

Unifying Logic and Probability: A New Dawn for AI?

Stuart Russell

¹ University of California, Berkeley CA 94720, USA

russell@cs.berkeley.edu

<http://www.cs.berkeley.edu/~russell>

² Université Pierre et Marie Curie, Laboratoire d'informatique de Paris 6,
4 place Jussieu, 75252 Paris Cedex 05, France

Abstract. Logic and probability theory are two of the most important branches of mathematics and each has played a significant role in artificial intelligence (AI) research. Beginning with Leibniz, scholars have attempted to unify logic and probability. For “classical” AI, based largely on first-order logic, the purpose of such a unification is to handle uncertainty and facilitate learning from real data; for “modern” AI, based largely on probability theory, the purpose is to acquire formal languages with sufficient expressive power to handle complex domains and incorporate prior knowledge. This paper provides a brief summary of an invited talk describing efforts in these directions, focusing in particular on *open-universe* probability models that allow for uncertainty about the existence and identity of objects.

Keywords: first-order logic, probability, probabilistic programming, Bayesian logic, machine learning.

1 Introduction

From its earliest days, AI adopted the idea of *declarative* system reasoning over explicitly represented knowledge with a general inference engine. Such systems require a formal language to express knowledge about the real world; and *the real world has things in it*. For this reason, in 1958, McCarthy [16] proposed first-order logic—the mathematics of objects and relations—as the foundation for what we now call “classical AI.”

The key benefit of first-order logic is its expressive power, which leads to concise—and hence easily learnable—models. For example, the rules of chess occupy 10^0 pages in first-order logic, 10^5 pages in propositional logic, and 10^{38} pages in the language of finite automata. The power comes from separating predicates from their arguments and quantifying over those arguments: so one can write rules about $On(p, c, x, y, t)$ (piece p of color c is on square x, y at move t) without having to fill in each specific value for $c, p, x, y,$ and t .

A second research tradition, sometimes called “modern AI,” developed around another important property of the real world: *pervasive uncertainty* about both

its state and its dynamics. Modern AI is based on probability theory, which provides principled methods for learning and making predictions from observations. The key advance underlying modern AI was the development of *Bayesian networks* [22] and the related family of undirected *graphical models* [6]. Bayes nets provided a formal language for probability models and enabled rapid advances in machine learning, vision, natural language understanding, and knowledge-based systems. The expressive power of Bayes nets is, however, limited. They assume a fixed set of *variables*, each of which can take on a value from a fixed *range*; thus, they are a *propositional* formalism, like Boolean circuits. The rules of chess and of many other domains are beyond them.

What happened next, of course, is that classical AI researchers noticed the pervasive uncertainty, while modern AI researchers noticed, or remembered, that the world has things in it. Both traditions arrived at the same place: *the world is uncertain and it has things in it*. To deal with this, we have to *unify logic and probability*.

But how? Even the meaning of such a goal is unclear. Early attempts by Leibniz, Bernoulli, De Morgan, Boole, Peirce, Keynes, Carnap, and Gaifman (surveyed in [8,10]) involved attaching probabilities to logical sentences. This line of work influenced AI research [9,3,14] but has serious shortcomings as a vehicle for representing knowledge. An alternative approach, arising from both branches of AI and from statistics, draws on the compositional semantics of Bayes nets. Some tools use programming constructs to build very large Bayes nets with repeated structure [7,4,15], while others adopt the syntactic and semantic devices of logic (composable function symbols, logical variables, quantifiers) to create declarative, first-order probabilistic languages [5,23,25,12,11].

Despite their successes, these approaches miss an important consequence of uncertainty in a world of things: there will be *uncertainty about what things are in the world*. Real objects seldom wear unique identifiers or preannounce their existence like the cast of a play. In the case of vision, for example, the existence of objects must be *inferred* from raw data (pixels) that contain no explicit object references at all. If, however, one has a probabilistic model of the ways in which worlds can be composed of objects and of how objects cause pixel values, then inference can propose the existence of objects given only pixel values as evidence. Similar arguments apply to areas such as natural language understanding, web mining, and computer security.

The difference between knowing all the objects in advance and inferring their existence and identity from observation corresponds to an important but often overlooked distinction between *closed-universe* languages such as SQL and logic programs and *open-universe* languages such as full first-order logic.

This distinction is best understood in terms of the *possible worlds* under each type of semantics. Figure 1(a) shows a simple example with two constants and one binary predicate. Notice that first-order logic is an open-universe language: even though there are two constant symbols, the possible worlds allow for 1, 2, or indeed arbitrarily many objects. A closed-universe language enforces additional assumptions that restrict the set of possible worlds:

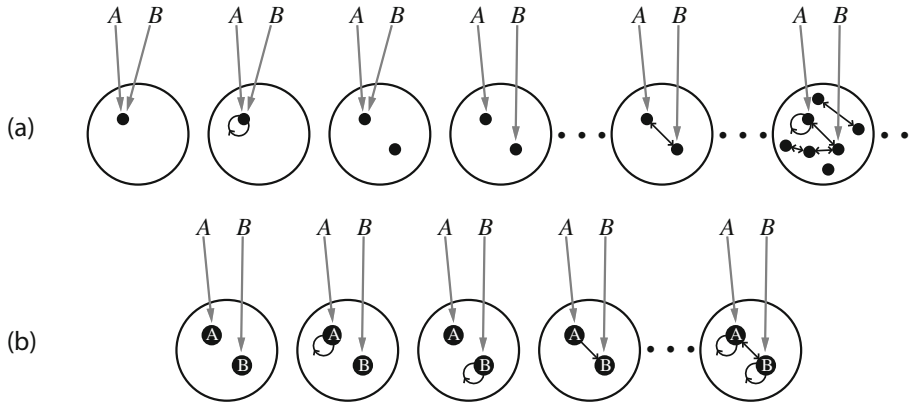


Fig. 1. (a) Some of the first-order possible worlds for a language with two constant symbols, A and B , and one binary predicate. Arrows indicate the interpretation of each constant symbol and the relations between objects. (b) The analogous figure under closed-universe semantics.

- The *unique names* assumption requires that distinct terms must refer to distinct objects.
- The *domain closure* assumption requires that there are no objects other than those named by terms.

These two assumptions mean that every possible world contains the same objects, which are in one-to-one correspondence with the ground terms of the language (see Figure 1(b)).¹

A formal probability model must specify the probability of every possible world given the vocabulary (predicates, functions, constants) of the model's syntactic representation. Obviously, the set of worlds under open-universe semantics is larger and more heterogeneous, which makes the task of defining open-universe probability models more challenging. The core part of the talk is concerned with a first-order, open-universe probabilistic language called *Bayesian logic* or BLOG [18,19]. BLOG was developed primarily as the PhD thesis research of Brian Milch [17]. The key results derived for BLOG are the following:

- Every well-formed BLOG model specifies a well-defined probability distribution over the possible worlds constructed from the vocabulary of the model.
- There exist Monte Carlo algorithms that provably converge (subject to technical conditions on the conditional distributions of the model) to the correct posterior probability for any first-order query for any well-formed BLOG model [20,1].

¹ The difference between open and closed universes can also be illustrated with a common-sense example. Suppose a system knows just two sentences, $Father(William) = Bill$ and $Father(Junior) = Bill$. How many children does Bill have? Under closed-universe semantics—e.g., in a database system—he has exactly 2; under open-universe semantics, between 1 and ∞ .

The generic algorithms (importance sampling and MCMC applied to a dynamically constructed ground representation) are often too slow for practical use on large models. Several avenues are being pursued for speeding up inference, including special-purpose block samplers for variables constrained by deterministic relationships [13], static analysis to identify submodels amenable to efficient inference, lifted inference to avoid grounding by manipulating symbolic distributions over large sets of objects [24,26], and compiler techniques to generate model-specific inference code.

More than two dozen BLOG models have been developed, covering a wide variety of standard machine learning models as well as applications including citation matching [21] and global seismic monitoring for the Comprehensive Nuclear Test-Ban Treaty [2].

2 Prospects

These are very early days in the process of unifying logic and probability. We need much more experience in developing models for a wide range of applications. Undoubtedly there are new modeling idioms, programming constructs, and inference algorithms to discover.

The development of Bayes nets in the late 1980s connected machine learning to statistics and reconnected (modern) AI with vision and language. It is possible that first-order probabilistic languages, which have both Bayes nets and first-order logic as special cases, can serve a similar, but more inclusive, unifying role.

Acknowledgments. The author is supported by the *Chaire Blaise Pascal*, funded by the l'État et la Région Île de France and administered by the Fondation de l'École Normale Supérieure. The research is also supported by DARPA contract FA8750-14-C-0011 for the APPRIL project under the PPAML program and previously by DARPA contract FA8650-11-1-7153 for the OUTBIDS project under the MSEE program.

References

1. Arora, N., Russell, S., de Salvo Braz, R., Sudderth, E.: Gibbs sampling in open-universe stochastic languages. In: UAI 2010 (2010)
2. Arora, N.S., Russell, S., Sudderth, E.: NET-VISA: Network processing vertically integrated seismic analysis. *Bull. Seism. Soc. Amer.* 103 (2013)
3. Bacchus, F.: *Representing and Reasoning with Probabilistic Knowledge*. MIT Press (1990)
4. Bessière, P., Mazer, E., Ahuactzin, J.M., Mekhnacha, K.: *Bayesian programming*. CRC (2013)
5. Breese, J.S.: Construction of belief and decision networks. *Computational Intelligence* 8, 624–647 (1992)
6. Darroch, J.N., Lauritzen, S.L., Speed, T.P.: Markov fields and log-linear interaction models for contingency tables. *The Annals of Statistics* 8(3), 522–539 (1980)

7. Gilks, W.R., Thomas, A., Spiegelhalter, D.J.: A language and program for complex Bayesian modelling. *The Statistician* 43, 169–178 (1994)
8. Hailperin, T.: Probability logic. *Notre Dame J. Formal Logic* 25(3), 198–212 (1984)
9. Halpern, J.Y.: An analysis of first-order logics of probability. *AIJ* 46(3), 311–350 (1990)
10. Howson, C.: Probability and logic. *J. Applied Logic* 1(3-4), 151–165 (2003)
11. Kersting, K., De Raedt, L.: Bayesian logic programs. In: *ILP 2000* (2000)
12. Koller, D., Pfeffer, A.: Probabilistic frame-based systems. In: *AAAI 1998* (1998)
13. Li, L., Ramsundar, B., Russell, S.: Dynamic scaled sampling for deterministic constraints. In: *AI/Stats 2013* (2013)
14. Lukasiewicz, T.: Probabilistic logic programming. In: *ECAI* (1998)
15. McCallum, A., Schultz, K., Singh, S.: FACTORIE: Probabilistic programming via imperatively defined factor graphs. *NIPS 22* (2010)
16. McCarthy, J.: Programs with common sense. In: *Proc. Symposium on Mechanisation of Thought Processes*. Her Majesty’s Stationery Office (1958)
17. Milch, B.: Probabilistic Models with Unknown Objects. Ph.D. thesis, UC Berkeley (2006)
18. Milch, B., Marthi, B., Sontag, D., Russell, S.J., Ong, D., Kolobov, A.: BLOG: Probabilistic models with unknown objects. In: *IJCAI 2005* (2005)
19. Milch, B., Russell, S.: Extending Bayesian networks to the open-universe case. In: Dechter, R., Geffner, H., Halpern, J. (eds.) *Heuristics, Probability and Causality: A Tribute to Judea Pearl*. College Publications (2010)
20. Milch, B., Russell, S.J.: General-purpose MCMC inference over relational structures. In: *UAI 2006* (2006)
21. Pasula, H., Marthi, B., Milch, B., Russell, S.J., Shpitser, I.: Identity uncertainty and citation matching. *NIPS 15* (2003)
22. Pearl, J.: *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann (1988)
23. Poole, D.: Probabilistic Horn abduction and Bayesian networks. *AIJ* 64, 81–129 (1993)
24. Poole, D.: First-order probabilistic inference. In: *IJCAI 2003* (2003)
25. Sato, T., Kameya, Y.: PRISM: A symbolic statistical modeling language. In: *IJCAI 1997* (1997)
26. Van den Broeck, G.: *Lifted Inference and Learning in Statistical Relational Models*. Ph.D. thesis, Katholieke Universiteit Leuven (2013)