

Research Article

Dynamic Knowledge Inference Based on Bayesian Network Learning

Deyan Wang ^{1,2}, Adam AmrilJaharadak ² and Ying Xiao¹

¹Management & Science University, Shah Alam, Malaysia

²School of Internet of Things Technology, Wuxi Institute of Technology, Wuxi, Jiangsu, China

Correspondence should be addressed to Deyan Wang; wangdy@wxit.edu.cn

Received 21 October 2020; Revised 27 November 2020; Accepted 28 November 2020; Published 7 December 2020

Academic Editor: Chenxi Huang

Copyright © 2020 Deyan Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

On the basis of studying datasets of students' course scores, we constructed a Bayesian network and undertook probabilistic inference analysis. We selected six requisite courses in computer science as Bayesian network nodes. We determined the order of the nodes based on expert knowledge. Using 356 datasets, the K2 algorithm learned the Bayesian network structure. Then, we used maximum a posteriori probability estimation to learn the parameters. After constructing the Bayesian network, we used the message-passing algorithm to predict and infer the results. Finally, the results of dynamic knowledge inference were presented through a detailed inference process. In the absence of any evidence node information, the probability of passing other courses was calculated. A mathematics course (a basic professional course) was chosen as the evidence node to dynamically infer the probability of passing other courses. Over time, the probability of passing other courses greatly improved, and the inference results were consistent with the actual values and can thus be visualized and applied to an actual school management system.

1. Introduction

In artificial intelligence research, one of the core issues lies in expressing the existing knowledge and applying the existing knowledge for analysis, processing, or inference in order to obtain new knowledge [1–3]. Among them, the expression and inference of uncertain knowledge is the most important and difficult [4, 5]. Uncertain knowledge representation can be divided into two categories. The first is a probability-based method, including a Bayesian network, dynamic causal network, and Markov network. The second one is a nonprobabilistic method, including fuzzy logic, evidence theory, and rough set theory, among others [6–10]. The Bayesian network was first proposed by Professor Judea Pearl of the University of California in the 1980s [6]. He extended the Bayesian network to expert systems and made it a common method for uncertain knowledge and data inference [7].

The paper selected 6 requisite courses in computer science as Bayesian network nodes to carry out Bayesian network structure learning and parameter learning. Then,

taking mathematics course as the evidence node, we carried out dynamic prediction of other course grades. The experimental results show that when mathematics course examination is passed, the probability of passing other courses will increase, which is consistent with the actual values.

2. Bayesian Network Definition

A Bayesian network is a directed graphical description based on a network structure. It is a combination of artificial intelligence, probability theory, graph theory, and decision theory. It uses a directed acyclic graph (DAG) with a network structure to express the dependence and influence degree of each information element. Among them, nodes are used to express each feature attribute, and the directed edges between connected nodes are used to express the dependence of each feature attribute. The conditional probability table (CPT) expresses the degree of influence between each feature attribute and combines the prior knowledge with the

sample information and the dependency relationship with the probability representation [6].

An extremely important property of Bayesian networks is that each node is independent of all its indirect predecessor nodes after the value of its immediate precursor node is determined.

The significance of this feature is to clarify that the Bayesian network can easily calculate the joint probability distribution. In general, we solve the multivariate non-independent joint conditional probability distribution using the following equation:

$$P(x_1, x_2, \dots, x_n) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2) \dots P(x_n|x_1, x_2, \dots, x_{n-1}). \quad (1)$$

In the Bayesian network, due to the aforementioned properties, the joint conditional probability distribution of any random variable combination can be defined as follows:

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{Parents}(x_i)). \quad (2)$$

Parents represent the joint probability of the direct precursor node of X_i , and other probabilities can be found from the CPT.

In Bayesian networks, there can be more than one directed path between nodes, and an ancestor node can influence its offspring nodes through different ways. When the appearance of an ancestor node leads to the generation of the result of a descendant node, it is a probability expression, not inevitable, so we need to add a conditional probability for each node. The probability of a node taking different attribute values under the condition of different value combinations of its parent node (direct cause node) constitutes the conditional probability table of the node. The initial root node occurs without any conditions, which is called unconditional probability.

3. Bayesian Network Learning

The core problem in Bayesian networks lies in Bayesian network learning. Bayesian network learning is the process of using a learning algorithm to obtain a Bayesian network that can truly express the relationship between the variables in the sample dataset. The Bayesian network comprises a network structure containing nodes and directed edges and the CPT representing the degree of dependence between nodes. Thus, Bayesian network learning is divided into two parts: structural learning and parametric learning. Structural learning is more difficult than parametric learning. Structural learning of Bayesian networks refers to the use of training datasets, combined with as much prior knowledge as possible to determine the appropriate Bayesian network topology.

3.1. Bayesian Network Structural Learning. Bayesian network structure learning mainly comprises two categories—the first is based on scoring and search methods and the second is based on conditional independence testing.

The scoring and search-based method has two elements: a scoring function and search algorithm. Cooper and Herskovits created the K2 algorithm [11] that uses the Bayesian score and hill-climbing search to obtain the optimal network structure under the given node order. In 1994, Remco proposed replacing the Bayesian scoring function in the K2 algorithm with the scoring function of minimum description length (MDL) in information theory and established the K3 algorithm [12]. In the same year, Lam and Bacchus proposed that the MDL be used as the criterion to learn the network structure through complete search and get rid of the constraint of prior information of node order [13]. Shan et al. [14] proposed the hill-climbing algorithm, which makes full use of the information of rings to optimize the network structure and demonstrates good learning effect. In recent years, there have been many studies on learning Bayesian networks. It presents an improved hybrid learning strategy that features parameterized genetic algorithms (GAs) to learn the structure of BNs underlying a set of data samples [15]. Wang and Jiang [16] studied the hybrid learning method of a Bayesian network structure based on the bee colony algorithm (CA) and genetic algorithm (GA) and selected the Asia network and car trouble-shooter network, and the efficiency of the algorithm was significantly improved in the process of increasing the number of samples. Cao et al. [17] studied the Bayesian network structure learning algorithm based on cloud genetic annealing. A fast learning method of Bayesian network structure based on attribute order was proposed to realize the rapid construction of Bayesian network structure [18].

The K2 algorithm effectively integrates prior information in the search process and demonstrates good time performance. It is a classic structure learning algorithm based on scoring search.

3.1.1. Scoring Function. The algorithm first needs to determine the order of the node variables in the network and propose a modular idea wherein the parent node set of each node is independent of each other. The K2 search algorithm designed according to this idea uses the hill-climbing heuristic algorithm to search the network structure under the assumption that the nodes are ordered and the prior probability of all network structures is equal. The parent node set is searched for each node in the given order, and the score of the local structure is increased by continuously adding a parent node to each node. The search stops until the node with the highest score is found for each node, and it is always required to maximize the score of the structure but the node is in order.

3.1.2. Search Strategy. The K2 algorithm uses greedy search to obtain the maximum value. We first assume that the random variables are ordered. If x_i precedes x_j , then there can be no edges from x_j to x_i . Simultaneously, we assumed that the maximum number of parent variables per variable is u . Each time the largest parent variable of the scoring function is selected and put into the set, the loop stops when the scoring function cannot be increased.

3.2. Bayesian Network Parameter Learning. The parametric learning of Bayesian networks refers to the determination of conditional probability density of each node for a given Bayesian network structure. This learning can be determined by expert knowledge or training sample data, and the incompleteness and inaccuracy of expert knowledge will affect the accuracy of the network parameters. Based on the previous structure learning, this section focuses on the analysis of the Bayesian network parameter learning method for complete sample data.

For complete training sample data, maximum a posteriori (MAP) and maximum likelihood estimation (MLE) are often used to learn Bayesian network parameters.

3.2.1. Maximum Likelihood Estimation (MLE). Maximum likelihood estimation (MLE) estimates the parameters of the model according to the data. The goal of MLE is to find a set of parameters to maximize the probability of the model producing the observed data. Simply put, MLE is to estimate the parameters (the environment in which the data are generated) based on the observed data.

Hypothesis data $X(x_1, x_2, \dots, x_n)$ are a set of samples of independent and identity distribution. Then, MLE chooses θ that maximizes the probability of observed data as follows:

$$\begin{aligned}\hat{\theta}_{MLE} &= \arg \max P(X; \theta) \\ &= \arg \max P(x_1; \theta)P(x_2; \theta) \dots P(x_n; \theta) \\ &= \arg \max \log \prod_{i=1}^n P(x_i; \theta) \\ &= \arg \max \sum_{i=1}^n \log P(x_i; \theta) \\ &= \arg \min - \sum_{i=1}^n \log P(x_i; \theta).\end{aligned}\quad (3)$$

3.2.2. Maximum A Posteriori (MAP). The maximum likelihood estimation is to obtain the parameter θ , which maximizes the likelihood function $P(X|\theta)$. The maximum posterior probability estimation is to find θ to maximize $P(X|\theta)P(\theta)$. The obtained θ is not only the maximum likelihood function but also the prior probability of θ itself. Then, MAP chooses θ that maximizes the probability of observed data as follows:

$$\hat{\theta}_{MAP} = \arg \min - \sum_{i=1}^n \log P(x_i; \theta) - \log P(\theta). \quad (4)$$

4. Dataset

In this study, the scores of 15 examination courses of 356 students who have graduated from network technology specialty are selected as the dataset. For the convenience of programming, the course adopts the form of abbreviation (see Table 1 for details).

The 15 courses in the table are listed with class hours, credits, and semester, which is convenient for the analysis of inference results.

5. Bayesian Network Dynamic Inference

Bayesian network inference is generally divided into two categories of precise inference and approximate inference from the methodological perspective. From the mode perspective, there are three most common types: causal, diagnostic, and support inferences [19]. The Bayesian network inference is widely used. Li et al. [20] considered a cost-sensitive Bayesian network and weighted K-nearest neighbor model to predict the duration of accidents. To minimize the negative impacts brought by floods, researchers propose a hierarchical Bayesian network-based incremental model to predict floods for small rivers [21]. Further, using the biological information from the literature to develop a Bayesian network along with a messaging passing algorithm, progress can be made in the treatment of breast cancer [22].

By applying K2 algorithm and MAP algorithm, the Bayesian network is constructed as shown in Figure 1.

The mean of the nodes is summarized in Table 2:

There are six nodes in the Bayesian network: C# (C), Java (J), Web (W), Database (D), Android (A), and Math (M). The directed side indicates the dependence between courses; the Bayesian network structure is constructed through the dataset of the students' course scores. For example, the performance of Math course will affect the C# course, and the probability values are used to describe the degree of impact between the different courses. The detailed probability values of different courses are listed in Tables 3–8.

For CPT, it only has unconditional probability if the node is the root node. It has the conditional probability if the node is not the root node. M and D are the root nodes, so they have the unconditional probability. C, J, W, and A are not the root nodes, so they have the conditional probability.

The CPT of the Bayesian network is as follows.

For this Bayesian network, the following inference problems need to be solved:

- (1) If the students have passed Math, how likely are they to pass C#?
- (2) If they have passed Math, how likely are they to pass Web?
- (3) If they have passed Android, how likely are they to pass Web?

Note: logically speaking, the third problem should not exist because it is impossible to infer the scores of previous courses from the courses in the following semester. This is just to explain how the bottom-up diagnostic inference works.

- (4) If they have passed Math, how likely are they to pass Database?
- (5) If they have passed Math, how likely are they to pass Android?

The inference algorithm based on message propagation is the exact inference algorithm proposed by Pearl in 1988

TABLE 1: Sample data structure.

Course	Attribute	Class hours	Credits	Semester
C++ Programming I	CI	48	3	1
C++ Programming II	CII	64	4	2
Design and Application of Web Database	DAWEB	64	4	5
Fundamentals of Computer Application	FCA	48	3	1
Network Basics	NB	64	4	1
Network Security Technology	NST	48	3	4
Network Switching and Routing	NSR	80	5	2
Application Mathematics I	AMI	64	4	1
Application Mathematics II	AMII	64	4	2
Basic English I	BE I	64	4	1
Basic English II	BEII	64	4	2
Integrated Communication Training	ICT	28	1	3
Network Engineering Project Training	NEPT	28	1	5
Integrated Network Training	INT	56	2	5
Wireless Network Deployment Week	WNDW	28	1	3

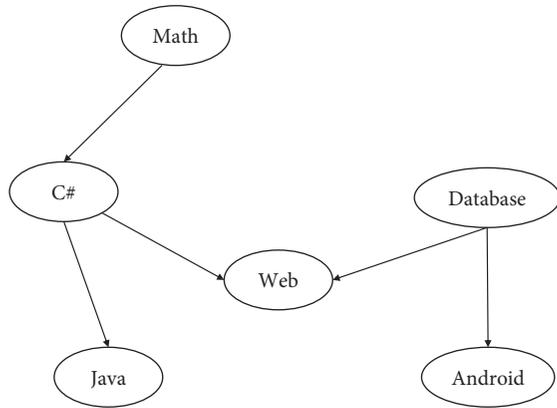


FIGURE 1: The Bayesian network.

TABLE 2: Node name, meaning, and abbreviation.

Node name	Node meaning	Node abbreviation
C#	C# programming	C
Java	Java programming	J
Web	Design and application of web database	W
Database	Database design and management	D
Android	Android app development	A
Math	Application mathematics	M

TABLE 3: Unconditional probability of M.

M	$P(M)$
Pass	0.800
Fail	0.200

TABLE 4: Unconditional probability of D.

D	$P(D)$
Pass	0.750
Fail	0.250

TABLE 5: Conditional probability of C at known node M.

M	C	$P(C M)$
Pass	Pass	0.850
Pass	Failure	0.150
Fail	Pass	0.550
Fail	Failure	0.450

TABLE 6: Conditional probability of J at known node C.

C	J	$P(J C)$
Pass	Pass	0.900
Pass	Failure	0.100
Fail	Pass	0.600
Fail	Failure	0.400

TABLE 7: Conditional probability of A at known node D.

D	A	$P(A D)$
Pass	Pass	0.980
Pass	Failure	0.020
Fail	Pass	0.010
Fail	Failure	0.990

TABLE 8: Conditional probability of W at known nodes C and D.

C	D	W	$P(W C, D)$
Pass	Pass	Pass	0.990
Pass	Pass	Failure	0.010
Pass	Failure	Pass	0.700
Pass	Failure	Failure	0.300
Fail	Pass	Pass	0.900
Fail	Pass	Failure	0.100
Fail	Failure	Pass	0.020
Fail	Failure	Failure	0.980

based on conditional independence [7]. According to the value of each node in the evidence node set E , the conditional probability distribution $P(x_{i,j}|e)$ of any node X in the Bayesian network can be obtained when different $x_{i,j}$ values

are considered. The belief propagation algorithm regards the summation operation of the variable elimination method as a message passing process, which solves the problem of repeated computation when solving multiple marginal distributions. In belief propagation algorithm, a node can send a message to another node only after receiving a message from all other nodes, and the marginal distribution of the node is proportional to the product of the message it receives [23]:

$$P(x_i) \propto \prod_{k \in n(i)} m_{ki}(x_i). \quad (5)$$

Among them, $m_{ki}(x_i)$ is expressed as follows:

$$m_{ij}(x_j) = \sum_{x_i} \psi(x_i, x_j) \prod_{k \in n(i) \setminus j} m_{ki}(x_i). \quad (6)$$

If there are no rings in the graph structure, the belief propagation algorithm can complete all message transmission in two steps:

- (1) Specify a root node that starts at all leaf nodes and delivers messages to the root node until the root node receives messages from all adjacent nodes.
- (2) From the root node, the message is transmitted to the leaf node until all the leaf nodes receive the message.

In the following section, the execution of the inference algorithm will be discussed.

5.1. Causal Inference. Causal inference is the deduction process from “cause” node to “result” node. The directed acyclic graph is represented as top-down inference of node probabilities. When the state of a node is known (evidence node), the probability distributions of its parent and child nodes are deduced. A causality analysis is conducted based on probability change, which is often used for prediction. The example of a causal inference with the Bayesian network is shown in Figure 1.

In order to infer, it is necessary to find out the total probability. The total probability is the sum of the probabilities of an event under different circumstances as

$$P(A) = \sum_{i=1}^n P(B_i)P(A|X). \quad (7)$$

For convenience, for a node, the p (+point) indicates that the student passed the course, P (-point) indicates that the student failed the course.

In the absence of some node information, the total probability of the nodes can be calculated.

$$\begin{aligned} P(+C) &= P(+M) * P(+C|+M) + P(-M) * P(+C|-M) \\ &= 0.8 * 0.850 + 0.2 * 0.550 \\ &= 0.79. \end{aligned} \quad (8)$$

$$P(-C) = 1 - P(+C) = 0.21.$$

In the absence of some node information, the probability of the students passing and failing C# is 0.79 and 0.21, respectively.

$$\begin{aligned} P(+W) &= P(+C) * P(+D) * P(+W|+C, +D) + P(+C) \\ &\quad * P(-D) * P(+W|+C, -D) + P(-C) \\ &\quad * P(+D) * P(+W|-C, +D) + P(-C) * P(-D) * P(+W|-C, -D) \\ &= 0.79 * 0.75 * 0.99 + 0.79 * 0.25 * 0.7 \\ &\quad + 0.21 * 0.75 * 0.9 + 0.21 * 0.25 * 0.02 \\ &= 0.867625. \end{aligned} \quad (9)$$

$$\begin{aligned} P(-W) &= 1 - P(+W) \\ &= 0.132375 \\ &= 0.867625. \end{aligned}$$

In the absence of some node information, the probability of the students passing and failing Web is 0.867625 and 0.132375, respectively.

$$\begin{aligned} P(+J) &= P(+C) * P(+J|+C) + P(-C) * P(+J|-C) \\ &= 0.79 * 0.900 + 0.21 * 0.600 \\ &= 0.837 \end{aligned} \quad (10)$$

$$P(-J) = 1 - P(+J) = 0.163.$$

In the absence of some node information, the probability of the students passing and failing Java is 0.837 and 0.163, respectively.

$$\begin{aligned} P(+A) &= P(+D) * P(+A|+D) + P(-D) * P(+A|-D) \\ &= 0.75 * 0.980 + 0.25 * 0.010 \\ &= 0.7375 \end{aligned} \quad (11)$$

$$P(-A) = 1 - P(+A) = 0.2625.$$

In the absence of some node information, the probability of the student passing and failing Android is 0.7375 and 0.2625, respectively.

For causal inference, in Bayesian network B, given the conditional probability of several nodes, the probability of occurrence of a node T is predicted:

- (1) For each node n in B that has not been processed, if it has the fact of occurrence, it will be marked as processed; otherwise, continue to the next step.
- (2) If one of its parent nodes has not been processed, the node is not processed; otherwise, continue to the next step.
- (3) According to the probability and conditional probability of all the parent nodes of node n , the probability distribution of node n is calculated, and the node n is marked as processed.
- (4) Repeat the above steps; the probability distribution of node T is the probability of its occurrence or not.

Now, the following question can be answered.

- (i) If the students passed Math, how likely are they to pass Java?

First, from Table 5, if the students have passed Math, the probability they have passed C# is 0.850. That is, after the

Math exam, the students can predict their C# score—if they passed Math, the probability they passed C# is 0.850; if they failed Math, the probability they failed C# is 0.550. Everyone should do their best to ensure that the probability they passed Math is high; Math is an important basic course for students majoring in C#.

$$\begin{aligned} P(+J) &= P(+C) * P(+J|+C) + P(-C) * P(+J|-C) \\ &= 0.850 * 0.900 + 0.150 * 0.600 \\ &= 0.855 \end{aligned} \quad (12)$$

If the students passed Math, the probability they passed Java is 0.855.

Now, the following question can be answered.

- (ii) If the students passed Math, how likely are they to pass Web?

From Table 5, if the students have passed Math, the probability they passed C# is 0.850. That is, after the Math exam, the student can predict their C# score—if they passed Math, the probability they passed C# is 0.850; if they failed Math, the probability they passed C# is 0.550. From Table 4, the probability the student passed Database is 0.75 and the probability they failed C# is 0.25. Thus,

$$\begin{aligned} P(+W) &= P(+C) * P(+D) * P(+W|+C, +D) + P(+C) * P(-D) * P(+W|+C, -D) \\ &\quad + P(-C) * P(+D) * P(+W|-C, +D) + P(-C) * P(-D) * P(+W|-C, -D) \\ &= 0.85 * 0.75 * 0.99 + 0.85 * 0.25 * 0.7 \\ &\quad + 0.25 * 0.75 * 0.9 + 0.25 * 0.25 * 0.02 \\ &= 0.949875. \end{aligned} \quad (13)$$

The probability they passed Web is 0.949875, if they passed Math.

Through these methods, the probability for any case can be predicted.

5.2. Diagnostic Inference. The diagnostic and support inferences are the processes of inferring the cause from the result. The reverse inference process involves passing informing from the network child node to the parent node in the network. When an event has occurred, the conditional probability distribution of “result” is used to solve the probability distribution of “cause.” It can effectively deduce the cause and cause probability. More commonly used inferences are pathological inference in the medical field and fault detection of systems and electronic devices. Further, because different semesters have different courses, it can also be used as diagnostic inference to estimate the probability of the later semester courses. An example of the diagnostic inference with the Bayesian network is shown in Figure 1.

For diagnostic inference, in Bayesian network B, given the conditional probability of several nodes, the probability of occurrence of a node T is predicted:

- (1) For each node n in B that has not been processed, if it has the fact of occurrence, it will be marked as processed; otherwise, continue to the next step.
- (2) If one of its child nodes has not been processed, the node is not processed; otherwise, continue to the next step.
- (3) According to the probability and conditional probability of all the child nodes of node n , the probability distribution of node n is calculated, and the node n is marked as processed.
- (4) Repeat the above steps; the probability distribution of node T is the probability of its occurrence or not.

Now, the following question can be answered.

- (i) If the students passed Android, how likely are they to pass Web?

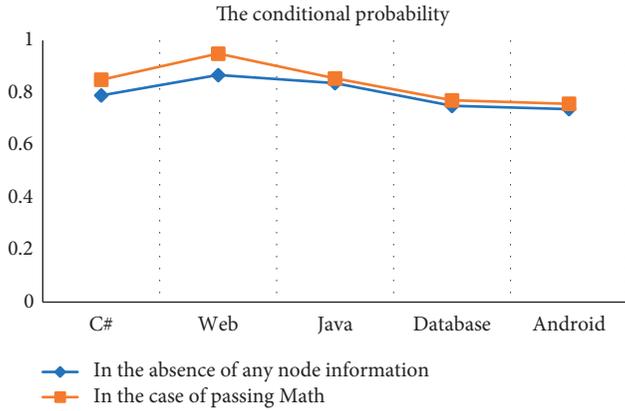


FIGURE 2: Change in conditional probability.

First, the total probability of passing Android should be calculated.

$$\begin{aligned}
 P(+A) &= P(+D) * P(+A|+D) + P(-D) * P(+A|-D) \\
 &= 0.75 * 0.980 + 0.25 * 0.010 \\
 &= 0.7375.
 \end{aligned}
 \tag{14}$$

Second, $P(+D|+A)$ should be solved. From Table 7, we can obtain $P(+A|+D)$; the famous Bayesian theorem given below can then be used [24]:

$$\begin{aligned}
 P(A|B) &= \frac{p(A)P(B|A)}{p(B)} \\
 P(+D|+A) &= \frac{P(+A|+D) * P(+D)}{P(+A)} \\
 &= \frac{0.98 * 0.75}{0.7375} \\
 &= 0.99661.
 \end{aligned}
 \tag{15}$$

$$P(-D|+A) = 0.00339.$$

Finally, the total probability of passing Web is solved.

$$\begin{aligned}
 P(+W) &= P(+C) * P(+D) * P(+W|+C, +D) + P(+C) * P(-D) * P(+W|+C, -D) \\
 &\quad + P(-C) * P(+D) * P(+W|-C, +D) \\
 &\quad + P(-C) * P(-D) * P(+W|-C, -D) \\
 &= 0.79 * 0.99661 * 0.99 + 0.79 * 0.00339 * 0.7 \\
 &\quad + 0.21 * 0.99661 * 0.9 + 0.21 * 0.00339 * 0.02 \\
 &= 0.9696879.
 \end{aligned}
 \tag{16}$$

Now, the following question can be answered.

- (ii) If the students passed Math, how likely are they to pass Database?

First, using the probability of the students passing Web (0.949875), we can calculate the probability of them passing Math.

$$\begin{aligned}
 P(+W) &= 0.949875, \\
 P(-W) &= 0.050125.
 \end{aligned}
 \tag{17}$$

Second, from Table 5, if the students have passed Math, the probability they passed C# is 0.850. That is, after the Math exam, the student can predict their C# score—if they passed Math, the probability they passed C# is 0.850; if they failed Math, the probability they passed C# is 0.550. Using the probability that they passed C#, we can calculate the edge condition probability of them passing Web, Database, and Math.

$$\begin{aligned}
 P(+W|+D) &= P(+C) * P(+W|+D, +C) \\
 &\quad + P(-C) * P(+W|+D, -C) \\
 &= 0.85 * 0.99 + 0.15 * 0.9 \\
 &= 0.9765.
 \end{aligned}
 \tag{18}$$

If they passed Math and Database, the probability they passed Web is 0.9765 and the probability they failed Web is 0.0235. Then, the condition probability of $P(+D|+W)$, using equation (13), can be solved as follows:

$$\begin{aligned}
 P(+D|+W) &= \frac{P(+W|+D) * P(+D)}{P(+W)} \\
 &= \frac{0.9765 * 0.75}{0.949875} \\
 &= 0.771023.
 \end{aligned}
 \tag{19}$$

If they passed Math, the probability they passed Database is 0.771023.

(iii) If the students passed Math, how likely are they to pass Android?

Solving this, if they passed Math, the probability they passed Database is 0.771023.

$$\begin{aligned} P(+A) &= P(+D) * P(+A|+D) + P(-D) * P(+A|-D) \\ &= 0.771023 * 0.98 + 0.228977 * 0.01 \\ &= 0.757892. \end{aligned} \quad (20)$$

If the students passed Math, the probability they passed Android is 0.757892.

The above calculation results are presented in a graph shown in Figure 2.

6. Conclusion

Bayesian networks are widely used because of their solid probability theory, flexible inference dynamic ability, and convenient decision-making mechanism. Based on the results of six courses and expert knowledge, a Bayesian network was constructed. The structure of the Bayesian network was consistent with that of expert knowledge. On the basis of structural learning and parameter learning, dynamic predictions of course performance were carried out. Our results are as follows.

When students did not take any course, the probability of passing each course was $P(+C) = 0.79$, $P(+W) = 0.867625$, $P(+J) = 0.837$, and $P(+A) = 0.7375$. Once the students passed Math, however, the conditional probability that they would pass other courses was $P(+C|+M) = 0.85$, $P(+W|+M) = 0.949875$, $P(+J|+M) = 0.855$, and $P(+A|+M) = 0.757892$.

We found that when the students passed Math, the conditional probabilities of other courses improved. This result is in line with real-world values. For students whose major is computer science, Math is the most important course. Thus, the Bayesian network based on the performance of the course can be used in the dynamic inference of course performance, and the inference results can also be used for visual programming.

It can be seen from the inference results that the performance of the ancestor node has a great impact on the performance of the offspring node. This model can be used in any school or educational management system. According to the current scores, the scores of the follow-up courses can be visualized, and the predicted results can be sent to students in time by WeChat or Email, which will greatly stimulate students' learning engagement.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the Project of the New Generation of Information Technology Innovation of Ministry of Education of People's Republic of China under grant no. 2018A02032 and 2017 Research Project on Higher Education Reform in Jiangsu Province under grant no. 2017JSJG283.

References

- [1] J. Giarratano and G. Riley, *Principle of Expert System and its Programming Design*, T. J. Yin, X. C. Liu, and Y. Tang, Eds., Machinery Industry Press, Beijing, China, 2000.
- [2] N. Friedman and M. Goldszmidt, "Building classifiers using Bayesian networks," in *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pp. 1277–1284, Portland, OR, USA, August 1996.
- [3] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, no. 2/3, pp. 131–163, 1997.
- [4] Z. Shi, "Review of expert system methods for uncertain information processing," *Systems Engineering and Electronic Technology*, vol. 6, pp. 68–76, 1990.
- [5] F. Li, "Comparison of imprecise reasoning models in expert systems," *Computer Engineering and Design*, vol. 4, pp. 52–59, 1990.
- [6] J. Pearl, "Fusion, propagation, and structuring in belief networks," *Artificial Intelligence*, vol. 29, no. 3, pp. 241–288, 1986.
- [7] J. Pearl, "Probabilistic reasoning in intelligent systems: networks of plausible inference," *SIAM Review*, vol. 32, no. 4, pp. 704–707, 1988.
- [8] N. Khakzad, F. Khan, and P. Amyotte, "Safety analysis in process facilities: comparison of fault tree and Bayesian network approaches," *Reliability Engineering & System Safety*, vol. 96, no. 8, pp. 925–932, 2011.
- [9] J. Neville, D. Jensen, and B. Gallagher, "Simple estimators for relational Bayesian classifiers," in *Proceedings of the Third IEEE International Conference on Data Mining*, pp. 609–612, IEEE, Melbourne, FL, USA, November 2003.
- [10] J. H. AlKhateeb, O. Pauplin, J. Ren, and J. Jiang, "Performance of hidden Markov model and dynamic Bayesian network classifiers on handwritten Arabic word recognition," *Knowledge-Based Systems*, vol. 24, no. 5, pp. 680–688, 2011.
- [11] G. F. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 9, no. 4, pp. 309–347, 1992.
- [12] R. R. Bouckaert, "A stratified simulation scheme for inference in Bayesian belief networks," in *Uncertainty Proceedings 1994*, pp. 110–117, Morgan Kaufmann Publishers, Burlington, USA, 1994.
- [13] W. Lam and F. Bacchus, "Learning Bayesian belief networks: an approach based on the MDL principle," *Computational Intelligence*, vol. 10, no. 3, pp. 269–293, 1994.
- [14] D. Shan, Q. Lv, F. Li, and L. Wang, "An effective mountain climbing algorithms in Bayesian network learning," *Journal of Chinese Mini-Micro Computer Systems*, no. 12, pp. 2457–2460, 2009.
- [15] C. Contaldi, F. Vafaei, and P. C. Nelson, "Bayesian network hybrid learning using an elite-guided genetic algorithm," *Artificial Intelligence Review*, vol. 52, no. 1, pp. 1–28, 2018.
- [16] F. Wang and Y. Jiang, "A hybrid algorithm for learning Bayesian network structure based on artificial bee colony and genetic algorithm," *Journal of Henan Normal University (Natural Science Edition)*, vol. 4, pp. 16–20, 2015.

- [17] R. Cao, H. Ni, and P. Zhang, "Bayesian network structure learning algorithm based on cloud genetic annealing," *Computer Science*, vol. 9, pp. 239–242, 2017.
- [18] L. Lv and Q. Lu, "Rapid method to build Bayesian network structure based on attribute order relation," *Computer Engineering and Design*, vol. 39, no. 9, pp. 2961–2966, 2018.
- [19] X. Zhou, "Research on the evaluation and analysis methods of operational effectiveness based on bayesian networks," Master's thesis, National University of Defense Technology, Changsha, China, 2008.
- [20] K. Li, Y. Han, J. Zhu, M. Tu, and L. Fan, "Predicting duration of traffic accidents based on cost-sensitive Bayesian network and weighted K-nearest neighbor," *Journal of Intelligent Transportation Systems*, vol. 23, no. 4, pp. 1–14, 2019.
- [21] Y. Wu, W. Xu, Q. Yu, J. Feng, and T. Lu, "Hierarchical Bayesian network based incremental model for flood prediction," in *Proceedings of the International Conference on Multimedia Modeling*, pp. 556–566, Thessaloniki, Greece, January 2019.
- [22] H. Vundavilli, A. Datta, C. Sima, J. Hua, R. Lopes, and M. L. Bittner, "Bayesian inference identifies combination therapeutic targets in breast cancer," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 9, pp. 2684–2692, 2019.
- [23] H. Zhou, *Machine Learning*, Tsinghua University Press, Beijing, China, 2015.
- [24] T. Bayes, "LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S," *Philosophical Transactions of the Royal Society of London*, vol. 53, pp. 370–418, 1763.