Information Sciences 233 (2013) 109–125

Contents lists available at SciVerse ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A review on evolutionary algorithms in Bayesian network learning and inference tasks

Pedro Larrañaga [a], Hossein Karshenas [a,*], Concha Bielza [a], Roberto Santana [b]

[a] Computational Intelligence Group, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo, 28660 Boadilla del Monte, Madrid, Spain

[b] Intelligent System Group, Department of Computer Science and Artificial Intelligence, University of the Basque Country, Paseo Manuel de Lardizabal 1, 20080 San Sebastian–Donostia, Spain

ARTICLE INFO

Article history:
Received 14 May 2012
Received in revised form 19 October 2012
Accepted 21 December 2012
Available online 29 January 2013

Keywords:
Probabilistic graphical model
Bayesian network
Evolutionary computation
Inference
Learning from data

ABSTRACT

Thanks to their inherent properties, probabilistic graphical models are one of the prime candidates for machine learning and decision making tasks especially in uncertain domains. Their capabilities, like representation, inference and learning, if used effectively, can greatly help to build intelligent systems that are able to act accordingly in different problem domains. Bayesian networks are one of the most widely used class of these models. Some of the inference and learning tasks in Bayesian networks involve complex optimization problems that require the use of meta-heuristic algorithms. Evolutionary algorithms, as successful problem solvers, are promising candidates for this purpose. This paper reviews the application of evolutionary algorithms for solving some NP-hard optimization tasks in Bayesian network inference and learning.

1. Introduction

Probability theory has provided a sound basis for many of scientific and engineering tasks. Artificial intelligence, and more specifically machine learning, is one of the fields that has exploited probability to develop new theorems and algorithms. A popular class of probabilistic graphical models (PGMs), Bayesian networks, first introduced by Pearl [105], combine graph and probability theories to obtain a more comprehensible representation of the joint probability distribution. This tool can point out useful modularities in the underlying problem and help to accomplish the reasoning and decision making tasks especially in uncertain domains. The application of these useful tools has been further improved by different methods proposed for PGM inference [86] and automatic induction [23] from a set of samples.

Meanwhile, the difficult and complex problems existing in real-world applications have increased the demand for effective meta-heuristic algorithms that are able to achieve good (and not necessarily optimal) solutions by performing an intelligent search of the space of possible solutions. Evolutionary computation is one of the most successful of these algorithms that has achieved very good results across a wide range of problem domains. Applying their nature-inspired mechanisms, e.g., survival of the fittest or genetic crossover and mutation, on a population of candidate solutions, evolutionary approaches like genetic algorithms [59] have been able to perform a more effective and diverse search of the vast solution space of difficult problems.

* Corresponding author. Tel.: +34 913363675; fax: +34 913524819.
E-mail addresses: pedro.larranaga@fi.upm.es (P. Larrañaga), hkarshenas@fi.upm.es (H. Karshenas), mcbielza@fi.upm.es (C. Bielza), roberto.santana@ehu.es (R. Santana).

0020-0255/$ - see front matter
http://dx.doi.org/10.1016/j.ins.2012.12.051

Some of the most relevant inference and learning problems in Bayesian networks are formulated as the optimization of a function. These problems usually have an intractable complexity and therefore are a potential domain for the application of meta-heuristic methods. The aim of this paper is to review how evolutionary algorithms have been applied for solving some of the combinatorial problems existing in the inference and learning of Bayesian networks.

The paper is organized as follows. Section 2 introduces Bayesian networks and reviews some of the inference and learning methods proposed for them. Section 3 presents the framework of evolutionary algorithms and discusses how they work. The main review of how evolutionary algorithms are used in Bayesian network learning and inference is given in Section 4. Finally, Section 5 concludes the paper.

2. Bayesian networks

This section gives an introduction to Bayesian networks and how they are used for representing probability distributions in discrete, continuous, and hybrid environments. It then briefly reviews some of the methods for inference and learning of Bayesian networks. The terminology and concepts adopted and introduced in this section are later used in the presentation of evolutionary algorithms for learning and inference in Bayesian networks. For more information on Bayesian networks and PGMs in general, see Koller and Friedman [74], and Larrañaga and Moral [83].

2.1. Probability-related notations

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a vector of random variables and $\boldsymbol{x} = (x_1, \ldots, x_n)$ a possible value combination for these variables. x_i denotes a possible value of X_i, the ith component of $\boldsymbol{X}$, and $\boldsymbol{y}$ denotes a possible value combination for the sub-vector $\boldsymbol{Y} = (X_{J_1}, \ldots, X_{J_k})$, $J = \{J_1, \ldots, J_k\} \subseteq \{1, \ldots, n\}$.

If all variables in $\boldsymbol{X}$ are discrete, $P(\boldsymbol{X} = \boldsymbol{x})$ (or simply $P(\boldsymbol{x})$) is used to denote the *joint probability mass* of a specific configuration $\boldsymbol{x}$ for the variables. The *conditional probability mass* of a specific value x_i of variable X_i given that $X_j = x_j$ is denoted by $P(X_i = x_i|X_j = x_j)$ (or simply $P(x_i|x_j)$). Similarly, for continuous variables, the *joint density function* will be denoted as $p(\boldsymbol{x})$ and the *conditional density function* by $p(x_i|x_j)$. When the nature of variables in $\boldsymbol{X} = (X_1, \ldots, X_n)$ is irrelevant, $\rho(\boldsymbol{x}) = \rho(x_1, \ldots, x_n)$ will be used to represent the generalized joint probability. Let $\boldsymbol{Y}$, $\boldsymbol{Z}$ and $\boldsymbol{W}$ be three disjoint sub-vectors of variables. Then, $\boldsymbol{Y}$ is said to be *conditionally independent* of $\boldsymbol{Z}$ given $\boldsymbol{W}$ (denoted by $I(\boldsymbol{Y}, \boldsymbol{Z}|\boldsymbol{W})$), iff $\rho(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{w}) = \rho(\boldsymbol{y}|\boldsymbol{w})$, for all $\boldsymbol{y}$, $\boldsymbol{z}$ and $\boldsymbol{w}$.

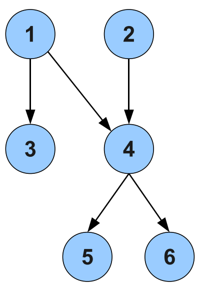
2.2. Bayesian network definition

A Bayesian network (BN) $\mathcal{B}(\mathcal{S}, \Theta)$ for a vector of variables $\boldsymbol{X} = (X_1, \ldots, X_n)$ consists of two components:

- A structure $\mathcal{S}$ represented by a directed acyclic graph (DAG), expressing a set of conditional independencies [30] between variables.
- A set of local parameters Θ representing the conditional probability distributions for the values of each variable given different value combinations of their parents, according to the structure $\mathcal{S}$.

Fig. 1a shows an example of a BN structure for a problem with six variables. For each variable X_i, $i = 1, \ldots, n$, structure $\mathcal{S}$ represents the assertion that X_i and its non-descendants, $ND(X_i)$, excluding its parents are conditionally independent given its parents, $\boldsymbol{Pa}_i$: i.e., $I(X_i, ND(X_i) \setminus \boldsymbol{Pa}_i|\boldsymbol{Pa}_i)$. This property is known as the Markov condition of BNs. Therefore, a BN encodes a factorization for the joint probability distribution of the variables

$$\rho(\boldsymbol{x}) = \rho(x_1, \ldots, x_n) = \prod_{i=1}^{n} \rho_{\mathcal{B}}(x_i|\boldsymbol{pa}_i), \tag{1}$$



X_1	1	1	1	2	2	2
X_2	1	2	3	1	2	3
$P(X_4=1 \mid X_1,X_2)$	θ_{411}	θ_{421}	θ_{431}	θ_{441}	θ_{451}	θ_{461}
$P(X_4=2 \mid X_1,X_2)$	θ_{412}	θ_{422}	θ_{432}	θ_{442}	θ_{452}	θ_{462}
$P(X_4=3 \mid X_1,X_2)$	θ_{413}	θ_{423}	θ_{433}	θ_{443}	θ_{453}	θ_{463}
$P(X_4=4 \mid X_1,X_2)$	θ_{414}	θ_{424}	θ_{434}	θ_{444}	θ_{454}	θ_{464}
$P(X_4=5 \mid X_1,X_2)$	θ_{415}	θ_{425}	θ_{435}	θ_{445}	θ_{455}	θ_{465}

(a) Structure (b) Conditional probability table

Fig. 1. An example of a Bayesian network structure and the parameters for one of its variables (X_4) assuming that $r_i = i + 1$.