Journal of Chemical and Pharmaceutical Research, 2014, 6(6):1315-1322



Research Article

ISSN: 0975-7384 CODEN(USA): JCPRC5

Study on displacement prediction of landslide based on neural network

Jian Huang, Zhihuan Liu and Ni Li

Hu Bei University of Arts and Science, P. R. China

ABSTRACT

In recent years, the economic losses caused by landslides were as high as several billion dollars. Therefore, the study of landslide risk has become a hot topic today. Landslide displacement prediction is a highly nonlinear and extremely complex issue. In most cases, it is difficult to use mathematical models to describe the process clearly. Data mining technology, which uncovers the hidden data patterns and models effectively, can be used for landslide displacement prediction. Series of system models are trained, validated, and applied to a landslide study along the Three Gorges and cases from the literature. Based on these monitoring data from deformation displacement of the Bazimen landslide in Zigui County, natural rainfall in the area of Shazhen and Guizhou, water level changes of the Three Gorges Dam, we propose a theoretical method suitable for analyzing the impact of landslide deformation. By drawing duration curve about each monitoring point displacement and predisposing factors on the valid data, we use regression analysis method to analyze correlation and hysteresis of landslide displacement variation and predisposing factors through the actual situation of monitoring monthly reports.

Key words: neural networks; spatial data mining; geological disasters; Regression analysis

INTRODUCTION

There are many unexpected geological hazards in recent years such as earthquakes, avalanches, landslides, mudslides, etc. Risk assessment of geological hazard has become one of the major issues of general interest [1]. The landslides as a kind of geological disasters, plain and simple, bring varies of serious threats to humans. Therefore the forecasting of landslide displacement is very meaningful.

Spatial data mining is a relatively new field with integration of multiple disciplines and a variety of related technologies, including artificial intelligence, database, pattern recognition techniques and so on [2-6]. Therefore, spatial data mining methods contain not only general data mining methods, as well as spatial database. There are already many research institutes and researchers into the field of space research in the data mining. Hao et al.[7] broken the landslide displacement into cycle terms and trend terms, and combined with the periodicity characteristics of time series to analyze cycle items of landslide displacement. Dujuan et al.[8] used BP neural network to predict its displacement based on the work of Hao. Li Qiang and Li Duan [9] proposed a time series analysis with capabilities of the forecasting complex systems in development trend, and adopted timing analysis method to establish the ARIMA model and the CAR model for landslide displacement dynamic fore-casts. E. Yesilnacar et al.[10]combined with logistic regression and neural network to overcome shortcomings of the statistical methods that could not effectively build models of complex geological disasters, deal with non-linear relationship between landslide hazard and predisposing factors. Xiangenjun [11] used rough set to dig out the inherent law of slope disaster activities from the historical slope data. Daifuchu [12] focused on the natural landslide spatial prediction in Hong Kong, adopted two or single type to support vector machine for spatial prediction of landslide hazard, and compared with Logistic regression models at the same time.

In addition, data mining technology has great advantages in terms of analyzing massive amounts of data. There is a research trend for a comprehensive prediction criterion of data mining and building forecasting model about how to choose the right method to introduce right theory to the study of geological disaster forecast. With the diversity of data mining and knowledge discovery methods, researches in the field of geological disaster prediction more use varies of data mining methods, compared with the conclusions in the same research area [13]. Due to early warning command, forecast necessities and integrated decision-making, this paper establishes a mining model suitable for landslide hazard data, studies spatial data mining about analysis of prediction of landslide hazard and landslide displacement, supporting geological disaster monitoring and early warning and emergency command. To be adopted, such as linear regression, neural networks and other data mining techniques, we process data automatically analyzed and summarized to extract critical data that assists prevention of geological disasters decisions, conduct a structural analysis model. With monitoring reports to analyze the actual experimental results, we verify the reliability and accuracy of the mining model applications. At last, according with final charts, visualization, and other forms of natural language, we establish a suitable data mining model for landslide hazard prediction, and present the knowledge discovery in an easy way for users to understand.

THE NATURE AND CHARACTERISTICS OF THE LANDSLIDE DATA

The data used for experiment is from monitoring data of landslide in area of Zigui county of the Three Gorges, where belongs to the subtropical monsoon climate, abundant rainfall, and more rainstorm, and also owns complicated geological features along the Yangtze River. Therefor, it has occurred geological disasters sometimes, such as collapses, landslides and debris flows. Bazimen landslide is active more frequently in recent years, so this paper chooses this area of landslide to analyze.

1. The property values of the landslide monitoring data have continuity, which is suitable for regression algorithm and fitting algorithms to build the model. For as-sociation rules algorithm in mining algorithms, we must discretize the landslide data in advance.

2. The impact of rainfall on landslide displacements is cumulative and time-delay. The generation of landslide displacement may be affected by heavy rainfall in a short time, or due to the cumulative rainfall over a period of time.

3. Changes in the water level of the Three Gorges dam are affected by natural rainfall and the impact of man-made water drainage activities. So, it must exclude the impact of human activity factors when researching the impact of the landslide displacement on dam water level.

4. Randomness, instability, and deformation characteristics of each formation mechanism of landslides are not all the same, so each landslide must be a detailed analysis based on the actual situation.

5. According to collation and analysis of the collected data about each landslide in Zigui county, as well as meteorological data and hydrological data, we summarize the following characteristics of the data.

THE DATA MINING APPLICATIONS OF BAZIMEN LANDSLIDE

Data Mining, known as knowledge discovery in databases (Knowledge Discovery from Database), deals with the observed data set which is usually large for analysis. Of course, the purpose is to find unknown information, summarize and present the data in the way that data owner can understand. This paper focuses on the contents of the application of the raw data for data mining on geological disaster monitoring. Appling the current data mining technology to extract useful information from the historical data, which is beneficial for geological disaster monitoring and disaster prediction, that can provide services for warning command system.

DATA PREPROCESSING

Existing data is dam water level daily monitoring data, including two monitoring points GPS measurement data of Bazimen landslide ZG110, ZG111. GPS measurement data is a monthly data which is different from the rainfall and the water level of the dam monitoring data in time granularity. Therefore, we need preprocessing and re-sampled monthly data, and select month time by the time intervals of GPS measurements, such as dam water level averages in March 2004 refers that water level measurements between February 2004 to March in GPS monitoring day, namely on February 11, 2004 to March 14. Rainfall data takes the cumulative value of the month. Dam water level fluctuation rate is used to measure the water level fluctuations. A positive value indicates the water level rises, a negative value indicates the water level drops. Landslide displacement rate is the monthly landslide deformation, which takes the average value of each observation point.

After we organize the raw data table, in order to prepare for the establishment of the model, the data is normalized to

eliminate the influence of dimension. Here the data is normalized to prepare for the later work with the neural network too. There is an approach we used that all the data is converted into [0,1] or [0.1,0.9] or [-1,1] between the number to cancel an order of magnitude difference between the dimensional data, and avoid a large difference in magnitude or larger prediction error fit when input and output data. In this paper, the experimental data with a time effect, that each record must follow chronologically. When selecting a training data set and test data sets, in accordance with the time sequence, selection of training data before eighty, twelve after the predicted data. Satisfying the above conditions assumed, so it will be its normalization.

LANDSLIDE DISPLACEMENT FACTOR

1) Displacement rate and dam water level and water level fluctuation rate

The average water level of the Three Gorges dam area changes in seasonal fluctuations. This is the main flood season from May to September every year, when happens to be a year of rainfall. To meet the arrival of flood season, the Three Gorges reservoir releases, the dam average water level drops. Every year from October to April is the second year storage period with less rainfall. Because of the need for power generation, the reservoir storages water, dam water level rises or remains on average at a certain water level. Through the relevant data analysis from two monitoring points of the Bazimen landslide, the trend of the cumulative displacement was rising at the beginning of storage water, but the rate of increase is growing slowly. As the water level continue rising, the displacement rate gets slower. When the water level rises to the highest point, the displacement rate will get faster. The water level is the highest point of a turning point. Before and after the water level the highest point, the water level on the rise and fall just two states, so the decrease in the water level rises or landslide displacement inevitably affects the rate of change.

2) The impact on the water level of the dam

The dam water level has an effect on seasonal variation, thus the monthly water level could effect by before. Here we use time series of ARMA model to analyze its impact by its own laws. Through repeatedly calculating, the average water level of the dam lag factor is 2. That is, the dam water level is affected by their first two phases data.

The average dam water level satisfies AR (2), the equation is below, as seen in Eq. 1.

$$w_t = 1.40859 w_{t-1} - 0.50807 w_{t-2} + 150.66445 + \mathcal{E}_t \tag{1}$$

In this model, R2 is 0.895, the fit of the model is very high. AIC = 259.768, SC = 267.398.

3) The rate of rainfall and displacement

The rainfall of the Bazimen landslide nearby has the seasonal variation, the landslides and displacement rate also vary with rainfall cyclical variation. Annual rate of landslide displacements has occurred to accelerate the increase of rainfall during the year. Annual maximum displacement rate occurred in the month before and after the biggest annual rainfall. It indicates that there is certain relevance between rainfall and landslide displacement rate, but the rainfall has a certain lag effect on landslide deformation.

4) Its own influence rainfall in the area of Huizhou

The rainfall in Guizhou shows seasonal variation, the monthly rainfall may be affected by prior rains. Here we use time series of ARMA model to analyze its impact by its own laws. Through repeatedly calculating, the average water level of the dam lag factor is 3. In the model, i = 1, it refers that making a first-order differential conversion on the rainfall data, then getting monthly rainfall data. In the model of IAR (3,1), the final state in the area has an affection on their monthly rainfall data of the first three. Based on above, the subsequent modeling process will consider the impact of the final state in the current month and previous rainfall areas of landslide displacements. The cumulative rainfall R1 meets the IAR (3,1), the equation is below, as seen in Eq. 2.

$$(1-B)R_1 = \frac{\mathcal{E}_t}{(1-0.22915B+0.04964B^2+0.0648B^3)} + 57.1369$$
(2)

R2=0.999, AIC=735.34421, SC=745.51739

LINEAR REGRESSION

In this section, the related interpretation includes the average water level w, dam water level fluctuation velocity dw, the landslide of Bazimen cumulative displacement x1, the landslide of Bazimen average displacement rate v1, cumulative rainfall of Guizhou R1, monthly rainfall of Guizhou R1 and their pre-variables such as dwt-2 indicates

the water level fluctuation speed ahead two phase and so on. Each phase of dam water level of fluctuation rate equals to the average level of the dam by subtracting the average water level on the phase of the dam, so when the dam water level fluctuation rate is positive, said the water level rises, otherwise the water level dropped, the absolute value of the speed of dam water level fluctuation in the period corresponds to the rate of rise or fall.

By stepwise regression method, the average water dam, the rainfall of the month and the former three month, water level fluctuation velocity and the month before the three fit following model, the equation is below, as seen in Eq. 3.

$$x1 = 0.858w - 0.438dw_{t-1} - 0.567dw_{t-2} - 0.823dw_{t-3} + 0.128$$
(3)

This model is R2 = 0.76, each coefficient of the above equation is detected by P.

Here Bazimen landslide cumulative average displacement is effected by the dam water level, the combined effects of dam water level fluctuation velocity. The rainfall of Guizhou is not considered among the state equation. From the model perspective, the landslide dam water level fluctuations in the rate of displacement with a negative correlation, and is subject to water level fluctuations of the first three. When the water level drops, and the drop amount increases, the landslide displacement will increase.

Here is a comprehensive evaluation of the results of this regression model, as seen in Fig. 1. Each evaluation of the model is seen in Tab. 1.



Fig. 1: Landslide Prediction Curve of Bazimen

Tab. 1: Model Evaluation Value

Minimum error	-0.27
Maximum error	0.22
Average error	0.0001
Average absolute error	0.076
Standard deviation	0.181

Overall fitting accuracy of the linear regression model is 47.25%. This linear regression model for landslide deformation phase shift, fits effect in gentle change part, namely water displacement of the landslide deformation is better fitting with accuracy of 51%, marked in Fig. 1. For the displacement of a large part of the volatility, it fits landslide displacement flood ineffective. In summary, the linear regression model for landslide deformation displacement of water is appropriate.

NEURAL NETWORK METHOD BUILD NEURAL NETWORK MODEL

BP neural network model to predict landslide Bazimen is divided into four steps: The first step is to determine the structure of the neural network; The second step views observational data of landslide displacement as training samples and then normalized; Third step is nerve network training; Fourth step makes use of the trained neural network to predict landslide displacement, and normalizes the input data. Firstly, data filtering mainly used to filter out some of the source data table unused variables, and then samples. Bazimen landslide data from January in 2004 to November in 2011 has a total of 92. We chose the former 80 as training data, the later 12 as the prediction data. First feature selection before modeling, according to the characteristics of the input variables with predictors variables of the Pearson correlation coefficient to select. Pearson coefficient values in [0.9,0.95] for the general importance, [0.95,1] is important. Thus select the input variables of the neural network. Then enter the first 80 data model for training. Training network is three layer network including input layer, a hidden layer and output layer. Five neurons in the input layer, three neurons in the hidden layer, one neuron in the output layer. The accuracy of

training network is up to 92.66%. After establishing a good neural network model, the data input the later twelve forecasting data to predict for Bazimen landslide.

There is a sudden increase significantly faster than the month before and after the month in June 2011. Except for the predictive value of the point, the model for flood displacement rate is generally better than the value of water displacement. It shows that the extent of the training network to fit the observed data is not very good, so the overall effect is predicted to be improved.

Landslide is a continuous process, the shape of the variable itself is bound to be affected by pre-deformation, which is the time effect of the landslide. The above model does not consider the effects of landslides own time. So here's an attempt to introduce trend term of landslide displacement of the average monthly, and error term of time series model to correct the above model.

Time	Average monthly rate of displacement	BP neural network predictive value
2010-12	0.111154	0.143243
2011-01	0.104648	0.139373
2011-02	0.168784	0.146493
2011-03	0.173122	0.148692
2011-04	0.191712	0.1563
2011-05	0.186135	0.191506
2011-06	0.9	0.280842
2011-07	0.263904	0.280038
2011-08	0.103718	0.197304
2011-09	0.220837	0.202106
2011-10	0.149574	0.166727
2011-11	0.148025	0.140472

Table 2. The Predicted Result Average Monthly Rate of Displacement of Bazimen Landslide

Table 3. Estimation Results

-0.137
0.619
-0.002
0.056
0.106

A lot of the time series of observed sample will show the trend, seasonal and random, or just showing three of the second or one. In this way, you can think of each time series, or after an appropriate time sequence function transformation, it can be broken down into three parts superimposed.

$$X_{t} = T_{t} + S_{t} + R_{t} \quad t = 1, 2, \cdots$$
(4)

Which $\{T_{i}\}$ is a trend item, $\{S_{i}\}$ is a seasonal item, $\{R_{i}\}$ is a random item. This time sequence is the superposition of three.

The average monthly Bazimen landslide displacement decomposed into trend term and seasonal items, and items will be added to the above trend BP neural network model for discussion. There are many time series decomposition methods, such as segmented trend method, the trend regression line method, quadratic trend, stepwise average method, regression, etc., The gradually average method is used in this paper. With the gradual trend term average twice stripped out, the formula is below, the results are shown in Fig. 2.

$$U_{n} = \frac{1}{6} \sum_{i=n-6}^{n} y_{i} \qquad T_{n} = \frac{1}{6} \sum_{i=n-6}^{n} U_{i}$$
(5,6)

The equation of Bazimen landslide average displacement suitable for ARIMA is in Eq. 7. ARIMA(1,1,1):

$$(1 - B)x = \frac{(1 - 0.575B)}{(1 - 0.258B)}\varepsilon_t + 12.511$$
R2=0.996, AIC=567.7, SC=575.33
(7)



Fig 2: Bazimen Landslide Trend Term Average Monthly rate of Displacement



Fig 3: Bazimen Average Displacement Curve Fitting Landslide

The time series equations of the landslide displacements of Bazimen can get the error term.

The error term and trend item were added to the BP network by order, we get BP network with trend items, BP network with ARIMA and BP network error term with trend term and the error term. The four experimental results compared within four neural networks are shown in Table 4.

	BP	TBP	EBP	TEBP
Training Network accuracy	92.66%	94.36%	97.03%	96.57%
Average error	-0.002	0.001	-0.002	0.001
Standard deviation	0.106	0.098	0.039	0.036
The overall prediction accuracy	27.2%	24%	34.8%	38%
Flood prediction accuracy	25.71%	20%	37.14%	42.86%

Tab 4: Four	Kinds of Neura	al Networks	Comparison	Results

From Fig. 4 to Fig. 7 and Tab. 4, the prediction of BP network with trend items and errors items on Bazimen landslide displacement is more accurate. Especially, the position of extreme points in Fig. 4, the normal BP network for this extreme point generally poorly fit. After joining the trend term and the error term, the overall fitting accuracy improves in BP network a little, the effect is not very obvious improvement. But from the following 4 compares the picture, you can clearly see that the prediction accuracy of landslide deformation displacement of the peak point of improves a lot. For example, the May in 2011, the predictive value of flood displacement is 0.194, the observed value is 0.186, the error is only by 4%.

The BP network with trend terms and terms, in the training process, the circumstance of fitting in the flood dis-placement points is more typical than BP network, which is also significantly improved. For example, the error of the predicted value of BP network in June 2010 up to 47.47%. After introducing trend backward, this error is reduced to 40%. After the introduction of the error term, the error at that point has improved reduces to 9.6%. After the introduction of the error is reduced to 5%.



Fig 4: Bazimen Landslide Average Monthly Rate of Displacement BP Network Prediction



Fig 5: BP Network Bazimen Landslide Displacement Rate with the Average Monthly Trend Term Predictions



Fig 6: Bazimen Landslide Average Monthly Rate of Displacement Error Term BP Network with ARIMA Forecasting Results



Fig 7: BP Network Bazimen Landslide with Average Monthly Rate of Displacement and the Error Term Trend Term Predictions

BP network model with a trend term and the error term for landslide deformation displacement stage, drastic change, fits effect, namely flood landslide deformation displacement fit better with accuracy of 42.86%, as shown in Figure 7 the marked section. For example, in May 2004, May and July and August in 2005, in June and August and September in 2006, in July and August in 2007, in September 2008, in June and August in 2009, in June and July in 2010, in May 2011, the prediction error is less than 10%. For the displacement of the flat part of the change, it is

fitting badly for water storage. In summary, BP network model with trend items and errors items on the flood landslide deformation displacement is more appropriate.

CONCLUSION

The landslide deformation displacement effected by rock-soil, the natural rainfall and reservoir water level fluctuations and other factors. In the field of landslide of landslide displacement, more researches are to estimate the stability of the landslide and the prediction of landslide deformation using univariate time series. In this paper, we takes into account the impact of water level fluctuations, and other factors on natural rainfall landslide deformation through a comprehensive analysis, which has better reflects the important role of external incentives for the development of landslide deformation.

By using multivariate regression model, the fitting curve of Bazimen landslide displacement is better in the phase of water storage, which is closer to the actual trend of displacement. The fitting results of the neural network are just its complementary. Application of BP network in the prediction of Bazimen landslide displacement, with high accuracy of its training network, but the prediction accuracy is not ideal. In order to achieve the purpose of short-term forecasts, the accuracy of the model remains to be further improved.

Future Works

In this paper, although the characteristics of Bazimen landslide data to establish a proper analysis and forecasting model, but there are many issues to be solved when suitable for all landslide further. Some observations due to the landslide displacement data point for the newly added point, data monitoring less, are not fully utilized. Therefore data processing could be improved. Since the particle size is the monthly data, the accuracy is not enough. If the landslide displacement of GPS can be accurate to days, the model experiments of landslide deformation will be more practical significance, which will play a greater role in disaster preparedness. Currently, we only experiment on the Bazimen landslide data, the follow-up model can be extended to other landslides in the other areas.

Acknowledgments

This work was financially supported by Science and Technology Projects Hubei Provincial Department of Education (Project ID: Q200625001).

REFERENCES

[1] Huang, R.Q., Xiang, X.Q., Ju, X.X, et al., Geological bulletin of china, 23(11):1078-1082, 2004.

[2] Martin E, HansPeter K, Jorg S., Spatial Data Mining: A *Lecture Notes in Computer Science*, Springer, 1997.
[3] Tang,L.B., Research on Progress of Spatial Data Mining [D]. *Graduate School of Chinese Academy of Sciences*, 2005.

[4] Xu,F.,Wang,Y.,Du,Y.,et al., Chinese Journal of Rock Mechanics and Engineering, 30(4):46-751, 2011.

[5] Wang,L., Geomatics & Spatial Information Technology, 36(6): 121-123, 2013.

[6] He,B.B.,Cui,Y.,Chen,C.H.,et al, Advances in Earch Science,26(6), 2011.

[7] Hao, X.Y., Hao, X.H, Xiong, H.M., et al, Journal of Engineering Geology, 7(3):279-283, 1999.

[8] Du,J.,Yin,K.L.,Chai,B., Chinese Journal of Rock Mechanics and Engineering,(09): 1783-1789, 2009.

[9] Li,Q.,Li,R.Y., Journal of Yangtze River Scientific Research Institute, 22(6), 2005.

[10] E.Yesilnacar, T.Topal. Engineering Geology, 79:251-266, 2005.

[11] Wang, G.Y., Cui, H.L., Li, Q., Rock and Soil Mechanics, 30(8): 2418-2422, 2009.

[12] Lin, D.C., An, F.P., Guo, Z.L., et al., Rock and Soil Mechanics, 32(1), 2011.

[13] Jose Z, Jesus P, Juan T. A UML Information and Software Technolog, 51:977-992, 2009.