International Journal of Approximate Reasoning 92 (2018) 295–320

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Blankets Joint Posterior score for learning Markov network structures

Federico Schlüter [a,*], Yanela Strappa [a], Diego H. Milone [b], Facundo Bromberg [a]

[a] DHARMa Lab, Dept of Information Systems, Facultad Regional Mendoza, Universidad Tecnológica Nacional, Mendoza, Argentina
[b] Research Institute for Signals, Systems and Computational Intelligence, sinc(i), FICH-UNL/CONICET, Santa Fe, Argentina

ARTICLE INFO

Article history:
Received 27 March 2017
Received in revised form 20 October 2017
Accepted 23 October 2017
Available online 27 October 2017

Keywords:
Markov network
Structure learning
Scoring function
Blankets posterior
Irregular structures

ABSTRACT

Markov networks are extensively used to model complex sequential, spatial, and relational interactions in a wide range of fields. By learning the Markov network independence structure of a domain, more accurate joint probability distributions can be obtained for inference tasks or, more directly, for interpreting the most significant relations among the variables. Recently, several researchers have investigated techniques for automatically learning the structure from data by obtaining the probabilistic maximum-a-posteriori structure given the available data. However, all the approximations proposed decompose the posterior of the whole structure into local sub-problems, by assuming that the posteriors of the Markov blankets of all the variables are mutually independent. In this work, we propose a scoring function for relaxing such assumption. The *Blankets Joint Posterior* score computes the joint posterior of structures as a joint distribution of the collection of its Markov blankets. Essentially, the whole posterior is obtained by computing the posterior of the blanket of each variable as a conditional distribution that takes into account information from other blankets in the network. We show in our experimental results that the proposed approximation can improve the sample complexity of state-of-the-art competitors when learning complex networks, where the independence assumption between blanket variables is clearly incorrect.

1. Introduction

A Markov network (MN) is a popular probabilistic graphical model that efficiently encodes the joint probability distribution for a set of random variables of a specific domain [1–3]. MNs usually represent probability distributions by using two interdependent components: an independence structure, and a set of numerical parameters over the structure. The first is a qualitative component that represents structural information about a problem domain in the form of conditional independence relationships between variables. The numerical parameters are a quantitative component that represents the strength of the dependences in the structure. There is a large list of applications of MNs in a wide range of fields, such as computer vision and image analysis [4–6], computational biology [7], biomedicine [8,9], and evolutionary computation [10,11], among many others. For some of these applications, the model can be constructed manually by human experts, but in many other problems this can become unfeasible, mainly due to the dimensionality of the problem.

* Corresponding author.
E-mail address: federico.schluter@frm.utn.edu.ar (F. Schlüter).

https://doi.org/10.1016/j.ijar.2017.10.018
0888-613X/

Learning the model from data consists of two interdependent problems: learning the structure; and given the structure, learning its parameters. This work focuses on the task of learning the structure, which is useful for a variety of tasks. The structures learned may be used to construct accurate models for inference tasks (such as the estimation of marginal and conditional probabilities) [12–14], and may also be interesting per se, since they can be used as interpretable models that show the most significant interactions of a domain [15–19]. The first scenario is known in practice as the density estimation goal of learning, and the second one is known as the knowledge discovery goal of learning [Chapter 16 [3]].

An interesting approach to MN structure learning is to use constraint-based (also known as independence-based) algorithms [20–23]. Such algorithms proceed by performing statistical independence tests on data, and discard all structures inconsistent with the tests. This is an efficient approach, and it is correct under the assumption that the distribution can be represented by a graph, and that the tests are reliable. However, the algorithms that follow this approach are quite sensitive to errors in the tests, which may be unreliable for large conditioning sets [20,3]. A second approach to MN structure learning is to use score-based algorithms [24,25,15,26]. Such algorithms formulate the problem as an optimization, combining a strategy for searching through the space of possible structures with a scoring function that measures the fitness of each structure to the data. The structure learned is the one that achieves the highest score in the search.

It is important to mention that both constraint-based and score-based approaches have been originally motivated by distinct learning goals. According to the existing literature [3], constraint-based methods are generally designed for the knowledge-discovery goal of learning [22,21], and their quality is often measured in terms of the correctness of the structure learned (structural errors). In contrast, most score-based approaches have been designed for the density estimation goal of learning [12–14], and they are in general evaluated in terms of inference accuracy. For this reason, score-based algorithms often work by considering the whole MN at once during the search, interleaving the parameter learning step. This makes them more accurate for inference tasks. However, since learning the parameters is known to be NP-hard for MNs [27], it has a negative effect on their scalability.

Recently, there has been a surge of interest towards efficient methods based on a strategy that follows a score-based approach, but with the knowledge discovery goal in mind. Basically, an undirected graph structure is learned by obtaining the probabilistic maximum-a-posteriori structure given the available data [28,19,29]. This hybrid strategy achieves scalability, as well as reliable performance. Such contributions consist in the design of efficient scoring functions for MN structures, expressing the problem formally as follows: given a complete training data set D, find an undirected graph $G^\star$ such that

$$G^\star = \arg\max_{G \in \mathcal{G}} \Pr(G|D), \tag{1}$$

where $\Pr(G|D)$ is the posterior probability of a structure given D, and $\mathcal{G}$ is the family of all the possible undirected graphs for the domain size. This class of algorithms has been shown to outperform constraint-based algorithms in the quality of the learned structures, with competitive computational complexities. The method proposed in this paper follows this approach.

Since there are no feasible exact methods for computing the posterior of MN structures, different approximations have been proposed. An important assumption commonly made by the current state-of-the-art methods is to suppose that the posterior of the structure is decomposable [30,31,3,28,19,29]. It means that the whole posterior can be computed as a product of the posteriors of the Markov blankets that compose the structure, which are smaller posteriors that can be computed independently. In fact, this is a good approximation that improves the efficiency of search. The research line of this work aims at designing a better approximation of the posterior, by relaxing this independence assumption. This work's contribution is the *Blankets Joint Posterior* (BJP), a scoring function that estimates $\Pr(G|D)$ as the joint posterior probability of the Markov blankets of G. This is achieved by formulating $\Pr(G|D)$ in a novel way that relaxes the independence assumption between the blankets. Essentially, the whole posterior is obtained by computing the posterior of the blanket of each variable as a conditional distribution that takes into account information from other blankets in the network. In our experiments we show that the proposed approximation can improve the sample complexity of state-of-the-art scores when learning networks with complex topologies, that commonly appear in real-world problems.

After providing some preliminaries, notations and definitions in Section 2, we introduce the BJP scoring function in Section 3. Section 4 presents the experimental results for several study cases. Finally, Section 5 summarizes this work, and poses several possible directions of future work.

2. Background

We begin by introducing the notation used for MNs. Then we provide some additional background about these models and the problem of learning their independence structure, and also discuss the state-of-the-art of MN structure learning.

2.1. Markov networks

Let V be a finite set of indexes, with lowercase subscripts for denoting particular indexes, e.g., $i, j \in V$, and uppercase subscripts for subsets of indexes, e.g., $W \subseteq V$. Let X_V be the set of random variables of a domain, denoting single variables as single indexes in V, e.g., $X_i, X_j \in X_V$ where $i, j \in V$. For a MN representing a probability distribution $P(X_V)$, its two components are denoted as follows: G, and θ. G is the structure, an undirected graph $G = (V, E)$ where the nodes $V = \{0, \ldots, n-1\}$ are the indices of each random variable X_i of the domain, and $E \subseteq \{V \times V\}$ is the edge set of the graph.