

# Analysis of the Max-Min Hill-Climbing Algorithm

Yue Wang

School of Mathematics and System Science, Shandong University of Science and Technology,  
Shandong 266590, China.

2654131436@qq.com

**Abstract.** The Bayesian network is one of the effective tools for study of knowledge representation and casual reasoning in the conditions of uncertain in the field of Artificial Intelligence. How to construct a Bayesian network structure from data have become a hot point in recent study. Max-Min Hill-Climbing (MMHC) algorithm is a newly Bayesian network structure learning algorithm. After a lot of simulation experiments, it has been corroborated that MMHC outperforms on average and in terms of various metrics several prototypical and state-of-the-art algorithms.

**Keywords:** Bayesian networks; Max-Min Hill-Climbing (MMHC).

## 1. Introduction

The Bayesian network was developed by artificial intelligence in the study of uncertainty, which was first proposed by Pearl [1] in 1988. Bayesian network learning [2-4] refers to the process of obtaining a Bayesian network by analyzing data, which includes parameter learning and structural learning.

The Bayesian network's structural learning process combines a priori information, including expert knowledge, to find the best network structure to fit the sample dataset. Since the 1990s, researchers have studied the learning problems of Bayesian networks from different angles, and proposed many classic Bayesian network learning algorithms. These algorithms can be roughly divided into three categories: based on search-and-score methods, based on dependency analysis methods and hybrid learning methods.

The Max-Min Hill-Climbing (MMHC) algorithm can be categorized as a hybrid method, using concepts and techniques from both approaches. MMHC first learns the skeleton (i.e., the edges without their orientation) of a Bayesian network using a local discovery algorithm called Max-Min Parents and Children (MMPC). It then orients the skeleton using a greedy Bayesian-scoring hill-climbing search. MMHC's skeleton identification phase is sound in the sample limit while the orientation phase does not provide any theoretical guarantees.

## 2. Bayesian Network

A Bayesian network is a directed acyclic graph with parameters, represented by a two-tuples  $\langle G, \Theta \rangle$ , where  $G = (V, E)$  represents a directed acyclic graph of relationship between nodes, called a Bayesian network structure, and node set  $V = \{X_1, X_2, \dots, X_n\}$  represents a random variable. The directed edge set  $E = \{e_{ij} \mid X_i \rightarrow X_j, i, j = 1, 2, \dots, n\}$  represents a direct dependency between variables;  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  represents the conditional probability of each node, called the Bayesian network parameter. The parameter  $\theta_i$  of node  $X_i$  represents the conditional probability distribution of its own and the set of parent nodes, ie.  $\theta_i = P(X_i \mid Pa(X_i))$ .

Any given Bayesian network satisfies the Markov condition, Therefore, the joint probability distribution of variable set  $V = \{X_1, X_2, \dots, X_n\}$  can be decomposed into the following form:

$$P(V) = P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i \mid Pa(X_i)).$$

### 3. The Max-Min Hill-Climbing Algorithm

The MMHC algorithm was proposed by Tsamardinos et al. [10] in 2006. The basic idea is to use conditional independence test to find the Parents and Children node sets of each node and reduce the search space. Then use the hill climbing search to search for the highest score in the structure space.

#### 3.1 The Max-min Parents and Children Algorithm

The max-min parents and children algorithm (MMPC) is a local causal discovery algorithm based on conditional independence testing, which returning the candidate parents and children (CPC) for each variable from a given data set using a Max-Min heuristic strategy. The max-min part of the algorithm name is reflected in the search for the candidate parents and children  $CPC_i$  of variable  $X_i$ . Under the given  $CPC_i$ , each of the variables that in the sets of  $V \setminus \{X_i, CPC_i\}$  has a minimum association value with  $X_i$ . Selecting the maximum from all minimum association values. The parents and children parts refer to its output. The result of MMPC is only the variable that has a dependency on the variable, that is, it may be its parent or its child. The specific direction of the edge needs to be determined by the scoring search of the second stage.

Through the above analysis, the specific steps of the MMHC algorithm can be described as follows:

Step1: Input the dataset  $D$ , variable set  $V = \{X_1, X_2, \dots, X_n\}$ , let  $i = 1$ .

Step2: Target variable  $X_i$ , and  $CPC_i = \phi$ .

Step3: for  $X_j \in V \setminus \{X_i, CPC_i\}$ , calculate the minimum association value of  $X_j$  and  $X_i$  under the condition  $CPC_i$  respectively, i.e.  $assoc_{X_j} = MinAssoc(X_j, X_i | CPC_i) = \min_{S \subseteq CPC_i} Assoc(X_j, X_i | S)$ , record the maximum minimum association values. Set the maximum value at the variable  $F$ , i.e.  $F = \arg \max_{X_j \in V} MinAssoc(X_j, X_i | CPC_i)$ ,  $assocF = \max_{X_j \in V} MinAssoc(X_j, X_i | CPC_i)$ .

Step4: (1) if  $assocF \neq 0$ , then  $CPC_i = CPC_i \cup F$ , go back to Step3;

(2) if  $assocF = 0$ ,  $CPC_i$  not change, the algorithm continues.

Step5: for all  $X_j \in CPC_i$  do

if  $\exists S \subseteq CPC_i$ , s. t.  $assoc(X_j, X_i | S) = 0$  then

$CPC_i = CPC_i \setminus \{X_j\}$

end if

return  $CPC_i$

end for

Step6: if  $i < n$ , let  $i = i + 1$ , and go to Step3;

else terminate algorithm.

#### 3.2 MMHC

Through the first stage identifies the parents and children set of each variable, and then performs a greedy hill-climbing search in the structure space of Bayesian networks. According to the idea of the scoring search method, the scoring function is first determined. Selecting the BDeu scoring function in the MMHC algorithm:

$$f_{BDeu}(G | D) = \log P(G) + \sum_{i=1}^n \sum_{j=1}^{q_i} \left[ \log \frac{\Gamma\left(\frac{\alpha}{q_i}\right)}{\Gamma\left(\frac{\alpha}{q_i} + N_{ij}\right)} + \sum_{k=1}^{r_i} \log \frac{\Gamma\left(\frac{\alpha}{r_i q_i} + N_{ijk}\right)}{\Gamma\left(\frac{\alpha}{r_i q_i}\right)} \right],$$

where  $n$  represents the number of variables;  $P(G)$  represents the prior probability of the structure;  $r_i$  represents the number of possible value assignments of variable  $X_i$ ;  $q_i$  represents the number of possible value assignments of  $Pa(X_i)$ ;  $N_{ijk}$  represents the number of cases in the database in which variable  $X_i$  has the value  $k$  and  $Pa(X_i)$  is instantiated as  $j$ ; a constant  $\alpha$  representing an equivalent sample size.

The greedy hill-climbing search begins with an empty graph. The edge addition, deletion, or direction reversal that leads to the largest increase in score is taken and the search continues in a similar fashion recursively. The important difference from standard greedy search is that the search is constrained to only consider adding an edge if it was discovered by MMPC in the first phase. When changes occur without an increase in the maximum score ever encountered during search, the algorithm terminates. The overall best scoring structure is then returned.

Through the above analysis, the specific steps of the MMHC algorithm can be described as follows:

Step 1: Input the dataset  $D$ , variable set  $V = \{X_1, X_2, \dots, X_n\}$ , and BDeu scoring function  $f_{BDeu}(G|D)$ .

Step 2: Call the MMPC algorithm to return the parents and children set  $CPC_i (i = 1, 2, \dots, n)$  of each variable.

Step 3: Starting from an empty graph perform Greedy Hill-Climbing with operators add-edge, delete-edge, reverse-edge. Only try operator add-edge  $Y \rightarrow X$  if  $Y \in CPC_X$ .

Step 4: Return the highest BDeu scoring, and the Bayesian network structure found.

## References

- [1]. Heckerman D, Geiger D, Chickering D. Learning Bayesian networks: The combination of knowledge and statistical data[J]. *Machine Learning*, 1995, 20(3): 194-243.
- [2]. Pearl J. Probabilistic Reasoning in Intelligent Systems[M]. Morgan Kaufmann: Network of Plausible Inference, 1988, 70(2): 1-86.
- [3]. Tsamardinos I, Brown E, Aleferis F. The max-min hill-climbing Bayesian network structure learning algorithm[J]. *Machine Learning*, 2006, 65(1): 31-78.
- [4]. Zhang Lianwen, Guo Haipeng. Introduction to Bayesian Network [M]. Beijing: Science Press, 2006.
- [5]. Wang Shuangcheng. Bayesian Network Learning, Reasoning and Application [M]. Shanghai: Lixin Accounting Press, 2010.