Learning Compact Markov Logic Networks With Decision Trees

Hassan Khosravi, Oliver Schulte*, Jianfeng Hu, and Tianxiang Gao
{hkhosrav, oschulte, jhfu, tga18}@sfu.ca

School of Computing Science
Simon Fraser University
Vancouver-Burnaby, B.C., Canada

Abstract. Markov Logic Networks (MLNs) are a prominent model class that generalizes both first-order logic and undirected graphical models (Markov networks). The qualitative component of an MLN is a set of clauses and the quantitative component is a set of clause weights. Generative MLNs model the joint distribution of relationships and attributes. A state-of-the-art structure learning method is the moralization approach: learn a 1st-order Bayes net, then convert it to conjunctive MLN clauses. The moralization approach takes advantage of the high-quality inference algorithms for MLNs and their ability to handle cyclic dependencies. A weakness of the moralization approach is that it leads to an unnecessarily large number of clauses. In this paper we show that using decision trees to represent conditional probabilities in the Bayes net is an effective remedy that leads to much more compact MLN structures. The accuracy of predictions is competitive with the unpruned model and in many cases superior.

1 Introduction

As relational data are very common in practice, an important goal is to extend machine learning techniques for them. Generative models represent probabilistic patterns over both links/relationships and attributes. A Markov Logic Network (MLN) is a set of 1st-order formulas, whose quantitative component is a set of weights, one for each clause. Domingos and Richardson show how an MLN can be interpreted as a template for a Markov random field whose nodes comprise ground atoms that instantiate the 1st-order formulas [1]. MLNs have achieved impressive performance on a variety of relational learning tasks. An open-source benchmark system for MLNs is the Alchemy package [2].

Structure Learning via Moralization. The recently introduced moralization approach [3] can be seen as a hybrid method that uses directed models for learning and undirected models for inference. This method learns a directed 1st-order Bayes net model for an input relational database. The Bayes net is then converted to an MLN using the moralization method, as described by Domingos and

* Supported by the Natural Sciences and Engineering Research Council of Canada. We are indebted to reviewers of the ILP conference for helpful comments.

Richardson [1, 12.5.3]. In graphical terms, moralization connects all co-parents, then omits edge directions. Converting the Bayes net to an undirected model to perform inference avoids the cyclicity problem, which is that there may be cyclic dependencies between the properties of individual entities. The learn-and-join algorithm of Khosravi *et al* upgrades propositional Bayes net learning to relational data in a very efficient way. Compared to predecessor MLN learning algorithms on several benchmark datasets, structure learning was orders of magnitude faster and substantially more accurate. A disadvantage of the moralization approach is that it adds a clause for each conditional probability parameter in the Bayes net. While this rich structure captures most of the relevant correlations in the data, the large number of clauses has several drawbacks. (i) The resulting MLN is harder for a user to understand. (ii) Parameter learning is slower. (iii) Inference is slower. (iv) The curse of dimensionality: As the number of weight parameters increase, parameter estimates are less accurate. This paper presents an extension of the moralization approach that produces significantly smaller MLN structures without sacrificing statistical power.

Decision Trees for Representing Local Independencies. It is well-known that because Bayes net graphs represent associations between random variables, rather than between specific values of these variables, they may fail to capture *local* independencies that hold conditional on specific values of the random variables [4]. A common way to represent local independencies is to replace each conditional probability table by a decision tree that predicts the probability of a child node value given values for its parents [4]. The main advantages of decision trees for relational models are as follows. (i) Many methods have been developed for learning decision trees that produce probability estimates [5]. (ii) Each tree branch corresponds to a conjunction of literals and is straightforwardly converted to an MLN clause.

Figure 1 illustrates the MLN clauses corresponding to a decision tree. $Int(S, I)$ means that the intelligence level of student S is I. $Ranking(S, R)$ means that the ranking of student S is R. $RA(P, S, V)$ means that student S is/is not an RA for professor P depending on whether the truth value V is True or False. $Pop(P, L)$ means that the popularity level of professor P is L.

Evaluation. We compared our learning algorithms with two state-of-the-art MLN learning methods (LHL and LSM [6]) using public domain datasets (MovieLens, Mutagenesis, Mondial, Hepatitis). As Table 2 shows, decision tree pruning is very effective in reducing the number of MLN clauses, by a factor of 5-25 depending on the dataset. It also shows that parameter learning in the pruned models is much faster than without pruning. The comparison with the unpruned moralized models and with LSM learning indicates that predictive accuracy with decision trees is competitive and in many cases superior.

Conclusion. Augmenting Bayes net learning with decision tree learning leads to a compact set of clauses that represent generative statistical patterns in a

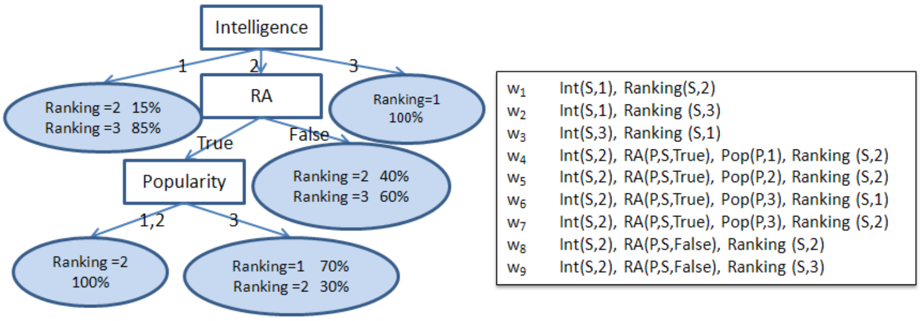

Fig. 1. A decision tree that specifies conditional probabilities for a *ranking* node in a 1st-order Bayes net and the corresponding MLN clauses generated from the decision tree.

Table 2 Left: 5-fold cross-validation estimate of the number of parameters in learned model. Right: 5-fold cross-validation estimate for average learning times in seconds. Runtimes for the moralization methods are given as (structure learning time + weight learning time).

	MBN + DT	MBN	LSM	LHL
MovieLens	39	327	10	NT
Mondial	102	2470	20	25
Mutagen	50	880	13	NT
Hepatitis	120	793	23	27

	MBN + DT	MBN	LSM	LHL
MovieLens	22 + 345	15 + 3401	34.03	NT
Mondial	9 + 18	4+ 1168	0.29	11524
Mutagen	18 + 274	12 + 4425	26.47	NT
Hepatitis	21 + 813	15 + 6219	10.94	72452

relational database. In our simulations on four benchmark relational databases, decision tree pruning significantly reduced the number of clauses, leading to faster and better estimates for the model parameters.

References

1. Domingos, P., Richardson, M.: Markov logic: A unifying framework for statistical relational learning. In: Introduction to Statistical Relational Learning. MIT Press (2007)
2. Kok, S., Summer, M., Richardson, M., Singla, P., Poon, H., Lowd, D., Wang, J., Domingos, P.: The Alchemy system for statistical relational AI. Technical report, University of Washington. (2009)
3. Khosravi, H., Schulte, O., Man, T., Xu, X., Bina, B.: Structure learning for Markov logic networks with many descriptive attributes. In: AAAI'10. (2010) 487–493
4. Boutilier, C., Friedman, N., Goldszmidt, M., Koller, D.: Context-specific independence in Bayesian networks. In: UAI'96. (1996) 115–123
5. Fierens, D., Ramon, J., Blockeel, H., Bruynooghe, M.: A comparison of pruning criteria for probability trees. Machine Learning **78**(1-2) (2010) 251–285
6. Kok, S., Domingos, P.: Learning markov logic networks using structural motifs. In: ICML'10. (2010) 551–558