Representational Power of Probabilistic-Logical Models: From Upgrading to Downgrading[*]

Kristian Kersting
Institute for Computer Science
Machine Learning Lab
Albert-Ludwigs University of Freiburg
Georges-Koehler-Allee, Building 079
79110 Freiburg, Germany

1 Position Statement

There is a diversity of *probabilistic-logical models* (PLM). No clear understanding of the relative advantages and limitations of different formalisms and their language concepts has yet emerged. To overcome this, we propose to *downgrade* highly expressive PLMs. This method has several advantages: one can profit from existing research on PLMs and inherit unique semantics, and inference and learning algorithms. Moreover, there is a clear relationship between the new PLM and its more expressive counterpart. No single existing approach is devalued.

2 Motivation

In recent years, there has been an increasing interest in probabilistic-logical models (PLMs). PLMs integrate probability theory with some first order logic. Traditionally, a probabilistic formalism like Bayesian networks or hidden Markov models is selected and *upgraded* by incorporating some logic such as entity-relationship (ER) models or Prolog. Real-world data applications have shown the potential of PLMs e.g. in query optimization [Getoor *et al.*, 2001], computational biology [Segal *et al.*, 2001; Kersting *et al.*, 2003] and web mining [Anderson *et al.*, 2002].

Despite these successes, the field of (learning) PLMs is quite complex and confusing. PLMs *"have been developed in several related, but different, subareas of artificial intelligence (reasoning under uncertainty, inductive logic programming, machine learning, and knowledge discovery and data mining)"* as stated by Lise Getoor and David Jensen in SRL-2003's CFP. Each subarea focuses on its own language concepts. Consider Table 1 which lists a subset of proposed formalisms [1]. The language concepts vary from acyclic to cyclic models, from logically structured dependencies among random variables to states, from finite to continuous random variables, and from functor-free languages to Prolog. They each have their respective merits. However, the *upgrading* mentality together with concentrating on particular language concepts makes a general understanding of PLMs and learning PLMs difficult – if not impossible.

3 Downgrading

Downgrading consists of two steps.

(Step 1) Choose a generally applicable (learning) PLM.

The PLM should cover the basic language concepts proposed in the different scientific subareas:

- **Relations** among random variables or states to model uncertainty. This subsumes interesting concepts such as referential and existential uncertainty.
- **Functors** allow to consistently encode temporal correlations (dynamic Bayesian networks), complex long distance correlations (stochastic grammars), *named by structure* entities as they are common in semi-structured data (e.g. XML), and general data structures (lists, trees, etc). Functors incorporate flexible memory capabilities.
- **Finite**, **discrete** and **continuous random variables** together provide compact models which are applicable in a broad field of applications such as classification, clustering, and regression.
- Often, e.g. in computational biology, one is interested not only to simulate but to gain insight into, and understand the underlying processes. Therefore, PLMs should be **interpretable**.
- Learning the PLM should facilitate to define and to specify both deterministic and probabilistic **background knowledge**. This not only makes it possible to specify the huge amount of expert knowledge often available but also to break complex questions into subtasks still taking care of dependencies among the subtasks.

It is likely that the very general PLM is prohibitively powerful for a problem at hand. Therefore,

(Step 2) downgrade it to strike the right balance between expressivity and learnability.

Compared to upgrading, downgrading has the following advantages. First, the downgraded PLM inherits unique semantics, and inference and learning algorithms. Second, downgrading does not focus on a particular PLM. Instead it systematically investigates the impact of language concepts. A general understanding of PLMs and learning PLMs is likely to emerge.

---

[*] This is a position statement for the IJCAI-2003 Workshop on Learning Statistical Models from Relational Data

[1] Avi Pfeffers's interesting PhD thesis provides some more references, [Pfeffer, 2000].

PLM	Probabilistic Formalism	Logic
Probabilistic Horn Abduction (PHA) [Poole, 1993]	Bayesian Networks	Prolog
PRISM [Sato, 1995]	Stochastic Grammars	Prolog
Stochastic Logic Programs (SLPs) [Muggleton, 1996; Cussens, 2000]	Stochastic Grammars	Prolog
Probabilistic Logic Programs (PLPs) [Ngo and Haddawy, 1997]	Bayesian Networks	Prolog
Bayesian Logic Programs (BLPs) [Kersting and De Raedt, 2001]	Bayesian Netwoks	Prolog
Relational Baysian networks (RBNs) [Jaeger, 1997]	Bayesian Networks	Relational
Probabilistic Relational Models (PRMs) [Friedman *et al.*, 1999]	Bayesian Networks	ER Models
Relational Markov Models (RMMs) [Anderson *et al.*, 2002]	Markov Models	Relational
Logical Hidden Markov Models (LOHMMs) [Kersting *et al.*, 2002]	Hidden Markov Models	Iterative Clauses

Table 1: A collection of probabilistic-logical models together with their underlying probabilistic and logical formalism.

Initial attempts of downgrading have been done. Restricting SLPs to iterative clauses leads in principle to LOHMMS [Kersting *et al.*, 2002]. [Sato and Kameya, 2001] propose an EM algorithm for parameter estimation of PRISMs showing that the algorithm exhibits the same complexity for hidden Markov models and stochastic context free grammars as the specialized counterparts.

4 Related Work

Downgrading is related to work comparing the expressivity of different PLMs [Kersting and De Raedt, 2001; Jensen and Neville, 2002]. Furthermore, downgrading is akin to contemporary considerations in the *inductive logic programming* and the *Baysian networks* communities. E.g. Kevin Murphy motivates the development of his Matlab *Baysian Network Toolbox* as follows: *"I was fed up with reading papers where all people do is figure out how to do exact inference and/or learning in a model which is just a trivial special case of a general Bayes net, e.g., input-output HMMs, coupled-HMMs, auto-regressive HMMs. My hope is that, by releasing general purpose software, the field can move on to more interesting questions"*, see `http://www.ai.mit.edu/~murphyk/Software/BNT/bnt.html`. For similar reasons, we initiated a repository for (learning) PLMs at `http://www.informatik.uni-freiburg.de/~kersting/plmr/`.

Acknowledgements

The position of the author was partially consolidated by the interesting discussion within the successful EU assessment project IST-2001-33053 (APRIL). The author would like to thank all collaborators and especially the reviewers. Special thanks to Lise Getoor and David Jensen for discussions about a PLM repository.

References

[Anderson *et al.*, 2002] C. R. Anderson, P. Domingos, and D. S. Weld. Relational Markov Models and their Application to Adaptive Web Navigation. In *Proceedings of the Eighth International Conference on Knowledge Discovery and Data Mining (KDD-2002)*, 2002.

[Cussens, 2000] J. Cussens. Parameter estimation in stochastic logic programs. *Machine Learning*, 44(3):245–271, 2000.

[Friedman *et al.*, 1999] N. Friedman, L. Getoor, D. Koller, and A. Pfeffer. Learning probabilistic relational models. In *Proceedings of Sixteenth International Joint Conference on Artificial Intelligence (IJCAI-1999)*, pages 1300–1307, 1999.

[Getoor *et al.*, 2001] L. Getoor, B. Taskar, and D. Koller. Selectivity estimation using probabilistic models. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 2001).*, 2001.

[Jaeger, 1997] M. Jaeger. Relational Bayesian networks. In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 266–273. Morgan Kaufmann, 1997.

[Jensen and Neville, 2002] D. Jensen and J. Neville. Schemas and models. In *Proceedings of the Multi-Relational Data Mining Workshop, 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2002.

[Kersting and De Raedt, 2001] K. Kersting and L. De Raedt. Towards Combining Inductive Logic Programming with Bayesian Networks. In *Proceedings of the 11th International Conference on Inductive Logic Programming*, volume 2157 of *LNAI*, pages 118–131. Springer, 2001.

[Kersting *et al.*, 2002] K. Kersting, T. Raiko, and L. De Raedt. Logical Hidden Markov Models (Extended Abstract). In *Proceedings of the First European Workshop on Probabilistic Graphical Models (PGM-02)*, Spain, November 2002.

[Kersting *et al.*, 2003] K. Kersting, T. Raiko, S. Kramer, and L. De Raedt. Towards Discovering Structural Signatures of Protein Folds based on Logical Hidden Markov Models. In *Proceedings of the Pacific Symposium on Biocomputing*, 2003.

[Muggleton, 1996] S. Muggleton. Stochastic logic programs. In L. De Raedt, editor, *Advances in Inductive Logic Programming*, pages 254–264. IOS Press, 1996.

[Ngo and Haddawy, 1997] L. Ngo and P. Haddawy. Answering queries from context-sensitive probabilistic knowledge bases. *Theoretical Computer Science*, 171:147–177, 1997.

[Pfeffer, 2000] A. J Pfeffer. *Probabilistic Reasoning for Complex Systems*. PhD thesis, Stanford University, 2000.

[Poole, 1993] D. Poole. Probabilistic Horn abduction and Bayesian networks. *Artificial Intelligence*, 64:81–129, 1993.

[Sato and Kameya, 2001] T. Sato and Y. Kameya. Parameter learning of logic programs for symbolic-statistical modeling. *Journal of Artificial Intelligence Research*, 15:391–454, 2001.

[Sato, 1995] T. Sato. A Statistical Learning Method for Logic Programs with Distribution Semantics. In *Proceedings of the 12th International Conference on Logic Programming (ICLP-1995)*, pages pp. 715 – 729, 1995.

[Segal *et al.*, 2001] E. Segal, B. Taskar, A. Gasch, N. Friedman, and D. Koller. Rich probabilistic models for gene expression. In *Proceedings of the 9th International Conference on Intelligent Systems For Molecular Biology (ISMB)*, 2001.