ESSENTIAL GRAPH AS VERY EFFICIENT TOOLS

GARRIDO ANGEL

Facultad de Ciencias UNED, Madrid, E-mail: agarrido@mat.uned.es

Abstract: Bayesian Nets (BNs, in abridged expression) represent joint probability distributions (JPDs). We says that two BNs are equivalent (denoted by X), if both represent the same JPD. Now, we can see some useful characterizations of the equivalence among BNs, and the usefulness of Essential Graphs (EGs) as representatives of equivalence classes of BNs, in AI; for instance, learning BNs.

Keywords: A. I., Knowledge Representation, Graph Theory, Bayesian Nets

1. INTRODUCTION

Let S and S´ two of such structures of BNs on V.

Then, we says that *S is equivalent to S´:* S X S´, for each parameterization, θ, of S, there exist another parameterization, θ´, of S´, such that:

$$P\left(V/_{S,\,\vartheta}\right) = P\left(V/_{S',\,\vartheta'}\right) \tag{1}$$

Therefore, S can represent every probability distribution which S´ represent and vice versa.

It verifies the following properties:

- Reflexive: B X B, ∀ B
- Symmetrical: if B X B´⇒ B´X B
- Transitive: if B X B´ and B´X B´´⇒ B X B´´

Therefore, it is an Equality or Equivalence Relation**,** defined on the BNs set. On such mathematical object, it is well established a partition in equivalence classes.

Th. S X S ⇔ both structures induce the same set of conditional independencies (according the Global Markov Property).

In a DAG, if we eliminate its directions to each directed edge, it remains their skeleton graph.

On a DAG, we says an immorality to each configuration of this type: $X \to Y \leftarrow Z$. Where we can observe the following directed edges: $X \to Z \text{ and } Z \leftarrow Y$

2. RESULTS

Th. Two models of BNs are X $\Leftrightarrow$ both have the same skeleton and the same immoralities.

But to have the same skeleton, it must to be equal its order (number of vertices-nodes) and the size (number of edges) among the considered graphs.

Th. Two models of BNs are X $\Leftrightarrow$ there exist a sequence of covered edge inversions, transforming one in another model.
We can denote the equivalence class of S by $[S]$.
It induces a partition into the set of BNs, B, in equivalence classes: $\Omega = B/X = \cup Bi$.

Let C be a class of DAGs Markov X among them. Then, their essential graph would be the lesser graph greater than every DAG that belongs to the class.

If we denote the essential graph as G*, this will be equivalent to say: $G^* = \cup\{G: G \in C\}$. Where such graph union is reached by the union of the nodes and edges of G:

$$V(G^*) = \cup\ V(G),\ \ E(G^*) = \cup\ E(G) \tag{2}$$

The directed edges connecting the same pair of nodes, but showing opposed directions, into two graphs belonging the same class, C, are substituted by a line.

So, G* will be the lesser of the upper bounds for every graph of the class represented.

Equivalence-Invariant Scoring Function: they are the scoring functions that given the same score value, when they are applied on equivalent models

Some functions permits to find the score of equivalence classes. But it is not so in every case.

Two models X of BNs ever have the same computational complexity. Because the inversion of essential edges no change the complexity class.

Given two BN models de RBs: S and S´, we said that S include to S´, if every assumptions of c. i. which are true for S´ they are also true for S.

Relations between Equivalence and Inclusion of BNs:

$$S\ X\ S' \Leftrightarrow (S \subset S') \wedge (S' \subset S) \tag{3}$$

And we said that S´ is strictly included into S, if S include to S´, but S´ is not included in S.

Th. A structure of BN, S, include to another, S´$\Leftrightarrow$ there exist a sequence of covered arc inversions and additions of arcs which transforms S in S´.

Cor.: A structure of BN, S, include to another, S´ $\Leftrightarrow$ S is able to represent any joint probability distribution that S´ can represent.

For acyclic directed graphs, G, of order three: O (G) = 3. Their size, T, depends of the class to which belongs, being $T(G) \in \{0, 1, 2, 3\}$.

In general, the number of possible structures, for BNs with **n** nodes, r (n), is given by the recurrence equation:

$$r(n) = \sum_{i=1}^{n} (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} r(n-i) \quad (4)$$

In the first case, to order n = 3:

$$\begin{aligned} r(3) &= \sum_{i=1}^{3} (-1)^{i+1} \binom{3}{i} 2^{i(3-i)} r(3-i) = \\ &= (-1)^2 \binom{3}{1} 2^{1(3-1)} r(3-1) + (-1)^3 \binom{3}{2} 2^{2(3-2)} r(3-2) + \\ &+ (-1)^4 \binom{3}{3} 2^{3(3-3)} r(3-3) = 36 - 12 + 1 = 25 \\ &(\neq estructures),\ because: r(0) = r(1) = 1\ \&\ r(2) = 3. \end{aligned} \quad (5)$$

When the graphs are of order: O(G) = 4, the size can be: T (G)$\in\{0,1, 2, 3, 4, 5, 6\}$. Applying the aforementioned recurrent equation, we obtain 543 different possible configurations:

$$\begin{aligned} r(4) &= \sum_{i=1}^{4} (-1)^{i+1} \binom{4}{i} 2^{i(4-i)} r(4-i) = \\ &= (-1)^2 \binom{4}{1} 2^{1(4-1)} r(4-1) + (-1)^3 \binom{4}{2} 2^{2(4-2)} r(4-2) + \\ &+ (-1)^4 \binom{4}{3} 2^{3(4-3)} r(4-3) + (-1)^5 \binom{4}{4} 2^{4(4-4)} r(4-4) = \\ &= 800 - 288 + 32 - 1 = 543 \quad different\ configurations \end{aligned} \quad (6)$$

Let S and P be a DAG and a JPD, respectively. Suppose which join to a S an edge X→Y not produce cycles. Then, we said that joining an edge to S is useless, if: X ⊥ Y / Pas (Y). In other case, we said that joining an edge to S is useful. In both cases, it is convenient to add: with respect to the probability distribution, P (w. r. t. P).

A scoring function is locally consistent, if adding a useful edge increase the score and adding a useless edge decrease the score.

As different DAGs can determine the same class of X of Markov, is of great interest know how much to improve their efficiency when we select for each class only one representative, instead of consider through exhaustive procedure each one of the DAGs.

For their computation, it was elaborated a program, due to Gillispie & Perlman´01, which permits the enumeration of the equivalence (X) of DAGs, according the equivalence criteria among BNs.

With her was computed the proportion of DAGs to classes of X is (asymptotically) of 3.75. That is, from classes to DAGs should be of 0.267. It is certainly a considerable reduction (to the fourth part), perhaps lesser tan the supposed (so, De Campos´06).

BIBLIOGRAPHY

[1] Anderson, Madigan and Perlman: “A characterization of Markov equivalence classe for acyclic digraphs”. Ann.Stat. 25, N. 2, p. 505-541, 1997.
[2] Chickering: “A transformational characterization of equivalent Bayesian Networks structures”. UAI´1995, p. 87-98.
[3] Ibid.: “Learning Equivalence Classes of Bayesian Network Structures”. JMLR, 2, p. 445-498, 2002.
[4] Gillispie and Perlman: “Enumerating Markov Equivalence Classes of Acyclic Digraphs”, UAI´2001, p. 171-177.

[5] Roverato and Studený: "A graphical representation of equivalence classes of AMP chain graphs", JMLR 7, p. 1045-1078, 2006.
[6] Studený: "Characterization of essential graphs by means of the operation of legal merging of components", Int. J. of Unc., Fuzziness and Knowledge-Based Systems 12, p. 43-62, 2004.
[7] Volf and Studený: "Using imsets for Learning Bayesian Networks", Proc. 10th Czech-Japan Seminar on Data Analysis, p. 178-189, 2007.
[8] Ibid.: "Graphical and Algebraic Representatives of Conditional Independence Models". Advances in PGMs, p. 55-80, 2007.