Markov Logic Networks: A Step Towards a Unified Theory of Learning and Cognition

Pedro Domingos
Department of Computer Science and Engineering
University of Washington
Seattle, WA 98195-2350, U.S.A.
pedrod@cs.washington.edu

To the best of our current knowledge, the cortex uses essentially the same learning and inference algorithms throughout. If this hypothesis is correct, discovering these algorithms should be the holy grail of machine learning. One way to pursue it is to focus on the neuroscience, attempting to understand and formalize what the "wetware" is doing. However, progress in this direction is hindered by our very limited current understanding of neurophysiology and (especially) neuroanatomy. Another approach is to focus on a simple task (e.g., digit recognition), develop learning algorithms that work on it, and hope they generalize to everything else the brain does. While this allows for a rapid experimental cycle, it seems unlikely that a single simple task will exhibit all the attributes of intelligence we need to capture, and we will need a lot of luck to find the "fundamental algorithms" in this way.

A third, and perhaps the most promising, approach, is to consider all the different things the brain does, and try to understand what they have in common. Is there a single set of capabilities that underlies vision, motion, language, common-sense reasoning, etc.? Viewed in this light, the entire AI and cognitive science literature becomes a rich source of potential clues. In particular, two themes appear repeatedly: the ability to handle noise, uncertainty, and incomplete information, which is well captured by probability and graphical models; and the ability to deal with complex situations, involving multiple objects, classes of objects, and relations among them, which is well captured by first-order logic.

A good place to start might then be to unify graphical models and first-order logic, develop learning and inference algorithms for the unified representation, and apply them to a wide variety of problems that people are good at. This is the direction we have pursued for the last few years. In particular, we have developed Markov logic networks (MLNs), a formalism that unifies logic and probability by assigning weights to first-order formulas and viewing them as templates for features of Markov networks [2].

We and other groups have made rapid progress in developing increasingly scalable and robust learning and inference algorithms for MLNs. The deep connections between logic and probability open up a rich field of possibilities. For example, the MC-SAT algorithm combines MCMC and satisfiability testing, is able to handle both deterministic and probabilistic dependencies, and is many orders of magnitude faster than Gibbs sampling and other MCMC techniques. Lifted belief propagation uses ideas from resolution and unification in first-order logic to greatly speed up BP. Using these algorithms, we routinely carry out inference in networks with millions of variables, billions of features, and large treewidths.

State-of-the-art learning algorithms for MLNs combine ideas from statistical learning and inductive logic programming. The former allow us to learn MLN weights, and the latter allow us to learn MLN structure. With the combination of the two, we can learn networks with thousands of latent variables and multiple layers of hidden structure. This would be very difficult under the standard assumption of i.i.d. (independent and identically distributed) data. However, logic allows us to easily model dependences between objects, potentially making the entire sample available to predict each hidden variable (as opposed to only the other properties of the same object). Further, using MLNs we can do transfer learning across domains that have no variables in common, but share structural regularities

[1]. The ability to do this is arguably the hallmark of human learning, and something that machine learning has previously lacked almost entirely.

The learning and inference algorithms developed to date are available in the open-source Alchemy package [4] and others. Alchemy has been downloaded over 5000 times, and a substantial research community has developed around MLNs and related representations [3]. MLNs have been successfully applied to natural language processing, robot navigation, video segmentation, activity recognition, social network analysis, and many other problems. In each of these applications, the MLNs used comprise only a few to tens of formulas, but outperform much more complex state-of-the-art systems. This is a direct result of the deep integration of logic, probability, learning and inference, and bodes well for the ultimate goal of finding the fundamental algorithms underlying learning and cognition.

As an example of what can be achieved, the USP system provides an end-to-end solution to the problem of reading free text, building an internal representation of its content, and answering questions based on it [6]. The learning is unsupervised; its only input is the dependency parses of the sentences in the text. USP works by recursively clustering expressions that occur with similar subexpressions, and learns both a semantic parser and a knowledge base of semantically parsed sentences. These can then be used to answer questions. USP is able to extract knowledge from very complex text, and greatly outperformed other systems on the Genia/PubMed corpus of biomedical abstracts.

A key research direction is tighter integration of learning and inference. Two MLNs with similar accuracy can have drastically different inference costs. Using inference complexity as a regularization penalty can thus have enormous benefits [5]. Conversely, because inference is a subroutine of MLN learning, targeting efficient representations can itself make learning much more tractable and reliable. In particular, we plan to explore the connections between deep learning and compact multilayer logical representations like Boolean decision diagrams.

References

[1] J. Davis and P. Domingos. Deep transfer via second-order Markov logic. In *Proceedings of the Twenty-Sixth International Conference on Machine Learning*, Montreal, Canada, 2009. Omnipress.

[2] P. Domingos and D. Lowd. *Markov Logic: An Interface Layer for Artificial Intelligence*. Morgan & Claypool, San Rafael, CA, 2009.

[3] L. Getoor and B. Taskar, editors. *Introduction to Statistical Relational Learning*. MIT Press, 2007.

[4] S. Kok, M. Sumner, M. Richardson, P. Singla, H. Poon, D. Lowd, J. Wang, and P. Domingos. The Alchemy system for statistical relational AI. Technical report, Department of Computer Science and Engineering, University of Washington, Seattle, WA, 2008. http://alchemy.cs.washington.edu.

[5] D. Lowd and P. Domingos. Learning arithmetic circuits. In *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence*, Helsinki, Finland, 2008. AUAI Press.

[6] H. Poon and P. Domingos. Unsupervised semantic parsing. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, Singapore, 2009. ACL.

Topic: Learning algorithms, graphical models.

Preference: Oral.