Prediction Method of College Students' Suicidal Tendency Based on Bayesian Network

Jiajun She[1, a], Wenxin Yang[2, b], Zuyi Liang[1, c], Xinyu Tan[2, d], Weihua Li[2, e, *]

[1]Institute of Rail Transit, School of Electrical Engineering, Jinan University, Zhuhai 519000, China;

[2]School of Electrical Engineering, Jinan University, Zhuhai 519000, China.

[a]13676087584@163.com, [b]y898802745@qq.com, [c]994047630@qq.com, [d]994047630@qq.com, [e]liweihua115@163.com

Abstract

The use of questionnaires to predict the suicidal tendency of college students is difficult to analyze, difficult to analyze all students individually, and the workload is large. This paper proposes a prediction method of college students' suicidal tendency based on Bayesian network, analyzes the suicide tendency questionnaires of all students separately, and screens out students whose suicidal tendency is higher than the threshold to facilitate timely psychological intervention. This paper firstly made an electronic version of the questionnaire based on the College Student Suicide Tendency Questionnaire compiled by Zheng Aiming and others in the Chinese Journal of Clinical Psychology, and then conducted a suicide tendency questionnaire survey on 305 students from the university and another university in Guangdong. Then we trained the built Bayesian network based on the survey sample. The results show that the Bayesian model proposed in this paper has high prediction accuracy and practicality for college students' suicidal tendency, and can provide theoretical basis for active suicide prevention and crisis intervention.

Keywords

Bayesian network, suicide tendency prediction, questionnaire.

1. Introduction

In the group of college students, due to the increasing academic, economic, interpersonal and emotional pressures and problems, some college students have gradually experienced different levels of psychological crisis, resulting in frequent suicides among college students. According to reports, the upper limit of suicidal ideation of Chinese college students can reach 40%, and suicidal ideation has become the primary factor of suicide [1]. The questionnaire survey method is used to predict the suicidal tendency of college students. After each questionnaire survey, how to accurately and quickly analyze thousands of questionnaires, and psychological intervention for students with high suicidal tendencies needs to be solved.

At this stage, after using the suicide tendency questionnaire survey on college students, a large number of questionnaire data is usually based on SPAA statistical software for general descriptive analysis, which can only roughly reflect the total suicidal tendency of the surveyed college students, but not individual students. This will not allow timely psychological interventions for a small number of students with a high suicidal tendency, which may lead to the emergence of dangerous behaviors, resulting in irreparable consequences.

In view of the above problems, this paper proposes a Prediction method of college students' suicidal tendency based on Bayesian network. By establishing and training the Bayesian network, each suicide tendency questionnaire can be quickly analyzed and the suicidal

tendency of the students participating in the questionnaire can be given. The experimental results show that the method is relatively accurate for predicting suicide tendency of college students, and the Bayesian network proposed in this paper has short modeling time and good practicability.

2. Basic Overview of Bayesian Networks

Bayesian network is an inductive reasoning method discovered by British pastor Bayes. It is a new reasoning method developed on the basis of estimation and hypothesis testing. Compared with the classical statistical inductive reasoning method, the Bayesian network needs to drive the sample information based on the knowledge and experience of the inferencer's past and draw conclusions [2,3].

When analyzing the suicide tendency questionnaire of college students, because each student's questionnaire answers have different probability, the questionnaire results have uncertain factors. The advantage of using Bayesian network is that it can handle intermittent and uncertain behavior data at any time. . Through modeling, the causal relationship between various sample information is established. The determination of causality enables us to make inferences and judgments on the questionnaires of students with high suicidal tendencies.

The Bayesian network, also known as the belief network, consists of three parts:<X,A,Θ>. Where <X, A> represents a structure G of a directed acyclic graph (DAG). X is a collection of nodes in the network, Xi∈X represents a random variable that limits the domain; A is a set of directed edges in the network, aij∈A represents a direct dependency between nodes, aij indicates the directed connection between Xi and Xj, Xi←Xj; Θ is the network parameter, θi ∈Θ represents the conditional probability distribution function associated with node Xi. The Bayesian network implies a conditional independence hypothesis that gives a node's set of parent nodes independent of all its non-descendant nodes. Therefore, the joint probability of all nodes represented by the Bayesian network can be expressed as the product of the conditional probability of each node, as shown in the following equation (1).

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

Where i=1, 2,...,n,π(Xi) represents the Xi parent node set. When the network structure G is given, the correlation between the nodes is also determined. Under this premise, combined with the network parameters Θ, a Bayesian network can uniquely determine the joint distribution probability of node X and obtain the inference result. Due to the conditional independence between nodes, the computational efficiency of Bayesian networks is much higher than other methods of calculating joint probability. use.

3. Preparation of the Suicide Tendency Questionnaire for College Students

Before building a Bayesian network model, we need to use questionnaires to collect the relevant data needed in the model to derive the prior probability of the top-level events in the model. In this paper, the paper on the suicide tendency of college students was compiled with reference to the paper by Zheng Aiming et al. (Zheng, Chen, Zhu, & Fu, 2012) [4] in the Chinese Journal of Clinical Psychology. There are 34 entries in the questionnaire. There are 4 factors in the questionnaire, which are Depression Levels (1-15 entries), Suicide Cognition (16-23 entries), Suicide Preparation (24-31 entries), and Despair (32-34 entries), the entries for this questionnaire and the factor loading for each entry in the questionnaire are shown in the following Figure 1.

Entry	Depression Level	Suicide Cognition	Suicide Preparation	Despair
1. I am disappointed with the future	0.80			
2. I want to commit suicide	0.78			
3. I feel that I am still dead.	0.78			
4. I feel that I have no ability to deal with any problems.	0.78			
5. I am not interested in everything.	0.72			
6. I am disappointed with myself	0.71			
7. If I die, others will live better.	0.68			
8. I am more easily irritated than usual.	0.66			
9. I feel it is difficult to make a decision.	0.66			
10. I am restless and difficult to keep calm	0.66			
11. I am doing things as hard as usual.	0.64			
12. My mind feels more confused than usual	0.62			
13. I feel tired for no reason	0.57			
14. I want to cry	0.49			
15. I feel depressed	0.49			
16. I don't think there will be suicide without difficulty.		0.70		
17. I think suicide will affect others.		0.68		
18. I think suicide can end the current pain.		0.65		
19. I think suicide can punish my fault.		0.60		
20. I think suicide can attract the attention of others.		0.56		
21. I think that suicide can retaliate against some people.		0.56		
22. I think suicide is a weak performance.		0.56		
23. I think we should cherish people's lives.		0.50		
24. I made all the preparations for suicide.			0.65	
25. I have hurt myself or committed suicide			0.65	
26. I am asking for help from people around me because of suicide.			0.64	
27. I have tried suicide			0.64	
28. I considered the time, place and method of suicide			0.64	
29. I hope that I will be killed in a car accident or other catastrophic event.			0.59	
30. I hope to end my life in a painless way.			0.56	
31. I am worried that I will put the idea of suicide into action.			0.49	
32. I feel that I am too tired to live.				0.67
33. I feel that I am a useless person.				0.81
34. My life has no meaning				0.47

Fig 1. Student suicide tendency questionnaire entries and factor load of each item

In this questionnaire, we also need to know the correlation between the four factors of the questionnaire and the correlation between each factor score and the total score. According to Zheng Aiming et al.'s paper, we can see that there is a significant positive correlation between factor scores, factor scores and total scores, as shown in Figure 2 below.

	Depression Level	Suicide Cognition	Suicide Preparation	Despair
Depression Level	1			
Suicide Cognition	0.435**	1		
Suicide Preparation	0.465**	0.523**	1	
Despair	0.287*	0.227*	0.211*	1
Total score	0.788**	0.723**	0.628**	0.415**

P.S. *P<0.05, **P<0.01

Fig 2. Correlation matrix of each factor and total score

4. Establishment and Training of Bayesian Network Model

4.1. Model Establishment

After assuming the independence condition, this paper uses the form of causal reasoning to derive the result from the top down and the cause, and find the probability of the result in the case of the cause. In order to establish a Bayesian network-based suicide tendency prediction model for college students, the tool used in this paper is Netica, which is a widely used Bayesian network analysis software. Netica provides three types of algorithms that can be used to calculate deductive conclusions, namely sample data statistics, expected maximum and gradient descent. This paper will build a model based on the functions provided by Netica using a combination of manual and sample data statistics [5]. This paper designs a Bayesian network model suitable for predicting the suicidal tendency of the questionnaire results. The nodes of the constructed model have 4 layers.

The nodes on the first level are 34 entries in the suicide tendency questionnaire. The node probability of this layer is the prior probability, which is determined according to the actual questionnaire.

The nodes of the second layer are 9 sub-factors. This layer constructs the corresponding 9 sub-factors as the node content according to the factor load of each item of the questionnaire (Depression Level, Suicide Cognition, Suicide Preparation), and divides the entries with similar factor load into one sub-factor. Among them, the factor "depression level" divided into 5 sub-factors (depression level 1, depression level 2, depression level 3, depression level 4, depression level 5). The first 4 entries of the questionnaire correspond to the sub-factor "depression level 1", the 5th and 6th entries of the questionnaire correspond to the sub-factor "depression level 2", and the 7th to 10th entries of the questionnaire correspond to the sub-factor "depression level 3" . The 11th to 13th entries of the questionnaire correspond to the sub-factor "depression level 4", the 14th and 15th entries of the questionnaire correspond to the sub-factor "depression level 5"; the factor "suicide cognition" is divided into 2 sub-factors (suicide cognition 1, Suicidal cognition 2). The 16th to 19th entries of the questionnaire correspond to the sub-factor "Suicide Cognition 1", the 20th to 23th entries of the questionnaire correspond to the sub-factor "Suicide Cognition 2"; the factor "Suicide Preparation" is divided into 2 sub-factors (Suicide Preparation 1, Suicide preparation 2), the 24th to 28th entries of the questionnaire correspond to the sub-factor "Suicide Preparation 1", and the 29th to 31st entries of the questionnaire correspond to the sub-factor "Suicide Preparation 2". Since the factor “despair” of the questionnaire corresponds to only three entries, it is not necessary to separate the sub-factors in this layer.

The third layer of nodes is the four factors of the questionnaire (Depression Level, Suicide Cognition, Suicide Preparation, and Despair). This layer takes the four factors of the questionnaire as the node content, and the node probabilities of the factors "Depression Level", "Suicide Cognition", and "Suicide Preparation" are obtained by weighting their corresponding sub-factors. The node probability of the factor "Despair" is determined by its corresponding three entries.

The content of the fourth layer of nodes is suicidal tendencies. There is only one node in this layer, and its probability is derived based on the proportion of the four factors in the questionnaire. In the above four-layer structure, after determining the content of the node, the node state must be determined. According to the meaning defined by the content of each layer node, the correspondence between the state value and the node variable determined in this paper is as shown in Table 1 below.

Table 1. Correspondence between node variables and state values

level	Node name	Entry1	Entry2	Entry3	…	Entry34
one	State value	Yes	Yes	Yes	…	Yes
		No	No	No	…	No
level two	Node name	DepressionLevel(1-5)		SuicideCognition(1-2)		SuicidePreparation(1-2)
	State value	Present		Present		Present
		Absent		Absent		Absent
level three	Node name	DepressionLevel	SuicideCognition	SuicidePreparation		Despair
	State value	Present	Present	Present		Present
		Absent	Absent	Absent		Absent
level four	Node name	SuicidalTendency				
	State value	Present				
		Absent				

After determining the node content and the node relationship, the next step is to perform probability allocation. Probability allocation includes two parts: one is to specify the prior probability for the top-level event without the parent node; the second is to specify the conditional probability for the event with the parent node.

When assigning the node probability of the first layer, this paper mainly calculates the prior probability based on the results of the college students' suicidal tendency questionnaire survey at the author's university and surrounding universities. The survey objectives of this survey mainly include students from two universities, a total of 305 students. Then based on the results of the survey, we can calculate the corresponding probability of the answer to each entry in the questionnaire and assign it to the first layer of the model.

The probability distribution method of the second layer node is as follows: Firstly, it is necessary to determine the influence probability ratio between the entries corresponding to the sub-factors, and the influence probability ratio between the items is mainly determined according to the factor load of each entry of the questionnaire. In the four entries corresponding to the sub-factor "depression level 1", the influence probability ratio between the four entries can be calculated as 40:39:39:39 according to the corresponding factor load value of each entry. Similarly, the probability ratio of the two entries corresponding to the sub-factor "depression level 2" can be calculated as 72:71. The probability ratio of the four entries corresponding to the sub-factor "depression level 3" is 34:33:33:33. The probability ratio of the three entries corresponding to the sub-factor "depression level 4" is 64:62:57. The probability ratio of the two entries corresponding to the sub-factor "depression level 5" is 1:1. The influence probability ratio of the four items corresponding to the sub-factor "Suicide Cognition 1" is 78:68:65:60; the influence probability ratio of the four items corresponding to the sub-factor "Suicide Cognition 2" is 28:28:28: 25; The probability ratio of the five entries corresponding to the sub-factor "Suicide Preparation 1" is 65:65:64:64:64; the probability ratio of the three entries corresponding to the sub-factor "Suicide Preparation 2" is 59:56: 49. After determining the influence probability ratio of each entry, the node probability of each sub-factor in the second layer node can be calculated by combining the maximum factor load value in the entry corresponding to each sub-factor.

The probability distribution method of the third layer node is similar to that of the second layer. Firstly, it is necessary to determine the influence probability ratio between the sub-factors corresponding to the factor (the probability distribution method of the factor "despair" is consistent with the second-layer node probability distribution method) The probability of influence between the sub-factors is calculated based mainly on the average factor load value

of the entry corresponding to each sub-factor. Therefore, the probability ratio of influence between the five sub-factors corresponding to the factor "depression level" is 157:143:133:122:98. Similarly, the probability ratio of the sub-factors corresponding to the factor "suicide cognition" can be calculated as 263:218. The probability ratio of the sub-factors corresponding to the factor "suicide preparation" is 483:410. The factor "despair" can be calculated according to the factor load value of the three items directly corresponding to it, and the influence probability ratio of the three items is 67:81:47. After determining the influence probability ratio of each item, the node probability of each factor in the third layer node can be calculated by combining the maximum load value of the sub-factor corresponding to each factor.

The node probability calculation of the fourth layer node "suicidal tendency" first needs to determine the influence probability ratio between each factor according to the correlation matrix of each factor and total score in the questionnaire. The probability ratios of the four factors "depression level", "suicide cognition", "suicide preparation", and "despair" in the questionnaire were set to 788:723:628:415. Then, with '1' as the maximum probability value of "suicidal tendency", the node probability of the third layer node can be calculated.

According to the node content and node probability determined above, a DAG of a 4-level structure can be constructed, as shown in Figure 3 below.

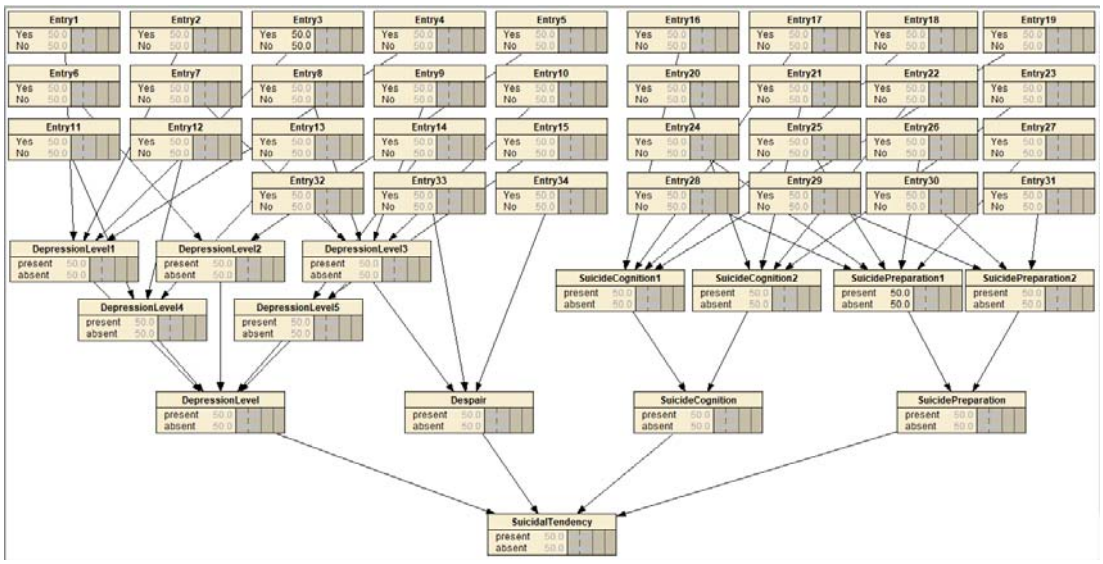

Fig 3. DAG initial map of college students' suicide tendency prediction model

Each box in the DAG represents an independent node variable of the college student suicide tendency prediction model, which includes the name, type, and status. Before training the DAG diagram, you need to complete the input of the state of the node variable. After determining the logical relationship and the influence probability ratio between each node variable and the precursor node, the calculated probability value is input into the DAG, and the initial probability of each node variable is obtained. As shown in Figure 4 below, the initial probability of suicidal tendency was 19.2%.

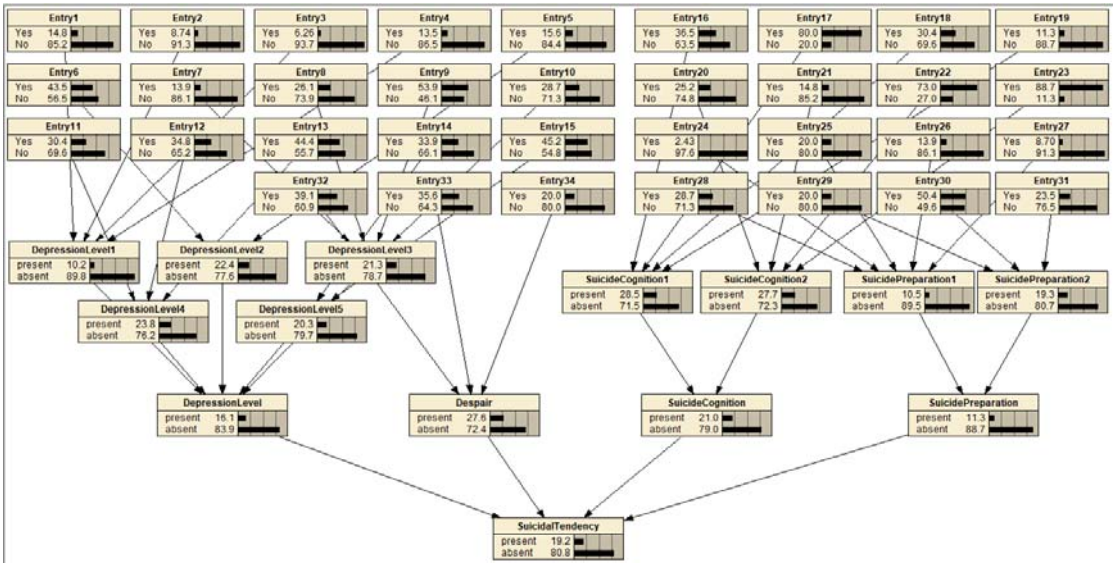

Fig 4. DAG initial probability map of college students' suicide tendency prediction model

4.2. Model Training

After the successful construction of the college student suicide tendency prediction model, we can see from Fig. 2 that the initial probability of college students' suicidal tendency is 23.6%. After the model is built, the next step is to train and analyze the model to predict the suicidal tendency of college students in practical applications, and finally prevent the occurrence of dangerous behaviors.

4.2.1. Prediction Analysis

When training the model, we can input the sample data used in this survey, or we can find new students to conduct a questionnaire survey, and then import their corresponding data for analysis. During the training model phase, we found two students with depression at the psychological counseling center of the author's university. One of the students was diagnosed with depression half a month ago and the other was already in the second stage of treatment.

After the questionnaire is surveyed for the first student, the answer of each questionnaire of the student is input into the model, that is, the initial probability value of the corresponding item is set to "Yes=100" or "No=100" according to the answer of the questionnaire. Then through the automatic deduction calculation, the first student's suicidal tendency was 45.8%. Then, based on the evidence that the first student suffered from depression, the initial probability value of the factor "depression level" in the third layer node was set to "present=100%", indicating the known state of the known evidence variable. Through the calculation of the post-drive node, the probability of suicidal tendency is 60.7%. The training scene is shown in Figure 5 and Figure 6 below.

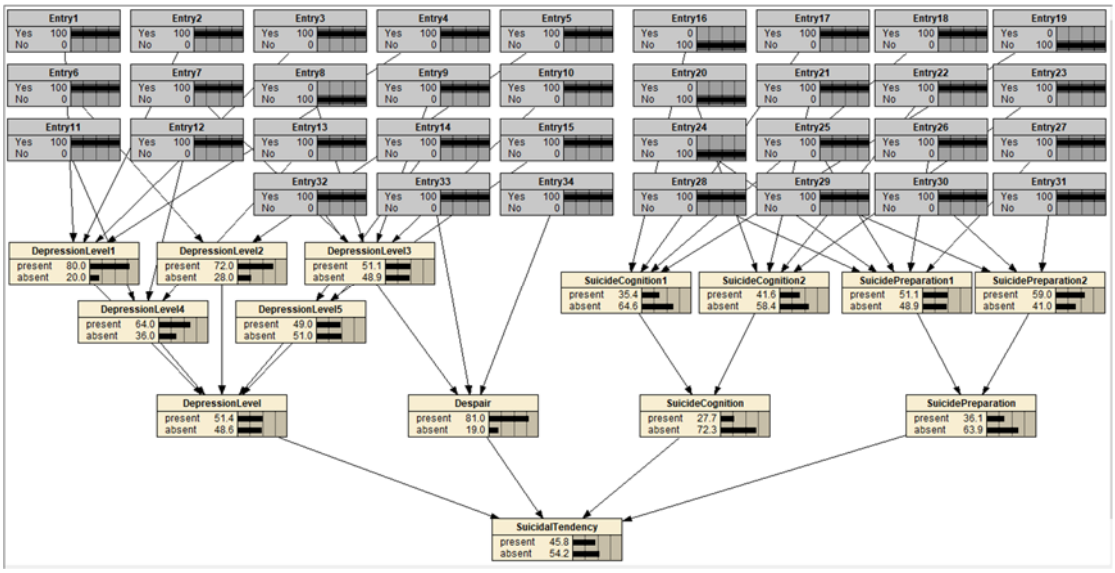

Fig 5. DAG of the first classmate test result

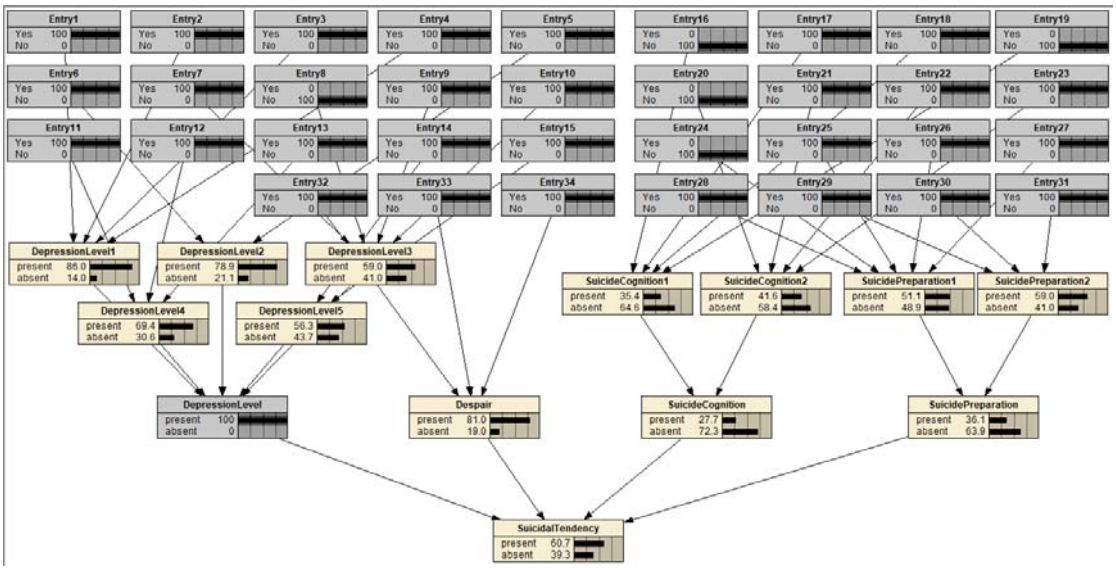

Fig 6. DAG training chart after new evidence is added

Similarly, the second student was surveyed using the same process and the results were entered into the model. Then through the automatic deduction calculation, the second student's suicidal tendency was 19.8%. Although the second student was already in the second stage of treatment for depression, but still did not heal, so the second student was still judged to have depression. The initial probability value of the factor "depression level" in the third layer node is set to "present=100%", indicating the known state of the known evidence variable. Through the calculation of the post-drive node, the probability of suicidal tendency is 43.1%. The training scene is shown in Figure 7 and Figure 8 below.

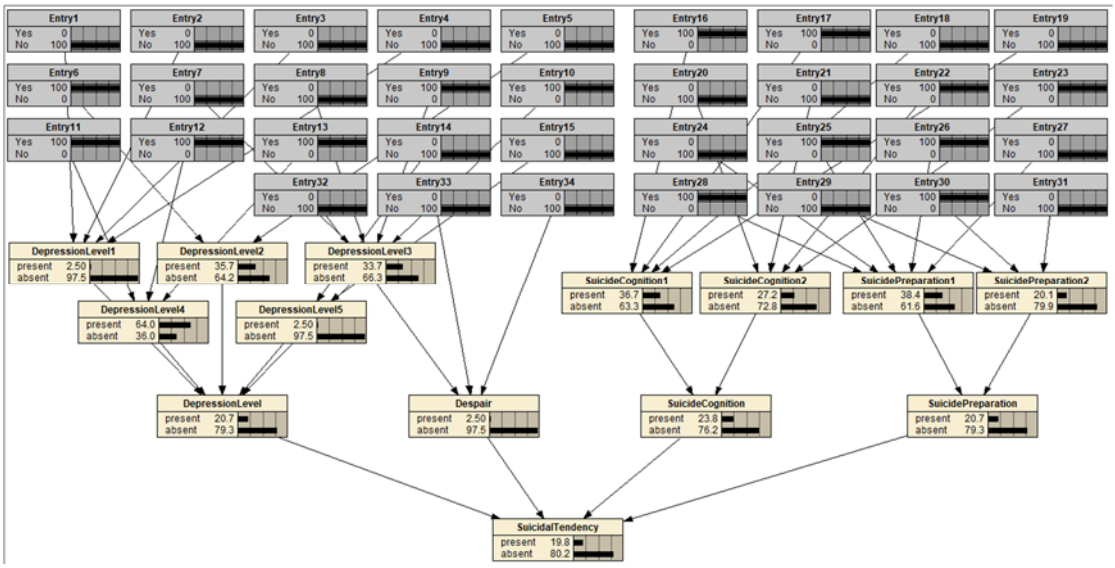

Fig 7. DAG of the second classmate test result

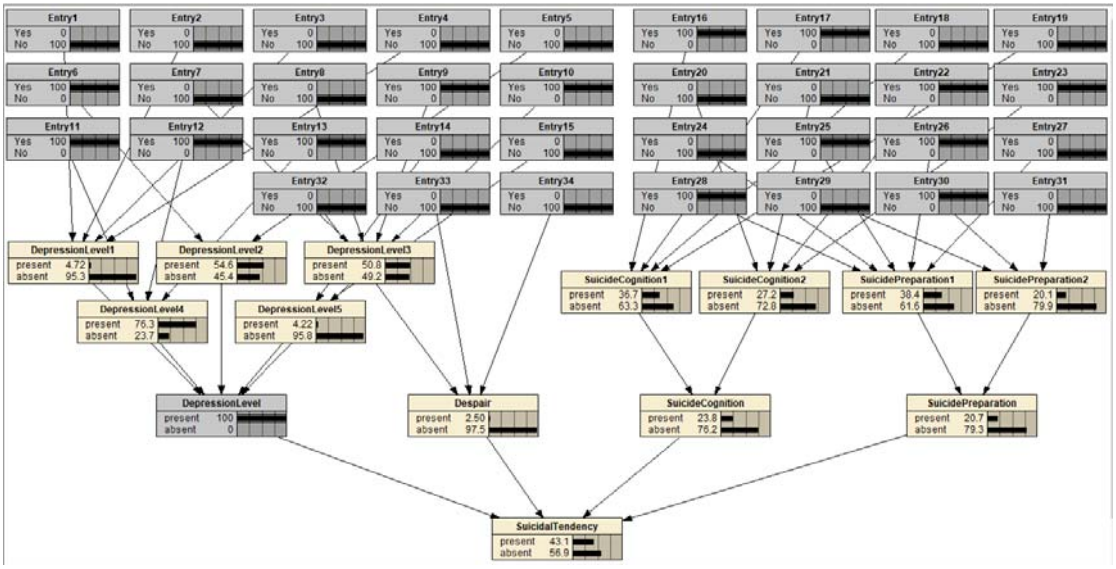

Fig 8. DAG chart after the new evidence is added

From the above tests of two students at different stages of depression, it can be found that for each questionnaire, as the number of "Yes" in the answer increases, the probability of "present" for the four factors will also increase, leading to the probability of suicidal tendencies also increases.

4.2.2. Reason Judgment

Through simulation, investigation, training, and analysis, the number of "Yes" in the answers of the first students was significantly higher than that of the second student. The probability of the suicide tendency of the post-drive node variable was also 17.6% higher than that of the second student. Through the actual psychological test of the first student, it also verified that the student has a higher suicidal tendency, which is consistent with the conclusion of the model algorithm with 60.7% suicidal tendency. Through the actual psychological test of the second student, it is basically consistent with its low suicidal tendency, which is in line with the conclusion of the model algorithm that 43.1% suicidal tendency is more consistent.

5. Conclusion and Thinking

The Bayesian network-based prediction method of college students' suicidal tendency can make up for the shortcomings of the traditional data processing of suicide tendency questionnaires. By importing the questionnaire data corresponding to each student into the model, the suicidal tendency of the student can be relatively accurately determined, so that outsiders can promptly intervene in the student.

Through repeated training and analysis of the Bayesian model, we can use 45% as the threshold for suicide tendency 'high' or 'low'. If the tested students have a suicidal tendency of more than 44%, they need outsiders to promptly intervene in the students to avoid the students as dangerous behaviors.

The use environment of the college students' suicide tendency prediction method proposed in this paper can be used in various universities. We can conduct suicide tendency questionnaires for college students on a regular basis, and then systematically analyze each student's suicidal tendency through Bayesian network by inputting each student's questionnaire answers into the model. At the same time, in the investigation and analysis of patients with depression, the third level of "depression level" can be set as "present=100%" as supplementary evidence, and the results can be used as one of the evidences of whether a depressed patient has a good psychological condition.

Acknowledgements

This work was supported by the Jinan University Rail Transit Research Institute.

References

[1] Zhang, B., & Duan, C. B., (2014). An Overview of College Students' Suicide Ideas. Youth and society, 62(10), 55-55.

[2] Li, S. H., & Zhang, J. (2015). Review of Bayesian networks structure learning. Application Research of Computers, 32(03), 641-646.

[3] Kabir, G., Suda, H., Cruz, A. M., Giraldo, F. M., & Tesfamariam, S. (2019). Earthquake-related Natech risk assessment using a Bayesian belief network model. Structure and Infrastructure Engineering, 15(6).

[4] Zheng, A. M., Chen, T. N., Zhu, T. T., & Fu, H. (2012). Development of College Students Suicide Inclination Questionnaire. Chinese Journal of Clinical Psychology, 20(06), 756-758.

[5] Chen, J., Jiang, Z. K., & Fu, J. Q. (2016). Construction of self-learning Bayesian network based on Netica. Journal of Electronic Measurement and Instrumentation, 30(11), 1687-1693.