A Utility and Information Based Heuristic for Argumentation[1]

Nir Oren and **Timothy J. Norman** and **Alun Preece**[2]

Abstract. While researchers have looked at many aspects of argumentation, an area often neglected is that of argumentation strategies. That is, given multiple possible arguments that an agent can put forth, which should be selected in what circumstances. In this paper, we propose a heuristic that implements one such strategy. The heuristic assigns a utility cost to revealing information, as well as a utility to winning, drawing and losing an argument. An agent participating in a dialogue then attempts to maximise its utility. After informally presenting the heuristic, we discuss some of its novel features, after which some avenues for future work are examined.

1 Introduction

Argumentation has emerged as a powerful reasoning mechanism in many domains. One common dialogue goal is to persuade, where one or more participants attempt to convince the others of their point of view. This type of dialogue can be found in many areas including distributed planning and conflict resolution, education and in models of legal argument. At the same time that the breadth of applications of argumentation has expanded, so has the sophistication of formal models designed to capture the characteristics of the domain. While many researchers have focused on the question of "what are the properties of an argument", fewer have looked at "how does one argue well".

In this paper, we propose a decision heuristic for an agent allowing it to decide which argument to advance. The basis for our idea is simple; the agent treats some parts of its knowledge as more valuable than other parts, and, while attempting to win the argument, attempts to minimise the amount of valuable information it reveals. This heuristic often emerges in negotiation dialogues, as well as persuasion dialogues in hostile setting (such as takeover talks or in some legal cases). Utilising this heuristic in arguments between computer agents can also be useful; revealing confidential information in an ongoing dialogue may damage an agent's chances of winning a future argument.

In the remainder of this paper, we will briefly describe the framework, provide an example as to its functioning, and then examine its features in more detail and look at possible extensions to our approach. First however, we will examine a number of existing approaches to strategy selection.

2 Background and related research

Argumentation researchers have recognised the need for argument selection strategies for a long time. However, the field has only recently started receiving more attention. Moore, in his work with the DC dialectical system [8], suggested that an agent's argumentation strategy should take three things into account:

- Maintaining the focus of the dispute.
- Building its point of view or attacking the opponent's one.
- Selecting an argument that fulfils the previous two objectives.

In most cases, there is no need for a strategy to maintain the focus of a dispute; many argumentation protocols are designed so as to fore this focus to occur. Nevertheless, this item should be taken into consideration when designing a general purpose strategy. The first two items correspond to the military concept of a strategy, i.e. a high level direction and goals for the argumentation process. The third item corresponds to an agent's tactics. Tactics allow an agent to select a concrete action that fulfils its higher level goals. While Moore's work focused on natural language argument, these requirements formed the basis of most other research into agent argumentation strategies.

In 2002, Amgoud and Maudet [1] proposed a computational system which would capture some of the heuristics for argumentation suggested by Moore. Given a preference ordering over arguments, the created agents which could follow a "build" or "destroy" strategy, either defending their own arguments or attacking an opponent's.

Using some ideas from Amgoud's work, Kakas et al. [7] proposed a three layer system for agent strategies in argumentation. The first layer contains "default" rules, of the form $utterance \leftarrow condition$, while the two higher layers provide preference orderings over the rules (effectively acting as meta-rules to guide dialogue). Assuming certain restrictions on the rules, they show that only one utterance will be selected using their system, a trait they refer to as determinism. While their approach is able to represent strategies proposed by a number of other techniques, it does require hand crafting of the rules. No suggestions are made regarding what a "good" set of rules would be.

In [2], Amgoud and Prade examined negotiation dialogues in a possibilistic logic setting. An agent has a set of goals it attempts to pursue, a knowledge base representing its knowledge about the environment, and another knowledge base which is used to keep track of what it believes the other agent's goals are. The authors then present a framework in which these agents interact which incorporates heuristics for suggesting the form and contents of an utterance, a dialogue game allowing agents to undertake argumentation, and a decision procedure to determine the status of the dialogue. They then suggest and formalise a number of strategies that an agent can follow.

Other notable mentions and formalisations of argumentation strategies can be found in [4, 10, 3]. In the latter, Bench-Capon identifies a number of stages in the dialogue in which an agent might be

[1] A more detailed version of this paper was presented at ECAI-2006

[2] Department of Computing Science, University of Aberdeen, AB24 3UE, Scotland, email: noren,tnorman,apreece@csd.abdn.ac.uk

faced with a choice, and provides some heuristics as to what argument should be advanced in each of these cases.

Apart from guiding strategy, heuristics have seen other uses in dialogue games. Recent work by Chesñevar et al. [5] has seen heuristics being used to minimise the search space when analysing argument trees. Argument schemes [13] are well used tools in argumentation research, and can be viewed as a form of heuristic that guides the reasoning procedure.

3 Confidentiality Based Argumentation

In many realms of argument, auxiliary considerations (apart from simply winning or losing the argument) come into play. In many scenarios, one such consideration involves hiding certain information from an opponent. In this section, we describe a utility based heuristic to guide an agent taking part in a dialogue while being careful about what information it reveals. When faced with a number of possible arguments that it can advance, we claim it should put forth the one that minimises the exposure of information that it would like to keep private. The limitations of our current approach, as well as extensions and refinements to it are discussed in Section 5.

This work emerged while investigating the properties of other formal argument systems (such as [6, 12, 11, 15]). It is thus based on our own formal argumentation system. We believe, and plan to show in future work, how our heuristic can be implemented in other, more widely accepted argumentation frameworks.

Our system can be divided into two parts; at the lower level lies the logical machinery used to reason about arguments, while at the higher level we have a dialogue game, definitions of agents and the environment, and the heuristic itself. In this section, we will informally discuss our framework. A formal definition of the system can be found in [9].

3.1 The Argumentation Framework

The framework underpinning our heuristic is very simple, but still allows for argumentation to takes place. Argumentation takes place over a language containing propositional literals and their negation. Arguments consist of conjunctions of premises leading to a single propositional conclusion. A conclusion a which requires no premises can be represented by the argument $(\{\top\}, a)$.

We are interested in the status of literals (given a set of arguments), rather than the status of the arguments themselves. We can classify a literal into one of three sets: *proven*, *in conflict*, and *unknown*. A literal is in conflict if we can derive both it and its negation from a set of arguments. It is (un)proven, if it can (not) be derived and it is not in conflict, and unknown if neither it, nor its negation can be derived.

Our derivation procedure is based on the forward chaining of arguments. We begin by looking at what can be derived requiring no premises. By using these literals as premises, we compute what new literals can be generated, and continue with this procedure until no further literals can be computed. At each step of the process, we check for conflicts in the derived literals. When a conflict occurs, the literal (and its negation) are removed from the derived set and placed into a conflict set. Arguments depending on these literals are also removed from the derivation procedure. At the end of the derivation procedure, we can thus compute all three classes of literals[3].

[3] A Prolog implementation of this framework is available at `http://www.csd.abdn.ac.uk/~noren`.

3.2 Agents, the Dialogue Game and the Heuristic

Agents engage in a dialogue using the argumentation framework described above in an attempt to persuade each other of certain facts. In our system, an agent is an entity containing a private knowledge base of arguments, a function allowing it to compute the cost of revealing literals, and a set of utilities specifying how much it would gain for winning, drawing or losing the argument. The dialogue takes place within an environment, that, apart from containing agents, contains a public knowledge base which holds all arguments uttered by the agents.

Our dialogue game proceeds by having agents take turns to make utterances[4]. An utterance consists of a set of logically linked individual arguments. Alternatively, an agent may pass, and the game ends when no new arguments have been introduced into the public knowledge by any of the participants during their turn (which means that a dialogue is guaranteed to end given assuming a finite number of arguments). Once this occurs, it is possible to determine the status of each agent's goal, allowing one to determine the net utility gain (or loss) of all the agents in the system.

An agent wins an argument if its goal literal is in the proven set, while it draws an argument if the goal literal is in the conflict set or unknown. Otherwise, an agent is deemed to lose the argument. The net utility for an agent is determined by subtracting the utility cost for all literals appearing in the conflict and knowledge set from the utility gained for winning/drawing/losing the game.

To determine what argument it should advance, an agent computes what the public knowledge base would look like after each of its possible utterances. Using the derivation procedure described previously, it determines whether making the utterance will allow it to win/draw/lose the dialogue, and, by combining this information with the utility cost for exposed literals, it computes the utility gain for every possible utterance. It then selects the utterance which will maximise its utility. If multiple such utterances exist, another strategy (such as the one described in [10]) can be used.

It should be noted that it is possible to remove literals from the conflict set by attacking the premises of the arguments that inserted them into the set (thus reinstating other arguments). The lack of a preference relation over arguments means that attack in our framework is symmetric. While limiting, we are still able to model a useful subclass of arguments.

Before discussing the properties of the system, we show how a dialogue might look when this heuristic is used.

4 Example

The argument consists of a hypothetical dialogue between a government and some other agent regarding the case for, or against, weapons of mass destruction (WMDs) existing at some location.

Assume that $Agent_0$ would like to show the existence of WMDs. Proving this gains it 100 utility, while showing that WMDs don't exist means no utility is gained. Uncertainty (i.e. a draw) yields a utility gain of 50. Furthermore, assume the agent begins with the following arguments in its knowledge base:

$$(\{\top\}, spysat), (\{\top\}, chemicals), (\{\top\}, news), (\{\top\}, factories)$$

$$(\{\top\}, smuggling), (\{smuggling\}, \neg medicine), (\{news\}, WMD)$$

$$(\{factories, chemicals\}, WMD), (\{spysat\}, WMD)$$

[4] Note that we place no restrictions on the number of arguing agents.

$(\{sanctions, smuggling, factories, chemicals\}, \neg medicine)$

We associate the following costs with literals:

$(spysat, 100)$	$(chemicals, 30)$
$(news, 0)$	$(\{medicine, chemicals\}, 50)$
$(smuggling, 30)$	$(factories, 0)$

Note that if both medicine and chemicals are present, the agent's utility cost is 50, not 80. Thus, if both $spysat$ and $chemicals$ are admitted to, the agent's utility cost will be 130.

The dialogue might thus proceed as follows:

$$\begin{array}{lll}
(\mathbf{1}) & Agent_0: & (\{\top\}, news), (\{news\}, WMD) \\
(\mathbf{2}) & Agent_1: & (\{\top\}, \neg news) \\
(\mathbf{3}) & Agent_0: & (\{\top\}, factories), (\{\top\}, chemicals), \\
 & & (\{factories, chemicals\}, WMD) \\
(\mathbf{4}) & Agent_1: & (\{\top\}, sanctions), \\
 & & (\{sanctions, factories, chemicals\}, \\
 & & \quad medicine), (\{medicine\}, \neg WMD) \\
(\mathbf{5}) & Agent_0: & (\{\top\}, smuggling), \\
 & & (\{sanctions, smuggling, factories, \\
 & & \quad chemicals\}, \neg medicine) \\
(\mathbf{6}) & Agent_1: & \{\} \\
(\mathbf{7}) & Agent_0: & \{\}
\end{array}$$

Informally, the dialogue proceeds as follows: $Agent_0$ claims that WMDs exist since the news says they do. $Agent_1$ retorts that he has not seen those news reports. $Agent_0$ then points out that factories and chemicals exist, and that these were used to produce WMDs. In response, $Agent_1$ says that due to sanctions, these were actually used to produce medicine. $Agent_0$ attacks this argument by pointing out that smuggling exists, which means that the factories were not used to produce medicines, reinstating the WMD argument. Both agents have nothing more to say, and thus pass. $Agent_0$ thus wins the game.

It should be noted that while $Agent_0$ is aware that spy satellites have photographed the WMDs, it does not want to advance this argument due to the cost of revealing this information. The final utility gained by $Agent_0$ for winning the argument is 20: 100 for winning the argument, less 30 for revealing $smuggling$, and 50 for the presence of the $chemicals$ and $medicine$ literals. Also, note that the fact that $Agent_1$ revealed the existence of medicines cost $Agent_0$ an additional 20 utility. While this makes sense in some scenarios, it can be regarded as counterintuitive in others. Extensions to overcome this behaviour are examined in the next section.

5 Discussion

As mentioned earlier, we created our own underlying framework, and one of our short term research goals involves mapping our heuristic into another, more widely used argumentation framework. Our framework shares much in common with the "sceptical" approach to argumentation; when arguments conflict, we refuse to decide between them, instead ruling them both invalid. The simplicity of our approach means that only specific types of arguments can be represented (namely, those whose premises are a conjunction of literals, and whose conclusion is a single literal). However, as seen in the example, even with this limitation, useful arguments can still emerge.

The way in which we represent the information "leaked" during the dialogue, as well as calculate the agent's net utility, while simple, allows us to start studying dialogues in which agents attempt to hide information. Until now, most work involving utility and argumentation has focused on negotiation dialogues (e.g. [14]). We propose a number of possible extensions to the work presented in this paper.

One simple extension involves the addition of a context to the agent's cost. In other words, given that fact A, B and C are known, we would like to be able to capture the notion that it is cheaper to reveal D and E together than as speech acts at different stages of the dialogue. Without some form of lookahead to allow the agent to plan later moves, this extension is difficult to utilise. Once some form of lookahead exists, the addition of opponent modelling can further enhance the framework. Experimentally, evaluating the effects of various levels of lookahead, as well as different forms of opponent modelling might yield some interesting results.

Currently, we do not differentiate between information which the agent has explicitly committed to, and information that the agent has not yet disagreed with. More concretely, assume that the public knowledge base contains the argument $(\{\top\}, A)$. If an agent makes use of this argument, perhaps by submitting the argument $(\{A\}, B)$, then it is committed to the fact that A is true. If however, it never puts forth arguments making use of the fact, then an opponent cannot know if the agent is actually committed to A or not. We plan to extend our formalism and heuristic to capture this interaction in the near future.

Another extension that emerges from this line of reasoning is the concept of lying. An agent might commit to A to win an argument, even if its knowledge base contains only $\neg A$. How best to deal with this situation is an open question.

6 Conclusions

In this paper, we proposed a heuristic for argumentation based on minimising the cost of information revealed to other dialogue participants. While such an argumentation strategy arises in many real world situations, we are not familiar with any application that explicitly makes use of this technique. To study the heuristic, we proposed an argumentation framework that allowed us to focus on it in detail. Several novel features emerged from the interplay between the heuristic and the framework, including the ability of an agent to win an argument that it should not have been able to win (if all information were available to all dialogue participants). While we have only examined a very abstract model utilising the heuristic, we believe that many interesting extensions are possible.

Acknowledgements

This work is partly funded by the DTI/EPSRC E-Science Core Program and British Telecom, via a grant for the CONOISE-G project (http://www.conoise.org), a multi-university collaboration between Aberdeen, Cardiff and Southampton universities, and BT.

We would also like to thank the anonymous reviewers for their very useful comments.

REFERENCES

[1] Leila Amgoud and Nicolas Maudet, 'Strategical considerations for argumentative agents (preliminary report).', in *NMR*, pp. 399–407, (2002).

[2] Leila Amgoud and Henri Prade, 'Reaching agreement through argumentation: a possiblistic approach', in *Proceedings of KR 2004*, (2004).

[3] Trevor J.M Bench-Capon, 'Specification and implementation of Toulmin dialogue game', in *Proceedings of JURIX 98*, pp. 5–20, (1998).

[4] C. Cayrol, S. Doutre, M.-C. Lagasquie-Schiex, and J. Mengin, '" minimal defence" : a refinement of the preferred semantics for argumentation frameworks', in *Proceedings of NMR-2002*, (2002).

[5] Carlos Iván Chesñevar, Guillermo Ricardo Simari, and Lluis Godo, 'Computing dialectical trees efficiently in possibilistic defeasible logic programming.', in *Logic Programming and Nonmonotonic Reasoning, 8th International Conference, LPNMR 2005, Diamante, Italy, September 5-8, 2005, Proceedings (LNCS 3662)*, eds., Chitta Baral, Gianluigi Greco, Nicola Leone, and Giorgio Terracina, Lecture Notes in Computer Science, pp. 158–171. Springer, (2005).

[6] Phan Minh Dung, 'On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games', *Artificial Intelligence*, **77**(2), 321–357, (1995).

[7] Antonis C. Kakas, Nicolas Maudet, and Pavlos Moraitis, 'Layered strategies and protocols for argumentation-based agent interaction.', in *ArgMAS*, pp. 64–77, (2004).

[8] David Moore, *Dialogue game theory for intelligent tutoring systems*, Ph.D. dissertation, Leeds Metropolitan University, 1993.

[9] Nir Oren, Timothy J. Norman, and Alun Preece, 'Arguing with confidential information', in *Proceedings of the 18th European Conference on Artificial Intelligence*, Riva del Garda, Italy, (August 2006). (To appear).

[10] Nir Oren, Timothy J. Norman, and Alun Preece, 'Loose lips sink ships: a heuristic for argumentation', in *Proceedings of the Third International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2006)*, pp. 121–134, Hakodate, Japan, (May 2006).

[11] John L. Pollock, 'Perceiving and reasoning about a changing world.', *Computational Intelligence*, **14**, 498–562, (1998).

[12] Henry Prakken and Giovanni Sartor, 'A dialectical model of assessing conflicting arguments in legal reasoning', *Artificial Intelligence and Law*, **4**, 331–368, (1996).

[13] Chris. A. Reed and Douglas N. Walton, 'Applications of argumentation schemes', in *Proceedings of the 4th Conference of the Ontario Society for the Study of Argument (OSSA2001), Windsor, Canada*, eds., H. V. Hansen, C. W. Tindale, J. A. Blair, R. H. Johnson, and Robert C. Pinto, p. CD ROM, (2001).

[14] Katia Sycara, 'Persuasive argumentation in negotiation', *Theory and Decision*, **28**(3), 203–242, (May 1990).

[15] Bart Verheij, 'DefLog: On the logical interpretation of prima facie justified assumptions', *Journal of Logic and Computation*, **13**(3), 319–416, (2003).