Incorporating Expert Judgement into Bayesian Network Machine Learning

Yun Zhou[1,2], **Norman Fenton**[1], **Martin Neil**[1], **Cheng Zhu**[2]
[1]Queen Mary University of London, UK
[2]National University of Defense Technology, China
{yun.zhou, norman, martin}@eecs.qmul.ac.uk; chengzhu@nudt.edu.cn

Abstract

We review the challenges of Bayesian network learning, especially parameter learning, and specify the problem of learning with sparse data.We explain how it is possible to incorporate both qualitative knowledge and data with a multinomial parameter learning method to achieve more accurate predictions with sparse data.

1 Review of Bayesian Network Learning

Constructing a Bayesian network (BN) from data is widely accepted as a major challenge in decision-support systems. For many critical risk analysis problems, decisions must be made where there is sparse or no direct historical data to draw upon, or where relevant data is difficult to identify. The challenge is especially acute when the risks involve novel or rare systems and events [Fenton and Neil, 2012] (e.g. think of novel project planning, predicting events like accidents, terrorist attacks, and cataclysmic weather events).

There are two typical categories of problems in learning BNs: one is parameter learning given a fixed graphical structure of the BN; and the other is structure learning, where the BN structure is unknown. Ideally, with sufficient data, classical learning algorithms like BDeu+MLE, PC+MLE or hybrid+MLE [Campos, 2007] can learn BNs that fit the true model in distribution and structure. However, these learning algorithms do not work when there is sparse data. To mitigate this problem, expert judgements are needed to supplement learning.

In the absence of data, experts are usually required to provide strong information like causality between nodes in structure learning, and specific numerical probability values of Node Probability Tables (NPTs) or Dirichlet priors in parameter learning. Such strong judgments can easily cause bias. Studies show that experts are often overconfident in providing qualitative knowledge rather than quantitative estimations [Druzdzel and van der Gaag, 2000]. Typical qualitative knowledge are constraints that limit the number of parents of a node (in structure learning) and equality/inequality relations among parameters in parameter learning; such constraints cut the search space significantly, and help escape local maxima. Because of the potential benefits, there is an increasing research interest in incorporating constraints into structure/parameter learning. Recent reseach developments have focused on triggering the necessary automated calculations and inferences to get more accurate BNs under these constraints.

For parameter learning, some approaches formulate this problem as a general constrained maximization problem, and outline the details of the classification of parameter constraint types ([Niculescu *et al.*, 2006] and [Liao and Ji, 2009]). Unfortunately, these approaches can be extremely inefficient for BNs with a large number of parameters. Nor can they handle exterior constraints among parameters, as discussed in [Feelders and van der Gaag, 2006] and [Tong and Ji, 2008]. The work of [Tong and Ji, 2008] has limited forms of constraints, while and the work of [Feelders and van der Gaag, 2006] can only learn the parameter for binary variables. Hence, our research is focused on the development of an extended BN graphical notation, and associated algorithms, to integrate judgements provided by domain experts in a much richer and less constrained way than the current state-of-the-art of modelling and tools supports.

2 Proposed Solution

For parameter learning, data statistics can be regarded as a sequence of independent Bernoulli experiments on different parameters. Given the number of Bernoulli experiments on a parameter, and observed results, the success probability of each value can be estimated by Bayesian inference. For each parameter in the BN, an auxiliary BN model called multinomial parameter learning model (MPL) can be created [Zhou *et al.*, 2013]. Then constraints can be integrated as additional nodes connected with this MPL.

Before inference, constraints and data observations are transferred as evidence in the separate BN, which are used to update the posterior distributions of target values. Inference refers to the process of computing the discretised posterior marginals of constrained values after obtaining the observations of its constraint. Because the parameter values are continuous, our approach requires the dynamic discretization inference algorithm [Neil *et al.*, 2007] to compute posteriors in this hybrid BN model. This algorithm has been implemented in the Agenarisk toolset [Agenarisk, 2013]. Any continuous node is implemented as a ‘simulation’ node meaning that its discretization will be calculated dynamically by the algorithm. The work presented here requires such an accurate dynamic discretization algorithm (other standard BN tools will not be able to achieve the same results), which provides the flexibility needed for inference with any constraint of interest.

In [Zhou *et al.*, 2013] we applied this approach withconstraints limited within a single parameter. Although the results were very promising, it limits the type of expert knowledge provided. We now extend the approach to handle a) constraints between parameters and b) uncertain relationships. Firstly, for an exterior constraint between two values from different parameters, we have a converging hybrid BN fragment, where the constraint node is the discrete child of two continuous parents that represent target values of different parameters. The constraint node is a binary (True/False) node with mathematical expressions that specify the constraint relationships between its parents. Secondly, for uncertain constraints, we can add soft evidence on them.

Suppose, for example, there are two parameters that represent the probabilities of getting cancer given smoker or not. An expert might assert that *P(Cancer|Smoker)>P(Cancer|Nonsmoker)* (95% confidence). Figure 1 shows the model (MPL with uncertain constraints, MPL-UC) for this simple example, where the true discrete probability distribution of these parameters are (0.75, 0.25) for *P1* and (0.80, 0.20) for *P2*, and their numbers of observed data for different values are (8, 2) and (7, 3) respectively.

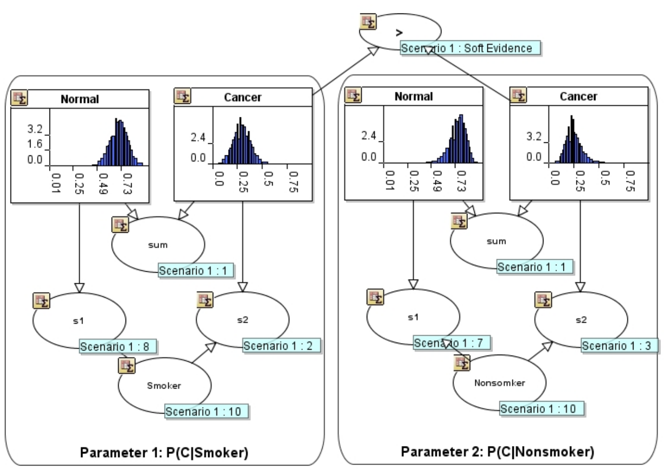

Figure 1: The MPL-UC model for two parameters, where the means of posterior distributions of success probabilities are elicited to update the former NPTs. The learning accuracy of MPL-UC is improved since its K-L distance is only 0.0166 compared with 0.0507 for MLE.

3 Progress

Our aim is to both reduce at source much unnecessary data collection and improve the results of analysis of data that is collected. In [Zhou *et al.*, 2013], we tested our method for parameter learning in a real-world problem called 'defects predication network'. We showed experimental results indicating that we achieved acceptable K-L distance under 50 data samples, whereas the MLE needs more than 500 data samples to achieve the equivalent result. For structure learning, we are developing methods to elicit constraints from generic, specific, or user-defined idioms in situations where little or no data is available [Fenton and Neil, 2012].

4 Future Work

We plan to incorporate our work into existing BN software (like Agenarisk), i.e. to provide a GUI to directly accumulate expert judgements. Also, we plan to develop a method to integrate judgements from multiple experts and handle inconsistent judgements. This would help to gain less one-sided domain knowledge.

Acknowledgments

This work was supported by Agenarisk, CSC/QM Joint PhD scholarship and National Natural Science Foundation of China (grant no. 71001105 and 61273322).

References

[Agenarisk, 2013] Agenarisk. Bayesian network and simulation software for risk analysis and decision support, http://www.agenarisk.com/, 2013.

[Campos, 2007] Luis M. Campos. A scoring function for learning bayesian networks based on mutual information and conditional independence tests. *Journal of Machine Learning Research*, 7(2):2149, 2007.

[Druzdzel and van der Gaag, 2000] Marek Druzdzel and Linda van der Gaag. Building probabilistic networks:where do the numbers come from? *IEEE Transactions on Knowledge and Data Engineering*, 12(4):481–486, 2000.

[Feelders and van der Gaag, 2006] Ad Feelders and Linda van der Gaag. Learning bayesian network parameters under order constraints. *International Journal of Approximate Reasoning*, 42(1):37–53, 2006.

[Fenton and Neil, 2012] Norman Fenton and Martin Neil. *Risk Assessment and Decision Analysis with Bayesian Networks*. CRC Press, New York, 2012.

[Liao and Ji, 2009] Wenhui Liao and Qiang Ji. Learning bayesian network parameters under incomplete data with domain knowledge. *Pattern Recognition*, 42(11):3046–3056, 2009.

[Neil *et al.*, 2007] Martin Neil, Manesh Tailor, and David Marquez. Inference in hybrid bayesian networks using dynamic discretization. *Statistics and Computing*, 17(3):219–233, 2007.

[Niculescu *et al.*, 2006] Radu Stefan Niculescu, Tom Mitchell, and Bharat Rao. Bayesian network learning with parameter constraints. *The Journal of Machine Learning Research*, 7:1357–1383, 2006.

[Tong and Ji, 2008] Yan Tong and Qiang Ji. Learning bayesian networks with qualitative constraints. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008.*, pages 1–8. IEEE, 2008.

[Zhou *et al.*, 2013] Yun Zhou, Norman Fenton, and Martin Neil. Bayesian network approach to multinomial parameter learning using data and expert judgements. *submitted for publication*, pages 1–12, 2013.