

基于SMOTE的盾构数据增强与地层分类预测研究

武文清

(中铁十四局集团大盾构工程有限公司, 江苏 南京 210017)

摘要: 目前盾构施工中,地质勘探精度较低,在盾构掘进时难以精确获取掌子面地层与不良地质情况,存在一定施工风险。依托某市快速连接隧道工程,提出基于机器学习的盾构机掘进地质情况实时识别方法。首先通过对地质数据进行相关性矩阵分析,选取部分盾构掘进参数作为模型输入,并按照复合地层情况进行分类整理作为模型输出,处理形成时序数据集和环代表数据集,用于对比数据量对分类结果的影响;其次,采用合成少数类过采样技术进行数据增强,解决数据不平衡的问题,分别形成环代表数据与时序数据的SMOTE增强数据集;最后,通过多种机器学习模型进行分类预测,并采用准确度(ACC)和混淆矩阵检验模型效果,对比不同数据集、不同模型导致的分类预测性能变化。获取时序数据进行数据集整理可以获得更大的数据量,保证机器学习模型的地层识别效果;通过SMOTE对地层数据进行处理,可实现不同类样本量的平衡与总样本量的扩充,所选模型中KNN、SVM、ANN模型分类效果得到进一步提高。

关键词: 盾构隧道;复合地层分类;数据增强;机器学习;地层预测

中图分类号: U452.1+1

文献标志码: A

文章编号: 1009-7716(2025)02-0266-07

Study on Shield Data Enhancement and Stratum Classification Prediction Based on SMOTE

WU Wenqing

(China Railway 14th Bureau Group Big Shield Engineering Co., Ltd., Nanjing 210017, China)

Abstract: Currently, in shield tunneling construction, the precision of geological exploration is relatively low, and it is difficult to accurately obtain the stratum of the palm surface and the bad geological condition during shield tunneling with some construction risks. Relying on a municipal rapid connector tunnel project, a method for real-time identification of geological conditions during shield tunneling based on machine learning is proposed. Firstly, by performing a correlation matrix analysis on geological data, several shield tunneling parameters are selected as model inputs. According to the situation of composite strata, the model is classified and sorted as output. Time series data set and ring representative data set are formed to compare the effect of data volume on classification results. Secondly, the synthetic minority oversampling technique (SMOTE) is used to enhance the data and to solve the problem of data imbalance, which forms SMOTE enhanced datasets for both ring representative data and time series data. Finally, various machine learning models are used for classification prediction, and accuracy (ACC) and confusion matrix are used to test the model effect. The classification prediction performance variations caused by different datasets and models are compared. A larger amount of data can be obtained by acquiring the time series data for data set collation to ensure the stratum identification effect of machine learning model. The use of SMOTE to process the stratum data can realize the balance of different sample sizes and the expansion of total sample sizes. The classification effects of KNN, SVM and ANN models have been further improved.

Keywords: shield tunnel; composite stratum classification; data enhancement; machine learning; stratum prediction

0 引言

随着我国经济的迅速发展,城市建设呈现立体

式发展趋势,地下隧道成为城市发展的重点。而盾构法采用盾构机掘进开挖,具有封闭性好、掘进速度快、同步衬砌等优点,得到广泛使用,且随着研究的深入,目前在超大直径、长距离、复杂地质上不断发展^[1-3]。然而,对于复杂地质问题,盾构法施工过程中还存在掘进地层感知困难的情况,这给工程施工

收稿日期: 2024-06-02

作者简介: 武文清(1983—),男,本科,高级工程师,从事隧道工程管理与研究工作。

带来了许多挑战。

在地下隧道施工前,为了避免不良地质对隧道施工的影响,通常会对地层进行地质勘探,并设计施工路线,尽可能了解掘进前方地层情况,从而避免风险情况发生。由于工程中地质勘探是按一定间隔设置钻孔勘探,所得地层情况不连续,精度会有一定降低,在非盾构法隧道施工中也会使用补勘手段,如振动噪声反演成像^[4]、电阻抗成像^[5]、地震CT法^[6]、跨孔电阻率CT法^[7]和电磁波CT法^[8]等。

由于盾构施工干扰较多、空间狭窄,一般勘探方法在盾构机机载使用方面存在明显不足,不能准确探测出不良地质的位置,即传统物探方法在准确性和泛用性方面有待提升,在掘进过程中有必要提高实时识别不良地质的精度,避免掘进风险。盾构机在掘进过程中,开挖面前方的地层变化是处于无法观察到的隐藏状态,而掘进参数数据是可以观察到的状态,掘进参数的变化可以反映开挖面前方地质情况和掘进机之间的相互作用关系,在一定程度上可以说是地质条件的外在表达。因此,利用机器学习方法建立盾构掘进参数和地层特征或地层分类之间的内在联系,进而可以建立基于掘进参数数据驱动的地层识别方法。

通过采用盾构掘进数据进行实时地层预测,不仅可以在施工过程中提供及时的地层信息,帮助施工人员调整相应的施工参数和策略,而且可以为后续工程设计和施工提供重要的参考依据。朱梦琦^[9]将盾构每环掘进时上升段的总推进力、刀盘功率、刀盘转矩、掘进速度、刀盘速度和刀盘转速等盾构掘进参数的均值和方差输入随机森林和AdaCost模型,进行围岩等级的预测,结果表明刀盘转速、撑靴压力、撑靴俯仰角、推进速度和撑靴滚动角等参数能更好地反映围岩状况;刘明阳^[10]将盾构机总推力、掘进速度、刀盘转速和刀盘扭矩输入局部切空间排列(LTSA)与极限梯度提升(XGboost)模型,进行地质类型识别预测,他利用LTSA算法从高维盾构机运行参数中提取内蕴低维特征作为分类模型输入参数,减少所选盾构运行参数中的冗余信息并保留其中的地质特征,解决了输入参数维度过高造成的识别模型准确度下降和训练效率降低的问题。

由于数据集总量小,地层分类往往存在数据量小且不平衡的问题,现有研究多是从机器学习模型的改进来解决。Hou Shaokang^[11]用网格搜索方法优化了多种机器学习模型及其叠加的集成分类器,通

过输入盾构操作参数,证明叠加集成分类器在小样本、不平衡样本数据的掌子面岩体等级实时预测上学习和泛化能力更强。

现有研究缺少对地层数据的数据增强的机器学习地层预测效果研究,本文基于工程实际,采用时序数据以提高数据量,并与环代表数据进行比较;同时基于合成少数类过采样技术(SMOTE)分别对环代表数据和时序数据中少数类样本进行增强处理,以解决数据不平衡的问题。

1 方法与原理

1.1 SMOTE 数据增强

由 N. V. Chawla^[12]等人(2002)提出,SMOTE 基于少数类样本进行插值,从而对少数类样本进行扩充增强,以实现各类别数据量的平衡。

具体原理可概括为:

对于少数类样本,找到该样本 K 个最近邻,其中 K 为预设参数。在选定样本间插值,得到新样本。再重复上述步骤,直至各类样本量均衡。

若采取线性插值,可用式(1)表示。

$$y_n = y_0 + \alpha(y - y_0) \quad (1)$$

式中: y_0, y_n, y 分别为原始样本、新样本、近邻样本, α 为随机权重系数,在 0~1 之间。

1.2 机器学习方法

地层识别是分类问题,且数据量较小,常见机器学习模型如 K 近邻方法、支持向量机、决策树、集成模型、神经网络等即可获得较好的预测性能,在此对增强前后的数据进行建模训练。各模型原理进行简要说明如下。

(1) K 近邻法^[13](k -nearest neighbor, KNN)能够快速高效地解决建立在特殊数据集上的预测分类问题,但 KNN 不产生模型,因此其准确性并不具备强可推广性。

(2) 支持向量机^[14](support vector machines, SVM)是一种二分类、小样本有监督学习模型。SVM 利用核函数将一个复杂的分类任务映射到更高维的数据空间,使之变为一个线性可分的任务。例如,在一个二维平面上的点,没法通过一条直线把两个类别完全区分开,但是选择合适的方法把这个二维平面映射到三维空间中,就有可能通过一个平面去完全区分^[15]。

(3) 决策树(Decision Tree, DT)是一类常见的机器学习方法,其中 ID3 算法^[16](及其改进 ID5 算

法^[17])、C4.5算法^[18]、CART算法^[19]等应用广泛。DT是经典的树形结构,易于理解,具有较好的模型可解释性。

(4)集成模型主要包括提升算法(Boosting)和袋装算法(Bagging)。Boosting^[20]利用基学习器之间串行的方式构造强学习器,从而加强整体模型的预测效果。常见的提升算法如AdaBoost,缩写自Adaptive Boosting。通过迭代降低对正确分类的基学习器损失权值,AdaBoost可以不断降低整体模型的错误率。Bagging^[21]是一种基于自助采样法的并行式集成学习方法,即对随机抽样得到的子数据集分别训练多个基学习器,并将其并行得到强学习器。

(5)神经网络^[22]在机器学习模型中是有监督学习一个较大的分支,又被冠以深度学习(DL)模型。人工神经网络通过模拟生物神经网络的结构和功能,由大量的神经元和之间的连接构成,包括输入层、隐含层、输出层等架构,具有较好的非线性拟合能力。

1.3 结果检验

地层分类识别为多分类问题,采用多分类问题的准确度ACC和混淆矩阵进行定量评价。其中准确度如式(2)所示:

ACC = 正确识别的样本量 / 总样本量 (2)

混淆矩阵(Confusion Matrix):混淆矩阵是一个更详细的评估工具,它展示了模型在不同类别上的分类结果,可以从混淆矩阵中计算准确度、精确度、召回率等指标。混淆矩阵图中包括真阳性率TPR(即召回率REC)和假阳性率FPR,如式(3)、式(4)所示:

TPR = REC = TP / (TP + FN) (3)

FNR = FN / (TP + FN) = 1 - TPR (4)

其中,预测结果分为:正样本被正确识别的数量 positive true(TP)、误报的负样本数量 positive failure(FP)、负样本被正确识别的数量 negative true(TN)、漏报的正样本数量 negative failure(FN)。

2 施工数据收集、处理与增强

2.1 数据收集与说明

2.1.1 地质数据

实例分析所用数据集来自某市越江盾构施工项目,包括盾构掘进参数数据、隧道设计数据、地层类型数据。根据区域地质及隧道设计资料,盾构穿越均为复合地层,采用顶部地层、中间地层、底部地层进行划分,并以掘进先后顺序进行编号,共有8种,

按环号来看,462环中,2号复合地层占比最高,达60.83%,4号复合地层占比位居其次,其他的复合地层占比较小,各地层统计参数及复合地层分类见表1、表2。

表1 隧道穿越地层编号与名称

层号	名称
2-2	淤泥质粉质黏土
2-2a	粉质黏土夹粉砂
2-3	粉砂
2-4	粉质黏土
2-5	粉砂
2-6	粉质黏土

表2 地层分类

复合地层序号	底层编号	中层编号	顶层编号	环数	占比/%
1	2-6		2-2a	10	2.16
2	2-6		2-5	278	60.17
3	2-6		2-4	13	2.81
4	2-5		2-4	63	13.64
5	2-5	2-4	2-3	19	4.11
6	2-5	2-2a/3/4	2-2	23	4.98
7	2-5	2-3	2-2a	29	6.28
8	2-3		2-2a	27	5.84
总计				462	100

2.1.2 盾构PLC参数

盾构PLC数据为盾构掘进时持续记录数据,导出数据频率为一分钟一条,为记录时间前一分钟的平均值,基于工程收集的掘进参数,初步整理推进速度、刀盘速度、贯入度、刀盘转矩、推进油缸压力、总推力、土压等参数。图1~图7为PLC数据进行每环代表数据提取后的纵向分布图。

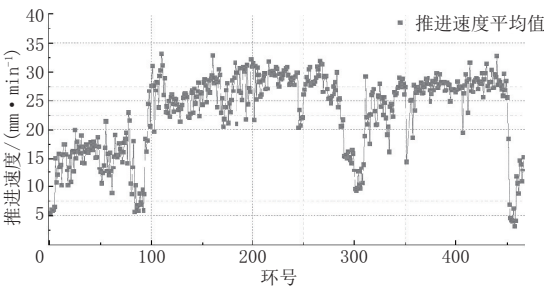


图1 推进速度平均值纵向分布图

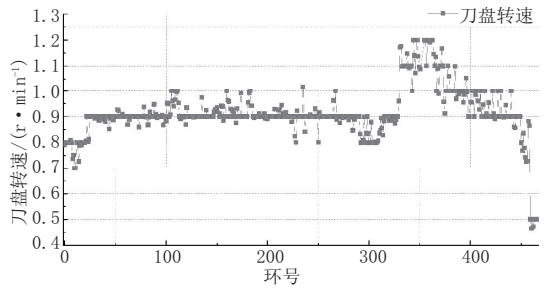


图2 刀盘速度纵向分布图

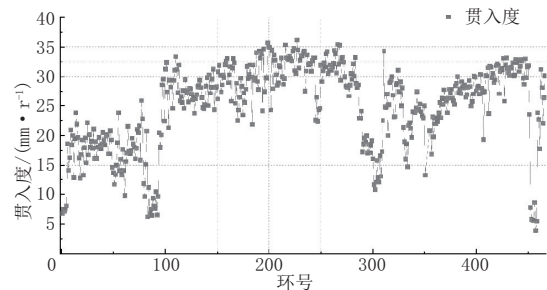


图3 贯入度纵向分布图

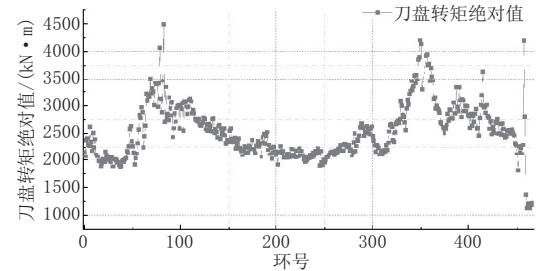


图4 刀盘转矩绝对值纵向分布图

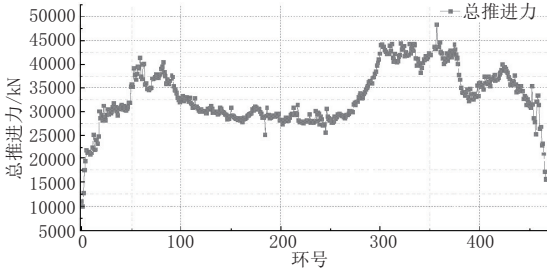


图5 掘进阶段总推进力纵向分布图

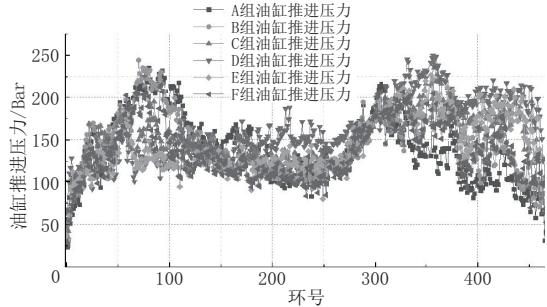


图6 掘进阶段推进油缸压力纵向分布图

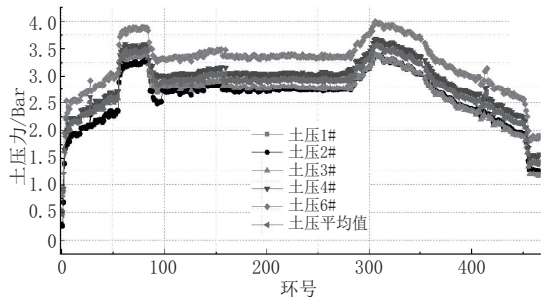


图7 掘进阶段土压纵向分布图

推进速度在稳定掘进后基本维持在 30 mm/min 左右,在 77~93 环、287~310 环因桩基与孤石存在导致贯入度降低,进而使推进速度平均值降低。

刀盘转矩在 1 120~4 200 kN·m 之间,总推进力在 20 000~45 000 kN 之间,推进油缸压力在 7.5~25 MPa 之

间,三者与隧道埋深呈正相关关系,即随埋深增大而增大,随埋深减小而减小。

在地面段,土压与埋深呈正相关关系,而进入江面段,土压则保持稳定,土压平均值在 0.3 MP 啊左右。

2.2 数据集构建

2.2.1 阶段划分

通过检验掘进速度平均值是否为 0,将数据划分为停机阶段与掘进阶段,并将停机阶段数据剔除。

2.2.2 参数选取

检验发现各项数据基本不符合正态分布。对掘进参数各数据项进行相关系数矩阵计算,采用斯皮尔曼相关系数,计算得到 ρ 和 P 值如图 8、图 9 所示, P 值计算值基本小于 0.01,即在 99% 的置信水平上认为各参数间具有相关性。仅有推进速度平均值与 C 组油缸推进压力和土压参数、贯入度与土压参数的 P 值较大,相关性不显著。由相关系数计算 ρ 可知地层分类标签与其他参数在 $-0.351\sim 0.634$ 之间,与部分参数可达强相关,可用上述参数进行模型输入;掘进速度平均值与贯入度、土压参数之间的相关性均较高,可各保留一个参数进行数据降维,在此保留掘进速度平均值和土压平均值。

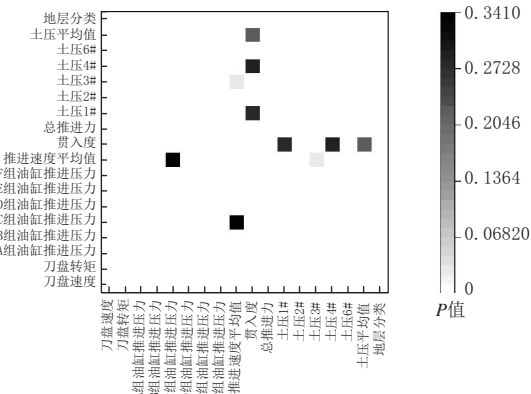


图8 相关系数矩阵(P 值)

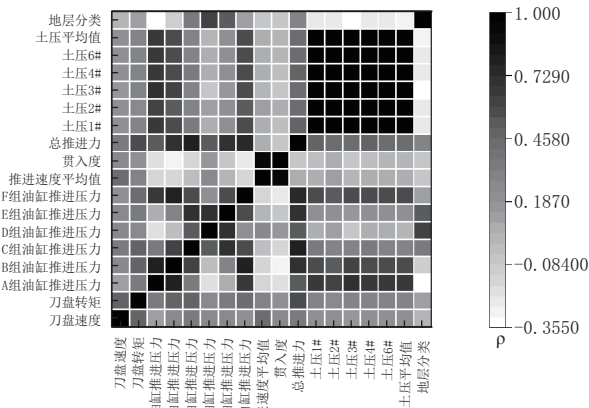


图9 相关系数矩阵(ρ)

因此,选取如下参数进行数据集构建:刀盘速度、刀盘转矩、A~F组油缸推进压力、总推进力、推进速度平均值、土压平均值、地层分类。

经过接收段数据的剔除与数据清洗,得到了6~462环地层识别时序数据集,输入为31 494×11矩阵,响应为31 494×1向量。基于时序数据集,进行环代表值计算处理,得到6~456环环代表值数据集,即451×12矩阵。

2.3 基于SMOTE的数据增强

基于SMOTE方法,采用Python编程,对环代表数据集与时序数据集中占比较少的数据进行数据增强以扩充数据,得到增强后的环代表增强数据集与时序增强数据集,分别为2 224×12矩阵、157 624×12矩阵,经过处理,类别1~8实现了数据量的平衡。

按增强前后时间进行排序,并进行统计参数计算,可得增强前后时序数据数据分布对比分析如图10、图11所示。通过SMOTE方法数据增强后,各参数的最大值和最小值没有变化,原少数类样本增多。

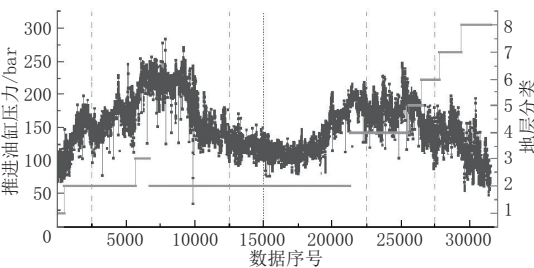


图 10 增强前时序数据分布

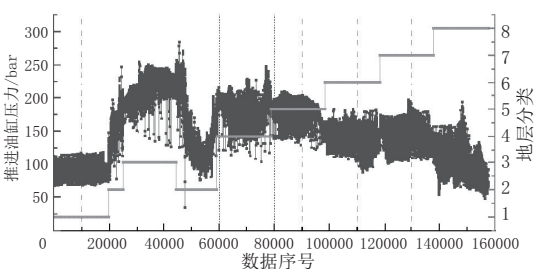


图 11 增强后时序数据分布

3 基于机器学习模型的地层预测研究

3.1 机器学习超参数设定

对于环代表数据,通过5折交叉验证并通过贝叶斯算法优化超参数,确定各模型最优超参数见表3。基于已确定超参数,将环代表数据、时序数据、SMOTE增强数据分别按6:2:2的比例划分训练集、验证集和测试集,使用训练集训练、验证集优化,最后用测试集的准确度和混淆矩阵进行模型评估。

表 3 模型超参数

机器学习模型	超参数项	设定值
KNN	邻点个数	1
	距离度量	Minkowski(三次)
	距离权重	反距离平方
SVM	核函数	二次
	框约束级别	987.900 5
	核尺度	自动
DT	多类方法	一对一
	最大分裂数	60
	分裂准则	基尼指数
集成模型(AdaBoost)	最大分裂数	13
	学习器数量	72
	学习率	0.945 16
集成模型(Bagging)	最大分裂数	17
	学习器数量	44
ANN	全连接层数	3
	全连接层大小	274/163/19
	激活函数	Tanh
	正则化强度	1.3612×10 ⁻⁷

3.2 模型测试结果

3.2.1 环代表原始数据

基于环代表数据集进行模型训练,所得测试集召回率见表4,准确度见表8,可知:

表 4 环代表原始数据测试集召回率 单位:%

地层类别	KNN	SVM	DT	AdaBoost	Bagging	ANN
1	100	100	100	100	50	100
2	96.4	98.2	92.7	94.5	90.9	96.4
3	100	100	100	100	100	100
4	91.7	91.7	9.7	91.7	91.7	100
5	100	100	100	100	100	100
6	75	100	100	100	100	100
7	83.3	83.3	100	83.3	100	100
8	75	100	100	100	100	100

(1)经由环代表数据集模型训练,各模型测试集的准确度在92.1%~97.8%之间,总体上准确度均较高,其中ANN的地层分类效果最强;

(2)样本中2号复合地层占比较多,分类效果较好,其召回率均在90%以上;其他复合地层由于数据量小,召回率波动较大,KNN、Bagging模型中1、6~8号复合地层分类效果较差,召回率最低达50%。

3.2.2 环代表SMOTE数据

基于环代表SMOTE增强数据集进行模型训练,所得测试集召回率见表5,准确度表见表8,可知:

(1)经由环代表增强数据集模型训练,各模型测

表 5 环代表 SMOTE 数据测试集召回率 单位: %

地层类别	KNN	SVM	DT	AdaBoost	Bagging	ANN
1	100	100	100	100	100	100
2	96.4	96.4	94.9	96.4	94.6	98.2
3	100	100	100	100	100	100
4	100	100	100	100	100	100
5	100	100	100	100	100	100
6	100	100	100	100	100	100
7	100	100	96.4	98.2	98.2	100
8	100	100	100	100	100	100

试集的准确度在 98.9%~99.8% 之间,总体上准确度均较高,其中 SVM 的地层分类效果最强;

(2)样本中 2 号复合地层占比较多,分类效果较好,其召回率均在 90% 以上;其他复合地层由于数据量小,召回率波动较大,DT 模型中 2 号复合地层分类效果较差,召回率最低达 95.9%。

3.2.3 时序原始数据

基于时序原始数据集进行模型训练,所得测试集召回率见表 6,准确度表见表 8,可知:

表 6 时序原始数据测试集召回率 单位: %

地层类别	KNN	SVM	DT	AdaBoost	Bagging	ANN
1	98.1	95.1	91.3	84.5	82.5	97.1
2	99.6	99.4	99	98.8	96.2	99.7
3	95.9	93.6	78.5	84.9	83.6	96.3
4	98.5	98.3	96.1	98	96.1	99
5	99	99.5	91.9	97.1	89	99.5
6	97.6	99.2	89.3	94.9	94.1	98.8
7	97.5	98.7	93.4	88	73.7	99.7
8	97.6	98.1	95.9	96.9	92.8	98.8

(1)经由时序数据集模型训练,各模型测试集的准确度在 93.9%~99.3% 之间,总体上准确度均较高,其中 ANN 模型地层分类效果最强;

(2)样本中 2 号复合地层占比较多,分类效果较好,其召回率均在 99% 以上;其他复合地层由于数据量小,召回率存在波动,DT、AdaBoost、Bagging 模型中 1、3、7 号复合地层分类效果较差,召回率最低达 73.7%。

3.2.4 时序 SMOTE 数据

基于时序 SMOTE 增强数据集进行模型训练,所得测试集召回率见表 7,准确度见表 8,可知:

(1)经由增强数据集模型训练,各模型测试集的准确度在 93.6%~99.9% 之间,总体上准确度均较高,其中 KNN、ANN 模型的地层分类效果最强;

(2)经由数据增强,样本中各复合地层占比相同,2 号复合地层其召回率有所下降,其他复合地层

表 7 时序 SMOTE 数据测试集召回率 单位: %

地层类别	KNN	SVM	DT	AdaBoost	Bagging	ANN
1	100	99.9	99.9	99.6	99.9	100
2	99.4	99	95	95.1	86.4	99.7
3	100	99.8	99.4	99.5	99.7	100
4	99.9	99.9	98.2	99	97.6	99.8
5	100	100	98.7	99.2	95.3	100
6	99.9	100	89.8	96.1	95.9	99.9
7	100	99.9	91.1	91.7	78.1	100
8	100	99.7	98.5	98.7	96.3	99.9

其召回率得到提升;而部分模型中召回率存在波动,AdaBoost、Bagging 模型中 2、7 号复合地层分类效果较差,召回率最低达到了 21.9%。

3.3 数据增强效果比较

统一不同数据集训练得到的准确度,见表 8,可知:

表 8 测试集准确度对比

测试集准确度/%	环数据	环增强数据	时序数据	时序增强数据
KNN	93.3	99.5	99.0	99.9
SVM	96.6	99.5	98.8	99.8
DT	94.4	98.9	96.6	96.3
Adaboost	94.4	99.3	97.1	97.3
Bagging	92.1	99.1	93.9	93.6
ANN	97.8	99.8	99.3	99.9

(1)随着环代表数据到环代表增强数据、时序数据、时序增强数据,数据量得到扩充,不同分类数据也更加均衡,模型准确度基本呈提升趋势,各模型地层分类效果均有提升,其中数据增强导致的均衡性对模型准确度提升更大;

(2)数据增强后,KNN、SVM、ANN 模型的分类能力显著提高,基本接近 100%,说明在样本量充足且不同样本平衡时 KNN、SVM、ANN 均能很好地完成地层分类。

而数据增强后,DT 的准确度提升较小,其对原少数类样本识别提高,但对 2 号复合地层样本识别能力下降,呈现模型的实际识别能力。

4 结论与展望

基于模型测试结果,总结如下:

(1)依据地层识别需要,基于时序数据进行数据集整理可以获得更大的数据量,保证机器学习模型的地层识别能力,也可用于实际施工中实时获取掘进数据并进行地层预测;

(2)通过 SMOTE 对数据进行处理,实现了不同

类间样本量的平衡,模型分类效果得到进一步提高,且消除了数据不平衡的影响后,可知KNN、SVM、ANN仍能很好地完成地层分类工作;且不论数据量是否充足、不同类数据量是否平衡,ANN均能较好地完成分类任务。

本文仍存在较大不足,所用复合地层分类参考了地勘资料,本身精确度不高;后续可在地层信息获取上进行连续、高精度的测量,以提升地层数据精度。

在数据增强方法上,后续可考虑使用不同的数据增强方法,比选出较好的方法,或是改进SMOTE中采样过程存在数据重复的缺点,以提升模型的分

参考文献:

- [1] 代洪波,季玉国.我国大直径盾构隧道数据统计及综合技术现状与展望[J].隧道建设(中英文),2022,42(5):757-783.
- [2] 高始军.超大直径长距离跨海高铁盾构隧道——广湛铁路湛江湾海底隧道项目介绍[J].隧道与地下工程灾害防治,2021,3(4):91-94.
- [3] 周昆.复杂地质超大直径越江盾构隧道——和燕路过江通道南段隧道工程A2标项目介绍[J].隧道与地下工程灾害防治,2020,2(1):105-107.
- [4] 薛爱民,李达,宋红星,等.应用振动噪声的频率谐振效应对地球进行成像[J].地质论评,2021,67(增刊1):47-48.
- [5] 刘宗毓.面向水质监测的动态电阻抗成像方法研究[D].天津:天津工业大学,2020.
- [6] 张腾飞.跨孔地震CT技术在大连地铁五号线岩溶段探测的应用研究[D].成都:西南交通大学,2019.
- [7] 张文俊.岩溶地层条件下的地铁跨孔电阻率CT超前预报方法研究[D].济南:山东大学,2014.
- [8] 杨刚,韦斯,张威.孔间电磁波CT在某水库防渗帷幕岩溶勘察中的应用[J].云南水力发电,2022,38(8):119-121.
- [9] 朱梦琦,朱合华,王昕,等.基于集成CART算法的TBM掘进参数与围岩等级预测[J].岩石力学与工程学报,2020,39(9):1860-1871.
- [10] 刘明阳,余宏淦,陶建峰,等.基于盾构机运行参数的局部切空间排列与Xgboost融合的地质类型识别[J].中南大学学报(自然科学版),2022,53(6):2080-2091.
- [11] HOU S, LIU Y, YANG Q. Real-time prediction of rock mass classification based on TBM operation big data and stacking technique of ensemble learning[J]. Journal of Rock Mechanics and Geotechnical Engineering, 2022, 14(1): 123-143.
- [12] CHAWLA N V, BOWYER K W, HALL L O, *et al.* SMOTE: synthetic minority over-sampling technique[J]. Journal of artificial intelligence research, 2002(16): 321-357.
- [13] COVER T, HART P. Nearest neighbor pattern classification[J]. IEEE transactions on information theory, 1967, 13(1): 21-27.
- [14] CORTES C, VAPNIK V. Support-vector networks[J]. Machine learning, 1995(20): 273-297.
- [15] 王远旭.基于机器学习算法的医疗数据处理与分析[D].厦门:厦门大学,2018.
- [16] QUINLAN J R. Induction of decision trees[J]. Machine learning, 1986(1): 81-106.
- [17] URGOFF P E. ID5: an incremental ID3[J]. Machine Learning Proceedings, 1988: 107-120.
- [18] QUINLAN J R. C4.5: programs for machine learning[M]. San Francisco: Morgan Kaufmann Publishers Inc, 2014.
- [19] GORDON A, BREIMAN L, FRIEDMAN J, *et al.* Classification and Regression Trees[J]. Biometrics, 1984, 40(3): 874.
- [20] SCHAPIRE R E. The boosting approach to machine learning: An overview[J]. Nonlinear estimation and classification, 2003: 149-171.
- [21] BREIMAN L. Bagging predictors[J]. Machine learning, 1996(24): 123-140.
- [22] MCCULLOCH W S, PITTS W. A logical calculus of the ideas immanent in nervous activity[J]. The bulletin of mathematical biophysics, 1943(5): 115-133.

(上接第265页)

- [15] 郭锋.抗震设计中有关场地的若干问题研究[D].武汉:华中科技大学,2010.
- [16] 吕红山,赵凤新.适用于中国场地分类的地震动反应谱放大系数[J].地震学报,2007(1):67-76.
- [17] MUKHERJEE S, GUPTA V K. Wavelet-based generation of spectrum-compatible time-histories[J]. Soil Dynamics and Earthquake Engineering, 2002, 22(9-12): 799-804.
- [18] SUAREZ L E, MONTEJO L A. Generacion de registros artificiales compatibles con un espectro de respuesta mediante la transformada wavelet[C]//The Proceedings of II Congreso Nacional de Ingenieria Sismica. Medellin: Beograd, 2003.
- [19] SUAREZ L E, MONTEJO L A. Generation of artificial earthquakes via the wavelet transform[J]. International Journal of Solids and Structures, 2005, 42(21-22): 5905-5919.