IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 29 November 2022, accepted 8 December 2022, date of publication 14 December 2022,
date of current version 20 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3229128

RESEARCH ARTICLE

A Bayesian Network Structure Learning Algorithm Based on Probabilistic Incremental Analysis and Constraint

HAORAN LIU[1], SHAOPENG CUI[1], SHENG LI[1], NIANTAI WANG[1], QIANRUI SHI[1], YANBIN CAI[1], LIYUE ZHANG[1], AND DAYAN LIU[2]

[1]Information Science and Engineering Department, Yanshan University, Qinhuangdao 066004, China
[2]School of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China

Corresponding author: Shaopeng Cui (csp15032358023@163.com)

This work was supported by the National Key Research and Development Program of China under Grant 2019YFB1707301.

ABSTRACT To address the problem of low efficiency of the existing hill-climbing algorithm in Bayesian network structure learning, this paper proposes a Bayesian network structure learning algorithm based on probabilistic incremental analysis and constraints. The algorithm constructs a suitable measure for representing the degree of node association in Bayesian networks based on the principle of random forest feature extraction; then uses the method to construct the initial Bayesian network structure and constrains the search space by setting a corresponding threshold for the probability increment centered on each node; finally takes the initial Bayesian network as the starting point and learns it by the forbidden hill-climbing search and BIC scoring method to obtain the optimal Bayesian network structure. Experimental results show that the correlation degree measure and mutual information proposed in this paper have an approximate correlation expression effect; compared with other Bayesian network structure learning algorithms of the same type, the method in this paper has a faster operation speed while ensuring the quality of the learned network. The experimental results show that the Bayesian network structure learning algorithm based on probabilistic incremental analysis and constraints is an effective and efficient Bayesian network structure learning algorithm.

INDEX TERMS Probability increment, Bayesian network structure learning, constrained search space, Hill-climbing algorithm.

I. INTRODUCTION

The interconnection between different things is the basis for nature to become a whole, and the interconnected nature of each element in the data set is the basis for forming a whole system. Therefore, studying the correlation between things belongs to the basic content of scientific research. In recent years, data analysis has been widely used in countless fields such as financial analysis [1] and medical research [2]. It can be seen that data analysis has penetrated all aspects of people's lives. Correlation is an important indicator in data analysis, which can present the degree of correlation between two different events in numerical form. So far, many correlation metrics have been proposed, such as mutual information (MI) estimation [3], [4], maximum correlation coefficient [5], and maximum information coefficient [6]. Correlation analysis methods are also often used in the process of learning Bayesian network structure and have good results. For example, use the chi-square test, MI to establish an initial network, use conditional independence test to establish the parent-child node set, etc., and then combine with search algorithms, such as hill-climbing algorithm, genetic algorithm [7], particle swarm algorithm [8], etc. to build a complete Bayesian network structure.

Bayesian Networks (BN), also known as Belief Network or Directed Acyclic Graphical Model (DAGM), consists of a Directed Acyclic Graph (DAG) and the corresponding conditional probability table, the nodes in the directed acyclic

The associate editor coordinating the review of this manuscript and approving it for publication was Alberto Cano.

graph represent random variables, the directed edges between nodes represent the dependencies relationships between nodes, the direction of the edges is from the parent node to the child node, the two nodes The strength of the relationship between the nodes is expressed by the conditional probability, and the information of the nodes without parent nodes is expressed by the prior probability. In recent years, Bayesian networks have been applied in many fields such as causal analysis [9], artificial intelligence [10], fault diagnosis [11], [12], and medical research [13], [14]. At present, Bayesian network structure learning has been proven to be an NP (Non-deterministic Polynomial Hard, NP-hard) problem. For this reason, many scholars have been concerned about how to learn the correct Bayesian network structure according to the data set. An in-depth study was carried out on the issue.

Tsamardinos et al. [15] proposed the Max-Min Hill-Climbing (MMHC) algorithm, which first uses the conditional independence criterion to obtain a network skeleton, and then obtains the final network through hill climbing to narrow the search space. The method improves the efficiency of the algorithm, but when the conditional independence is applied, its complexity will increase exponentially with the increase of the number of nodes, which makes its effect in large networks not ideal; Gámez et al. [16], [17] Constrained Hill Climbing (CHC) algorithm, Fast Constrained Hill Climbing (Fast-CHC) algorithm and their improved algorithms, these algorithms limit the subsequent search space by dynamically changing candidate structure sets to improve the search rate, but are limited by the Due to the defects of the constraint method itself, these algorithms are easy to exclude the correct structure from the candidate structure, which leads to the reduction of the accuracy of the algorithm; Liu et al. [18] proposed an improved hill-climbing algorithm (Simplify Hill-Climbing, SHC), The algorithm builds the maximum weight spanning tree (MWST) through MI, and then finds the optimal Bayesian network structure through the combination of the hill-climbing algorithm with only edge addition and edge rotation operations and the BIC score. Since there is no edge reduction operation in this algorithm, the final result is prone to multilateral situations, moreover, the algorithm uses prior knowledge for initial network orientation; Liu et al. [19] proposed a hill-climbing algorithm based on the combination of V-structure & log-likelihood function orientation and tabu search. This algorithm greatly improves the performance indicators of the hill-climbing algorithm, but it uses MI. The method of establishing the MWST forces all nodes to be connected as a whole in the Car network, and there will inevitably be at least two redundant edges in the initial network, which brings unnecessary trouble to the subsequent mountain climbing search.

To improve the efficiency of the existing Bayesian network structure learning algorithm while ensuring the accuracy of the hill climbing algorithm, this paper proposes a Hill Climbing Algorithm Based on Probabilistic Incremental Analysis and Constraints (PHC). The main research contents of this paper are as follows:

Based on the principle of random forest feature selection, a correlation measure with good correlation capability and suitable for learning the Bayesian network structure is established. The average probability increment percentage (APIP) of each state between node pairs is obtained by changing the state of one node and observing the probability change of each state of other nodes through the method of control variables, and the value is used as an index to measure the degree of association between nodes.

The PHC algorithm is constructed in three steps. Firstly, the MWST [18], [20] is constructed based on the APIP values among the nodes in the data set. Then, a threshold is set for the APIP values to constrain the search range. Finally, a modified hill-climbing algorithm and a forbidden search strategy are used as the starting point to construct the structure of the ex-optimal Bayesian network.

In this paper, we experimentally demonstrate that APIP is more suitable for constrained Bayesian network structure learning algorithms than the currently used MI, and also demonstrate that the PHC algorithm has comparable or even higher rates of effectiveness than other improved hill climbing algorithms.

II. RESEARCH ON PHC ALGORITHM

A. THE STRUCTURE OF BAYESIAN NETWORKS AND THE INPUT AND OUTPUT OF THE PHC ALGORITHM

The Bayesian network is a directed acyclic structure graph, usually denoted by $G = (\boldsymbol{V}, \boldsymbol{E})$. $\boldsymbol{V}$ denotes the set of all nodes in the network, $\boldsymbol{V} = (X_1, X_2, \cdots, X_n)$, and $\boldsymbol{E}$ denotes the set of all directed edges in the network, each edge starts from one node and points to another node.

The PHC algorithm we study is applicable to the learning of Bayesian network structures under discrete variables. Therefore, we define an $n \times m$ matrix $DATA$ as the input data, $DATA$ can be expressed as a combination of m column vectors with the expression $DATA = \left(\boldsymbol{X}_1^T \ \boldsymbol{X}_2^T \ \cdots \ \boldsymbol{X}_m^T\right)$, where each column vector is a complete data with the expression $\boldsymbol{X}_k^T = \left(x_{1k} \ x_{2k} \ \cdots \ x_{nk}\right)^T$, among them, $k \in (1, 2, \cdots, m)$, x_{ik} represents the state value of the node X_i in the k th set of data, $i \in (1, 2, \cdots, n)$. x_{ik} is a state variable with t possible state values, that is, the values of x_{ik} are discrete.x_{ik} takes the value shown in equation (1):

$$x_{ik} = \begin{cases} s_1, p_1 \\ s_2, p_2 \\ \vdots \\ s_t, p_t \end{cases} \tag{1}$$

among them, s_l represents the state value that x_{ik} ($l \in (1, 2, \cdots, t)$) may take, and p_l represents the probability that node i is in state s_l. In summary, the expression of $DATA$ is

shown in Equation (2).

$$\text{DATA} = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} \tag{2}$$

among them, x_{ij} is the value that has been determined.

The output results in a Bayesian network structure, usually represented by a matrix named DAG. The size of the DAG is determined by the number of nodes in the Bayesian network structure. If the number of nodes in the Bayesian network is n, the size of the corresponding DAG is $n \times n$. The only elements in the DAG are the numbers 0 and 1, and its expression is shown in formula (3):

$$\text{DAG} = \begin{pmatrix} g_{11} & \cdots & g_{1n} \\ \vdots & \ddots & \vdots \\ g_{n1} & \cdots & g_{nn} \end{pmatrix} \tag{3}$$

among them,$i, j \in (1, 2, \cdots n)$. if $g_{ij} = 0$ and $g_{ji} = 0$, it means that there is no connection between node X_i and node X_j, that is, there is no edge connection between node X_i and node X_j; if $g_{ij} = 1$ and $g_{ji} = 0$, it means that node X_i is The parent node of node X_j, that is, there is an edge connection between node X_i and node X_j, and the direction of the edge is from node X_i to node X_j; if $g_{ij} = 1$ and $g_{ji} = 1$, it means that there is an edge connecting node X_i and node X_j, But this edge has no direction.

B. FORMULA DERIVATION OF APIP

Commonly used methods to measure the correlation between two nodes are MI, conditional entropy, chi-square test, etc. This paper proposes a new correlation measurement method APIP. The calculation principle is shown in formula (4):

$$\text{PI}(X, Y) = \frac{|P(X) - P(X|Y)|}{P(X)} \tag{4}$$

among them, PI_{XY} represents the probability increment of event X when event Y is known; $P(X)$ represents the probability of event X occurring; $P(X|Y)$ represents the probability of event X when event Y has already occurred the probability of the event occurring. The value of PI_{XY} can reflect the degree of influence of event Y on event X, and the degree of influence is positively correlated with the value of PI_{XY}. Equation (4) can be simplified to Equation (5):

$$\text{PI}(X, Y) = \left|1 - \frac{P(X|Y)}{P(X)}\right| \tag{5}$$

similarly, the calculation formula of PI_{YX} can be obtained as shown in formula (6):

$$\text{PI}(Y, X) = \left|1 - \frac{P(Y|X)}{P(Y)}\right| \tag{6}$$

according to Bayes' theorem

$$P(Y|X)P(X) = P(X|Y)P(Y) \tag{7}$$

we can get:

$$\frac{P(X|Y)}{P(X)} = \frac{P(Y|X)}{P(Y)} \tag{8}$$

substitute equation (11) into equation (8) to get:

$$\text{PI}(Y, X) = \left|1 - \frac{P(Y|X)}{P(Y)}\right| \tag{9}$$

therefore:

$$\text{PI}(X, Y) = PI(Y, X) \tag{10}$$

It can be seen that the correlation obtained by the probability incremental analysis method is symmetrical, and this property is consistent with the results obtained by other correlation calculation methods. The results are consistent.

If each node of the Bayesian network is regarded as an event, then occurrence and non-occurrence are all its states. Equation (5) is sufficient for correlation analysis, but in practical application analysis, the state of the node There are often more than two. Therefore, to facilitate the calculation, the formula (5) needs to be extended to be used for the correlation calculation of nodes with more states.

Definition 1: The possible states of the node X are $x_1, x_2, \cdots, x_{t_1}$, Y node may be in the state of $y_1, y_2, \cdots, y_{t_2}$, There are N state combinations between nodes X and Y. The calculation formula of N is shown in formula (11):

$$N = t_1 t_2 \tag{11}$$

in this case, the degree of association between nodes X and Y should be expressed in the form of an average probability increment percentage, its calculation formula is shown in formula (13):

$$\text{APIP}(X, Y) = \frac{\sum_{i=1}^{t_1}\sum_{j=1}^{t_2}\left|1 - \frac{P(x_i|y_j)}{P(x_i)}\right|}{N} \times 100\% \tag{12}$$

among them, $X \neq Y$, x_i indicates that node X is in the ith state ($i \in (1, 2, \ldots, t_1)$); y_j indicates that node Y is in the jth state ($j \in (1, 2, \ldots, t_2)$), $APIP_{XY}$ represents the average of probability increment percentage between node X and node Y.

From formula (13), it can be determined that $APIP_{XY}$ takes values in the range$[0. + \infty)$, and only when $APIP_{XY}$ is equal to zero, node X and node Y are independent of each other. The above are the same theoretical properties as MI, and the calculation of MI is shown in formula (14) – (16):

$$H(X) = -\sum_{i=1}^{t_1} P(x_i) \log P(x_i) \tag{13}$$

$$H(X, Y) = -\sum_{i=1}^{t_1}\sum_{j=1}^{t_2} P(x_i, y_j) \log P(x_i, y_j) \tag{14}$$

$$MI(X, Y) = H(X) + H(Y) - H(X, Y) \tag{15}$$

Among them, $H(X)$ denotes the conditional entropy of variable X and $H(X, Y)$ denotes the joint entropy of variable X and variable Y.

Analyzed from the point of view of the principle, MI calculates the magnitude of information between two variables, which can also be equated to the Kullback-Leibler Divergence (D_{KL}) of the joint and marginal distributions between two variables, which is calculated in formula (17):

$$\mathrm{MI}(X, Y) = D_{KL}(P(X, Y) \| P(X) P(Y)) \tag{16}$$

in contrast, APIP measures the extent to which a change in one variable causes a change in the probability distribution of another variable.

Combining formulas (14), (15), and (16) yields a formula equivalent to equation (16), the expression of which is shown in formula (17):

$$\mathrm{MI}(X, Y) = \sum_{i=1}^{t_1} \sum_{j=1}^{t_2} P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i)P(y_j)} \tag{17}$$

equation (13) can also be replaced by an equivalent formula as shown in formula (18):

$$\mathrm{APIP}(X, Y) = \left(\frac{1}{N} \sum_{i=1}^{t_1} \sum_{j=1}^{t_2} \left| 1 - \frac{P(x_i, y_j)}{P(x_i)P(y_j)} \right| \right) \times 100\% \tag{18}$$

let

$$v_{ij} = \frac{P(x_i, y_j)}{P(x_i)P(y_j)} \tag{19}$$

from the meaning of APIP (APIP indicates the probability change of one variable after the change of another variable) combined with formula (18), it can be concluded that after the variables X and Y are determined, the probability change of variable Y after the change of variable X is correlated with v_{ij}. APIP is calculated by summing all the v_{ij} and taking the mean value to indicate the correlation degree of X and Y. However, MI puts all the v_{ij} into a non-linear change function ($\log v_{ij}$) and then multiplies it by a value ($P(x_i, y_j)$) that varies with i and j before summing them, which has disrupted the original law of change transmission from the principle. Therefore, by ordering the MI values between a node and other nodes from largest to smallest, the resulting order of nodes no longer has the regularity of influence from largest to smallest. But APIP retains this regularity intact.

C. SCORING FUNCTION

The commonly used scoring functions for Bayesian network structure learning are AIC [21], BIC [22], and MDL [23]. Among them, AIC and BIC are relatively similar. Usually, the definitions of AIC and BIC are shown in equations (21) and (22):

$$AIC = 2K - 2\ln(L) \tag{20}$$

$$BIC = K\ln(n) - 2\ln(L) \tag{21}$$

among them, k represents the number of parameters of the model, L represents the likelihood function, n represents the number of samples in the data set, and the first half of the right side of the equation is the penalty term of the scoring function. From the perspective of punishment, BIC considers the number of samples, and its punishment is larger than that of AIC, which can prevent the model from being too complex due to too many samples.

The BIC scoring function is also known as the Bayesian Information Criterion. Its calculation formula is shown in formula (23):

$$\mathrm{BIC}(G|D) = \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{t_i} m_{ijk} \lg \frac{m_{ijk}}{m_{ij}} - \frac{q_i (t_i - 1)}{2} \lg m \tag{22}$$

among them, n represents the number of nodes in the network, q_i represents the total number of parent node state combinations of node X_i, t_i represents the total number of all values of the variable X_i, m is the total number of samples, m_{ijk} is the data set D that satisfies $X_i = k$ and the parent The node state combination is the number of samples of $m_{ij} = \sum_{k=1}^{r_i} m_{ijk}$.

The BIC scoring function can effectively evaluate the learned Bayesian network structure, and the results are presented in the form of negative values. The higher the score, the better the fit between the structure and the data.

D. GENERATE INITIAL NETWORK

The process of learning the Bayesian network structure by the PHC algorithm requires first obtaining an initial network G consisting of undirected edges by correlation degree analysis, which is expressed in the same form as the DAG.

First, the APIP values of each pair of nodes in the input data are calculated by equation (13) and put into the matrix W. The expression of W is shown in formula (23):

$$W = \begin{pmatrix} w_{11} & w_{21} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nm} \end{pmatrix} \tag{23}$$

Among them, W denotes an $n \times n$ matrix and w_{ij} denotes the APIP value of the node X_j and the node X_i. To facilitate subsequent calculations, when $i = j$, let $w_{ij} = 0$.

Then, the values in W are used as weights to construct MWST and obtain the matrix G. Because the subsequent process of learning Bayesian network structure is equivalent to the process of continuously changing the values of elements in G. After the learning of Bayesian network structure is finished G becomes DAG, so the expression of G is the same as that of DAG, and the meanings of elements in it are also the same.

E. INITIAL NETWORK ORIENTATION METHODS

Orientation of G is required before entering the hill-climbing search phase, and since the V-structure orientation method is more accurate than the unilateral orientation method [20], we first use the V-structure orientation method to orient G.

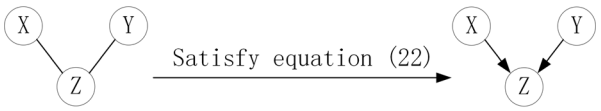


FIGURE 1. Schematic diagram of V-structure orientation.

TABLE 1. Comparison of relevant elements in G before and after V-structure orientation satisfying equation (22).

Before determining the direction	After determining the direction
$g_{XZ}=1$	$g_{XZ}=1$
$g_{ZX}=1$	$g_{ZX}=0$
$g_{YZ}=1$	$g_{YZ}=1$
$g_{ZY}=1$	$g_{ZY}=0$

The basic principle of the V-structure orientation method is shown in formula (25), and the undirected connection diagrams of nodes X, Y, and Z are shown in Figure 1:

$$\mathrm{APIP}(X, Y|Z) > \mathrm{APIP}(X, Y) \tag{24}$$

Among them, $\mathrm{APIP}(X, Y|Z)$ denotes the probability increment percentage (degree of association) of X and Y in the presence of known Z, which is calculated as shown in formula (26):

$$\mathrm{APIP}(X, Y|Z) = \frac{\sum_{k=1}^{c}\sum_{i=1}^{a}\sum_{j=1}^{b}\left|1-\frac{P(x_i|y_j,z_k)}{P(x_i|z_k)}\right|}{Nc}\times 100\% \tag{25}$$

Among them, the two nodes X and Y are defined in the same way as in formula (13), the possible states of Z nodes take the values $z_1, z_2, \cdots, z_c$, and z_k denotes the value of Z. When formula (23) is satisfied between X, Y, and Z, the changes of element values in the corresponding positions in G before and after orientation are shown in Table 1.

After V-structure orientation, there will be some unoriented non-V-structure edges left in G. Other methods are needed to orient the remaining edges. Since the causal phenomenon is not obvious when only one edge is considered, it is difficult to determine the direction of this edge when only one edge is considered. In order to make the initial network have the initial condition of the hill-climbing search process (all edges must have directions), we use the common and good orientation Conditional Relative Average Entropy (CRAE) [24] orientation method to orient the remaining undefined edges.

The calculation formula of CRAE [24] is shown in formulas (26):

$$\mathrm{CRAE}\left(X_j \to X_i\right) = \frac{H\left(X_i|X_j\right)}{t_i H\left(X_i\right)} \tag{26}$$

among them, t_i represents the number of values of node X_i. $H(X_i)$ and $H\left(X_i|X_j\right)$denotes the entropy and conditional entropy of X_i, If $CRAE\left(X_i \to X_j\right) \geq CRAE\left(X_j \to X_i\right)$, the direction is $X_i \to X_j$, otherwise the direction is $X_j \to X_i$.

TABLE 2. Request rules.

Symbols	Value	Meaning
ADD_{ij}	1	Request to add the edge $i \to j$
	0	No request
RVS_{ij}	1	Request to invert the edge $i \to j$
	0	No request
DEL_{ij}	1	Request to delete the edge $i \to j$
	0	No request

F. CONSTRAINT RULE SETTING

To improve the efficiency of hill-climbing algorithm, in addition to constructing an initial network before conducting the hill-climbing search, constraints need to be set for the search range of the hill-climbing search, which can improve the overall efficiency of the algorithm. However, if the constraint is too strong it will cause the correct result to be excluded from the constraint range, which will reduce the accuracy of the algorithm, so choosing the appropriate constraint method and constraint strength is the key to ensuring the accuracy of the algorithm and improve the overall efficiency of the algorithm.

For a node, in general, the degree of association of the node directly associated with it is stronger than that of the node indirectly associated with it. Therefore, from the theoretical point of view, the constraint matrix can be generated by setting the threshold value of the degree of association to exclude the pairs of nodes with weaker degree of association and then using this matrix to constrain the range of the mountain climbing search, which may obtain a better constraint effect. To make the constraint more applicable to each node, we constrain the column with the maximum value of each column in W as the reference, and the constraint method is shown in formula (27):

$$\begin{cases} \mathrm{CM}_{ij} = 1, & W_{ij} \geq \alpha \times \max(W_j) \\ \mathrm{CM}_{ij} = 0, & W_{ij} < \alpha \times \max(W_j) \end{cases} \tag{27}$$

Among them, CM denotes a constraint matrix of size $n \times n$, which does not change as the hill-climbing search proceeds, where the element values are 0 or 1; CM_{ij} denotes the value of the element in the i th row and j th column in CM; a denotes the constraint strength; and max $\left(W_j\right)$ denotes the maximum value of the j th column in W.

To visually represent the process and principle of constraint validation, we define to consider the hill-climbing search as an organic combination of three requests, the operation contents and corresponding symbols are shown in the Table 2:

During the hill-climbing search process, turning the direction of a currently existing edge or deleting an already existing edge will not add a new edge to the current network structure; only when an edge is added to the current network structure, a new edge will be introduced; therefore, CM only needs to work when a new edge needs to be added. The CM effective cases correspond to the following Table 3:

TABLE 3. CM and request synergy constraint rules.

ADD_{ij}	$CM_{ij} \| CM_{ji}$	Operation to be performed
1	1	add the edge $i \rightarrow j$
1	0	No operation to perform
0	1	No operation to perform
0	0	No operation to perform

TABLE 4. Experimental environment configuration table.

lab environment	Parameter
Processor	Intel(R) Pentium(R) CPU G4400
Main frequency	3.30GHz
Memory	8.00 GB
Operating system	Windows10 64 bit
Compiler Environment	MATLAB2020a

The experimental environment is shown in Table 4.
among them, $CM_{ij}||CM_{ji}$denotes the result of taking the logical or operation of CM_{ij} and CM_{ji}.

As can be seen from Table 3, the add-edge operation is performed only when $ADD_{ij} = 1$ and one of CM_{ij} and CM_{ji} is equal 1. In other cases, no operation will be performed.

In summary, the method proposed in this subsection aims to limit the search space and reduce the running time of the algorithm by setting a threshold value for the APIP between all the node pairs, to exclude the weakly connected node pairs from the search range and keep the strongly connected node pairs. In this method, if the value of α (threshold) is too large, although it can achieve the purpose of reducing the running time of the algorithm, it may cause some node pairs that should exist connection relations to be excluded from the search range, reducing the accuracy of the final result; if the value of α is too small, although it can ensure that the accuracy of the final result does not drop, but the effect of reducing the running time of the algorithm is not obvious. Therefore, if we want to weigh the accuracy and running time, we need to choose a suitable value of α to reduce the running time of the algorithm as much as possible while ensuring that the accuracy of the result is not reduced.

G. THE FLOW OF THE PHC ALGORITHM

The PHC algorithm is divided into three stages:

In the first stage, the initial network structure is constructed. First, the APIP value between each pair of nodes is calculated according to the input data set, and the probability increment matrix W is generated. Then, the undirected initial network structure G is obtained by analyzing W. Finally, the V-structure orientation method and the CRAE orientation method were used to orient the G.

In the second stage, a hill-climbing search is performed starting with G. The hill-climbing search process is implemented step by step, and at each step, all the new structures that can be generated are recorded and scored. Each of these steps consists of three operations; the first operation is to add an edge that is not in the network and is within the constraint; the second operation is to invert an existing edge; the third operation is to delete an existing edge. The structure with the highest rating among all the possible generated structures within the search range of this step is selected as the starting point for the next climb. After iterating with this rule until the scoring value is not rising, the program enters the third stage of the PHC algorithm.

The third stage is the forbidden climbing search, which still needs to be performed under the constraints of CM. First, let $G^* = G$, put G^* into the taboo set TB, record the score of G^* as max_score, and then perform a hill-climbing search on G^*. In each hill-climbing search step of this stage, if there is no score higher than max_score, replace G^* with the network structure with the highest score among all the results of this step so as to substitute the new G^* into the next hill-climbing in the search. If a structure with a higher score than max_score is found in the 15-step [19] search, assign it to G and return to the second stage to continue the hill-climbing search, and if no network structure with a higher score is found, make DAG $= G$ and output DAG.

The PHC algorithm pseudo-code is shown as algorithm 1.

III. EXPERIMENTAL RESULTSLAB

A. LAB EIRONMENT

B. EVALUATION INDICATORS

To verify the effectiveness of the APIP method, we judge whether the APIP method has the ability to analyze the degree of correlation between variables through the effectiveness of the initial network constructed by both the APIP method and the MI method using the MWST strategy. Since the initial network has no direction, this paper only considers whether the correlation of the edges in the initial network is correct. There are three evaluation indicators, namely correct edges, missing edges, and redundant edges.

To verify the effectiveness of the algorithm, this paper selects the evaluation indicators F1 (F1-score), Hamming Distance (HD), and TP (number of correct edges) as measurement tools for judging the quality of the structure. Among them, the calculation formula for F1 and HD is as follows:

$$\text{recall} = \frac{TP}{TP + FN} \tag{28}$$

$$\text{precision} = \frac{TP}{TP + FP} \tag{29}$$

$$\text{F1} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \tag{30}$$

$$\text{HD} = \text{FP} + \text{FN} \tag{31}$$

among them, TP represents the number of edges in the network learned by the algorithm that is the same as those in the standard network; FN represents the number of edges that exist in the standard network but not in the network learned by the algorithm, that is, the number of missing edges and reverse edges The sum of the number; FP represents the number of edges that exist in the network learned by the algorithm

Algorithm 1 DAG = PHC (DATA, ns, α)

Input: DATA: sample data set; ns: set of the total number of states for each node variable in the sample data set; α: constraint factor

Output: DAG: the Bayesian network structure corresponding to the dataset

$n \leftarrow size$(ns);
for each integer $i \in [1, n]$ **do**
  for each integer $j \in [1, n]$ **do**
    $W \leftarrow [W_{ij}]$; / $*$ W_{ij} represents the APIP value between node X_i and node X_j*/
  end for
end for
$G \leftarrow$ *Build MWST by* W; /* According to the method in [19] */
$G \leftarrow$*determine the direction of the edge in* G; /* Refer to part II, subsection D */
$CM \leftarrow$*Determine the constraint range according to* W *and* α;/*follow the method in equation (27) */
$r \leftarrow 0$;
while1 **do**
  if $r = 0$ **do**
    $[G, r] \leftarrow$*carry out mountain climbing search with* G *as the starting point, and continuously update the structure of* G *until the score does not increase, let* $r = 1$.
  else if $r = 1$ **do**
    $G^* = G$;
    $[G^*, r] \leftarrow$ *start a tabu search with* G^* *as a starting point, but update* G^* *only if a structure with a higher score than* G *is found. Until the search exceeds the specified step size, let* $r = 2$; *or find a structure with a score greater than* G, *let* $G = G^*$ *and* $r = 0$;
  else if $r = 2$ **do**
    $DAG \leftarrow G$;
    break;
  end if
end while
return *DAG*;

but does not exist in the standard network, that is, the sum of the number of reverse edges and redundant edges; recall represents the recall rate; precision represents the precision rate; HD Represents the gap between the network learned by the algorithm and the standard network.

C. ALGORITHM INTRODUCTION

To verify the effectiveness and advancement of the PHC algorithm, several other improved hill-climbing algorithms were selected for comparison, among which are the HC algorithm, CHC algorithm, Fast-CHC algorithm, Tabu algorithm, MMHC algorithm, and VTH algorithm, Table 5 shows the profiles of these algorithms.

TABLE 5. Introduction to the contrast algorithm.

Algorithm name	Algorithm Introduction
HC	The classic climbing algorithm
CHC	An improved version of the hill climbing algorithm combined with dynamic constraints
Fast-CHC	Improved and upgraded version of CHC algorithm
Tabu	Hill-climbing algorithm using taboo tables to improve the global search capability of the algorithm
MMHC	A hybrid algorithm combining since analysis and classical hill climbing
VTH	The initial network is first directed using a joint V-structure and log-likelihood function, followed by a hill-climbing algorithm combined with forbidden search

TABLE 6. Four kinds of network structure parameters in Asia, Car, Child, and Alarm.

Network name	Number of nodes	number of edges
Asia	8	8
Car	12	9
Child	20	25
Alarm	37	46

D. EXPERIMENTAL RESULTS

Asia, Car, Child, Alarm, these four standard networks are often used to test the effectiveness of the Bayesian structure learning algorithm. The parameters of the network structure of these four networks are shown in Table 6.

To test whether the APIP value has the ability to measure the degree of association between variables, the MI value was selected as a reference, and a comparative experiment was conducted on the MWST construct- ed by using the APIP value and the MI value. We selected four standard Bayesian networks of Asia, Car, Child, and Alarm. Each network took 1,000 sets of data, 5,000 sets of data, and 10,000 sets of data, a total of 12 sets of comparative experiments, and each set of comparative experiments had 45 data sets, the experimental results are the average of 45 experimental results. The experimental results are shown in Figure 2.

As can be seen from figure 2, except for the Asia network and the Alarm network, in the MWST constructed by the APIP method and the MI method, the numbers of correct edges, missing edges, and redundant edges are almost the same in other networks; In the network experimental results, the number of correct edges obtained by the APIP method is less than that obtained by the MI method. Since the MWST strategy can only obtain N-1 edges, where N is the number of nodes in the network, so a slight decrease in the number of correct edges leads to an increase in the number of wrong edges, making the APIP and MI results appear to differ significantly, but not significantly.; in the experimental results of the Alarm network, in the case of 10,000 sets of data, the experimental results of the APIP method are better than the MI method.

To make the PHC algorithm narrow the search range and achieve a high running speed while ensuring the effectiveness of the algorithm, we conducted experiments on the value

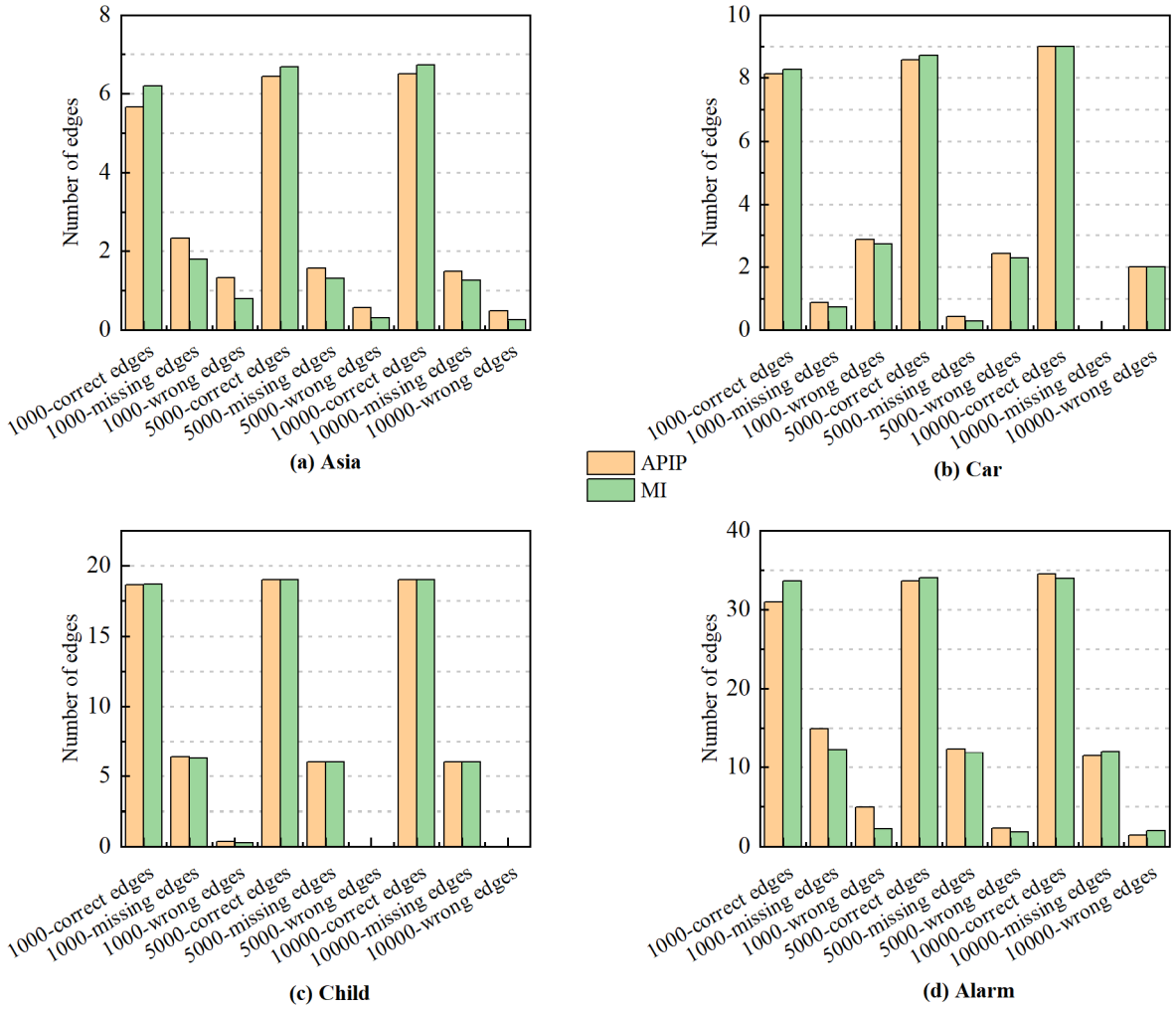


FIGURE 2. Comparison of experimental results using APIP and MI to build an initial network respectively.

of α. In this paper, the value of α is set to the interval of 0.05. The value of α was tested in four networks: Asian, car, child, and alarm, with 5000 sets of data. In addition to this, the same constraint strategy was used on MI for the same experiments, and the experimental results are shown in Figure 3 and Figure 4.

From Figure 3 and Figure 4, it can be seen that the APIP constraint method can reduce the running time of the algorithm with a certain degree of guaranteeing the accuracy of the PHC algorithm and has a more stable constraint effect, while the MI constraint method cannot reduce the running time with guaranteeing the accuracy of the PHC algorithm.

From analyzing the learning results of Asia, Car, and Alarm in Figure 3, it can be concluded that when the value of α is 0.2, it can make the PHC algorithm's accuracy not reduced and greatly improve its efficiency. The inclusion of the constraint strategy even improves the accuracy of the PHC algorithm. When the value of α is 0.05, the quality of the learning results of the PHC algorithm in all the networks participating in the test can be guaranteed, but at this time the value of α is too small to act as a constraint. After a comprehensive analysis of the accuracy of the experimental results and the running time, the PHC algorithm chooses 0.2 as the value of α, which allows the algorithm to maximize the efficiency in the structure learning of most networks. In the following experiments, the α value was all 0.2.

To more visually demonstrate the advanced nature of using APIP restricted search, under the same constraint strategy, we plotted the Pareto curves (F1-Time) using APIP and using MI, respectively, as shown in Figure 5. We can clearly see that the method using APIP restricted search can spend less time in Asia's structure learning and learning the network structure with higher scores. In the structure learning results of the Car network, it is easy to see that the method using APIP-constrained search can get higher-rated network structures at the same time. In the network structure learning results of Child and Alarm, the results of these two methods are similar, but in this aspect of the stability of the ratings of the network,

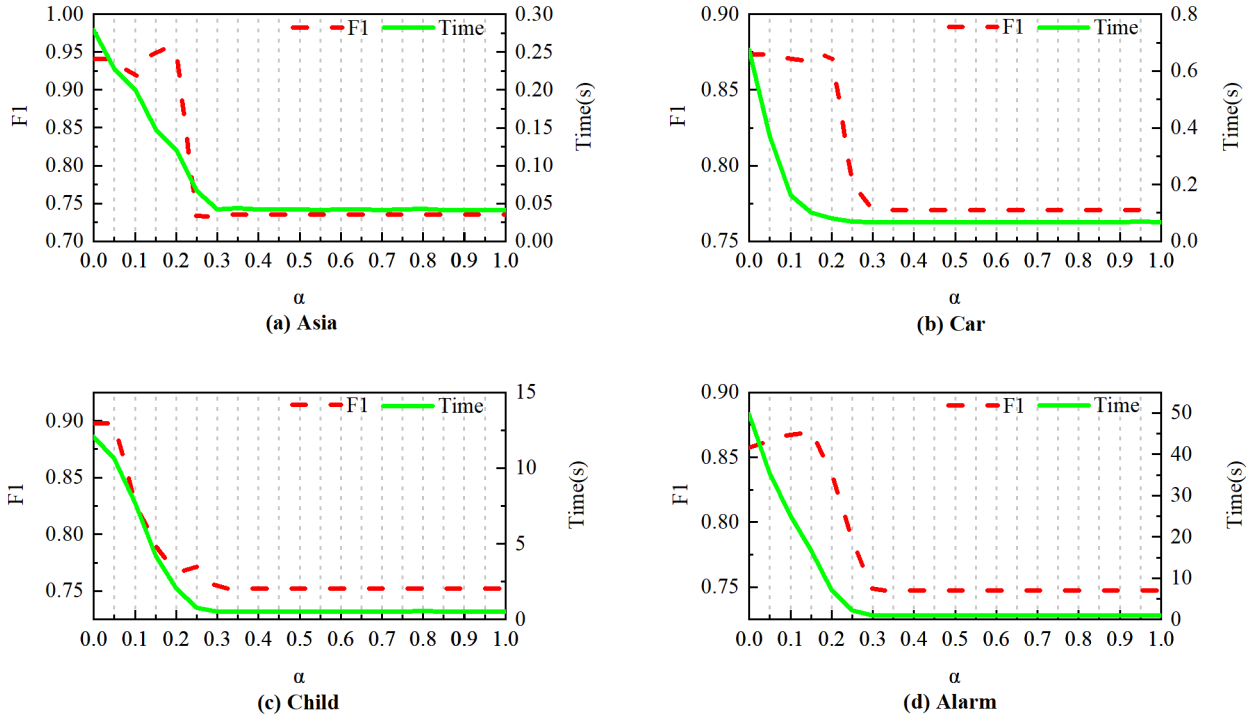


FIGURE 3. F1 and running time of PHC algorithm influenced by a for the case of CM generation using APIP.

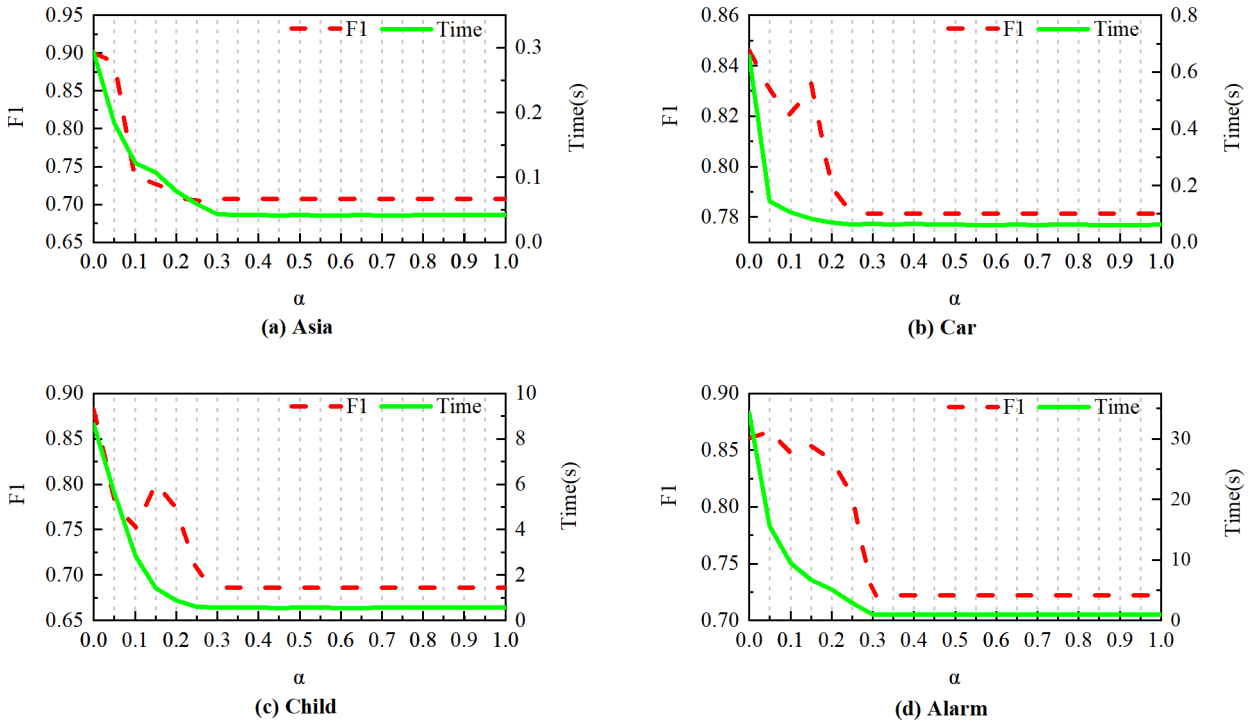


FIGURE 4. F1 and running time of PHC algorithm influenced by a for the case of CM generation using MI.

the method using APIP-constrained search is slightly better. The reason for this experimental result may be that the MI formula does not preserve the sequential law of influence degree transmission between nodes, while APIP effectively preserves the mutual influence degree between nodes (see Section II, Part B for the detailed analysis of the principle).

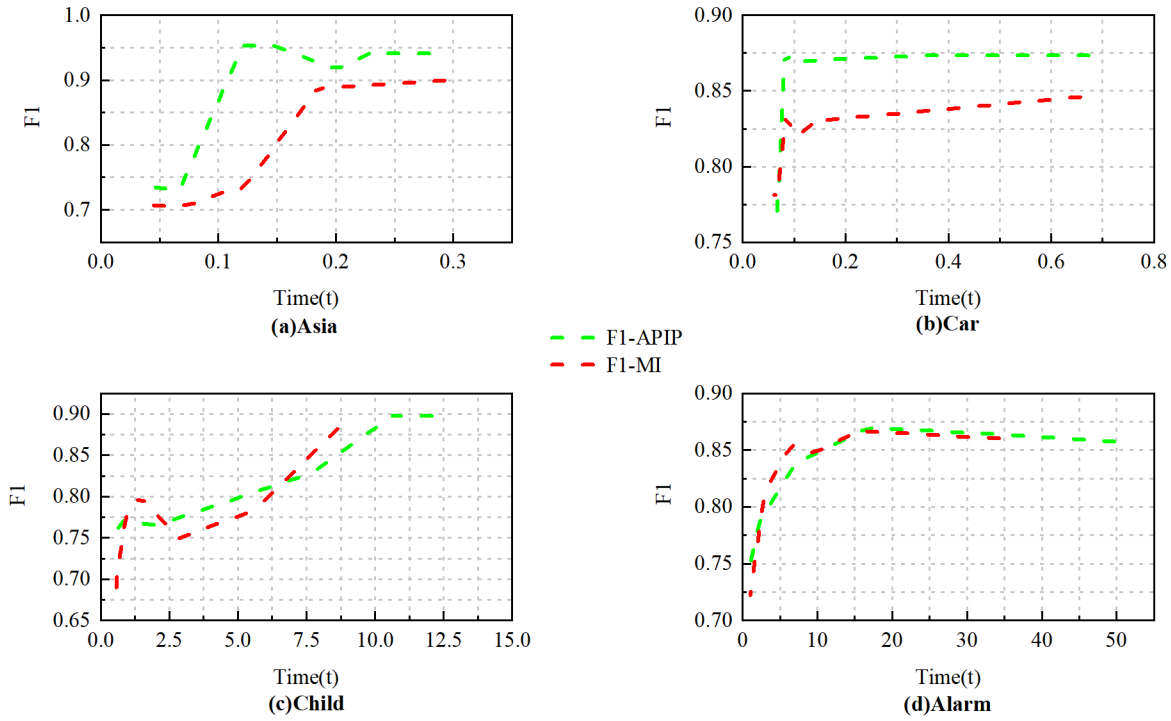


FIGURE 5. Pareto curves of F1 and algorithm runtime Time with APIP and MI as constraints, respectively.

To visualize the running time comparison of the PHC algorithm with and without introducing constraints, the experiments were conducted by making a = 0 and a = 0.2 in four networks with 1000, 5000, and 10000 as the sample capacity, respectively. The experimental results are shown in Figure 6. As can be seen from Figure 6, the running time of the constrained PHC algorithm is significantly shortened. After the constraints, the average running time of the PHC algorithm in the Asia network is reduced by about 38.2%; the average running time of the Car network is reduced by 77.1% per month; the running time of the Child network is reduced by about 43.8%; the running time of the Alarm network is reduced by about 55.1%. It can be seen that the effectiveness of the constraint strategy is confirmed.

To analyze the performance of the PHC algorithm, we compared the PHC algorithm with other algorithms on four networks of Asia, Car, Child, and Alarm. 45 experiments were carried out independently under the conditions of 1000 sets of data, 5000 sets of data, and 10000 sets of data for each network, and the average value of the 45 experiments was taken as the final experimental result under each condition, Car, Child, Alarm standard network generation. The experimental results of the F1 value, HD, and TP value are shown in Table 7 to Table 10, and the experimental results of running time are shown in Figure 7.

It can be seen from Table 7 to Table 10 and Figure 7 that in the experiments of the Asia and Car networks, the PHC algorithm performs very well, especially in the case of 5000 groups and 10000 groups of data. The F1 value, HD value, and TP value of the learned Bayesian network structure are the best among all the algorithms involved in the experiment. Because the PHC algorithm is similar to the VTH algorithm, the reason for this result may be that the PHC algorithm is climbing a mountain. The process of searching plays a role in the handling of isolated nodes; from the point of time, the PHC algorithm is not as fast as HC and MMHC in the Asia network, and not as fast as MMHC in the Car network, but its results are much better than those. In the experimental results of the Child network, the effect of the PHC algorithm is worse than other algorithms when the data volume is 1000, but not as good as the TVH algorithm and the MMHC algorithm when the data volume is 5000 and 10000. The reason may be that the PHC algorithm adds constraints. When the amount of data is low, the data cannot accurately represent the characteristics of the standard network. If the search range is too large, the wrong local structure will be difficult to be identified by the scoring function. It has an adverse effect on the accuracy of hill-climbing search; but as the amount of data increases, the ability of the data to express standard network features becomes stronger, and the wrong local structure within the search range is easily identified by the scoring function, and it can be seen from Figure 7, the constraints are not effective in the Child network, so the PHC algorithm is not as good as the MMHC algorithm and the VTH algorithm when the amount of Child network data is large. In the experimental results of the Alarm network, relatively speaking, the effects of the PHC algorithm and the VTH algorithm are similar. Although the results of the F1

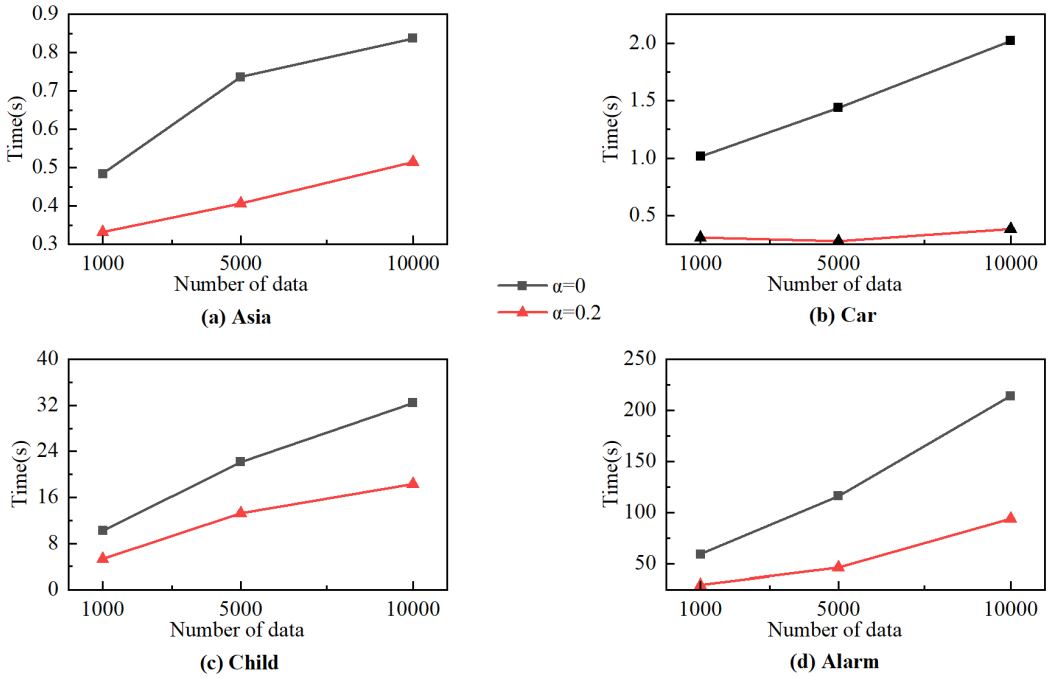


FIGURE 6. The relationship between the running time of the PHC algorithm and the amount of data under constraints and without constraints.

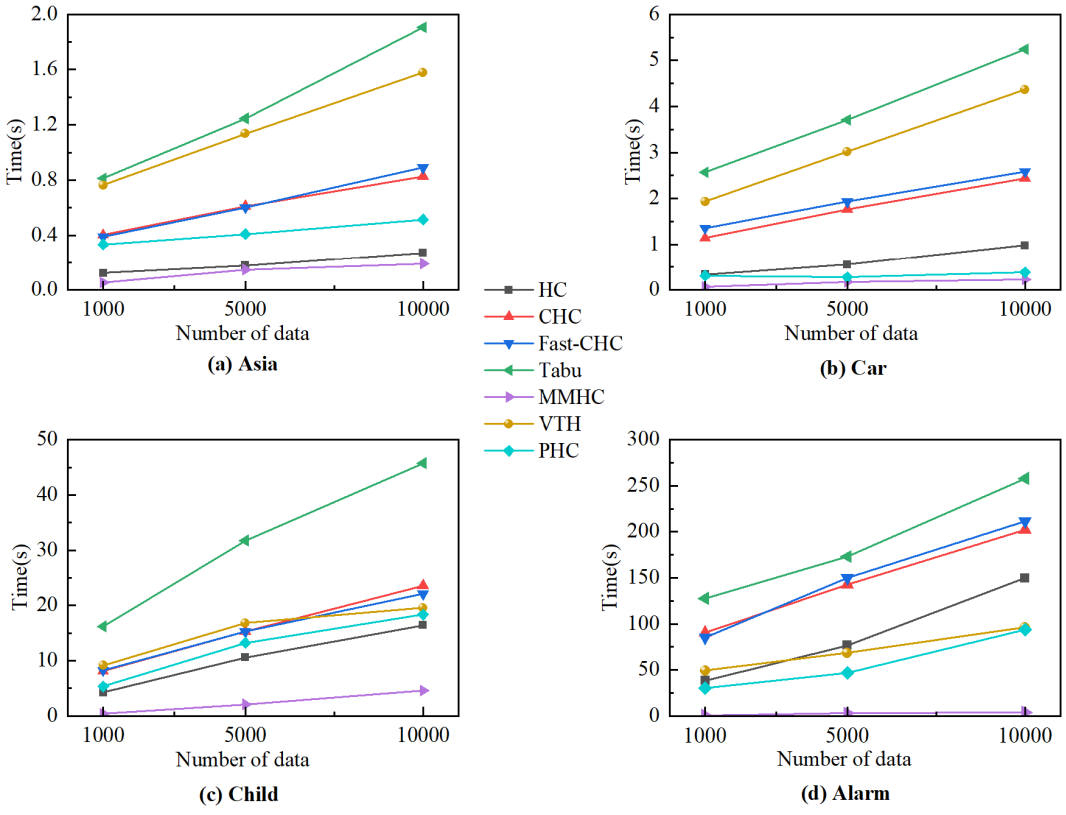


FIGURE 7. Comparison of running time between PHC algorithm and other algorithms.

value, HD value, and TP value of the VTH algorithm are better, the PHC algorithm only When the F1 value is reduced by 3.09%, the running time is reduced by about 24.4%; compared with other algorithms except for VTH, the results of the PHC algorithm are significantly better than those of other algorithms.

TABLE 7. Experimental results of PHC algorithm and other algorithms in Asia network.

Algorithm	Asia-1000			Asia-5000			Asia-10000		
	F1	HD	TP	F1	HD	TP	F1	HD	TP
HC	0.355	6.29	2.84	0.374	6.00	3.00	0.373	6.00	3.00
CHC	0.548	4.80	4.31	0.485	5.44	3.95	0.540	4.78	4.44
Fast-CHC	0.545	4.82	4.29	0.48	5.47	3.93	0.545	4.76	4.47
Tabu	0.601	4.29	4.69	0.547	4.78	4.40	0.570	4.51	4.62
MMHC	0.693	3.71	5.42	0.792	2.09	6.15	0.803	1.80	6.29
VTH	**0.817**	**2.11**	**6.20**	0.861	1.53	6.55	0.896	1.13	6.89
PHC	**0.784**	**2.87**	**6.38**	**0.885**	**1.29**	**7.09**	**0.924**	**0.84**	**7.40**

TABLE 8. Experimental results of PHC algorithm and other algorithms in Car network.

Algorithm	Car-1000			Car -5000			Car -10000		
	F1	HD	TP	F1	HD	TP	F1	HD	TP
HC	0.518	4.98	4.84	0.522	5.07	4.95	0.526	5.00	5.00
CHC	0.506	5.22	4.75	0.590	4.36	5.62	0.610	4.22	5.84
Fast-CHC	0.509	5.18	4.78	0.584	4.47	5.55	0.615	4.22	5.87
Tabu	0.563	4.58	5.24	0.649	3.62	6.07	0.700	3.44	6.27
MMHC	**0.722**	**2.60**	**6.42**	0.825	1.60	7.42	0.816	1.67	7.35
VTH	**0.704**	**2.98**	**6.44**	0.773	2.18	7.02	0.740	2.49	6.75
PHC	**0.707**	**2.89**	**6.49**	**0.859**	**1.33**	**7.76**	**0.914**	**0.78**	**8.22**

TABLE 9. Experimental results of PHC algorithm and other algorithms in Child network.

Algorithm	Child-1000			Child -5000			Child -10000		
	F1	HD	TP	F1	HD	TP	F1	HD	TP
HC	0.612	9.22	14.2	0.679	9.84	17.7	0.700	9.22	18.3
CHC	0.631	12.2	14.9	0.664	10.5	17.1	0.702	9.16	18.2
Fast-CHC	0.636	12.5	15.0	0.664	10.5	17.1	0.706	9.02	18.3
Tabu	0.696	10.1	16.3	0.743	7.82	19.0	0.770	6.89	19.7
MMHC	**0.778**	**7.09**	**18.1**	0.857	3.87	21.3	0.866	3.42	21.6
VTH	**0.789**	**7.35**	**18.2**	**0.931**	**2.04**	**23.4**	**0.897**	**3.13**	**22.9**
PHC	**0.793**	**7.24**	**18.2**	0.795	6.53	19.4	0.764	7.56	19.0

TABLE 10. Experimental results of PHC algorithm and other algorithms in Alarm network.

Algorithm	Alarm-1000			Alarm -5000			Alarm -10000		
	F1	HD	TP	F1	HD	TP	F1	HD	TP
HC	0.327	40.3	15.1	0.339	41.4	17.2	0.320	43.3	16.6
CHC	0.415	35.1	19.3	0.426	35.4	21.4	0.409	36.9	21.0
Fast-CHC	0.407	35.5	18.9	0.426	35.4	21.4	0.409	37.0	20.9
Tabu	0.417	35.1	19.3	0.437	34.6	21.9	0.427	35.8	21.7
MMHC	0.681	18.2	31.5	0.795	11.2	37.0	0.729	14.3	34.2
VTH	**0.742**	**15.9**	**32.9**	**0.896**	**6.93**	**40.8**	**0.934**	**4.18**	**43.1**
PHC	**0.718**	**16.8**	**32.2**	**0.864**	**7.84**	**39.9**	**0.911**	**5.22**	**42.0**

IV. CONCLUSION

In this paper, the concept of the APIP algorithm and the PHC algorithm is proposed, the calculation formula of APIP is deduced, and experiments are used to verify that APIP has the ability to express the degree of association and the effectiveness of the PHC algorithm. In Bayesian network learning, APIP and MI have approximate correlation degree expression ability; the PHC algorithm has excellent learning ability, and the threshold of its constraint strategy is suitable for most networks, but in individual networks, due to constraints Excessive force leads to poor algorithm performance. In the future, other methods can be used to constrain the

search range of the algorithm, so that the algorithm has better adaptability.

ACKNOWLEDGMENT

The authors would like to thank the editors and the anonymous reviewers whose insightful comments have helped to improve the quality of this paper considerably.

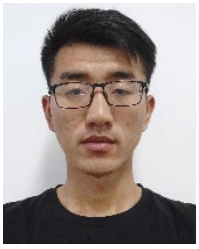
REFERENCES

[1] M. U. Habib, "An econometric analysis of relationship between economic growth and petroleum consumption in India," *Jindal J. Bus. Res.*, vol. 8, no. 2, pp. 118–127, Dec. 2019.
[2] *Heart Disorders and Diseases—Angina Pectoris; Researchers at Department of Cardiology Target Angina Pectoris (Correlation Study on the Blood Pressure Variability, Serum High-Sensitivity C-Reactive Protein and Homocysteine of Patients With Stable Angina Pectoris and Hypertension)*, Chemicals & Chemistry, 2020.
[3] G. A. Darbellay and I. Vajda, "Estimation of the information by an adaptive partitioning of the observation space," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1315–1321, May 1999.
[4] Y.-I. Moon, B. Rajagopalan, and U. Lall, "Estimation of mutual information using kernel density estimators," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 52, no. 3, pp. 2318–2321, Sep. 1995.
[5] L. Breiman and J. H. Friedman, "Estimating optimal transformations for multiple regression and correlation," *J. Amer. Stat. Assoc.*, vol. 80, no. 391, pp. 580–598, Sep. 1985.
[6] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti, "Detecting novel associations in large data sets," *Science*, vol. 334, no. 6062, pp. 1518–1524, Dec. 2011.
[7] B. Sun and Y. Zhou, "Bayesian network structure learning with improved genetic algorithm," *Int. J. Intell. Syst.*, vol. 37, no. 9, pp. 6023–6047, Sep. 2022.
[8] B. Sun, Y. Zhou, J. Wang, and W. Zhang, "A new PC-PSO algorithm for Bayesian network structure learning with structure priors," *Exp. Syst. Appl.*, vol. 184, Dec. 2021, Art. no. 115237.
[9] W. Huang, H. Li, H. Fan, and Y. Qian, "Causation mechanism analysis of excess emission of flue gas pollutants from municipal solid waste incineration power plants by employing the fault tree combined with Bayesian network: A case study in Dongguan," *J. Cleaner Prod.*, vol. 327, Dec. 2021, Art. no. 129533.
[10] S. Gerassis, E. Giráldez, M. Pazo-Rodríguez, Á. Saavedra, and J. Taboada, "AI approaches to environmental impact assessments (EIAs) in the mining and metals sector using AutoML and Bayesian modeling," *Appl. Sci.*, vol. 11, no. 17, p. 7914, Aug. 2021.
[11] N. Liu, M. Hu, J. Wang, Y. Ren, and W. Tian, "Fault detection and diagnosis using Bayesian network model combining mechanism correlation analysis and process data: Application to unmonitored root cause variables type faults," *Process Saf. Environ. Protection*, vol. 164, pp. 15–29, Aug. 2022.
[12] Z. Liu, K. Lv, C. Zheng, B. Cai, G. Lei, and Y. Liu, "A fault diagnosis method for rolling element bearings based on ICEEMDAN and Bayesian network," *J. Mech. Sci. Technol.*, vol. 36, no. 5, pp. 2201–2212, May 2022.
[13] X. Xiong, H. Xu, S. Wang, X. Liao, X. Yi, K. Jin, H. Lei, S. Bai, S. Qiu, and L. Yang, "Association of novel androgen receptor axis-targeted therapies with diarrhea in patients with prostate cancer: A Bayesian network analysis," *Frontiers Med.*, vol. 8, Jan. 2022.
[14] Y. Fan, Y. Mulati, L. Zhai, Y. Chen, Y. Wang, J. Feng, W. Yu, and Q. Zhang, "Diagnostic accuracy of contemporary selection criteria in prostate cancer patients eligible for active surveillance: A Bayesian network meta-analysis," *Frontiers Oncol.*, vol. 11, Jan. 2022.
[15] I. Tsamardinos, L. E. Brown, and C. F. Aliferis, "The max-min hill-climbing Bayesian network structure learning algorithm," *Mach. Learn.*, vol. 65, no. 1, pp. 31–78, 2006.
[16] J. A. Gámez, J. L. Mateo, and J. M. Puerta, "Learning Bayesian networks by Hill climbing: Efficient methods based on progressive restriction of the neighborhood," *Data Mining Knowl. Discovery*, vol. 22, nos. 1–2, pp. 106–148, Jan. 2011.
[17] J. A. Gámez, J. L. Mateo, and J. M. Puerta, "One iteration CHC algorithm for learning Bayesian networks: An effective and efficient algorithm for high dimensional problems," *Prog. Artif. Intell.*, vol. 1, no. 4, pp. 329–346, 2012.
[18] L. Haoran et al., "Research on the fault diagnosis model of rotary kiln based on Bayesian improvement algorithm," *J. Instrum.*, vol. 36, no. 7, pp. 1554–1561, 2015.
[19] L. Haoran, N. Wang, Y. Wang, L. Zhang, Z. Su, W. Liu, and X. Zhao, "Bayesian network structure algorithm based on V-structure & log-likelihood function orientation with forbidden Hill climbing," *Electron. Inf. Lett.*, vol. 43, no. 11, pp. 3272–3281, 2021.
[20] L. Haoran et al., "A simplified greedy algorithm for structure learning in Bayesian networks," *Small Microcomputer Syst.*, vol. 36, no. 2, pp. 306–309, 2015.
[21] M. Zichuan, S. Hussain, Z. Ahmad, O. Kharazmi, and Z. Almaspoor, "A new generalized Weibull model: Classical and Bayesian estimation," *Comput. Syst. Sci. Eng.*, vol. 38, no. 1, pp. 79–92, 2021.
[22] S. Behjati and H. Beigy, "Improved K2 algorithm for Bayesian network structure learning," *Eng. Appl. Artif. Intell.*, vol. 91, May 2020, Art. no. 103617.
[23] J. Suzuki, "A construction of Bayesian networks from databases based on an MDL principle," 2013, *arXiv:1303.1486*.
[24] K. Liu, Y. Cui, J. Ren, and P. Li, "An improved particle swarm optimization algorithm for Bayesian network structure learning via local information constraint," *IEEE Access*, vol. 9, pp. 40963–40971, 2021.

HAORAN LIU received the B.S. degree from the School of Electrical Engineering, Yanshan University, Qinhuangdao, China, in 2002, the M.S. degree from the School of Information Science and Technology, Yanshan University, in 2004, and the Ph.D. degree from the Institute National des Sciences Appliqués de Rouen, in 2009. He is currently a Professor with the College of Information Science and Technology, Yanshan University. His current research interests include wireless sensor network protocol architectures, and industrial fault detection and prediction.

SHAOPENG CUI received the bachelor's degree from Yanshan University, in 2019, where he is currently pursuing the master's degree with the School of Information Science and Engineering. His research interests include artificial intelligence and machine learning.

SHENG LI received the bachelor's degree from Shijiazhuang University, in 2020. She is currently pursuing the master's degree with the School of Information Science and Engineering, Yanshan University. Her research interests include artificial intelligence and data quality prediction of CNC machine tools.



NIANTAI WANG received the bachelor's degree from the University of Jinan, in 2019. He is currently pursuing the Ph.D. degree with the School of Information Science and Engineering, Yanshan University. His research interests include Bayesian network learning and feature selection.

QIANRUI SHI received the bachelor's degree from Shijiazhuang University, in 2021. She is currently pursuing the master's degree with the School of Information Science and Engineering, Yanshan University. Her research interests include artificial intelligence and machine learning.

YANBIN CAI received the bachelor's degree from the China University of Geosciences (Beijing). He is currently pursuing the master's degree with the School of Information Science and Engineering, Yanshan University. His research interests include artificial intelligence and machine learning.

LIYUE ZHANG received the B.S. degree in electronic information and engineering from Hengshui University, Hengshui, China, in 2017. He is currently pursuing the Ph.D. degree with Yanshan University. His research interests include Bayesian networks, intelligent algorithm, data modeling, and industrial fault detection and prediction.

DAYAN LIU received the B.S. and M.S. degrees from the University of Lille 1, France, in 2005 and 2007, respectively, and the Ph.D. degree in applied mathematics from the University of Lille 1 & INRIA LilleNord Europe, in October 2011. Since then, he has held postdoctoral positions at Arts et Métiers ParisTech, France, and the King Abdullah University of Science and Technology (KAUST), Saudi Arabia. He is currently an Associate Professor with the French National Institute of Applied Sciences (INSA), Centre Val de Loire, where he belongs to the Control Team, PRISME Laboratory. His main research interests include estimation and identification for integer order linear systems and fractional order linear systems.