Learning Bayesian network parameters from small data sets: A further constrained qualitatively maximum a posteriori method

Zhi-gao Guo, Xiao-guang Gao, Hao Ren, Yu Yang, Ruo-hai Di, Da-qing Chen

Department of System Engineering, Northwestern Polytechnical University, China
Division of Computer Science and Informatics, London South Bank University, UK

Abstract

To improve the learning accuracy of the parameters in a Bayesian network from a small data set, domain knowledge is normally incorporated into the learning process as parameter constraints. MAP-based (Maximum a Posteriori) methods that utilize both sample data and domain knowledge have been well studied in the literature. Among all the MAP-based methods, the QMAP (Qualitatively Maximum a Posteriori) method exhibits the best learning performance. However, when the data is insufficient, the estimation given by the QMAP often fails to satisfy all the convex parameter constraints, and this has made the overall QMAP estimation unreliable. To ensure that QMAP estimation does not violate any given parameter constraints and to further improve the learning accuracy, a FC-QMAP (Further Constrained Qualitatively Maximum a Posteriori) algorithm is proposed in this paper. The algorithm regulates QMAP estimation by replacing data estimation with a further constrained estimation via convex optimization. Experiments and theoretical analysis show that the proposed algorithm outperforms most of the existing parameter learning methods (namely, Maximum Likelihood, Constrained Maximum Likelihood, Maximum Entropy, Constrained Maximum Entropy, Maximum a Posteriori, and Qualitatively Maximum a Posteriori).

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

A Bayesian network (BN) is typically a directed acyclic graph representing a model that combines probability theory and graphical model theory. The BN was systematically introduced in 1988 [1] by Judea Pearl. Following an intensive research for about 30 years worldwide, BNs have become a powerful tool for uncertainty analysis and have been applied to deal with a wide range of issues, including gene analysis [2], robot control [3], fault diagnoses [4], target tracking [5], signal processing [6], ecosystem modeling [7] and educational measurement [8].

Usually, to construct a BN from data, reasonably-sized samples are required, depending on the topology and complexity of the network. If sufficient samples are available, the construction of an accurate BN is easy and can be accomplished...
by traditional methods such as maximum likelihood (ML) [9]. Unfortunately, collecting a large amount of data is difficult for some decision-making problems under certain circumstances, such as rare disease diagnosis [10], earthquake prediction [11] and parole assessment [12]. In such cases, knowledge from domain experts is often considered as supplementary information in constructing the network.

In practice, domain experts usually feel comfortable with providing qualitative parameter constraints in the form of $p_1 > 0.8$, $p_1 \approx p_2$, $p_1 > p_2$, $(p_1 + p_2) > (p_3 + p_4)$ [13], etc., where $p_1$, $p_2$, $p_3$ and $p_4$ denote parameters in a Bayesian network. Although such constraints look simple, they can be very effective for improving the BN modeling accuracy, especially when the given data set is small. In this paper, we focus on learning BNs in cases where the sample size is small but a domain expert is involved in the learning process by providing domain knowledge about constraints on the relevant network parameters. By incorporating qualitative parameter constraints with the limited data, we provide an improved maximum a posteriori estimation method.

The remainder of the paper is organized as follows: In Section 2, the related works on parameter learning are introduced, especially works with both sample data and parameter constraints. The BN parameter learning problem studied in this paper is formalized and described in detail in Section 3. An improved maximum a posteriori estimation algorithm and a theoretical analysis on the algorithm are presented in Section 4. In Section 5, a detailed numerical example is given to illustrate the principle of the proposed algorithm comparatively along with the ML and QMAP algorithms. In Section 6, a set of simulation experiments are presented with four benchmark Bayesian networks to compare the performance of the proposed algorithm with other parameter learning algorithms. Finally, Section 7 summarizes the main conclusion and findings of the paper and briefly explores several interesting directions for future research.

### 2. Related work

Methods for BN parameter learning from small data sets can be mainly grouped into two types: MAP-based methods and non-MAP-based methods. In non-MAP-based methods, the BN parameters are computed by the optimization or the regulation of constrained parameter estimation models. Among those methods, Wittig [14] proposed a constrained parameter learning algorithm. This algorithm can be applied to cross-distribution parameter constraints,¹ which define the relative relations between a pair of parameters over two different distributions. First, a set of parameter constraint models are constructed from qualitative expert knowledge. Then, an optimization model consisting of an entropy function and the parameter constraints is built. Finally, the built optimization model is optimized using the adaptive probabilistic networks method. Altendorf [15] also discussed a parameter learning method applicable to the cross-distribution constraints. What is interesting about this method is that it defines an objective function that integrates the parameter constraint model into an entropy function, and the function is then solved using the gradient-descent algorithm. Feelders [16] proposed an isotonic regression estimation method concerned with the cross-distribution constraints. Their algorithm employs the ML method to learn a set of initial parameters at the beginning, and then elicits parameter orders from parameter constraints. The initial parameters are regulated by the algorithm so that the regulated parameters satisfy all the parameter orders. Isozaki [17] suggested a minimum free energy method that is suitable for axiomatic parameter constraints.² Essentially, this method starts by constructing a free energy function. This energy function consists of the Kullback–Leibler divergence and an entropy function, and is used as the objective function. Furthermore, the energy function and parameter constraints are integrated by the Lagrange multipliers, and the gradient-descent method is employed to solve the problem. The constrained maximum likelihood method was proposed by Campos [18]. This method works to any convex parameter constraints. Parameter constraints are created from domain expert knowledge, and a convex optimization model is then constructed with likelihood function and parameter constraints. The model can be optimized using the convex optimization method. Campos discussed the constrained maximum entropy method applicable to any convex parameter constraints [19]. In this method, an imprecise Dirichlet model that combines the prior information and the data set is created as a supplementary parameter constraint. Further, a convex optimization model containing an entropy function and convex parameter constraints is constructed. The convex optimization method can be applied to solve the formulated model. Zhou [20] suggested a method for dealing with the cross-distribution constraints, named constrained optimization with flat prior. An objective function is considered in this method that combines the likelihood function and the penalty function derived from constraint violations. The objective function is solved using the sequential quadratic programming and the solutions are taken as the optimal parameters.

In MAP-based methods, BN parameters are computed as linear interpolation values of the sample observations and the prior information. Among them, the qualitative maximum a posteriori method [21] has been designed to tackle any convex parameter constraints. The method requires a certain amount of possible parameters to be sampled from parameter constraints using the rejection–acceptance sampling strategy. In addition, hyper-parameters of the prior Dirichlet distribution are determined as the products of a equivalent sample size and the mean values of the sampled parameters. The optimal parameters are then computed as the interpolations of the sample observations and the hyper-parameters. A method, named

---

¹ Cross-distribution constraints are very common in real-world problems and have been referred to as “monotonicity constraints”, “order constraints” or “monotonic influence constraints” in other studies [14] [15] [20].

² Axiomatic parameter constraints are from the law of probability and are hence not required to be provided by the domain experts.

Please cite this article in press as: Z.-g. Guo et al., Learning Bayesian network parameters from small data sets: A further constrained qualitatively maximum a posteriori method, Int. J. Approx. Reason. (2017), http://dx.doi.org/10.1016/j.ijar.2017.08.009
دریافت فوری
متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات

ISIArticles
مرجع مقالات تخصصی ایران