Ontology for the Intelligence Community (OIC-2007)
November 28-29, 2007 · Columbia, Maryland

Ontological Support for Bayesian Evidence Management

Michael N. Huhns and Marco G. Valtorta
Center for Information Technology
University of South Carolina
Columbia, SC 29208

Abstract

This paper describes our work on an integrated system that can assist analysts in exploring hypotheses using Bayesian analysis of evidence from a variety of sources. The hypothesis exploration is aided by an ontology that represents domain knowledge, events, and causality for Bayesian reasoning, as well as models of information sources for evidential reasoning. We are validating the approach via a tool, Magellan, that uses Bayesian models for an analyst's prior and tacit knowledge about how evidence can be used to evaluate hypotheses.

1. Introduction

Much of the extensive work on ontologies to date has focused on modeling and representing the world of objects. The ontologies needed for our research supporting the management of hypotheses and evidence for analysts, however, must additionally model events and causality. Less work has been done on this aspect of ontologies. In this paper we show how concepts from a causal ontology can be used directly as variables in Bayesian networks and how the attributes of the causal concepts can be used in matching evidence to the variables. Moreover, subclass relationships in the ontology enable the extension of Bayesian reasoning over types.

2. Bayesian Reasoning for Evidence Management

There are numerous real-world situations about which an analyst might wish to hypothesize and investigate, but it would be impractical to encode all of them explicitly in a support system for analysts. Instead, our approach is to represent fragments of situations and provide a mechanism for combining them into a wide variety of more complete ones [1,4]. The combination occurs dynamically as evidence about a situation becomes available or as an analyst revises or enters new hypotheses. A fragment is represented as a Bayesian network with nodes for hypotheses, events, and evidence, and links for relating them. Our ability to combine the fragments into more complete situation models is dependent on having a consistent terminology in which the fragments are described. The focus of our work has been on (1) defining and representing the terminology, including terms of a domain and terms for evidence in that domain, (2) capturing new fragments from a variety of sources, and (3) incorporating the terminology and BN fragments into an integrated end-to-end tool, Magellan.

2.1. Capturing the Terminology and Prior Knowledge for a New Domain

Intelligence analysts are concerned primarily with hypotheses that involve cause-and-effect. These are best supported by an ontology emphasizing events and their causal relationships, along with a hypothetical world of possible events, actions, and causes. However, causal relationships must be interpreted in the context of real-world objects and their properties, which can be represented in a conventional ontology such as those that are part of SUMO. The evidence for reasoning about hypotheses can come from a variety of sources, and the acquisition of evidence and events from these sources must also be represented, constituting a third kind of ontological representation describing the information sources. Figure 1 depicts the three ontological models we use for modeling situations, relating situations to background knowledge about objects in the world, acquiring evidence, and assessing the likelihood of the situations using Bayesian reasoning. Our tool, Magellan, uses Protégé for capturing the ontologies, RDF for representing the terminology, XMLBIF for representing the causal relationships, and RDF and RDQL for requesting evidence from information sources.

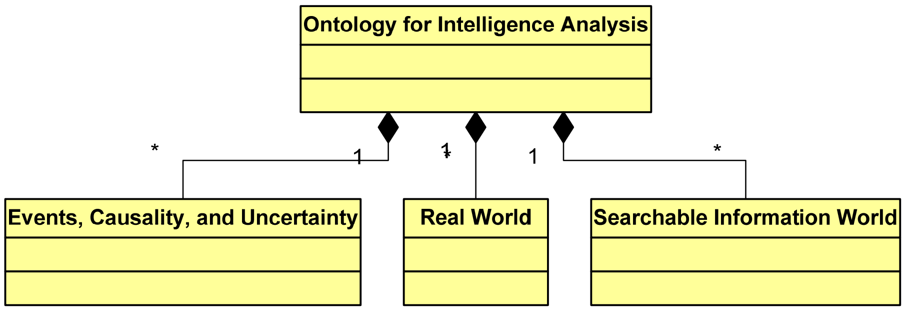

Figure 1. An ontology for intelligence analysts has three related parts, corresponding to the world of causality and hypothetical events needed for Bayesian reasoning, the real world of things needed to model situations, and the world of information and information sources needed for evidence management

Causality is a special relationship among events for which certain properties hold probabilistically. For example, causality is logically irreflexive and asymmetric, but probabilistically transitive. Causal models are very useful, because they allow prediction of the effect of interventions [3,5].

New variables are added to the causal and event portion of an analyst's ontology using Protégé, so that all of the nodes in a Bayesian network fragment are represented in a standard and consistent terminology. We extend SUMO with this terminology, so that we can take advantage of SUMO's existing description of general knowledge of the world. Each variable has a set of identifying attributes, which are used to combine fragments (fragments can be combined only if their attributes unify) [4].

Probabilities are assigned to events in the fragment by performing experiments, estimating beliefs, or counting outcomes. Once assigned, they are updated by conditioning on evidence using Bayes rule and the laws of probability. The fragments are stored in a repository, where they can be matched with evidence and combined with other fragments to produce models of situations that are as complete, accurate, and specific as possible.

We also represent in the information source ontology the level of credibility of items of evidence, and provide a Bayesian interpretation of credibility. We define *evidence* to be a collection of findings, each of which describes the state of a Bayesian network variable, and distinguish three kinds [7]:

1. A *hard finding* specifies that the variable has a particular value.
2. A *soft finding* is a distribution on the states of a variable, usually corresponding to an "objective" statistical distribution that is not expected to change within a scenario.
3. A *virtual finding* is a likelihood ratio corresponding to the credibility associated to an evidence source, such as a witness. Unlike soft findings, virtual findings allow for an update of the posterior probability of the evidence variable.

Our modified version of ACH1 [6] is used by an analyst to enter the appropriate hypotheses and any initial evidence that might be available. The terminology available to the analyst is provided via drop-down menus as shown in Figure 2, where the menu entries are the ontology terms from

Protégé. The resultant ACH [2] matrix is converted automatically into a bipartite Bayesian network, with initial probabilities assigned based on the relevance factors assigned to cells of the matrix. The network is saved into the repository of fragments.

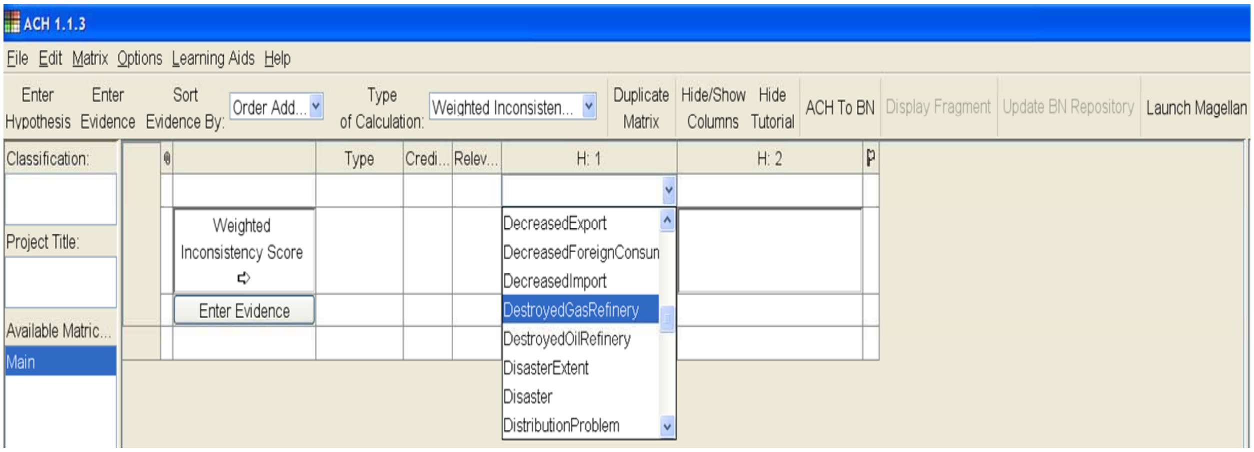

Figure 2. The extended ACH interface is integrated with the ontology of events through pull-down menus

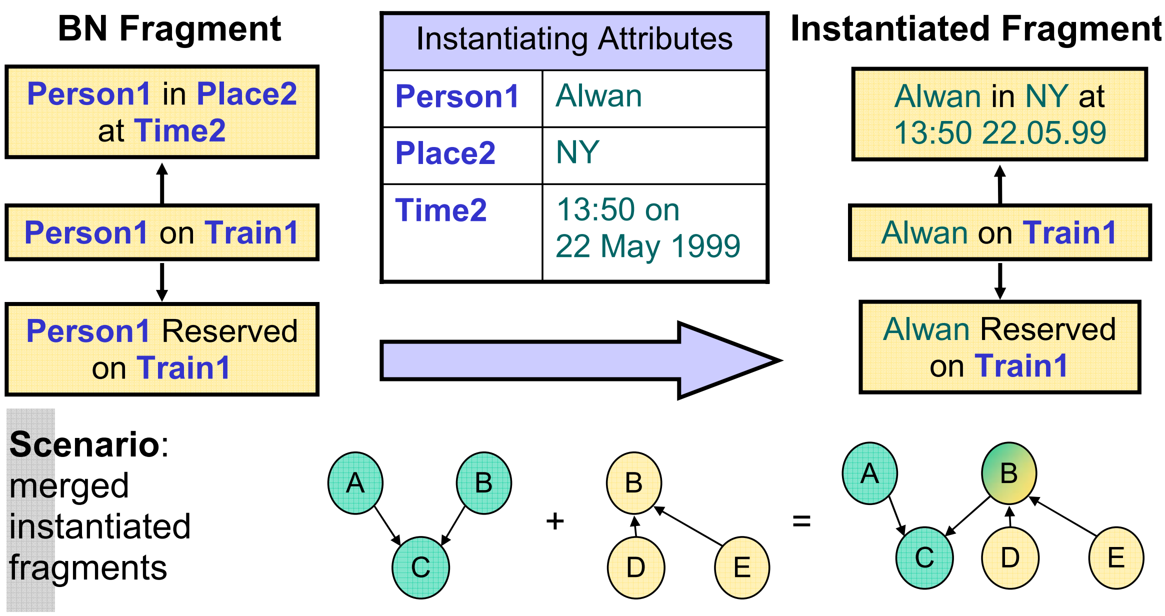

Figure 3. Fragments (templates) are merged based on instantiating evidence

3. Architecture for Bayesian Reasoning

Figure 4 shows an architecture for Bayesian reasoning, which would be used as follows. Based on initial triggering messages, or based on a hypothesized situation that an analyst would like to investigate, an appropriate scenario represented as a Bayesian model is chosen by the analyst and a corresponding form is shown listing initial evidence and the domain variables for the scenario. The evidence values for the variables can be supplied automatically from the triggering messages or can be entered by the analyst. The Bayesian reasoning component, using a value-of-information calculation, then determines which pieces of evidence would be most useful in confirming or denying the analyst's hypothesis. A request for this evidence is sent to the analyst, who returns the result to the Bayesian reasoner for incorporation into the scenario, and the likelihood of the analyst's hypothesis is reassessed. The process is repeated until the analyst decides to stop or there is no more evidence available that changes the plausible outcomes.

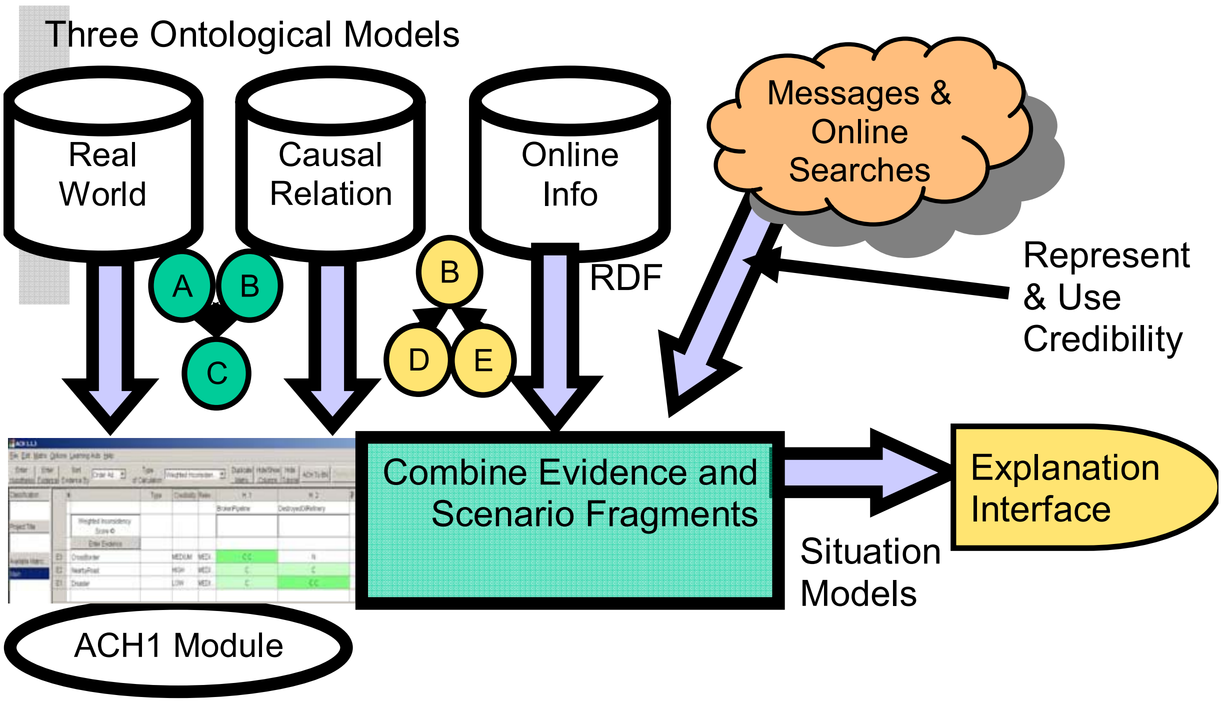

Figure 4. Magellan architecture for Bayesian Reasoning used to explore an analyst's hypotheses

4. Conclusion

As we continue to increase the functionality of our Bayesian reasoning system, we will improve our representation of events and causality, and increase the capabilities for the application of prior and tacit knowledge to the exploration of analysts' hypotheses.

5. Acknowledgements

This work was funded in part by the Disruptive Technology Office Collaboration and Analyst System Effectiveness (CASE) Program, contract FA8750-06-C-0194 issued by Air Force Research Laboratory (AFRL). The views and conclusions are those of the authors, not of the US Government or its agencies.

6. References

[1] John Cheng, Ray Emami, Larry Kerschberg, Eugene Santos, Jr., Qunhua Zhao, Hien Nguyen, Hua Wang, Michael Huhns, Marco Valtorta, Jiangbo Dang, Hrishikesh Goradia, Jingshan Huang, and Sharon Xi, "OmniSeer: A Cognitive Framework for User Modeling, Reuse of Prior and Tacit Knowledge, and Collaborative Knowledge Services," *Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS38)*, Hawaii, January, 2005.

[2] Richards J. Heuer, Jr., *Psychology of Intelligence Analysis*, Center for the Study of Intelligence (at http://www.cia.gov/csi/books/19104/index.html), 1999.

[3] Yimin Huang and Marco Valtorta, "Pearl's Calculus of Intervention is Complete," *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI-06)*, pp. 217-224, 2006.

[4] Katherine Laskey and Suzanne Mahoney, "Network Fragments: Representing Knowledge for Constructing Probabilistic Models," *Proceeding of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, pp. 334-341, 1997.

[5] Judea Pearl, *Causality: Modeling, Reasoning, and Inference*, Cambridge, England: Cambridge University Press, 2000.

[6] Peter Pirolli and Lance Good, "Evaluation of a Computer Support Tool for Analysis of Competing Hypotheses," Palo Alto, CA: Palo Alto Research Center (UIR Technical Report), 2004.

[7] Marco G. Valtorta, Y.-G. Kim, and Jirka Vomlel, "Soft Evidential Update for Multiagent Systems," *International Journal of Approximate Reasoning*, 29, 1, pp. 71-106, January 2002.