

A Bayesian Network Learning Algorithm Based on Independence Test and Ant Colony Optimization

JI Jun-Zhong¹ ZHANG Hong-Xun¹ HU Ren-Bing¹ LIU Chun-Nian¹

Abstract To solve the drawbacks of the ant colony optimization for learning Bayesian networks (ACO-B), this paper proposes an improved algorithm based on the conditional independence test and ant colony optimization (I-ACO-B). First, the I-ACO-B uses order-0 independence tests to effectively restrict the space of candidate solutions, so that many unnecessary searches of ants can be avoided. And then, by combining the global score increase of a solution and local mutual information between nodes, a new heuristic function with better heuristic ability is given to induct the process of stochastic searches. The experimental results on the benchmark data sets show that the new algorithm is effective and efficient in large scale databases, and greatly enhances convergence speed compared to the original algorithm.

Key words Uncertainty modeling, Bayesian network structure learning, ant colony optimization (ACO), conditional independence test

Bayesian network (BN) is an important theory model within the community of artificial intelligence, and also a powerful formalism to model the uncertainty knowledge in practise. Recently, learning a BN structure from data has received considerable attentions and researchers have proposed various learning algorithms^[1-16]. Especially, there are three efficient approaches using the stochastic search mechanism to tackle the problem of learning Bayesian network. The first one uses genetic algorithm (GA)^[5, 7], the second one applies evolutionary programming (EP)^[8, 11, 13], and the third one employs ant colony optimization (ACO)^[6, 9].

To solve the drawbacks of the ant colony optimization for learning Bayesian networks^[6] (ACO-B), this paper proposes a Bayesian network structure learning algorithm based on the conditional independence test and ant colony optimization (I-ACO-B), which not only employs constraint knowledge to reduce the search space, but also takes it as heuristic knowledge to induct the process of stochastic searches. First, the new algorithm uses order-0 independence tests to effectively restrict the available scope of candidate arcs, reduce the space of candidate solutions, and induce ants to avoid many unnecessary searches. And then, by combining the global score increase of a solution with the local mutual information between nodes, a new heuristic function with better heuristic ability is given to induct the process of stochastic searches. The experimental results on the benchmark data sets show that the new algorithm is effective and efficient in large scale databases, and greatly enhances convergence speed compared to the original algorithm.

The paper is organized as follows. In Section 1, we present the background of Bayesian networks and the basic idea of the ant colony optimization for learning Bayesian networks. In Section 2, we describe our new algorithm in detail. Section 3 reports our experimental results. Finally, we conclude the paper in Section 4.

1 Bayesian network structure learning based on ant colony optimization

1.1 Bayesian networks

A Bayesian network (BN) can be denoted as a triple

group $\langle X, A, \Theta \rangle$, where $\langle X, A \rangle$ defines a directed acyclic graph (DAG) structure G , X is the set of nodes, $X_i \in X$ represents a random variable in a special domain; A is a set of directed arcs, $a_{ij} \in A$ describes a direct probabilistic dependency between X_i and X_j , $X_i \leftarrow X_j$; and $\Theta = \{\theta_i\}$ is a set of parameters, $\theta_i = p(X_i | \Pi(X_i))$ is the conditional probability distribution of X_i given the parent set of the variable X_i . As the graph structure G qualitatively characterizes the independence relationship among random variables, and the conditional probability distribution quantifies the strength of dependencies between a node and its parent nodes. Thus, Bayesian network $\langle X, A, \Theta \rangle$ uses a graph structure and a set of parameters to encode uniquely the joint probability distribution of the domain variables $X = \{X_1, X_2, \dots, X_n\}$:

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

1.2 Bayesian network structure learning

The structure of a BN reflects the underlying probabilistic dependence relations among the nodes (corresponding attributes of data) and a set of assertions about conditional independencies. The problem of learning a BN structure can be stated as follows: given a sample data set $D = \{X[1], X[2], \dots, X[N]\}$, 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. Actually, there are two basic realization mechanisms. The first one is an approach based on constraints^[2-3], which poses the learning process as a constraint satisfaction problem, and then constructs a network structure by testing the conditional independence relations. The second one is score-and-search approach^[1, 4-9], which poses the learning problem as a structure optimization problem. Namely, it uses a score metric to evaluate every candidate network structure, and then, finds a network structure with the best score. Though the implement of the former approach is relatively simple, the computations for high-order testings are complex and irresponsible. Moreover, the precision of learning a model is hard to ensure, thus the score-and-search approach gradually becomes a popular approach for learning Bayesian networks.

Since the parent nodes of each node in a BN, $\Pi(X_i) = \{X_k : k \in \Phi(i)\}$, are only selected from the set of nodes preceding the current node in a node ordering, namely, $\Phi(i) \subset \{1, 2, \dots, i-1\}$ (i denotes the sequence number

Received December 24, 2007; in revised form April 15, 2008
Supported by National Natural Science Foundation of China (60496322), Natural Science Foundation of Beijing (4083034), and Scientific Research Common Program of Beijing Municipal Commission of Education (KM200610005020)

1. Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology, College of Computer Science and Technology, Beijing University of Technology, Beijing 100124, P. R. China
DOI: 10.1016/S1874-1029(08)60077-4

of a node), 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 then the complexity of a BN structure space is $n! \cdot 2^{n(n-1)/2}$ for the case of a node ordering unknown. Obviously, it is intractable for the complete search based on score to find the global best solution when n is large. Recently, some researches proposed some effective algorithms^[4,10,12] with the restriction of having a complete node ordering. Unfortunately, these algorithms still perform complete searching in the worst case, so they are unfitted to learn a BN structure without a complete node ordering.

Though some improved hill-climbing algorithms^[15–16] can 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 technology has provided effective and feasible methods to tackle the problem, and genetic algorithms, simulated annealing^[14], evolutionary programming and ant colony optimization have been applied to learning Bayesian networks one after the other. These methods perform stochastically iterative search and find the global optimal solution by means of simulating various natural phenomena. In the following, we introduce the ACO-B algorithm, which is an effective ant colony optimization for learning Bayesian networks.

1.3 Learning Bayesian networks using ACO (ACO-B)

Ant colony optimization (ACO), proposed by Dorigo in 1990^{s[17–18]}, is a new meta-heuristic search algorithm, which is often used to solve combinatorial optimization problems. The mechanism is the simulation of the intelligent behaviors of real ant colonies looking for food. The frame work of ACO has gradually grown up^[19–21] for many years, and there are many successful applications in a wide range of different fields^[22], such as data mining, machining learning, and bioinformatics. ACO-B algorithm^[6] is a score-and-search approach based on the ant colony optimization for learning Bayesian networks, whose main idea is to use the $K2$ metric as a score measure $f(G : D) = \sum_{i=1}^n f(X_i, \Pi(X_i))$ to evaluate a BN structure, and induce ants to search the global maximum in a feasible solution space.

Let a be the number of ants in an ant colony, $\tau_{ij}(t)$ be the pheromone intensity associated with the directed arc a_{ij} at time t , and the initial pheromone intensity of every directed arc be a constant value C , i.e., $\tau_{ij}(0) = C$. During constructing a solution, each ant k ($k = 1, 2, \dots, a$) starts from the empty graph G_0 (arcs-less DAG) and proceeds by adding an arc at one time. The construction process of a BN for an ant is shown in Fig. 1, where the current state G_h of an ant is a graph with all nodes $X_i \in X$, exactly h arcs and no directed cycle. Suppose there are m candidate directed arcs. In terms of the pheromone and heuristic information of candidate arcs, the ant selects the s -th arc a_{ij} as a new component of a solution, thus the new state by adding an arc a_{ij} can be denoted as $G_{h+1} = G_h \cup \{a_{ij}\}$. Once there is no way to make the score of a BN structure more higher by adding an arc, the construction process is ended and the ant gets its solution G_g .

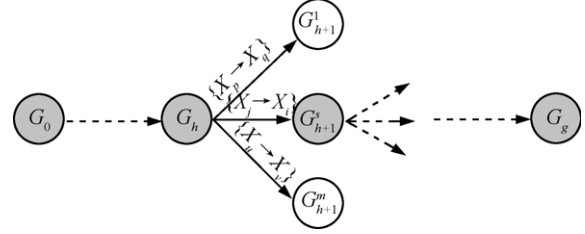


Fig. 1 The construction process of a BN for an ant

The detailed process of constructing a solution can be described as follows. At time t , the probabilistic transition rule that an ant k selects a directed arc a_{ij} from the current candidate arcs is defined as

$$i, j = \begin{cases} \arg \max_{r,l \in DA_k(t)} \{[\tau_{rl}(t)] \cdot [\eta_{rl}(t)]^\beta\}, & \text{if } q \leq q_0 \\ I, J, & \text{otherwise} \end{cases} \quad (2)$$

where $\eta_{rl}(t)$ represents the heuristic information of the directed arc a_{rl} , β is the weighted coefficient which controls $\eta_{rl}(t)$ to influence the selection of arc. $DA_k(t)$ ($r, l \in DA_k(t)$) is the set of all candidate arcs that satisfy constraint conditions and heuristic information is larger than zero, q_0 ($0 \leq q_0 \leq 1$) is an initial parameter that determines the relative importance of exploitation versus exploration, q ($q \in [0, 1]$) is a random number; I and J are a pair of nodes randomly selected according to the probabilities in (3), with $\alpha = 1$.

$$p_{ij}^k(t) = \begin{cases} \frac{[\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}(t)]^\beta}{\sum_{r,l \in DA_k(t)} [\tau_{rl}(t)]^\alpha \cdot [\eta_{rl}(t)]^\beta}, & \text{if } i, j \in DA_k(t) \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where parameter α depicts the relative importance of the pheromone $\tau_{rl}(t)$ left by the real ants. As the learning goal is to achieve the best BN structure whose $K2$ score is the maximum, the heuristic information function of a directed arc can be interpreted as the greatest increase produced in $K2$ score when the arc is added to the graph. Since the metric $K2$ is decomposable, the heuristic information function can be defined as

$$\eta_{ij}(t) = f(X_i, \Pi(X_i) \cup X_j) - f(X_i, \Pi(X_i)) \quad (4)$$

After each iteration of the ant colony is performed, ACO-B algorithm will carry out the pheromone updating process, which includes local and global updating steps. First, while building a solution, if an ant selects an arc a_{ij} , then the pheromone level of the corresponding arc is changed in the following way

$$\tau_{ij}(t+1) = (1 - \psi)\tau_{ij}(t) + \psi\tau_0 \quad (5)$$

where τ_0 is a constant related with the initial solution, $0 < \psi \leq 1$ is a parameter that controls the pheromone evaporation. And then, from all feasible solutions, the algorithm finds the best solution obtained so far by means of the $K2$ metric, and performs the global updating for each arc of the current best solution, the global updating rule is

$$\tau_{ij} = (1 - \rho)\tau_{ij} + \rho\Delta\tau_{ij} \\ \Delta\tau_{ij} = \begin{cases} \frac{1}{|f(G^+ : D)|}, & \text{if } a_{ij} \in G^+ \\ \tau_{ij}, & \text{otherwise} \end{cases} \quad (6)$$

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات