Ontologies and Probabilities: Working Together for Effective Multi-INT Fusion

Eric Little (Director, Center for Ontology and interdisciplinary Studies, D'Youville College), Kathryn B. Laskey (Center of Excellence in C4I, George Mason University), Terry Janssen (Lockheed Martin Corporation)

Formal ontologies are becoming an essential tool for Intelligence analysis. An ontology provides upper- and domain-level category systems for decomposing and relating objects, object attributes/properties, temporal events, and relations of interest to the intelligence analyst. A great deal of intelligence analysis focuses on understanding and reacting to instance-level report data of varied fidelity on numerous kinds of entities, events and relations. Some of these entities and relations may not be represented explicitly within the ontology's categorical structure, but may be ingested from ancillary systems with which the ontology must interoperate. To an increasing degree, intelligence analysts rely on fusing reports from many different sources. These include reports from different kinds of sensors, processed intelligence from various systems and databases, and human intelligence. The common vocabulary and precisely specified semantics of formal ontologies is a critical enabling factor for interoperability. The promise of multi-INT fusion is that individually noisy and unreliable indicators can be brought together to form a common operating picture (COP) of a given situation [1]. Because the reports being combined may vary greatly in quality, it is essential to account for source quality in combining reports. This requires understanding data quality and applying methodologies for combining information that make use of data quality in a sound and principled manner. Probabilistic reasoning is a well-understood, theoretically sound, and generally applicable method for combining evidence from multiple sources of varying reliability. Computational probabilistic reasoning is a well-established and growing field of research and application (e.g., [2, 3, 4]. Probability has shown its value across a wide range of applications, and many qualitative and heuristic approaches to combining information have been explained as "fast and frugal" approximations to the normative probabilistic solution [5]. Until recently, there has been little research on marrying the fields of formal ontology and probabilistic reasoning. However, this situation is changing (e.g., [6]). This paper will address the question of how formal ontologies can best be combined with probability theory to provide theoretically sound and practically useful semantic technology for multi-INT fusion. We will investigate theoretical concerns associated with the connections between logics associated with formal ontology (e.g., description logic, common logic, first-order logic) and those of probabilistic mathematics. The goal is to provide a high-level discussion of the issues involved with combining ontologies and probabilistic systems as a basis for dialog between these two communities, and to identify a broadly construed research agenda for their mutual development and interaction. The authors of this paper argue the necessity of articulating a clear theoretical foundation as a basis for later development of specific methodologies and languages.

An important question, therefore, is how probabilistic formalisms such as Bayesian Networks can be merged with formal ontologies. Probabilistic theories produce qualified conclusions, graded by numerical measures of plausibility. By contrast, formal ontologies have focused on purely logical reasoning that leads to definite conclusions. Formal ontological categories are related to one another in definite, law-governed ways, and are understood as possessing a purely binary truth functionality. Formal ontologies are useful for data integration, particularly at the upper-most levels, because they provide for a logical structure for various categories and relations, independent of any particular material knowledge of a given domain [7]. In this sense, a formal ontology provides a means of understanding all types of objects, attributes and relations associated to one another within a given domain by understanding the most basic formal structures they share in common [8].

The question of how formal ontologies can be merged with probabilistic reasoning rests on first defining which items are *in* an ontology per se and which items are *associated with* the ontology (e.g., the reasoning engine, the query language, the results analyzer, etc.). An upper ontology provides asserted facts about the ontic world, meaning the world of 'general being' as opposed to any distinctive philosophical or scientific theory of that world – the ontological. So, according to this approach, an upper ontology provides a type of assumed god's eye view of reality, independent of human observations. By their very nature, human observations presume certain epistemic (i.e., mind- or knowledge-dependent assertions about reality (e.g., as discussed in the lengthy philosophical debates between realist and conceptualist theories of reality) [9-11]. At the upper-most levels, for example, an ontology normally contains non-recursive categorical relations such as: a TerroristAgent is_a Person, an IED is_a Explosive, an ObjectShape is_dependent_on Substance). The lattice of types and subtypes is a logical structure that is generally taken as given by both logical and probabilistic domain theories. While there may be competing upper ontologies, each with its own type lattice, generally within an ontology, there is no uncertainty associated with the categorical relations.

However, the situation changes when we consider the problem of categorizing instance data. As an example, consider an individual who is declared a Person-Of-Interest, and is being observed to assess whether or not he is engaging in terrorist activities. Information relevant to this problem includes, for example, the network of individuals with whom he associates, his religious affiliation, purchases he has recently made (e.g., materials that could be used to manufacture explosive devices), phone calls to individuals suspected of plotting an attack, etc. The decomposition of Person-Of-Interest into sub-categories of Terrorist and non-Terrorist is a purely logical assumption. However, categorizing an individual as a terrorist or non-terrorist would make use of probabilistic information, such as the base rate of terrorist versus non-terrorist individuals within the relevant population, the likelihood of the pattern of attributes and activities given that the individual is a terrorist versus a non-terrorist, and the credibilities of the reports on which we are basing our inferences about his attributes and activities (e.g., [12]).

Probability is an essential tool for performing this kind of inference in a systematic and principled manner. To perform this kind of reasoning, a system needs the basic categorical knowledge typically encoded in an ontology, and also the likelihood information needed by the probabilistic reasoner. This likelihood information can be obtained from statistical summaries of past instance data, from the judgment of experienced experts, from physical characteristics of sensing systems, or from some combination of the above. This information must be represented in computational form to be processed by probabilistic reasoning algorithms. Increasingly, with the proliferation of distributed fusion systems and web services, it is becoming important to represent this likelihood information not just internally within a given system, but also for consumption by other systems with which it interoperates. Data quality information must be represented as metadata associated with a web service. When a service returns a result on a situation-specific query, it often must return not just a most likely conclusion, but also information on the uncertainty associated with the conclusion, and also pedigree information to provide the consumer with an audit trail regarding how the conclusion was reached. Interoperating systems require not just shared vocabulary for domain concepts, but also shared vocabulary for communicating statistical regularities pertaining to categories in the ontology, as well as uncertainties associated with instance-level reasoning results, and pedigree information about how conclusions were reached.

An important research issue is how to combine the categorical and relational knowledge typically represented in ontologies with the likelihood knowledge required for multi-INT fusion. Tantamount to this research agenda is the analysis of how quantitative probabilistic reasoning interacts with qualitatively linked ontological categories. The goal of the current paper is to examine varied approaches to the interactions between ontologies and probabilistic systems, rather than present a clear-cut solution for implementing these kinds of systems (e.g., in multi-sensor fusion applications and the like). Likelihood information must be represented in computational form and combined appropriately with the categorical and relational knowledge contained in ontologies. There have been proposals (e.g., [12]) for augmenting ontology languages to represent probabilistic information. Others (e.g., [13]) have made use of non-probabilitic ontologies to represent structural features of a domain, and have incorporated probabilistic information from outside the ontology to construct a probabilistic model.

To illustrate how probabilistic reasoning can be combined with ontological reasoning, we have developed a simple probabilistic model for multi-INT fusion to identify and head off a potential terrorist attack. The representation language is multi-Entity Bayesian Networks (MEBN) [14]. MEBN Fragments (MFrags) represent small, separable components of probabilistic knowledge about the domain. These MFrags draw on knowledge about ontically existent objects, events and relations, which form *contexts* within which probabilistic reasoning is performed. In the full paper and the presentation, we will describe the case study problem in detail, describe how the likelihood information is used in conjunction with the categorical and relational knowledge, and discuss the question of how to combine probabilistic and ontological technology for problems of this kind.

Acknowledgement
The authors gratefully acknowledge support for this research from Lockheed Martin Corporation.

References

[1] Hall, D & Llinas, J. (eds.). (2001) Handbook of Multisensor Data Fusion, Boca Raton: CRC Press.

[2] Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA, Morgan Kaufmann.

[3] Bacchus, F. (1990). *Representing and Reasoning with Probabilistic Knowledge: A Logical Approach to Probabilities*, Boston, MA, MIT Press.

[4] S. Dzeroski and N. Lavrac, eds. (2001) *Relational Data Mining*. Berlin, Springer-Verlag.

[5] Martignon, L. and K. B. Laskey (1999). Taming Wilder Demons: Bayesian Benchmarks for Fast and Frugal Heuristics. In *Simple Heuristics that Make us Smart*. The ABC Group, Oxford University Press.

[6] Costa, P., Laskey, K.B., and Lukasiewicz, T. Uncertainty Representation and Reasoning in the Semantic Web, under review for inclusion in J. Cardoso and M. Lytras (eds), *Semantic Web Engineering in the Knowledge Society*, IRM Press, forthcoming.

[7] Smith, B & Smith, D.W. (eds.). (1995) *The Cambridge Companion to Husserl* - Cambridge: Cambridge University Press. pp. 27-29.

[8] Little, E. (2006) "The Need for Metaphysically-based Ontologies in Higher-level Information Fusion Applications," in IFOMIS Report 14 (04/06) Johansson, Klein, Roth-Berghofer (eds.) *WSPI 2006: Contributions to the Third International Workshop on Philosophy and Informatics* (online at: CEUR-WS.org).

[9] Little, E. & Vizenor, L. (2006) "Principles for the Development of Upper Ontologies in Higher-level Information Fusion Applications," in *Proceedings of the 2006 Formal Ontology for Information Systems Conference*, Baltimore, Maryland, November 9-11.

[10] Little, E. & Rogova, G. (2006) "An Ontological Analysis of Threat and Vulnerability," in *Proceedings of the FUSION 2006-9th International Conference on Multisource Information Fusion*, July 10-13, Florence, Italy.

[11] Little, E. & Rogova, G. (2005) "Ontology Meta-Model for Building A Situational Picture of Catastrophic Events," in *Proceedings of the FUSION 2005-8th International Conference on Multisource Information Fusion*, July 25-29, Philadelphia, PA.

[12] Wright, E. and Laskey, K.B. Credibility Models for Multi-Source Fusion, *Proceedings of the Ninth International Conference on Information Fusion*, July 2006.

[13] Poole, D. and Smyth, C. Ontologically-Grounded Qualitative Probabilistic Matching, *Proceedings of the Fifth Annual Bayesian Modeling Applications Workshop*, Vancouver, Canada, July 2006.

[14] Laskey, K.B. (in press) MEBN: A Language for First-Order Bayesian Knowledge Bases. *Artificial Intelligence*.