A Landslide Stability Calculation Method Based on Bayesian Network

Jiang Tingyao
The College of Computer and Information Technology,
China Three Gorges University, Yichang, Hubei
Province, China
e-mail: jiangty@ctgu.edu.cn

Wand Dinglong
The College of Computer and Information Technology,
China Three Gorges University, Yichang, Hubei
Province, China
e-mail: 2007114102@ctgu.edu.cn

*Abstract***—The calculation method of landslide stability is a critical issue to landslide research. Because the landslide is an unbalanced and unstable complex system, meanwhile the interactions among the various factors that composed a landslide are uncertain and random, the landslide stability calculation method based on probability becomes the future trend. This paper presents a new landslide stability calculation method based on Bayesian Network, which applies K2 algorithm and Bayesian method to learn Bayesian network's structure and parameters. In the established Bayesian network model a joint tree inference algorithm is used to analyze and calculate the landslide stability under the effects of slope height, slope angle, bulk density, angle of internal friction, cohesion and etc. The paper uses 5-fold Cross-Validation to verify the accuracy in the proposed model. Compared with the calculation method based on Support Vector Machine (SVM), its reliability is higher and prediction results are better. The proposed model can directly represent the interaction mechanism between the decision-making behavior and effect factors.**

Keywords- Bayesian network;Landslide research; Stability calculation

I. INTRODUCTION

The landslide is a kind of important geological disaster, which poses a serious threat to the national economy and people's lives and property. China is one of countries that the geological disasters occurred frequently and severely. In order to avoid or reduce landslide disasters on people's lives and property loss or the impact of construction, we need to analyze and forecast the landslide stability for the later protection ready. Landslide is affected by many factors and evolved to be a nonlinear dynamical system. Predicting landslide is a challenging task, which makes the calculation method of landslide stability become a hot issue of landslide research.

The main content of landslide stability research is to calculate the landslide stability. Several methods have been proposed [1-8], which can be divided into three categories: numerical analysis, limit equilibrium method and probabilistic method. Limit equilibrium method is the earliest and the most commonly used quantitative analysis method at present. The advantages of this method are simple and intuitive, and it has many years of practical experience, seizing main aspects of the problem. However, this method makes simplifying assumptions in mechanics and oversimplifies the landslide as rigid. Numerical analysis includes the finite element method, the boundary element method, the discrete element method, the interface element method, the unbounded element method, etc. This method avoids the shortcoming of the limit equilibrium method that it oversimplified the landslide as the rigid body, and approximately analyzes deformation and damage mechanisms of the slope from the stress and strain [9]. Limit equilibrium method and Numerical analysis is mathematics-mechanical model and deterministic model, which will not be able to calculate without fixed parameters. Actually landslide's calculation parameters have uncertainty and randomness, and landslide is an uneven, unstable, and complex system, which has continuous exchanges with the external environment of material, energy and information [10]. The stochastic characteristics of slope's elements were introduced in probabilistic analysis in the early 1970s. It was considered that strength parameters of the slope analysis were consistent with a probability distribution function $F=(c,\varphi,\gamma,u,H,\mu\cdots)$. The concept of safe limits was also introduced and Janes combined it with the maximum entropy principle for calculating the failure probability of each slider, and then calculate the unstable probability of the entire landslide [8].

Since all the influence factors of landslide are uncertain, and a certain relationship between them. In order to analyze their effects on the landslide stability, this paper presents a calculation method for landslide stability based on Joint probability. The calculation method calculate the stability of landslide state by the probability of occurrence, through constructing a Bayesian network model which reflects the interdependent relationship between variables and analyzing the uncertain relationship between the various factors affecting landslide stability.

This paper defines a Bayesian network model that contains five variables affected the landslide stability. It begins with identifying factors variables affected the landslide stability, as well as the relationships between these variables. Then it uses the K2 algorithm to learn the Bayesian network model, and uses Bayesian method for parameters learning to obtain the conditional probability (CPT). Finally, it validate the model by using 5-fold Cross Validation. Compared to support vector machine, the results indicate that this model is better.

II. BAYESIAN NETWORK

Bayesian network is a modeling and analyzing tool to solve the problem of uncertainty in artificial intelligence, called probabilistic causal network. Probabilistic inference is one of its based tasks. Bayesian network provides a reasoning method based on a probability distribution to support and cause discovery in an uncertainty environment or under incomplete information. Bayesian network theorem is a result with the conditional probability and marginal probability distribution of the random variables from probability theory, which can represent the joint probability distribution function of a group of variables in the graph [11-14].

A typical Bayesian network consists of directed acyclic graph (DAG) and conditional probability table (CPT) [15]. Markov nature and chain rule applied in Bayesian network greatly simplifies the process of solving the joint probability. Bayesian network modeling is mainly divided into two processes, the qualitative and quantitative process to a specific question. Qualitative process is to solve the problem of the relationship between variables, build topology of the network. Quantitative process is to solve the conditional probability of each node in the network.

Bayesian network has the following advantages:

1) Bayesian network can really effectively deal with incomplete data;

2) Combing with other technology, Bayesian network can perform causal analysis;

3) Bayesian network can make a priori knowledge and data combination of organic;

4) Bayesian network can effectively avoid overfitting the data.

III. BAYESIAN NETWORK MODEL FOR PREDICTING LANDSLIDE STABILITY

There are many factors that affect the landslide stability, divided into internal factors and the external influence factors. The internal factors mainly include: landform, stratum structure, rock and soil characteristics, in-situ stress, hydrological characteristics, vegetation conditions, etc; The external factors mainly include: climatic conditions, groundwater dynamics, vibration situation, artificial engineering, etc. But when it came to certain areas and certain type of landslide, we need to do some trade-offs or refinement on influence factors [16]. This paper mainly takes soil bulk density $\gamma(kN/m^3)$, cohesion $c(kPa)$, internal friction angle $\varphi(°)$, slope angle $\psi(°)$, slope height $H(m)$ as the main influence factors of the landslide stability. We regard the state of the stability as the class variable, others are attribute variables, that is, the Evaluation index of the stability. And the state of the landslide stability is divided into 5 classes: stable, relatively stable, generally stable, unstable, extremely unstable, represent in the numbers 1, 2, 3, 4, 5. Table Ⅰ lists the relationship between evaluation index and the corresponding rank.

TABLE I. THE RELATIONSHIP BETWEEN EVALUATION INDEX AND THE CORRESPONDING RANK

Stability rank / Evaluation index	1	2	3	4	5
	stable	relatively stable	generally	unstable	extremely unstable
slope height(m)	<6	6~12	12~18	18~24	>24
angle(°)	<15	15~25	25~35	35~45	>45
bulk density (kN/m^3)	>18	17~18	16~17	15~16	<15
internal friction angle(°)	>35	30~35	25~30	20~25	<20
cohesion (kPa)	>50	40~50	30~40	20~30	<20

This paper processes as following for the rank of state of stability: The stability of grade 1, 2, 3 is processed into the stable of the landslide, set a value of 1; the stability of grade 4, 5 is processed into sliding state, set the value of 2.Using the rules in the table1, then divided the sample data into 5 sections, and represented the section by 1, 2, 3, 4, 5

A. Date collection and pretreatment

In this paper, the data in literature [17] and the literature [9] is used as the training data and predicted data. According to the above rules this paper disperses data collected, using 2 represent sliding and 1 represent stable. The influencing factors are respectively 1, 2, 3, 4, 5 according to the grading rules and practical data interval.

B. Structure learning

This paper uses K2 algorithm to do structure learning which uses greedy algorithm to deal with the model selection problem: firstly, define a score function of evaluating the network structure, and then, start from a network, according to the predetermined maximum number of the parent node and the node order, choose the highest score as the node's parent node. K2 algorithm uses the posterior probability as a score function:

$$p(D \mid B_s) = \prod_{i=1}^{n} score(i, pa_i) \tag{1}$$

Where $score(i, pa_i) = \prod_{j=1}^{q_i} [\frac{\Gamma(\partial_{ij})}{\Gamma(\partial_{ij} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\partial_{ijk} + N_{ijk})}{\Gamma(\partial_{ijk})}]$

K2 pseudo-code is as follows:

Algorithm1 $k2(X, \rho, \mu, \vartheta)$

Input:

$X = \{X_1, X_2, ..., X_n\}$: A set of variables

ρ : A variable order(Let it consistent with the variable subscript)

μ : Variable upper bound on the number of father nodes

ϑ : A complete set of data

Output:

A Bayesian Network

```
1. ζ ← boundless graph is consisted by nodesX1, X2, . . . , Xn 2. for j=1 to n
3.   π_j ← φ;
4.   V_old ← CH(<X_j, π_j >| ϑ);
5. while (true)
6.   i ← arg max_{1≤i<j, X_i∉π_j} CH(<X_j, π_j ∪ {X_i} >| ϑ)
7.   V_new ← CH(<X_j, π_j ∪ {X_i} >| ϑ)
8.   if(V_new > V_old and |π_j|<μ)
9.     V_old ← V_new ;
10.    π_j ← π_i ∪ {X_i} ;
11.    add X_i → X_j in ϑ ;
12.  else
13.    break;
14.  end if
15. end while
16. end for
17.Estimate ζ 's parameter θ ;
18. return ( ζ , θ );
```

This paper uses the K2 algorithm of GeNIe2.0 software to learn the Bayesian Network model, after repeated variable selection and sorting adjustment, here comes the final structure graph shown in fig.1.This network structure consists of six nodes and several connections. Six nodes are six variables, including a variable been going to be analyzed, namely the stability state. Connections between the nodes indicate the mutual influence between them. Cohesion and slope angle affect the landslide stability together shown in Figure. 1.

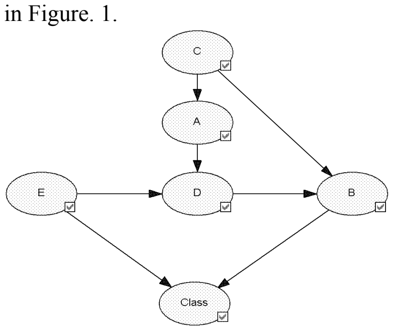

Figure 1. A-slope height, B- slope angle, C- bulk density, D-internal friction angle, E- cohesion, Class-landslide stability.

C. Parameters learning

Parameter learning is to determine the conditional probability distribution of each node under a given Bayesian Network topology. There are two common parameter learning methods: Maximum likelihood estimation (MLE) and Bayesian method. Where the basic idea of Bayesian Method is: given a distribution with location parameter and a complete collection of instance data *c*, θ is a random variable with a prior distribution $p(\theta)$, can be estimated based on previous knowledge, or that $p(\theta)$ is a uniform distribution. The information of parameter *c* has changed that represents by *p*, called the posterior probability of the parameter *c*. The task of Bayesian parameter learning is to compute the posterior probability.

Take $p(\theta)$ as Dirichlet distribution:

$$p(\theta) = Dir(\theta \mid \alpha_1, \alpha_2, \cdots \alpha_r) = \frac{\Gamma(\alpha)}{\prod_{k=1}^{r} \Gamma(\alpha_k)} \prod_{k=1}^{r} \theta_k^{\alpha_{k-1}} \quad (2)$$

Where $\alpha = \sum_{k=1}^{r} \alpha_k, \alpha_k > 0, k = 1, 2, \cdots r; \alpha_1, \alpha_2, \cdots, \alpha_r$ called Hyper Parameter. $\Gamma(\alpha)$ Gamma function; and *Dir*(.) Dirichlet distribution function.

Sample Probability is:

$$p(D) = \int p(\theta) p(D \mid \theta) d\theta = \int \frac{\Gamma(\alpha)}{\prod_{k=1}^{r} \Gamma(\alpha_k)} \prod_{k=1}^{r} \theta_k^{\alpha_k - 1} \times \prod_{k=1}^{r} \theta_k^{N_k} d\theta$$
$$= \frac{\Gamma(\alpha)}{\Gamma(\alpha + N)} \cdot \prod_{k=1}^{r} \frac{\Gamma(\alpha_k + N_k)}{\Gamma(\alpha_k)} \quad (3)$$

The posterior probability of parameter also is Dirichlet distribution:

$$p(\theta \mid D) = \frac{p(\theta) p(D \mid \theta)}{p(D)} = \frac{\frac{\Gamma(\alpha)}{\prod_{k=1}^{r} \Gamma(\alpha_k)} \prod_{k=1}^{r} \theta_k^{\alpha_{k-1}} \times \prod_{k=1}^{r} \theta_k^{N_k}}{\frac{\Gamma(\alpha)}{\Gamma(\alpha + N)} \cdot \prod_{k=1}^{r} \frac{\Gamma(\alpha_k + N_k)}{\Gamma(\alpha_k)}}$$
$$= \frac{\Gamma(\alpha + N)}{\prod_{k=1}^{r} \Gamma(\alpha_k + N_k)} \cdot \prod_{k=1}^{r} \theta_k^{\alpha_k + N_k - 1} \quad (4)$$
$$= Dir(\theta \mid \alpha_1 + N_1, \cdots, \alpha_r + N_r)$$

With Bayesian method to learn Bayesian network parameters, we use Matlab to obtain the results for the conditional probability of landslide stability (CPT) in the following Table Ⅱ:

TABLE II. THE CONDITIONAL PROBABILITY OF CLASS NODE (THE STATE OF LANDSLIDE STABILITY)

B(slope angle)	E(cohesion)	P(Class=1|B,D)	P(Class=2|B,D)
1	1	0.9808	0.019
1	2	0.5	0.5
1	3	0.5	0.5
1	4	0.5	0.5
1	5	0.5	0.5
2	1	0.9808	0.0192
…	…	…	…
5	1	0.9934	0.0066
5	2	0.5	0.5
5	3	0.5	0.5
5	4	0.0050	0.9950
5	5	0.0033	0.9967

D. Reason and predict landslide stability

The premise of using Bayesian network inference is to construct Bayesian network model from the original data. Actually, it is mining data from the original data. First, find out the definite network structure that corresponds most to the original data, and then calculate the conditional probability of the nodes according to the causation of the network structure. The inference principle of Bayesian network must base on Bayesian theorem:

$$p(B_s \mid D) = \frac{p(D, B_s)}{p(D)} = \frac{p(B_s) p(D \mid B_s)}{p(D)}$$

After getting the Bayesian network, we can use it to predict the landslide stability. This is task of Bayesian network causal inference, which is calculating the probability distribution of the rest of the node in the network with the value of the known collection of nodes. It is also calculating the posterior probability distribution.

For the general Bayesian network, in the worst case, all method is exponential, and the complexity of reasoning mainly depends on the network structure. Therefore the design of efficient reasoning methods has been the hot topic of scholars. The reasoning methods can be roughly divided into accurate reasoning and approximate reasoning. There are three main kinds of inference forms for Bayesian network: causal reasoning, diagnostic reasoning, and support reasoning.

In this paper, the task of causal reasoning of Bayesian network is to predict the landslide stability. Currently, Joint Tree algorithm proposed by Jensen is commonly used, which can not only resolve the reasoning under simply connected networks, but also accomplish reasoning under multiply connected networks. This algorithm exhibits outstanding convenience especially when reasoning is conducted in a network where multiple interrogating nodes exist.

The main idea of joint tree inference algorithm is to convert the Bayesian network into a tree, and then calculate the probability by defining the process of message passing on joint tree.

This experiment uses the joint tree inference engine of BNT toolkit to perform precise reasoning for the 10 instances of landslides in literature [9]. A random sample *X* = *(A (slope height) = 1, B (slope angle) = 2, C (bulk density) = 3, D (internal friction angle) = 5, E (cohesion) = 5)*, has the result *P(X|Class=1)*= 0.3008，*P(X|Class=2)*= 0.6992. Therefore the sample identified slide more likely.

IV. Analysis

In order to analyze the predictive accuracy, the experiment uses 5-fold Cross Validation to test the classified results：The initial data sets are randomly divided into five disjoint subsets $S_1, S_2, \cdots, S_5$, equally to the formula $S_1 \cap S_2 \cap \cdots \cap S_5 = \varnothing$, and the size of each subset is basically the same. Learning and testing are carried out five times. Taking the i-th iteration S_i is used as test set, and the rest sub-sets are used to train the classifier. The final outcome of the assessment is the average accuracy rate of taking 5 correct classifications to divide the total number of samples in the initial data.

To compare with other forecasting methods, experiment uses the same number of samples and variables, and uses support vector machine (SVM) method, extraction of which 1/5 of the data to predict, and the rest of the data used as training data on Matlab and Libsvm. After five iterations, the comparison results shown in Table III.

TABLE III. Results of individual indicators

Method	TP	TN	FP	FN	Detection rate (%)	False detection rate (%)	Total accuracy rate (%)
Bayesian network	26	16	9	5	83.87	36	75
SVM	27	4	19	6	81.82	82.61	55.36

TP (True Position): The number of correct and affirmative, the number of destroyed landslide forecasted as destruction landslide;

TN (True Negatives): The number of correct and negative, the number of stable landslide forecasted as stable landslide;

FP (False Positives): The number of false and correct, the number of stable landslide forecasted as destruction landslide;

FN (False Negatives): The number of false and negative, the number of destroyed landslide forecasted as stable landslide;

Detection rate: TP/ (TP+FN);

False detection rate: FP/ (FP+TN);

Overall accuracy rate: (TP+TN) / (TP+FN+FP+TN).

From the indicators and the training sample, Bayesian network inference has higher credible results than the SVM. The result of Bayesian network is better and higher reliability for predicting landslide stability than SVM. Due to the small sample size, the detection rate reached 83.87%, the overall accuracy rate of 75%, the results relative to the ideal. It can be concluded from this, using Bayesian network to predict the stability of the landslide has a good credibility.

V. Conclusions

Based on Bayesian network, this paper presents a landslide stability prediction model and achieves a state warning for an actual landslide. Through the prediction of landslide probability, it can judge the landslide stability of the landslide. Results show that using Bayesian network model for the prediction of landslide has a higher reliability and a better credibility than the SVM. The detection rate of this model reaches 83.87%, and the overall accuracy rate reaches 75%. It can be concluded that using Bayesian network to predict the stability of the landslide has a good credibility. In future it should take more factors into consideration, such that the last landslide may impact the next landslide, the groundwater may affect the bulk density, and the rainfall may affect groundwater.

Acknowledgment

This paper is supported by the National Natural Science Foundation of China (Grant No. 41172298).

References

[1] Li Jing, “The thinking and summarizing of Landslide stability research quo, ” West-China Exploration Engineering, 2006 Section 8.

[2] Luan Yongping, Gao Yongqiang, “Landslide Stability Analysis Method., ” West-China Exploration Engineering, 2007,pp.35-37.

[3] Shi WeiMin, Zheng Ying-ren,Tang Boming, “Landslide Stability Evaluation Methods, ” Rock and Soil Mechanics, 2008, 24 (4), pp. 545-548.

[4] Xia Yuanyou,Li Mei, “Slope Stability Evaluation Method Research and Development Trends,” Rock Mechanics and Engineering, 2002,21(7), pp.1087-1091.

[5] Yang Xuetang,Wang Fei, “ Slope Stability Evaluation Method Research and Development Trends,” Geotechnical Engineering Technique, 2004, 18 (2),pp.103-106.

[6] Luo Wenqiang, “Slope stability probability theory and methods,” Rock Mechanics and Engineering,1999, 18 (2), pp. 240 - 240.

[7] Tang Yichuan , He Yucheng, “Overview on evaluation methods of slope stability analysis,” Shanxin Architecture, 2010.12(36),pp.66-68

[8] Wang Guoti, “Slope stability and landslide analysis method for stress [M] ,” Beijing : Science and Technology Press, 2012.

[9] Huang Guangdong, “Landslide Modeling and Analysis Based on Intelligent Algorithm[D] ,” Beijing. China University of Geosciences, 2011.

[10] Cui Zhengquan,Li Ning, “ Slope Engineering -a latest development in theory and practice[M] ,” Beijing: China Water Power Press, 1999.

[11] Xiao Qinkun,Gao Song, “Bayesian Networks in Intelligent Information Processing[M] ,”Beijing: National Defense Industry Press, 2012,1

[12] He Wei,Pan Quan,Zhang Hongcai, “The Development and Prospect of Bayesian network structure learning,” Information and Control,2004,33 (2),pp.185-190.

[13] Zhang Bingli,Pei Yahui, “Bayesian Network Model Overview ,”Computer & Information Technology,2008,16 (5) ,pp.41-42.

[14] Jensen FV. An Introduction to Bayesian Networks [M]. New York: Springer , 1996.

[15] Huang Youping, “Bayesian Network Research[D] ,” Beijing: Institute of Computing Sciences, 2005.

[16] Xia Yuanyou, Zhu Ruigeng, “Slope Stability Analysis Expert System Development ,” Catastrophology, 1997.

[17] NK Sachertorte, P. R. Shanley, “Slope stability maximum likelihood estimation ,”Foreign Metal Mining Magazine,1995,(3),pp.6-12