International Journal of Approximate Reasoning 55 (2014) 1043–1071

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Learning Bayesian network structure: Towards the essential graph by integer linear programming tools

Milan Studený [a,*], David Haws [b]

[a] Institute of Information Theory and Automation of the ASCR, Prague, Czech Republic
[b] IBM T.J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, NY 10598, USA

ARTICLE INFO

Article history:
Available online 8 October 2013

Keywords:
Learning Bayesian network structure
Integer linear programming
Characteristic imset
Essential graph

ABSTRACT

The basic idea of the geometric approach to learning a Bayesian network (BN) structure is to represent every BN structure by a certain vector. If the vector representative is chosen properly, it allows one to re-formulate the task of finding the global maximum of a score over BN structures as an integer linear programming (ILP) problem. Such a suitable zero-one vector representative is the *characteristic imset*, introduced by Studený, Hemmecke and Lindner in 2010, in the proceedings of the 5th PGM workshop. In this paper, extensions of characteristic imsets are considered which additionally encode chain graphs without flags equivalent to acyclic directed graphs. The main contribution is a polyhedral description of the respective domain of the ILP problem, that is, by means of a set of linear inequalities. This theoretical result opens the way to the application of ILP software packages. The advantage of our approach is that, as a by-product of the ILP optimization procedure, one may get the *essential graph*, which is a traditional graphical BN representative. We also describe some computational experiments based on this idea.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Learning a *Bayesian network* (BN) structure is the statistical task of model choice, where the candidate statistical structural models are ascribed to acyclic directed graphs. A score-and-search approach to this learning task consists in maximization of a *quality criterion* $\mathcal{Q}$, also called a *score* or a *scoring function*, which is a real function of the (acyclic directed) graph G and the observed database D. The value $\mathcal{Q}(G, D)$ says how much the BN structure defined by the graph G, that is, the statistical model ascribed to G, fits the database D. Note that some researchers in machine learning are accustomed to identify a BN structure with the respective equivalence class of graphs and prefer to talk about learning *an equivalence class* of Bayesian networks.

Two important technical assumptions on the criterion $\mathcal{Q}$ were pinpointed in the literature in connection with computational aspects of the above-mentioned maximization task. Since the goal is to learn a BN structure, $\mathcal{Q}$ should be *score equivalent* [3], which means it ascribes the same value to equivalent graphs, that is, to graphs defining the same BN structure. The other assumption is that $\mathcal{Q}$ is *decomposable*, which means $\mathcal{Q}(G, D)$ decomposes into contributions which correspond to factors in the factorization according to the graph G. Typically, it is required that $\mathcal{Q}$ is additively decomposable [5], that is, $\mathcal{Q}(G, D)$ is the sum of such local contributions, called *local scores*.

The geometric approach is to represent every BN structure by a certain vector so that such a criterion $\mathcal{Q}$ becomes an affine function of the vector representative, that is, a linear function plus a constant term. This idea was introduced by Studený already in 2005 [22] and then deepened by Studený, Vomlel and Hemmecke in 2010 [23]. A suitable (uniquely

* Corresponding author.
E-mail addresses: studeny@utia.cas.cz (M. Studený), dchaws@gmail.com (D. Haws).

0888-613X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ijar.2013.09.016

determined) zero-one vector BN representative seems to be the *characteristic imset*, introduced recently by Studený, Hemmecke and Lindner [24,12].

Jaakkola et al. [13] and Cussens [6,7] came independently with an analogous geometric approach. The main difference is that they used certain special zero-one vector codes of (acyclic) directed graphs to represent (non-uniquely) BN structures. On the other hand, they made much more progress with the practical application of *integer linear programming* (ILP) tools. To overcome technical problems with the exponential length of their vectors they utilized the idea of the reduction of the search space developed by de Campos et al. [9,10], based on a particular form of databases and quality criteria occurring in practice.

We have compared [28] both methods of BN structure vector representation and found that the characteristic imset can be viewed as a (many-to-one) linear function of the above mentioned zero-one codes of directed graphs. We also found [28] that by transforming the polyhedral approximation used by Jaakkola et al. [13] and Cussens [7] one also gets a polyhedral approximation for the characteristic imset polytope. However, an unpleasant finding was that, while there is a one-to-one correspondence for their acyclicity-encoding inequalities, their basic non-negativity and equality constraints are transformed to much higher number of inequalities. Thus, direct transformation of their inequalities does not lead to a suitable ILP problem formulation in terms of characteristic imsets.

Lindner [17] in her thesis dealt with the problem of finding a workable LP relaxation of the characteristic imset polytope, that is, an outer polyhedral approximation such that the only vectors with integer components satisfying the considered inequalities are the characteristic imsets. To overcome/avoid this problem she used the idea of extending the characteristic imset with additional components, which is a useful standard trick in combinatorial optimization. Her additional components allowed her to encode acyclic directed graphs inducing the characteristic imset. She also reported on some computational experiments based her approach.

This paper is another step on the way to develop a consistent ILP approach based on characteristic imsets. Being inspired by Lindner [17], we introduce an extended vector BN structure representative which includes the characteristic imset and, moreover, encodes a certain special graph (equivalent to an acyclic directed graph). The main theoretical result is a *polyhedral characterization* of the domain of the respective ILP problem. Note that the objective in this ILP problem only depends on the characteristic imset and the additional components play an auxiliary role; this is different from the approach based on direct encoding of graphs [13,7].

More specifically, a set of linear inequalities is presented such that the only vectors with integer components in the polyhedron specified by those inequalities are the above mentioned extended characteristic imsets. The presented inequalities are classified in four groups. The number of inequalities in the first two groups is polynomial in the number of variables, that is, in the number of nodes of the graph, while the number of remaining inequalities is exponential. However, provided that the length of the vector representatives is limited/reduced to a quasi-polynomial number by the idea presented by de Campos and Ji [10], the number of inequalities in the third group can be reduced to a quasi-polynomial number as well. The inequalities in the fourth group correspond to acyclicity restrictions. In general, they cannot be reduced to a polynomial number, but the method of iterative constraint adding may be applied to solve the respective ILP problem. The overall number of our inequalities is lower than Lindner [17] provided.

The main advantage of our approach is that once an optimal solution is found one can use the obtained (extended) characteristic imset to get the corresponding *essential graph*, which is known as a standard (unique) graphical BN representative [2]. This graph can be obtained as the solution of a secondary ILP problem: it has another objective but shares with the main ILP problem the first two groups of inequalities. Thus, the second ILP problem has both polynomial length of vectors and polynomial number of inequalities.

We have also performed some computational experiments whose aim was to verify the feasibility of the approach. This was confirmed but the experiments also indicated that the number of our inequalities is too high to be able to compete with the running times presented at GOBNILP web page [32]. On the other hand, the experiments confirmed the assumption that the second ILP problem, used to get the essential graph, is easily solvable.

In Section 2 we recall basic concepts and some special results on chain graphs we use later in the paper. In Section 3 more details on the ILP approach to learning BN structure are given; specifically, we describe how to compute the objective in the characteristic imsets case (from local scores) and how to utilize the search space reduction [10] in the context of characteristic imsets. Section 4 describes the theoretical basis of our specific ILP approach; we say which graphs are encoded in our extended characteristic imsets, list/comment the inequalities and formulate the main results. In Section 5 we describe both phases to learn the optimal essential graph. Section 6 deals with two specific methods to solve the main ILP problem that we tested in our computational experiments, described then in Section 7. In Conclusions we comment the results and discuss further perspectives. Appendix A contains the proofs.

2. Basic concepts

Let N be a finite non-empty set of *variables*; to avoid the trivial case, assume $|N| \geqslant 2$. In statistical context, the elements of N correspond to random variables in consideration; in graphical context, they correspond to nodes.

Download English Version:

https://daneshyari.com/en/article/397896

Download Persian Version:

https://daneshyari.com/article/397896

Daneshyari.com