International Journal of Approximate Reasoning 68 (2016) 108–126

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Learning marginal AMP chain graphs under faithfulness revisited

Jose M. Peña [a,*], Manuel Gómez-Olmedo [b]

[a] ADIT, IDA, Linköping University, Sweden
[b] Dept. Computer Science and Artificial Intelligence, University of Granada, Spain

ARTICLE INFO

Article history:
Received 19 December 2014
Received in revised form 9 August 2015
Accepted 29 September 2015
Available online 9 October 2015

Keywords:
Chain graphs
AMP chain graphs
MVR chain graphs
Structure learning

ABSTRACT

Marginal AMP chain graphs are a recently introduced family of models that is based on graphs that may have undirected, directed and bidirected edges. They unify and generalize the AMP and the multivariate regression interpretations of chain graphs. In this paper, we present a constraint based algorithm for learning a marginal AMP chain graph from a probability distribution which is faithful to it. We show that the marginal AMP chain graph returned by our algorithm is a distinguished member of its Markov equivalence class. We also show that our algorithm performs well in practice. Finally, we show that the extension of Meek's conjecture to marginal AMP chain graphs does not hold, which compromises the development of efficient and correct score+search learning algorithms under assumptions weaker than faithfulness.

1. Introduction

Chain graphs (CGs) are graphs with possibly directed and undirected edges, and no semidirected cycle. They have been extensively studied as a formalism to represent independence models, because they can model symmetric and asymmetric relationships between the random variables of interest. However, there are three different interpretations of CGs as independence models: The Lauritzen–Wermuth–Frydenberg (LWF) interpretation [11], the multivariate regression (MVR) interpretation [8], and the Andersson–Madigan–Perlman (AMP) interpretation [2]. It is worth mentioning that no interpretation subsumes another: There are many independence models that can be represented by a CG under one interpretation but that cannot be represented by any CG under the other interpretations [2,24]. Moreover, although MVR CGs were originally represented using dashed directed and undirected edges, we like other authors prefer to represent them using solid directed and bidirected edges.

Recently, a new family of models has been proposed to unify and generalize the AMP and MVR interpretations of CGs [17]. This new family, named marginal AMP (MAMP) CGs, is based on graphs that may have undirected, directed and bidirected edges. This paper complements that by Peña [17] by presenting an algorithm for learning a MAMP CG from a probability distribution which is faithful to it. Our algorithm is constraint based and builds upon those developed by Sonntag and Peña [23] and Peña [16] for learning, respectively, MVR and AMP CGs under the faithfulness assumption. It is worth mentioning that there also exist algorithms for learning LWF CGs under the faithfulness assumption [12,27] and under the milder composition property assumption [19]. In this paper, we also show that the extension of Meek's conjecture to MAMP

* Corresponding author.
E-mail addresses: jose.m.pena@liu.se (J.M. Peña), mgomez@decsai.ugr.es (M. Gómez-Olmedo).

http://dx.doi.org/10.1016/j.ijar.2015.09.004
0888-613X/

CGs does not hold, which compromises the development of efficient and correct score+search learning algorithms under assumptions weaker than faithfulness.

Finally, we should mention that this paper is an extended version of that by Peña [18]. The extension consists in that the learning algorithm presented in that paper has been modified so that it returns a distinguished member of a Markov equivalence class of MAMP CGs, rather than just a member of the class. As a consequence, the proof of correctness of the algorithm has changed significantly. Moreover, the algorithm has been implemented and evaluated. This paper reports the results of the evaluation for the first time.

The rest of this paper is organized as follows. We start with some preliminaries in Section 2. Then, we introduce MAMP CGs in Section 3, followed by the algorithm for learning them in Section 4. In that section, we also include a review of other learning algorithms that are related to ours. We report the experimental results in Section 5. We close the paper with some discussion in Section 6. All the proofs appear in Appendix A at the end of the paper.

2. Preliminaries

In this section, we introduce some concepts of models based on graphs, i.e. graphical models. Most of these concepts have a unique definition in the literature. However, a few concepts have more than one and we opt for the most suitable in this work. All the graphs and probability distributions in this paper are defined over a finite set V. All the graphs in this paper are simple, i.e. they contain at most one edge between any pair of nodes. The elements of V are not distinguished from singletons.

If a graph G contains an undirected, directed or bidirected edge between two nodes V_1 and V_2, then we write that $V_1 - V_2$, $V_1 \to V_2$ or $V_1 \leftrightarrow V_2$ is in G. We represent with a circle, such as in $V_1 \multimap V_2$ or $V_1 \circ\!\!-\!\!\circ V_2$, that the end of an edge is unspecified, i.e. it may be an arrowhead or nothing. If the edge is of the form $V_1 \multimap V_2$, then we say it has an arrowhead at V_2. If the edge is of the form $V_1 \to V_2$, then we say that it has an arrowtail at V_1. The parents of a set of nodes X of G is the set $pa_G(X) = \{V_1 | V_1 \to V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The children of X is the set $ch_G(X) = \{V_1 | V_1 \leftarrow V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The neighbors of X is the set $ne_G(X) = \{V_1 | V_1 - V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The spouses of X is the set $sp_G(X) = \{V_1 | V_1 \leftrightarrow V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The adjacents of X is the set $ad_G(X) = ne_G(X) \cup pa_G(X) \cup ch_G(X) \cup sp_G(X)$. A route between a node V_1 and a node V_n in G is a sequence of (not necessarily distinct) nodes $V_1, \ldots, V_n$ such that $V_i \in ad_G(V_{i+1})$ for all $1 \le i < n$. If the nodes in the route are all distinct, then the route is called a path. The length of a route is the number of (not necessarily distinct) edges in the route, e.g. the length of the route $V_1, \ldots, V_n$ is $n-1$. A route is called descending if $V_i \to V_{i+1}$, $V_i - V_{i+1}$ or $V_i \leftrightarrow V_{i+1}$ is in G for all $1 \le i < n$. A route is called strictly descending if $V_i \to V_{i+1}$ is in G for all $1 \le i < n$. The descendants of a set of nodes X of G is the set $de_G(X) = \{V_n |$ there is a descending route from V_1 to V_n in G, $V_1 \in X$ and $V_n \notin X\}$. The strict ascendants of X is the set $san_G(X) = \{V_1 |$ there is a strictly descending route from V_1 to V_n in G, $V_1 \notin X$ and $V_n \in X\}$. A route $V_1, \ldots, V_n$ in G is called a cycle if $V_n = V_1$. Moreover, it is called a semidirected cycle if $V_n = V_1$, $V_1 \to V_2$ is in G and $V_i \to V_{i+1}$, $V_i \leftrightarrow V_{i+1}$ or $V_i - V_{i+1}$ is in G for all $1 < i < n$. A cycle has a chord if two non-consecutive nodes of the cycle are adjacent in G. The subgraph of G induced by a set of nodes X is the graph over X that has all and only the edges in G whose both ends are in X. Moreover, a triplex $(\{A, C\}, B)$ in G is an induced subgraph of the form $A \multimap B \leftarrow\!\circ\, C$, $A \multimap B - C$ or $A - B \leftarrow\!\circ\, C$.

A directed and acyclic graph (DAG) is a graph with only directed edges and without semidirected cycles. An AMP chain graph (AMP CG) is a graph whose every edge is directed or undirected such that it has no semidirected cycles. A MVR chain graph (MVR CG) is a graph whose every edge is directed or bidirected such that it has no semidirected cycles. Clearly, DAGs are a special case of AMP and MVR CGs: DAGs are AMP CGs without undirected edges, and DAGs are MVR CGs without bidirected edges. We now recall the semantics of AMP and MVR CGs. A node B in a path ρ in an AMP CG G is called a triplex node in ρ if $A \to B \leftarrow C$, $A \to B - C$, or $A - B \leftarrow C$ is a subpath of ρ. Moreover, ρ is said to be Z-open with $Z \subseteq V$ when

- every triplex node in ρ is in $Z \cup san_G(Z)$, and
- every non-triplex node B in ρ is outside Z, unless $A - B - C$ is a subpath of ρ and $pa_G(B) \setminus Z \neq \emptyset$.

A node B in a path ρ in an MVR CG G is called a triplex node in ρ if $A \multimap B \leftarrow\!\circ\, C$ is a subpath of ρ. Moreover, ρ is said to be Z-open with $Z \subseteq V$ when

- every triplex node in ρ is in $Z \cup san_G(Z)$, and
- every non-triplex node B in ρ is outside Z.

Let X, Y and Z denote three disjoint subsets of V. When there is no Z-open path in an AMP or MVR CG G between a node in X and a node in Y, we say that X is separated from Y given Z in G and denote it as $X \perp_G Y | Z$. The independence model represented by G, denoted as $I(G)$, is the set of separations $X \perp_G Y | Z$. In general, $I(G)$ is different depending on whether G is an AMP or MVR CG. However, it is the same when G is a DAG.