23rd. INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE

August 3-9, 2013, Beijing, China

Beijing International Convention Center (BICC)

print

User: Oliver Schulte (GzmDXfjg)

- Start page
- Edit user data
- Logout

PCMember

- Reviewer Management
- View assigned papers
- Ballot and discussion

Author

- View own papers

Review Feedback and Response

- The purpose of author feedback is to point out technical errors or significant misunderstandings. it is not to dispute the opinions of the reviewers, or to explain how criticisms can be addressed!
- **Feedback can be entered only by contact authors.**
- Do not expect replies to your feedback in the final reviews. Rest assured that feedback identifying true flaws or misunderstandings will be taken into account in final decisions.
- Please do not provide feedback explaining how you can address the referees' concerns or criticisms. It is our assumption already that concerns will be addressed in the final version.
- Feedback should be brief and specific.

Paper ID	1290
Paper authors	Oliver Schulte
Paper title	A Log-Linear Model for Bayes Nets Applied to Relational Data
Paper subtitle	
Track	Main Track (MT)
Paper Type	Technical Paper
Keywords	MT>Machine Learning::Relational Learning ** MT>Machine Learning::Learning Graphical Models
Abstract	We describe a new log-linear multi-relational model for directed graphs (Bayes nets) that provides fast parameter learning and accurate predictions. Log-linear models are widely used for multi-relational data. They are usually associated with undirected graphical models (Markov nets) \cite{Taskar2002,Domingos2009}. The key new feature of our model is that it uses the frequencies of multi-relational features as predictor variables; previous models use feature counts. Frequencies scale counts by dividing by the size of the relevant relational neighborhood. Our experiments show that frequencies provide substantially more accurate predictions than counts. We provide a novel sampling semantics for the frequency model, based on a random instantiation of the target node's Markov blanket. We compared Bayes net learning with our model on five benchmark databases with state-of-the-art Markov net methods (Alchemy weight learning, MLN-Boost). Bayes net learning is fast (parameter learning took seconds vs. hours). The predictive accuracy of the Bayes net log-linear models was competitive, in most cases superior.
Download	

Comments to author(s)

SUMMARY

The main contribution is a new specification of the distribution for relational Bayes nets. Frequencies instead of counts are used in the distribution to avoid the issues of scale mismatch. Semantics of using the frequencies is presented which have an intuitive meaning. Finally, experiments are described clearly showing benefits of this new approach.

ORIGINALITY:

This work suggests provides novel semantics for the use of frequencies rather than feature counts in the log linear distribution of relational Bayes nets.

SIGNIFICANCE:

Significant for relational models.

TECHNICAL QUALITY:

The paper is well written and clear. The idea seems straightforward but the detailed experiment results show a positive impact on performance measured by the accuracy in a 5-fold cross validation setup.

The theoretical reasoning and semantics presented also has an intuitive meaning when counts are replaced by frequencies.

Minor Comments

1) I had some trouble understanding the example in figure 3 for the frequency model, how are the frequencies (i.e. p_{ijk}) values when computing the local regression probability i.e. the example database used to compute it is not specified. Is it the data in example 1.

2) A small sketch of the theorem proof might be useful.

Summary of review

Well written paper with a simple easy to understand idea that gives good results in practice.

You may optionally submit feedback that will be visible to the reviewers and PC members. The length of your feedback is limited to 150 words.

Please enter your feedback here:

Back

Submit

Comments to author(s)
SUMMARY:
The paper describes a log-linear inference model for relational Bayesian Networks. The model uses the logarithm of the Bayesian Network's CPT as weights for inference in a form of regression. Instead of using counts induced by the grounding, the paper suggest using frequencies, i.e. counts divided by the number of possible groundings, in the regression which has an interpretation in terms of a randomized process. The resulting model is evaluated on five different benchmarks and compared to count based regression in Bayesian Networks and MLNs. Additionally, the frequency model is compared to MLNs which were learned with a state-of-the-art MLN learner, i.e. MLN-Boost.

RELEVANCE:
The paper is relevant to the AI community as directed and undirected graphical models play key roles in modeling all kinds of problems. Hence, any improvement in quality or running time helps tackling many different problems and extends the understanding of probabilistic models.

ORIGINALITY:
The main contribution of the paper suggests to replace counts by frequencies, i.e. counts divided by the number of possible groundings. Besides that, the paper also gives an interpretation for this adaption in terms of a so called "random regression". The novelty of the paper is somewhat limited, as Schulte[2011] already discuss the usage of frequencies instead of counts in pseudo log-likelihood calculation. Nevertheless, Schulte's work focuses on learning and the authors show that frequencies are a well working enhancement in inference as well.

SIGNIFICANCE:
The work is important as it introduces a new model based on well known and widely used techniques. It is an interesting idea to replace counts by frequencies to tackle the scale balance problem. It shows improvements compared to the known models and the introduction of frequencies provides an interesting point of view. Nevertheless, the insights gained by the replacement of the counts are limited and a similar discussion has been given already. As counts are very common in relational models, it would have been interesting to discuss, whether replacing frequencies by counts can be used as a general paradigm. Moreover, the experimental results unfortunately
do not present significance test. Indeed, the many of the results look significant but then others do not. The authors should really provide a significance analysis.

TECHNICAL QUALITY:
Generally, the methodology of the evaluation is sound. The running time for learning and the quality of the frequency model are compared to the count models in BNs and MLNs on five different datasets. It is shown that the time required for parameter learning is much faster for the frequency model on all problem instances as there is a closed form solution. The authors use conditional log-likelihood as well as accuracy to measure the quality of the competing models. For both measures, the frequency model shows an improvement in almost all cases. Additionally, the frequency model is also compared to state-of-the-art MLN learning by means of MLN-Boost. Here, the frequency model still achieves comparable results (although as explained above a significance analysis is missing).

The grounding is not explained carefully, hence, the examples are not easy to follow. It is not clear why the first example in Sec. 3 has 2x2 possible groundings while the second example has 3 possibles groundings. Is the number of possible groundings not equal to the product of the domain sizes? Should not the first example then have 3x3 possible groundings? Also, the number of true groundings for the first example is said to be 2. Following Fig. 1, the instantiation g(Sam), g(Anna), f(Sam, Anna) seems to be the only true grounding.

The proof of theorem 5.1 is omitted due to space constraints. At least a few sentences on the intuition of the proof would have been desirable and possibly helpful. A few more sentences on the count regression equation and its derivation in terms of grounded functors would also be good to clarify the calculation in Fig. 3. It is also not clear, how the exact inference is done precisely.

READABILITY AND ORGANIZATION:
Generally, the paper is well written and cleanly structured. Nevertheless, different aspects could have been explained more detailed while other parts could have been written shorter to gain more space for other aspects. Changing size of the section titles should also be avoided, as the space could have been gained by different means.

Fig. 1 states that g(Sam)=M. Should not then in Fig. 3 g(Sam)=M instead of W? If so, the local conditional distribution is not correct either. This is only a small issue but makes the derivation of the count regression model more difficult. Generally, the examples look like as they need a careful revision.

Summary of review
The strong points of the paper are in the contribution of an interesting modification of the count based formula. The evaluation of the model is sound and compares running times as well as the accuracy on several datasets. The new model is proven to be fast while achieving good results. The contribution though is somewhat limited in its originality and significance. Additionally, the running examples seem rather confusing as the example is not consistent. Combining this with partially short explanations, some parts are difficult to follow.

Comments to author(s)
SUMMARY:

The paper describes a new variety of a discriminative directed probabilistic relational model. Main novel feature is that conditional probabilities for the target variable are defined in terms of a random instantiation of the free variables, rather than in terms of a product (or sum) taken over all instantiations.

ORIGINALITY/SIGNIFICANCE:

The paper proposes yet another formalism in an already densely populated area. Still, the idea of replacing aggregation/combination in relational models by a random grounding semantics is novel, and could be useful in some cases.

TECHNICAL QUALITY:

I found the technical presentation in the theoretical sections 3-5 rather sloppy. Clearly, the authors try to minimize the complexity of notation. However, since the definition of the frequencies p_ijk, and their use in Eqn. (4.3) is the key technical aspect of the paper, it would be good to provide precise definitions and notation. As it is, it becomes quite difficult to correctly read (4.3), where due to the streamlined notation there is no apparent connection between the y,x values on the left, and the right side of the equation. Also, it is not clear whether p^Y_ijk is any different from p_ijk. I presume not. But then the critical dependence of the right side of (4.3) on Y is through the range of the summation index i which is determined by Y. This is not at all reflected in the chosen notation. It seems to me that once full and accurate definitions and notation are provided, Theorem 5.1 becomes self-evident.

The problems are compounded by the fact that the examples are not carefully worked out. In the example at the end of section 3, I think that the first number of true groundings n_ijk should be 1, not 2 (only X=Sam, Y=Anna), and that the number of possible groundings is 3x3, not 2x2. The example computations shown in Figure 3 also are quite unclear. How does the taking of the square root on the right side of the figure relate to equation (4.3)? Apparently in this Figure the authors make use of the geometric mean as announced in Section 2, but this does not seem to correspond to anything presented in sections 3 and 4. Another discrepancy between what is written in the text, and what is displayed in the equations is towards the end of section 4: "we only consider instances that are linked to the target node by some relationship type (i.e., we do not consider negated links)." How is that implemented by equation (4.3)?

In the introduction the authors say that they will "describe theoretical considerations and experimental results to support this claim [that frequencies lead to more accurate predictions than counts]." I don't really see in the paper very clear theoretical considerations. There is the very broad claim that the use of frequencies solves the "scale imbalance problem". However, as also observed in section 6.4, it is not even clear that the scale imbalance problem exists, since in parameter learning weights will already be adjusted according to the multiplicity of groundings for different features. Furthermore, it is not very clear how different the proposed approach is from using the 'mean' combining rule -- especially when looking at equation (4.3), where there is no evident use of geometric means.

The experimental results do appear to support the case for the frequency-based regression. However, it is very difficult to interpret and draw conclusions from the reported performance results. Few details are given about the actual prediction task that were performed, and no comparison is made with published results for these tasks for other state-of-the art methods. The comparison with the 'MBN' method in Table 3 looks less relevant, since the way the MLN here is constructed may very well be sub-optimal for MLNs. The comparison in Section 7 seems more pertinent.

It should also be pointed out that the big time advantage in the parameter learning comes from the fact that the BN parameters are fitted directly by taking normalized counts. This only works if data is complete (perhaps "made" complete via a closed-world assumption). If data is incomplete (the more realistic scenario), then some iterative optimization technique would also have to be used, and the time advantage over MLN weight learning may be lost.

READABILITY AND ORGANIZATION:

On the surface, the paper is easy to read. The English is good, and the overall structure and presentation is fine. As detailed above, the presentation is not so good when it comes to presenting the technical details.

Summary of review
A potentially useful paper with too many flaws in technical presentation.

You may optionally submit feedback that will be visible to the reviewers and PC members. The length of your feedback is limited to 150 words.
Please enter your feedback here:

Back Submit