Computational Statistics and Data Analysis 55 (2011) 3135–3147

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Structural learning for Bayesian networks by testing complete separators in prime blocks

Ping-Feng Xu [a,b], Jianhua Guo [a,*], Man-Lai Tang [c]

[a] Key Laboratory for Applied Statistics of MOE and School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, Jilin Province, China
[b] School of Basic Science, Changchun University of Technology, Changchun 130012, Jilin Province, China
[c] Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong, China

ARTICLE INFO

Article history:
Received 3 March 2010
Received in revised form 10 June 2011
Accepted 14 June 2011
Available online 22 June 2011

Keywords:
Bayesian network
Complete separator
Conditional independence
Moral edge
Prime block
Structural learning

ABSTRACT

In this paper, we consider how to recover the structure of a Bayesian network from a moral graph. We present a more accurate characterization of moral edges, based on which a complete subset (i.e., a separator) contained in the neighbor set of one vertex of the putative moral edge in some prime block of the moral graph can be chosen. This results in a set of separators needing to be searched generally smaller than the sets required by some existing algorithms. A so-called structure-finder algorithm is proposed for structural learning. The complexity analysis of the proposed algorithm is discussed and compared with those for several existing algorithms. We also demonstrate how to construct the moral graph locally from, separately, the Markov blanket, domain knowledge and *d*-separation trees. Simulation studies are used to evaluate the performances of various strategies for structural learning. We also analyze a gene expression data set by using the structure-finder algorithm.

1. Introduction

Bayesian networks play an important role in decision making and statistical inference and have been applied in many fields such as machine learning and bioinformatics. They offer powerful knowledge representations for independence, conditional independence and causal relationships among variables in a given domain (Whittaker, 1990; Lauritzen, 1996; Cowell et al., 1999; Pearl, 2000; Spirtes et al., 2000; Jensen and Nielsen, 2007).

Briefly, a Bayesian network is a directed graph with no directed cycles (i.e., a directed acyclic graph, denoted as DAG) $G = (V, E)$ with a probability distribution P. The vertices in G usually represent random variables $X = (X_v)_{v\in V}$, and P is the joint probability distribution of X with $P(x) = \prod_{v\in V} P(x_v|x_{pa(v)})$, where $P(x_v|x_{pa(v)})$ is a conditional distribution and $pa(v)$ is the set of parents of v. The DAG is its qualitative component which represents dependence and independence relationships. That is, the absence of some directed edges represents the existence of certain conditional independence relationships among variables, and the presence of edges represents the existence of direct dependence relationships or causal relationships. The joint probability distribution is its quantitative component that represents the strength of association between variables.

For a Bayesian network $G = (V, E)$ with a distribution P, if subsets A and $B \subset V$ are d-separated by S in G (see the corresponding definition in Section 2) then X_A and X_B are conditionally independent given X_S with respect to P. In this paper, we assume that all conditional independencies among variables implied in the true distribution can be indicated by

* Corresponding author. Tel.: +86 431 85098576; fax: +86 431 85098237.
E-mail addresses: xupf900@gmail.com (P.-F. Xu), jhguo@nenu.edu.cn (J. Guo), mltang@math.hkbu.edu.hk (M.-L. Tang).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2011.06.017

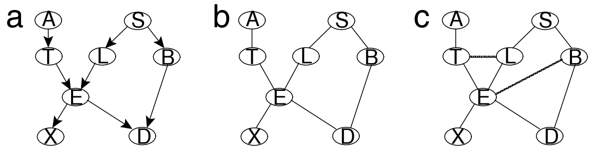

Fig. 1. (a) DAG, (b) skeleton, and (c) moral graph (with gray lines being moral edges) of the Asia network.

d-separations of some DAGs, which is called the faithfulness assumption (Spirtes et al., 2000). Two DAGs over the same variable set are said to be Markov equivalent if they represent the same conditional independencies among variables. An equivalence class of Markov equivalent DAGs is represented by a partially directed DAG (PDAG) in which the directed edges are common to every DAG, while the undirected edges may be oriented in one way in some DAGs and in another way in other DAGs. Hence, the goal of structural learning is to construct a partially directed graph to represent the equivalence class from an observed data set.

In this paper, we consider the problem of the learning structure of a Bayesian network. This problem has been widely discussed by many authors (e.g. Cowell et al., 1999; Spirtes et al., 2000) and references therein). There are two major kinds of algorithms for structural learning, namely constraint-based approaches (Spirtes and Glymour, 1991; Verma and Pearl, 1990; Xie et al., 2006; Xie and Geng, 2008; Ma et al., 2008; Liu et al., 2010) and search-and-score approaches (Chickering, 2002; Heckerman et al., 1995). We focus on the constraint-based approaches in this article.

For many constraint-based algorithms, two major steps are usually adopted to recover the DAG structure (Xie and Geng, 2008). First, they learn the moral graph of the target DAG by applying Markov boundary learning algorithms, where the Markov boundary for a variable u is defined to be the set of variables composed of u's parents, its children, and its children's other parents (Pearl, 1988). Second, they perform further independence tests for deleting the moral edges and orienting edges on the basis of the moral graph learned in the first step. Therefore, as mentioned in (Xie and Geng, 2008), in a constraint-based algorithm, searching for separators of pairs of variables is a major challenge for the orientation of edges and for the structure recovery of a DAG. Here a separator is a subset of variables given which the variable pairs are conditionally independent. Verma and Pearl (1990) proposed the inductive causation (IC) algorithm that searches for a separator from all possible subsets of the vertex set. Spirtes and Glymour (1991) proposed the PC algorithm, which is a general systematic way of searching for a separator in increasing order of cardinality within a constraint-based framework. However, searching for a separator is limited to those vertices that are adjacent to the vertex pairs in the PC algorithm. Geng et al. (2005) decomposed the moral graph into subgraphs (i.e., prime blocks), and then searched for a separator in all possible subsets in prime blocks.

In this paper, we present a more accurate characterization of moral edges in the moral graph. We show that a separator corresponding to a putative moral edge can be obtained in the set of complete subsets contained in the neighbor set of one vertex of the putative moral edge in some prime block. Using this result, we propose a structure-finder algorithm for structural learning from the moral graph, which substantially improves on the IC algorithm and the decomposition approach proposed by Geng et al. (2005). We also discuss how to construct moral graphs using marginal data locally. The rest of this paper is organized as follows. In Section 2, we introduce notation and definitions. In Section 3, we give the characterizations of moral edges, and present the structure-finder algorithm and a complexity analysis. Section 4 discusses how to construct the moral graph locally from, separately, the Markov blanket, domain knowledge and d-separation trees. Simulation studies are conducted to demonstrate the performance of our algorithm and existing algorithms in Section 5. We analyze a gene expression data set in Section 6. Brief conclusions are drawn in Section 7. All proofs will be presented in the Appendix.

2. Notation and definitions

In this section, we provide some basic technical terminologies that are sufficient for understanding this article. For more details, see Cowell et al. (1999) and Lauritzen (1996).

Let $G = (V, E)$ be a directed graph without any directed cycles (DAG). A vertex $\alpha \in V$ ($\beta \in V$) is called a parent (child, respectively) of β (α) if the directed edge $(\alpha, \beta) \in E$. The set of all parents of α is denoted by $pa(\alpha)$. A path from α to β is a sequence $[\alpha = \alpha_0, \alpha_1, \ldots, \alpha_n = \beta]$ of distinct vertices such that $(\alpha_{i-1}, \alpha_i) \in E$ or $(\alpha_i, \alpha_{i-1}) \in E$ for $i = 1, \ldots, n$. If there is a directed path from α to β, we call α (β) an ancestor (descendant, respectively) of β (α). The set of ancestors of β is denoted by $an(\beta)$. For any subset $B \subseteq V$, suppose that $An(B) = [\cup_{b \in B} an(b)] \cup B$.

For a DAG $G = (V, E)$, its skeleton $G^u = (V, E^u)$ is an undirected graph obtained by dropping the directions of the edges in G. The moral graph $G^m = (V, E^m)$ of G is formed by connecting vertices that have a common child, and then making all edges in the graph undirected. The fill-in edge (α, β) is called a moral edge, that is, $(\alpha, \beta) \notin E^u$, and α and β have a common child. A triple $(\alpha, \upsilon, \beta)$ of distinct vertices is called a υ-structure if (α, υ) and $(\beta, \upsilon) \in E$, but $(\alpha, \beta) \notin E^u$.

In a DAG $G = (V, E)$, we call γ a collider (non-collider) in a path π if arrows of π (do not, respectively) meet head to head at γ. A path π is said to be d-separated by $S \subseteq V$ in G if it contains a vertex $\gamma \in \pi$ such that either (i) $\gamma \in S$ and γ is a non-collider in π, or (ii) $\gamma \notin S$ and γ does not have any descendants in S, and also γ is a collider in π. Two subsets $A, B \subset V$ are said to be d-separated by S, denoted by $A \perp\!\!\!\perp B | S[G]$, if every path from a vertex in A to a vertex in B is d-separated by S.