A FIRST-ORDER BAYESIAN LOGIC FOR HETEROGENEOUS DATA SETS

Gábor HULLÁM
Advisors: Péter ANTAL, György STRAUSZ

I. Introduction

In many areas of intelligent data analysis, especially in the field of biomedicine, the statistical inference about semantic properties of a model using heterogeneous data sets became an important issue. The combination of first-order probabilistic logic (FOPL) and the Bayesian framework provides a wide variety of different alternatives for the specification of semantic priors, models and queries. This field however still has many open problems.

Our long-term goal is to provide a unified probabilistic framework for inferring with heterogeneous data using hierarchical models and the power of first-order logic for formulating queries about the models. In this paper we describe a methodology behind a system we are currently working on, which aims to implement such inferences. The methodology contains the following major components: (1) A hybrid Bayesian semantics over first-order sentences which enables a model-based embedding of complex posteriors. (2) A stochastic annotated graph-grammar to define priors for hierarchical Bayesian networks. (3) A probabilistically linked model-network for the fusion of heterogeneous models. Here we focus on the hybrid Bayesian semantics and on the fusion of heterogeneous models.

II. FOPL

In the fields of knowledge representation information is often expressed in logic statements since they have clear, easily interpretable semantics. They lack however the ability to represent uncertainty which can be done efficiently by another set of tools: probabilistic graphical models, more specifically Bayesian networks (BN). Since many applications require both the descriptive power of first-order logic and the ability to represent uncertainty, the interest in combining the two approaches and developing a unified representation has grown significantly in the recent years.

A. First-order logic

First-order logic allows the use of predicates and quantifiers, thus it supports statements about groups of objects. This expressive power is needed in many application areas despite its computationally demanding algorithms. Furthermore, it prevents such obvious incorporations of probabilistic information as in propositional logic.

The basic inference problem in first-order logic is to determine whether a knowledge base (KB) entails a formula F, (i.e., if F is true in all worlds where KB is true). Unfortunately, inference in first-order logic is only semidecidable. Therefore, most of the knowledge bases are constructed using a restricted subset of first-order logic which in turn affects the representational capabilities in a negative way. Due to these limitations, pure first-order logic has limited applicability to practical problems.

B. Bayesian Networks

Bayesian networks(BNs) represent the modeled domain by a directed acyclic graph (DAG), where the nodes are stochastic variables (the entities of the domain) and the edges can be regarded as direct probabilistic dependencies. Besides the DAG structure there is a conditional dependency model assigned to each variable (in the following, we will use the terms variable and node as synonyms), describing how the given node depends on its parents. In a sense, the Bayesian network itself implements a hierarchic model structure: the structure of the DAG is the higher-order discrete component, while the local dependency models, described through numeric parameters form the lower-order continuous component. In case of discrete Bayesian networks (where every variable is discrete) the most common form of the local dependency models are the *conditional probability tables* (CPTs), which describe the probability distribution of the child variable conditional on its parents.

C. Earlier Works

Two of the early works that attempted to unify logic and probability was [1], which introduced a combination of the domain-frequency approach and the possible worlds approach, and [2], which described knowledge-based model construction (KBMC). Its underlying concept is to order dedicated networks to different queries or modeling situations by constructing them on-the-fly from a background knowledge base.

The Relational Bayesian network (RBN) [3] is another possible way to represent probabilistic first-order statements. In RBNs every predicate is represented by one node in the network and each of them has a probability formula assigned to it describing its probability distribution conditional on its parents.

Another FOPL approach is the BLOG (Bayesian Logic) language [4], which defines a probability distribution over model structures of a typed first-order logical language.

Other examples of probabilistic extentions to first-order reasoning are Stochastic logic programs [5], Bayesian logic programs [6], and Markov logic networks [7].

III. FOPL by model-based embedding of posteriors

The general approaches summarized earlier define distributions from elementary probabilities over worlds (object, relations, etc.), possible worlds (literals) or knowledge bases (syntactic statements), where the result can be interpreted as an extended monolithic BN defined hierarchically and/or recursively. This section describes a more modest hybrid approach for fusing a certain logical knowledge, which mainly contains factual free-text information and complex unnormalized posteriors for hierarchically defined BNs.

To introduce a hybrid approach to FOPL first consider the probabilistic model-based semantics for propositions. In propositional logic, the joint value-assignments (atomic events) $\boldsymbol{x}$ are the canonic representations of the worlds under various interpretations (resulting in the possible worlds). The set of models of a propositional knowledge base $\mathcal{K}$ is the set of worlds $\boldsymbol{x}$ where $\mathcal{K}$ is true in a given interpretation. The knowledge base semantically entails a sentence α (denoted with $\mathcal{K} \models \alpha$) if α is true in all models of $\mathcal{K}$. That is the knowledge base defines a truth-value for the possible worlds and the truth-value of the sentence is defined by $\forall x\ \mathcal{K}(\boldsymbol{x}) \rightarrow \alpha(\boldsymbol{x})$. The definition of probability as truth-value of a sentence is similar to this, as the probabilistic knowledge base defines a distribution over the models and the truth-value of the sentence can be defined as the expectation $E_{p(\boldsymbol{x}|\mathcal{K})}[\alpha(\boldsymbol{x})]$ expressing its coverage by the models of $\mathcal{K}$ (normalized to the model of the KB).

In first-order logic this approach requires a distribution over worlds with interpretations M containing potentially varying number of objects and predicate and functional relations between them. Whereas this task is an open research problem, earlier works addressed several restricted cases, such as the works on the relational probability models [8, 9].

A frequently applicable hybrid approach defines the distribution over the models $p(M)$ of a logical knowledge base using an additional probabilistic model. The logical knowledge base $\mathcal{K}^l$ describes the certain knowledge in the domain and defines the set of models (possible worlds) $\mathcal{M}(\mathcal{K}^l) = \{M : \mathcal{K}^l(M)\ is\ true\}$. The probabilistic knowledge base $\mathcal{K}^p$ expresses the remaining uncertain knowledge over these worlds by defining a distribution over these models $p(M|M \in \mathcal{M}(\mathcal{K}^l)$. That is the uncertain knowledge only weights models but it does not narrow the set of models any further. So the probability of a sentence α is defined as the expectation of its truth in valid worlds $\boldsymbol{M}$.

$$p(\boldsymbol{M} : \alpha(\boldsymbol{M})|\mathcal{K}^l, \mathcal{K}^p) = E_{p(\boldsymbol{M}|\mathcal{K}^l, \mathcal{K}^p)}[\alpha(\boldsymbol{M})] = \sum_{M \in \mathcal{M}(\mathcal{K}^l)} \alpha(\boldsymbol{M}) p(\boldsymbol{M}|\mathcal{K}^p) \tag{1}$$

If the models vary only in a well-defined respect such as a given object, this regularity can be used to define the distribution over the models based on a distribution over this respect.

IV. Augmented BNs for multiple data sets and models

The need for the joint analysis of multiple, heterogeneous data sets calls for the usage of Bayesian statistics and probabilistic graphical models, especially Bayesian meta modeling, which offers a normative solution for the problem.

We propose two assumptions to derive a practical Bayesian meta model. First we do not model uncertainty over the structural and parametric dependency of the models, i.e. we fix a prior S^M, θ^M and we do not perform Bayesian inference/learning over them. Second we assume exclusive correspondence between the random variables of a domain and its model class. Furthermore to simplify presentation we omit the parameters assuming that they are averaged out for the discrete models.

These assumptions result in a two-layered structure using the augmented BN representation. The upper layer is a BN with nodes $\mathcal{M}_1, \ldots, \mathcal{M}_K$ representing model classes, and their joint distribution $p(\mathcal{M}_1, \ldots, \mathcal{M}_K)$. The lower layer contains the observations $D_{N_i}^{(i)}$ connected as children to the appropriate model class without interconnections, which conforms to our assumption about exclusivity and exchangeability.

An important consequence of these assumptions that inference is decomposed as follows:

$$p(\mathcal{M}_1, \ldots, \mathcal{M}_K|D_{N_1}^{(1)}, \ldots, D_{N_K}^{(K)}) \propto p(D_{N_1}^{(1)}, \ldots, D_{N_K}^{(K)}|\mathcal{M}_1, \ldots, \mathcal{M}_K) p(\mathcal{M}_1, \ldots, \mathcal{M}_K) \tag{2}$$

$$= \prod_{i=1}^{K} \prod_{l=1}^{N_i} p(D_{N_i}^{(i,l)}|\mathcal{M}_i) \prod_{i=1}^{K} p(\mathcal{M}_i|Pa(\mathcal{M}_i)). \tag{3}$$

It means that the effect of observations can be computed independently for each model class and can be incorporated as a virtual evidence into inferences at the upper layer among the model classes, e.g. to compute a marginal for a given model class $p(\mathcal{M}_i|D_{N_1}^{(1)}, \ldots, D_{N_K}^{(K)})$ using MC methods.

In integrated learning from heterogeneous sources however, rescaling of belief for the sources is advisable to express our confidence in them, which can be easily achieved by flattening/peakening the virtual evidences for the model classes.

The specification of the local probability dependency models $p(\mathcal{M}_i|Pa(\mathcal{M}_i))$ in the upper layer of the model-augmented BN can be done by using feature sets, such as parental sets $Pa(\mathcal{M}_i)$, the directed edges and undirected edges.

V. The Bayes^3 system

The Bayes^3 system, upon which we are currently working on, aims to offer a unified probabilistic framework for inferring with heterogeneous data using hierarchical models and the power of first-order logic for formulating queries about the models. It is based on the before mentioned components: the hybrid Bayesian semantics (see Section III.), the probabilistically linked model-network (see Section IV.) and posteriors for BNs $p(M_i|D_{N_i}^{(i)})$, potentially with so called hyper-posteriors corresponding to their hierarchical, modular definition H_i, $p(M_i, H_i|D_{N_i}^{(i)})$.

The extension of this hybrid method to the general case when the models of the $\mathcal{K}^l$ vary in multiple, but well-defined respects is straightforward.

A. Inference

We can identify two inference methods suited for this model-based FOPL. The first is basically an enumeration method, iterating over the models, evaluating the truth-value of the sentence weighted with the probability of the model or iterating through only the models where the target is true. The second method directly estimates the expectation of its truth using an MC method with the indicator function of the truth-value of the target sentence.

B. Examples

For example using the posterior over BN structures given a particular data set the following statement has well-defined probability given the data sets

$$\forall v_1, v_2 : Var(v_1) \wedge Var(v_2) \wedge SimilarText(v_1, v_2) \Rightarrow \neg Independent(v_1, v_2, targetDAG) \quad (4)$$

where the predicate $Independent(v_1, v_2, G)$ means that the variables v_1, v_2 are not confounded or causally not connected in the DAG G.

Analogously, the posterior of a sentence including elements of the hierarchical definition of BNs exist, such as the occurrence of the ***M'*** module in the target model M_i

$$p(M' \subseteq M_i | D^{(1)}_{N_1}, \ldots, D^{(K)}_{N_K}) = \sum_{\substack{M_j \\ i \neq j}} p(M_1, \ldots, M' \subseteq M_i, \ldots, M_K | D^{(1)}_{N_1}, \ldots, D^{(K)}_{N_K}) \quad (5)$$

VI. Conclusion

The fusion of heterogeneous information resources, particularly the integration of electronic prior knowledge such as knowledge bases and free-text with expertise and experimental data is of vital importance. We presented two components of the Bayes^3 framework that will allow a practical and still normative Bayesian fusion of literature, experimental data and expertise. Currently we are working on the implementation of these methods.

Acknowledgement

I am grateful to Peter Antal for many valuable discussions relating to this topic.

References

[1] J. Y. Halpern, "An analysis of first-order logics of probability," *Artificial Intelligence*, 46:311–350, 1990.

[2] M. Wellman, J. Breese, and R. Goldman, "From knowledge bases to decision models," *The Knowledge Engineering Review*, 7(1):35–53, 1992.

[3] M. Jaeger, "Relational bayesian networks," *Proc. of the 13th Conference on Uncertainty in Artificial Intelligence (UAI-1997)*, 1997.

[4] B. Milch, B. Marthi, and S. Russell, "Blog: Relational modeling with unknown objects," in *ICML 2004 Workshop on Statistical Relational Learning and Its Connections to Other Fields*, 2004.

[5] S. Muggleton, "Stochastic logic programs," in *Proceedings of the 5th International Workshop on Inductive Logic Programming*, L. De Raedt, Ed., pp. 254–264. Department of Computer Science, Katholieke Universiteit Leuven, 1995.

[6] K. Kersting and L. D. Raedt, "Bayesian logic programs," in *Proceedings of the Work-in-Progress Track at the 10th International Conference on Inductive Logic Programming*, J. Cussens and A. Frisch, Eds., pp. 138–155, 2000.

[7] M. Richardson and P. Domingos, "Markov logic networks," *Machine Learning*, 62:107–136, 2006.

[8] M. Jaeger, "Complex probabilistic modeling with recursive relational bayesian networks," *Annals of Mathematics and Artificial Intelligence*, pp. 297–308, 2002.

[9] A. Pfeffer and D. Koller, "Semantics and inference for recursive probability models," in *Proc. of the 17th National Conference on Artificial Intelligence (AAAI), Madison, Wisconsin*, 2000.