent* ag_i*. A labelling profile is a tuple* $\mathcal{L} = (L_1, \ldots, L_n)$*.*

Example 3 (Flatmates' opinions) *Figure 2(b) graphically depicts Alan's, Bart's, and Cathy's labellings (noted as* L_1, L_2, L_3 *respectively), each one appearing next to the corresponding arguments in the* $TODF$*'s graphical representation in Figure 2(a).*

2.3 Coherent argument labellings

Given an argument a, we will define:

- its set of attacking arguments as $A(a) = \{b \in \mathcal{A} | b \mapsto a\}$, and
- its set of defending arguments as: $D(a) = \{c \in \mathcal{A} | c \Vdash a\}$

Thus, the labelling of arguments in $A(a) \cup D(a)$ compose the indirect opinion on a.

Given an argument labelling L and a set of arguments $S \subseteq A$, we can quantify the number of accepted arguments in S as:

$$\mathtt{in}_L(S) = |\{b \in S \,| L(b) = \mathtt{in}\}|$$

and the number of rejected arguments in S as:

$$\mathtt{out}_L(S) = |\{b \in S \,| L(b) = \mathtt{out}\}|$$

Thus, given an argument a, we can readily quantify its accepted and rejected defending arguments as $\mathtt{in}_L(D(a))$ and $\mathtt{out}_L(D(a))$ respectively. Moreover, we can also quantify its accepted and rejected attacking arguments as $\mathtt{in}_L(A(a))$ and $\mathtt{out}_L(A(a))$ respectively. Now we are ready to measure the *positive* and *negative support* contained in the indirect opinion of a given argument as follows.

Definition 7 (Positive support) *Let $a \in \mathcal{A}$ be an argument and L a labelling on $\mathcal{A}$. We define the* positive (pro) support *of a as: $Pro_L(a) = \mathtt{in}_L(D(a)) + \mathtt{out}_L(A(a))$.*

Definition 8 (Negative support) *Let $a \in \mathcal{A}$ be an argument and L a labelling on $\mathcal{A}$. We define the* negative (con) support *of a as: $Con_L(a) = \mathtt{in}_L(A(a)) + \mathtt{out}_L(D(a))$.*

Notice that the positive support of an argument combines the strength of its accepted defending arguments with the weakness of its rejected attacking arguments in the argument's indirect opinion. As a dual concept, the negative support combines accepted attacking arguments with rejected defending arguments.

We now introduce our notion of coherence by combining the positive and negative support of an argument. We say that a labelling is coherent if the following conditions hold for each argument: (1) if an argument is labelled accepted (in) then it cannot have more negative than positive support (the majority of its indirect opinion supports the argument); and (2) if an argument is labelled rejected (out) then it cannot have more positive than negative support (the majority of its indirect opinion rejects the argument).

Definition 9 (Coherence) *Given a $TODF = \langle \mathcal{A}, \mapsto, \Vdash, \tau \rangle$, a* coherent labelling *is a total function $L : \mathcal{A} \to \{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}$ such that for all $a \in \mathcal{A}$ with $A(a) \cup D(a) \neq \emptyset$ it satisfies: i) $L(a) = \mathtt{in} \implies Pro_L(a) \geq Con_L(a)$; and ii) $L(a) = \mathtt{out} \implies Pro_L(a) \leq Con_L(a)$.*

Example 4 *Again, considering our example and its labellings from Figure 2(b) (L_1, L_2, L_3), we note that just L_1, L_2 belong to the subclass of its coherent argument labellings $Coh(TODF)$.*

3 The aggregation problem

Definition 10 (Labelling discussion problem) *Let $Ag = \{ag_1, \cdots, ag_n\}$ be a finite non-empty set of agents, and $TODF = \langle \mathcal{A}, \mapsto, \Vdash, \tau \rangle$ be a target-oriented discussion framework. A labelling discussion problem is a pair $\mathcal{LDP} = \langle Ag, TODF \rangle$.*

Given an $\mathcal{LDP}$, our aim is to find how to aggregate the individuals' labellings into a single labelling that captures the opinion of the collective.

Definition 11 (Aggregation function) *An aggregation function for a labelling discussion problem $\mathcal{LDP} = \langle Ag, TODF \rangle$ is a function $F : \mathbf{L}(TODF)^n \longrightarrow \mathbf{L}(TODF)$, being $\mathbf{L}(TODF)$ the class of the argument labellings of $TODF$.*

Plainly, an aggregation function F takes a labelling profile representing all agents' opinions and yields a single labelling computed from the individual labellings. Such aggregation function is key to assessing the collective decision over the target.

Definition 12 (Decision over a target) *Let $\mathcal{LDP} = \langle Ag, TODF \rangle$ be a labelling discussion problem, $\mathcal{L}$ a labelling profile, and F an aggregation function for the $\mathcal{LDP}$. The decision over the target of the $TODF$ is the label $F(\mathcal{L})(\tau)$.*

The literature on Social Choice theory has identified fair ways of adding votes. These can be translated into formal properties that an aggregation function is required to satisfy [9]. Based on [4], here we formally state what we consider to be the most desirable property for an aggregation function that allows to assess the decision over the target of a target-oriented discussion framework. Thus, we consider an aggregation function $F(\mathcal{L})$ to be **Collective coherent** (CC) *iff* $F(\mathcal{L}) \in Coh(TODF)$ for all $\mathcal{L} \in \mathbf{L}(TODF)^n$, being $Coh(TODF)$ the subclass of coherent argument labellings.

Notice that if an aggregation function does not produce a coherent labelling, there is at least some argument whose collective label (direct opinion) is in contradiction with its indirect opinion. Thus, the aggregation would not be reliable.

4 The coherent aggregation function

In order to define an aggregation function to compute the collective labelling, we first introduce notation to quantify the direct positive and negative support of an argument. Let $\mathcal{L} = (L_1, \cdots, L_n)$ be a labelling profile and a an argument. We note:

- The *direct positive support* of a as $\mathtt{in}_{\mathcal{L}}(a) = |\{ag_i \in Ag \, | L_i(a) = \mathtt{in}\}|$; and
- The *direct negative support* of a as $\mathtt{out}_{\mathcal{L}}(a) = |\{ag_i \in Ag \, | L_i(a) = \mathtt{out}\}|$.

Next, we define our chosen aggregation function: the *coherent aggregation function*. The main purpose of this function is to compute a coherent aggregated labelling, and hence fulfil the collective coherence property. that is, to yield a rational outcome that is free of contradiction.

Definition 13 (Coherent aggregation function) *Let $\mathcal{L}$ be a labelling profile. For each argument a the coherent function over $\mathcal{L}$ is defined as:*

$$CF(\mathcal{L})(a) = \begin{cases} \texttt{in} & , IO(\mathcal{L})(a) + DO(\mathcal{L})(a) > 0 \\ \texttt{out} & , IO(\mathcal{L})(a) + DO(\mathcal{L})(a) < 0 \\ \texttt{undec} & , IO(\mathcal{L})(a) + DO(\mathcal{L})(a) = 0 \end{cases}$$

where DO (direct opinion) and IO (indirect opinion) functions are defined as:

$$DO(\mathcal{L})(a) = \begin{cases} 1 & , \texttt{in}_{\mathcal{L}}(a) > \texttt{out}_{\mathcal{L}}(a) \\ 0 & , \texttt{in}_{\mathcal{L}}(a) = \texttt{out}_{\mathcal{L}}(a) \\ -1 & , \texttt{in}_{\mathcal{L}}(a) < \texttt{out}_{\mathcal{L}}(a) \end{cases}$$

If $A(a) \cup D(a) = \emptyset$ then $IO(\mathcal{L}) = 0$, Otherwise:

$$IO(\mathcal{L})(a) = \begin{cases} 1 & , Pro_{CF(\mathcal{L})}(a) > Con_{CF(\mathcal{L})}(a) \\ 0 & , Pro_{CF(\mathcal{L})}(a) = Con_{CF(\mathcal{L})}(a) \\ -1 & , Pro_{CF(\mathcal{L})}(a) < Con_{CF(\mathcal{L})}(a) \end{cases}$$

Notice that to compute our CF on a single argument a we need to compute first the CF of its descendants. The acyclic characterisation of our $TODF$ prevents endless recursion.

Example 5 (Flatmates' discussion) *Back to our example involving a flatmates' discussion, we use the coherent aggregation function to obtain the aggregated opinion of the provided labellings (see Figure 2(b)). Figure 3 shows the results of the aggregation and the decision over the target as produced by CF. We observe that the flatmates collectively accept arguments a_1 and a_3, whereas argument a_2 becomes undecidable. Finally, the decision over the norm is to accept it.*

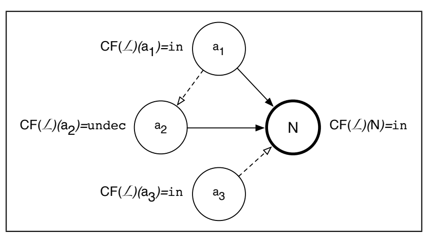

Fig. 3. Flatmates example: aggregated labellings (and decision over target N) computed by CF.

Proposition 14 *CF satisfies the collective coherence property.*

Proof. Let a be an argument such that $CF(\mathcal{L})(a) = \texttt{in}$. From Definition 13 we know that $IO(\mathcal{L})(a) + DO(\mathcal{L})(a) > 0$. Thus, there are three possibilities: (i) $DO(\mathcal{L})(a) = 1$ and $IO(\mathcal{L})(a) = 1$; (ii) $DO(\mathcal{L})(a) = 1$ and $IO(\mathcal{L})(a) = 0$; or (iii) $IO(\mathcal{L})(a) = 0$ and $DO(\mathcal{L})(a) = 1$. Since $IO(\mathcal{L})(a) \geq 0$ in all cases, this implies that $Pro_{CF(\mathcal{L})}(a) \geq Con_{CF(\mathcal{L})}(a)$, and hence CF satisfies the coherence property. The proof goes analogously for the case $CF(\mathcal{L})(a) = \texttt{out}$.

5 Conclusions and future work

This paper formalises the problem of taking collective decisions by proposing a target oriented discussion framework and a novel aggregation function that combines opinions. We show that such function satisfies coherence, a valuable social choice property.

We are currently studying other social choice properties, such as anonymity, non-dictatorship, or supportiveness. Regarding the operationalisation of our problem, we are also working on an algorithm for the computation of the decision over a target.

Finally, as for future work, we first plan to extend our Target Oriented Decision Framework (TODF) to permit loops, and hence ease rebuttal, a common feature of argumentation systems. Morevoer we will also pursue to provide more fine-grained means of computing argument support.

References

1. City of Barcelona participation portal. https://decidim.barcelona, 2016.
2. City of Reykjavík participation portal. http://reykjavik.is/en/participation, 2016.
3. Leila Amgoud, Claudette Cayrol, Marie-Christine Lagasquie-Schiex, and Pierre Livet. On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems*, 23(10):1062–1093, 2008.
4. Edmond Awad, Richard Booth, Fernando Tohmé, and Iyad Rahwan. Judgment aggregation in multi-agent argumentation. *CoRR*, abs/1405.6509, 2014.
5. Trevor J. M. Bench-Capon and Paul E. Dunne. Argumentation in artificial intelligence. *Artif. Intell.*, 171(10-15):619–641, 2007.
6. Martin Caminada. On the issue of reinstatement in argumentation. In *Logics in artificial intelligence*, pages 111–123. Springer, 2006.
7. Martin WA Caminada and Dov M Gabbay. A logical account of formal argumentation. *Studia Logica*, 93(2-3):109–145, 2009.
8. Claudette Cayrol and Marie-Christine Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In *Symbolic and quantitative approaches to reasoning with uncertainty*, pages 378–389. Springer, 2005.
9. Franz Dietrich. A generalised model of judgment aggregation. *Social Choice and Welfare*, 28(4):529–565, 2007.
10. Simone Gabbriellini and Paolo Torroni. Microdebates: Structuring debates without a structuring tool1. *AI Commun.*, 29(1):31–51, 2015.
11. Mark Klein. Enabling large-scale deliberation using attention-mediation metrics. *Computer Supported Cooperative Work*, 21(4-5):449–473, 2012.
12. Mark Klein and Gregorio Convertino. A roadmap for open innovation systems. *Journal of Social Media for Organizations*, 2(1):1, 2015.
13. Iyad Rahwan, Guillermo R. Simari, and Johan Benthem. *Argumentation in artificial intelligence*, volume 47. Springer, 2009.
14. Vishanth Weerakkody and Christopher G Reddick. *Public sector transformation through e-government: experiences from Europe and North America*. Routledge, 2012.