International Journal of Approximate Reasoning 69 (2016) 147–167

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Structural learning of Bayesian networks by bacterial foraging optimization

Cuicui Yang [a], Junzhong Ji [a,*], Jiming Liu [b,1], Jinduo Liu [a], Baocai Yin [a]

[a] College of Computer Science and Technology, Beijing University of Technology, Beijing, China
[b] Department of Computer Science, Hong Kong Baptist University, Kowloon Tong, Hong Kong

ARTICLE INFO

Article history:
Received 15 October 2014
Received in revised form 29 October 2015
Accepted 3 November 2015
Available online 27 November 2015

Keywords:
Bacterial foraging optimization
Bayesian networks
Structural learning
Swarm intelligence

ABSTRACT

Algorithms inspired by swarm intelligence have been used for many optimization problems and their effectiveness has been proven in many fields. We propose a new swarm intelligence algorithm for structural learning of Bayesian networks, BFO-B, based on bacterial foraging optimization. In the BFO-B algorithm, each bacterium corresponds to a candidate solution that represents a Bayesian network structure, and the algorithm operates under three principal mechanisms: chemotaxis, reproduction, and elimination and dispersal. The chemotaxis mechanism uses four operators to randomly and greedily optimize each solution in a bacterial population, then the reproduction mechanism simulates survival of the fittest to exploit superior solutions and speed convergence of the optimization. Finally, an elimination and dispersal mechanism controls the exploration processes and jumps out of a local optima with a certain probability. We tested the individual contributions of four algorithm operators and compared with two state of the art swarm intelligence based algorithms and seven other well-known algorithms on many benchmark networks. The experimental results verify that the proposed BFO-B algorithm is a viable alternative to learn the structures of Bayesian networks, and is also highly competitive compared to state of the art algorithms.

1. Introduction

A Bayesian network (BN) is one of the most effective theoretical models to represent uncertainty of knowledge in artificial intelligence. A BN uses a graphical model to depict conditional independence relations among random variables in a domain and encode the joint probability distribution of random variables [1]. Given a BN and observations of some variables, the values of other unobserved variables can be predicted by probabilistic inference. Therefore, systems successfully use this paradigm to model practical problems in many different areas, such as medical diagnosis, natural language processing, forecasting, biology, and control [2].

Learning a BN structure automatically from data has received much attention, and variety of learning algorithms have been proposed [3–29]. These algorithms all adopt either the dependency analysis or score and search approaches. Dependency analysis is a constraint satisfaction problem, and employs a statistical method to judge dependency and independency relationships among variables and thereby constructs a BN [19]. Score and search is an optimization problem, and employs a

* Corresponding author.
E-mail addresses: yangcc_2008@163.com (C. Yang), jjz01@bjut.edu.cn (J. Ji), jiming@comp.hkbu.edu.hk (J. Liu).
[1] Fellow, IEEE.

http://dx.doi.org/10.1016/j.ijar.2015.11.003
0888-613X/

search method to probe the space of BN structures and a metric to constantly evaluate each candidate network structure until the best metric value is obtained [14]. Unfortunately, both approaches have fatal drawbacks. Dependency analysis needs to perform an exponential number of dependency tests that are usually complex and unreliable, and it is hard to ensure learning quality. In contrast, learning a BN structure by score and search becomes an NP-hard problem as the number of variables increases [30]. Once the space of candidate networks becomes large, nearly all exact searches are inappropriate for BN structural learning. Although some heuristic algorithms, such as iterated local search [3], K2 [25], and hill climbing [26,27] algorithms can address the problem of large search spaces, they often become trapped in local optima.

To solve these problems, several stochastic algorithms based on global optimization mechanisms have been introduced for structural learning of BNs in recent years. These algorithms can be divided into two categories [31]: 1) The evolutionary algorithm, which draws inspiration from evolution and natural genetics and includes evolutionary programming, genetic algorithm, evolution strategy, and genetic programming. Evolutionary programming [4,5] and genetic algorithm [8,9,17] based methods are effective ways with which a BN structure can be successfully learned. 2) The swarm intelligence algorithm, which is a nature inspired optimization technology that consists of particle swarm optimization (PSO) [32,33], ant colony optimization (ACO) [34,35], artificial bee colony optimization (ABC) [36], and bacterial foraging optimization (BFO) [37]. ACO, ABC, and PSO have proven their effectiveness at learning a BN structure from the data [6,10,23]. Their common feature is the use of a meta-heuristic search mechanism to explore the BN structural space while a scoring metric is applied to evaluate the fitness of candidate networks.

BFO is a swarm intelligence algorithm developed by Passino in 2002 [37,38], which simulates the foraging behavior of *Escherichia coli* bacteria. The basic principle is that bacteria move through either tumbling or swimming to maximize the energy consumed by eating as many nutrients as they can. As the smallest creatures on earth, bacteria contain many clever optimization mechanisms. Thus, BFO has unique good performance, and has been successful in a wide variety of optimization tasks since it was proposed [39–44]. However, to date this optimization technology has not been applied to learning BN structures.

Existing structural learning methods PSO-B, ACO-B, and ABC-B (based on the PSO, ACO, and ABC swarm intelligence algorithms, respectively) have some latent drawbacks. PSO-B keeps track of two types of optimal solutions, which makes it easily trapped in local optima. ACO-B and ABC-B employ pheromones to construct solutions. Although the positive feedback mechanism behind the pheromone can effectively guide the search for superior solutions, if the pheromone is over used, it may overpower a better solution, and the risk of the algorithms becoming trapped in local optima is high. However, BFO does not contain mechanisms that make an algorithm easily trapped, and has a high probability of escaping from local optima. Hence, we propose a new BN structural learning method, BFO-B, based on BFO.

In BFO-B, each bacterium constantly looks for a network structure with a better metric value using three optimization mechanisms: chemotaxis, reproduction, and elimination and dispersal. The chemotaxis locally optimizes each feasible solution, reproduction applies survival of the fittest to candidate solutions, and elimination and dispersal allows jump out of a local optima. The three mechanisms maintain a balance between exploitation and exploration and make it possible to obtain a global optimal or near optimal solution. To verify BFO-B performance, we conducted a series of experiments on many benchmark networks, investigating the effects of key parameters on the algorithm performance, contributions of different mechanisms to the algorithm performance, and performance comparisons with two swarm intelligence based algorithms and seven other algorithm types. The experimental outcomes verify that BFO-B is a promising approach to learn BN structures from data, highly competitive compared with two state of the art swarm intelligence based methods, and significantly superior to other methods.

Section 2 of this paper briefly introduces BNs, the K2 scoring metric of BNs, and BFO. Section 3 presents the details of the BFO-B algorithm, and the verification experiments are described and outcomes are discussed in Section 4. Section 5 summarizes our conclusions and possible future directions.

2. Preliminaries

2.1. BNs

A BN, also known as a belief network or a causal network, is a directed acyclic graph (DAG), which qualitatively characterizes the dependent and independent relationships among random variables, and uses a set of probability parameters to quantify the strength of the dependencies between each node and its parent nodes. It can be denoted as $G = (X, A)$, where $X = \{X_1, X_2, \cdots, X_i, \cdots, X_n\}$ is a set of nodes, X_i is a random variable, $A = \{a_{ij}\}$ is a set of arcs, and a_{ij} describes a direct dependence relationship between X_i and X_j. A set of conditional probability parameters is also associated with each non-root node, $P\left(X_i|\prod(X_i)\right)$, where $\prod(X_i)$ is a parent set of X_i, which quantifies how much X_i depends on its parents. Thus, a BN can be uniquely encoded using the joint probability distribution of the variable set $X = \{X_1, X_2, \cdots, X_n\}$,

$$P(X_1, X_2, \cdots, X_n) = \prod_{i=1}^{n} P(X_i|\prod(X_i)). \tag{1}$$