International Journal of Approximate Reasoning 53 (2012) 1336–1349

Contents lists available at SciVerse ScienceDirect

International Journal of Approximate Reasoning

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

Characteristic imsets for learning Bayesian network structure

Raymond Hemmecke [a], Silvia Lindner [a], Milan Studený [b,*]

[a] *Zentrum Mathematik, Technische Universität München, Munich, Germany*
[b] *Institute of Information Theory and Automation of the ASCR, Prague, Czech Republic*

ARTICLE INFO

Article history:
Available online 6 May 2012

Keywords:
Learning Bayesian network structure
Essential graph
Standard imset
Characteristic imset
LP relaxation of a polytope

ABSTRACT

The motivation for the paper is the geometric approach to learning Bayesian network (BN) structure. The basic idea of our approach is to represent every BN structure by a certain uniquely determined vector so that usual scores for learning BN structure become affine functions of the vector representative. The original proposal from Studený et al. (2010) [26] was to use a special vector having integers as components, called the *standard imset*, as the representative. In this paper we introduce a new unique vector representative, called the *characteristic imset*, obtained from the standard imset by an affine transformation.

Characteristic imsets are (shown to be) zero-one vectors and have many elegant properties, suitable for intended application of linear/integer programming methods to learning BN structure. They are much closer to the graphical description; we describe a simple transition between the characteristic imset and the *essential graph*, known as a traditional unique graphical representative of the BN structure. In the end, we relate our proposal to other recent approaches which apply linear programming methods in probabilistic reasoning.

1. Introduction

Bayesian networks are basic graphical models, used widely both in statistics [13] and artificial intelligence [18]. These statistical models of *conditional independence* (CI) structure are described by acyclic directed graphs whose nodes correspond to (random) variables in consideration. It may happen that two different graphs describe the same statistical model, that is, they are *Markov equivalent*. A classic result [10,31] says that two acyclic directed graphs are Markov equivalent iff they have the same adjacencies and *immoralities*, which are special induced subgraphs over three nodes.

A quite important topic is learning *Bayesian network* (BN) *structure* [16], which is determining the statistical model on the basis of observed data. Although there are learning methods based on statistical CI tests, contemporary score and search methods are based on maximization of a suitable *quality criterion* $\mathcal{Q}$, also named a *scoring criterion* or simply a *score* by some authors. It is a real function of the (acyclic directed) graph G and the observed database D. The value $\mathcal{Q}(G, D)$ measures how well the BN structure defined by G fits the database D. Two important technical assumptions on the criterion $\mathcal{Q}$ emerged in the literature in connection with computational methods dealing with this maximization task: $\mathcal{Q}$ should be *score equivalent* [3] and (additively) *decomposable* [5].

Representing the BN structure by any of the acyclic directed graphs defining it leads to a non-unique description causing later identification problems. Thus, researchers calling for methodological simplification proposed to use a unique representative for each particular BN structure. A classic unique graphical representative is the *essential graph* [1] of the corresponding Markov equivalence class of acyclic directed graphs, which is a special graph allowing both directed and undirected edges.

* Corresponding author.
E-mail addresses: hemmecke@ma.tum.de (R. Hemmecke), slindner@ma.tum.de (S. Lindner), studeny@utia.cas.cz (M. Studený).

0888-613X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ijar.2012.04.001

The basic idea of an algebraic approach to the description of CI structures [22] is to represent them by certain vectors with integer components, called *imsets*. In the context of learning Bayesian networks this led to the proposal to represent each BN structure uniquely by a so-called *standard imset*. The advantage of this algebraic approach is that every score equivalent and decomposable quality criterion becomes an affine function (= linear function plus a constant) of the standard imset (see Chapter 8 in [22]).

A geometric view was offered in [26], where it was shown that the standard imsets over a fixed set of variables N are vertices (= extreme points) of a certain polytope, called the *standard imset polytope*. These results allow one to use the tools of polyhedral geometry in the area of learning Bayesian nets, because they transform the learning task to a *linear programming* (LP) problem [19], namely to optimize a linear function over a bounded polyhedron.

In this paper, we propose an alternative vector representative of the BN structure, called the *characteristic imset*. It is a vector obtained from the standard imset by a one-to-one affine transformation which maps integer vectors to integer vectors (in both directions). Thus, every score equivalent and decomposable criterion is an affine function of the characteristic imset and the set of characteristic imsets is the set of vertices of a polytope, called the *characteristic imset polytope*. Every characteristic imset is a zero-one vector, that is, it has only zeros and ones as its components. Moreover, it is very close to the graphical description: both adjacencies and immoralities are encoded by certain components of the characteristic imset (see Corollary 2 for details).

We establish a simple relation of the characteristic imset to any acyclic directed graph defining the BN structure and to the respective essential graph as well. More specifically, we provide a formula for the characteristic imset on basis of any chain graph defining the BN structure which has no *flag*, which is a special induced subgraph over three nodes. In particular, this makes it possible to get the characteristic imset immediately on the basis of the essential graph. We also consider the converse task of reconstructing the essential graph from the characteristic imset and provide a polynomial algorithm (in the number $|N|$ of variables) for it.

If we restrict our attention to *decomposable models* [13], interpreted as BN structures, then the characteristic imset has quite a simple form. The situation is particularly transparent in the case of (models induced by) *undirected forests*: then the edges in the graph correspond to ones in the characteristic imset. Thus, one can use the well-known *greedy algorithm* [20] to learn these special graphical models; this gives an elegant geometric interpretation to the classic heuristic procedure proposed by Chow and Liu [6].

The structure of the paper, which is based on [27], is as follows. In Section 2 we recall some of the definitions and relevant results. In Section 3 we introduce the characteristic imset and derive the above mentioned observations on it. Section 4 is devoted to the transition between the essential graph and the characteristic imset. Section 5 contains comments on learning decomposable models. In Section 6 we relate characteristic imsets to other zero-one vector structure representatives which have recently appeared in the literature [17,11]. We also discuss there the idea of intended future application of this approach to practical learning, motivated by [11,7,9]. In Section 7 we briefly mention our preliminary computational experiments, and, in Conclusions we discuss the perspectives.

2. Basic concepts

We tacitly assume that the reader is familiar with basic concepts from polyhedral geometry. Throughout the paper N is a finite non-empty set of *variables*; to avoid the trivial case we assume $|N| \geq 2$. In statistical context, the elements of N correspond to random variables in consideration; in graphical context, they correspond to nodes.

2.1. Graphical concepts

Graphs considered here have a finite non-empty set of nodes N and two types of edges: directed edges, called *arrows*, denoted like $i \rightarrow j$ or $j \leftarrow i$, and undirected edges. No loops or multiple edges are allowed between two nodes. If there is an edge between nodes i and j, we say they are *adjacent*.

Given a graph G over N and a non-empty set of nodes $A \subseteq N$, the *induced subgraph* of G for A has just those edges in G having both end-nodes in A. An *immorality* in G is an induced subgraph (of G) for three nodes $\{a, b, c\}$ in which $a \rightarrow c \leftarrow b$ and a and b are not adjacent. A *flag* is another induced subgraph for $\{a, b, c\}$ in which $a \rightarrow b$, b and c are adjacent by an undirected edge and a and c are not adjacent.

A set of nodes $K \subseteq N$ is *complete* in G if every pair of distinct nodes in K is adjacent by an undirected edge. To avoid confusion note that some authors [13] may use this term to name a set in which every pair of distinct nodes is adjacent, no matter whether by a directed or an undirected edge. However, in this paper we do need the stronger concept of (an undirected) complete set. A maximal complete set is called a *clique*.

A set $C \subseteq N$ is *connected* if every pair of distinct nodes in C is connected via an undirected path. Maximally connected sets are called (undirected) *components*. Of course, the components of G provide a natural partition of N.

A graph is *directed* if all its edges are arrows. A directed graph G over N is called *acyclic* if there exists an ordering $b_1, \ldots, b_{|N|}$ of all its nodes which is consistent with the direction of arrows, that is, $b_i \rightarrow b_j$ in G implies $i < j$.

A graph is *undirected* if all its edges are undirected. An undirected graph is called *chordal*, or *decomposable*, if every (undirected) cycle of the length at least four has a chord, that is, an edge connecting two non-consecutive nodes in the cycle.