Abstract

We propose a novel framework for performing quantitative Bayesian inference based on qualitative knowledge. Here, we focus on the treatment in the case of inconsistent qualitative knowledge. A hierarchical Bayesian model is proposed for integrating inconsistent qualitative knowledge by calculating a prior belief distribution based on a vector of knowledge features. Each inconsistent knowledge component uniquely defines a model class in the hyperspace. A set of constraints within each class is generated to describe the uncertainty in ground Bayesian model space. Quantitative Bayesian inference is approximated by model averaging with Monte Carlo methods. Our method is firstly benchmarked on ASIA network and is applied to a realistic biomolecular interaction modeling problem for breast cancer bone metastasis. Results suggest that our method enables consistently modeling and quantitative Bayesian inference by reconciling a set of inconsistent qualitative knowledge.

1. Introduction

Bayesian reasoning provides a probabilistic approach to inference. In Bayesian framework, quantities of interest are described by probabilities and optimal decisions can be made by reasoning about these probabilities together with the observation or evidence. Bayesian reasoning is important to machine learning because it provides a quantitative approach to weighting the evidence supporting alternative hypotheses. Numerous algorithms have been proposed for learning the Bayesian network structure and parameter from the observed data. These algorithms produce a single Bayesian model by maximizing its probability given the training data, i.e. maximum a posterior approximation. In realistic problem, learning Bayesian model by training data requires relatively large amount of observed data comparing to the size of network. However, the data basis is often very sparse and it is hardly sufficient to select one adequate model due to the model uncertainty, thus, selecting a single model may induce overfitting to the data and can lead to strongly biased inference results. It is therefore preferable to adopt a full Bayesian approach with model averaging.

Besides the training data, the prior background knowledge provides many ways to adjust uncertainties. The prior background knowledge includes qualitative and quantitative knowledge which describes the entities and their relationships with different levels of abstraction. Quantitative knowledge can be exemplified by a probability elicitation procedure from a domain expert. In most domains, this is particularly difficult due to the limitations of expert knowledge in this level. In contrast, qualitative knowledge, which only provides loose constraints with uncertainty on the entities and their relations exist in many science and engineering domains. For example, in biomedicine, the statement: “Gene CTGF, IL11 and OPN cooperatively activate bone metastasis in breast cancer”, entities are gene CTGF, IL11, OPN and Bone metastasis in breast cancer, their qualitative relation: cooperatively activate. In some cases, there are properties which further specify the qualitative relationship. In “The risk of lung cancer among smokers is approximately 10 times higher than non-smokers”.

$^\star$ An abbreviated version of some portions of this article appeared in Chang and Stetter (2007b) as part of the IJCNN 2007 Conference Proceedings, published under IEE copyright.

$^\ast$ Corresponding author.

E-mail address: chang.rui@hotmail.com (R. Chang).
Smoking cause lung cancer and the influence is 10 times higher to non-smokers. In recent studies (Chang & Stetter, 2007a, 2007b), it is shown that qualitative knowledge can be used and translated into a set of constraints on the Bayesian model space. This set of constraints defines the model uncertainty in structure and parameter space respectively. The model uncertainty represented by the qualitative knowledge enables the full Bayesian approach where a class of Bayesian networks which are consistent with the semantics of the set of qualitative hypotheses are drawn according to the model uncertainty. The probabilistic network inference and reasoning can be derived by performing quantitative prediction and inference in each of the Bayesian model and these quantitative results are averaged weighted by the model posterior probability. This approach has been successfully applied to both well-known benchmark model and real-world application. However, one significant drawback of this qualitative knowledge-driven probabilistic network modeling and inference approach is its incapability of dealing with inconsistent qualitative knowledge. It is well-known that knowledge are often inconsistent, i.e., in the same domain, there may exist contradicting qualitative statements on dependency, causality and parameters over a set of entities. Therefore, it is imperative to develop methods for reconciling inconsistent qualitative knowledge and for modeling Bayesian networks and performing quantitative prediction. In this paper, we propose a novel framework for performing quantitative Bayesian inference with model averaging based on the inconsistent qualitative statements as a coherent extension of framework of quantitative Bayesian inference based on a set of consistent hypotheses introduced in Chang and Stetter (2007b). Our method interprets the qualitative statements by a vector of knowledge features whose structure can be represented by a hierarchical Bayesian network. The prior probability for each qualitative knowledge component is calculated as the joint probability distribution over the features and can be decomposed into the production of the conditional probabilities of the knowledge features. These knowledge components define multiple Bayesian model classes in the hyperspace. Within each class, a set of constraints on the ground Bayesian model space can be generated. Therefore, the distribution of the ground model space can be decomposed into a set of weighted distributions determined by each model class. This framework is used to perform full Bayesian inference which can be approximated by Monte Carlo methods, but is analytically tractable for smaller networks and statement sets.

In our approach, qualitative knowledge is well formulized to construct the prior distribution over the structure and parameter space. The inputs are only qualitative statements with certain fuzzy cause-effect relationships between the domain variables. The uncertainty in these statements are modeled and translated into belief distribution on the Bayesian model space. Therefore, each Bayesian model in the space produces a quantitative prediction which will be weighted by the models belief assignment in the space. The novelty of our approach is that we take advantage of model uncertainty modeling and full Bayesian approach with model averaging to produce quantitative predictions and inference based on only qualitative information. Further quantitative information from data is not needed. This is especially useful when there are sparse data in the domain. Moreover, the approach described in this paper is a coherent extension of the methods introduced in Chang and Stetter (2007b), in that inconsistent qualitative knowledge can be integrated to model the uncertainty in the knowledge space, thus, multiple classes of Bayesian networks can be constructed for performing the quantitative inference weighted
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