International Journal of Approximate Reasoning 93 (2018) 74–87

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Tractability of most probable explanations in multidimensional Bayesian network classifiers ☆

Marco Benjumeda *, Concha Bielza, Pedro Larrañaga

Computational Intelligence Group, Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid, Spain

ARTICLE INFO

Article history:
Received 1 December 2016
Received in revised form 4 October 2017
Accepted 9 October 2017
Available online 27 October 2017

Keywords:
Multidimensional classification
Bayesian network classifiers
Most probable explanation complexity
Machine Learning

ABSTRACT

Multidimensional Bayesian network classifiers have gained popularity over the last few years due to their expressive power and their intuitive graphical representation. A drawback of this approach is that their use to perform multidimensional classification, a generalization of multi-label classification, can be very computationally demanding when there are a large number of class variables. Thus, a key challenge in this field is to ensure the tractability of these models during the learning process.

In this paper, we show how information about the most common queries of multidimensional Bayesian network classifiers affects the complexity of these models. We provide upper bounds for the complexity of the most probable explanations and marginals of class variables conditioned to an instantiation of all feature variables. We use these bounds to propose efficient strategies for bounding the complexity of multidimensional Bayesian network classifiers during the learning process, and provide a simple learning method with an order-based search that guarantees the tractability of the returned models. Experimental results show that our approach is competitive with other methods in the state of the art and also ensures the tractability of the learned models.

1. Introduction

Bayesian network classifiers [1] are one of the most widely used machine learning tools to address the problem of classification. Classification consists of assigning an instance to a class that is described by a set of features.

Multidimensional Bayesian network classifiers (MBCs) [2] extend Bayesian network classifiers to the problem of multidimensional classification. Multidimensional classification involves assigning an instance to a set of classes (instead of a single class) given the value of the set of features. This problem is common in several domains like text categorization (a text can be assigned to multiple topics), medicine (a patient may suffer from several diseases) or system monitoring (a system may break down from multiple failures).

MBCs are Bayesian networks (BN) with a restricted topology, where no arcs from feature variables to class variables are allowed. Each MBC is composed of a class subgraph, a bridge subgraph, and a feature subgraph (see Section 2). Inference in MBCs may have a high computational cost for some structures, even when the class and feature subgraphs are restricted to trees or polytrees.

☆ This paper is part of the Virtual special issue on the Eighth International Conference on Probabilistic Graphical Models, Edited by Giorgio Corani, Alessandro Antonucci, Cassio De Campos.

* Corresponding author.
E-mail addresses: marco.benjumeda.barquita@upm.es (M. Benjumeda), mcbielza@fi.upm.es (C. Bielza), pedro.larranaga@fi.upm.es (P. Larrañaga).

https://doi.org/10.1016/j.ijar.2017.10.024
0888-613X/

Although there is work in the literature addressing the problem of computational complexity in MBCs, the focus has not been on taking advantage of the most common type of queries of such models. In this paper, we study the computational complexity of most probable explanations (MPEs) and marginals of class variables in MBCs when an instantiation of the feature variables is given. The paper also provides upper bounds on the complexity of these models given additional restrictions on their structure that limit the treewidth of a transformation of it that we call the pruned graph.

Class-bridge (CB) decomposable MBCs [3] are capable of dividing the MPE problem into multiple simpler subproblems that can be computed independently in each of the MBC components. We prove that CB-decomposability can also be used to efficiently bound the complexity of MBCs during the learning process. We propose a learning method that uses these properties to search for tractable MBCs in the space of topological orderings.

This paper is an extended version of the work published in [4]. We extend the theoretical results to marginal computations, provide an alternative strategy for bounding the complexity of MBCs in the learning method, and extend the experiments using additional performance measures and real-world datasets. The rest of the paper is organized as follows. Section 2 describes MBCs, introduces CB-decomposability, and reviews previous work on inference complexity and learning in MBCs. Section 3 presents the new theoretical results with respect to the complexity of computations of MPEs and marginals in MBCs. Section 4 describes the method proposed for learning tractable MBCs. Section 5 reports the experimental results. Section 6 draws some conclusions and suggests future research lines.

2. Background

2.1. Multidimensional classification with Bayesian networks

A Bayesian network $\mathcal{B}$ represents a joint probability distribution over a set of random variables $\mathcal{V} = \{V_1, \ldots, V_n\}$. It is composed of a directed acyclic graph (DAG) $\mathcal{G}$ that represents the conditional dependences among the variables in $\mathcal{V}$, and a set of parameters $\Pr(V_i|\mathbf{Pa}_{\mathcal{G}}(V_i))$ (we use $\mathbf{Pa}_{\mathcal{G}}(V_i)$ to refer to the parents of V_i in $\mathcal{G}$) that represent the conditional probability distributions (CPDs) of each $V_i \in \mathcal{V}$ conditioned on its parents in $\mathcal{G}$. A joint probability distribution that satisfies the Markov condition with $\mathcal{G}$ is given by

$$\Pr(V_1, \ldots, V_n) = \prod_{i=1}^{n} \Pr(V_i|\mathbf{Pa}_{\mathcal{G}}(V_i)) \ . \tag{1}$$

Van der Gaag and de Waal [2] introduced *multidimensional Bayesian network classifiers* as an extension of Bayesian classifiers to multidimensional classification. MBCs are a special case of Bayesian networks with a restricted structure topology. They are defined as follows:

Definition 1. An MBC is a Bayesian network $\mathcal{B}$ over a set of variables $\mathcal{V} = \{V_1, V_2, \ldots, V_n\}$, where $\mathcal{V}$ is partitioned into two sets $\mathcal{C} = \{C_1, \ldots, C_d\}$, $d \geq 1$, of class variables and $\mathcal{F} = \{F_1, \ldots, F_m\}$, $m \geq 1$, of feature variables ($d + m = n$). The arcs in $\mathcal{G}$ are partitioned into three subsets, A_C, A_F, A_B, such that:

- $A_C \subseteq \mathcal{C} \times \mathcal{C}$ is composed of the arcs between the class variables having a subgraph $\mathcal{G}_C = (\mathcal{C}, A_C)$ – class subgraph – of $\mathcal{G}$ induced by $\mathcal{C}$.
- $A_F \subseteq \mathcal{F} \times \mathcal{F}$ is composed of the arcs between the feature variables having a subgraph $\mathcal{G}_F = (\mathcal{F}, A_F)$ – feature subgraph – of $\mathcal{G}$ induced by $\mathcal{F}$.
- $A_B \subseteq \mathcal{C} \times \mathcal{F}$ is composed of the arcs from the class variables to the feature variables having a subgraph $\mathcal{G}_B = (\mathcal{V}, A_B)$ – bridge subgraph – of $\mathcal{G}$ induced by $\mathcal{V}$ connecting class and feature variables.

Fig. 1 shows an example of the structure of an MBC and its corresponding subgraphs.

The problem of multidimensional classification in MBCs involves getting the most probable explanation (MPE) of the class variables given an instantiation of the feature variables, which is given by

$$\mathbf{c}^* = \arg\max_{\mathbf{c} \in \Omega_{\mathcal{C}}} \Pr(\mathbf{c}|\mathbf{f}) = \arg\max_{\mathbf{c} \in \Omega_{\mathcal{C}}} \Pr(\mathbf{c}, \mathbf{f}) \ , \tag{2}$$

where $\mathbf{f}$ is an instantiation of $\mathcal{F}$ and $\Omega_{\mathcal{C}}$ is the set containing all the possible configurations of $\mathcal{C}$.

2.2. Class-bridge decomposable multidimensional Bayesian network classifiers

An MBC is *class-bridge decomposable* [3] if it can be decomposed into multiple connected components, where each component is composed of all the nodes that are connected by an undirected path in $\mathcal{G}_C \cup \mathcal{G}_B$. Basically, the components of an MBC are the connected graphs obtained after removing the arcs of the feature subgraph from this MBC.

Definition 2. A CB-decomposable MBC is a BN $\mathcal{B}$ whose class subgraph and bridge subgraph are decomposed into r maximal components such that: