ECML-PKDD 2012

European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases

September 24th - 28th, 2012, Bristol, United Kingdom

Reviews For Paper

Track	1. Research track
Paper ID	500
Title	Log-Linear Inference Models for Bayes Nets Applied to Relational Data

---

Masked Reviewer ID: Assigned_Reviewer_3

Review:

Question	
Give your own brief summary of the paper:	This paper presents a new framework for statistical relational learning. This framework is a log-linear Bayes network based on conditional probability. Advantages of this model are its interpretability and efficiency. Moreover it demonstrates good accuracy on several benchmarks.
State three strong points of the paper:	- Very good presentation - Sound and original framework - Convincing performances
State three weak points of the paper:	- A few more explanations would be appreciated to understand some points and reproduce this work. - Needs a strong background in probabilistic graphical models - Focuses on learning the parameters and predicting attributes. No space to deal with predicting links and learning the structure (but the authors refer to their learn-and-join algorithm with the reference and URL of the implementation).
Indicate to what extent the paper has	Very much

the potential to inspire the research community by introducing new and relevant problems, concepts, solution strategies, and ideas:	
Indicate to what extent the paper contributes to solving a problem widely recognized as both challenging and important:	Very much
Indicate to what extent the paper addresses a novel area of impact of machine learning and data mining:	Somewhat
Give detailed comments justifying your	This is a very good paper. The presentation is better than usual in this domain (statistical relational learning), given that it needs a strong background in probabilistic graphical models.

evaluation of the paper, covering scientific rigour and correctness, challenges overcome, quality and reproducibility of the experiments (if any), and presentation:	The approach is sound, effective and efficient. I suggest the following improvements in the presentation: - define co-parents, page 5 - define adjacencies and instantiating adjacencies, page 6 - explain and illustrate "a database instance specifies a unique value for each ground node. - explain n_{ijk}(V=v). Isn't it redundant? What is the difference between the instantiation ijk and the instantiation V=v? - Figure 2 is not references, thus not explained in the text. What do blue/black arrow mean in the bayes net? - Last two paragraph of section 4 are for readers with a stronger backgrond/experience, or would require more explanations. Typo: bidrected (page 7)

Masked Reviewer ID: Assigned_Reviewer_4

Review:

Question

Give your own brief summary of the paper:	The paper presents a novel approach for considering log-linear inference models for the application of Bayes networks to relational data. Differently from previous approaches, the authors propose to use frequency instead of counts in the definition of predictive features.
State three strong points of the paper:	1) Theoretical considerations 2) Interesting topic
State three weak points	1) no empirical evidence that the frequency model outperforms the count model

of the paper:	2) weak experimental evaluation 3) originality
Indicate to what extent the paper has the potential to inspire the research community by introducing new and relevant problems, concepts, solution strategies, and ideas:	Somewhat
Indicate to what extent the paper contributes to solving a problem widely recognized as both challenging and important:	Somewhat
Indicate to what extent the paper addresses a novel area of impact of machine	Not at all

learning and data mining:	
Give detailed comments justifying your evaluation of the paper, covering scientific rigour and correctness, challenges overcome, quality and reproducibility of the experiments (if any), and presentation:	The paper is well motivated and the authors report some interesting theoretical considerations. However, the impression is that the paper strongly overlaps what is already proposed in the literature and reports only limited extensions. The authors assimilate recursive dependencies with autocorrelation. In general, this is not necessarily true since a recursive dependency is only a necessary condition for autocorrelation (but not a sufficient condition). Results are not convincing and do not show a clear advantage of the proposed approach over its direct counterpart. This contrasts with the conclusions the authors draw: "the empirical findings make a strong case for recommending the frequency model over the count model for CP". Indeed, for CP the frequency model outperforms the count model in three datasets out of five and in such cases, the improvement is very limited (except for MovieLens). The situation is ever worst in the case of mbn, where the frequency model looses in three cases and gives the same results in two cases. Finally, comparison is performed on a limited set of datasets and the paper does not report a comparison with state of the art MRDM approaches. For example, for Mutagenesis

	other systems reach 89 - 90 % of accuracy. Is the setting the same? Do the authors use the 188 regression friendly molecules? Which background knowledge did they use?

Masked Reviewer ID: Assigned_Reviewer_5

Review:

Question

Give your own brief summary of the paper:	The paper addresses the relational regression problem. In the paper a new log-linear regression model is proposed that applies Bayes nets to relational data. Moreover, the paper suggests to re-scale the features to the instances frequency in order to reduce the possible dominance of those feature associated with more instances. The authors studied the model using 5 benchmark data sets and the experimental results confirmed their idea and conclusions.
State three strong points of the paper:	- The paper is well written and clearly organized - The experiments support the idea/conclusions of the paper and confirm the promised advantages of the model proposed.
State three weak points of the paper:	- lack of novelty - small impact
Indicate to what extent the paper has the potential to inspire the research	Somewhat

community by introducing new and relevant problems, concepts, solution strategies, and ideas:	
Indicate to what extent the paper contributes to solving a problem widely recognized as both challenging and important:	Somewhat
Indicate to what extent the paper addresses a novel area of impact of machine learning and data mining:	Not at all
Give detailed comments justifying your evaluation of the paper, covering	The quality of the exposition is really impressive. However, as said in 3, the main problem of the paper may be its limited novelty. It is not seen that the work of the paper would have some broad impact on machine learning.

scientific rigour and correctness, challenges overcome, quality and reproducibility of the experiments (if any), and presentation:	