Addressing argumentation puzzles with model-based diagnosis

Sileno, G.; Boer, A.W.F.; van Engers, T.M.

Published in:
CEUR Workshop Proceedings

Addressing argumentation puzzles with model-based diagnosis (extended abstract)

Giovanni SILENO [a,1], Alexander BOER [a] and Tom VAN ENGERS [a]
[a] *Leibniz Center for Law, University of Amsterdam*

Keywords. Explanation, Model-based diagnosis, Argumentation, Agent-roles

1. Introduction

In law, and, consequently, in AI & Law, argumentation, scenario-modeling, and the combination of both, are the traditional ways of theorizing about judicial reasoning and legal truth, while probabilistic reasoning has traditionally been treated with suspicion[2]. Nevertheless, because of the growing relevance of forensic scientific evidence, a proper integration of probabilistic reasoning into the argumentation process is increasingly a debated problem.

Pollock presents in [1] a lucid philosophical critique on how probabilistic methods approach the problem of *justification*, in the form of some interesting legal puzzles. He gives the following case: *Jones says that the gunman had a moustache. Paul says that Jones was looking the other way and did not see what happened. Jacob says that Jones was watching carefully and had a clear view of the gunman.* This is an example of "collective defeat" (Paul vs Jacob), which results in a "zombie argument" (Jones'). From this story, Pollock targets some intuitive properties. (1) Given the conflict of witnesses, we should not believe to Jones' claim carelessly. (2) If we consider Paul more trustworthy than Jacob, Paul's claim should be justified, but to a lesser degree. (3) Conversely, if Jacob had confirmed Paul's claim, its "degree of justification" should have increased.[3] Pollock gives then a preliminary, elaborated proposal for degrees of justification, based on "probable probabilities". Working with a different – in one sense opposite – perspective, we have found an alternative solution to his quest.

2. Methodology

Argumentation is generally perceived as operating at a meta-level, concerned with support and attack relationships between *claims*, rather than between *messages* and *explana-*

[1] Corresponding author: g.sileno@uva.nl.

[2] In Nulty & Ors v Milton Keynes Borough Council [2013] the court puts the point concisely and – to many indignant scientists – provocatively: "you cannot properly say that there is a 25 per cent chance that something *has happened*: Hotson v East Berkshire Health Authority [1987]. Either it has or it has not".

[3] We slightly changed the third one, in order to make use of the same story.

tions. In fact, common argumentation theories treat messages directly as claims, i.e. constructions based only on the *story* level of narrative acts [2]. We propose to consider the relation between an individual message and an explanation, and the space of hypothetical explanations.

Messages are *speech acts*, and as such, they are generated and interpreted depending on the knowledge and intentions of the participants. Thus, the quest for a solution to a case requires not only an investigation into the structures and processes that made the occurrence of the case possible, but also into the process of elicitation and evaluation of explanations of the case.

In our approach, we emphasize agents' *positions*. We encourage the modeller to consider scenarios from the perspective of the participants, through the elicitation of *agent-roles*, which refer to prototypical patterns of behaviour in the target social domain.[4] Some of them represent normal behaviours, while others are associated to faulty, non compliant ones, in the sense of being at fault with the (normatively characterized) behaviour of the social system.[5]

Fundamental concepts An *observation O* consists of three elements: 1) a set of *scenario agents*, including an observer, 2) a set of messages between the observer and other agents, and 3) a temporal ordering relationship on messages (e.g. indexed on reception time). An observation becomes a *diagnostic problem* if it is surprising/alarming to the observer. Given a certain social context, an *explanation E* (or interpretation) is a multi-agent system, and consists of three elements: 1) a set of scenario agents, embodying *agent-roles*, 2) a set of messages between the agents, and 3) a (partial) temporal ordering relationship on messages. Given an observation, the observer/interpreter should be able to generate a set of explanations. An explanation may include 1) additional agents beyond the observed ones, 2) the merging of multiple agents into one agent, or 3) the splitting of an observed agent into multiple agents. To determine the relative value of an explanation E, given O, we calculate the confirmation value of O for explanation E with the measure proposed in [5], permitting ordinal judgments about explanations[6]:

$$c(O,E) = \frac{P(O|E) - P(O|\neg E)}{P(O|E) + P(O|\neg E)} \tag{1}$$

Operationalization Our methodology can be applied in three steps. First, we create executable models of the prototypical agent-roles. Second, we generate all explanatory hypotheses, allocating known agent-roles to the scenario agents. Third, we evaluate all explanations, given the messages reported to the observer/interpreter.

[4] An *agent-role* is a social intentional entity provided with certain beliefs, rationality and goal-oriented plans of actions, dual to specific social dispositions. It may be epistemically associated to multiple identities. An agent-role may produce unsuccesful outcomes too, because of faulty inputs, incomplete knowledge or wrong processing.

[5] It is worth to observe that compliance and non-compliance are qualifications relative to the position of the diagnostic agent in the social system. In a world of liars, people telling the truth would fail in respect to the social practice of systematically lying.

[6] $p(\neg E)$ is the probability that E is not the case. If $c(O,E)$ approaches 1 (-1), the observation O confirms (disconfirms) the explanation E. If c is equal to 0, the observation O is irrelevant. Put in words, with this measure, an observation confirms an explanation if it is predicted by the explanation *and* discriminates the explanation from its alternatives.

	(1) Jacob attacks Paul			(2) Jacob attacks Paul			(3) Jacob supports Paul		
	E_9	E_{24}	E_{26}	E_9	E_{24}	E_{26}	E_{13}	E_{20}	E_{30}
Jones tells the truth	true	false	false	true	false	false	true	false	false
Paul tells the truth	false	true	false	false	true	false	false	true	false
Jacob tells the truth	true	false	true	true	false	true	false	true	false
Jones saw the gunman	true	false	true	true	false	true	true	false	true
gunman had a moustache	true	false	false	true	false	false	true	false	false
$P(\texttt{k(Paul)})$		0.5			0.8			0.5	
$P(\texttt{k(Jacob)})$		0.5			0.5			0.5	
$c(O_2, E)$	0,72	0,72	0,72	0,76	0,81	0,76	0,72	0,72	0,72
$c(O_3, E)$	0.88	0.88	0.88	0.88	0.95	0.88	0.88	0.88	0.88

Table 1. Confirmation factors in Pollock's puzzles

3. Results

We apply this method to Pollock's puzzle. We consider $2^5 = 32$ possible scenarios, three scenario agents (Jones, Paul, Jacob) and two agent-roles: truth-tellers (`k`) or liars (`-k`). The outcome is summarized on Table 1, reporting only explanations confirmed by the complete observation. The following results show how we have obtained the properties targeted in the introduction. (1) Assuming indifference toward hypotheses, our approach confirms *to the same degree* hypotheses in which the gunman has a moustache, and not. (2) Using for instance $P(\texttt{k(Paul)}) = 0.8 > P(\texttt{k(Jacob)}) = 0.5$, the hypothesis in which Paul is telling the truth is the one confirmed to the greater degree. (3) Seeing that Jacob confirms what said by Paul, we observe that the confirmation factor of the hypothesis they both support increases, *just as much* as the hypotheses in which they are both lying. The third point is an important consequence of indifference towards prior probabilities. For instance, it allows us to consider – with the same strength – the possibility of organized crime schemes.

Obviously, our easy solution does not solve Pollock's argumentation puzzles within the rules of his game, but it clearly demonstrates the added value of our model-based diagnosis framework, proposed first in [3,4], in the field of law.

References

[1] J. L. Pollock. Reasoning and probability. *Law, Probability and Risk*, 6(1-4):43–58, 2007.

[2] G. Sileno, A. Boer, and T. van Engers. Analysis of legal narratives: a conceptual framework. In *JURIX 2012: 25th Int. Conf. Legal Knowledge and Information Systems*, 143–146, 2012.

[3] A. Boer and T. van Engers. Diagnosis of multi-agent systems and its application to public administration. *Lecture Notes in Business Information Processing*, 97:258–269. Springer, 2011.

[4] A. Boer and T. van Engers. An agent-based legal knowledge acquisition methodology for agile public administration. *ICAIL 2011: 13th Int. Conf. AI & Law*, 171–180, 2011.

[5] K. Tentori, V. Crupi, N. Bonini, and D. Osherson. Comparison of confirmation measures. *Cognition*, 103(1):107–119, 2007.