

坝肩岩体质量 LDA-KNN 分类模型

荀鹏¹, 李娟², 魏玉峰¹, 李常虎³, 范文东³

- 地质灾害防治与地质环境保护国家重点实验室(成都理工大学), 成都 610059;
- 四川省能源地质调查研究所, 成都 610070;
- 中国电建集团西北勘测设计研究院有限公司, 西安 710065

[摘要] 工程岩体质量分级评价对工程的安全、设计、经济效益等有重要影响。针对当前岩体划分方法中存在不确定性, 人为因素干扰和忽视了传统定性分级中对岩体质量评价的重要性等问题, 本文通过在工程实际中搜集样本建立数据库, 从工程的实际需求出发, 选择岩体完整性系数(K_v)、结构面间距(D)、岩石质量指标(RQD)等合适的评价指标, 通过引入 LDA(Linear Discriminant Analysis)降维方法和 K 近邻分析(K-Nearest-Neighbor, KNN)相结合的多分类模型, 实现了岩体的非线性分级预测。通过定性定量相结合实现了岩体多因素, 多指标的综合分级, 并解决了多指标判断时信息冗余, 复杂程度高的问题。与其他判别方案相比较, 模型得出的结果准确率高, 符合工程实际, 减少了人为因素的影响, 体现出较强的预测判别能力。该研究为水电站大坝坝肩处的平硐岩体质量划分提出了一种可行的预测方案。

[关键词] 岩体结构; 岩体质量分级; 线性降维; K 近邻算法; 分类模型

[分类号] TV223.1

[文献标志码] A

The LDA-KNN model to classify quality of rock mass of a dam abutment

XUN Peng¹, LI Juan², WEI Yufeng¹, LI Changhu³, FAN Wendong³

- State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059, China;
- Sichuan Institute of Energetical and Geological Survey, Chengdu 610070, China;
- PowerChina Northwest Engineering Corporation Limited, Xi'an 710065, China

Abstract: The classification and evaluation of the engineering quality of rock mass has an important impact on the safety, design, and economic benefits of construction projects. In this study, we responded to the problems posed by uncertainty, interference by human factors, and the neglect of traditional qualitative grading in prevalent models used to assess the quality of the rock mass. We

[收稿日期] 2023-09-06。

[基金项目] 国家自然科学基金(42072303)。

[第一作者] 荀鹏(1998—), 男, 硕士研究生, 研究方向: 地质工程、岩土工程, E-mail: 1249707274@qq.com。

[通信作者] 李娟(1983—), 女, 高级工程师, 研究方向: 地质工程、岩土工程, E-mail: 332535445@qq.com。

[引用格式] 荀鹏, 李娟, 魏玉峰, 等, 2024. 坝肩岩体质量 LDA-KNN 分类模型[J]. 成都理工大学学报(自然科学版), 51(2): 281-290+302.

Xun P, Li J, Wei Y F, et al., 2024. The LDA-KNN model to classify quality of rock mass of a dam abutment [J]. Journal of Chengdu University of Technology (Science & Technology Edition), 51(2): 281-290+302.

collected samples of drift rock at the abutment of a hydropower dam to establish a database of rock mass, selected the spacing between structural planes (D), coefficient of integrity of the rock (K_v), and rock quality index (RQD) as indices of evaluation based on an analysis of actual projects, and introduced the linear discriminant analysis-based method of dimension reduction and the K-Nearest-Neighbor multi-level classification model to develop a nonlinear model to classify rock mass. A combination of qualitative and quantitative analyses was used to comprehensively classify the rock mass based on multiple factors and indicators, and to solve the problems of the redundancy of information and the complexity of judgment based on multiple indicators. The proposed model generated more accurate results than previously developed schemes that were in line with engineering practice, reduced the influence of human factors, and exhibited a strong predictive capability.

Key words: rock mass structure; classification of rock mass quality; linear discriminant analysis; K-Nearest-Neighbor analysis; classification model

工程岩体质量等级分类是各类岩石工程基础工作的重要环节之一,它反映了岩体结构固有的物理力学特性,是一种对工程影响范围内的岩体,包括大坝地基岩体(徐建荣等,2021;孙鹏昌等,2022),建筑地基岩体,地下工程围岩等,按其工程地质条件的优劣以分级方式进行简化,评价工程岩体稳定性的方法。

在水利水电工程建设中,坝基(肩)岩体质量分级是对不同工程部位各级岩体的可利用性进行判别的基础,由此可进一步选择合理的建基面及嵌入深度。从工程的实际需求出发,选取合适的指标对岩体进行合理准确分级具有重要的现实意义。早期工程岩体分级依靠单指标判断,如 Terzaghi 岩体载荷指标分级、普氏系数法、和 RQD 法等。此类方法仅考虑了岩体的单一属性,忽视了工程岩体的复杂性,使得评价结果与实际情况之间存在较大差异。工程岩体分级开始由单指标判断走向考虑多指标的综合分级。当前,国际岩体分类方法中应用广泛的有 Q 系统分类(Fereidooni et al., 2015)和 RMR(Abdessattar et al., 2021)系统分类;而在中国,工程中常用的岩体分类方法主要为《工程岩体分级标准》(GB/T 50218—2014)中的 BQ 法(中华人民共和国水利部,2014;Wu et al., 2023)和《水力发电工程地质勘察规范》(GB 50287—2016)中提出的围岩与坝基(肩)的岩体工程地质分类方法(中国电力企业联合会,2016)。

随着计算机科学算法的进一步发展,越来越多的综合评价方法被应用到岩体的分类评价中。有学者(帅青燕和何亚伯,2013;Wang and Guo,

2019)将云模型理论应用于岩体质量分级,并通过进一步改进多维云模型理论来克服岩体分级中的模糊性问题(周坦等,2019);还有通过对蒙特卡洛模拟和 TOPSIS 方法耦合,针对岩体参数的不确定性,通过概率描述的可靠度给出分级结果(李绍红等,2017);一些学者(胡建华等,2012;陈顺满等,2018)通过粗糙集理论耦合不同的系统工程分析方法来解决指标权重确定的主观性问题;还有一些学者(周宗青等,2017;Tu et al., 2019)引入可拓学原理,建立了岩体质量等级预测的物元可拓模型,取得了不错的结果;同时在智能算法的普及下,支持向量机(Zheng et al., 2020)、迁移学习算法(Qin et al., 2019)、人工神经网络(郭少文等,2015;荆少东等,2022)和距离判别分析法(姚银佩等,2010;钱兆明等,2016)等方法也开始被应用于岩体质量分级中,通过训练样本数据,拟合得到非线性模型,能够在岩体质量分级中获得较为准确的结果。

但以上研究多注重指标权重的选取与指标中分类区间的模糊性研究,方法中评价指标权重的计算主观性较强,不确定性方法的应用较为复杂;人工神经网络等机器学习算法对样本分布有要求,超参数确定较为困难,且以上方法均缺少对大量数据的分类处理研究;忽视了传统定性分级中对岩体质量评价的重要性,同时考虑多指标计算会使分类评价复杂化,在实际工程上操作较为繁琐,可以引入降维算法解决这一问题。

本文通过对大量数据的分析,建立一种直接从指标数据到分类结果的坝肩岩体质量的综合评价模型,可以达到减小人为因素干扰和解决变量

和结果之间的不确定性的目的,同时通过对分类模型进行改进以减小误判率。运用分类算法,学习并结合传统的岩体质量评价中采用的经验、定性判断结果,基于实际工程中测量收集的数据,建立 LDA 降维方法和 KNN 相结合的多分类模型以实现岩体的非线性分级预测,得到一种可靠、简单的岩体质量评价预测方法。

1 数据集的获取与模型理论基础

1.1 数据集的获取

本文所采用数据来源于西藏那曲市某拟建水电站站址区勘探结果。该水电站两侧岩体为致密、坚硬的二长花岗岩,初拟坝高 265 m,属于高拱坝。拱坝的受力特点以及超过 200 m 的坝高使该工程对坝基岩体质量的要求极高。为了对坝肩处的岩体结构、岩体风化、岩体卸荷、岩体质量等级进行准确划分,进而综合评价岩体的工程性质,在两侧山体的勘探平洞中收集了大量数据(图 1)。

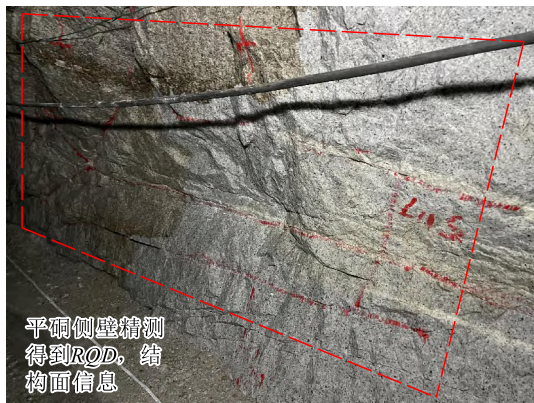


图 1 数据获得方法

Fig. 1 Method of data acquisition

1.2 岩体质量评价指标的选择

工程岩体质量受诸多因素影响,在工程实际中,需要在传统的单因素方法基础上建立岩体质量多因素指标分类体系。

当前应用于规范中的坝基岩体分级指标主要有岩石单轴抗压强度 R_b 、 RQD 值、 D 、 K_v 等。工程实践中还使用变形模量、岩体抗剪断强度参数 f' 和 c' 以及块度模数、岩体块度系数、岩体块度指数、吕荣值等;同时还需要考虑到岩体风化卸荷程度,岩体的蚀变状态等难以量化的指标。在工程应用中,以上的指标不可能全部采用。所以,只能根据具体工程的特点,选取有代

表性的指标进行岩体质量分类。因此,如何从这些众多指标中选取具有较强代表性的指标至关重要(边毓,2019)。

结合工程部位岩体实际情况,评价岩体质量时以国家标准《水力发电工程地质勘察规范》(GB 50287—2016)中提出的坝基岩体工程地质分类为基本依据。该规范中岩体工程地质分类的核心是岩体结构,对岩体工程性质评价主要内容为岩体的抗滑、抗变形能力。同时,为了更加符合工程实际,还要做到指标数据的易于获取;最后将工程地质分类设定为 I, II, III₁, III₂, IV 5 个质量级别,根据国标中提出的岩体特征中岩体结构,结构面发育程度进行定性划分,如 I 类岩体其特征呈整体状、巨厚层状,结构面不发育,延展性差,多闭合,工程性质好,属优良高混凝土坝地基,而 IV 类岩体呈碎裂结构,结构面很发育,且多张开或夹岩屑泥屑,不宜作高混凝土坝地基。局部存在该类岩体时应做专门处理。

综合考虑,对坝址区两岸坝肩岩体选取岩体结构面间距(D)、岩体完整性系数(K_v)、岩石质量指标(RQD)作为岩级的主要划分指标。

1.3 数据集的构建

根据工程实践,共收集站址区内 30 条勘探平洞接近 5 000 m 的数据进行训练学习,对每 5 m 洞段进行精细的测量,以 PD10 为例(表 1)。通过详细的现场调查,判断岩体的风化卸荷程度,结构面发育状况和岩体结构类型,从而对岩体工程地质分级做出定性判断,将现场定性分级作为分类标签并结合特征数据集进行学习。

收集到的各特征指标数据统计分布(图 2),可以看出数据分布情况并不相同,一些模型对于数据分布有一定要求,这时还需要对特征指标进行转换或寻找新的指标,而 KNN 模型的优点在于不需要对数据的分布做出任何假设,是典型的非参数模型,适合本文中的数据分类。

根据测量数据进行单因素指标的岩体分级,得到各级岩体占比情况(图 3),可以看出分别按照 D 、 RQD 、 K_v 进行单因素定量分级判断,得出的分类情况存在较大差异,会出现同一段的岩体指标对应着不同分级的情况,而且以 RQD 分级为例;岩体较为完整,强度较高的 II 类岩体占比达到 91%,与现场勘探情况不符,仅凭单因素定量分级无法准确对岩体质量进行评判。

表 1 训练样本数据集(部分)
Table 1 Dataset of training samples (partial)

PD10 硇段/m	评价指标			定性分级	PD10 硇段/m	评价指标			定性分级
	D/m	RQD/%	K_v			D/m	RQD/%	K_v	
0~5	0.09	81.0	0.237	IV	50~55	1.02	96.6	0.792	II
5~10	0.17	85.4	0.274	IV	55~60	1.14	96.2	0.843	I
10~15	0.53	91.6	0.476	III ₂	60~65	0.98	96.4	0.780	II
15~20	0.35	93.6	0.621	III ₂	65~70	0.49	95.6	0.831	III ₁
20~25	0.29	87.0	0.545	III ₂	70~75	0.66	95.8	0.780	II
25~30	0.33	93.0	0.614	III ₂	75~80	0.53	96.6	0.887	II
30~35	0.58	94.2	0.675	II	80~85	0.48	93.8	0.769	III ₁
35~40	0.47	91.6	0.694	III ₁	85~90	0.49	93.4	0.843	III ₁
40~45	0.29	86.4	0.629	III ₂	90~95	0.48	91.8	0.792	III ₁
45~50	0.44	94.8	0.748	III ₁	95~100	0.55	93.4	0.792	II

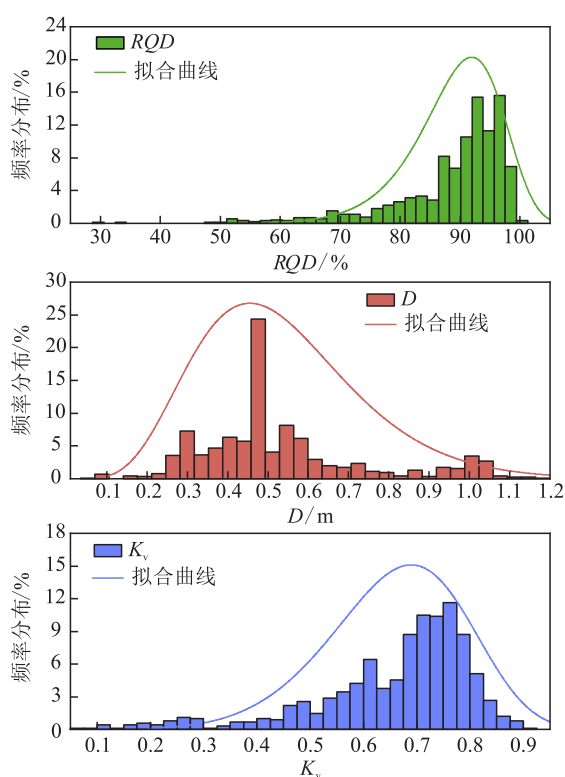


图 2 频率分布直方图

Fig. 2 Histogram of frequency distribution

1.4 模型的理论基础

线性判别分析 LDA 又称为 Fisher 线性判别 (Prakash and Rajkumar, 2021), 主要思想是将一个高维空间中的数据投影到一个较低维的空间中, 且投影后要保证各个类别的类内方差小而类间均值差别大。这意味着同一类的高维数据投影到低维空间后相同类别的聚在一起, 而不同类别之间相距较远, 并得到最佳的投影方向的表达式, 在该子空间中展现了数据的最佳可分离性。对于

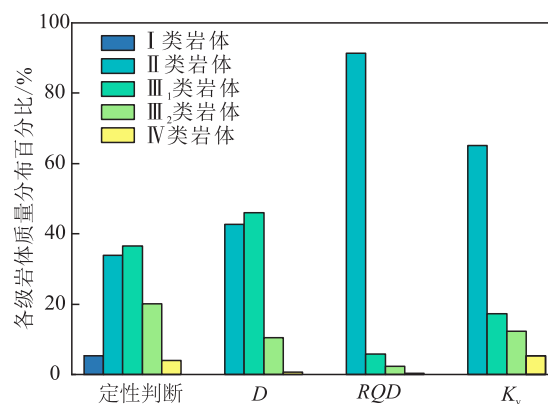


图 3 不同指标判断下各级岩体占比统计图

Fig. 3 Statistical chart of the ratios of rock mass at various levels according to judgments based on different indices

岩体质量分级来说, 需要在已确定样本岩级分级的情况下, 分别计算不同分级之间的类间离散度 S_b 和相同分级之间的类内离散度 S_w , 得到矩阵 $S_w^{-1}S_b$ 。矩阵中 $S_w^{-1}S_b$ 中的 d 个最大特征值及其对应的 d 个特征向量 (w_1, w_2, \dots, w_d) 即为所求的最优投影矩阵 W 。

LDA 算法其本身也可以作为分类器进行数据的分析预测, 但在多分类情况下更常见的是利用 LDA 特征提取的效果配合其他分类算法使用。

KNN 算法 (毋雪雁等, 2017) 是一种有监督的懒惰机器学习 (Lazy Learning) 算法, 应用场景非常广泛, 如船舶识别 (Jin et al., 2023), 航空航天工程 (Cui et al., 2023), 洪水预报 (陈鑫等, 2022), 图像文本分类等各种不同领域均可适用。其基本思想是在训练集中找到和测试集中的新数据最接近的 k 个样本, 根据这 k 个最近邻的类别投票决定测试样本的类别。

该方法常用于离散型分类问题,对于多分类问题可能出现不同标签下统计得到的最近邻样本相同的情况,导致无法判别,故本文决定将离散型标签量化,使其转换为连续变量(表 2),并将 KNN 用于回归问题,测试集中要预测的样本值通过求与它距离最近的 k 个点的距离值的加权平均值得到。

表 2 标签量化与样本值结果分类

Table 2 Results of classification of label quantization and sample values

岩体质量分级	标签量化结果	结果分类
I 类岩体	1.0	(0.5,1.5]
II 类岩体	2.0	(1.5,2.5]
III ₁ 类岩体	3.0	(2.5,3.5]
III ₂ 类岩体	4.0	(3.5,4.5]
IV 类岩体	5.0	(4.5,5.5]

KNN 是典型的非参数模型即不用假设样本数据的分布,适用于大量数据的非线性分类计算。本文采用欧氏距离作为特征空间中的 2 个实例点的距离度量,如下式:

$$D(x, y) = \left(\sum_{l=1}^n |x^{(l)} - y^{(l)}| \right)^{\frac{1}{2}} \quad (1)$$

式中: x, y 为 2 个样本向量; D 为 2 个样本之间的距离; n 为样本的特征数。

对于 KNN 算法,超参数 k 值的选取非常重要,较小的 k 值会使得对预测样本的预测错误概率更大,较大的 k 值能够减小噪声的影响,但会使类别之间的界限变得模糊,适当的 k 值特别重要。可以使用采用网格搜索(Qi and Tang, 2018)和 5-折交叉验证法对超参数进行优化,其中网格搜索法本质为一种穷举法,及对每一个 k 值单独建立模型计算准确率 ACC,为了避免与待判样本较远的训练实例也会对预测起作用,使预测发生错误,在实际应用中,考虑训练集数目的影响, k 值一般从较小的数值中选取,为保证分类准确, k 值最大取值不超过 26。5-折交叉验证(图 4)是将原始数据分成 5 组,将每个子集数据分别做一次验证集,其余的 4 组子集数据作为训练集,这样会得到 5 个模型。这 5 个模型分别在验证集中评估结果并可以得到一个平均准确率 \overline{ACC} (张化进等, 2022)。该方法作为一种使用广泛模型评估、模型选择方法,在本文中作为最优的超参数 k 值选取的依据,是模型建立这一部分的重要步骤。

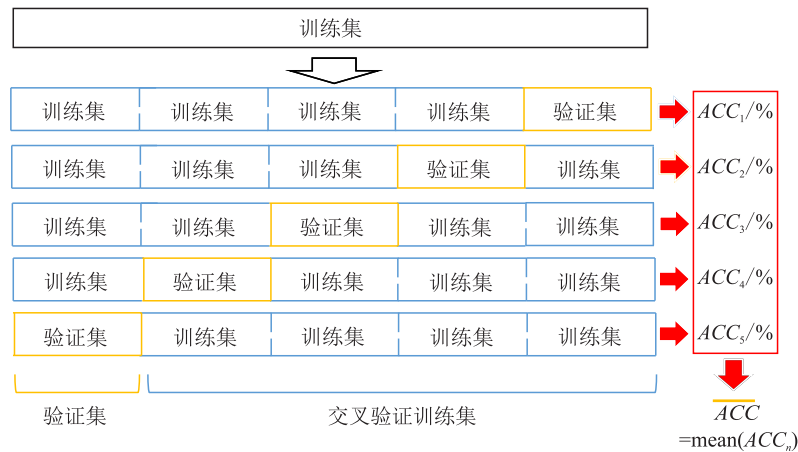


图 4 5-折交叉验证法

Fig. 4 Five-fold cross-validation

2 模型的建立

(1)将现场收集的所有数据作为训练集进行模型的训练学习,另外保留 2 条勘探平洞的数据作为测试集进行岩体质量评价分类模型的可靠性验证。

由于 KNN 算法的核心是通过计算样本之间的距离完成分类,如果一个特征值域范围非常大,那么距离计算就主要取决于这个特征,从而与实

际情况相悖,首先应该对数据进行归一化处理,使得预处理的数据被限定在一定的范围内,如下:

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

式中: x_{max} 为样本数据的最大值; x_{min} 为样本数据的最小值; x 为样本数据; x_{new} 为归一化后的样本数据。

对数据进行线性降维,需要输入站址区勘探平洞中所收集的归一化后的数据集:

$$E = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_m, y_m)\} \quad (3)$$

其中,任意样本 x_i 为 n 维向量,及数据的特征指标个数为 n 。结合上文所述,特征指标分别为结构面间距(X_1)、岩石质量指标(X_2)、完整性系数(X_3); $y_i \in \{C_1, C_2, \dots, C_k\}$ 。前文将岩体质量分为 5 级,故 $k=5$; m 为数据集样本个数。

可以分别计算类内散度矩阵 S_w 和类间散度矩阵 S_b :

$$S_w = \sum_{j=1}^k \sum_{x \in M_j} (x - \mu_j)(x - \mu_j)^T \quad (4)$$

$$S_b = \sum_{j=1}^k N_j (\mu_j - \mu)(\mu_j - \mu)^T \quad (5)$$

式中: μ_j 为第 j 类样本的均值向量; μ 为所有样本的均值向量; M_j 为第 j 类样本的集合; N_j 为第 j 类样本的个数; $j=1, 2, \dots, k$ 。

目的是将高维空间的数据投影到低维空间,假设其投影到的维度为 d ,对应的基向量为 (w_1, w_2, \dots, w_d) ,得到关于 W 的函数 $F(W)$:

$$F(W) = \frac{\text{tr}(W^T S_b W)}{\text{tr}(W^T S_w W)} \quad (6)$$

式中: W 为基向量组成的 $n \times d$ 矩阵。

为了使投影后类间离散度最大和相同分级之间的类内离散度最小,需要求出使得函数 $F(W)$ 最大时的自变量矩阵 W 的值。此时矩阵中 $S_w^{-1} S_b$ 中的 d 个最大特征值及其对应的 d 个特征向量 (w_1, w_2, \dots, w_d) 即为所求的最优投影矩阵 W 。

得到新的数据集 $D' = DW$ 。由数据集的散点图可知,5 组不同分类数据的分散程度低。将原本 3 个维度的数据进行低维度投影,寻找出最优的投影矩阵。

$$D' = DW \quad \downarrow$$

$$(Y_1 \ Y_2) = (X_1 \ X_2 \ X_3) \begin{pmatrix} 0.988 & 0.529 \\ 0.045 & -0.259 \\ 0.148 & -0.808 \end{pmatrix} \quad (7)$$

式中: W 为最佳投影矩阵; X_1, X_2, X_3 为上文所述的原特征指标数据; Y_1, Y_2 为原指标经过线性变换后得到的新的特征指标数据。

如图 5, LDA 算法使得不同类别的数据变得较为集中,从而得到了具有良好可分性的低维度样本,用于后续的分类建模和预测。

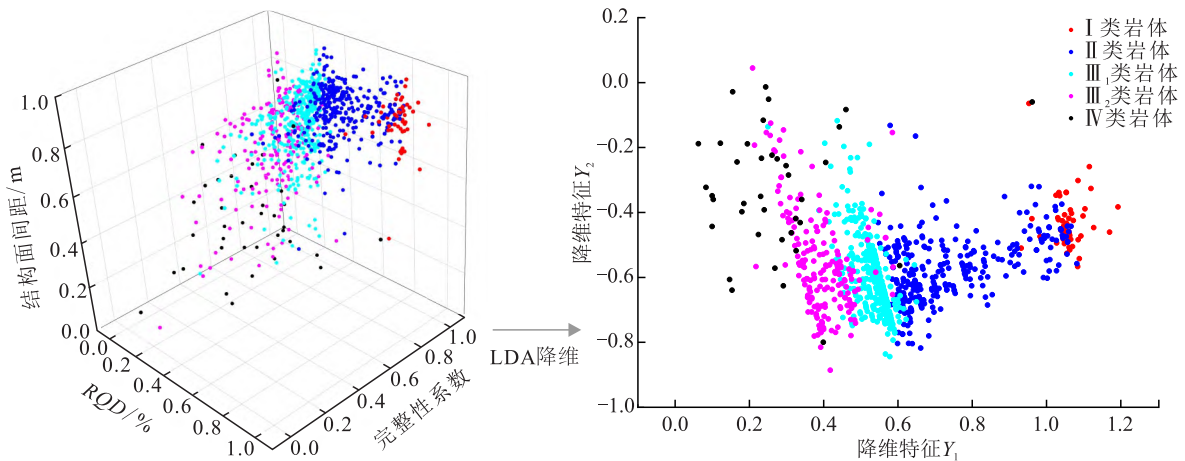


图 5 降维效果展示
Fig. 5 Effect of dimension reduction

(2)在完成降维后,采用 KNN 算法进行分类模型的建立和样本分类的预测,首先输入降维优化后的特征矩阵,建立分类器模型。

使用 KNN 算法对待判样本 x 进行分类。对于某一个样本点 x 来说,使用 $KNN(x)$ 来表示与 x 最近的 k 个临近样本点欧式距离的集合:

$$KNN(x) = \{D(x, x_1), D(x, x_2), \dots, D(x, x_k)\} \quad (8)$$

使用高斯函数定权(毕京学等,2017),使得距离近的临近样本获得的权值较高,距离远的样本获得的权值较小。将 k 个近邻欧氏距离进行 Z-score 标准化处理得到 D^* 使其符合参数为 $(0, 1)$ 的正态分布,使用高斯函数计算权值 ω_i :

$$\omega_i = \frac{e^{(-D_i^* / 2)}}{\sqrt{2\pi}} \quad (i=1, 2, \dots, k) \quad (9)$$

增加权重影响后,输出测试集样本的加权平均值。

$$f(\mathbf{x}) = \frac{\sum_{i=1}^k \omega_i g(\mathbf{x}_i)}{\sum_{i=1}^k \omega_i} \quad (i=1, 2, \dots, k) \quad (10)$$

式中: $g(\mathbf{x}_i)$ 为样本 \mathbf{x}_i 所对应的标签量化后的值。

则可以定义判别函数为:

$$G(\mathbf{x}) = \begin{cases} \text{I 类岩体} & f(\mathbf{x}) \in (0.5, 1.5] \\ \text{II 类岩体} & f(\mathbf{x}) \in (1.5, 2.5] \\ \text{III}_1 \text{ 类岩体} & f(\mathbf{x}) \in (2.5, 3.5] \\ \text{III}_2 \text{ 类岩体} & f(\mathbf{x}) \in (3.5, 4.5] \\ \text{IV 类岩体} & f(\mathbf{x}) \in (4.5, 5.5] \end{cases} \quad (11)$$

(3) 通过分层采样的方式从 5 种岩体质量等级的数据中分别选取 80% 的数据作为测试集, 20% 的数据作为验证集测试分类器, 通过 5-折交叉验证的方法评估模型在 k 不同取值下的性能, 进行调参。

如前文所示, 为保证分类准确, 选择的最大近邻个数不超过 26 个。采用 MATLAB 编写程序运算后可得到每次验证下不同 k 值的分类准确率结果。

最后的分类准确率结果如图 6 所示。可以看

到, 交叉验证结果表示 LDA-KNN 模型最高准确率在 91% 以上, 随着 k 值增大, 算法的分类效果会逐渐提升; 但在增大到某个值后, 随着的进一步增大, 算法的分类效果会逐渐下降。当 $k=15$ 时, 分类效果最好。同时模型的综合准确率可达 88%。结果表明模型的准确性和可靠性较高。

LDA 降维方法结合 KNN 算法分析所制作的分类器(图 7), 其基本思路是利用 KNN 算法将降维方法处理过的样本数据进行判别, 两者相结合以提高准确率。模型计算过程均可使用 MATLAB 软件进行编程计算, 十分简便, 且适用于大量数据的处理。

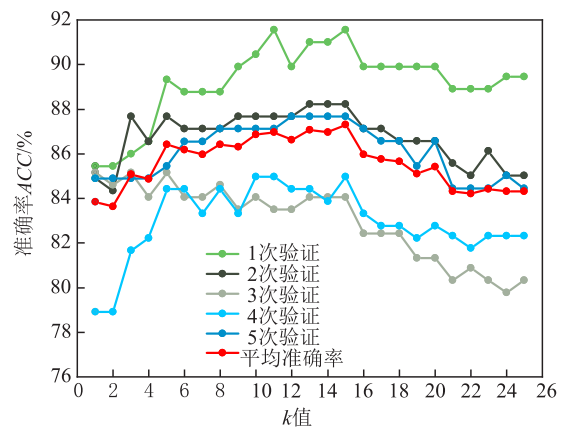


图 6 确定超参数 k 值

Fig. 6 Determination of the hyper-parameter k

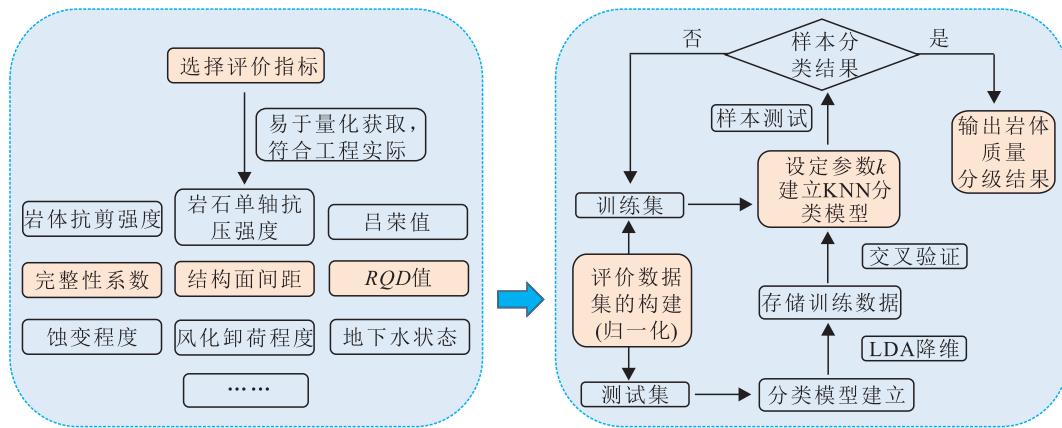


图 7 数据处理与算法流程

Fig. 7 Data processing and algorithmic flow

3 模型的应用

为进一步验证该方法在工程实际中对岩体质量评价的有效性, 将构建完成的 LDA-KNN 算法分类模型对测试集中的样本进行判别, 与未进行

LDA 降维优化的 KNN 模型判断结果进行比较。PD25, PD33 分别为长度 170 m 和 150 m 的勘探平洞, 位于坝址区右岸, 是为了进一步得到右岸坝肩处岩体质量信息所新开挖的平洞。利用先开挖平洞信息数据库所构建的分类模型对岩体信息进

行预测,并对模型进行精度验证和泛化能力的评价,采用相对均方根误差(RMSE)和决定系数(R^2)作为模型的性能度量,公式如下:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (\hat{y}_i - y_i)^2}{\sum_{i=1}^m (\bar{y}_i - y_i)^2} \quad (13)$$

式中: y_i 为样本实际值; \hat{y}_i 为样本估计值; m 为待判样本数。

得到的结果如图 8。

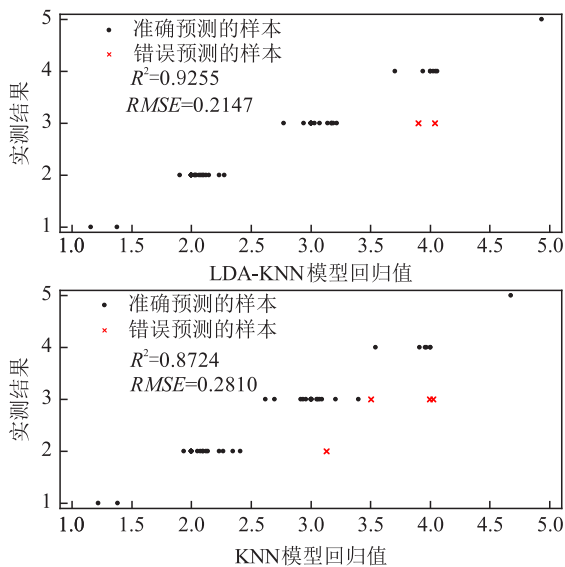


图 8 模型结果对比

Fig. 8 Comparison of the results of different models

由图 8 可知,2 种模型决定系数均大于 0.8,均取得了较好的拟合结果。其中,LDA-KNN 模型 R^2 达到了 0.93,说明改进后的 KNN 模型相比与传统的 KNN 模型表现出更高的精度,同时均方根误差下降了 23.6%,改进后模型的错判数明显减少,表明在岩体质量分级领域,改进后的模型具有更好的估测潜力。

继续研究模型的可靠性,将现场定性判断,常规评价结果和未进行 LDA 降维优化的 KNN 模型判断结果作为参考,进行对比,分析本文方法的工程可靠性。

通过对比不同方法下的分类结果(图 9),可以得出该模型对定性分级的学习分类结果非常理想;根据规范得出的常规评价方法结果较为保守,如 PD33 在 25~40 m 洞段处,模型分类结果将 III₂ 类岩体评价为 III₁ 类岩体,但总体上与常规评价结果具有很好的一致性;PD25 在 65~75 m 洞段处,其分类结果与常规评价结果同为 III₂ 类岩体,但现场定性判断结果却为 III₁ 类岩体,这个结果体现了 LDA-KNN 算法对岩体评价的优化效果。

以定性分级结果作为样本的实际值计算准确率为 97%,以常规评价结果作为样本的实际值计算准确率为 92%,说明该模型在岩体质量评价中适用性较高,其评价结果可为复杂岩体的综合评价结果提供复核与优化依据。最后,相比于未改进的 KNN 分类模型,可以看到引入 LDA 降维算

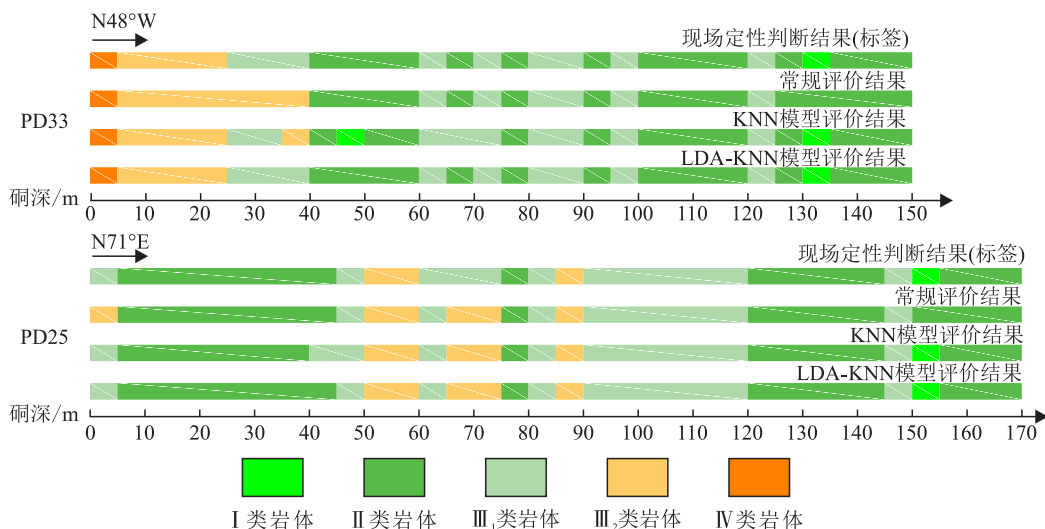


图 9 不同方法预测结果

Fig. 9 Results of predictions by different methods

- (增刊 2):427—432.
- Qian Z M, Ren G F, Chu F J, *et al.*, 2016. Rock mass quality classification based on PCA and Fisher discrimination analysis[J]. *Rock and Soil Mechanics*, 37(S2):427—432. (in Chinese)
- 帅青燕, 何亚伯, 2013. 基于云模型的坝基岩体质量综合评价[J]. *东南大学学报(自然科学版)*, 43(增刊 1): 54—58.
- Shuai Q Y, He Y B, 2013. Comprehensive evaluation on rock quality of dam foundation based on cloud model [J]. *Journal of Southeast University (Natural Science Edition)*, 43(S1):54—58. (in Chinese)
- 孙鹏昌, 卢文波, 杨招伟, 等, 2022. 白鹤滩坝肩边坡开挖爆破损伤预测研究[J]. *水力发电学报*, 41(10): 30—41.
- Sun P C, Lu W B, Yang Z W, *et al.*, 2022. Prediction of rock damage induced by blasting excavation in high rock slope of Baihetan Dam abutment[J]. *Journal of Hydroelectric Engineering*, 41(10):30—41. (in Chinese)
- 毋雪雁, 王水花, 张煜东, 2017. K 最近邻算法理论与应用综述[J]. *计算机工程与应用*, 53(21):1—7.
- Wu X Y, Wang S H, Zhang Y D, 2017. Survey on theory and application of K-Nearest-Neighbor algorithm[J]. *Computer Engineering and Applications*, 53(21): 1—7. (in Chinese)
- 徐建荣, 赖道平, 吴关叶, 等, 2021. 适应柱状节理玄武岩坝基的特高拱坝结构研究[J]. *水力发电学报*, 40(3):155—164.
- Xu J R, Lai D P, Wu G Y, *et al.*, 2021. Study on the super high arch dam structure adapting to the columnar basalt base [J]. *Journal of Hydroelectric Engineering*, 40(3):155—164. (in Chinese)
- 姚银佩, 李夕兵, 宫凤强, 等, 2010. 加权距离判别分析法在岩体质量等级分类中的应用[J]. *岩石力学与工程学报*, 29(增刊 2):4119—4123.
- Yao Y P, Li X B, Gong F Q, *et al.*, 2010. Application of weighted mahalanobis distance discriminant analysis method to classification of rock mass quality[J]. *Chinese Journal of Rock Mechanics and Engineering*, 29(S2):4119—4123. (in Chinese)
- 中国电力企业联合会, 2016. 水力发电工程地质勘察规范:GB 50287—2016[S]. 北京:中国计划出版社.
- China Electricity Council, 2016. Code for Hydropower Engineering Geological Investigation: GB 50287—2016 [S] Beijing:China Planning Press. (in Chinese)
- 张化进, 吴顺川, 韩龙强, 等, 2022. 基于改进 Dempster-Shafer 证据理论的岩体质量异质集成评价方法[J]. *岩土力学*, 43(增刊 1):532—541.
- Zhang H J, Wu S C, Han L Q, *et al.*, 2022. Heterogeneous ensemble evaluation method of rock mass quality based on improved Dempster-Shafer evidence theory [J]. *Rock and Soil Mechanics*, 43(S1):532—541. (in Chinese)
- 周坦, 胡建华, 匡也, 2019. 基于模糊 RES—多维云模型的岩体质量评判方法与应用[J]. *中国有色金属学报*, 29(8):1771—1780.
- Zhou T, Hu J H, Kuang Y, 2019. Rock mass quality evaluation method and application based on fuzzy RES-multidimensional cloud model[J]. *The Chinese Journal of Nonferrous Metals*, 29(8):1771—1780. (in Chinese)
- 周宗青, 李术才, 李利平, 等, 2017. 岩体质量等级分类预测方法及其工程应用[J]. *中南大学学报(自然科学版)*, 48(4):1049—1056.
- Zhou Z Q, Li S C, Li L P, *et al.*, 2017. Classification method of rock mass quality and its engineering application[J]. *Journal of Central South University (Science and Technology)*, 48(4):1049—1056. (in Chinese)
- 中华人民共和国水利部, 2014. 工程岩体分级标准: GB/T 50218—2014[S]. 北京:中国计划出版社.
- Ministry of Water Resources of the People's Republic of China, 2014. Standard for Engineering Classification of Rock Mass:GB/T 50218—2014[S]. Beijing:China Planning Press. (in Chinese)
- Abdessattar L, Neguritsa D L, Bedr S, *et al.*, 2021. Determination and quality classification of rock mass of the Diatomite mine, Algeria[J]. *News of the Ural State Mining University*, 61(1):17—24.
- Fereidooni D, Khanlari G R, Heidari M, *et al.*, 2015. Assessment of a modified rock mass classification system for rock slope stability analysis in the Q-system [J]. *Earth Sciences Research Journal*, 19(2): 147—152.
- Jin S Y, Su Y, Guo C J, *et al.*, 2023. Offshore ship recognition based on center frequency projection of improved EMD and KNN algorithm [J]. *Mechanical Systems and Signal Processing*, 189.
- Cui L F, Zhang Q Z, Shi Y, *et al.*, 2023. A method for satellite time series anomaly detection based on fast-DTW and improved-KNN [J]. *Chinese Journal of Aeronautics*, 36(2):149—159.

- 周博,汪华斌,赵文锋,等,2012.黏性材料细观与宏观力学参数相关性研究[J].岩土力学,33(10):3171—3175.
- Zhou B, Wang H B, Zhao W F, *et al.*, 2012. Analysis of relationship between particle mesoscopic and macroscopic mechanical parameters of cohesive materials [J]. *Rock and Soil Mechanics*, 33(10):3171—3175. (in Chinese)
- Athanassiadis A G, Miskin M Z, Kaplan P, *et al.*, 2014. Particle shape effects on the stress response of granular packings[J]. *Soft Matter*, 10(1):48—59.
- Fonseca J, O'Sullivan C, Coop M R, *et al.*, 2013. Quantifying the evolution of soil fabric during shearing using scalar parameters [J]. *Géotechnique*, 63(6):818—829.
- Hall S A, Bornert M, Desrues J, *et al.*, 2010. Discrete and continuum analysis of localised deformation in sand using X-ray μ CT and volumetric digital image correlation[J]. *Géotechnique*, 60(5):315—322.
- Liu L F, Zhang Z P, Yu A B, 1999. Dynamic simulation of the centripetal packing of mono-sized spheres [J]. *Physica A: Statistical Mechanics and its Applications*, 268(3-4):433—453.
- Oda M, 1972. Initial fabrics and their relations to mechanical properties of granular material [J]. *Soils and Foundations*, 12(1):17—36.
- Thornton C, 2000. Numerical simulations of deviatoric shear deformation of granular media [J]. *Géotechnique*, 50:43—53.
- Xu W J, Hu L M, Gao W, 2016. Random generation of the meso-structure of a soil-rock mixture and its application in the study of the mechanical behavior in a landslide dam[J]. *International Journal of Rock Mechanics & Mining Sciences*, 86:166—178.

(责任编辑:黄维婷)



(上接第 290 页)

- Prakash P N S, Rajkumar N, 2021. Improved local fisher discriminant analysis based dimensionality reduction for cancer disease prediction[J]. *Journal of Ambient Intelligence and Humanized Computing*, 12:8083—8098.
- Qi C C, Tang X L, 2018. A hybrid ensemble method for improved prediction of slope stability[J]. *International Journal for Numerical and Analytical Methods in Geomechanics*, 42(15):1823—1839.
- Qin Y G, Luo Z Q, Wang J, *et al.*, 2019. Evaluation of goaf stability based on transfer learning theory of artificial intelligence[J]. *IEEE Access*, 7:96912—96925.
- Tu W F, Li L P, Li S C, *et al.*, 2019. Research on the application of dynamic weighting on the rock mass quality rating [J]. *Arabian Journal of Geosciences*, 12(3):87.
- Wang J C, Guo J, 2019. Research on rock mass quality classification based on an improved rough set-cloud model[J]. *IEEE Access*, 7:123710—123724.
- Wu A Q, Zhao W, Zhang Y H, *et al.*, 2023. A detailed study of the CHN-BQ rock mass classification method and its correlations with RMR and Q system and Hoek-Brown criterion [J]. *International Journal of Rock Mechanics and Mining Sciences*, 162:105290.
- Zheng S, Jiang A N, Yang X R, *et al.*, 2020. A new reliability rock mass classification method based on least squares support vector machine optimized by bacterial foraging optimization algorithm[J]. *Advances in Civil Engineering*, 2020:3896215.

(责任编辑:黄维婷)