Applied Soft Computing 11 (2011) 3373–3384

Contents lists available at ScienceDirect

Applied Soft Computing

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

A hybrid method for learning Bayesian networks based on ant colony optimization

Junzhong Ji*, Renbing Hu, Hongxun Zhang, Chunnian Liu

College of Computer Science and Technology, Beijing University of Technology, Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology, Beijing 100124, China

ARTICLE INFO

Article history:
Received 7 June 2009
Received in revised form 23 August 2010
Accepted 3 January 2011
Available online 12 January 2011

Keywords:
Bayesian networks
Ant colony optimization
Variable search space
Heuristic
Function
Simulated annealing strategy

ABSTRACT

As a powerful formalism, Bayesian networks play an increasingly important role in the Uncertainty Field. This paper proposes a hybrid method to discover the knowledge represented in Bayesian networks. The hybrid method combines dependency analysis, ant colony optimization (ACO), and the simulated annealing strategy. Firstly, the new method uses order-0 independence tests with a self-adjusting threshold value to reduce the size of the search space, so that the search process takes less time to find the near-optimal solution. Secondly, better Bayesian network models are generated by using an improved ACO algorithm, where a new heuristic function is introduced to further enhance the search effectiveness and efficiency. Finally, an optimization scheme based on simulated annealing is employed to improve the optimization efficiency in the stochastic search process of ants. In a number of experiments and comparisons, the hybrid method outperforms the original ACO-B which uses ACO and some other network learning algorithms.

1. Introduction

Bayesian networks (BNs) are important probabilistic models within the field of artificial intelligence, and also powerful formalisms to model the uncertainty in the real world. A Bayesian network uses a graphical model to depict conditional independence among random variables in the domain and encodes the joint probability distribution. Given a network and observations of some variables, the values of other unobserved variables can be predicted by a probabilistic inference. Nowadays, many systems have been constructed based on this paradigm in a variety of different areas including vision recognition, medial diagnosis, trouble-shooting, information retrieval and so on.

With the development and popularity of BNs, earning BN structure from data has received considerable attention, and researchers have proposed various learning algorithms [1–13]. Generally, these algorithms can be classified into two main categories [3]: the dependency analysis approach, and the score-and-search approach. The first poses BN learning as a constraint satisfaction problem, and constructs a BN by dependency tests [2,3]. The second poses BN learning as an optimization problem, and uses a search method to find a network structure with the best score where a scoring metric is employed to evaluate candidate networks [1,4]. Unfortunately, both approaches have their own drawbacks. For example, the first approach has to perform an exponential number of dependency tests and some test results of higher order are unreliable, while the second approach often traps in a local optimum due to huge search spaces and the limitation of search methods. To solve these problems, new algorithms have been developed in recent years. For instance, there are three efficient approaches using a meta-heuristic mechanism to get the global near-optimum in the candidate network space. The first uses Genetic Algorithm (GA) [5,7], the second applies Evolutionary Programming (EP) [8,11], and the third employs ant colony optimization (ACO) [6,9]. Moreover, there is a research focus [10,11] that combines basic ideas of the dependency analysis approach and the score-and-search approach. These hybrid methods first use a dependency analysis method to reduce the search space of candidate solutions, then employ a score-and-search method to search in the reduced space. Different methods in dependency analysis and score-and-search phases can be used, which compose different hybrid methods.

In this paper, we propose a hybrid method to learn BNs. The hybrid method consists of two phases, namely, the Conditional Independence (CI) test phase and the search phase. In the CI test phase, order-0 independence tests with a self-adjusting threshold value are conducted to dynamically restrict search spaces of feasible solutions, so that the search process in the next phase can be accelerated while keeping good solution quality. In the search phase, an improved ACO for learning BNs is used to find good models. Here we use two techniques: 1. A new heuristic function

* Corresponding author.
E-mail address: jjz01@bjut.edu.cn (J. Ji).

1568-4946/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2011.01.009

combining the global score-increase of a solution with local mutual information between nodes is introduced to enhance the search effectiveness and efficiency. 2. An optimization strategy based on a Metropolis rule of simulated annealing is employed to further improve the optimization efficiency in the stochastic searching of ants. We call our new method hybrid ant colony optimization for Bayesian network learning (HACO-B). In a number of experiments, we perform an analytical study to compare the new method to ACO-B and some other network learning algorithms. The experimental results on benchmark data sets show that the hybrid algorithm outperforms the original ACO-B and some other network learning algorithms.

The paper is organized as follows. In Section 2, we present the background of Bayesian networks and the basic idea of the ant colony optimization for learning Bayesian networks. In Section 3, we describe our new algorithm in detail. Section 4 reports our experimental results. Finally, we conclude the paper in Section 5.

2. Background

2.1. Bayesian networks

A Bayesian network is a Directed Acyclic Graph (DAG) $G = \langle X, A \rangle$, where each node $X_i \in X$ represents a random variable in a domain, and each arc $a_{ij} \in A$ describes a direct dependence relationship between two variables X_i and X_j. Associated with each node X_i, is a conditional probability distribution represented by $\theta_i = P(X_i \mid \prod(X_i))$, which quantifies how much the node X_i depends on its parents $\prod(X_i)$. As the graph structure G qualitatively characterizes the independence relationship among random variables, the conditional probability distribution quantifies the strength of dependencies between a node and its parent nodes. It can be proved that a Bayesian network $\langle X, A \rangle$ uniquely encodes the joint probability distribution of the domain variables $X = \{X_1, X_2, \ldots, X_n\}$:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \Pi(X_i)) \quad (1)$$

2.2. Learning Bayesian network structures

The structure of a BN reflects the underlying probabilistic dependence relations among the nodes (corresponding perhaps to a database attribute) and a set of assertions about conditional independencies. The problem of learning a BN structure can be stated as follows: given a sample data $D = \{X[1], X[2], \ldots, X[N]\}$ where $X[i]$ is an instance of domain variables, the learning goal is to find the BN structure that best matches D. During the past decade, people have proposed many algorithms on learning Bayesian network structure. As mentioned above, there are two basic mechanisms. The first is an approach based on the dependency analysis [2,3], which takes the learning process as a constraint satisfaction problem, and then constructs a network structure by testing the conditional independence relations. The second is score-and-search approach [1,4], which takes the learning problem as a structure optimization problem. The latter uses a score metric to evaluate every candidate network structure, and then finds a network structure with the best score. Though the implementation of the former approach is relatively simple, the computations for high-order tests are complex and unreliable. Moreover, the precision of the learned model from the dependency analysis approach is hard to ensure, thus the score-and-search approach is gradually becoming a popular approach for learning Bayesian networks.

Given a node ordering, the parent nodes of each node in a BN, $\prod(X_i) = \{X_k : k \in \Phi(i)\}$, are only selected from the set of nodes preceding the current node X_i, namely, $\Phi(i) \subseteq \{1, 2, \ldots, i-1\}$, thus the number of possible parent sets is 2^{i-1} for each node X_i. Further, the number of possible structures for a BN with n nodes is $2^{n(n-1)/2}$ when a node ordering is known, and the complexity of a BN structure space is $n!\ 2^{n(n-1)/2}$ for the case of an unknown node ordering. Obviously, it is intractable for the complete search based on a score to find the global optimal solution when n is large. In the last few years, researchers proposed some effective algorithms [4,10,12] assuming a complete node ordering. Unfortunately, these algorithms still perform complete searching in the worst case, and they are unfit to learn a BN structure without a complete node ordering.

Though some improved hill-climbing algorithms [12,13] can also solve the problem of learning a BN structure with an unknown node ordering, they usually get a local optimal solution of the model. Recently, the development of stochastic search technologies has provided an effective and feasible method to tackle the problem. Genetic algorithms [5,7], evolutionary programming [8,11] and ant colony optimization [6,9] have been applied to learning Bayesian networks, respectively. These methods perform stochastically iterative searches and find the global best solution by means of simulating various natural phenomena.

2.3. Learning Bayesian networks using ACO (ACO-B)

2.3.1. Ant colony optimization

Ant colony optimization (ACO) is a meta-heuristic search algorithm, which was first proposed by Dorigo et al. in the 1990s [14,15]. Since then ACO has attracted a large number of researchers. As the theoretical framework of ACO has grown up in recent years [16–18], ACO is becoming popular, and it often gives satisfactory results for various optimization problems in a wide range of domains [19–21], such as data mining, machine learning, bioinformatics and multiple objective optimization problems. In addition, ACO plays a more and more important role in combination with other meta-heuristic mechanisms to effectively solve many NP-complete problems [22,23].

Initially, ACO was inspired by the observation of real ants looking for food. Ethnologists observed that ants can find the shortest path from their nest to the feeding food source by exploring and exploiting pheromone information, which has been deposited on the path when they traversed. They then can choose routes based on the amount of pheromone. Namely, ants communicate information about food source via pheromone, which they secrete as they move along. The larger amount of pheromone is deposited on a route, the greater is the probability of selecting the route by ants. Thus, when one ant finds a good short path from the nest to a food source, other ants are more likely to follow this path, and such a self-strengthening behavior eventually leads all the ants following the shortest path. The idea of the ACO is to mimic this behavior with artificial ants walking around the graph representing the problem to solve. While constructing the solutions, each artificial ant finds a solution starting from a start node and moving to feasible neighbor nodes step-by-step. During the process, the pheromone also evaporates over time, so that pheromone trails of infrequently traveled paths become weaker while frequently traveled paths are reinforced. Moreover, artificial ants not only imitate the learning behavior described above, but also employ problem-specific heuristic information to govern them to search towards neighbor nodes stochastically. Based on this mechanism, an effective ACO algorithm with the $K2$ metric for learning Bayesian networks, called ACO-B, is proposed in [6].

2.3.2. K2 metric

The $K2$ metric is a well-known evaluation measure for learning Bayesian networks from data, which uses a Bayesian scoring metric to measure the joint probability of a BN. The scoring metric is