Research Article

Sandy Soil Liquefaction Prediction Based on Clustering-Binary Tree Neural Network Algorithm Model

Yu Wang and Jiachen Wang
School of Civil Engineering, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Jiachen Wang; wangjiachen@bjtu.edu.cn

Received 19 September 2021; Accepted 25 October 2021; Published 13 December 2021

Academic Editor: Zhiyong Chen

Copyright © 2021 Yu Wang and Jiachen Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The neural network algorithm is a small sample machine learning method built on the statistical learning theory and the lowest structural risk principle. Classical neural network algorithms mainly aim at solving two-classification problems, making it infeasible for multiclassification problems encountered in engineering practice. According to the main factors affecting sand liquefaction, a sand liquefaction discriminant model based on a clustering-binary tree multiclass neural network algorithm is established using the class distance idea in cluster analysis. The model can establish the nonlinear relationship between sand liquefaction and various influencing factors by learning limited samples. The research results show that the hierarchical structure based on the clustering-binary tree neural network algorithm is reasonable, and the sand liquefaction level can be categorized accurately.

1. Introduction

Earthquake-induced sand liquefaction is a very destructive phenomenon of geological disasters, which can bring about damages to the farmland, roads, bridges, various civil buildings, and water conservancy facilities, causing great harm to the social economy and human life safety. Therefore, sand liquefaction under the action of earthquake has always been one of the hot issues in the field of geotechnical engineering and disaster geology. As earthquakes are inevitable, the possibility of seismic liquefaction assessment and prediction of the site are the focus of the research on this issue [1–4].

The current discrimination methods for seismic liquefaction are mainly divided into two categories: the empirical analysis method based on the investigation data of the sand liquefaction disaster at the earthquake site and the experimental analysis method based on the field or indoor test.

The commonly used empirical analysis methods mainly include the standard penetration method, the critical void ratio, and the energy discrimination method [5]. The experimental analysis methods mainly include the Seed–Idriss simplification method [6], shear wave velocity method [7], static cone penetration method [8], dynamic simple shear test, shaking table test [9, 10], and dynamic triaxial test [11, 12].

In recent years, the application of machine learning-based research methods in the field of earth sciences has gained a lot of attention [13–20]. Traditional geological research is gradually moving closer to the fields of data deep mining, big data, and artificial intelligence, such as in coal mine water inrush prediction [21] and research on landslides [22, 23].

In the study of seismic liquefaction, Xue and Yang examined the potential of support vector machines (SVMs) for assessing liquefaction potential based on cone penetration test (CPT) field data [24]. A hybrid model based on a combination of SVMs and particle swarm optimization (PSO) was proposed in the study to improve the forecasting performance. Chou and Thedja proposed a novel classification system integrating swarm and metaheuristic intelligence, i.e., a smart firefly algorithm (SFA), with a least-square support vector machine (LSSVM), to provide decision-makers with timely warnings of geotechnical hazards.
Advances in Civil Engineering

2 Advances in Civil Engineering

2.1. Introduction to Neural Network Algorithm.

One of the most popular feedforward artificial neural networks is the multilayer perceptron (MLP), which is composed of a large number of nodes connected to each other by imitating the transmission of electrical signals between biological neurons in the natural world. Among them, multilayer perceptron (MLP) is the most popular feedforward artificial neural networks. A typical MLP neural network usually contains an input layer, a hidden layer (one or more layers), and an output layer. The input layer and the input variables have the same number of nodes; the hidden layer contains the weight values of the connections between neurons and the corresponding threshold function, which are used to transmit and process the signal; and the output value is the final data to be obtained.

The mathematical representation of the feedforward neural network is as follows:

\[ y_k(x) = \sum_{i=1}^{M} \omega_{ih} \times T_r(z) + b_{ih}, \]  
\[ z = \sum_{i=1}^{D} \omega_{ho} \times x_i + b_{ho}, \]

where \( x \) is the input parameter, \( \omega_{ih} \) and \( \omega_{ho} \) are the input layer-hidden layer weight and hidden layer-output layer weight, respectively, \( b_{ih} \) and \( b_{ho} \) are the deviation parameter, \( M \) is the number of nodes in the hidden layer, \( d \) is the number of nodes in the input layer, and \( T_r(z) \) is the transfer function that performs a nonlinear transformation on the summation input.

2.2. Neural Network Algorithm Method of Multiclassification Based on Clustering-Binary Tree.

The basic neural network algorithm is aimed at two-classification problems, and the multiclassification problem is more common in engineering practice. The multiclass neural network algorithm method based on the binary tree is one of the effective methods to solve the multiclassification problem. As shown in Figure 1, the binary tree structure classifier solves a complex multiclass problem into multiple two-class problems. A multiclass problem can be transformed into multiple two-category problems in various forms with various corresponding binary tree structures. For a k-category problem, the number of binary trees that can be constructed is \( N_k = \prod_{i=1}^{k-1} 2 * i - 1, k > 2 \). As shown from Figure 1, the classification performance of the neural network algorithm subclassifier of the upper node has a greater impact on the generalization of the entire classification model. If a classification error occurs at a node, the error will continue, resulting in a phenomenon of “error accumulation,” making the next-level node lose its classification meaning. It can be seen that constructing a reasonable binary tree structure is essential to the correct classification of the model. In order to make the binary decision tree have the optimal level of performance and construct a reasonable binary tree hierarchy, the samples must be divided into two groups at the decision point in an approximate optimal level method even if the separability between the two type sample types in the

At present, in the field of geotechnical engineering, machine learning methods are often used to consider problems such as binary classification. In contrast, the problems encountered in geotechnical practice are often multiclassification problems. To date, multiclassification research is still in its infancy. The application of neural network algorithms in geotechnical engineering is rarely reported in the literature. This paper introduces multiple types of neural networks into the evaluation of sand liquefaction, extracts characteristic parameters, and adopts the cluster analysis method of class distance to establish a sand soil based on the cluster-binary tree multiclass neural network algorithm. The liquefaction discriminant model has achieved satisfactory evaluation results.

2. Neural Network Algorithm

2.1. Introduction to Neural Network Algorithm. The neural network algorithm is a typical supervised machine learning algorithm, which is composed of a large number of nodes connected to each other by imitating the transmission of electrical signals between biological neurons in the natural world. Among them, multilayer perceptron (MLP) is one of the most popular feedforward artificial neural networks. A typical MLP neural network usually contains an input layer, a hidden layer (one or more layers), and an output layer. The input layer and the input variables have the same number of nodes; the hidden layer contains the weight values of the connections between neurons and the corresponding threshold function, which are used to transmit and process the signal; and the output value is the final data to be obtained.

The mathematical representation of the feedforward neural network is as follows:

\[ y_k(x) = \sum_{i=1}^{M} \omega_{ih} \times T_r(z) + b_{ih}, \]  
\[ z = \sum_{i=1}^{D} \omega_{ho} \times x_i + b_{ho}, \]
upper layer is strong. In so doing, the distance between the cluster centers of the two groups of samples is the largest, and the divergence of each group of samples is the smallest.

2.2.1. Definition of Distance. For given A class and B class \((x_a \in A, x_b \in B)\), the distance between two samples can be expressed as \(d(x_a, x_b)\). In the linear case, the distance between the two samples \(x_a\) and \(x_b\) is defined as follows:

\[
d(x_a, x_b) = ||x_a - x_b|| = \sqrt{\sum (x_a^n - x_b^n)^2}.
\] (4)

In the case of nonlinearity, the distance between the two samples \(x_a\) and \(x_b\) is defined as follows:

\[
d(x_a, x_b) = \sqrt{K(x_a, x_a) - 2(x_a, x_b) + K(x_b, x_b)}.
\] (5)

This paper adopts the idea of class distance in cluster analysis as the algorithm for generating binary trees. First, the class farthest from others is segmented, and the optimal hyperplane constructed at this time has good generalization.

Algorithm flow of multiclass neural network based on clustering-binary tree is as follows:

1. Calculate the distance between classes according to formulas (4) or (5): \(d_{i,j} (i, j = 1, 2, \ldots, k, i \neq j)\).
2. For each class, there are \((k - 1)\) class spacing values from other classes, and the class spacing values of each class are sorted in ascending order from small to large and renumbered. For example, the spacing \(d_{i,j} (i, j = 1, 2, \ldots, k, i \neq j)\) between the \(i\)-th class and other classes is sorted as \(d^1_i \leq d^2_i \leq \cdots \leq d^k_i\).
3. Compare the minimum class spacing of each category \(d^i_q (i = 1, 2, \ldots, k)\), and sort them in descending order from largest to smallest. When there are two or more categories with the same minimum category spacing \(d^i_q\), the sizes \(d^i_q\) can be compared again and so on. For example, when \(d^{i+1}_q > d^i_q\), sort the \(q\)-th category in front of the \(p\)-th; if the first \(d^i_q = d^i_p\), then compare \(d^i_q\) and \(d^j_p\); if \(d^i_q < d^i_p\), sort the \(p\)-th category in front of the \(q\)-th category so that the larger the difference, the more the category partition. Thus, one can get the arrangement of all categories \(n_1, n_2, \ldots, n_k\).

4. According to the generated binary tree, a binary neural network training algorithm is used to construct the optimal hyperplane of the binary node. Take the \(n_1\)-th sample as the positive sample set and other samples as the negative sample set. The neural network training algorithm is used to construct the binary neural network subclassifier at the root node. The \(i\)-th subclassifier takes the \(i\)-th sample as a positive sample and uses the \(i + 1, i + 2, \ldots, k\)-th sample as a negative sample to train the neural network \(i\) subclassifier until \((k - 1)\)-th subclassifier separates the \((k - 1)\)-th category from the \(k\)-th category. In this way, the recognition of multiple types of problems can be completed.
5. The multiclassification classification ends and outputs the result.

3. Sand and Soil Discriminant Model Based on the Neural Network Algorithm

3.1. Analysis of Influencing Factors and Model Establishment. Sand liquefaction is a form of foundation failure caused by earthquake damages. It is an important content of geological engineering evaluation in earthquake areas. The liquefaction of saturated sand is a very complicated process with many influencing factors.

It can be roughly divided into three categories:

(i) The nature of the sand layer, such as the type of soil, particle composition, and degree of compaction
(ii) The buried depth of the soil layer and the groundwater level
(iii) Seismic factors, such as earthquake intensity, epicenter distance, and earthquake duration

This paper proposes representative parameters from many influencing factors as the distinguishing index of sand liquefaction, including the seismic intensity \(I\), epicenter distance, and groundwater level.anni.
distance $R$, sand buried depth $d$, groundwater level $d_w$, standard penetration number $N$, average particle size $D_{50}$, nonuniformity coefficient $C_u$, and the dynamic shear stress ratio $(\tau_d/\sigma'_v)$. The eight variables are used as classification indicators. Such selection helps avoid that attributes with a large value range are more dominant than those with a small value range.

First, by performing normalization preprocess on each attribute of the sample data and adjusting linearly to $[-1, +1]$, the sand liquefaction evaluation can be divided into 4 grades, which are severe liquefaction (I), moderate liquefaction (II), slight liquefaction (III), and no liquefaction (IV).

The use of neural network algorithms to discriminate sand liquefaction is to find the nonlinear relationship between influencing factors and sand liquefaction. Assuming that $n$ samples are collected, the problem is abstracted as the mapping from attribute set $X$ to classification set $Y$; that is, the index matrix is $X_{n \times 8}$, $Y = \{I, II, III, IV\}$. This paper sorts out 40 typical examples from the literature of the Wenchuan Earthquake and divides them into training sample sets and test sample sets. Among them, 30 samples are used for learning machine learning, as shown in Table 1, and the other 10 samples are used to test the model performance.

### 3.2. Sample Learning

There are 4 levels of sand liquefaction, indicating a 4-classification problem. A decision binary tree with 3 classifiers needs to be established to divide the sample into 4 subsets. Through the learning of 30 samples, various classification functions are obtained, and the sand liquefaction discriminant model is established based on the neural network algorithm of the clustering-binary tree.

The parameters of the neural network algorithm are mainly kernel function parameters and penalty factor $C$. In the research, it is found that for different training parameters, the learning efficiency and generalization ability of neural network algorithms are different. In this paper, the radial basis function is selected as the kernel function:

$$K(x, x_i) = \exp \left( -\frac{||x - x_i||^2}{\sigma^2} \right),$$

where $x_i$ and $x$ are the training input and $\sigma$ is the width of the kernel function.

In this way, the training parameters of the support vector machine in this paper are $\sigma$ and $C$. The size of $\sigma$ affects the output response interval of the sample. When $\sigma$ is small, the response interval is narrow, leading to less risky
classification. The penalty factor $C$, which represents the degree of punishment for errors, is a compromise between training error and promotion ability by controlling training accuracy. The larger the $C$ is, the greater the penalty for misclassification is.

The value of $\sigma$ and $C$ can be determined according to the samples during the training process. This paper uses the variable increment iteration method based on the number of misclassified samples to obtain $\sigma = 0.28$ and $C = 300$. Through the learning of 30 samples, the error-free classification of the sample is given in Table 1.

3.3. Sample Prediction. When using the clustering-binary tree neural network algorithm for classification, starting from the root classifier, the attributes of the unknown samples are substituted into the classification function to test their values one by one.

When the test value is 1, it reaches the leaf node and stops moving forward, indicating that the sample is the class represented by the leaf node. When the test value is $-1$, one must go down the branch until reaching a certain leaf node. The other 10 samples are input as test samples into the decision binary tree, and the classification of each sample is shown in Table 2. One can see that the learning results are consistent with the actual ones. Therefore, the method in this paper can correctly distinguish multiple types of problems, indicating that the promotion performance of the support vector machine is good.

4. Conclusion

In this paper, a multiclass neural network is introduced into the discrimination of sand liquefaction. During the modelling process, it was found that the promotion performance of the multiclass neural network model based on the binary tree is related to the binary tree structure. The learning efficiency and promotion performance vary with the binary tree structure and the model parameter. This paper adopts the idea of class distance in cluster analysis as the algorithm of generating binary tree and establishes a multiclass neural network to predict the sand liquefaction based on cluster-binary tree. The model establishes the nonlinear relationship between sand liquefaction and various influencing factors through the learning of limited samples. The research results show that the required number of classifiers in the clustering-binary tree neural network is small (only 3 classifiers are needed for 4 types of problems), the repetitive training samples are few, the binary tree hierarchy is reasonable, the classification accuracy is high, and the generalization is good. Thus, the proposed model predicts the liquefaction level accurately and can be used for sandy soil.

The neural network algorithm method is based on the VC theory of statistical theory and the principle of minimum structural risk. It seeks the best compromise between complexity and learning ability based on limited sample information and has a good generalization ability. As a new machine learning method, the neural network algorithm has broad application prospects in geotechnical engineering and other fields because of its unique advantages.

**Data Availability**

The data are generated from the field and can be available from the corresponding author upon request.

**Conflicts of Interest**

The authors declare that there are no conflicts of interest regarding the publication of this paper.

**Acknowledgments**

This work was supported by Beijing Jiaotong University.

**References**

[1] Y. Zhang, Y. Xie, Y. Zhang, J. Qiu, and S. Wu, “The adoption of deep neural network (DNN) to the prediction of soil liquefaction based on shear wave velocity,” *Bulletin of Engineering Geology and the Environment*, vol. 80, no. 6, pp. 5053–5060, 2021.

[2] Y.-g. Zhang, J. Qiu, Y. Zhang, and Y. Wei, “The adoption of ELM to the prediction of soil liquefaction based on CPT,” *Natural Hazards*, vol. 107, no. 1, pp. 539–549, 2021.
[3] Y. Zhang, J. B. Qiu, Y. G. Zhang, and Y. L. Xie, “The adoption of a support vector machine optimized by GWO to the prediction of soil liquefaction,” *Environmental Earth Sciences*, vol. 80, 2021.

[4] Y. Zhang, J. Qiu, Y. Zhang, and R. Liao, “The establishment of a constitutive model of sand under monotonic loading by adopting the support vector machine (SVM),” *Arabian Journal for Science and Engineering*, vol. 25, pp. 1–15, 2021.

[5] M. Ishac and A. Heidebrecht, “Energy dissipation and seismic liquefaction in sands,” *Earthquake Engineering & Structural Dynamics*, vol. 10, no. 1, pp. 59–68, 2010.

[6] J. Wang, T. Zuo, X. Li, Z. Tao, and J. Ma, “Study on the fractal characteristics of the pomegranate Biotite Schist under impact loading,” *Geoﬁuids*, vol. 2021, Article ID 1570160, 8 pages, 2021.

[7] R. Dobry, R. Ladd, and D. Powell, *Prediction of Pore Water Pressure Buildup and Liquefaction of Sands during Earthquakes by the Cyclic Strain Method*, National Bureau of Standards Building Science Series, Annapolis, MA, USA, 1982.

[8] P. Robertson and C. Wride, “Evaluating cyclic liquefaction potential using the cone penetration test,” *Canadian Geotechnical Journal*, vol. 35, no. 3, pp. 442–459, 1988.

[9] X. Ling, C. Wang, and Z. Wang, “Study on large-scale shaking table proportional model test for free-ground liquefaction arisen from earthquake,” *Earthquake Engineering & Engineering Vibration*, vol. 23, no. 6, pp. 138–143, 2003.

[10] I.-S. Ha, S. M. Olson, M.-W. Seo, and M.-M. Kim, “Evaluation of relquefaction resistance using shaking table tests,” *Soil Dynamics and Earthquake Engineering*, vol. 31, no. 4, pp. 682–691, 2011.

[11] H. Yu, W. Mao, and M. Huang, “Triaxial tests on the ﬂuidic behavior of post-liquefaction sand,” *Environmental Earth Sciences*, vol. 67, no. 8, pp. 2325–2330, 2012.

[12] H. Qing, L. Heng, and Y. Zhang, “Liquefacation evaluation based on dynamic triaxial test and soil seismic response,” *Journal of Geodesy & Geodynamics*, vol. 242017, in Chinese.

[13] Y. Zhou, D. Zhao, and B. Li, “Fatigue damage mechanism and deformation behaviour of granite under ultrahigh-frequency cyclic loading conditions,” *Rock Mech Rock Eng.*, vol. 54, 2021.

[14] A. I. Lawal, S. Kwon, O. S. Hammed, and M. Adebayo Idris, “Blast-induced ground vibration prediction in granite quarries: an application of gene expression programming, ANFIS, and sine cosine algorithm optimized ANN,” *International Journal of Mining Science and Technology*, vol. 31, no. 2, pp. 265–277, 2021.

[15] V. A. Temeng, Y. Y. Ziggah, and C. K. Arthur, “A novel artifical intelligent model for predicting air overpressure using brain inspired emotional neural network,” *International Journal of Mining Science and Technology*, vol. 30, no. 5, pp. 683–689, 2020.

[16] R. Kumar, P. K. Mandal, A. Narayan, and A. J. Das, “Evaluation of load transfer mechanism under axial loads in a novel coupler of dual height rock bolts,” *International Journal of Mining Science and Technology*, vol. 31, no. 2, pp. 225–232, 2021.

[17] X. Sun, C. Zhao, Y. Zhang, F. Chen, S. Zhang, and K. Zhang, “Physical model test and numerical simulation on the failure mechanism of the roadway in layered soft rocks,” *International Journal of Mining Science and Technology*, vol. 31, no. 2, pp. 291–302, 2021.

[18] L. Adam, “Proactive interburden fracturing using UIS drilling with validation monitoring,” *International Journal of Mining Science and Technology*, vol. 31, no. 1, pp. 3–7, 2021.