Probabilistic Relational Learning and Inductive Logic Programming at a Global Scale

David Poole

Department of Computer Science,
University of British Columbia
Vancouver, BC, Canada
http://www.cs.ubc.ca/~poole/

Abstract. Building on advances in statistical-relational AI and the Semantic Web, this talk outlined how to create knowledge, how to evaluate knowledge that has been published, and how to go beyond the sum of human knowledge. If there is some claim of truth, it is reasonable to ask what evidence there is for that claim, and to not believe claims that do not provide evidence. Thus we need to publish data that can provide evidence. Given such data, we can also learn from it. This talk outlines how publishing ontologies, data, and probabilistic hypotheses/theories can let us base beliefs on evidence, and how the resulting world-wide mind can go beyond the aggregation of human knowledge. Much of the world's data is relational, and we want to make probabilistic predictions in order to make rational decisions. Thus probabilistic relational learning and inductive logic programming need to be a foundation of the semantic web. This talk overviewed the technology behind this vision and the considerable technical and social problem that remain.

Keywords: Statistical relational AI, Probabilistic Relational Learning, Semantic Science, Lifted Probabilistic Inference, World-Wide Mind

To make decisions, we should base them on the best information available; we need to be able to find all relevant information and condition on it effectively. We have called the technology to support decisions "semantic science" [7, 8], based on the semantic web, which is an endeavor to make all of the world's knowledge accessible to computers, and using scientific methodology to make predictions. Figure 1 shows the main components of semantic science. Ontologies are used to define the vocabulary of data and hypotheses. These are needed to enable us to find and use all relevant information. Observational data, which depends on the world and the ontologies, are published. Such data sets can be very heterogenous, at widely varying levels of abstraction and detail. Hypotheses that make probabilistic predictions are also published. Hypotheses are not created in isolation, but depend on some training data. Hypotheses can be judged by their prior plausibility and how well they predict the data. Given a new case, various hypotheses are combined to form models that can be used to make predictions on that case. Given a prediction, users can ask what hypotheses were used to make that prediction, and for each hypothesis, users can find the relevant data to evaluate the hypothesis. In this way decisions can be based on all of the applicable evidence.

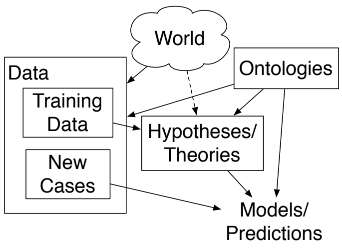

Fig. 1. Semantic Science

Typically data is not just a set of mappings from features into a fixed set of values, as is often assumed by traditional machine learning, but often refers to individuals that are only referred to by name; it is the properties of these individuals and the relationships among these individuals that is important for prediction. Following the publication of what could be argued was the first probabilistic relational language [1, 2, 5], the combination of logical and probabilistic reasoning, and probabilistic programming languages [6] has blossomed. There are still many open fundamental problems for representations, inference and learning. [3] proposed the problem of lifted inference: carrying out probabilistic inference reasoning about classes of individuals as a unit, without reasoning about them individually. Another problem is where models refer to individuals in terms of the roles they fill, but the data does not label the observed individuals with roles [4]. There is still lots of exciting research to be done!

References

1. Poole, D.: Probabilistic Horn abduction and Bayesian networks. Artificial Intelligence 64(1), 81–129 (1993)
2. Poole, D.: The independent choice logic for modelling multiple agents under uncertainty. Artificial Intelligence 94, 7–56 (1997), special issue on economic principles of multi-agent systems
3. Poole, D.: First-order probabilistic inference. In: Proc. Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-03). pp. 985–991. Acapulco, Mexico (2003)
4. Poole, D.: Logical generative models for probabilistic reasoning about existence, roles and identity. In: 22nd AAAI Conference on AI (AAAI-07) (July 2007)
5. Poole, D.: The independent choice logic and beyond. In: De Raedt, L., Frasconi, P., Kersting, K., Muggleton, S. (eds.) Probabilistic Inductive Logic Programming: Theory and Application. LNAI 4911, Springer Verlag (2008)
6. Poole, D.: Probabilistic programming languages: Independent choices and deterministic systems. In: Dechter, R., Geffner, H., Halpern, J. (eds.) Heuristics, Probability and Causality: A Tribute to Judea Pearl, pp. 253–269. College Publications (2010)
7. Poole, D., Smyth, C., Sharma, R.: Semantic science: Ontologies, data and probabilistic theories. In: da Costa, P.C., d'Amato, C., Fanizzi, N., Laskey, K.B., Laskey, K., Lukasiewicz, T., Nickles, M., Pool, M. (eds.) Uncertainty Reasoning for the Semantic Web I. LNAI/LNCS, Springer (2008)
8. Poole, D., Smyth, C., Sharma, R.: Ontology design for scientific theories that make probabilistic predictions. IEEE Intelligent Systems 24(1), 27–36 (Jan/Feb 2009)