Prediction of Collapse for Tunnel Portal Slopes using General Linear Model

Yong Baek,[1] Jong-Hwa Na,[2] O-Il Kwon,[3] Yong-Seok Seo[4] and Ja-Hae Jung[5]
1) Geotechnical Engineering Research Department., Korea Institute of Construction Technology, 2311 Daehwa-Dong, Ilsan-Gu, Goyang-Si, Gyeonggi-Do 411-712, Republic of Korea (baek44@kict.re.kr)
2) Department of Information and Statistics, Chungbuk University, 12 Gaeshin-Dong, Heungduk-Gu, Cheongju-Si, Chungbuk, 361-763 Republic of Korea (cherin@cbu.ac.kr)
3) Geotechnical Engineering Research Department., Korea Institute of Construction Technology, 2311 Daehwa-Dong, Ilsan-Gu, Goyang-Si, Gyeonggi-Do 411-712, Republic of Korea (kwonoil@kict.re.kr)
4) Department of Earth and Environmental Sciences, Chungbuk University, 12 Gaeshin-Dong, Heungduk-Gu, Cheongju-Si, Chungbuk, 361-763 Republic of Korea (ysseo@cbu.ac.kr)
5) Geotechnical Engineering Research Department., Korea Institute of Construction Technology, 2311 Daehwa-Dong, Ilsan-Gu, Goyang-Si, Gyeonggi-Do 411-712, Republic of Korea (jhjung@kict.re.kr)

Abstract

This study was intended to develop a high reliable technique by statistically processing on-site data with a general linear model, providing the basic data for construction, analysis of stability and establishment of maintenance measures for portal slopes in the future. This study evaluated the stability of a tunnel's portal slope using a quantified technique, which is based on a general linear model. The important scores of each independent variable were allocated by using the ranges of the quantified values, based on the predicted coefficient of regression and the scores for categories of each independent variable were allocated so that those are equally spaced. The quantification model obtained from the results of evaluating the total data used for the quantification process provided precise results. In addition, it is expected that a more detail subdivision of response variables and sufficient data would produce a better stability evaluation standard.

Keywords: prediction, tunnel portal slope, general linear model, evaluation

Introduction

Tunnels connecting roads in the country were essential considering the fact that mountains account for 70% and more of the territory. However, collapses of tunnel portal slopes and exit/entrance slopes have caused a huge loss nationally, which has wasted the required budget on repairs. The review of stability in constructing tunnels, therefore, is becoming more important and has been studied intensively (Yoo, C. S., 2004). On the other hand, a study about tunnel portal slopes needed to be studied further. General slopes have been managed by the slope maintenance system or the routine investigation system (KISTEC, 2003) on each of expressway, national highway and railroad.

This study was intended to develop a high reliable technique by statistically processing on-site data with a general linear model, providing the basic data for construction, analysis of stability and establishment of maintenance measures for portal slopes in the future.

Concept of General Linear Model

A collapse forecasting quantification technique using statistical processes, as a tunnel portal slope maintenance method, was utilized to evaluate the status. The general linear model for data in which a response variable is continuous and independent variables are continuous or categorical is executed from the following sequences. Assuming that the actual values are $y = (y_1, y_2 \cdots y_n)'$, the following linear model is probable.

$$y = X\beta + e$$

The predicted value, based on the least square method for the above linear model, of coefficient of regression is given as below.

$$\beta = (X'X)^{-1}X'y$$

pp. 523–527

From a view of quantification, the prediction for the efficient of regression can be interpreted as a process to get β a maximum value of the following correlation coefficient.

$$corr(y, X\beta) \tag{1}$$

If clarifying it when the linear model includes the intercept term, it can be expressed in the following way

$$\begin{aligned} X &= (1, X_1, \cdots, X_p) : n \times p \text{ Matrix} \\ y &= (y_1, \cdots y_n)' : n \times 1 \text{ vector} \\ \beta &= (\beta_0, \beta_1, \cdots, \beta_p)' = (\beta_0, \beta_1') : p \times 1 \text{ Vector} \end{aligned}$$

and it could be assumed for convenience's sake that X and y are centered matrix and vector. At the point, the correlation coefficient in the formula (1) is,

$$corr(y, X\beta) = \frac{y'X\beta_1}{[y'y]^{1/2}[(X\beta_1)'(X\beta_1)]^{1/2}}$$

and β, the max value of it can be expressed as the following objective function and constraints.

$$\max_\beta y'X\beta_1 \text{ subject to } \beta_1'(X'X)\beta_1 = \text{constant}$$

The solution in the above formula is a value in which the following Lagrange function is maximized.

$$L(\beta_1, : \lambda) = y'X\beta_1 - \lambda\beta_1'(X'X)\beta_1 - \beta_1'(\widetilde{X}, \widetilde{X})\beta_1 = \lambda\text{Lagrange multiplier}$$

Differenciating a function $L(\cdot)$ w.r.t. β_1 and equate 0, we obtain the following;

$$\partial L/\partial\beta = X'y - 2\lambda(X'X)\beta_1 = 0$$

If we take $\lambda = 1/2$, the predicted value of β_1 is as follows.

$$\widetilde{\beta_1} = (X'X)^1 X'y$$

In this case, if a constant term, β_0 is equal to $\bar{y}$ the predicted β is;

$$\widetilde{\beta} = (X'X)^1 X'y$$

The above linear model assumes that independent variables are continuous, but even though they are mixed with categorical variables, it can be similarly applied by introducing dummy variables. That is, the procedure to express a categorical variable to dummy variable is as follows. In general, a categorical variable, ν with as many as k categories can be expressed as $w_1, \cdots, w_{k-1}$, the dummy variable of $(k-1)$.

$$\begin{array}{ccc} \nu = 1 & \Leftrightarrow & w_1 \text{ and } w_2 = \cdots = w_{k-1} = 0, \\ & \vdots & \\ \nu = k-1 & \Leftrightarrow & w_{k-1} = 1 \text{ and } w_1 = \cdots = w_{k-2} = 0, \\ \nu = k & \Leftrightarrow & w_1 = w_2 = \cdots = w_{k-2} = w_{k-1} = 0 \end{array}$$

Important variable quantification techniques

In this study, for independent variables affecting the stability of tunnel portal slopes, mixed variables including 7 categorical variables and 2 continuous variables were used. For the response variable, continuous values expressing the probability of collapse by the researchers are used. That is, $y = 1$ (stable) represents the first 15% that collapse is rarely probable, $y = 5$ (monitoring) does the first 30–15% that collapse is so probable and $y = 6$ (measure, unstable) does the first 15% that collapse is also very probable. The detail descriptions about these variables are summarized in Table 1.

The goodness-of-fit test of the model and the predicted results for the coefficients of regression are presented in Table 2. In the goodness-of-fit inspection, it is reasonable that this model is significant at 5% significance level as F-value was 10.88 and significant probability was $< .0001 **$. The correlation coefficient in the formula (1), given by the square root of the determination coefficient R^2, is as high as $\sqrt{0.7462} = 0.8638$ and the estimated standard deviation is 1.1904, the Root MSE in Table 2.

Table 1. Important variables in the stability evaluation of tunnel's portal slopes

Category	Categorical variables							Continuous variables		Response variable
	Weather	Vegetation	Convex	Seepage	Measure	Capability of collapse	Previous collapse	Slope hight	Slope angle	y : Status of portal slopes
1	S.W	Good	Even	Dry	Good	None	None	5~ 18	25~ 47	Stable (1)
2	M.W	Fair	Normal	Damp	Fair	Low	Rockfall	19~ 31	48~ 59	Monitoring (5)
3	H.W	Poor	Convex	Wet	Poor	Moderate	Surface failure	32~	60~	measure, unstable(6)
4	C.W	-	-	Flow	-	High	Collapse	-	-	-

Table 2. Goodness of fit test of the model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	308.2978	15.4149	10.88	<.0001**
Error	74	104.8600	1.4170		
Corrected Total	94	413.1579			
R-Square	Coeff Var	Root MSE	Y Mean		
0.7462	33.2 609	1.1904	3.5790		

Table 3. Range of quantified values of continuous variables

Variable	Category	Section	No. of Data	Median	Calculation of the range of quantified value
Slope height	1	5~18	32	14.85	R(high) = 0.022*(45.5-14.85) =0.6743
	2	19~31	31	24.18	
	3	32~	32	45.5	
Slope angle	1	25~47	32	41	R(angle) = -0.008*(63-41) = -0.176
	2	48~59	31	55	
	3	60~	32	63	

The range of quantified values for categorical variables could be obtained from (max. value - min. value) of the predicted values by dummy variables. In addition, the range of the quantified values for continuous variables could be obtained from $R(X) =$ (coefficient of regression estimated from a variable$X$)* (sectionalized range of a variable X). At the moment, a 'sectionalized range' can be defined as a difference between the median in the max section and the median in the min section. The reason why a continuous variable is divided into several sections is because it is necessary to consider fairly with categorical variables. In the study, the slope height for a continuous variable and the angle were divided into 3 sections. That is, the data totals 95, so 32 was allocated to the first section, 31 for the middle and 32 for the last, to realized the balanced categorization. The ranges of quantified values are calculated as in Table 3.

Scoring collapse probability and evaluating this model

The above ranges of quantified values tell the importance of each independent variable. In Table 4, the importance of each variable is allocated in proportion to the range(total sum: 100 points) and the scores in a category are allocated to maintain regular intervals. Therefore, the stability for a tunnel's portal slope

Table 4. Collapse probable scores

Variable		Slope height	Slope angle	Weather	Vegetation	Convex	Seepage	Measure	Capability of collapse	Previous collapse	Total
Range		0.674	0.176	1.230	1.029	0.300	4.089	0.674	3.931	0.599	-
Importance score of each variable		5	2	10	8	2	32	5	31	5	100
Contribution score for categories	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-
	2	2.50	1.00	3.33	4.00	1.00	10.67	2.50	10.33	1.67	-
	3	5.00	2.00	6.67	8.00	2.00	21.33	5.00	20.67	3.33	-
	4	-	-	10.00	-	-	32.00	-	31.00	5.00	-

Table 5. Evaluation of the model

Collapse probability score	No. of data	Actual values (y)				Collapse probability
		1	5	6	Total	
0 ~ 10	20	18	2	0	20	Low
10 ~ 20	20	16	3	0	20	
20 ~ 30	18	3	13	2	18	
30 ~ 40	8	0	7	1	8	
40 ~ 50	16	0	11	5	16	
50 ~ 60	11	0	8	3	11	
60 ~ 70	1	0	1	0	1	
70 ~ 80	1	0	0	1	1	High
Total	95	37	45	13	95	

can be measured, based on a 100 points-system. It could be understood that a lower score represents higher stability and vice versa. In addition, to divide scores according to stability levels requires an additional process to select the related threshold.

The total data of 95 used for the analysis were evaluated by using the contribution scores for categories in Table 4. Table 5 shows 5 sections divided according to collapse probability with the number of data and actual values. It could be understood that a lower score has a lower y and vice versa.

Conclusion

This study evaluated the stability of a tunnel portal slope using a quantified technique, which is based on a general linear model. The important scores of each independent variable were allocated by using the ranges of the quantified values, based on the predicted coefficient of regression and the scores for categories of each independent variable were allocated so that those are equally spaced. The quantification model obtained from the results of evaluating the total data used for the quantification process provided precise results. In addition, it is expected that a more detail subdivision of response variables and sufficient data would produce a better stability evaluation standard.

References

Korean National Railroad, 2003, Volume 5, Tunnel design handbook (roadbed), pp. 1–42

Lee, P. H., Shin, J. S., Kim, I. H., Kim, W. S. Kim, Y. G., 2003, “Design criteria of environmental friendly tunnel portal considering road safety”, No.4, Volume 5, Tunnel Technology, Korean Tunnelling Association,

pp. 18–33
Korea Infrastructure Safety & Technology Corporation, 2003, “Detailed guides for safety check and precision safety diagnosis (cutting slopes)”, Volume 1, Ministry of Construction and Transportation, pp. 118
Korea Institute of Construction Technology, 1999, “Road design handbook (I)”, Volume 1, Ministry of Construction and Transportation, pp. 600–623
Yoo, C. S., Jeong, Y. W., Kim, J. H., Park, Y. J., Yoo, J. H., 2004, “Development and Application of Risk Assessment System according to the Construction of Tunnel based on GIS”, Korea Geotechnical Society, Volume 20, No.1, pp. 49–59.
Korea Institute of Construction Technology, 2002, “Development and Operation of Road Cut Slope Management System IV”, Ministry of Construction and Transportation, pp. 153–173