On the Use of Target Sets for Move Selection in Multi-Agent Debates

Dionysios Kontarinis [1] and **Elise Bonzon** [2] and **Nicolas Maudet** [3] and **Pavlos Moraitis** [4]

Abstract. In debates, agents are faced with the problem of deciding how to best contribute to the current state of the debate in order to satisfy their own goals. Target sets specify minimal changes on the current state of the debate that are required to achieve such goals, where changes are the addition and/or deletion of attacks among arguments. However, agents may not have the ability to implement all the actions prescribed by a target set, nor to rely on others to help them to do so. In this short paper we provide evidence that this notion is still a useful criterion for move selection.

1 INTRODUCTION

In recent years, the study of the collective aspects of argumentation has seen a surge of interest in AI. In on-line settings in particular, we may consider that a group of agents gradually construct a *weighted argumentation system* [6], where the weight attached to an edge reflects the number of agents who have committed to a given attack. Such settings raise new challenges for argumentation theory [10, 4, 7].

In practice such debates may be (more or less flexibly) regulated, to ensure that they remain focused, and that some fairness is guaranteed among the agents. Something missing though is a study of the dynamics of debates regulated by such protocols: it is not clear how agent strategies could change the outcome of debates. In [2] a very simple dynamic is investigated, based on a direct notion of relevance inspired by [9]. The authors exhibited in particular that in the absence of coordination and with a myopic behavior, agents can actually play against their own interest. This justifies the fact that some "guidance" might be useful to agents, without assuming any sort of explicit coordination among agents. Recently, the notion of *target sets* has been proposed in the litterature [1]. Roughly speaking, a target set specifies the minimal change necessary to achieve an argumentative goal in the debate. The intuition is that agents should be better off following their target set recommendations. One challenge though is that target sets may prescribe multiple changes, and in general it is impossible to assume that agents have the opportunity to make all these changes.

In this short paper we provide evidence that this notion is still a useful criterion for dialogue move selection.

[1] Paris Descartes University, LIPADE, France, email: dionysios.kontarinis@parisdescartes.fr
[2] Paris Descartes University, LIPADE, France, email: elise.bonzon@parisdescartes.fr
[3] UPMC, LIP6, France, email: nicolas.maudet@lip6.fr
[4] Paris Descartes University, LIPADE, France, email: pavlos@mi.parisdescartes.fr

2 ARGUMENTATIVE DEBATES

In this work we study the dynamics of argumentative debates, focusing on minimal change achieving a goal. The setting of the debate is the following: There is an arbitrary, finite set of participating agents, denoted Ag. Each agent $i \in Ag$ has a private Dung **argumentation system** [5], denoted $AS_i = \langle A, R_i \rangle$. For the sake of simplicity, we assume that all agents share the same set of arguments A, but they may disagree on the validity of the attacks between those arguments. The debate focuses on the status (w.r.t. a given semantics) of a single argument, called the $issue$. The agents use a common system called $Gameboard$ (GB in short), where they play *moves*. A move is a vote on some attack of the GB, and it can be positive (if the agent believes the attack is valid) or negative (otherwise). According to the votes cast, some attacks will be collectively considered valid, while others invalid. Every agent's objective is to have, at the end of the debate, the issue's status on the Gameboard coincide with the issue's status in his private system.

In Dung's framework, the *acceptability of an argument* depends on its membership to some sets, called extensions. Several *acceptability semantics* have been defined [5] in [5]. In what follows, we will say that an argument $a \in A$ is **credulously accepted** (resp. **sceptically accepted**) w.r.t. system AS under semantics S, denoted $S_\exists(a, AS)$ (resp. $S_\forall(a, AS)$), iff a belongs to at least one (resp. to every) extension of AS under the S semantics.

The GB stores all the opinions expressed by the agents during the debate and aggregates them. We will not focus on how the agents' opinions are gathered and aggregated: The debate is typically regulated by some protocol, while the agents' expertise may play a role in the aggregation of opinions. What is of interest for us in this paper is that the GB can give rise to a single argumentation system, which will allow us to draw the debate's conclusions. To keep notation simple, we also denote GB the argumentation system which contains, at every time during the debate, all the attacks which are collectively considered valid (*e.g.* attacks supported by a majority). Note that a single move made by an agent may not directly affect the argumentation system of the GB.

3 FOCUS ON TARGET SETS

At this point we turn our attention to possible strategic considerations of agents in this type of debates. Which are the attacks of the gameboard the voting agents should focus on and try to add/remove? The aim of the following analysis is to provide insight on how to vote in order to achieve a goal. We focus on target sets [1], which represent the minimal change on an argumentation system, achieving a goal.

[5] Due to lack of space, we do not detail these well known semantics here.

<table>
<tr><td rowspan="4">$GB' = \Delta(GB, m)$, and no action of m belongs in a target set of GB</td><td rowspan="2">$\forall t' \in \mathbb{T}(GB')$</td><td>$\nexists t \in \mathbb{T}(GB)$ such that $t' \subset t$</td><td>Property 1</td></tr>
<tr><td>$\exists t \in \mathbb{T}(GB)$ such that $t \subseteq t'$</td><td>Property 2</td></tr>
<tr><td>$\forall t \in \mathbb{T}(GB)$</td><td>$\exists t' \in \mathbb{T}(GB')$ such that $t \subseteq t'$</td><td>Property 3</td></tr>
<tr><td colspan="2">$|\mathbb{T}(GB')| \geq |\mathbb{T}(GB)|$</td><td>Property 4</td></tr>
</table>

Table 1. Properties of playing outside target sets

3.1 Basic definitions of target sets

A target set is a minimal set of *actions* on an argumentation system allowing to achieve a given argumentative goal. We assume argumentative goals are related to the issue of the debate (denoted d).

An **action on** GB is a tuple $((x, y), s)$ with $(x, y) \in A \times A$ and $s \in \{+, -\}$, which stands respectively for adding or removing an attack on the GB. The **resulting GB** after playing a set of actions m, is denoted as $\Delta(GB, m) = \langle A, R^m \rangle$. A **goal** $g = S_X(d)$ (resp. $\neg S_X(d)$) is **satisfied** in GB iff $S_X(d, GB)$ holds (resp. if $S_X(d, GB)$ does not hold) [6]. Finally, m is a **successful set of actions for goal** g iff g is satisfied in $\Delta(GB, m)$. We denote by $\mathbb{M}(GB)$ the set of all successful sets of actions. m is a **target set for goal** g iff m is a $\subseteq$-minimal set of actions from $\mathbb{M}(GB)$. We denote by $\mathbb{T}(GB)$ the set of all target sets.

We note that the computation of such modifications in an argumentation system has recently been studied [8, 3]. Here we instead analyze the evolution of target sets when it cannot be assumed that an agent can play all the actions prescribed by a target set.

3.2 Playing outside target sets

We first turn our attention into what happens if we play a set of actions which contains no action of any target set, for a given goal. Intuitively, in that case we will not get closer to the goal. Moreover, such an action could even lead us farther away from the goal, in the sense that even more actions will be needed in order to achieve it.

Let m be a set of actions on a gameboard GB such that *m does not contain any action of any target set of GB*. After playing m, the goal remains unsatisfied in the resulting gameboard $GB' = \Delta(GB, m)$,[7] while the set of target sets changes and is denoted $\mathbb{T}(GB')$. The properties obtained when such a set of actions m is played are illustrated in Table 1. The proofs are omitted due to lack of space.

Property 1 states that, for every new target set, there is no old target set which is a strict superset of the new one. Thus, no target set "shrinks" when playing m. Property 2 states that, for every new target set, there exists an old target set which is a subset of it. Property 3 states that, for every old target set, there exists a new target set which is a superset of the old one. Thus, no target set "disappears" when playing m. Finally, Property 4 states that the cardinality of the new set of target sets is greater or equal to the cardinality of the old set of target sets.

Thus, if a set of actions m which does not contain any action of any target set of $\mathbb{T}(GB)$ is played, then according to Property 2, every new target set in $\mathbb{T}(GB')$ will be bigger than some old target set of $\mathbb{T}(GB)$. In that sense, it will become harder (or at least not easier) to satisfy the goal under consideration, even if the cardinality of the set of target sets may grow.

[6] Where $X \in \{\exists, \forall\}$.

[7] Because if m was successful, then a subset of m would be a target set. Impossible, since m contains no actions of any target set.

3.3 Playing in target sets

What happens if we play a set of actions m belonging to a target set? Let $t \in \mathbb{T}(GB)$ and let $m \subset t$. If we play m, then $t \setminus m$ will become a target set of the new gameboard $\Delta(GB, m)$. However, we are uncertain about the other target sets of GB: some may shrink too, but others may remain unchanged, or even grow. Moreover, the cardinality of $\mathbb{T}(GB)$ could decrease, remain unchanged, or increase in $\Delta(GB, m)$. Therefore, playing in a target set shrinks (at least) that target set, regardless of what happens to the other target sets. In that sense, at least one "path" towards the satisfaction of the goal becomes shorter, whereas this is not the case if we play outside target sets.

4 CONCLUSION

We provide some evidence that target sets are legitimate to be used for move selection, even when agents cannot play the whole set of actions prescribed. However, the limit of this study is that "playing in target sets" can mean many different things: there are typically various options, which may lead to different strategies. Is it always better to play in the smallest target set for instance? We plan to conduct an experimental study of different strategies based on target sets.

REFERENCES

[1] G. Boella, D. Gabbay, A. Perotti, L. van der Torre, and S. Villata, 'Conditional labelling for abstract argumentation', in *TAFA'11*, (2011).

[2] E. Bonzon and N. Maudet, 'On the outcomes of multiparty persuasion', in *Proc. of AAMAS'11*, pp. 47–54, (May 2011).

[3] S. Coste-Marquis, S. Konieczny, J.-G. Mailly, and P. Marquis, 'On the revision of argumentation systems: minimal change of arguments statuses', in *KR'14*, (2014).

[4] S. Coste-Marquis, S. Konieczny, P. Marquis, and M.A. Ouali, 'Weighted attacks in argumentation frameworks', in *Proc. of KR'12*, (2012).

[5] P. M. Dung, 'On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games', *Artificial Intelligence*, **77**(2), 321–358, (1995).

[6] P. E. Dunne, A. Hunter, P. McBurney, S. Parsons, and M. Wooldridge, 'Weighted argument systems: Basic definitions, algorithms, and complexity results', *Artif. Intell.*, **175**(2), 457–486, (2011).

[7] S. Egilmez, J. Martins, and J. Leite, 'Extending social abstract argumentation with votes on attacks', in *proc. of TAFA'13*, pp. 16–31, (2013).

[8] D. Kontarinis, E. Bonzon, N. Maudet, A. Perotti, L. van der Torre, and S. Villata, 'Rewriting rules for the computation of goal-oriented changes in an argumentation system', in *CLIMA'13*, pp. 51–68, (2013).

[9] H. Prakken, 'Coherence and flexibility in dialogue games for argumentation', *Journal of Logic and Computation*, **15**, 347–376, (2005).

[10] F. Toni and P. Torroni, 'Bottom-up argumentation', in *TAFA'11*, pp. 249–262, (2011).