

基于无监督方法确定岩土参数取值

阮永芬, 李鹏辉, 朱强, 王勇, 闫明

Determination of geotechnical parameters based on the unsupervised learning method

RUAN Yongfen, LI Penghui, ZHU Qiang, WANG Yong, and YAN Ming

在线阅读 View online: <https://doi.org/10.16030/j.cnki.issn.1000-3665.202207046>

您可能感兴趣的其他文章

Articles you may be interested in

基于主成分分析法的Q2黄土湿陷特性研究

A study of the collapsibility of Q2 loess based on principal component analysis

王玉涛, 刘小平, 曹晓毅 水文地质工程地质. 2020, 47(4): 141-148

基于层次聚类算法的孔压静力触探土体分类方法及试验研究

Soil classification method and experimental research onCPTU based on the hierarchical clustering algorithm

邱敏, 宋友建, 丛璐, 梅年峰, 王闫超 水文地质工程地质. 2019, 46(3): 117-117

基于自组织神经网络的污染场地多监测指标相关性分析

Correlation analysis of multiple monitoring indicators of contaminated site based on self-organizing map

马春龙, 施小清, 许伟伟, 任静华, 王佩, 吴吉春 水文地质工程地质. 2021, 48(3): 191-202

使用长短期记忆神经网络进行花岗岩变形破坏阶段的判别

Determination of granite deformation and failure stages using the long short term memory neural network

陶雪杰, 徐金明, 王树成, 王亚磊 水文地质工程地质. 2021, 48(3): 126-134

双参数强度折减法研究中存在的问题分析

Analyses of the existing problems in the double parameters reduction method

陈子玉, 宋彦辉, 严豪, 陈康达 水文地质工程地质. 2019, 46(2): 125-125

一种改进的地下水模型结构不确定性分析方法

An improved method of groundwater model structural uncertainty analysis

孙晓卓, 曾献奎, 吴吉春, 孙媛媛 水文地质工程地质. 2021, 48(6): 24-33



关注微信公众号, 获得更多资讯信息

DOI: 10.16030/j.cnki.issn.1000-3665.202207046

阮永芬, 李鹏辉, 朱强, 等. 基于无监督方法确定岩土参数取值 [J]. 水文地质工程地质, 2023, 50(4): 149-159.

RUAN Yongfen, LI Penghui, ZHU Qiang, *et al.* Determination of geotechnical parameters based on the unsupervised learning method[J]. Hydrogeology & Engineering Geology, 2023, 50(4): 149-159.

基于无监督方法确定岩土参数取值

阮永芬¹, 李鹏辉¹, 朱强², 王勇³, 闫明⁴

(1. 昆明理工大学建筑工程学院, 云南昆明 650500; 2. 中铁十六局集团有限公司北京轨道交通建设工程有限公司, 北京 101100; 3. 昆明军龙岩土工程有限公司, 云南昆明 650214; 4. 中铁二十局集团第五工程有限公司, 云南昆明 650000)

摘要: 随着城市工程建设的发展, 建筑工程事故问题愈发突出, 采用传统方法求取的岩土参数区间无法满足实际工程需要。基于无监督学习思想, 选取工程性质较差的泥炭质土, 结合工程经验选用 8 个物理指标作为输入集, 利用主成分分析 (principle components analysis, PCA) 算法实现多样本多参数去耦合的降维处理, 得出各物理指标相关性及其敏感度, 结合其相关性及其敏感度赋予不同埋深泥炭质土物理指标的综合评价。利用 k-means 聚类分析泥炭质土物理指标、综合评价及工程特性之间的关系, 为岩土参数选取提供理论基础。采用监督学习方法——BP 神经网络算法分析无监督结果, 验证 (PCA—k-means) 算法模型的合理性。将通过聚类分析得到的正态样本利用多种截尾法优化, 得到可靠取值区间, 并将取值结果与实际工程取值比较, 验证了该模型工程参数取值的合理性。该算法模型具有较好的工程应用价值, 所得研究结果可为工程勘察、设计、施工参数取值提供参考, 也能为岩土参数取值分析提供新的分析方法。

关键词: 主成分分析; k-means 聚类; BP 神经网络; 截尾法; 岩土参数

中图分类号: TU443

文献标志码: A

文章编号: 1000-3665(2023)04-0149-11

Determination of geotechnical parameters based on the unsupervised learning method

RUAN Yongfen¹, LI Penghui¹, ZHU Qiang², WANG Yong³, YAN Ming⁴

(1. Faculty of Civil Engineering and Mechanics, Kunming University of Science and Technology, Kunming, Yunnan 650500, China; 2. China Railway 16 Bureau Group Metro Engineering Construction Co. Ltd., Beijing 101100, China; 3. Kunming Junlong Geotechnical Engineering Co. Ltd., Kunming, Yunnan 650214, China; 4. No.5 Engineering Corporation Limited of CR20G, Kunming, Yunnan 650000, China)

Abstract: With the development of urban engineering construction, the issue of construction engineering accidents has become more and more prominent. The geotechnical parameter interval obtained by using the traditional methods cannot meet the needs of actual engineering. Based on the idea of unsupervised learning, the peaty soil with the worst engineering properties is considered, and 8 physical indexes are selected as the input set. The principal component analysis (PCA) algorithm is used to realize the dimensionality reduction of multi-sample and multi-parameter decoupling, and the correlation and sensitivity of each physical index is obtained. Combined with

收稿日期: 2022-07-28; 修订日期: 2022-10-20

投稿网址: www.swdzgcdz.com

基金项目: 中铁二十局集团第五工程有限公司科研计划项目 (CR2005-5-JS-2021-009)

第一作者: 阮永芬 (1964-), 女, 博士, 教授, 主要从事岩土工程方面的研究。E-mail: ryy64@163.com

通讯作者: 李鹏辉 (1999-), 男, 硕士研究生, 主要从事岩土工程方面的研究。E-mail: 1017343481@qq.com

its correlation and sensitivity, the comprehensive evaluation value of physical indexes of peat soil with different buried depths is given. The k-means clustering is used to analyze the relationship among physical index, and comprehensive evaluation value and engineering characteristics of peaty soil provide a theoretical basis for the selection of geotechnical parameters. The supervised learning method-BP neural network algorithm is used to analyze the unsupervised results and verify the accuracy of the (PCA—k-means) algorithm model. The normal samples obtained by clustering analysis are optimized by a variety of truncation methods to obtain a reliable value range, and the value results are compared with the actual engineering values to verify the rationality of the model engineering parameters. The algorithm model is of good engineering application value. The research results can provide references for engineering investigation, design and construction parameter values, and also provide a new analysis method for geotechnical parameter value analyses.

Keywords: principal component analysis; k-means clustering; BP neural networks; truncated method; geotechnical parameters

昆明滇池流域广泛分布由泥炭质土、粉土、黏土、粉砂互层组成的特殊湖相沉积软土,这类软土不仅土体类别离散,物理指标也非常离散,导致工程性质较复杂,容易引发工程事故,同时为工程勘察、设计及施工时参数的选取带来很多困难。这类软土中,泥炭质土最为特殊,具有孔隙比大、含水率高、压缩性大等特点,工程性质极差^[1-4]。实际工程中参数取值是否合理直接影响后续工程施工安全。因此,得到合理可靠的岩土参数值成为亟待解决的问题。

为降低泥炭质土给工程带来的施工困难与安全隐患,目前很多学者致力于研究泥炭质土物理力学参数相关性。Kolay 等^[5]对马来西亚 6 个不同地点的重塑泥炭质土进行试验研究,其试验结果与他人研究结果吻合较好,即随着有机质含量的增加,泥炭质土的液限、纤维含量、最优含水率均增加,比重、密度均减少。Price 等^[6]研究了泥炭质土的物理性质,对密度、分解度、持水能力、泥炭土纤维含量及种类与压缩性之间的关系进行分析。Santagata 等^[7]对有机质含量达 40%~60%的原状土、重塑泥炭质土及人造无机黏土进行恒载及增量加载一维固结试验,分析研究了压缩系数、渗透系数与孔隙比的变化规律以及土的结构性。裴利华等^[8]研究了有机质含量对无定形泥炭质土物理力学性质的影响,为工程应用提供理论基础。

随着岩土参数相关性研究内容的深入,越来越多的学者采用一元及多元线性方程对岩土参数进行拟合,为设计参数的合理取值提供参考^[9-11]。归纳现有研究成果可知,通过线性拟合得到的岩土参数值考虑的影响因素较少,不能真实反映工程状况。机器学习算法可快速高效处理多参数耦合问题,但是研究发现监督学习严重依赖外部因素,难以处理噪声样本,需

排除噪音点后使用,导致工程适用性不强^[12-13]。无监督学习作为解决分类及回归问题的常用方法之一,目前已被多次应用到岩土工程。例如,阮永芬等^[14]采用无监督学习理论中的主成分分析(principle components analysis, PCA)算法和模糊 C-均值聚类算法(fuzzy C-means algorithm, FCM),即 PCA-FCM 算法,对盾构施工地表沉降进行预估,很大程度上提高了预估正确率。宫凤强等^[15]采用主成分分析与距离分析识别法识别矿山水源,极大提高了判别正确率。任玉晓^[16]采用无监督学习理论对隧道地震波速进行反演,为地震波速深度学习反演在实际数据中的应用提供了可行手段。但是,无监督学习理论在岩土参数方面应用较少且不系统,且计算时未能将多指标样本联系在一起,未考虑多参数指标对土的工程性质影响^[17-18]。

为解决以上问题,采用昆明会展中心场地地质勘测资料,将不同埋深下的泥炭质土 8 个物理指标 381 组数据引入无监督学习理论中,利用 PCA 算法和 k-means 聚类分析将物理指标及工程特性联系起来,为工程参数取值提供参考。将无监督学习理论的结果与 BP 神经网络得到的结果进行分析,进一步验证建立模型的适用性;并对优化后样本指标采用多种截尾法确定其样本取值区间,与实际工程综合建议值比较分析,检验 PCA—k-means 模型的合理性。

1 算法原理

1.1 主成分分析(PCA)

PCA 算法的目标是通过某种线性投影,将高维的数据映射到低维的空间中,并期望在所投影的维度上数据的信息量最大,以便使用较少的数据维度,同时保留住较多的原数据点特性^[19]。采用文献 [14] 中

PCA-FCM模型中的相关PCA算法实现数据降维,其基本步骤如下所述。

为消除不同指标参数量纲的影响,首先对指标参数进行Z-score标准化处理,计算如下:

$$x_{ij} = \frac{x'_{ij} - \bar{x}_j}{\sigma_j}, i = 1, 2, \dots, n, j = 1, 2, \dots, m \quad (1)$$

式中: x'_{ij} ——第*i*个样本第*j*个指标的实测值;

\bar{x}_j 、 σ_j ——第*j*个指标的均值与标准差;

n——样本个数;

m——指标空间维度。

对样本指标标准化后,构建协方差矩阵*C*如下:

$$C = \begin{bmatrix} \text{Cov}(x_1, x_1) & \cdots & \text{Cov}(x_1, x_m) \\ \text{Cov}(x_2, x_1) & \cdots & \text{Cov}(x_2, x_m) \\ \vdots & \vdots & \vdots \\ \text{Cov}(x_m, x_1) & \cdots & \text{Cov}(x_m, x_m) \end{bmatrix} \quad (2)$$

式中: $\text{Cov}(x_j, x_{j'})$ ——第*j*个与第*j'*个指标间的协方差。

协方差的计算公式见式(3)(4),其大小决定了不同指标间的关系:当其值>0时,两样本间为正相关;当其值<0时,两样本间为负相关。

$$\text{Cov}(x_j, x_j) = 1, j = 1, 2, \dots, m \quad (3)$$

$$\text{Cov}(x_j, x_{j'}) = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j) - (x_{ij'} - \bar{x}_{j'})}{n-1} \quad (4)$$

计算矩阵*C*的特征值(λ_j)和其相对应的特征向量(μ_j),可得 $C\mu_j = \lambda_j\mu_j$ 。将 λ_j 从大到小排列,计算指标贡献率(c_j)以及累计贡献率(c_k)。为保证前*k*个主成分信息能代表原始变量的大部分信息,通常取前*k*个主成分使得 c_k 达到85%以上。当 c_k 达到85%时,信息误差在允许范围之内。因此,取 $c_k > 85\%$ 时的前*k*项代表原表征指标样本。 c_j 和 c_k 计算公式如下:

$$c_k = \sum_{j=1}^k c_j, c_j = \frac{\lambda_j}{\sum_{j=1}^m \lambda_j} \times 100\% \quad (5)$$

对样本指标降维后形成新的坐标空间,由选定的前*k*个指标 y_k 组成的矩阵*Y*可以表示为:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} = PX = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{km} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \quad (6)$$

式中:*P*——主成分得分系数矩阵;

X——*m*维列向量。

1.2 k-means 聚类分析

k-means聚类主要用于解决未知量的分类问题,

目前已被用于解决多个领域分类问题。其算法基本原理为随机确定*k*个初始点作为簇质心,然后计算数据样本中的每个点与每个簇质心间的距离,依据此距离对样本进行分配;之后将每簇的质心更改为该簇内所有点的平均值;经过多次迭代计算,簇质心最终趋于稳定^[20-21]。k-means聚类的优点为运算速度快,操作简单,容易理解。算法的基本步骤如下:

(1)从*n*个样本($x_1, x_2, x_3, \dots, x_n$)中随机选取*k*个初始聚类中心 $C_i (i = 1, 2, \dots, k)$ 。

(2)采用欧式距离公式分别计算其余样本 x_j 与各个聚类中心 C_i 的距离,将样本对象分配至聚类最近的簇中。欧式距离公式如下:

$$d(x_j, C_i) = \left[\sum_{p=1}^m (x_{jp} - C_{ip})^2 \right]^{1/2} \quad (7)$$

式中: x_{jp} 、 C_{ip} —— x_j 、 C_i 的第*p*个属性值。

(3)所有样本分配完成后,观察是否会发生点重新分配的现象,若没有则样本分配稳定聚类结束,否则重新计算新的簇中心。

(4)当聚类中心稳定不再变化时停止,并输出聚类结果。

1.3 BP神经网络

BP神经网络是一种多层的前馈神经网络,是监督学习的常用方法之一。其过程主要分为两个阶段,第一阶段是信号的前向传播,从输入层经过隐含层,最后到达输出层;第二阶段是误差的反向传播,从输出到隐含层,最后到输入层,然后依次调节隐含层到输出层、输入层到隐含层的权重和偏置。结合文献[22-23]中的BP神经网络算法,对数据进行预测分析。算法基本步骤如下:

(1)网络初始化,选取合适的激活函数*f(x)*,常用的激活函数为双曲正切“S”型函数:

$$f(x) = \frac{1 - e^{-x}}{1 + e^{-x}} \quad (8)$$

(2)对隐含层输出(H_j)进行计算:

$$H_j = f\left(\sum_{i=1}^n w_{ij}x_i + a_j\right), j = 1, 2, \dots, l \quad (9)$$

式中: w_{ij} ——输入层到隐含层的权重;

a_j ——输入层到隐含层的偏置。

l——隐含层的节点数。

(3)对输出层输出(O_k)进行计算:

$$O_k = \sum_{j=1}^m w_{jk}H_j + b_k, k = 1, 2, \dots, l_1 \quad (10)$$

式中: w_{jk} ——隐含层的输出层的权重;

b_k ——隐含层到输出层的偏置;

l_1 ——输出层的节点数。

(4) 计算其误差 (E):

$$E = \frac{1}{2} \sum_{k=1}^{l_1} (Y_k - O_k)^2 \quad (11)$$

式中: Y_k ——期望输出, 记 $Y_k - O_k = e_k$, 则

$$E = \frac{1}{2} \sum_{k=1}^{l_1} e_k^2 \quad (12)$$

(5) 权值更新, 计算如下:

$$w_{ij} = w_{ij} + \eta H_j (1 - H_j) x_i \sum_{k=1}^{l_1} w_{jk} e_k \quad (13)$$

$$w_{jk} = w_{jk} + \eta H_j e_k \quad (14)$$

(6) 偏置更新

隐含层到输出层的偏置更新:

$$a_j = a_j + \eta H_j (1 - H_j) x_i \sum_{k=1}^{l_1} w_{jk} e_k \quad (15)$$

$$b_k = b_k + \eta e_k \quad (16)$$

输入层到隐含层的偏置更新:

$$b_k = b_k + \eta H_j (1 - H_j) \sum_{k=1}^{l_1} w_{jk} e_k \quad (17)$$

计算结束后, 判断算法是否收敛。若结果收敛, 则算法结束; 否则重复上述计算操作, 直到结果收敛。

2 案例分析

2.1 参数指标确定

对大量地勘报告进行整理和分析, 选取地层分布具有代表性的昆明会展中心场地进行研究。此场地属于滇池盆地湖相沉积平原地貌, 在勘测范围内, 地层主要由表层填土、第四系冲洪积及湖相沉积层组成, 共有 4 层泥炭质土。地层构造如图 1 所示。分别对各层不同深度泥炭质土取样, 选取 381 个样本进行研究。选择密度 (ρ)、比重 (G_s)、含水率 (ω)、液限 (ω_L)、塑限 (ω_p)、孔隙比 (e)、压缩系数 ($a_{1,2}$)、有机质含量 (w_u) 等 8 个物理指标进行分析。每层泥炭质土的物理指标参数如表 1 所示。

2.2 参数相关性分析

根据式 (1) 对泥炭质土样本指标进行标准化处理后进行 KMO 和巴利特球形检验, 得到 KMO 取样适切性量数为 0.897 (>0.50), 显著性概率为 0.000 (<0.05),

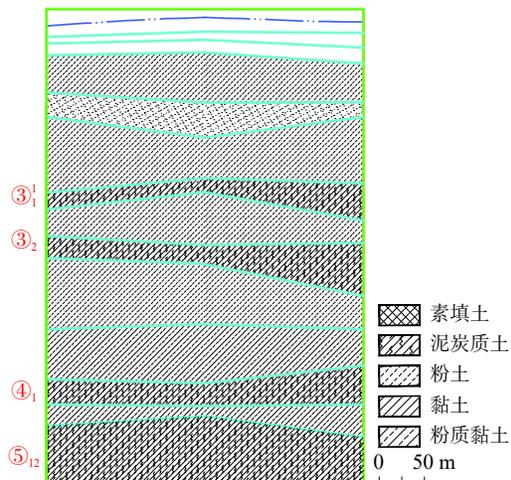


图 1 会展中心地层构造图

Fig. 1 Stratigraphic diagram of the convention center

表 1 各层泥炭质土的物理指标统计表

Table 1 Statistical table of physical indexes of peaty soil in each layer

地层编号	ρ /($\text{g}\cdot\text{cm}^{-3}$)	G_s	ω /%	ω_L /%	ω_p /%	e	$a_{1,2}$ /MPa $^{-1}$	w_u /%
③ ₁	1.61	2.30	110.1	135.4	100.0	2.49	1.62	24.33
③ ₂	1.42	2.42	101.2	119.2	87.2	2.40	1.32	18.36
④ ₁	1.25	1.95	145.7	182.1	146.4	2.93	1.49	40.10
⑤ ₁₂	1.27	1.97	130.6	174.3	140.6	2.52	1.12	38.77

满足主成分分析法的基本要求。通过式 (2)、(3)、(4) 得到样本指标的协方差矩阵 C , 将 C 中元素平方可得到指标间的相关性系数 (R^2), R^2 随指标间相关性增高而增大。将各指标间 R^2 统计于表 2, 并通过式 (18) 对不同指标进行线性拟合。

$$R^2 = \frac{\sum_{i=1}^n (Y_i - \bar{y})^2}{\sum_{i=1}^n (Y_i - \bar{y})^2 + \sum_{i=1}^n (Y_i - y_i)^2} \quad (18)$$

式中: y_i 、 \bar{y} ——实测数据值及均值;

Y_i ——拟合曲线的计算值。

以泥炭质土的 ρ 和 w_u 为例, 分析 ω 、 ω_p 与 $a_{1,2}$ 间的线性拟合关系, 结果如图 2、图 3 所示。 ρ 与 ω 、 ω_p 、 e 的线性关系明显, 相关系数均大于 0.8, 与 $a_{1,2}$ 的线性关系较弱, 相关系数为 0.62。 w_u 与 ρ 、 ω 、 ω_L 的相关系数均大于 0.8, 线性关系较明显。结合表 2 可知, ρ 与除 G_s 、 $a_{1,2}$ 之外的其余几个物理指标相关性较高, 可通过线性关系表达。 G_s 与 $a_{1,2}$ 相关性最低, 相关系数仅为 0.42。 ρ 和 G_s 与其他物理指标为负相关, 其余物理指标之间为正相关。同理, 其余物理指标相关性关系均可通过表 2 得到。综上可知, 该场地泥炭质土物理

表 2 不同物理指标间的相关性系数

Table 2 Correlation between different physical indicators

物理指标	ρ	G_s	ω	ω_L	e	a_{1-2}	ω_p	w_u
ρ	1.00							
G_s	0.64	1.00						
w	0.87	0.55	1.00					
w_L	0.89	0.59	0.94	1.00				
e	0.85	0.85	0.94	0.87	1.00			
a_{1-2}	0.62	0.42	0.81	0.68	0.85	1.00		
w_p	0.85	0.58	0.82	0.93	0.58	0.49	1.00	
w_u	0.83	0.66	0.73	0.78	0.70	0.45	0.80	1.00

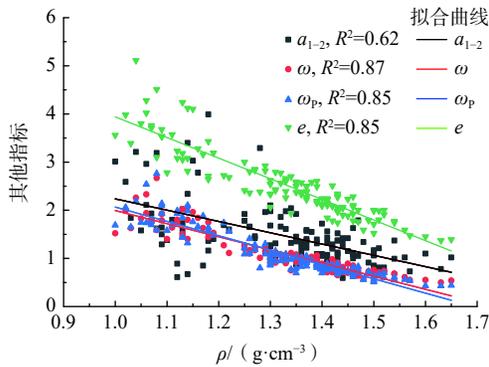


图 2 ρ 与 ω 、 ω_L 、 e 、 a_{1-2} 关系曲线

Fig. 2 Relationship of ρ and ω , ω_L , e and a_{1-2}

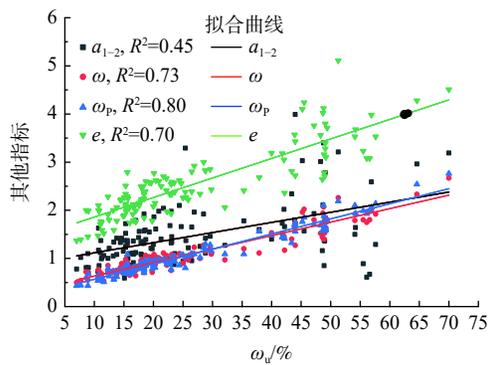


图 3 w_u 与 ω 、 ω_L 、 e 、 a_{1-2} 关系曲线

Fig. 3 Relationship of w_u and ω , ω_L , e and a_{1-2}

指标之间存在一定的相关性规律。

2.3 因子分析及敏感性分析

计算协方差矩阵 C 的特征值和特征向量, 通过式 (5)(6) 计算主成分的贡献率、累计贡献率和得分系数矩阵。为确定主成分 k 值个数, 观察特征值、贡献率及累计贡献率随 k 值变化情况, 如图 4、图 5 所示。由图可知, 当 $k > 3$ 时, 曲线趋于平稳, 且累计贡献率大于 85%, 故取前 3 个主成分为降维后的样本指标, 代表原始样本数据的全部信息参与计算。利用式 (19) 计算各指标的敏感度 (r_j), 将 r_j 统计到表 3。由表 3 可

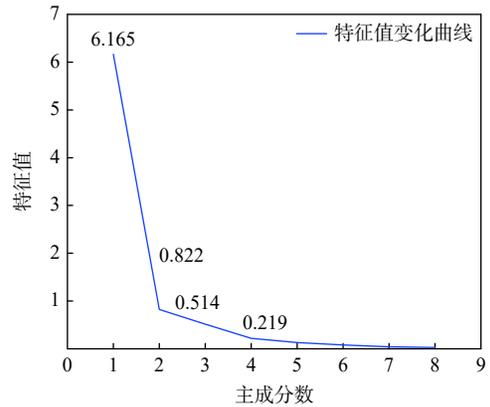


图 4 特征值变化规律

Fig. 4 Variation of eigenvalues

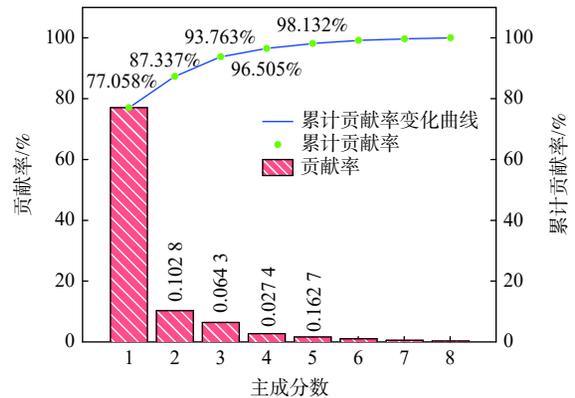


图 5 累计贡献率

Fig. 5 Cumulative contribution

知, 按 r_j 由高到低排序, 各物理指标依次为: e 、 ω 、 ω_L 、 ω_p 、 ρ 、 w_u 、 a_{1-2} 、 G_s , 当研究泥炭质土的物理指标变化规律时, 可优先考虑分析 e 、 ω 与 ω_L 的变化规律。

$$r_j = \frac{\sum_{c=1}^k \lambda_c |p_{jc}|}{\sum_{j=1}^m \sum_{c=1}^k \lambda_c |p_{jc}|} \quad (19)$$

为得到不同埋深泥炭质土的综合评价价值, 可利用主成分得分系数矩阵 P 采用回归法分别计算 3 个主成分的因子得分函数 F_1 、 F_2 、 F_3 , 计算公式如式 (20) 所示。将因子得分函数 F_1 、 F_2 、 F_3 代入式 (21) 计算不同埋深下的泥炭质土的综合评价价值 (Q)。将不同埋深泥炭质土的综合评价价值由低到高进行排列表 4。由于样本数量较多, 直接分析综合评价价值与物理指标变化规律不方便; 而物理指标之间存在某种线性关系, 故将样本数量缩小, 利于观察其变化趋势, 因此令 10 个样本为 1 组, 计算各组物理指标平均值。计算结果如表 4 所示。

表 3 各物理指标敏感度统计表
Table 3 Sensitivity statistics of each physical index

物理指标编码	$X_1(\rho)$	$X_2(G_s)$	$X_3(\omega)$	$X_4(\omega_L)$	$X_5(e)$	$X_6(a_{1-2})$	$X_7(\omega_p)$	$X_8(w_u)$
敏感度符号	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8
敏感度值	0.234 9	0.206 9	0.254 0	0.242 8	0.260 8	0.215 1	0.241 0	0.228 1

$$F_k = \sum_{j=1}^m (p_{kj}x_j), k = 1, 2, 3 \quad (20)$$

$$Q = (F_1c_1 + F_2c_2 + F_3c_3)/(c_1 + c_2 + c_3) \quad (21)$$

由排序结果可知,随着综合评价值的升高,敏感性较高的 e 、 ω 、 ω_L 与 w_u 整体呈递增的趋势,而 ρ 及 G_s 大体呈递减的趋势。这种现象是由泥炭质土特殊的成分——有机质所引起的。由于有机质内存在腐殖酸,泥炭质土的结合水层较厚,有机质与水膜联合形成架空结构,使泥炭质土具有较高的 e 、 ω 值。由于有机质含量的重度小于黏土颗粒的重度,故随着 w_u 的增多,泥炭质土的 ρ 及 G_s 逐渐较小。

对埋深 20~35 m 与 45~55 m 的泥炭质土与其评价值进行分析,发现上层泥炭质土各物理指标稳定,而下层泥炭质土物理指标参数离散性较大。这是由

于随着埋深的增加,泥炭质土发生煤化作用,上覆沉积物产生静压力排出大量孔隙水,泥炭质土转化为褐煤,导致参数空间变异性增大,故离散性增强^[1]。

泥炭质土物理指标取值除受深度影响外,还受其土层厚度、地理位置、气候气象、地形地貌及水文地质条件的影响,这导致我国不同地区的泥炭质土物理指标值有所差异。但这些地区的泥炭质土工程性质均较差,具有较高的触变性、流变性、低透水性、高压缩性及低强度等特点。建筑物基础容易产生地面不均匀沉降,进而影响工程施工进度。而泥炭质土不良的工程性质与物理指标具有一定的相关性,高含水率、高压缩系数、高孔隙比的泥炭质土工程性质较差。由表 4 可知,由多个物理指标得到的综合评价价值与其工程性质具有相关性。

表 4 各组泥炭质土综合评价排序
Table 4 Ranking of comprehensive evaluation of the peaty soil

组类	综合评价价值	物理指标平均值							
		$\rho/(g \cdot cm^{-3})$	$\omega/\%$	G_s	e	$\omega_L/\%$	$\omega_p/\%$	a_{1-2}/MPa^{-1}	$w_u/\%$
1	-1.295	1.565	57.590	2.517	1.536	68.690	42.690	0.722	8.100
2	-1.162	1.533	67.980	2.397	1.565	71.730	44.910	0.812	12.670
3	-1.096	1.499	68.280	2.428	1.699	76.180	48.270	1.103	11.690
4	-1.035	1.483	72.120	2.436	1.824	84.800	49.560	0.977	12.270
5	-0.928	1.452	73.690	2.238	1.882	92.400	58.890	0.850	13.678
6	-0.946	1.454	78.070	2.426	1.971	90.320	53.550	1.121	12.100
7	-0.900	1.455	82.180	2.426	1.957	95.363	62.181	0.833	13.318
...
34	1.293	1.115	211.200	1.796	4.406	228.270	149.400	3.802	43.750
35	1.496	1.134	220.200	1.164	4.761	228.710	144.100	4.718	44.346
36	1.566	1.117	232.700	1.311	4.818	238.770	149.310	5.031	40.817
37	1.861	1.121	247.000	0.996	5.327	249.400	156.300	6.006	41.710
38	2.574	1.079	287.182	0.990	5.839	304.936	207.800	5.978	56.455

3 结果检验及参数分析

3.1 聚类结果分析

为验证上述结论,采用聚类分析法分析其物理指标值与其综合评价价值关系。将 PCA 转化后的主成分数据作为分级判据指标输入到 k-means 聚类算法时,需考虑聚类个数 k 的影响, k 直接影响聚类效果和主要信息完整性。对分判指标采用不同 k 值进行分析,

随着 k 的增加,会有一类样本基本不会发生变化,因此可将这类样本认定为最优子类。当 $k=3$ 时,样本分类趋于稳定,分类结果明显,且样本点较为集中。故选取 $k=3$ 为最优分类数。分类结果如图 6 所示。

确定好聚类数 $k=3$ 后,对输出结果施加外部标签(综合评价价值),按外部标签由小到大将输出类数进行规范化排列,结果如图 7 所示。结合表 4 及图 7 可知,分类结果与外部标签之间相关性较优,外部标签虽存

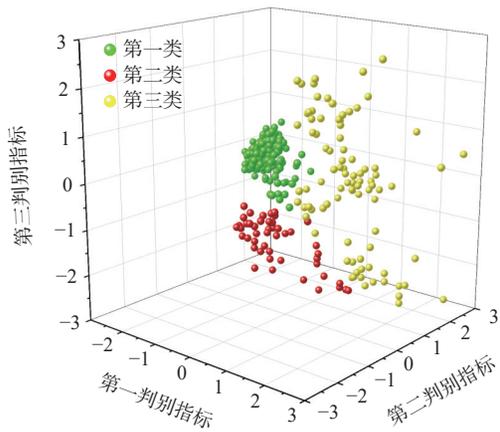


图 6 判别指标分类结果

Fig. 6 Discriminative index classification results

在混淆区但混淆样本较少, 存在混淆样本点的原因是岩土参数变异性、数据本身存在一定误差。大部分样本点集中在第一类, 保留了原样本的主要信息, 综合评价价值较低。

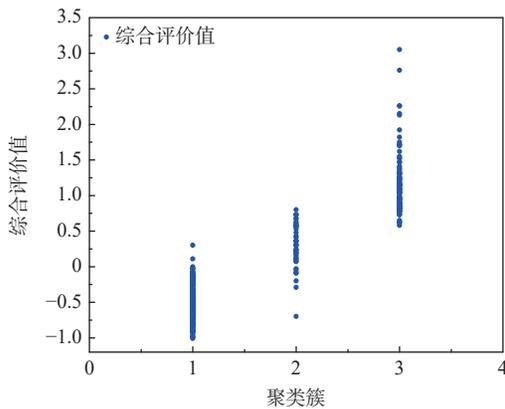


图 7 综合评价分类结果

Fig. 7 Results of comprehensive evaluation

为验证综合评价价值与物理指标值之间的关系, 选择相关性较高且敏感度较高的 e 、 ω 与 ω_1 3 个指标分析 $k=3$ 时物理指标的变化规律。结果如图 8 所示。将图 7、图 8 联系起来分析可发现, 随着综合评价价值升高, e 、 ω 与 ω_1 基本呈线性增长趋势, 且在空间分类界限较为明显。同理, 由表 2 中不同物理指标相关性, 可判断剩余物理指标的变化规律。综上可知, 综合评价价值与物理指标值具有一定的相关性, 综合评价价值越大, 泥炭质土的工程性质越差, 引起工程事故的可能性更高。结合勘测报告所给的会展中心地基土的物理力学指标评价可知, 下层泥炭质土指标离散性较大, 孔隙比大, 压缩性高, 不宜作为桩端持力层使用。由 PCA 法计算的这类土层综合评价价值较高(工程特性

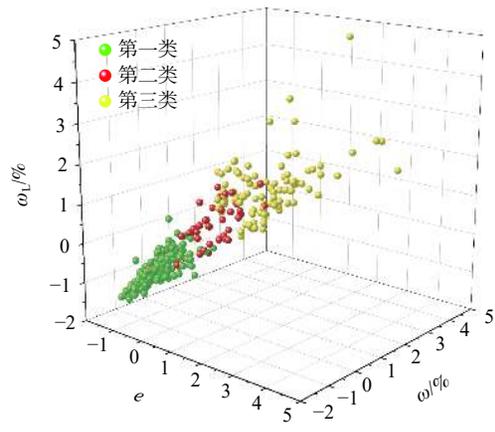


图 8 e 、 ω 与 ω_1 分类结果

Fig. 8 Classification results of e , ω and ω_1

差), 验证了本文综合评价价值的工程应用可行性。

3.2 BP 神经网络验证

为验证综合评价价值理论上的有效性, 将无监督学习结果输入到 BP 神经网络分析法进行分析。通过与新的数据使用方法得到的结果比较, 验证无监督理论的可行性。为保证模型结果的准确性, 将样本分为训练与检验两部分, 训练样本占 70%, 检验样本占 30%。当两部分模型误差均小于 0.05 时, 认为此模型可进行神经网络分析。分析结果如图 9、图 10 所示。

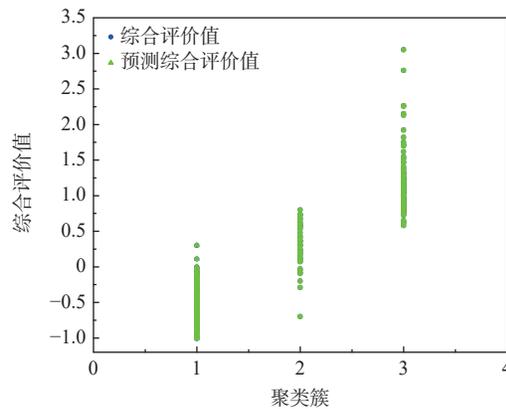


图 9 分类结果对比

Fig. 9 Comparison of classification results

由图 9、图 10 可知, 无监督学习的综合评价价值分类结果与 BP 神经网络得到的预测分类结果基本一致, 原综合评价价值与回归得到的预测综合评价曲线基本拟合。通过 2 种方法的对比, 证明了综合评价价值的正确性。监督学习理论目标明确且效果容易评估, 通过采用监督学习方法评估无监督学习的结果为一种可靠方法。监督学习方法可以从多个方面验证无监督学习的准确性, 增强理论依据, 并为无监督学习理

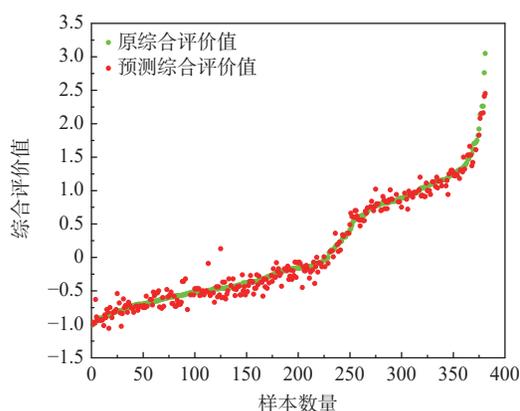


图 10 综合评价对比

Fig. 10 Comprehensive evaluation and comparison

论应用到岩土工程提供理论基础。

3.3 参数分布分析

在工程中岩土参数的选取会直接影响到工程的

安全性,故参数既不能过大也不能过小,应处于合适的范围之内。为解决这个问题,宫凤强等^[24]针对不同的样本分布提出了相应的区间长度原则。随着参数样本数量的增加,岩土参数逐渐服从某种分布。而滇池流域泥炭质土物理指标离散性较大,很大程度上不满足正态及对数正态分布,导致无法得到可靠的参数区间范围。为保证工程设计、施工、勘察的安全性,应尽可能合理缩小物理力学指标取值范围,得到可靠安全的参数取值。

以 e 、 w 与 w_L 3 个指标为例分析其样本分布,如图 11 所示。分别采用正态分布与对数正态分布曲线拟合 3 个指标的样本分布,对比可知,3 个物理指标均不满足这 2 种分布,分别采用不同截尾法,即 3σ (标准差)、 c_{33} (参考偏度双侧调整)、 c_3 (参考偏度单侧调整)原则计算参数区间范围,结果如表 5 所示。

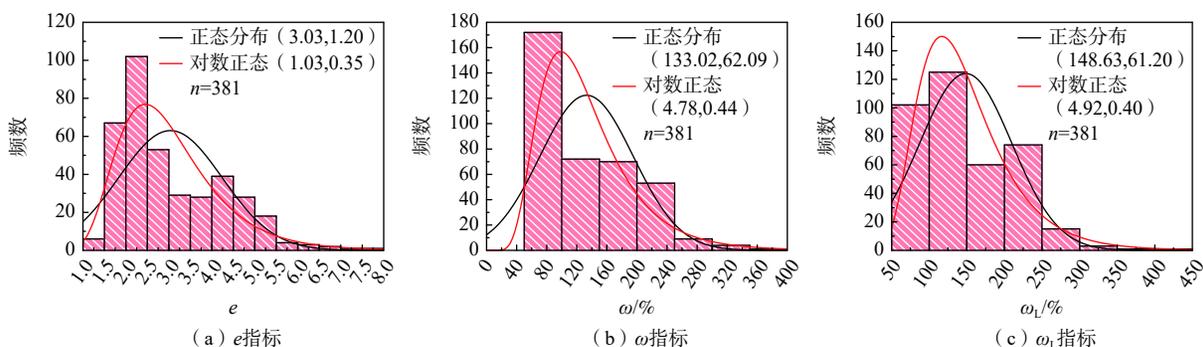


图 11 3 个物理指标原样本分布图

Fig. 11 Original sample distribution map of three physical indicators

表 5 不同截尾法的原物理指标区间范围

Table 5 Original physical index range of different censoring methods

物理指标	均值	标准差	偏度系数	变异系数	3σ 区间范围	c_{33} 区间范围	c_3 区间范围
e	3.03	1.20	0.94	0.39	[0,6.63]	[0.56,7.76]	[0, 7.76]
$w/\%$	133.02	62.09	0.93	0.46	[0,319.29]	[4.48,377.03]	[0,377.03]
$w_L/\%$	148.63	61.26	0.87	0.41	[0,332.41]	[18.14,385.71]	[0,385.71]

由表 5 可知,由多种截尾法得到的 3 个物理指标的参数区间范围较大,导致在实际工程中参数取值不可靠,无法较好应用到工程中。针对这个问题,将综合评价应用到参数选取中,对保留样本主要信息及综合评价较低(工程特性较好)的第一类样本指标进行分析,结果如图 12 所示。

由图 12 可知,经过截尾法优化后,3 个物理指标参数区间范围均在原样本数据区间基础上进一步集中,样本正态分布曲线与对数正态分布曲线拟合度提高。为检验样本分布合理性,采用 K-S 检验法对图 12

进行检验并与原样本的正态检验结果对比,结果见表 6。由表 6 可知,聚类前 3 个指标 P 值均小于 0.05,但聚类后 3 种指标远大于 0.05,且 P 值提高了很多数量级,分类结果显著,聚类后 3 个指标参数样本分布为正态分布可接受。采用 3σ 、 c_{33} 、 c_3 原则计算聚类后物理指标参数区间范围,同理其余物理指标均可按相同方法分析,结果如表 7 所示。

由表 7 可知,聚类后的新样本保留了原样本主要信息,且正态性有显著提高。由不同截尾法求得的区间范围较之前有很大程度的缩小,取值更加方便,与

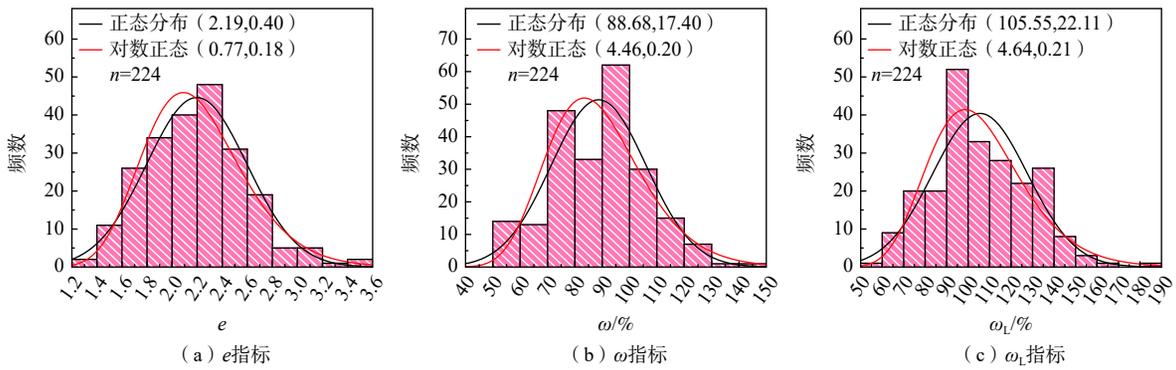


图 12 分类后 3 个物理指标样本分布图

Fig. 12 Sample distribution map of three physical indicators after classification

表 6 物理指标样本检验结果

Table 6 Normal test result of the physical index sample

指标	分布类型	D值	P值	检验结果
e (图11)	正态分布	0.155	1.81×10^{-4}	排除
e (图12)	正态分布	0.090	1.00	接受
ω (图11)	正态分布	0.189	2.36×10^{-12}	排除
ω (图12)	正态分布	0.064	0.710	接受
ω_1 (图11)	正态分布	0.146	1.62×10^{-7}	排除
ω_1 (图12)	正态分布	0.072	0.190	接受

注: D值为检验统计量; P值为渐近显著性水平, P值大于0.05时, 接受原假设。

工程实用值更加接近。综上可知, 通过 PCA—k-means 法将物理指标降维, 赋予物理指标一个外部标签(综合评价), 综合评价可衡量泥炭质土工程特性的好坏。工程特性较差的泥炭质土指标离散性较大, 样本分布不规律, 导致物理指标区间范围较大, 取值不方便, 在实际工程中不能较好应用。通过聚类分析得到综合评价值较低的样本, 这类样本不仅数量多, 保留了原样本的主要信息, 样本分布较规律, 得到的区间范围更加可靠, 方便工程取值。

表 7 不同截尾法的新样本区间范围

Table 7 Range of new sample intervals for different censoring methods

物理指标	均值	标准差	偏度系数	变异系数	3 σ 区间范围	c_{33} 区间范围	c_3 区间范围
e	2.19	0.40	0.39	0.18	[0.99, 3.39]	[1.45, 3.55]	[0.99, 3.55]
ω /%	88.68	17.40	0.39	0.18	[36.48, 140.88]	[43.26, 147.66]	[36.48, 140.88]
ω_1 /%	105.55	22.11	0.28	0.20	[39.22, 171.88]	[45.41, 178.07]	[39.22, 178.08]
ρ ($\text{g} \cdot \text{cm}^{-3}$)	1.40	0.08	0.08	0.06	[1.13, 1.66]	[1.15, 1.67]	[1.13, 1.67]
G_s	2.25	0.45	-3.08	0.20	[0.89, 3.61]	[1.15, 3.55]	[1.15, 3.61]
a_{1-2} /MPa $^{-1}$	1.24	0.56	1.56	0.45	[0, 2.93]	[0.43, 3.81]	[0, 2.93]
ω_w /%	68.76	18.06	0.38	0.26	[14.57, 122.97]	[21.48, 129.88]	[14.57, 129.88]
w_w /%	17.63	7.64	2.02	0.43	[0, 40.57]	[10.17, 56.06]	[0, 56.06]

在实际工程中, 为保证施工安全, 可以通过多种原位试验, 得到各土层物理力学指标设计参数综合值。现对会展中心 4 层泥炭质土的物理指标设计参数进行分析, 并利用本文方法分析其合理性, 结果如表 8 所示。4 层泥炭质土物理指标设计参数综合值与表 7 所得各指标参数值相差不大, 根据各层物理指标综合值得到的综合评价属第一类, 说明其参数取值较合理, 证明通过 PCA—k-means 模型得到的物理指标值是合理性的。

4 结论

(1) 通过协方差矩阵研究多个物理指标可知, 不

同物理指标间相关性满足一定的规律, 压缩系数与其他几个物理指标相关性较低, 含水率、液限和孔隙比与其他物理指标相关系数较高。

(2) 利用主成分分析法对多个物理指标参数降维

表 8 各层泥炭质土设计参数检验结果

Table 8 Test results of design parameters of the peaty soil in each layer

地层编号	ρ /($\text{g} \cdot \text{cm}^{-3}$)	ω /%	e	I_L	a_{1-2} /MPa $^{-1}$	综合评价	所属类
③ $_1^1$	1.37	103.0	2.41	0.15	1.63	-0.23	第一类
③ $_2$	1.39	91.1	2.33	0.12	1.30	-0.26	第一类
④ $_1$	1.20	142.3	2.92	-0.14	1.47	-0.08	第一类
⑤ $_{12}$	1.26	127.7	2.53	-0.28	1.12	-0.16	第一类

分析,减小时间与人工成本。利用各指标敏感度及相关性赋予泥炭质土样本综合评价价值,利用综合评价价值衡量工程特性,可为工程参数选取提供理论基础。

(3)采用监督学习——BP神经网络方法验证PCA—k-means模型得到的综合评价价值及分类结果,对比可知2种方法所得结果基本一致。采用监督学习方法多个方面验证PCA—k-means模型准确性,可以增强理论依据,并且为无监督学习理论应用到岩土工程提供理论基础。

(4)针对离散性较大的泥炭质土样本参数困难这个问题,利用k-means聚类选取综合评价价值较低的一类样本,将不规则样本转化为规律分布样本,样本正态性增加,进一步缩小参数取值区间,并与工程设计参数综合值比较分析,验证其参数取值合理性,可为工程参数取值提供便利。

参考文献 (References) :

- [1] 阮永芬,刘岳东,王东,等.昆明泥炭与泥炭质土对建筑地基的影响[J].昆明理工大学学报(理工版),2003,28(3):121-124. [RUAN Yongfen, LIU Yuedong, WANG Dong, et al. Effect of Kunming's peat & peaty soil on the building foundation[J]. Journal of Kunming University of Science and Technology, 2003, 28(3): 121-124. (in Chinese with English abstract)]
- [2] 柴春阳,张广泽,胡婷.丽香铁路冰水沉积型泥炭土特性及工程对策分析[J].地下空间与工程学报,2017,13(增刊2):757-761. [CHAI Chunyang, ZHANG Guangze, HU Ting. Characteristics of ice-water deposition peat in Li-Xiang railway and analysis on engineering countermeasures[J]. Chinese Journal of Underground Space and Engineering, 2017, 13(Sup 2): 757-761. (in Chinese with English abstract)]
- [3] 李强,李刚.沪苏浙高速公路泥炭质土的工程性质研究[J].公路工程,2010,35(6):99-103. [LI Qiang, LI Gang. Research on the engineering property of peat quality soil in Hu-Su-Zhe highway[J]. Highway Engineering, 2010, 35(6): 99-103. (in Chinese with English abstract)]
- [4] 谢俊文.泥炭的工程地质特性[J].工程勘察,1981,9(5):31-34. [XIE Junwen. Engineering geological characteristics of peat[J]. Geotechnical Investigation & Surveying, 1981, 9(5): 31-34. (in Chinese with English abstract)]
- [5] KOLAY P K, AMINUR M R, TAIB S N L, et al. Correlation between different physical and engineering properties of tropical peat soils from Sarawak[C]//Soil Behavior and Geo-Micromechanics. Reston, VA: American Society of Civil Engineers, 2010: 56-61.
- [6] PRICE J S, CAGAMPAN J, KELLNER E. Assessment of peat compressibility: Is there an easy way?[J]. Hydrological Processes, 2005, 19(17): 3469-3475.
- [7] SANTAGATA M, BOBET A, JOHNSTON C T, et al. One-dimensional compression behavior of a soil with high organic matter content[J]. Journal of Geotechnical and Geoenvironmental Engineering, 2008, 134(1): 1-13.
- [8] 裴利华,杨醒宇,桂跃,等.有机质含量及组分对泥炭土物理力学性质影响[J].水文地质工程地质,2022,49(2):77-85. [PEI Lihua, YANG Xingyu, GUI Yue, et al. Influence of organic matter content and ingredient on the physical and mechanical properties of peat soils[J]. Hydrogeology & Engineering Geology, 2022, 49(2): 77-85. (in Chinese with English abstract)]
- [9] 严春风,刘东燕,朱可善.岩石抗剪强度指标互相关性的敏感度分析[J].水文地质工程地质,1997,24(6):45-47. [YAN Chunfeng, LIU Dongyan, ZHU Keshan. Susceptibility analysis to the correlativity of shear strength indices of ricks[J]. Hydrogeology & Engineering Geology, 1997, 24(6): 45-47. (in Chinese with English abstract)]
- [10] 程豪,田林.昆明市某场地不同埋深下泥炭质土物理力学指标统计规律研究[J].贵州大学学报(自然科学版),2022,39(4):92-98. [CHENG Hao, TIAN Lin. Statistical law of physical and mechanical indexes of peat soil under different burial depths in a site in Kunming[J]. Journal of Guizhou University (Natural Science), 2022, 39(4): 92-98. (in Chinese with English abstract)]
- [11] 丁祖德,付江,李夕松,等.昆明泥炭质土的物理力学指标特征及相关性分析[J].公路工程,2018,43(4):86-91. [DING Zude, FU Jiang, LI Xisong, et al. Properties of physical and mechanical indexes and correlation analysis of peaty soil in Kunming area[J]. Highway Engineering, 2018, 43(4): 86-91. (in Chinese with English abstract)]
- [12] 陈玉萍,袁志强,周博,等.遗传算法优化BP网络在滑坡灾害预测中的应用研究[J].水文地质工程地质,2012,39(1):114-119. [CHEN Yuping, YUAN Zhiqiang, ZHOU Bo, et al. Application of back propagation neural networks with optimization of genetic algorithms to landslide hazard prediction[J]. Hydrogeology & Engineering Geology, 2012, 39(1): 114-119. (in Chinese with English abstract)]
- [13] 马壮壮,束龙仓,季叶飞,等.基于遗传算法的BP神经

- 网络计算岩溶水安全开采量[J]. 水文地质工程地质, 2016, 43(1): 22 - 27. [MA Zhuangzhuang, SHU Longcang, JI Yefei, et al. Calculation of Karst water safe yield by using BP neural network based on genetic algorithm[J]. Hydrogeology & Engineering Geology, 2016, 43(1): 22 - 27. (in Chinese with English abstract)]
- [14] 阮永芬, 张虔, 乔文件, 等. 盾构施工地表沉降的无监督学习预估方法[J/OL]. 安全与环境学报, (2022-05-09)[2022-09-05]. <http://doi.org/10.13637/j.issn.1009-6094.2022.0351>. [RUAN Yongfen, ZHANG Qian, QIAO Wenjian, et al. Unsupervised learning prediction method of ground settlement in shield construction[J/OL]. Journal of Safety and Environment, (2022-05-09)[2022-09-05]. (in Chinese with English abstract)]
- [15] 宫凤强, 鲁金涛. 基于主成分分析与距离判别分析法的突水水源识别方法[J]. 采矿与安全工程学报, 2014, 31(2): 236 - 242. [GONG Fengqiang, LU Jintao. Recognition method of mine water inrush sources based on the principal element analysis and distance discrimination analysis[J]. Journal of Mining & Safety Engineering, 2014, 31(2): 236 - 242. (in Chinese with English abstract)]
- [16] 任玉晓. 地震波速无监督深度学习反演方法及其在隧道超前探测中的应用[D]. 济南: 山东大学, 2021. [REN Yuxiao. Unsupervised deep-learning inversion method for seismic velocity and its application in tunnel forward-prospecting[D]. Jinan: Shandong University, 2021. (in Chinese with English abstract)]
- [17] 阮永芬, 魏德永, 高骏, 等. K-means聚类分析高原湖相沉积软土参数[J]. 昆明理工大学学报(自然科学版), 2020, 45(1): 85 - 91. [RUAN Yongfen, WEI Deyong, GAO Jun, et al. K-means cluster analysis of soft soil parameters of plateau sedimentary[J]. Journal of Kunming University of Science and Technology (Natural Science), 2020, 45(1): 85 - 91. (in Chinese with English abstract)]
- [18] 孔令奇, 李翠娟. 岩土参数概率分布的系统推断研究[J]. 工业建筑, 2022, 52(1): 129 - 136. [KONG Lingqi, LI Cuijuan. Study on system inference of probability distribution of geotechnical parameters[J]. Industrial Construction, 2022, 52(1): 129 - 136. (in Chinese with English abstract)]
- [19] 殷瑞刚, 魏帅, 李晗, 等. 深度学习中的无监督学习方法综述[J]. 计算机系统应用, 2016, 25(8): 1 - 7. [YIN Ruigang, WEI Shuai, LI Han, et al. Introduction of unsupervised learning methods in deep learning[J]. Computer Systems & Applications, 2016, 25(8): 1 - 7. (in Chinese with English abstract)]
- [20] 孙吉贵, 刘杰, 赵连宇. 聚类算法研究[J]. 软件学报, 2008, 19(1): 48 - 61. [SUN Jigui, LIU Jie, ZHAO Lianyu. Clustering algorithms research[J]. Journal of Software, 2008, 19(1): 48 - 61. (in Chinese with English abstract)]
- [21] 张麟, 潘红岩. 聚类分析算法应用研究[J]. 数字技术与应用, 2016(10): 143-145. [ZHANG Lin, PAN Hongyan. Research on application of clustering analysis algorithm[J]. Digital Technology and Application, 2016(10): 143 - 145. (in Chinese)]
- [22] 刘荣. 人工神经网络基本原理概述[J]. 计算机产品与流通, 2020(6): 35. [LIU Rong. Overview of basic principles of artificial neural network[J]. Computer Products & Circulation, 2020(6): 35. (in Chinese)]
- [23] 袁冰清, 程功, 郑柳刚. BP神经网络基本原理[J]. 数字通信世界, 2018(8): 28 - 29. [YUAN Bingqing, CHENG Gong, ZHENG Liugang. Basic principle of BP neural networks[J]. Digital Communication World, 2018(8): 28 - 29. (in Chinese with English abstract)]
- [24] 宫凤强, 侯尚骞, 李夕兵. 岩土参数截尾分布的正态信息扩散推断方法[J]. 武汉大学学报(工学版), 2016, 49(5): 661 - 667. [GONG Fengqiang, HOU Shangqian, LI Xibing. Truncated distribution deduction method for geotechnical parameters based on normal information diffusion method[J]. Engineering Journal of Wuhan University, 2016, 49(5): 661 - 667. (in Chinese with English abstract)]

编辑: 刘真真