ArgTrust: Decision Making with Information from Sources of Varying Trustworthiness

(Demonstration)

Simon Parsons[1,2]   Elizabeth Sklar[1,2]   Jordan Salvit[2]   Holly Wall[1]   Zimi Li[2]
[1]Brooklyn College and [2]The Graduate Center
The City University of New York, New York, USA
parsons@sci.brooklyn.cuny.edu,sklar@sci.brooklyn.cuny.edu,jordan@jordansalvit.com,
holly.e.wall@gmail.com,zimili.sjtu@gmail.com

ABSTRACT

This work aims to support decision making in situations where sources of information are of varying trustworthiness. Formal argumentation is used to capture the relationships between such information sources and conclusions drawn from them. A prototype implementation is demonstrated, applied to a problem from military decision making.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Coherence & co-ordination; multiagent systems.*

Keywords

Trust, Argumentation, Human-Machine Interfaces

1. INTRODUCTION

In the abstract, our problem domain is one in which a human decision maker has to incorporate information that comes from a variety of sources, where those sources are of varying trustworthiness. In such situations, the decision maker not only has to weigh the pros and cons of various courses of action, but also has to take into account the reliability of the information that she is using, where that reliability depends upon the trustworthiness of the sources.

To make this problem statement more concrete, in this paper we use the following example, loosely based on [3]. In this example, a decision is being made about whether to carry out an operation in which a combat team will move into a region to try to apprehend a high value target (HVT) believed to be in a village in the region.

We have the following information. If there are enemy fighters in the area, then an HVT is likely to be in the area. If there is an HVT in the area, and the mission will be safe, then the mission should go ahead. If the number of enemy fighters in the area is too large, the mission will not be safe. UAVs that have flown over the area have provided images that appear to show the presence of a significant number of campfires, indicating the presence of enemy fighters. The

quality of the images from the UAV is not very good, so they are not very trusted. A reconnaissance team that infiltrated the area saw a large number of vehicles in the village that the HVT is thought to be inhabiting. Since enemy fighters invariably use vehicles to move around, and the reconnaissance team is highly trusted, this is strong evidence for the presence of many enemy fighters. Informants near the combat team base claim that they have been to the area in question and that a large number of fighters are present. In addition, we have the default assumption that missions will be safe, because in the absence of information to the contrary, we believe that the combat team will be safe.

Thus there is evidence from UAV imaging that sufficient enemy are in the right location to suggest the presence of an HVT. There is also some evidence from informants that there are too many enemy fighters in the area for the mission to be safe. Since informants are paid, their incentive is often to make up what they think will be interesting information and so they are not greatly trusted. However, this conclusion is supported by the findings of the reconnaissance team who are highly trusted. There is conflicting information in this example scenario, and no clear "right answer"—precisely the type of scenario we address in the work demonstrated here.

2. FORMAL REPRESENTATION

Elsewhere [4] we have described the formal argumentation system that we have developed for reasoning with this kind of information. This allows part of the scenario to be represented as follows[1]:

$$InArea(campfires)$$
$$Safe(mission)$$
$$InArea(campfires) \Rightarrow InArea(enemy)$$
$$InArea(enemy) \Rightarrow HVT$$
$$HVT \wedge Safe(mission) \Rightarrow Proceed(mission)$$

From this information, we can construct arguments such as:

$$\left(\left\{ \begin{array}{l} InArea(campfires), \\ InArea(campfires) \Rightarrow InArea(enemy), \\ InArea(enemy) \wedge Safe(mission) \Rightarrow HVT, \\ HVT \Rightarrow Proceed(mission) \end{array} \right\}, \right.$$
$$\left. Proceed(mission) \right)$$

---

[1]We stress that this is purely illustrative — a real model of this example would not only represent all the information, but also be considerably more detailed.
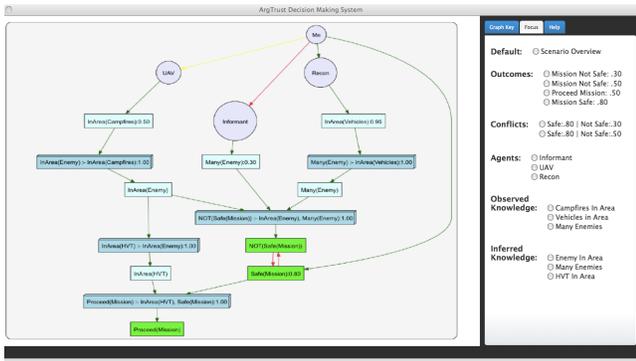
Figure 1: The interface providing a high-level view

Figure 2: The interface providing a more focussed view

which is an argument for the mission proceeding, based on the fact that there are campfires in the area, these suggest enemy fighters, that enemy fighters suggest the presence of an HVT, and that the presence of an HVT (along with the default assumption that the mission will be safe) suggests that the mission should go ahead.

We can build other arguments from the full information that is available. For example, from the recon team's information, we can conclude that there are many enemies in the area and hence the mission will not be safe. This conflicts with the previous argument by undermining the assumption about the mission being safe. The argumentation system includes rebuttal as well as undermining. The system also captures trust between the information sources. Trust values become weights on arguments, and the weights are used to resolve attacks into defeats [1].

3. IMPLEMENTATION

The system currently takes as input an XML file. Two types of data are supplied: a specification of each individual's knowledge in the form of defeasible Horn clauses, and a specification of how much each source of information is trusted. This information can be modified through a command-line interface. The trust information specifies the individuals involved (including "me", the decision maker) and the trust relationships between them, including the level of trust (specified as a number between 0, no trust, and 1, completely trustworthy). The current implementation uses these values to compute the trust that one agent places on another, employing either TidalTrust [2] or the mechanism described in [5].

Given the input data, the system can answer queries about whether a given conclusion can be derived by a given agent using the grounded semantics to establish the acceptability status of arguments. The system is invoked from the Unix command line, and generates output in the form of an annotated `dot`[2] description. This can be converted to any graphical format.

Since displaying all the available information rapidly overwhelms the user, we are working on approaches to allow the user to navigate the graphical interface. Our current prototype interface provides a number of views of the set of arguments and associated information. For example, it allows the user to look at a high-level view of the relationship

_____
[2] http://www.graphviz.org/
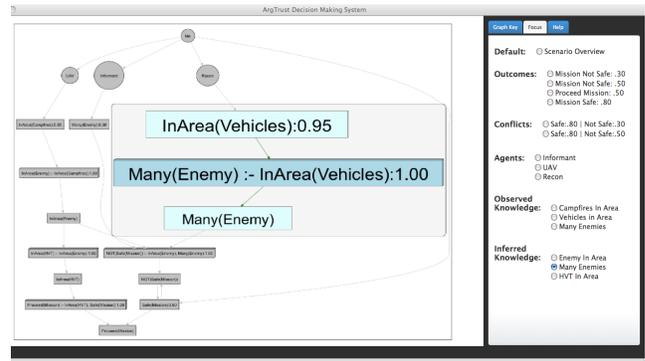
between the arguments (Figure 1) and also to focus on the detail of a specific argument (Figure 2)

Examining the arguments reveals that the conflict in this case is between the information from the recon team and the informants — which support the fact that there are many enemy fighters present — and the information about campfires that comes from the UAV — which supports the presence of an HVT and hence that the mission should go ahead. Further consideration of the situation can then focus on the reliability of these pieces of data. Were it the case, for example, that the information from the UAV imagery was considered more reliable than the information from the recon team, then there would be a case for proceeding. Such "what if" reasoning is supported by our implementation through the ability to add and retract assertions at the command line.

Acknowledgement

4. REFERENCES

[1] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artifical Intelligence*, 34(3), 2002.

[2] J. Golbeck. *Computing and Applying Trust in Web-based Social Networks*. PhD thesis, Univ of Maryland, College Park, 2005.

[3] S. Naylor. *Not a Good Day Day to Die: The Untold Story of Operation Anaconda*. Berkley Caliber Books, 2005.

[4] Y. Tang, K. Cai, P. McBurney, E. Sklar, and S. Parsons. Using argumentation to reason about trust and belief. *Journal of Logic and Computation*, 22(5), 2012.

[5] Y. Wang and M. P. Singh. Trust representation and aggregation in a distributed agent system. In *Proc of the 21st National Conference on Artificial Intelligence*, 2006.