Integer linear programming approach to learning Bayesian network structure: towards the essential graph

Milan Studený (Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic)

Abstract The basic idea is to represent every Bayesian network (BN) structure by a certain vector. This 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. Suitable such a zero-one vector representative is the *characteristic imset*, introduced in [1]. Here, extensions of characteristic imsets are considered which additionally encode chain graphs without flags equivalent to acyclic directed graphs. The domain of the respective ILP problem is specified in terms of linear inequalities. The advantage of this particular approach is that, as a by-product, one may get the *essential graph*, which is a traditional graphical BN representative.

Preliminaries

Bayesian network structure

N, $2 \leq |N| < \infty$ a set of *variables*

X_i, $2 \leq |\mathsf{X}_i| < \infty$ for each $i \in N$ individual *sample spaces*

Bayesian network structure (= BN structure) defined by an acyclic directed graph G (over N) is the class of probability distributions on $\prod_{i\in N} \mathsf{X}_i$ that are Markovian with respect to G.

Different graphs could be *Markov equivalent* (= define the same BN structure).

Quality criteria for learning

A *database* D is a sequence $x_1, \ldots, x_\ell$, $\ell \geq 1$ of elements in $\prod_{i\in N} \mathsf{X}_i$.

A *quality criterion* is a real function $\mathcal{Q}$ of two variables: of an acyclic directed graph G and of a database D.

It is *score equivalent* if, for any fixed D, $\mathcal{Q}(G,D) = \mathcal{Q}(H,D)$, for any pair of Markov equivalent graphs G and H.

It is additively *decomposable* if $\mathcal{Q}(G,D) = \sum_{i\in N} q_D(i|\mathrm{pa}_G(i))$, where $\mathrm{pa}_G(i)$ denotes the set of *parents* of $i \in N$ in G.

Examples

Schwarz's *Bayesian information criterion* (BIC)

Bayesian Dirichlet Equivalence (BDE) score

Characteristic imset

Zero-one vector with components indexed by sets in $\mathcal{P}_2(N) \equiv \{S \subseteq N; |S| \geq 2\}$.

The *characteristic imset* for an acyclic directed graph G over N:

$$c_G(S) = 1 \quad \text{iff} \quad \exists\, i \in S \text{ such that } S \setminus \{i\} \subseteq \mathrm{pa}_G(i).$$

One has $c_G = c_H$ iff G and H are Markov equivalent. Every score equivalent and decomposable criterion $\mathcal{Q}$ has the form

$$\mathcal{Q}(G,D) \,=\, \mathcal{Q}(G^{\emptyset},D) + \sum_{S\subseteq N,\, 2\leq|S|} r_D^{\mathcal{Q}}(S)\cdot c_G(S)\,,$$

Here, $r_D^{\mathcal{Q}} \in \mathbb{R}^{\mathcal{P}_2(N)}$ is the *revised data vector*.

Essential graph

The *essential graph* G^* of a Markov equivalence class $\mathcal{G}$ is a hybrid graph with the adjacencies and arrows shared throughout $\mathcal{G}$.

Lemma 1 Let $\mathcal{G}$ be an equivalence class of acyclic directed graphs over N and $\mathcal{H}$ an equivalence class of chain graphs without flags such that $\mathcal{G} \subseteq \mathcal{H}$. Then G^* is the largest graph in $\mathcal{H}$.

Extended BN vector representative

Definition We encode a hybrid graph H over N by a zero-one vector (a_H, c_H) given as follows: for distinct $i,j \in N$ and $S \subseteq N$, $|S| \geq 2$,

$$a_H(i\to j) = 1 \quad \Longleftrightarrow \quad i \to j \text{ in } H,$$

$$c_H(S) = 1 \quad \Longleftrightarrow \quad H_S \text{ has a super-terminal component},$$

which means, there exists $\emptyset \neq K \subseteq S$ undirected clique in H_S such that $\forall\, j \in S\setminus K$, $\forall\, i \in K$ one has $j \to i$ in H.

The list of inequalities

The *basic non-negativity inequalities* are:

(b.1) $\forall\, i,j \in N$ distinct $\quad 0 \leq a(i\to j)$,

(b.2) $\forall\, S \subseteq N$, $|S| = 3,4 \quad 0 \leq c(S)$.

The *consistency inequalities* mainly relate the a-part to the c-part:

(c.1) $a(i\to j) + a(j\to i) \leq c(ij)$,

(c.2) $c(ij) \leq 1$,

(c.3) $2\cdot c(ijk) \leq 2\cdot c(ij) + a(i\to k) + a(j\to k)$,

(c.4) $a(i\to j) + c(jk) \leq 1 + c(ijk) + a(j\to k)$,

(c.5) $a(i\to j) + c(jk) + c(ik) \leq 2 + a(i\to k) + a(k\to j)$.

The *extension inequalities* determine uniquely $c(S)$ for $|S| \geq 4$:

(e.1) $\forall\, S \subseteq N$, $|S| \geq 3 \quad \sum_{i\in S} c(S\setminus\{i\}) \leq 2 + (|S|-2)\cdot c(S)$,

(e.2) $\forall\, S \subseteq N$, $|S| \geq 4 \quad (|S|-1)\cdot c(S) \leq \sum_{i\in S} c(S\setminus\{i\})$.

The *acyclicity inequalities* are then as follows:

(a.1) $\forall\, S \subseteq N$, $|S| \geq 4 \quad \sum_{T\subseteq S,\, 2\leq|T|} c(T)\cdot(-1)^{|T|} \leq |S|-1$.

LP relaxation

Theorem 1 A vector (a,c) with integer components satisfies (b.1)-(b.2), (c.1)-(c.5), (e.1)-(e.2) and (a.1) iff there exists (uniquely determined) chain graph H without flags equivalent to an acyclic directed graph over N such that $(a,c) = (a_H, c_H)$. (the proof is in [2])

Theorem 1 also holds with (a.1) replaced by a simplified version:

(a.1*) $\forall\, S \subseteq N$, $|S| \geq 4 \quad \sum_{T\subseteq S,\, 2\leq|T|\leq 3} c(T)\cdot(-1)^{|T|} \leq |S|-1\,.$

Summary of the whole procedure

- A pre-processing step is *pruning*, whose result should be values $r_D^{\mathcal{Q}}(S)$ for $S \in \mathcal{T}$, where $\mathcal{T} \subseteq \{S \subseteq N;\ |S| \geq 2\}$ is a "small" class of sets closed under subsets.
- The first ILP problem to be solved is to maximize the function

$$(a,c) \mapsto \sum_{S\in\mathcal{T}} r_D^{\mathcal{Q}}(S)\cdot c(S)$$

  over the domain of vectors with integral components specified by (b.1)-(b.2),(c.1)-(c.5),(e.1)-(e.2) and (a.1). As concerns (a.1), the method of *iterative constraint adding* can be applied.
- The second ILP problem is, with fixed c, to minimize

$$(a,c) \mapsto \sum_{i,j\in N, i\neq j} a(i\to j)$$

  under the same constraints. By Lemma 1, the solution encodes the *essential graph* G^* corresponding to c.

References [1] M.Studený, R.Hemmecke, S.Lindner: Characteristic imset: a simple algebraic representative of a BN structure. In *5th PGM* (2010), 257-264.

[2] M.Studený: LP relaxations and pruning for characteristic imsets. Research report n. 2323, ÚTIA, Prague (2012), 30 pages.

Acknowledgement This research has been supported by the grants GAČR n. 201/08/0539.