Learning directed probabilistic logical models from relational data

Daan Fierens^a

^a *Department of Computer Science, Katholieke Universiteit Leuven, Celestijnenlaan 200A, 3001 Heverlee, Belgium*
E-mail: daan.fierens@cs.kuleuven.be

Data that has a complex relational structure and in which observations are noisy or partially missing poses several challenges to traditional machine learning algorithms. One solution to this problem is the use of so-called probabilistic logical models (models that combine elements of first-order logic with probabilities) and corresponding learning algorithms. In this thesis we focus on directed probabilistic logical models. We show how to represent such models and develop several algorithms to learn such models from data.

Keywords: probabilistic logic learning, Bayesian networks, decision trees

1. Introduction

Machine learning is concerned with algorithms that construct models from data. Two kinds of models that have received special attention are probabilistic models and logical models.

- With *probabilistic models* we mean generative models, that is, models that define a probability distribution on a set of random variables. The main advantage of such models is their ability to deal with domains in which observations are noisy and/or partially missing. A well-known kind of such models are *Bayesian networks* [8].
- With *logical models* we mean models that use elements of first-order logic or related formalisms (such as logic programming) [1]. The advantage of such models is their ability to deal with domains that have a complex structure characterized by multiple types of entities and relations.

Since real-world domains often have both of the above characteristics, there is a growing interest in combining the advantages of both kinds of models through the use of so-called *probabilistic logical models* [2]. Of course, such models require specialized learning algorithms. The goal of this thesis [3] was to develop such algorithms for a particular kind of probabilistic logical models, namely directed models.

2. Directed probabilistic logical models

Probabilistic models are sometimes distinguished based on whether they are directed (e.g. Bayesian networks) or undirected (e.g. Markov networks). This distinction can also be made for probabilistic logical models. In this thesis we focus on directed probabilistic logical models. To represent such models we introduce the formalism of *Logical Bayesian Networks (LBN)*. This formalism can be seen as an extension of Bayesian networks to multi-relational domains [4].

Like Bayesian networks, the LBN formalism is centered around the notion of *conditional probability distributions (CPDs)*. In Bayesian networks, one possible format for representing CPDs is that of *probability trees* (decision trees that predict probability distributions). This is quite popular because probability trees can compactly represent certain distributions, and this compactness can be exploited to speed-up inference and to learn more accurate models [7]. In the LBN formalism we upgrade this approach by representing CPDs as first-order logical probability trees (probability trees in which the tests in internal nodes are first-order logic queries). We then represent a probabilistic logical model as a set of such CPDs, one for each predicate symbol in the vocabulary. As for Bayesian networks, this set of CPDs needs to be acyclic in order for the model to be valid.

3. Learning algorithms

The directed probabilistic logical models that we consider essentially consist of a set of first-order logical probability trees. Hence, any learning algorithm for such models requires a learning algorithm for probability trees. For this reason, we first study in detail the problem of learning probabilistic trees. Next, we consider the problem of learning the directed probabilistic logical models themselves.

3.1. Learning probability trees

Ideally, the learning algorithm for probability trees should yield accurate and relatively compact trees, and should be relatively efficient (since learning a single probabilistic logical model requires learning many probability trees). Of course, these three measures (accuracy, compactness, efficiency) counteract, so a trade-off needs to be found. Since these three measures are all heavily influenced by the pruning criterion used in the learning algorithm, we perform a detailed study of six pruning criteria [5]. We experimentally evaluate each of the pruning criteria on relational data as well as on attribute-value (“non-relational”) data. Based on the conclusions of these experiments, we retained two pruning criteria for further use in our algorithms for learning directed probabilistic logical models (namely the Bayesian Information Criterion and randomization [5]).

3.2. Learning directed probabilistic logical models

First we consider the problem of learning *non-recursive* directed probabilistic logical models [6]. We introduce two algorithms for learning such models. The first algorithm is based on searching over directed acyclic graphs, and is relatively close to some existing learning algorithms. The second algorithm is based on searching over orderings, an idea that has already been used for learning Bayesian networks, but that had not yet been applied to probabilistic logical models. Experiments on relational data show that the two algorithms are comparable in terms of the quality of the learned models, and that searching over orderings is significantly more efficient.

In a next step we show how the above two algorithms can also be used for learning *recursive*

models under the simplifying assumption that the relationship that controls the recursion is known in advance (and is acyclic). We also introduce an algorithm for learning recursive models which does not require this assumption. We do this by generalizing the above algorithm for searching over orderings [9]. Experiments on relational data show that the new algorithm can indeed learn useful recursive dependencies, but for learning non-recursive dependencies the original algorithm is superior.

Acknowledgements

Daan Fierens is supported by the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT Vlaanderen).

References

- [1] L. De Raedt, *Logical and Relational Learning*, Springer, 2008.
- [2] L. De Raedt and K. Kersting, Probabilistic logic learning, *SIGKDD Explorations* 5(1) (2003), 31–48.
- [3] D. Fierens, *Learning Directed Probabilistic Logical Models from Relational Data*, Ph.D. thesis, Department of Computer Science, Katholieke Universiteit Leuven, 2008. <http://hdl.handle.net/1979/1833>
- [4] D. Fierens, H. Blockeel, M. Bruynooghe and J. Ramon, Logical Bayesian networks and their relation to other probabilistic logical models, in: *Proc. of the 15th International Conference on Inductive Logic Programming*, LNCS, Vol. 3625, Springer, 2005, pp. 121–135.
- [5] D. Fierens, J. Ramon, H. Blockeel and M. Bruynooghe, A comparison of approaches for learning probability trees, in: *Proc. of 16th European Conference on Machine Learning*, LNCS, Vol. 3720, Springer, 2005, pp. 556–563.
- [6] D. Fierens, D., J. Ramon, M. Bruynooghe and H. Blockeel, Learning directed probabilistic logical models: Ordering-search versus structure-search, in: *Proc. of 18th European Conference on Machine Learning*, LNCS, Vol. 4701, Springer, 2007, pp. 567–574.
- [7] N. Friedman and M. Goldszmidt, Learning Bayesian networks with local structure, in: M. Jordan (Ed.), *Learning in Graphical Models*, Kluwer Academic Publishers, 1998, pp. 421–459.
- [8] R. Neapolitan, *Learning Bayesian Networks*, Prentice Hall, 2003.
- [9] J. Ramon, T. Croonenborghs, D. Fierens, H. Blockeel and M. Bruynooghe, Generalized ordering-search for learning directed probabilistic logical models, *Machine Learning* 70(2–3) (2008), 169–188.