Selection of representative embankments based on rough set-fuzzy clustering method

OU Bin¹, LIN Zhi-xiang²,⁵, FU Shu-yan⁴¹ and GAO Sheng-song⁴

¹College of Water Conservancy, Yunnan Agricultural University, Kunming 650201, China
²Engineering Management Bureau, Water Resources Department of Yunnan Province, Kunming 650021, China
³State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China
⁴College of Mechanical and Electrical Engineering, Yunnan Agricultural University, Kunming 650201, China
⁵lzx7808@163.com

Abstract. The premise condition of comprehensive evaluation of embankment safety is selection of representative unit embankment, on the basis of dividing the unit levee the influencing factors and classification of the unit embankment are drafted. Based on the rough set-fuzzy clustering, the influence factors of the unit embankment are measured by quantitative and qualitative indexes. Construct to fuzzy similarity matrix of standard embankment then calculate fuzzy equivalent matrix of fuzzy similarity matrix by square method. By setting the threshold of the fuzzy equivalence matrix, the unit embankment is clustered, and the representative unit embankment is selected from the classification of the embankment.

1. Introduction

The embankments are generally longer they will generally choose some representative dike sections to evaluate the safety of dike. At present, there are not many researches on the selection of representative embankments which are mainly random sampling method[1] and geological conditions prospecting method. Random sampling method is adopted to select random grades or random selection of embankments. For example, Zhang Qing-bin divides the dike segments with the same features into 1km evaluation unit when the existing embankments of hidden danger can be shortened appropriately. The accuracy of random sampling method depends on the number of units selected. When the embankment is longer, the workload will be large so is suitable for the shorter embankment[2]. The exploration method of geological conditions mainly uses the geological prospecting data to select the general geological condition or the history of the dangerous section of the embankment to evaluate the overall security state of the embankment[3]. The method of workload is simple but also easy to ignore the parts of the embankments which results in the inaccuracy of the comprehensive safety evaluation. The above method, each has its advantages and disadvantages, this article on the basis of some unit embankments are selected, then used to measure the similarity of the various indexes of the embankment on rough set-fuzzy clustering method that representative embankments can be selected more comprehensively and rationally.
2. Criterion of unit embankment division
The embankment is composed of a number of unit embankments, and the basic principles of division include: (1) the once dangerous embankment is divided into one unit embankment, (2) there are no dangerous embankments according to the main features of the embankment section (structural form, geological conditions, supporting forms, etc.), and the same embankments are divided into the same units[1].

3. Index classification of unit embankment
The unit embankment contains many characteristic indexes, in this paper, the five core indexes of embankment are selected, including the content of soil clay, the dry density of levee, the slope ratio, the structure of the embankment and the hidden danger, such as the ratio of first 3 indexes belong to quantitative indexes, the last belong to the qualitative indexes. The qualitative indexes are divided into 4 categories by table 1.

| Index classification | Class1 | Class2 | Class3 | Class4 |
|----------------------|--------|--------|--------|--------|
| Foundation embankment| A single structure consisting of thick silt loam | The upper soft clay has a small thickness and a large thickness of the lower layer | The multi-structure of sandy soil mainly | The multi-structure of loam soil mainly |
| Hidden troubles in dikes | Clearly visible cracks, caves, and so on, and developed cracks, caves, and so on, but no development trend | There are slight cracks, no caves, etc | No hidden troubles |

4. Based on rough set - fuzzy cluster analysis of unit embankment classification
A 4-tuple is commonly used to represent information system in rough set theory[4,5], which $U$ is all objects in the data set; $A$ is the full attribute of data set, $A = C \cup D$; $C$ is the set of conditional attribute; $D$ is a collection of decision attributes; $V$ is the range of attributes; $f : U \times A \to V$ is the information function, specifying the attribute value of each object in $U$.

The theory is applied to the unit embankment chosen clustering, $U$ is the collection of all unit embankments, $C$ is collection of core index attribute set, which contains the quantitative index and qualitative index, The value $V$ is the assignment of all attributes. The form of rough set decision table for unit embankments is shown in table 2.

| Unit embankments | $C_1$ | $C_2$ | ... | $C_p$ |
|------------------|------|------|-----|------|
| $x_1$            | $c_{11}$ | $c_{12}$ | ... | $c_{1p}$ |
| $x_2$            | $c_{21}$ | $c_{22}$ | ... | $c_{2p}$ |
| ...              | ...   | ...   | ... | ...   |
| $x_n$            | $c_{n1}$ | $c_{n2}$ | ... | $c_{np}$ |

In table 2, $C_1$-$C_p$ represent the set of all indexes, which can describe the characteristics of unit embankments from different angles. However, since their dimensions and magnitudes are not consistent, the original index values need to be processed. This paper studies the quantitative indicators using normalized processing [6].

For quantitative and qualitative indexes in rough set decision table, a suitable similarity measurement formula is needed to construct the fuzzy similarity matrix. The similarity of unit embankments can be measured by the distance between them. For quantitative indicators, the usual distance formula is Minkowski distance, Manhattan distance, Euclidean distance, Mahalanobis distance, Chebyshev distance, etc. In this paper the Euclidean distance formula is adopted.
Where: \( d_E \) is the Euclidean distance of the quantitative index for the unit embankment; \( c_{ik} \) and \( c_{jk} \) are the \( k \)th quantitative index value of unit \( X_i \) and \( X_j \) respectively; \( w \) is number of quantitative indexes.

For qualitative indicators, define the function \((a, b)\) as

\[
f(a, b) = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases}
\]

The distance about the qualitative index between the unit embankment \( X_i \) and \( X_j \) is

\[
d_H(X_i, X_j) = \sum_{k=1}^{q} f(c'_{ik}, c'_{jk})
\]

Where: \( c'_{ik} \) and \( c'_{jk} \) are the \( k \)th quantitative index value of unit \( X_i \) and \( X_j \) respectively; \( q \) is the number of quantitative indexes.

The above analysis shows that the similarity measure formula of unit embankment \( X_i \) and \( X_j \) is

\[
r_{ij} = r(X_i, X_j)
\]

\[
= \theta \left( 1 - \frac{d_E(X_i, X_j)}{\max_{X_a, X_b \in U} d_E(X_a, X_b)} \right) + (1 - \theta) \left( 1 - \frac{d_H(X_i, X_j)}{\max_{X_a, X_b \in U} d_H(X_a, X_b)} \right)
\]

Where: \( \theta \) is the proportion of quantitative index to the total indexes; \( \max_{X_a, X_b \in U} d_E(X_a, X_b) \) is the maximum quantitative index distance of the unit embankment; \( \max_{X_a, X_b \in U} d_H(X_a, X_b) \) is the maximum qualitative index distance of the unit embankment.

The formula (4) can establish the n-order fuzzy similarity matrix \( R \), which satisfies the following two properties: ① self-reflexivity, \( r_{ii} = 1 \); ② Symmetry, \( r_{ij} = r_{ji} \). The matrix \( R \) shows the degree of similarity between the unit embankments.

Fuzzy clustering analysis uses fuzzy similarity relation or fuzzy equivalence relation to divide the unit embankment into different classes according to certain standard. High similarity between the same class of unit embankment, less similarity between different classes. At present, there are many fuzzy clustering methods based on fuzzy division concepts, such as transitive closure method, maximal tree method and net-making method, etc. [6].

The transitive closure method is defined by: \( R \) is the relation on the nonempty set \( A \), if the relationship \( R \) on \( A \) satisfies the following conditions: ① \( R \subseteq R^2 \) ② \( R \) is reflexive (symmetric and transmitted) , If \( R \) is reflexive (symmetric and delivered), then there is \( R \subseteq R \). \( R \) is called the reflexive (symmetric, transitive) closure of \( R \), denoted by \( t(R) \). The matrix \( R \) established by formula (4) satisfies the reflexivity and symmetry, and does not satisfy the transitivity, that is not satisfied \( R \circ R \subseteq R \). So we need to find the transitive closure \( t(R) \) of \( R \).

The literature theorem 1 shows that[5], if \( R \) is a fuzzy similarity matrix, there is a minimum natural number \( k \), so that the transitive closure, \( t(R) = R^k \) and \( h \) is greater than \( k \) for all natural numbers, always have \( R^h = R^k \). Based on this, the method is used to calculate the transitive closure of fuzzy similarity matrix, namely fuzzy equivalent matrix. Then the threshold \( \lambda \) of the fuzzy equivalent matrix is set, and cluster of the unit embankment is according to it. During the clustering process, \( \lambda \) can be reduced continuously until the appropriate number of classes is obtained.
5. Application example

5.1 Project profile
An embankment project is 12.92km long, which is an important flood protection project in the plain area of the drainage basin. The main embankment of the wide of top embankment is 6m, the average height of the embankment is 12.5 m, the upstream slope is 1:2.5, the downstream is restricted by the topography, the slope ratio is 1:2 and 1:2.5, and the embankment is the loam soil. The embankment foundation has the boulders, gravel sand and a small amount of sand and sandy loam. The upper part is sandy loam, powdery sand, cohesive soil and silt soil. The surface layer is the most recently deposited silty soil, which has been washed and deposited by water. Because the water flow is slow and stable, the sedimentary regularity and stratification are obvious, and the layers and interstratification, thin layer distribution and horizontal bedding development are the main characteristics.

5.2 Impact index classification
According to the unit dike division of basic principles, the total length of 12.92 km dikes were divided into some 12 units, according to the body soil clay content, dry density, downstream slope ratio, embankment foundation structure, hidden dangers, etc. unit embankment classification index statistics as shown in table 3.

| unit embankments | clay content (%) | dry density (g/cm³) | downstream slope ratio | hidden danger | foundation structure |
|------------------|------------------|---------------------|------------------------|---------------|---------------------|
| x1               | 9.20             | 1.45                | 2                      | 4             | 1                   |
| x2               | 9.20             | 1.43                | 2                      | 4             | 2                   |
| x3               | 14.40            | 1.44                | 2                      | 1             | 2                   |
| x4               | 9.70             | 1.52                | 2                      | 1             | 4                   |
| x5               | 9.90             | 1.57                | 2                      | 4             | 4                   |
| x6               | 21.10            | 1.63                | 2                      | 4             | 2                   |
| x7               | 12.70            | 1.58                | 2                      | 2             | 2                   |
| x8               | 14.30            | 1.64                | 2                      | 4             | 2                   |
| x9               | 15.10            | 1.66                | 2.5                    | 4             | 3                   |
| x10              | 13.40            | 1.55                | 2.5                    | 4             | 1                   |
| x11              | 12.00            | 1.58                | 2.5                    | 3             | 2                   |
| x12              | 21.10            | 1.63                | 2                      | 4             | 2                   |

5.3 Unit embankment classification
The quantitative and qualitative indexes in table 3 are treated by formula (1) and (3) respectively, then the formula (4) is adopted to establish the fuzzy similarity matrix of the unit embankment.
\[ R = \begin{pmatrix}
1.0000 & 0.3845 & 0.5270 & 0.69368 & 0.30368 \\
0.3845 & 0.78751 & 0.53685 & 0.706085 & 0.59063 \\
0.5270 & 0.53685 & 0.75321 & 0.76728 & 0.431942 \\
0.69368 & 0.706085 & 0.76728 & 0.64416 & 0.100000 \\
0.30368 & 0.59063 & 0.431942 & 0.64416 & 0.100000
\end{pmatrix} \]

\[ R^+ = \begin{pmatrix}
1.0000 & 0.3845 & 0.5270 & 0.69368 & 0.30368 \\
0.3845 & 0.78751 & 0.53685 & 0.706085 & 0.59063 \\
0.5270 & 0.53685 & 0.75321 & 0.76728 & 0.431942 \\
0.69368 & 0.706085 & 0.76728 & 0.64416 & 0.100000 \\
0.30368 & 0.59063 & 0.431942 & 0.64416 & 0.100000
\end{pmatrix} \]

Then the fuzzy similarity matrix \( R \) is modified by the flat method to obtain the following matrix: after \( R \) is tested, it satisfies reflexivity, symmetry and \( R^+ \subseteq R \), so \( R \) is a fuzzy equivalent matrix that satisfies the transmission. Taking \( \lambda = 0.7 \) is equal to 0.7, and you get \( A \) intercept matrix B. Take \( \lambda = 0.7 \), and get the \( A \) intercept matrix \( R_A \).

\[ R_A = \begin{pmatrix}
1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0
\end{pmatrix} \]

\[ r_{ij} = \begin{cases} 1 & r_{ij} \geq \lambda \\
0 & r_{ij} < \lambda \end{cases} \]

The embankment is divided into five classes which include \( \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\} \), \( \{x_1\} \), \( \{x_6, x_{12}\} \), \( \{x_7\} \) and \( \{x_{10}\} \). By analyzing the cluster analysis of embankments, the 12 unit embankments were divided into five categories which one selected from each class.

6. Conclusion

According to the historical risk situation and embankment section and structural support condition, the rough set-fuzzy clustering method is used to measure the similarity measure and the quantitative index of the unit embankment. The fuzzy similarity matrix of the standard embankment is constructed, then fuzzy equivalent matrix corresponding to the fuzzy similarity matrix is calculated by the square method. By adjusting the appropriate threshold, the unit embankment is classified and the representative unit embankment is selected as the basis for the comprehensive safety evaluation of the embankment.

Acknowledgments

This work is jointly supported by the Open Foundation of State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering(Grant No.2016491811), Scientific Research Foundation of Yunnan Provincial Department of Education(Grant No.2016ZZX109).
References

[1] ZHANG Qing-ming, XU Shuai, ZHOU Yang. Research on Units Division of Embankment Engineering Safety Assessment [J]. Yellow River, 2012, 34(01): 16-18.

[2] ZHANG Qing-bin. Evaluation on Dike Aging in the Lower Reaches of Yellow River [J]. Bulletin of Soil and Water Conservation, 1999, (05): 21-23.

[3] XU Ru-hai. Application of Fuzzy Theory and Gray Systematic Theory in the Embankment Bank Slope Stability Analyzing [D]. Nanning: Guangxi University, 2006.

[4] AO Liang-zhong, ZHAO Wei, XIA Cun-jiang. Cluster Analysis of the Aero-engine Status Based on Fuzzy Information Fusion [J]. Journal of Xi'an Aeronautical University, 2015, 34(5): 20-25.

[5] SUN Hui-Qin, XIONG Zhang. Fuzzy Cluster Based on Rough Set and Result Evaluating [J]. Journal of Fudan University (Natural Science), 2004, 43(5): 819-822.

[6] XIAO Han–guang, CAI Cong-zhong. Comparison study of normalization of feature vector [J]. Computer Engineering and Applications, 2009, 45(22): 117-119.