

Crowdsourcing for Argument Verification in Intelligence Analysis: A Position Paper

Alice Toniolo¹, Robin Wentao Ouyang², Timothy J. Norman¹, and Mani Srivastava^{2,*}

¹ Department of Computing Science, University of Aberdeen, Scotland, UK

² University of California, Los Angeles, CA, USA

Abstract. We discuss crowdsourcing as a means to evaluate the benefits of employing argument schemes for making sense of information in intelligence analysis. In this position paper, we propose two sets of crowdsourcing experiments to measure the effectiveness of argument schemes in forming valid arguments, and critical questions to prevent cognitive biases.

1 Crowdsourced Analysis of Arguments

Models of argumentation have increasingly been employed in human decision-making systems to facilitate good reasoning [1]. Stereotypical patterns of reasoning, such as argument schemes derived from analyses of human reasoning, are commonly employed in such systems for determining whether arguments are well-formulated, and carry some weight towards the plausibility of the conclusions [8]. In our research, we employ argument schemes to support collaborative intelligence analysis within an agent-based model of reasoning that aims to facilitate the identification of plausible hypotheses about events [7].

In order to maximise the effectiveness of systems that employ computational models of argumentation, there is a real need to rigorously evaluate their use in human decision support [2]. In this research, we propose the use of crowdsourcing for verifying good reasoning and to study the benefits of argument schemes in guiding the construction of well-formulated arguments. Crowdsourcing is a time-cost effective method to perform tasks that are hard for machines. Moreover, it provides methods for dealing with biased and unreliable reports [6]. Crowdsourcing has recently been applied to different aspects of argumentation: for *argument mining*, where the crowd is employed to identify arguments in transcripts [4]; and for *crowdsourcing arguments* about risk assessment via online argument mapping systems to support public debate. In the context of intelligence analysis, crowdsourcing offers a twofold advantage: controlled experiments may be used to evaluate the effectiveness of argument schemes in constructing arguments from information; and crowds can be leveraged to give timely feedback on whether arguments are valid and present plausible conclusions. This process is referred to as *crowdsourced analysis of arguments* and is fundamental to identifying assumptions and biases when the analysis is to be shared with decision-makers or collaborators. In this paper, we propose two sets of experiments to develop an effective strategy for guiding users in employing argument schemes; e.g. by suggesting premises necessary to understand the claims. Further, we investigate cognitive biases by studying the mitigating effects of critical questions (CQs) and the conditions for which analysts are more likely to be affected by bias. The results may highlight a strategy to counteract biases via a combination of prioritisation of critical questions and probabilistic methods to reduce the influence of biased opinions.

2 Experiments

Argument Validation through Crowdsourcing. Our system for supporting intelligence analysis [7] employs argument schemes [8] as guidelines to link together information and draw plausible conclusions. This differs from other applications where schemes are used to analyse previously stated arguments. Here we investigate the question *Would argument schemes be a reliable method to structure inferences and draw plausible conclusions from incomplete information?* This is important when the analysis is shared with others; e.g., collaborators may not be able to understand conclusions or the inferential rule may inaccurately be considered valid due to unstated premises (*enthymemes*[8]).

* This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001.

In initial experiments, we may ask the crowd to select plausible conclusions from unlinked information. This is tested against tasks where information is proposed via argument schemes. For example, given “I1: Some valuable equipment was stolen from the biology lab”, “I2: A criminal gang G involved in narcotics has returned to town”, and “I3: There was a protest against the use of mice”, we may ask “Who stole the equipment?”. A superficial answer may state that the gang is guilty, but the *argument for the identification of actors* suggests that gang and protesters have at least a feature (Q) required to bring about the robbery. Hence we would expect the crowd to state Q for both actors; the protesters could have gained access to the lab given that they had information about the mice. We may ask the crowd to annotate information with types of premises, derived from schemes. The effectiveness of the scheme is shown by its ability to guide the crowd to state the premises required for an inference to be defeasibly valid, and for conclusions to be plausible. Further, we may vary the type of scheme used to study which is best suited for the question asked; e.g. an *argument from cause to effect* is more likely to promote an answer regarding how the equipment was stolen. Finally, we may investigate the effects of enthymemes on the conclusion drawn according to different types of missing premises (i.e. evidence, common-knowledge, or assumptions [8]).

Examining Bias through Crowdsourcing. The sensemaking process may be affected by cognitive biases such as confirmation, vividness of information, and perception of cause and effects [5]. Here, we investigate: *Would critical questions help analysts prevent different cognitive biases when drawing conclusions?*

In the basic task, the crowd is asked to report an opinion about conclusions of argument schemes. We test different biases with different schemes. Initially, we must identify situations in which the responses are biased and compare these results against opinions when introducing critical questions. For example, we consider a biased perception of cause and effects. We received information “P1: The airspace control unit has replaced a radar”, and “P2: Ten more UAVs than usual were detected”. It is likely that the conclusion of an *abductive argument from effect to cause* “P1 is a plausible cause for P2” would be accepted. There is, however, no evidence for a new radar being the cause of an improved airspace control. The question *CQ1: Is there evidence for a causal relation between the events?* should guide a negative opinion on the conclusion. *CQ1* directly prevents the bias, but *CQ2: Is there any other cause that leads the effect to occur?* may also avoid pitfalls. We intend to explore strategies for selecting these questions. Moreover, there are factors that increase the likelihood of this type of bias, including overestimation of influential actors, analysis of secondhand information, and features of the crowd such as background, geographical provenance and so on. These factors may help us explore probabilistic models of aggregation of responses to systematically mitigate bias [6].

In conclusion, we believe that crowdsourced analysis will guide the development of models to provide effective support for reasoning with evidence in intelligence analysis. Moreover, crowdsourced analysis would bring useful insights for validating crowdsourced arguments in maps or debates (e.g. [3]).

References

1. Bench-Capon, T., Dunne, P.: Argumentation in artificial intelligence. *Artificial Intelligence* 171(10-15), 619–641 (2007)
2. Cerutti, F., Tintarev, N., Oren, N.: Formal arguments, preferences, and natural language interfaces to humans: an empirical evaluation. In: *Proc. of the 21st European Conference on Artificial Intelligence* (2014)
3. Fox, J., Hazlewood, V., Elson, T.: Use of argumentation and crowdsourcing techniques for risk assessment and policy development. In: *Proc. of the 11th Workshop on Argumentation in Multi-Agent Systems* (2014)
4. Ghosh, D., Muresan, S., Wacholder, N., Aakhus, M., Mitsui, M.: Analyzing argumentative discourse units in online interactions. In: *Proc. of the 1st Workshop on Argumentation Mining*. pp. 39–48 (2014)
5. Heuer, R.J.: *Psychology of intelligence analysis*. US Government Printing Office (1999)
6. Ouyang, R.W., Srivastava, M., Toniolo, A., Norman, T.J.: Truth discovery in crowdsourced detection of spatial events. In: *Proc. of the 23rd Conference on Information and Knowledge Management*. pp. 461–470 (2014)
7. Toniolo, A., Norman, T.J., Etuk, A., Cerutti, F., Ouyang, R.W., Srivastava, M., Oren, N., Dropps, T., Allen, J.A., Sullivan, P.: Supporting reasoning with different types of evidence in intelligence analysis. In: *Proc. of the 14th International Conference on Autonomous Agents and Multiagent Systems* (2015)
8. Walton, D., Reed, C., Macagno, F.: *Argumentation schemes*. Cambridge University Press (2008)