SDM 2013

13th SIAM International Conference on Data Mining

May 02-04, 2013, Austin, TX, USA

Reviews For Paper

Paper ID 101

Title Relational Random Regression for Bayes Nets

Masked Reviewer ID: Assigned_Reviewer_1

Review:

Question

Overall Rating	Reject
Detailed Comments	In this paper, the authors consider an inference model for Bayesian networks with cycles. Although I am not familiar with this kind of topics, the contributions of this paper seem somehow minor. Also, it seems not so relevant to the SDM audiences although I know the design procedure for bayesian networks is an important topic for analyzing data based on prior knowledge. The main contribution of the paper is to develop a model based on, what they call, random regression, to avoid the difficulty when involving cycles. Their method seems to work well empirically. However, in this point, first, it seems a small contribution to the existing researches because it looks that the authors just replace the regression model (although they give some amount of considerations). It is unclear why this replacement reasonable and works well for the situations with cycles from the contexts.

	Moreover, there exist several works on methods that work even under the situations where cycles exist and thus I cannot see why their scheme is useful among such existing frameworks.

Masked Reviewer ID: Assigned_Reviewer_2

Review:

Question

Overall Rating	Neutral
Detailed Comments	The paper presents a novel relational log-linear inference model for Bayes nets. Specifically, it shows a novel relational inference method for relational models with cyclic dependencies. The key idea is to define the random regression log-probability of a target node value as the expected log-probability for a random instantiation of the node's Markov blanket. An experimental evaluation shows that the predictive performance of the new approach is comparable with other approaches. Relational models and in turn relational inference is currently receiving a lot of attention. For most directed relational models that essentially induce a large Bayesian network for a given finite set of constants, however, cyclic dependencies are a

major
obstacle since they do not result in a Bayesian network.
In this context, the paper proposed a novel semantic
and inference approach that essentially treats the
log-probability of a target node value as the expected
log-probability for a random instantiation of the node's
Markov blanket.

This is a simple but also interesting idea. The paper
is in principle well written but fails short on related work.
The insights and experimental evalution presented just shows
it works but unforunately do not go beyond a simple evaluation.

For instance, to which extend does the paper really goes beyond

Stéphane Ross, Daniel Munoz, Martial Hebert, J. Andrew Bagnell:
Learning message-passing inference machines for structured prediction.
CVPR 2011: 2737-2744

Daniel Lowd: Closed-Form Learning of Markov Networks from Dependency Networks.
UAI 2012: 533-542

Tushar Khot, Sriraam Natarajan, Kristian Kersting, Jude W. Shavlik:
Learning Markov Logic Networks via Functional Gradient Boosting. ICDM 2011: 320-329

Jennifer Neville, David Jensen: Relational Dependency Networks.
Journal of Machine Learning Research 8: 653-692 (2007)

Niels Landwehr, Kristian Kersting, Luc De Raedt:
nFOIL: Integrating Naïve Bayes and FOIL. AAAI 2005: 795-800

that also all (to some extend) explore the link between undircted models
respecticaly log-linear models
and cyclic directed models. Indeed, many (but not all) have been
cited in the paper but the differences are not stated. They are also
not investigated in the experimental section. Moreover, it is unclear
whether the MLNs have been learned in a discriminative fashion or not.
Also, it is not clear whether the MLN and BNs used in the experimental
evaluation follow (as much as possible) the same structure. Clarifying is
imporant since the differences in performance are often rather small
and might very well be due to the different structure and learning
approaches. Also, it is not clear whether the

difference are always
significant. Please provide some significance test.

Related to this, what exactly is the expectation in "the expected value of this log-sum is ... " on Page 4? Just above you say that you randomly select constants. Is it this expectation? Since you break the variable bindings, you may need a lot of samples to reduce the variance. In other words, I am wondering about the indpendence assumption you make by using the simple log-linear model here.
As far as I see you treat literals independently. So why the clauses in first place? Could this also explain the sometimes rather low improvements in performance?

To summarize, I am missing a little bit the effort to look behind the scenes, to investigate the reasons why the proposed model works or not. This is also reflected in the introducing and discussion of the differences approaches. But then the experimental results actually show that frequencies are doing better.
This is surprising since first the authors explain that differences are cool but than they do not pay off.

However, that "the closed form for random regression that avoids constructing a ground network" is really appealing and quite

interesting but not experimentally
explored. Is there really an efficiency gain?
Indeed, Figure 6 indicates this
but here you use the standard MLN inference
approach. However, there are actually
quite a lot of advanced inference techniques for
MLNs. It particular, many of them
try to avoid the time consuming grounding,
which may explain some of the high running
times in your case, see e.g. Chris Re VLDB paper.

The intro highlights that "an important
observation for this paper is that
the choice of parameter learning method
interacts with the choice
of features vs. counts as predictors". This
obervation is not well
investigated in the experimental section. In the
end, it is clear
since a regression learner is involved. I was
expecting some deeper insight
that explains why a particular feature should be
used. Moreover, it is also
obvious that one should avoid an exponential
scaling of influence if
employing a regression approach. This is exactly
why people have considered
combination and aggregation functions. A deeper
insight, e.g. would be
why other representation cannot deal well with
scaling.

Nevertheless, the insight that log-linear models (
or actually logistic regression)

	could actually help speeding up inference in relational models is highly interesting. Although there are several papers that point into this direction (see also above), the present paper is the first going into that direction. This is highly interesting.

Masked Reviewer ID: Assigned_Reviewer_3
Review:

Question

Overall Rating	Neutral
Detailed Comments	The paper examines Bayes net inference on relational database based on relational Markov network. The main contribution is proposing a different form of inference parameters as an extension to original moralized Bayes net methodology. Strong points: 1. The paper is well-written. 2. The proposed model seems reasonable and can be utilized as a novel framework for related topics. 3. The model can achieve comparable accuracy with great improvement on computational efficiency. Weak points:

1. The novelty of the proposed work is limited: it only utilizes existing techniques without novel contributions in theory or modeling.

2. The experiments are weak: only one baseline is compared and the improvement is marginal for some datasets.

3. It would be better if the paper can be organized more clearly: reduce the heavy review of background concepts and try to focus on authors' own contribution.

4. Simply doing the scaling for log probability instead of counts may lose the sparsity of the original data. It may help to consider this drawback and provide some pro-and-con comparison analysis for count/frequency models.

In summary, the paper has some interesting points, but could be improved in terms of novelty and experiments.