Electronic Supplementary Materials

For https://doi.org/10.1631/jzus.A2500581

Evaluating and predicting tunnel overbreak by integrating cloud modeling and fuzzy evaluation theory

Shibin YAO[1], Jian ZHOU[1], Biao HE[2], Chuanqi LI[1]

[1]*School of Resources and Safety Engineering, Central South University, Changsha 410083, China*
[2]*Civil, Structural & Environmental Engineering, University College Cork, Cork T12 K8AF, Ireland*

Section S1 Data description and visualization

This section provides a visual overview of the 523 records collected from the HXT tunnel, including the distributions of the selected input indicators and the output variable used for overbreak evaluation and prediction.

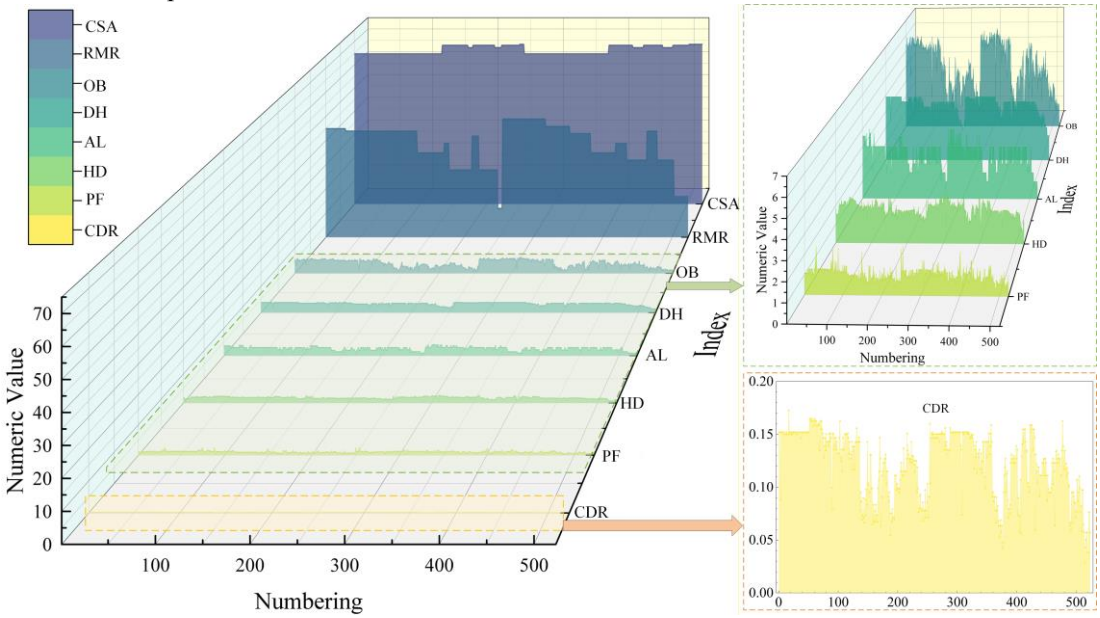


Fig. S1 HXT tunnel data visualization.

Section S2 Detailed description of weighting methods

To identify the most suitable evaluation theory for tunnel engineering and overbreak assessment, this study will employ five commonly used objective evaluation methods: coefficient of variation (CV), multi-correlation coefficient (MCC), criteria importance through intercriteria correlation (CRITIC), entropy weight (EW), principal component analysis (PCA), along with a novel objective evaluation method based on ridge regression optimization (RR), and one subjective evaluation (SE) method to determine the weights and establish models. By comparing the accuracy of these models, the optimal evaluation method will be selected, and a final CE model will be developed through the integration of both subjective and objective evaluation methods. The basic principles of each evaluation method are outlined below: First, an indicator matrix is established. If there are m evaluation indicators and n evaluation objects, the original evaluation matrix X is defined as follows, where X_{ij} represents the j-th indicator for the i-th evaluation object:

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{S1}$$

Section S2.1 Coefficient of variation method

The CV method is a technique used to measure the variability or dispersion of data. The weight assigned to each evaluation determine by the degree of variation in the corresponding data values for the evaluated objects. A higher variation indicates a greater contribution to the overall evaluation. The basic calculation process of the CV method can be divided into three steps:

(1) Calculate the standard deviation of each indicator.

$$S_{\mathrm{j}} = \sqrt{\frac{\sum_{i=1}^{n} (x_{ij} - \overline{x}_j)^2}{n}} \tag{S2}$$

(2) Calculate the CV for each indicator.

$$v_{\mathrm{j}} = \frac{S_j}{\overline{x}_j} \tag{S3}$$

(3) Normalize the CV.

$$w_{\mathrm{j}} = \frac{v_j}{\sum_{j=1}^{m} v_j} \tag{S4}$$

where S_{j} represents the correlation coefficient between indicators;

n represents the number of evaluation subjects;

x_{ij} represents the *j*-th indicator for the *i*-th evaluation object;

S_j represents the standard deviation of the *j*-th indicator;

v_{j} represents the coefficient of variation of the *j*-th indicator;

$\overline{x}_j$ represents the mean of the *j*-th indicator;

w_{j} represents the weight of the *j*-th indicator.

After that, the normalized values obtained for the corresponding indicators are used as the weights assigned to each indicator.

Section S2.2 Multi-correlation coefficient method

The multi-correlation coefficient (MCC) method determines indicator weights according to the degree of redundancy among indicators. If one indicator can be well explained by the remaining indicators, its independent information content is limited and its assigned weight should be relatively small. Conversely, if an indicator contains more independent information, a larger weight should be assigned. The calculation procedure is described as follows:

(1) First, the correlation coefficient matrix among the *m* indicators is constructed as:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mm} \end{bmatrix} \tag{S5}$$

(2) Second, the squared multiple correlation coefficient of the *j*-th indicator with respect to the other indicators is calculated as:

$$R_j^2 = r_j^T R_{-j}^{-1} r_j \tag{S6}$$

(3) Third, the independent information coefficient of the *j*-th indicator is expressed as:

$$\rho_j = 1 - R_j^2 \tag{S7}$$

(4) Finally, the weight of each indicator is obtained by normalizing the independent information coefficient:

$$w_j = \frac{\rho_j}{\sum_{j=1}^{m} \rho_j} \qquad j = 1, 2, \ldots, m \tag{S8}$$

where r_{ij} is the Pearson correlation coefficient between the *i*-th and *j*-th indicators;

R_{-j} denotes the correlation matrix of the remaining *m*-1 indicators;

r_j denotes the correlation vector between the *j*-th indicator and those indicators;

R_j^2 is the squared multiple correlation coefficient of the *j*-th indicator;

ρ_j is the independent information coefficient of the *j*-th indicator;

w_j is the final weight assigned to the *j*-th indicator.

Section S2.3 Criteria importance through intercriteria correlation method

The CRITIC method determines the weights entirely based on the data itself, focusing on the comparative strength of evaluation indicators and the conflict between them. The specific calculation procedure is as follows:

(1) First, the raw data are normalized to eliminate dimensional effects. For a positive indicator, the normalized value is calculated as:

$$x'_{ij} = \frac{x_{ij} - x_{j,\min}}{x_{j,\max} - x_{j,\min}} \tag{S9}$$

For a negative indicator, the normalized value is calculated as:

$$x'_{ij} = \frac{x_{j,\max} - x_{ij}}{x_{j,\max} - x_{j,\min}} \tag{S10}$$

(2) Second, the variability of the *j*-th indicator is measured using the standard deviation of the normalized values:

$$S_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x'_{ij} = \overline{x}'_j)^2} \tag{S11}$$

(3) Third, the conflict of the *j*-th indicator with the other indicators is defined as:

$$R_j = \sum_{k=1, k \neq j}^{m} (1 - r_{jk}) \tag{S12}$$

(4) Fourth, the information content of the *j*-th indicator is calculated as:

$$C_j = S_j \times R_j \tag{S13}$$

(5) Finally, the weight of each indicator is obtained by normalizing the information content:

$$W_{\mathrm{j}} = \frac{C_j}{\sum_{j=1}^{m} C_{\mathrm{j}}} \tag{S14}$$

Where x_{ij} is the value of the *j*-th indicator for the *i*-th sample;

x'_{ij} is the normalized value of the *j*-th indicator for the *i*-th sample;

$x_{j,\max}$ represents the maximum value of the *j*-th indicator;

$x_{j,\min}$ represents the minimum value of the *j*-th indicator;

S_j is the standard deviation of the normalized values of the *j*-th indicator;

$\bar{x}'_j$ is the mean normalized value of the *j*-th indicator;

r_{jk} is the correlation coefficient between the *j*-th and *k*-th indicators;

R_{j} is the conflict measure of the *j*-th indicator with respect to the other indicators;

C_{j} represents the information content of the *j*-th indicator;

W_{j} represents the weight of the *j*-th indicator.

Section S2.4 Entropy weight method

In the traditional concept of information, entropy represents the degree of disorder in a system, measuring the effective reliability of the data. The entropy method determines the weight of each indicator based on the amount of information contained in the data of each indicator. The greater the difference between indicators, the smaller the entropy, the larger the amount of information transmitted, and thus the greater the corresponding weight. Given *m* evaluation indicators and *n* evaluation objects, the three steps to calculate the weight using the entropy method are as follows.

(1) Normalization

$$a_{\mathrm{ij}} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}} \tag{S15}$$

(2) Calculate the entropy value for each indicator

$$H_{\mathrm{j}} = -k\sum_{i=1}^{n} a_{ij} \ln a_{ij} \left(k = 1/\ln n\right) \tag{S16}$$

(3) Convert the entropy value into weights.

$$w_{\mathrm{j}} = \frac{1 - H_j}{m - \sum_{j=1}^{m} H_j} \tag{S17}$$

where x_{ij} represents the value of the *j*-th indicator for the *i*-th object;

a_{ij} represents the standard value of the *j*-th indicator for the *i*-th object;

H_{j} represents the entropy value of the *j*-th indicator;

n represents the number of evaluation objects;

m represents the number of indicators;

w_{j} represents the weight of the *j*-th indicator.

Section S2.5 Principal component analysis method

PCA reduces the dimensionality of originally correlated indicators and recombines them into a new set of uncorrelated indicators. These new indicators are mutually independent while retaining the information from the original indicators. The PCA calculation involves the following three steps.

(1) Solve the correlation matrix of the original data matrix (Eq. (S5)).

(2) Calculate the eigenvalues λ and eigenvectors A of the correlation matrix.

$$(R-\lambda I)=0 \tag{S18}$$

(3) Determine the principal components based on the eigenvalues and calculate the corresponding weights.

Section S2.6 Subjective evaluation method (expert judgment)

Due to the multiple factors and complex relationships involved in overbreak, as well as the incomprehensible mechanisms underlying the phenomenon, it is essential to incorporate subjective weights to optimize and complement the limitations of objective weights. In this study, the expert judgment method, a subjective weighting technique, is used to determine the weights of each indicator under subjective conditions. A panel consisting of one frontline technical expert, two industry professors, and two researchers is invited to independently assign weights to the seven identified indicators. The final subjective evaluation weights are then derived by averaging the results from each expert.

Section S2.7 Comprehensive evaluation method

The comprehensive weighting method seeks to optimize the combination of weights by leveraging the strengths of both subjective and objective evaluation approaches. In this study, objective weights are evaluated based on their classification accuracy under the overbreak grading criterion, and the set of weights corresponding to the highest accuracy is selected. These objective weights are then combined with those derived from the subjective weighting method, and their average is calculated to obtain the final integrated weights. The accuracy of this combined weighting approach was subsequently calculated to build the final CE model.

Section S3 Comparison of evaluation accuracy among different weighting methods

This section presents a supplementary comparison of the evaluation accuracy obtained using different weighting methods under the established overbreak grading criterion.

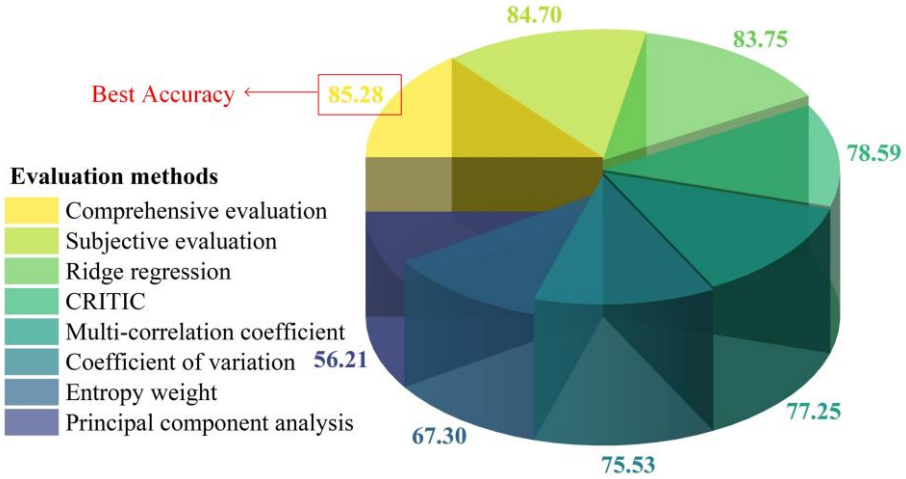


Fig. S2 Accuracy of different evaluation models.