IJA:7917

International Journal of Approximate Reasoning ••• (••••) •••–•••

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Concise representations and construction algorithms for semi-graphoid independency models

Stavros Lopatatzidis [a,*], Linda C. van der Gaag [b]

[a] *Ghent University, SYSTeMS Research Group, Technologiepark-Zwijnaarde 914, 9052 Zwijnaarde, Belgium*
[b] *Utrecht University, Department of Information and Computing Sciences, Princetonplein 5, 3584 CC Utrecht, The Netherlands*

ARTICLE INFO

Article history:
Received 25 November 2015
Received in revised form 22 June 2016
Accepted 24 June 2016
Available online xxxx

Keywords:
Conditional independence
Semi-graphoid axioms
Closure
Closure representation
Dominant independence statements

ABSTRACT

The conditional independencies from a joint probability distribution constitute a model which is closed under the semi-graphoid properties of independency. These models typically are exponentially large in size and cannot be feasibly enumerated. For describing a semi-graphoid model therefore, researchers have proposed a more concise representation. This representation is composed of a representative subset of the independencies involved, called a basis, and lets all other independencies be implicitly defined by the semi-graphoid properties. An algorithm is available for computing such a basis for a semi-graphoid independency model. In this paper, we identify some new properties of a basis in general which can be exploited for arriving at an even more concise representation of a semi-graphoid model. Based upon these properties, we present an enhanced algorithm for basis construction which never returns a larger basis for a given independency model than currently existing algorithms.

1. Introduction

Mathematical models capturing joint probability distributions over sets of random variables are employed in numerous real-world applications. Especially probabilistic graphical models have become quite popular as appropriate models for describing distributions for problems in a range of societal fields. The practicability of computing probabilities of interest from these models typically derives from inference algorithms which exploit the modelled independency relation among the variables involved [5,6]. Independency relations embedded in joint probability distributions and their (concise) representation have therefore been subjects of extensive studies [2–4,9–11].

Pearl and his co-researchers were among the first to formalise qualitative properties of probabilistic independency in an axiomatic system [7,8]. The axioms from this system, which are known as the *semi-graphoid axioms*, are often looked upon as derivation rules for generating new independencies from a basic set of independency statements; any set of independencies that is closed under finite application of these rules is then called a *semi-graphoid independency model*. Semi-graphoid models constitute a quite general class among the various types of conditional independency model, in the sense that they provide for describing any independency relation embedded in a real-world probability distribution, even if this distribution is not strictly positive.

* Corresponding author.
E-mail addresses: stavros.lopatatzidis@ugent.be (S. Lopatatzidis), L.C.vanderGaag@uu.nl (L.C. van der Gaag).

http://dx.doi.org/10.1016/j.ijar.2016.06.011

A range of computational problems on semi-graphoid independency models are being addressed in the literature. Two closely-related problems are the *implication problem* and the *representation problem*. The implication problem is the problem of deciding whether a given statement of independency can be derived from a given set of such statements [14]. The representation problem is the problem of finding a small subset of independency statements that fully describes a given independency model. Semi-graphoid independency models in general include exponentially many statements. Representing these models by enumeration of their element independencies therefore is not feasible in practice. Studený was the first to propose a concise representation of an independency model, based on the semi-graphoid axioms [12,13]. The idea is to explicitly enumerate a representative subset of independency statements from a semi-graphoid model and let all other independencies be defined implicitly through the derivation rules; such a representative subset of statements is termed a *basis* for the model at hand. Studený designed an efficient algorithm for computing a basis for a semi-graphoid model from a given starting set of independency statements, which was later improved by Baioletti and his co-researchers [1].

In this paper, we revisit the representation of semi-graphoid independency models, and show that the subset of independency statements which have to be represented explicitly, can often be further reduced in size. We introduce the new notion of maximal non-symmetric basis for this purpose, with an associated algorithm for its computation. Our algorithm is shown to never result in a larger basis for a given independency model than existing algorithms. Also, the intermediate bases constructed in the various iterations of our algorithm will never be larger than those constructed by existing algorithms. Our enhanced algorithm as a consequence improves not just upon the size of the resulting basis for representation but upon the runtime complexity of its construction as well.

The paper is organised as follows. We provide some preliminaries on semi-graphoid independency models in Section 2, and review concise representations and their associated construction algorithms in Section 3. In Section 4 we detail, among other notions and properties, our notion of maximal non-symmetric basis. Section 5 then describes our enhanced algorithm for basis construction, and demonstrate the practicability of our algorithm by means of a number of example independency models. The paper ends with our concluding observations in Section 6.

2. Semi-graphoid independency models

We briefly review semi-graphoid independency models [7,13], and thereby introduce our notational conventions. We consider a finite, non-empty set S of random variables. A *triplet* θ over S is a statement of the form $\theta = \langle A,\ B \mid C\rangle$, where $A, B, C \subseteq S$ are mutually disjoint subsets of S with $A, B \neq \varnothing$; we will use $X_\theta = A \cup B \cup C$ to refer to the triplet's set of variables. A triplet $\langle A,\ B \mid C\rangle$ states that the sets of variables A and B are mutually independent given the set C; in view of a joint probability distribution Pr over S, the triplet states that $\Pr(A, B \mid C) = \Pr(A \mid C) \cdot \Pr(B \mid C)$. The set of all triplets over S is denoted by $S^{(3)}$. A (sub-)set of triplets now constitutes a *semi-graphoid independency model* if it satisfies the four so-called semi-graphoid properties stated in the following definition.

Definition 1. A *semi-graphoid independency model* is a subset of triplets $J \subseteq S^{(3)}$ which satisfies the following properties:

G1: if $\langle A,\ B \mid C\rangle \in J$, then $\langle B,\ A \mid C\rangle \in J$ *(symmetry);*
G2: if $\langle A,\ B \mid C\rangle \in J$, then $\langle A,\ B' \mid C\rangle \in J$ for any non-empty subset $B' \subseteq B$ *(decomposition)*;
G3: if $\langle A,\ B_1 \cup B_2 \mid C\rangle \in J$ with $B_1 \cap B_2 = \varnothing$, then $\langle A,\ B_1 \mid C \cup B_2\rangle \in J$ *(weak union)*;
G4: if $\langle A,\ B \mid C \cup D\rangle \in J$ and $\langle A,\ C \mid D\rangle \in J$, then $\langle A,\ B \cup C \mid D\rangle \in J$ *(contraction).*

The four semi-graphoid properties jointly convey the idea that learning irrelevant information does not alter the independencies among the variables discerned [7]. The weak union property G3 for example, states that learning information about B_2 which is known to be irrelevant with respect to A given C cannot help irrelevant information about B_1 to become relevant to A. We note that the contraction rule G4 cannot always be applied to two arbitrarily chosen triplets; we will return to this observation in Section 4.1.

The semi-graphoid properties of independency are often viewed, and referred to, as derivation rules for generating (new) triplets from a given set of triplets. Given a starting set of triplets $J \subseteq S^{(3)}$ and a designated triplet $\theta \in S^{(3)}$, we write $J \vdash^* \theta$ if the triplet θ can be derived from J by finite application of the semi-graphoid rules G1, G2, G3 and G4. We note that the four rules thereby induce a derivational relation among the triplets of a semi-graphoid model. A variety of problems on semi-graphoid independency models are being studied by building upon this derivational relation. A well-known problem is the *implication problem* [14], which is the problem of deciding whether a specific triplet θ can be derived from a given set of triplets J. More formally, the problem asks whether a given triplet θ is included in the *closure* of a triplet set J, where the notion of closure is defined as follows.

Definition 2. Let $J \subseteq S^{(3)}$ be a set of triplets. Then, the *closure* of J, denoted by $\overline{J}$, is the set of all triplets $\theta \in S^{(3)}$ such that $J \vdash^* \theta$.

Another, closely related problem on semi-graphoid independency models is the *representation problem* [13], which is the problem of finding a (small) subset of triplets J that fully describes a given semi-graphoid independency model M. More formally, the problem asks for a *basis* for a given independency model, where the notion of basis is defined as follows.