

Simple Rules for Probabilistic Commonsense Reasoning

Adam Purtee and Lenhart Schubert

University of Rochester
Rochester, New York
{apurtee,schubert}@cs.rochester.edu

Abstract

A long-sought goal of AI is the mechanization of human-like reasoning and argumentation based not only on firm knowledge supporting deduction but also generic conditional knowledge supporting tentative or probabilistic inferences. Arguably, most of the general knowledge people possess is of this type. We propose probabilistic “amplifying rules”, “Bayes rules”, and “categorization rules”, combined with an algebraic conception of probability, as a basis for commonsense probabilistic reasoning, i.e., the generation of new, probabilistically qualified propositions from a commonsense knowledge base. The proposed framework allows for both “hard” (e.g., universally quantified) knowledge and knowledge in the form of the above types of rules, expressing that under certain specified conditions, a conclusion is warranted to some degree, or the members of some spectrum of alternative conclusions are warranted to varying degrees.

Our rules are meant to be quantitative formalizations of generic statements about the world. Such statements, in sentential form, come in two distinct types: ones that generalize over individuals that realize *kinds* (such as the kind, dogs, or the kind Rottweilers); and ones that generalize over events or situations (e.g., dogs often bark when agitated; people usually use paint rollers to paint walls; etc.). We can view many generic sentences as statistical claims about the world (sometimes quite explicitly so, as in *One out of four times, the airlines will lose your checked baggage when you travel*). We regard our rules as grounded in statistical claims as well, although, when applied to particular instances, the probabilities they lead to are no longer statistical but rather degrees of belief. Our proposed rules are able to capture a wide variety of statistical knowledge expressed in natural language.

More specifically, amplification/attenuation rules are used when a direct causal (or anti-causal) relationship is known to hold. E.g., watering the lawn daily is known to cause grass to grow, whereas sowing salt into the soil has the opposite effect. The net effect of an a-rule is monotonic, in the sense that even when their corresponding parameters are arbitrarily small, the likelihood of the consequent is monotonically increased (or decreased, in the case of attenuation rules).

An example application is the claim that for each common friend between a pair of persons, the likelihood that the pair are friends increases. We can capture this with a single amplification rule, and the consequences thereof are that as the number of common friends increases, so does the likelihood of transitive friendship, and that the strength of the boost in friendship likelihood decays with the distance, i.e., the number of intermediary friendships necessary to establish a transitive connection.

Two key processes are portrayed in this example: the accumulation of evidence from common friendships, and the reduced evidential strength derived from distant transitive connections in comparison with more direct ones. These complementary processes find expression in our approach as evidence combination and forward chaining, and we have developed formal rules for such reasoning with each of our three classes of proposed rules.

Categorization rules are intended to capture probability distributions over mutually exclusive alternatives, and are particularly well-suited to modeling proportional partitioning of kinds (of entities or events) into sub-kinds.

Bayesian rules are so-named to evoke the familiar Bayesian rule of conditioning, upon which they are based. The direction of their effect can be positive or negative, depending both on the rule parameter and the prior probability of the modified event.

Empirically, our rules are able to exactly match the outputs of an artificially constructed, medical diagnosis-themed Bayesian network, and in principle can represent more general Bayesian networks as well.

Further, we verified that with a simple, two-rule baseline, we are able to match performance of Richardson and Domingos [1] on the UW-CSE dataset, which consists of predications about individual members of a computer science department, publication authorship, teaching roles, and teaching-assistant roles, using these to infer advisor-advisee relations by means of Markov Logic Networks (MLNs). We obtain an area under the precision-recall curve of 0.225 as compared to the best reported results in that work of 0.220. While methods enhanced with structure-learning techniques later improved upon the original MLN results, our results provide an empirical verification of the utility of our approach. We emphasise that these results are preliminary, and represent baseline performance with a-rules only.

Our work connects two distinct threads of research into logical representations suitable for capturing the kinds of knowledge expressed through natural language [2], as well as an algebraic characterization of probability which is closely connected to logical and-or-not networks [3].

The existing approaches that come closest to enabling such general reasoning as our method are nonmonotonic reasoning (NMR) and production systems (also called rule-based systems or expert systems). NMR methods extend deductive methods to allow conclusions to be drawn tentatively in the absence of contrary evidence. However, they fail to allow for the quite subtle *quantitative* intuitions that people have about the world around them, thanks to both “statistical experience” and verbally transmitted knowledge.

Our framework for commonsense reasoning refines NMR methods by allowing for rules whose consequent is affirmed with some degree of certainty (or with a “spectrum” of consequents affirmed with various degrees of certainty). In relation to production systems, our rules operate on logical, formally interpretable formulas, are restricted to a set of simple, natural types, allow for alternatives in their conclusions, and interact in well-motivated ways when applied in sequence or in parallel.

Being built upon an extension of first-order logic, our method is compatible with quantified and ground deduction, and supplies transparent provenance for arguments about belief strengths. We strictly generalize path-based methods of nonmonotonic reasoning which only operate with “usually is-a” and “usually is-not-a” relationships. Our numerical parameters have natural interpretations as conditional probabilities, and thanks to its expressivity, our representation is well-suited to working with knowledge expressed through natural language.

Acknowledgements

The work was supported in part by DARPA grant W911NF-15-1-0542.

References

- [1] Matthew Richardson and Pedro Domingos. Markov logic networks. *Mach. Learn.*, 62(1-2):107–136, February 2006.
- [2] Lenhart K. Schubert and Chung Hee Hwang. Natural language processing and knowledge representation. chapter Episodic logic meets Little Red Riding Hood: a comprehensive natural representation for language understanding, pages 111–174. MIT Press, Cambridge, MA, USA, 2000.

- [3] Lenhart K Schubert. A new characterization of probabilities in bayesian networks. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pages 495–503. AUAI Press, 2004.