Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

Scaling-Up Inference in Markov Logic

Deepak Venugopal
Department of Computer Science
The University of Texas at Dallas
dxv021000@utdallas.edu

Introduction

Markov logic networks (Domingos and Lowd 2009) (MLNs) combine the power of first-order logic and probabilistic graphical models and as a result are ideally suited for solving large, complex problems in application domains (e.g., NLP, computer vision, the Web, etc.) that have both rich relational structure and large amount of uncertainty. However, inference in these rich, relational representations is quite challenging. Specifically, MLNs are defined as a set of weighted first-order logic formulas and given a set of constants that represent objects in the domain of interest, they represent a large Markov network which have a potential for each grounding of each first-order formula. For example, the simple formula $\forall x, y, z$ $\texttt{Friends}(x, y) \wedge \texttt{LivesIn}(y, z) \Rightarrow \texttt{Safe}(x, z)$ which says that if you have several friends in a neighborhood z, you feel safe in the neighborhood z, yields ten million formulas assuming that there are 1000 people in the domain and 10 neighborhoods. Although inference techniques for Markov networks (graphical models) have come a long way and are able to tackle much larger problems than ever before, several inference tasks for complex real-world MLNs are currently out of reach of even the most advanced techniques. The aim of this thesis is to advance the state-of-the-art in MLN inference (Kok et al. 2006; Niu et al. 2011), enabling it to solve much harder and more complex tasks than is possible today. To this end, I will develop techniques that exploit logical structures and symmetries that are either explicitly or implicitly encoded in the MLN representation and demonstrate their usefulness by using them to solve hard real-world problems in the field of natural language understanding.

Progress to Date

Effective Sampling in Relational Models

MLNs typically contain large number of logical dependencies. In the presence of such dependencies, popular sampling algorithms such as Gibbs sampling and importance sampling do not converge (or converge very slowly). We proposed two approaches to address this problem.

Our first approach called `GiSS` (Venugopal and Gogate 2013b) combines Gibbs sampling with an advanced importance sampling approach called SampleSearch (Gogate and Dechter 2011), to jump across regions/clusters in the sampling-space that are fractured due to deterministic dependencies. Specifically, SampleSearch leverages SAT solvers to sample the clusters while Gibbs sampling samples within the clusters.

Our second approach, which is presented in (Venugopal and Gogate 2013a), combines blocking and collapsing, two widely used methods to improve the convergence of Gibbs sampling, especially when the variables in the target distribution have logical dependencies and/or correlations. It turns out that combining blocking and collapsing optimally and tractably is a hard problem due to the complex interplay between them. We formulated this as an optimization problem and showed that a solution to the optimization problem yields an adaptive sampler that is superior in terms of accuracy and convergence to existing approaches.

Lifted Inference

Propositional inference algorithms treat MLNs as a regular graphical model, i.e., they work on the completely ground MLN, forgetting that it has relational structure. In contrast, *lifted inference* algorithms (Poole 2003; de Salvo Braz 2007; Gogate and Domingos 2011) exploit symmetries in the relational representation and are far more scalable. In our prior work, we lifted two widely used sampling algorithms, Blocked Gibbs sampling (Jensen, Kjaerulff, and Kong 1995) and importance sampling to the first-order level.

Lifted Blocked Gibbs (LBG) (Venugopal and Gogate 2012) blocks first-order atoms such that *exact lifted inference* within each block is tractable given all other blocks. In contrast to blocking over propositional variables, it turns out that increasing the block size over first-order atoms in some cases allows us to exploit more symmetries, reducing the complexity of Gibbs sampling. LBG works on a first-order *Gibbs cluster graph* where each cluster/block passes a lifted message (sufficient statistics) to its neighbors. Lifted samples are generated by performing exact inference within a cluster incorporating the messages from its neighbors. Our evaluation demonstrated that LBG is far superior to propositional approaches in terms of scalability and convergence.

In lifted importance sampling (Gogate, Jha, and Venugopal 2012), we draw lifted samples from a proposal distribution instead of sampling individual groundings. We showed that sampling from such a lifted space reduces the variance of estimates derived from the samples and developed a scalable method for constructing the proposal that applies several lifted inference rules approximately.

Improving Scalability through Approximate Symmetries

A key problem with lifted inference methods is that they tend to work well only when the MLN has specific symmetric structure, which is often not the case in real-world applications. Moreover, evidence breaks symmetries, further diminishing their performance. As a result, in practice, for arbitrary MLN/evidence structures, lifted inference is as scalable as propositional inference.

In our recent ECML paper (Venugopal and Gogate 2014b), we proposed a general, practical approach to scale-up inference in MLNs without any restrictions on its structure or evidence. We utilized standard unsupervised machine learning approaches such as KMeans to cluster objects based on approximate symmetries and generated a "compressed" MLN in which the objects are replaced by their cluster-centers. To learn these approximate symmetries, we developed a distance measure between two distinct objects in a domain based on the evidence presented to the MLNs.

In a related paper (Venugopal and Gogate 2014a), we showed how to scale-up importance sampling for arbitrary MLN/evidence structures and at the same time provided several approximation guarantees. Specifically, we used lifted Gibbs sampling to tractably sample from an informed proposal that we constructed using the "compressed" MLN. Further, we approximated the *importance weight* of each sample tractably such that we obtain asymptotically unbiased estimates.

Application: Event Extraction

An effective approach to event extraction in NLP is *joint inference*, namely, methods that reason about relational dependencies between events. In our recent paper (Venugopal et al. 2014), we developed a joint inference based event extraction system using MLNs. However, it turns out that the key linguistic features that are useful in event extraction are extremely high dimensional and therefore learning them effectively using MLNs requires an infeasible amount of data. SVMs on the other hand, lack the ability to perform joint inference but are very well-suited to handle high-dimensional features. To get the best of both worlds, we learned high dimensional features using SVMs and encoded them as low-dimensional soft-evidence in MLNs. Further, by exploiting the MLN's structural properties, we ensured the computational feasibility of joint inference. On three BioNLP datasets, our system was better or on par with the best systems and outperformed all previous MLN-based systems.

Future Work

A counting problem which we call `#SATGround` (counting the satisfied groundings of a first-order formula given assignments ground atoms) is arguably the main bottleneck in several MLN inference and learning algorithms. Specifically, variants of `#SATGround` manifest themselves in algorithms such as Gibbs sampling, MaxWalkSAT, voted perceptron, contrastive divergence and pseudo-likelihood learning. Existing MLN inference/learning systems (Kok et al. 2006; Niu et al. 2011) naively solve `#SATGround` greatly diminishing their scalability. We want to connect `#SATGround` to solution counting in graphical models and leverage decades of advances in space/time efficient counting strategies to develop a family of scalable MLN inference/learning algorithms.

References

de Salvo Braz, R. 2007. *Lifted First-Order Probabilistic Inference*. Ph.D. Dissertation, University of Illinois, Urbana-Champaign, IL.

Domingos, P., and Lowd, D. 2009. *Markov Logic: An Interface Layer for Artificial Intelligence*. Morgan & Claypool.

Gogate, V., and Dechter, R. 2011. SampleSearch: Importance sampling in presence of determinism. *Artificial Intelligence* 175(2):694–729.

Gogate, V., and Domingos, P. 2011. Probabilistic Theorem Proving. In *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence*, 256–265. AUAI Press.

Gogate, V.; Jha, A.; and Venugopal, D. 2012. Advances in Lifted Importance Sampling. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*. AAAI Press.

Jensen, C. S.; Kjaerulff, U.; and Kong, A. 1995. Blocking Gibbs Sampling in Very Large Probabilistic Expert Systems. *Internat. J. Human-Computer Studies* 42:647–666.

Kok, S.; Sumner, M.; Richardson, M.; Singla, P.; Poon, H.; and Domingos, P. 2006. The Alchemy System for Statistical Relational AI. Technical report, University of Washington.

Niu, F.; Ré, C.; Doan, A.; and Shavlik, J. W. 2011. Tuffy: Scaling up Statistical Inference in Markov Logic Networks using an RDBMS. *PVLDB* 4(6):373–384.

Poole, D. 2003. First-Order Probabilistic Inference. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, 985–991.

Venugopal, D., and Gogate, V. 2012. On Lifting the Gibbs Sampling Algorithm. In *Proceedings of the 26th Annual Conference on Neural Information Processing Systems*, 1664–1672.

Venugopal, D., and Gogate, V. 2013a. Dynamic Blocking and Collapsing for Gibbs Sampling. In *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence*.

Venugopal, D., and Gogate, V. 2013b. GiSS: Combining Gibbs Sampling and SampleSearch for Inference in Mixed Probabilistic and Deterministic Graphical Models. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*.

Venugopal, D., and Gogate, V. 2014a. Scaling-up Importance Sampling for Markov Logic Networks. In *Proceedings of the 28th Annual Conference on Neural Information Processing Systems (NIPS)*.

Venugopal, D., and Gogate, V. 2014b. Evidence-Based Clustering for Scalable Inference in Markov Logic. In *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD*.

Venugopal, D.; Chen, C.; Gogate, V.; and Ng, V. 2014. Relieving the Computational Bottleneck: Joint Inference for Event Extraction with High-Dimensional Features. In *Empirical Methods on Natural Language Processing*.