Hill climbing algorithm for Bayesian network structure

Cite as: AIP Conference Proceedings **2479**, 020035 (2022); https://doi.org/10.1063/5.0099793
Published Online: 26 July 2022

Ria Puan Adhitama and Dewi Retno Sari Saputro

View Online
Export Citation

© 2022 Author(s).

Hill Climbing Algorithm for Bayesian Network Structure

Ria Puan Adhitama[1, a)] and Dewi Retno Sari Saputro[1, b)]

[1]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Sebelas Maret University*
Ir. Sutami 36A Street, Surakarta, 57126, Central Java, Indonesia.

[a)] Corresponding author: ria.puan@student.uns.ac.id
[b)]dewiretnoss@staff.uns.ac.id

Abstract. Bayesian Network (BN) model is a method developed to describe causal relationships between variables in a system. BN is a form of Probabilistic Graphical Model or a simple probabilistic graph built from Bayes probability theory and graph theory. There are two approaches used to construct the BN structure, namely the constraint-based method and the score-based method. The scored-based method is a method that assigns a score to each structure in BN and tries to maximize the scoring with several heuristic search algorithms. The scoring is executed through the usage of Bayesian Information Criterion (BIC) scoring function. In this study, scored-based totally is solved through the Hill Climbing (HC) algorithm. This algorithm is a value-based algorithm in a directed graph space and includes a heuristic search method that works greedily. The results of the study show that the use of the score-based method followed by the HC algorithm and the BIC scoring function can build the BN structure.

INTRODUCTION

One of the techniques used in data classification is Bayesian Network (BN). According to Simanjuntak, BN is a model in the form of a graph that can represent the level of dependence of a random variable with its graph structure [16]. BN is one of the artificial intelligence methods that can be applied in intelligent applications such as machine learning, weather forecasting, voice recognition, signal processing, medical diagnosis, and other intelligent system applications. One of the methods like the BN method that can be used in machine learning based on training data regarding conditional probabilities. BN is one or other of the Probabilistic Graphical Model (PGM) or a simple probabilistic graph. BN is also defined as a simple probability-based data modeling method built on probabilistic theory and graph theory. Probability theory (probabilistic) is immediately correspond to the data, while graph theory is immediately correspond with the form of obtained representation. BN represents a causal relationship between the variables contained in the BN structure [12].

There are two approach methods used to construct the BN structure, namely the dependency analysis method (constraint-based) and search and scoring (scored-based). The method of dependency analysis (constraint-based) is a structure that represents a set of conditional freedoms between variables, while the search and scoring (scored-based) method is a structure that represents a joint probability distribution. Since the structure is formed, the evaluation process can be continued with the BIC scoring function. Bayesian Information Criterion (BIC) is a method of selecting the best structure from other model choices. The BIC thought became first delivered via way of means of Schwarz (1978). The Bayesian Information Criterion (BIC) or Schwarz Information Criterion additionally referred to as Schwarz's Bayesian Information Criterion or Schwarz Bayesian Criterion is a gauge for version choice amongst a constrained set of models. The version with the bottom BIC is preferred. It is primarily based totally in element at the probability characteristic and carefully associated with the Akaike data criterion (AIC).

Several types of algorithms can be used to the scored-based approach, one of which is the Hill Climbing (HC) algorithm. HC is a value-based learning algorithm in a directed graph space and belongs to a heuristic search method that works greedily. The heuristic function accustomed to valuate the situation so that the desired solution can be obtained. Therefore, in this article, the HC algorithm is discussed in building the BN structure.

The International Conference on Mathematics and Learning Research (ICOMER) 2021
AIP Conf. Proc. 2479, 020035-1–020035-7; https://doi.org/10.1063/5.0099793
Published by AIP Publishing. 978-0-7354-4359-4/$30.00

RESEARCH METHOD

This study conducted a study on build the BN structure using the score-based method followed by the HC algorithm. The research method used in this study is a literature review of reference books, journals, and writings on BN, score-based method, BIC scoring function, and HC algorithm.

MATERIALS AND METHODS

Bayesian Network

Thomas Bayes all over in 1950, found the Bayes theory. The theory of probability conditions that issue into bill the probability of a causa depending on other occurrences [1],[11],[13]. According to Saputro et al., the concept states that destiny occurrences may be expected at the situation that preceding occasions have occurred [1]. In 1988, Judea Pearl victimised Bayes' theory to get the probability of an event beingness artificial by another case with a graph, which is now known as the Bayesian Network (BN) [18]. BN consists of nodes that represent random variables and edges that represent the relationship of dependence between random variables [22].

BN is one method for unpredictability event [8]. This method uses probability theory as a formal framework for unpredictability event in Artificial Intelligence. BN is also referred to as a simple probability-based data modeling method built on probabilistic theory and graph theory. The probabilistic theory is immediately correlated to data as the same time as graph theory is immediately correlated to the shape of illustration to be obtained.

BN is a probabilistic graphical version that represents a sequence of variables and the connection among variables. The connection among these variables shows the probability of the relationship between related and unrelated events [19]. According to Heckerman, BN contain of two main parts, explained as [8]

1) The BN graph structure is called the Directed Acyclic Graph (DAG), that is a directed graph with out a directed cycle. DAG include nodes and edges. Nodes constitute random variables and edges constitute a right away dependency courting and also can also be interpreted as a right away effect (cause and effect) among variables have been been connected. The absence of an edge indicates a conditional freedom relationship among variables. This DAG can represent knowledge about an uncertain domain. Specifically, every node in the graph is a random variable, while the ends among nodes represent the probabilistic dependence of the corresponding random variables. This dependency condition in the graph is often estimated using statistics known as computational methods. Therefore, BN combines principles from statistics, computer science, probability theory, and graph theory.
2) The set of parameters defined a conditional probability distribution for each variable which the nodes are correspond to a random variable. Each node is connected to a set of conditional probabilities, $P(x_i|A_i)$ with x_i is the variable associated with the node and A_i is the set of parents in the graph.

BN is built using a statistical approach known as Bayes' theorem. Conditional probability is the reckoning of the probability of a phenomenon A if it is investigated that phenomenon B has happened is indicated by $P(A|B)$. This theorem is used to reckon the probability of a data set a certain class based on the inference of existing data. If it is associated with the probability of the disease symptom, B is the symptom of the disease and A is the type of the disease. Bayes theory formula is written as

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

or it cold be rewrite as

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}$$

with :

$P(A|B)$: posterior probability, that is the probability Y occurs after probability X has occurred
$P(B|A)$: likelihood, that is the probability X occurs after probability Y has occurred
$P(A)$: prior, that is the probability of the event Y
$P(B)$: that is the probability of the event X

BN also performs probabilistic decision-making (inference). Probabilistic inference is a step to expecting the cost of an unknown variable immediately with the aid of using the use of the values of different regarded variables. Probabilistic inference can be performed if first acquired Joint Probability Distribution (JPD) of all variables that occur simultaneously. JPD is the chance of all variable activities taking place simultaneously. Probabilistic inference may be accomplished if the BN has constructed so that the first thing to do is build the BN structure based on obtained incident data.

BN consists of (i) a structure known as a directed acyclic graph (DAG) in which every node represents a random variable and edges constitute structured probabilistic dependencies related to the Markov conditions; and (ii) series of conditional probability distributions described for every variable given its parent withinside the graph [4],[5]. Due to the character of this graphic, BN is right to be used in complicated probabilistic relationships current in lots of real-globabl matters.

Build A Bayesian Network Structure

After studying the basic concepts in building a BN structure, then in building a BN structure, there are two stages, as follow

1) Structural production or additionally called the qualitative stage, that is searching out the connection among the variables being modeled. The BN structure is a DAG that can represent a pattern from a set of data. Representation in the form of graphs are doing by identifying information concepts that are appropriate to the matter. Furthermore, these concepts are known as the set of variables. Then, this set of variables is represented as nodes in the graph. The relationship between variables is stated explicitly with the edges contained in the graph.
2) Parameter estimation or known as the quantitative stage, that is calculating chance values. After the BN structure is formed, parameters and dependency relationships between nodes are determined using expert knowledge.

The approach in Building Bayesian Network Structure

According to Sitohang and Saptawati, there are two approaches in construct the BN structure, namely (i) the scored-based method [17]. In this method, BN is considered as the most suitable variable probability distribution for the data; and (ii) the constraint-based method. This method is used to locate the BN shape through figuring out the conditional independence courting among variables. This method considers BN as a set of conditional freedom relationships among variables. The constraint-based method is a method that studies the network structure by analyzing the probability relationship possessed with the Markov property of BN testing conditional independence and then building a graph structure that fits the D-separation concept. The ensuing model on is frequently interpreted as a causal model even if observe with observational facts as in [9],[14]. On the other hand, the scored-based method is a method that assigns a rating to every candidate in BN and attempts to maximise the scoring by several heuristic seek algorithms. According to Scutari et al. when using real-world data with simulation data, it appears that the constraint-based method is neither greater green nor greater touchy to mistakes than the scored-based method [15]. Thus, it is better to use the scored-based method to build the BN structure.

AIC And BIC Scoring Function

According to Carvalho, determining the quality measure of BN should be calculated in numerous one of a kind ways [7]. This ends in the generalization of the Minimum Description Length (MDL) scoring feature which is written as follows

$$\emptyset(B|T) = LL(B|T) - f(N)|B|$$

with $f(N)$ is a non-negative penalization function. If $f(N) = 1$, then the Akaike Information Criterion (AIC) scoring function can be used

$$AIC(B|T) = LL(B|T) - |B|.$$

If $f(N) = \frac{1}{2}\,log(N)$, so Bayesian Information Criterion (BIC) scoring function can be used, which coincides with the MDL score. If $f(N) = 0$, so LL score was obtained.

Bayesian Information Criterion (BIC) is a way of choosing the best version from several other model choices. The BIC theory turned into first delivered through Schwarz (1978), this BIC makes use of nomerous Bayesian theorems in order that BIC is likewise referred to as Schwarz's Bayesian Information Criterion (SBIC) and a few additionally name it the Schwarz Bayesian Criterion (SBC). When becoming a version, it's far feasible to boom the opportunities through including parameters, however doing so may also bring about overfitting. Both BIC and AIC try and clear up this hassle through introducing a penalty time period for the variety of parameters withinside the version. Punishment terms are greater in BIC than in AIC. When choosing the best model with this BIC method, the model with the smallest BIC value is selected. The smaller the BIC value, the better the model obtained.

Hill Climbing Algorithm

Hill Climbing (HC) is one of the methods used to resolve the nearest search problem [20]. The manner it really works to decide the subsequent step via way of means of setting the factor as a way to appear as near as feasible to attain the target. The checking out system is executed the use of heuristic functions. The technology of subsequent states is highly dependending on remarks from the test procedure. The check withinside the shape of a heuristic feature will display how properly the wager fee is taken in competition to different feasible conditions [21]. Two types of HC are slightly different, namely Simple Hill Climbing and Steepest-Ascent Hill Climbing.

According to Suyanto, HC is often used if there is a good heuristic function to evaluate the state. Two types of HC are slightly different, namely Simple Hill Climbing (simple HC) and Steepest Ascent Hill Climbing (HC with the sharpest/steepest slope) [6]. Simple Hill Climbing simply chooses a new state that has a better ("steep") path than the previous paths without taking into other paths that are more "steep". While Steepest Ascent Hill Climbing, as the name implies, will evaluate all states that are under the current state and choose the state with the most "steep" path.

In generate the closest search path using the HC method, nodes or points are needed to compare the distance. The states are obtained from paths that have two or more path intersections. The previous states are stored first so that they can provide the level of accuracy of the path. In the implementation process, this method requires the location of each user and state to be used as a path to compare the distance. The node distance that obtained will always be compared with other states until it gets the best value from other states. If in the process of finding the best state from the previous one, the position of the initial state will change. The process will be carried out until the occupied node is the destination node. The state that has passed will not be read again.

The Hill Climbing (HC) algorithm is the only score-based learning algorithm that performs a greedy search on a unidirectional graph space as in [9],[14]. Optimization in implementing with the aid of using default the use of rating caching, rating decomposability, and rating equivalence to lessen the quantity of duplicated tests [14]. One of the advantages of using the HC algorithm it can be used in almost all search procedures [9],[14]. The rating characteristic is on the whole rating-equivalent, in order that networks having the identical chance distribution are assigned the identical rating [2],[14]. According to Choon and Chandrasekaran, HC is a value-based learning algorithm that is commonly used in directed graph space [3]. HC is also used to find local optima and to improve the results of better structures.

The algorithm for Simple Hill Climbing Search is described as

1. Evaluated preliminary state. If this state is a purpose state, then go back this state as a solution and exit the program. If this state is not a purpose state, continue the process with the preliminary state as the current state.
2. Repeat until a solution is found or until no new operator (production rule) can be applied to the current state
    a. Select an operator that has now no longer been carried out to the current state and observe the operator to provide a new state.
    b. New state evaluation.
        i. If this state is a purpose state, then return this state as a solution and exit the program.
        ii. If this state is not always a purpose state however better than the current state, then make this state the current state.
        iii. If this state is not better than the current state, return to step 2a.

Scored-Based Method With BIC-Hillclimb Algorithm

Search and scoring technique (score-based). In this technique, a great graph shape is searched from the statistics used. A shape constructed from a aggregate of present nodes. The seek become executed with the aid of using the hunt technique, then keep with the scoring feature in its evaluation. The creation method is executed iteratively, beginning with a graph with out edges, observed with the aid of using a seek technique to feature edges, and prevent then there's no new shape this is better than the preceding shape.

The process of assigning a score to the BN structure according to Margaritis is written as [10]

$$Score(\mathcal{G},\mathcal{D}) = Pr(\mathcal{G}|\mathcal{D}) \tag{1}$$

The score-based method is used to maximize the score with $\mathcal{G}$ being a directed acyclic graph (DAG) from BN and $\mathcal{D}$ is the main data set. Equation (1) can be changed to a more suitable form by using Bayes' law, which is written as

$$Score(\mathcal{G},\mathcal{D}) = Pr(\mathcal{G}|\mathcal{D}) = \frac{Pr(\mathcal{D}|\mathcal{G})Pr(\mathcal{G})}{Pr(\mathcal{D})} \tag{2}$$

To maximize the result of equation (2), it is necessary to maximize the value of the numerator, because the denominator does not depend on $\mathcal{G}$. Here is an illustration of the BN structure with the HC algorithm as shown in Figure 1.

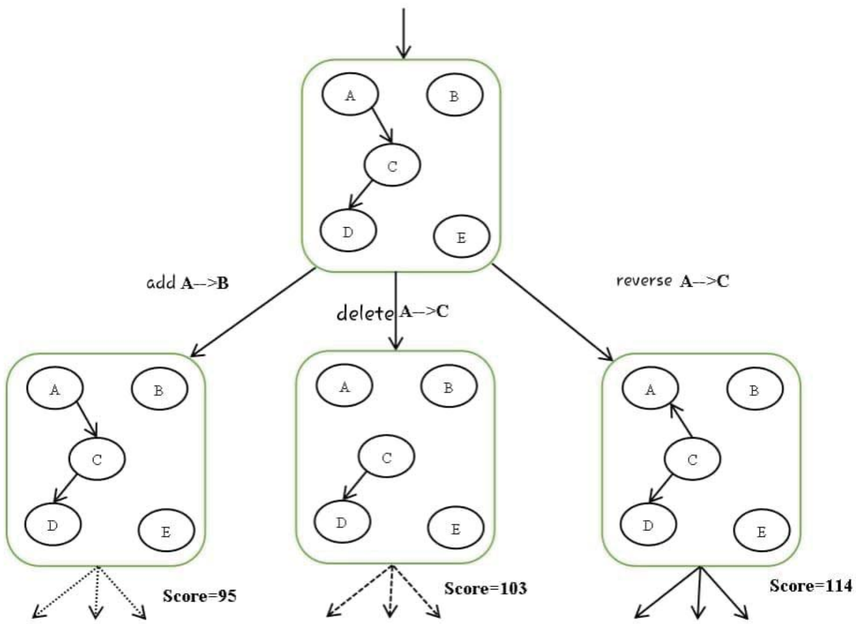

FIGURE 1. BN Structure With HC Algorithm

Based on Figure 1, dissimilar scores result in different BN structures. After that, it is continuing by calculative the BIC scoring function. The BIC formula is written as

$$BICscore(\mathcal{G},\mathcal{D}) = \log\Pr(\mathcal{D}|\hat{\boldsymbol{p}}\mathcal{G}) - \frac{d}{2}\log N$$

According to Margaritis [10], the hillclimb algorithm on a data set $\mathcal{D}$ is described as shown

1. $\mathcal{E} \leftarrow \emptyset$
2. $\mathcal{T} \leftarrow$ Conditional Probability Tables (E, $\mathcal{D}$)
3. $\mathcal{B} \leftarrow \langle \mathcal{U}, \mathcal{E}, \mathcal{T} \rangle$
4. score $\leftarrow -\infty$
5. do calculation on :
   a. maximum score $\leftarrow$ score

b. to each set (X, Y) continue step

c. to each $\mathcal{E}' \in \{\mathcal{E} \cup \{X \to Y\},$
$\mathcal{E} - \{X \to Y\},$
$\mathcal{E} - \{X \to Y\} \cup \{Y \to X\}\}$

d. $\mathcal{T}' \leftarrow$ Conditional Probability Tables $(\mathcal{E}', \mathcal{D})$

e. $\mathcal{B}' \leftarrow \langle \mathcal{U}, \mathcal{E}', \mathcal{T}' \rangle$

f. newscore $\leftarrow$ BICscore $(\mathcal{B}', \mathcal{D})$

g. in case newscore > score, so

$\mathcal{B} \leftarrow \mathcal{B}'$

score $\leftarrow$ newscore

6. in case score > maximumscore

7. return at step 3

at same time $\hat{\boldsymbol{p}}$ is the set of maximum likelihood estimates connate the BN structure, $\boldsymbol{d}$ is the parameter of the Gaussian multivariate, $\mathcal{E}$ is the collection of edges in the BN, $\mathcal{T}$ is the collection of parameters, $\mathcal{B}$ is the BN containing the DAG and the parameters.

CONCLUSION

Supported at the consequences of the discussion , it could be concluded that BN is one of the Probabilistic Graphical Model (PGM) or a easy probabilistic graph that's describe as a easy probability-based data modeling approach construct from probabilistic theory and graph theory. The HC algorithm is one of the algorithms that could be used to continue the score-based method. In the cognitive manner appoint of the scored-based adjustment that's so observed through the HC algorithm with the BIC scoring function it could be used to buid the BN structure.

ACKNOWLEDGMENTS

The authors would like to thank Mathematical Soft Computing Research Group for their moral support and incisive comments to improve this article. The research described in this paper is supported by a fundamental grant from the Research and Community Service Institute of Universitas Sebelas Maret through a letter of agreement for the implementation of the research implementation of the non-APBN Fund No. 260 / UN27.22 / HK.07.00 / 2021.

REFERENCES

1. D. R. S. Saputro, P. Widyaningsih, F. Handayani, and N. A. Kurdhi, *Prior and Posterior Direchlet Distributions On Bayesian Networks (BNs),* AIP Conference Proceedings 1872, 020036 (2017).
2. D. M. Chickering, *A Transformational Characterization of Equivalent Bayesian Network Structures*. University of California. Los Angeles.
3. O. H. Choon, and K. Chandrasekaran, *A Bayesian Network Approach to Identify Factors Affecting Learning of Additional Mathematics*. Jurnal Pendidikan Malaysia, 40(2), 185-192 (2015)
4. A. H. C. Correia, J. Cussens, C. P. Campos, *On Pruning for Score-Based Bayesian Network structure Learning*.
5. J. Cussens, *Bayesian Network Learning With Cutting Planes,* University of York, Deramore Lane. UK.
6. Suyanto, *Artificial Intelligence : Searching, Reasoning, Planing, Learning,* Informatika, Jakarta (2014).
7. A. M. Carvalho, *Scoring Function for Learning Bayesian Networks,* INESC-ID Tec. Rep 54 (2009).
8. D. Heckerman, *A Tutorial on Learning With Bayesian Networks*. Redmond WA 98052 Microsoft Research : Advanced Technology Division (1996).
9. T. A. Kung, and M. S. Mohamad, *Using Bayesian Network to Construct Gene Regulatory Networks from Microarry Data*. UTM Jahar Bahru : Jurnal Teknologi (2012).
10. D. Margaritis, *Learning Bayesian Network Model Structure from Data,* University of Carnegie Mellon. Pittsburgh (2003).
11. J. M. Bernando and A. F. M. Smith, *Bayesian Theory,* Wiley, New York, (1994).

12. R. E. Neapolitan, *Learning Bayesian Network*. University of Northeastern Illinois University. Chicago (2004).
13. J. O. Berger, *Statistical Decision Theory and Bayesian Analysis,* 2nd Edition, Springer-Verlag, New York, (1985).
14. M. Scutari, *Learning Bayesian Networks With The Bnlearn R Package*. University of Padova, 35(3), (2010).
15. M. Scutari, C. E. Graafland, and J. M. Gutierrez, *Who Learns Better Bayesian Network Structures : Constraint-Based, Score-Based or Hybrid Algorithms?* . Proceedings of Machine Learning Research, 72, pp : 416-427 (2018).
16. H. T. A. Simanjuntak, *Pengembangan Algoritma CB Untuk Konstruksi Struktur Bayesian Network Dari Data Tidak Lengkap,* SESINDO : Jurnal Sistem Informasi ITS, pp : 71-77 (2010).
17. B. Sitohang and G. A. P. Saptawati, *Improvement of CB & BC Algorithms (CB* Algorithm) for Learning Structure of Bayesian Networks as Classifier in Data Mining,* Jurnal ICT Institute Teknologi Bandung, 1(1), 29-41 (2007).
18. J. Pearl, *Probabilistic reasoning in Intelligent Systems : Networks of Plausible Inference*, pp : 505-507 (1988).
19. K. W. Przytula and D. Thompson, *Construction of Bayesian Network for Diagnostic. IEEE,* (2000).
20. Russel and Norvig, *Artificial Intelligence : A Modern Approach,* Prentice Hall, USA (2003).
21. S. Kusumadewi, *Artificial Intelligence (Teknik & Aplikasinya),* Yogyakarta : Graha Ilmu (2003).
22. G. Cooper and E. Herskovits. "*A Bayesian Method for The Induction of Probabilistic Networks from Data*". Machine Learning, 9:309–347, 1992. URL http://smi-web.stanford.edu/ pubs/SMI_Reports/SMI-91-0355.pdf. Appeared as 1991 Technical Report KSL-91-02 for the Knowledge Systems Laboratory, Stanford University, SMI-91-0355.