Proportional dependencies and asymptotics of probabilistic representations

Felix Weitkämper[*]

Institut für Informatik
Ludwig-Maximilians-Universität München
Oettingenstr. 67
80538 München
felix.weitkaemper@lmu.de

Abstract. In this contribution we give a general overview of our work in progress on incorporating proportional dependencies in statistical relational frameworks. We describe the problem of domain-size dependent probabilities that proportional dependencies set out to tackle and explain how they help to solve that problem and how they are implemented in relational logistic regression models, Markov logic networks and general relational Bayesian networks. We then pose the open problem of how to transfer those concepts and results to probabilistic logic programming settings such as Problog.

1 Introduction

In the last 20 years, Statistical Relational Artificial Intelligence (StarAI) has developed into a promising approach for combining the reasoning skills of classic symbolic AI with the adaptivity of modern statistical AI. It is not immediately clear, however, how StarAI behaves when transitioning between domains of different sizes. This is a particularly pertinent issue for StarAI since the template-like design of statistical relational formalisms, that are presented independently of grounding to a concrete domain, is one of their main attractions. Furthermore, scalability has proven a key barrier to their widespread deployment in applications.

The topic of scaling across domain sizes has therefore received some attention from the literature, and general patterns of behaviour have become clear. On the one hand, Jaeger and Schulte [3] have provided very limiting necessary conditions under which domain size does not affect inference in different StarAI approaches. On the other hand, Poole et al. [6] have characterised the scaling behaviour of both Markov Logic Networks (MLN) and Relational Logistic Regression (RLR) on a small class of formulas on which the inferences turn out to be asymptotically independent of the learned or supplied parameters. This characterisation was extended and partly corrected by Mittal et al. [5] with considerable analytic and

---

numerical effort. They also present a proposal to mitigate the domain-size dependence in MLN by scaling the weights associated with formulas according to the size of the domain, calling the resulting formalism Domain-size Aware Markov Logic Networks (DA-MLN). With similar computational effort, they prove that asymptotic probabilities in DA-MLN are dependent on the supplied parameters for some example cases. However, a general and systematic investigation of this dependence is still lacking

In this short note, we give a general overview of the work we have done so far to alleviate this problem, and outline the difficulties experienced when trying to transfer our results to the setting of probabilistic logic programming as exemplified by the ProbLog language..

2 Proportional dependencies

2.1 General concept

A common example in statistical relational AI is that of smoking and friendship groups, where one is more likely to be a smoker if one has friends that also smoke. In the simplest possible representation, this could be modelled as a domain of one's friends with a single unary predicate "smokes" and then one proposition "I_smoke". The probability of "I_smoke" would then depend on the number of friends which smoke (i. e. objects in the domain for which "smokes" is true). It is easy to see that if there is a fixed probability of 'smokes(x)', then with more friends the number of smoking friends will increase proportionally. Thus, if every smoking friend has a certan likelihood of making me smoke, I will eventually smoke with arbitrarily high probability if I have many friends.

This asymptotic behavior (that the probability of "I_smoke" converges to 1 as the domain size increeases) is completely independent of the likelihood that one given friend convinces me of smoking, or the percentage chance that any given friend smokes. It might rather be appropriate, therefore, to consider the *proportion* of my friends that smoke as the determining factor, rather than the raw number. Then, the asymptotic behaviour does indeed depend both on the likelihood of any given friend smoking and on the effect that smoking in my friends has on me as such.

This proportional dependence is asymptotically well-behaved, at least for directed models and a fragment of undirected Markov Logic Networks. An interesting question is however, to what extent proportional dependencies are adequate to practical cases of changing domain sizes. While this will generally depend on the domain and on the interpretation of the data, proportional representation is particularly suitable for randomly sampled subdomains, since random sampling conserves mean proportions.

The way we model such proportional dependencies depends on the knowledge representation formalism used. In [10] we have shown that for weight-parametrised directed models (such as the RLR formalism developed by Kazemi et al. [4]) scaling the weight parameters with the domain size leads to such a proportional influence and that indeed the asymptotic properties are both readily

computable and depend on the factors parametrising the model (such as weights and probabilities of the parent formulas). We will briefly outline the scope and limitations of our approaches to proportional dependencies in RLR, MLN and Relational Bayesian Networks (RBN). For the sake of brevity, we refer to the respective literature (e. g. [4,8,2] respectively) for a description of the formalisms themselves.

2.2   RLR

In an RLR model, the probability of an atomic formula at a child node depends on the truth values of the atoms of the parent nodes, grounded for all those variables that do not (also) occur in the child. Thus the maximum number of connections in the ground network is given by $\prod_x |D_x|$ where the product is taken over all variables occurring in a parent node that do not occur in the child node. To ensure proportion, we then have to scale the sum of the weights of a formula by that number.

In [10] we show that when we scale the weight parameters, the asymptotic probabilities of all literals vary with the weights in an intuitive way. Furthermore, the asymptotic distribution can in fact be computed directly over a *proportional relational belief network*, a Bayesian network that incorporates deterministic dependencies wherever the law of large numbers provides certain proportions almost surely.

2.3   MLN

In Markov logic networks, where weight scaling with the domain size had first been suggested by Mittal et al. [5], there is the added complexity that the number of connections induced by a formula differs between the atoms in a formula. Therefore, each formula has a connection *vector* associated with it, rather than just a single number. When scaling, then, one has to aggregate the entries into a single scaling factor. Mittal et al. [5] chose the maximum as a scaling factor, and they show examples in which the asymptotic limit probability of an atom in a formula depends on the weight of the formula. However, in all those examples, this atom coincides with the atom of maximal connection number. In [10], we show using examples that where the number of connections of a literal in a formula has a lower order of magnitude than the maximum, the asymptotic probability will tend to 0.5 regardless of the weight of the formula. Thus, weight scaling in MLN is most appropriate when the number of connections is the same for all literals in a formula, and we believe that in those cases the asymptotic probabilities will always depend in a transparent way on the weights provided. This is however still work in progress.

2.4   RBN

In the RBN formalism which differs from RLR in that it allows for different combination rules rather than the default sigmoidal aggregator used in RLR,

proportional reasoning is implemented by the 'mean' combination function. This behaves similarly to the weight-scaling in RLR; however, since it applies the mean directly on the level of the probabilities rather than on the level of weights, it scales linearly rather than logistically with proportions.

2.5 Projectivity

In [3], Jaeger and Schulte introduce the notion of projectivity as the strongest form of well-behavedness under increasing domain size: The induced probability distribution on a set is independent of which superset it is embedded in. The authors also show that a certain fragment of unscaled statistical regression models are projective; they are those which only involve local influences. The question naturaly arises whether the asymptotically well-behaved formalisms we introduce here are in fact projective. However, this is generally not the case: Consider a proportional RLR, say, with structure $Q(y) \longrightarrow P(x)$, a probability of $Q(y)$ of 0.5 and a probability of $P(x)$ given by the atomic formula $Q(y)$ with weight 2. Now consider the conditional probability $\Pr(P(a)|Q(a))$. On a domain of size 1, $Q(a)$ implies that all y satisfy $Q(y)$, and the probability evaluates to around 0.88. On a very large domain, however, $Q(a)$ does not significantly impact the proportion of y for which $Q(y)$ holds, and therefore the probability will limit to the asymptotic unconditional probability of $P(a)$, which turns out to be 0.73.

3 What about probabilistic logic programming?

While Problog is akin to Bayesian networks in its inherent directedness (see [7]), the semantics is very different to the relational Bayesian networks above. While the main attraction of the RBN are the combination rules, which determine the probabilities at interior nodes from their parents, current ProbLog semantics only attach probabilities to the ground facts, all of which are evaluated independently, and then model all dependencies deterministically.

This is in fact a general trait of logic programming languages that has its root in Sato's distribution semantics [9], which makes it difficult to adapt the methods outlined above in an obvious way; the scaling of the weights is indeed performed everywhere but at the leaves, which gels badly with a syntax and semantics that confines the probabilities to just those leaf nodes.

One approach one might consider would be to use the translation to weighted Boolean formulas provided by Fierens et al. [1]. In fact, the current implementation of ProbLog (ProbLog2), outlined there, converts the weighted Boolean formula to a Markov logic network and then uses a standard MLN algorithm to solve the approximate inference task. Thus one might think to apply weight scaling at this level of the architecture. The issue here is, however, that all this is done *after* grounding the logic program, and in fact, grounding is applied at a rather high level: indeed, even a statement such as '0.5 :: q(X) :- person(X)' is only considered as syntactic sugar for the individual ground probabilistic facts.

Thus, combining ProbLog with proportional reasoning seems to require a greater change to the underlying architecture than in any of the frameworks mentioned above.

References

1. Fierens, D., van den Broeck, , G., Renkens, D., Shterionov, D., Gutmann, B., Thon, I., Janssens, G. and de Raedt, L. (2015) Inference and learning in probabilistic logic programs using weighted Boolean formulas. *Theory and Practice of Logic Programming* 15:3, 358 - 401

2. Jaeger, M. (2002) Relational Bayesian Networks: a Survey. *Electronic Transactions in Artificial Intelligence* 6.

3. Jaeger, M. and Schulte, O. (2018) Inference, learning and population size: Projectivity for SRL models. *CoRR* abs/1807.00564.

4. Kazemi, S. M., Buchman, D., Kersting, K., Natarajan, S. and Poole, D. (2014a) Relational Logistic Regression. In Baral, C., Giacomo G. D., Eiter, T. (Eds.) *Principles of knowledge representation and reasoning: Proceedings of the Fourteenth International Conference, Vienna, Austria, July 20-24, 2014.* Palo Alto, CA: AAAI Press.

5. Mittal, H., Bhardwaj, A., Gogate, V. and Singla, P. (2019) Domain-size aware Markov logic networks. In: Chaudhuri, K., Sugiyama, M. (Eds.) *The 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019, Naha, Japan, 16-18 April 2019.* Proceedings of machine learning research, vol. 89:3216-3224.

6. Poole, D., Buchman, D., Kazemi, S. M., Kersting and K., Natarajan, S. (2014) Population size extrapolation in relational probabilistic modelling. In: Straccia, U., Cali, A. (Eds.) *Scalable Uncertainty Management - 8th International Conference, Oxford, UK, September 15-17, 2014.* Lecture Notes in Computer Science vol. 8720:292-305. Berlin: Springer.

7. de Raedt, L., Kersting, K., Natarajan, S. and Poole, D. (2016) *Statistical relational artificial intelligence: logic, probability and computation.* Synthesis lectures on artificial intelligence and machine learning. Morgan & Claypool.

8. Richardson, M. and Domingos, P. (2006). Markov Logic Networks. *Machine Learning* 62 (1–2): 107–136.

9. Sato, T. (1995) A statistical learning method for logic programs with the distribution semantics. In Sterling, L. S. (Ed.) *Logic Programming: The 12th International Conference.* Cambridge, MA: MIT Press.

10. Weitkämper, F. (2020) Scaling the weight parameters of Markov logic networks and relational logistic regression models. epub.ub.uni-muenchen.de/71690.