International Journal of Approximate Reasoning 51 (2010) 573–586

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

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

A geometric view on learning Bayesian network structures

Milan Studený [a,*], Jiří Vomlel [a], Raymond Hemmecke [b]

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

ARTICLE INFO

Article history:
Received 27 November 2008
Accepted 20 April 2009
Available online 28 January 2010

Keywords:
Learning Bayesian networks
Standard imset
Inclusion neighborhood
Geometric neighborhood
GES algorithm

ABSTRACT

We recall the basic idea of an algebraic approach to learning Bayesian network (BN) structures, namely to represent every BN structure by a certain (uniquely determined) vector, called a *standard imset*. The main result of the paper is that the set of standard imsets is the set of vertices (=extreme points) of a certain polytope. Motivated by the geometric view, we introduce the concept of the *geometric neighborhood* for standard imsets, and, consequently, for BN structures. Then we show that it always includes the *inclusion neighborhood*, which was introduced earlier in connection with the greedy equivalence search (GES) algorithm. The third result is that the global optimum of an affine function over the polytope coincides with the local optimum relative to the geometric neighborhood.

To illustrate the new concept by an example, we describe the geometric neighborhood in the case of three variables and show it differs from the inclusion neighborhood. This leads to a simple example of the failure of the GES algorithm if data are not "generated" from a perfectly Markovian distribution. The point is that one can avoid this failure if the search technique is based on the geometric neighborhood instead. We also found out what is the geometric neighborhood in the case of four and five variables.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

The motivation for this theoretical paper is learning a Bayesian network (BN) structure from data by the method of maximizing a quality criterion (=the score and search method). By a *quality criterion*, also named a *score metric* or simply a *score* by some other authors, we mean a real function $\mathcal{Q}$ of the BN structure, usually represented by a graph G, and of the database D. The value $\mathcal{Q}(G,D)$ "evaluates" how the BN structure given by G fits the observed database D.

An important related question is how to represent a BN structure in the memory of a computer. Formerly, each BN structure was represented by an arbitrary acyclic directed graph defining it, which led to the non-uniqueness in its description. Later, researchers calling for methodological simplification came up with the idea to represent every BN structure with a unique representative. The most popular graphical representative is the *essential graph*. It is a chain graph describing shared features of acyclic directed graphs defining the BN structure. The adjective "essential" was proposed by Andersson et al. [2], who gave a graphical characterization of (graphs that are) essential graphs.

Since direct maximizing a quality criterion $\mathcal{Q}$ seems, at least at first sight, to be infeasible, various *local search methods* have been proposed. The basic idea is that one introduces a neighborhood relation between BN structure representatives, also named *neighborhood structure* by some authors [3]. Then one is trying to find a local maximum with respect to the chosen neighborhood structure. This is an algorithmically simpler task because one can utilize various greedy search techniques for this purpose. On the other hand, the algorithm can get stuck in a local maximum and fail to find the global maximum. A

* Corresponding author.
E-mail address: studeny@utia.cas.cz (M. Studený).

0888-613X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2010.01.014

typical example of these techniques is the *greedy equivalence search* (GES) algorithm proposed by Meek [12]. The neighborhood structure utilized in this algorithm is the *inclusion neighborhood*, which comes from the conditional independence interpretation of BN structures. Chickering [5] proposed a modification of the GES algorithm, in which he used essential graphs as (unique) BN structure representatives.

There are two important technical requirements on a quality criterion $\mathcal{Q}$ brought in connection with the local search methods, namely to make them computationally feasible. One of them is that $\mathcal{Q}$ should be *score equivalent* [3], which means it ascribes the same value to equivalent graphs (=defining the same BN structure, that is, having ascribed the same essential graph). The other requirement is that $\mathcal{Q}$ should be *decomposable* [5], which means that $\mathcal{Q}(G,D)$ decomposes into contributions which correspond to the factors in the factorization according to the graph G.

The basic idea of an algebraic approach to learning BN structures, presented in Chapter 8 of [18], is to represent both the BN structure and the database with real vectors. More specifically, an algebraic representative of the BN structure defined by an acyclic directed graph G is a certain integer-valued vector u_G, called the *standard imset* (for G). It is also a unique BN structure representative because $u_G = u_H$ for equivalent graphs G and H. Another advantage of standard imsets is that one can read practically immediately from the differential vector $u_G - u_H$ whether the BN structures defined by graphs G and H are neighbors in the sense of inclusion neighborhood. However, the crucial point is that every score equivalent and decomposable criterion $\mathcal{Q}$ is an affine function (=linear function plus a constant) of the standard imset. More specifically, it is shown in Section 8.4.2 of [18] that one has

$$\mathcal{Q}(G,D) = s_D^{\mathcal{Q}} - \langle t_D^{\mathcal{Q}}, u_G \rangle,$$

where $s_D^{\mathcal{Q}}$ is a real number, $t_D^{\mathcal{Q}}$ a vector of the same dimension as the standard imset u_G (these parameters both depend solely on the database D and the criterion $\mathcal{Q}$) and $\langle *, * \rangle$ denotes the scalar product. The vector $t_D^{\mathcal{Q}}$ is named the *data vector* (relative to the criterion $\mathcal{Q}$).

We believe that the above-mentioned result paves the way for future application of efficient linear programming methods in the area of learning BN structures. This paper is a further step in this direction: its aim is to enrich the algebraic approach by a geometric view. One can imagine the set of all standard imsets over a fixed *set N of variables* (=the set of nodes for graphs) as the set of points in the corresponding Euclidean space (of a higher dimension). The main result of the paper is that it is the set of vertices (=extreme points) of a certain polytope P. We derive this geometric fact from former theoretical results on BNs. A consequence of the result is as follows: since every "reasonable" quality criterion $\mathcal{Q}$ can be viewed as (the restriction of) an affine function on the corresponding Euclidean space, the task to maximize $\mathcal{Q}$ over BN structures is equivalent to the task to maximize an affine function over the above-mentioned polytope.

This maximization problem has been treated thoroughly within the linear programming community. Deep algorithmic theory was developed and fast software codes are available that can handle problems with vectors having thousands or even millions of components. A classic tool to solve linear programming problems is the *simplex method* [15]. One of possible interpretations of this method is that it is a kind of an augmentation algorithm (=a search method), in which one moves between vertices of a polytope along its edges (in the geometric sense) until an optimal vertex is reached. Although it has not yet been decided whether the simplex method can be modified to get a polynomial-time algorithm,[1] it performs extremely well in practice.

In order to apply the (classic) simplex method, one needs an explicit description of the polytope via finitely many linear inequalities. Such a description always exists by Weyl–Minkowski theorem [15], which says that any polytope can equivalently be introduced as a (bounded) polyhedron, that is, the intersection of finitely many (affine) half-spaces. Note that the implicit knowledge about these inequalities is often enough to solve a linear program at hand. As concerns the standard imset polytope P, for $|N| = 3$ and $|N| = 4$ a minimal such system has 13 and 154 inequalities, respectively. However, it is already a challenge to existing software packages to find such a minimal inequality description of P for $|N| = 5$ (given by 8782 vertices). Thus, for general $|N|$, one definitely needs to classify these inequalities implicitly in order to apply the classic tools from linear programming.

Because such a "polyhedral" description of the polytope P is not available for arbitrarily high $|N|$ we propose an alternative approach that mimics the walk along the edges in the simplex method. The basic idea is to introduce the concept of *geometric neighborhood* for standard imsets, and, therefore, for BN structures as well. The standard imsets u_G and u_H will be regarded as (geometric) neighbors if the line-segment connecting them is an edge of the polytope P in the geometric sense. An important observation is that the above-mentioned inclusion neighborhood is always contained in the geometric one. Nevertheless, the crucial fact is that, for any affine function, its local maximum relative to the geometric neighborhood is necessarily its global maximum over the polytope. We give the proof of both these observations in the paper.

Thus, once one succeeds in characterizing explicitly or implicitly the geometric neighborhood structure, one can apply the following augmentation algorithm: start at a standard imset (=a vertex of P), for example $u_G = 0$, and keep moving to (geometrically) adjacent standard imsets (=via edges of P) with a higher value of the criterion $\mathcal{Q}$ until one reaches its local maximum relative to the geometric neighborhood. Thus, by the above-mentioned observation, the *global maximum* of $\mathcal{Q}$ over the standard imsets must be found. Note that the resulting standard imset can then be transformed to the corresponding essential graph by a polynomial-time algorithm described in [19].

[1] This is a long-standing open question in linear programming.

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

✓ 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