Expert Systems With Applications 55 (2016) 361–373

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

When and where to transfer for Bayesian network parameter learning

Yun Zhou [a,b,*], Timothy M. Hospedales [a], Norman Fenton [a]

[a] Risk and Information Management (RIM) Research Group, Queen Mary University of London, United Kingdom
[b] Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, China

ARTICLE INFO

Keywords:
Bayesian networks parameter learning
Transfer learning
Bayesian model comparison
Bayesian model averaging

ABSTRACT

Learning Bayesian networks from scarce data is a major challenge in real-world applications where data are hard to acquire. Transfer learning techniques attempt to address this by leveraging data from different but related problems. For example, it may be possible to exploit medical diagnosis data from a different country. A challenge with this approach is heterogeneous relatedness to the target, both within and across source networks. In this paper we introduce the Bayesian network parameter transfer learning (BNPTL) algorithm to reason about both network and fragment (sub-graph) relatedness. BNPTL addresses (i) how to find the most relevant source network and network fragments to transfer, and (ii) how to fuse source and target parameters in a robust way. In addition to improving target task performance, explicit reasoning allows us to diagnose network and fragment relatedness across Bayesian networks, even if latent variables are present, or if their state space is heterogeneous. This is important in some applications where relatedness itself is an output of interest. Experimental results demonstrate the superiority of BNPTL at various scarcities and source relevance levels compared to single task learning and other state-of-the-art parameter transfer methods. Moreover, we demonstrate successful application to real-world medical case studies.

1. Introduction

Bayesian networks (hereafter referred to by the abbreviation BNs) have proven valuable in modeling uncertainty and supporting decision making in practice (Fenton & Neil, 2012; Pearl, 1988). However, in many applications it is hard to acquire sufficient examples to learn BNs effectively from data. For example, in a small hospital or country there may be insufficient data to learn an effective medical diagnosis network. However, directly applying a network learned in another domain may be inaccurate or impossible because the underlying tasks may have quantitative or qualitative differences (e.g., care procedures vary across hospitals and countries). In this paper we investigate leveraging BNs in different but related domains to assist learning a target task with scarce data. This is an important capability in at least two distinct scenarios: (i) those where the source tasks are the same as the target, but have different specific statistics (e.g., due to different demographic statistics in another country), and (ii) those where the source tasks are related to the target in a *piecewise* way, (the target and source tasks are not the same, but share common sub-graphs, e.g., two hospitals share a subset of procedures; or two diseases share a subset of symptoms).

The proposed contribution falls under the topical area of transfer learning (Pan & Yang, 2010; Torrey & Shavlik, 2009) (also known as domain adaptation), which aims to significantly reduce data requirements by leveraging data from related tasks. Transfer has been successfully applied in a variety of machine learning areas for example, recommendations (Pan, Xiang, & Yang, 2012), classification (Li, Jin, & Long, 2012; Ma, Luo, Zeng, & Chen, 2012) and natural language processing (Collobert & Weston, 2008). Central challenges include computing *when to transfer* (transfer or not depending on relevance), *from where* (which of multiple sources of varying relevance) (Eaton, desJardins, & Lane, 2008; Mihalkova & Mooney, 2009) and *how* (how to fuse source and target information). These are crucial to ensure that transfer is helpful, and avoid 'negative transfer' risk (Pan et al., 2012; Seah, Ong, & Tsang, 2013a). Despite the popularity of transfer learning, limited work (Luis, Sucar, & Morales, 2010; Niculescu-mizil & Caruana, 2007; Oyen & Lane, 2012) has been done on transfer learning of BNs. Outstanding challenges in BN transfer include dealing automatically with from where to transfer, transferring in the presence of latent variables and transferring between networks with heterogeneous state spaces. In this paper we introduce the first framework that resolves these issues in a BN context, leveraging the structured nature of

* Corresponding author at: Risk and Information Management (RIM) Research Group, Queen Mary University of London, United Kingdom. Tel.: +4407429481207.

E-mail addresses: yun.zhou@qmul.ac.uk, zhouy.nudt@gmail.com (Y. Zhou), t.hospedales@qmul.ac.uk (T.M. Hospedales), n.fenton@qmul.ac.uk (N. Fenton).

http://dx.doi.org/10.1016/j.eswa.2016.02.011
0957-4174/

BNs for piecewise transfer, so multiple sources of partial relevance and potentially heterogeneous state spaces can be exploited.

In this paper we assume the target and source domain structures are provided[1] and concentrate on the challenges of learning the target network parameters in the presence of latent variables and from multiple sources of varying – continuous and/or piecewise – relevance. Importantly, we do not require that the source and target networks correspond structurally, or that node names are shared. Our novel solution involves splitting the target and source BNs into fragments (sub-graphs) and then reasoning explicitly about both network-level and fragment-level relatedness. Reasoning simultaneously about both is important, because pure fragment-level relatedness risks over-fitting if there are many sources. We achieve this via an Expectation Maximization (EM) style algorithm that alternates between (i) performing a Bayesian model comparison to infer per-fragment relatedness and (ii) updating a source network relatedness prior. This solves when and from where to transfer at both coarse and fine-grained level. Finally, the actual transfer is performed per-fragment using Bayesian model averaging to robustly fuse the source and target fragments, addressing how and how much to transfer. In this way we can deal robustly with a variety of transfer scenarios including those where the source networks are: (i) highly relevant or totally irrelevant, (ii) have the same or heterogeneous state spaces and (iii) uniform or piecewise (varying per sub-graph) relevance. Our explicit network and fragment relatedness reasoning also provides a diagnostic of which networks/domains are similar, and which subgraphs are common or distinct. This is itself an important output for applications where quantifying relatedness, and uncovering the source of heterogeneity between two domains is of interest (e.g., revealing differences in treatment statistics between hospitals). To evaluate our contribution, we conduct experiments on six standard networks from a BN repository, comparing against various single task baselines and prior transfer methods. Finally, we apply our method to transfer learning in two real-world medical networks.

2. Related work

Expert elicitation. An advantage of BNs is their interpretable nature means that experts can define variables, structure and parameters in the absence of data. Nevertheless, learning BNs from data is of interest because there are many situations for which there is no available expert judgment, or where it may not be possible to elicit the conditional probability tables (CPTs). Studies have therefore tried to bridge the gap between these two paradigms. Most typically, experts specify a semantically valid network structure, and CPTs are learned from data. Recently, expert specified qualitative constraints on CPTs have been exploited to improve parameter learning. This is done, for example, via establishing a constrained optimization problem (Altendorf, 2005; de Campos & Ji, 2008; de Campos, Zeng, & Ji, 2009; Liao & Ji, 2009; Niculescu, Mitchell, & Rao, 2006) or auxiliary BNs (Khan, Poupart, & Agosta, 2011; Zhou, Fenton, & Neil, 2014a, 2014b). In this study we exploit the ability of experts to easily specify a network structure and focus on transfer to improve quantitative estimation of parameters.

CPTs combination. When there is limited training data, researchers have attempted to construct CPTs from different relevant sources of information. Given a set of CPTs involving the same variables, conventional methods to aggregate them are linear aggregation (i.e., weighted sum) and logarithmic aggregation (Chang & Chen, 1996; Chen, Chiu, & Tseng, 1996; Genest & Zidek, 1986). Based on this, the work of (Luis et al., 2010) introduced the DBLP (distance based linear pooling) and LoLP (local linear pooling) aggregation methods by considering the CPTs' confidences and similarities learnt from the original datasets. This method highlighted the importance of measuring the weights/confidences of different CPTs. However, the method is a too simplistic heuristic: confidence values depend only on the CPT entry size and dataset size, without considering the fit to the target training data.

Transfer learning. Transfer learning in general is now a well studied area, with a good survey provided by Pan and Yang (2010). Extensive work has been done on transfer and domain adaptation for flat machine learning models, including unsupervised transfer and analysis of relatedness (Duan, Tsang, Xu, & Chua, 2009; Eaton et al., 2008; Seah et al., 2013a; Seah, Tsang, & Ong, 2013b). However, these studies have generally not addressed one or more of the important conditions that arise in the BN context addressed here, notably: transfer with heterogeneous state space, piece-wise transfer from multiple sources (a different subset of variables/dimensions in each source may be relevant), and scarce *unlabeled* target data (thus precluding conventional strategies that assume ample unlabeled target data, such as MMD (Huang, Smola, Gretton, Borgwardt, & Scholkopf, 2007; Seah et al., 2013b)).

Transfer learning in BNs. In the context of transfer learning in BNs, the multi-task framework of Niculescu-mizil and Caruana (2007) considers structure transfer. However, it assumes that all sources are equally related and simply learns the parameters for each task independently. Kraisangka and Druzdzel (2014) construct BN parameters from a set of regression models used in survival analysis. However, this method cannot be generalized to transfer between BNs. The transfer framework of (Luis et al., 2010) covers a more similar parameter transfer problem to ours and proposes a method to fuse source and target data. However, the heuristic CPT fusion used assumes every source is both relevant and equally related. It is not robust to the possibility of irrelevant sources and does not systematically address when, from where, and how much to transfer (as shown by our experiments where this method significantly underperforms ours). The study (Oyen & Lane, 2012) considers multi-task structure learning, again with independently learned parameters. They investigate network/task-level relatedness, showing transfer performs poorly without knowledge of relatedness. However, they address this by using manually specified relatedness. Finally, a recent study (Oates, Smith, Mukherjee, & Cussens, 2014) improves this by automatically inferring the network/task-level relatedness. However, they do not consider information sharing of parameters. In contrast, we explicitly learn about both network and fragment-level relatedness from data. None of these prior studies cover transfer with latent variables or heterogeneous state spaces.

A related area to BN transfer is transfer in Markov Logic Networks (MLNs) (Davis & Domingos, 2009; Mihalkova, Huynh, & Mooney, 2007; Mihalkova & Mooney, 2009). In contrast to these studies, our approach has the following benefits: We can exploit multiple source networks rather than exactly on each; we automatically quantify source relevance and are robust to some or all irrelevant sources (rather than assuming a single relevant source); these MLN studies use the transferred clauses directly rather than weighting the resulting transfer by estimated relevance.

3. Model overview

3.1. Notation and definitions

In a BN parameter learning setting, a domain $\mathcal{D} = \{V, G, D\}$ consists of three components: variables $V = \{X_1, X_2, X_3, \ldots, X_n\}$

[1] This is easiest to elicit from experts, and is moreover required in many domains such as medicine where the structure must be semantically meaningful to be acceptable to end users.