



南海天然气水合物智能识别方法与应用

田冬梅, 杨胜雄, 刘鑫, 李沅衡, 胡广, 曹荆亚, 周军明, 邓雨恬

Intelligent identification and application of gas hydrate in South China Sea

TIAN Dongmei, YANG Shengxiong, LIU Xin, LI Yuanheng, HU Guang, CAO Jingya, ZHOU Junming, and DENG Yutian

在线阅读 View online: <https://doi.org/10.16562/j.cnki.0256-1492.2024092401>

您可能感兴趣的其他文章

Articles you may be interested in

南海神狐海域天然气水合物微观赋存特征的超分辨率CT图像识别

Super-resolution CT image recognition of micro-occurrence characteristics of natural gas hydrates from Shenhu area in northern South China Sea

海洋地质与第四纪地质. 2024, 44(3): 149-159

南海神狐海域非均质性天然气水合物储层的分频反演

Frequency-divided inversion method of heterogenous natural gas hydrates reservoir in the Shenhu area, South China Sea

海洋地质与第四纪地质. 2020, 40(6): 106-120

基于岩石物理模拟与声学实验识别孔隙—裂隙充填型水合物

Identification of pore-filling and fracture-filling hydrate by petrophysical simulation and acoustic experiment

海洋地质与第四纪地质. 2020, 40(6): 208-218

南海神狐海域天然气水合物饱和度的数值模拟分析

Numerical modeling of gas hydrate saturation for the Shenhu area, South China Sea

海洋地质与第四纪地质. 2021, 41(2): 210-221

波形反演在天然气水合物中的应用研究进展

Application of full waveform inversion to gas hydrate research

海洋地质与第四纪地质. 2022, 42(4): 207-221

海洋天然气水合物储层特性及其资源量评价方法

Characteristics of marine gas hydrate reservoir and its resource evaluation methods

海洋地质与第四纪地质. 2021, 41(5): 44-57



关注微信公众号, 获得更多资讯信息

田冬梅, 杨胜雄, 刘鑫, 等. 南海天然气水合物智能识别方法与应用[J]. 海洋地质与第四纪地质, 2024, 44(6): 25-33.

TIAN Dongmei, YANG Shengxiong, LIU Xin, et al. Intelligent identification and application of gas hydrate in South China Sea[J]. Marine Geology & Quaternary Geology, 2024, 44(6): 25-33.

南海天然气水合物智能识别方法与应用

田冬梅¹, 杨胜雄¹, 刘鑫², 李沅衡¹, 胡广¹, 曹荆亚¹, 周军明¹, 邓雨恬¹

1. 南方海洋科学与工程广东省实验室(广州), 广州 511458

2. 广州海洋地质调查局, 广州 511458

摘要: 天然气水合物是一种重要的能源资源, 具有能量高、储量大、分布广和埋藏浅等优势。准确地从地层中识别出天然气水合物储层是应用天然气水合物资源的必要前提。本文围绕水合物勘探识别的难点问题, 结合海洋-地质-人工智能学科交叉技术, 以具有显示度的地球物理属性参数为基础, 研究并提出了有效的含水合物地层识别技术方法, 在中国南海东沙海域研究区进行了方法的验证, 选择了几种较为常用的机器学习算法, 例如随机森林、Bagging、AdaBoost、和最近邻(KNN)算法, 对水合物变化灵敏度较高的纵波速度和密度属性进行数据分析, 通过训练优化不同算法模型参数, 对比不同算法模型的识别分类效果。结果表明, 这几种算法都能够较好地对地层中是否含有水合物进行区分, 其中KNN性能最好, 表明基于机器学习手段能够提高天然气水合物的识别精度和准确性。

关键词: 水合物; 识别; 机器学习; 地震属性; 南海

中图分类号: P714

文献标识码: A

DOI: 10.16562/j.cnki.0256-1492.2024092401

Intelligent identification and application of gas hydrate in South China Sea

TIAN Dongmei¹, YANG Shengxiong¹, LIU Xin², LI Yuanheng¹, HU Guang¹, CAO Jingya¹, ZHOU Junming¹, DENG Yutian¹

1. Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, China

2. Guangzhou Marine Geological Survey, Guangzhou 511458, China

Abstract: Gas hydrate is an important ideal energy source, with advantages of high energy, large reserves, wide distribution, and shallow burial. Accurate identification of gas hydrate reservoirs and estimation of hydrate saturation are the prerequisite for the application of gas hydrate resources. This study focuses on the difficult issues of hydrate identification, combining the interdisciplinary technologies of oceanology, geology, and artificial intelligence. Effective methods of hydrate-bearing strata identification were proposed based on the geophysical attributes, and verified in the Dongsha area of South China Sea. Machine-learning algorithms were used to analyze whether the sediment contains gas hydrates. Several commonly used machine-learning algorithms were selected, including random forest, Bagging, AdaBoost, and KNN; and data were analyzed based on the P-wave velocity and density attributes that are more sensitive to hydrate existence. The parameters of different algorithms were trained and optimized, and the effects of different algorithms on the identification and classification were compared. All these algorithms could do good on whether there is hydrate in the sediment, of which KNN algorithm was shown the best. Therefore, machine-learning-based methods could improve the identification accuracy of gas hydrate.

Key words: hydrate; identification; machine learning; seismic attributes; South China Sea

天然气水合物是一种似冰状固体化合物, 由水分子和气体分子在低温高压条件下生成, 具有可燃性, 因此也被称为“可燃冰”^[1]。水合物的燃烧效率极高, 并且污染小, 被视为是未来化石燃料的替代

能源, 美国、日本、韩国、印度等国家均已制定国家水合物勘探研究计划^[2-4]。中国在近年来水合物勘探开发上也取得了较为显著的成效, 2017年神狐海域的成功试采标志着中国成为全球第一个实现了

资助项目: 国家自然科学基金国家重大科研仪器研制项目“海底地震与电磁同步探测系统关键技术及验证样机”(42327901); 国家自然科学基金项目“南海北部高富集天然气水合物储层特征与成藏控制机理研究”(U2244224); 广州市基础与应用基础研究项目“基于地震属性海域天然气水合物识别方法研究”(2023A04J0916)

作者简介: 田冬梅(1995—), 女, 博士, 主要从事海洋地球物理、天然气水合物识别计算研究, E-mail: dongmeitian@gmlab.ac.cn

通讯作者: 刘鑫(1983—), 男, 硕士, 高级工程师, 主要从事海洋地质环境相关研究, E-mail: lxin05@gmlab.ac.cn

收稿日期: 2024-09-24; **改回日期:** 2024-12-23. 张现荣编辑

在海域可燃冰开采中获得连续稳定产气的国家,同年,水合物被列为中国第173个矿种^[5-9]。水合物在全球分布的范围十分广泛(图1),主要存在于海底沉积层和永久冻土层中。据估算,世界可燃冰总资源量大约相当于全球已知煤、石油和天然气总资源量的2倍,储量十分可观^[10]。然而在复杂海底地质环境下勘探水合物具有相当大的困难,因此,准确勘探识别水合物和估算水合物储量一直以来都是国内外的研究热点问题。

地震勘探技术一直以来都是水合物识别与评价的重要地球物理方法^[11]。早期通过识别地震剖面上的似海底反射层(BSR)来判断地层中的天然气水合物稳定带底界,BSR通常呈现出与海底平行、极性相反、高振幅强反射、与沉积地层相交等特征,但地震剖面上的BSR与水合物分布并不是——对应的^[12-14]。由于水合物的存在会导致地层的速度升高,所以有学者提出用速度异常的方法来识别水合物,但该方法仍有一定的问题,如有其他矿物充填对应的测井曲线上也有速度异常升高,因此速度异常的方法也存在指示不明的问题。后来,Liu等^[15]提出利用纵横波速度增量比的方法来识别地层中的水合物,在布莱克海台ODP164航次的实际数据中取得较好的应用效果。Tian和Liu^[16]发现通过纵横波速度交汇图的方法也可以区分出含水合物地层,以及用纵横波速度增量交汇图的方法估算水合物饱和度范围,并在俄勒冈水合物海岭的实际数据

中进行验证,结果与实际情况相一致。传统的水合物识别方法往往需要综合的判断分析,处理解释的时间长,水合物饱和度计算准确率低,亟待提出一套高效准确的水合物识别计算方法。随着人工智能的高速发展,机器学习作为人工智能领域的核心算法,分为监督学习、非监督学习和深度学习,无论是在理论研究还是实践应用方面都取得了巨大的突破^[17-18]。目前通过机器学习的方法也解决了很多地质问题,例如,王迪等^[19]利用深度学习方法定量地进行了烃类预测;Bai和Tan^[20]采用机器学习方法进行储层参数预测并取得较好结果;陈钢花等^[21]采用了双向长短时记忆神经网络有效解决了致密滩坝的储层岩性识别问题;Chen等^[22]利用径向基函数神经网络对神狐海域水合物饱和度进行预测得到了较好的应用效果;Zhu等^[23]基于机器学习方法对水合物的赋存类型进行了识别并估算了水合物饱和度。因此,本文考虑在中国南海海域借助人工智能的手段来进一步高效准确识别水合物资源。

1 研究方法

本研究主要通过几种不同算法,包括AdaBoost算法、随机森林算法、Bagging算法、和最近邻(KNN)算法,对测井数据中的纵波速度和密度数据进行模型训练和测试,从而区分含水合物地层与非

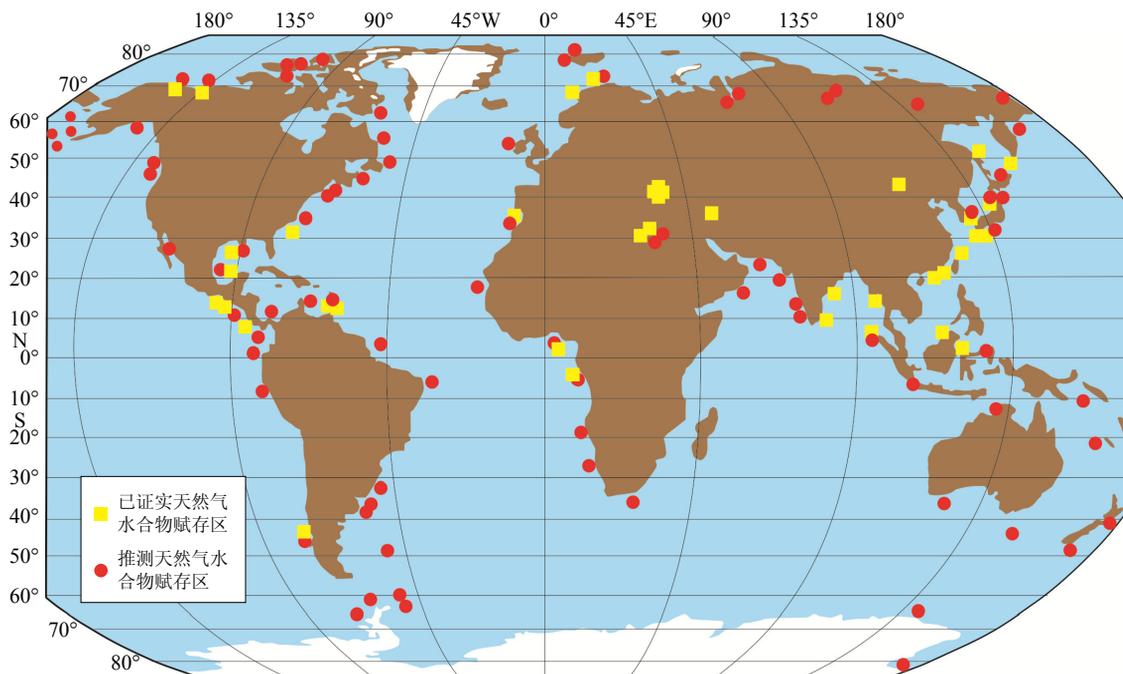


图1 全球天然气水合物分布图(据文献[3]修改)

Fig.1 Global gas hydrate distribution (modified from reference [3])

含水合物地层。

1.1 算法选择

1.1.1 AdaBoost 算法

AdaBoost 是基于 Boosting 的一种集成学习算法。该算法的核心思想是通过调整样本分布和弱分类器权值, 将一系列弱分类器集成为一个分类精度高的强分类器, 改变分类器在已有样本分布上的优势, 提高泛化能力(模型在未曾见过的数据上的表现能力, 即模型对于新的数据的适应能力)。AdaBoost 算法通过改变训练集中各样本的权重实现迭代过程, 根据每次训练集中各样本是否分类正确及上次总体分类的准确率, 综合确定各样本的权重, 把修改过权重的新数据集送给下层分类器进行训练, 并将每次训练所得分类器融合起来, 形成最终的决策分类器^[24]。

1.1.2 随机森林算法

随机森林(Random Forest, 简称 RF)是一种集成多棵决策树的有监督学习算法。RF 以决策树为基学习器构建 Bagging 集成的基础上, 进一步在决策树的训练过程中引入了随机属性选择^[25]。随机森林使用 bootstrap 方法构建 n 个训练集, 每个训练集对应生成一个决策树, 总体就有 n 个决策树, 因为每个决策树的数据集都不相同, 所以每棵树又有少量区别。最后对所有的决策树的预测结果取平均减少预测的方差, 提高在测试集上的性能表现。该算法具有简单、容易实现、计算开销小等特点, 在很多现实任务中展现出强大的性能。

1.1.3 Bagging 算法

Bagging 算法是一种用来提高弱分类器准确率的集成方法, 该算法对训练集实施有放回的随机样本重新采样, 形成多个与训练集规模相近但各不相同的训练子集, 在此基础上形成多个基分类器, 在多个基分类器的基础上使用多数投票原则^[26-27]。Bagging 方法具有较大的随机性, 但当采样次数足够多时会生成足够丰富的基本模型, 可以得到十分精确的结果。通常在减小方差方面 Bagging 方法效果明显, 因此 Bagging 方法对训练集微小变化敏感的模型可以提高其稳定性。

1.1.4 KNN 算法

KNN 算法是一种有监督的机器学习算法, 通过遍历训练集样本, 计算测试数据与训练样本点属性间的距离, 找到与测试数据最近的 K 个样本点, 采取少数服从多数的原则确定测试数据的类别^[28]。KNN 算法尤其适用于训练样本量较大且样本之间

的差异性相对显著的海量数据分类, 优势明显, 准确度非常高。

1.2 评价指标

为了更好地确定水合物地层识别模型的预测精度和泛化能力, 利用准确率、精确率、召回率、 F_1 分数、ROC 曲线来对分类模型进行性能评估。

识别水合物的过程是一个二分类问题, 结果只有真和假两种。一般用 T(True)表示预测结果正确(与实际相符), F(False)表示预测结果不正确(与实际不符), P(Positive)表示真实的正样本, N(Negative)表示真实的负样本, 因此预测结果分为四种, 如表 1 所示:

表 1 评价结果指标意义
Table 1 Significance of evaluation results indicators

	预测为真	预测为负
实际为真	TP	FN
实际为负	FP	TN

准确率(Accuracy)用来衡量分类正确的比例, 虽然准确率可以判断总的正确率, 但在样本不平衡的情况下, 并不能作为很好的指标来衡量分类结果:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

精确率(Precision)是指预测为正类样本数中实际为正类的比例:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

召回率(Recall)是度量有多少个正类样本数被判定为正类:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

精确率与召回率可能是相互矛盾的, 因此需要对二者进行综合考虑, 因此引入 F 分数, 定义如下:

$$F_{\beta} = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}} \quad (4)$$

当式中 $\beta=1$ 时, F 分数为 F_1 分数。 F_1 是精确率和召回率的调和均值。当式中 $\beta>1$ 时, 召回率的权重高于精确率。相反, 当 $\beta<1$ 时精确率的权重高于召回率。

接受者操作特征(ROC)曲线是反映模型敏感性与特异性连续变化的综合指标。该曲线的横坐标为预测为正但实际为负的样本占有所有负例样本的比例; 曲线的纵坐标为预测为正且实际为正的样本占有所有正例样本的比例。

$$TPR = \frac{TP}{TP+FN} \quad (5)$$

$$FPR = \frac{FP}{FP+TN} \quad (6)$$

ROC 曲线下与坐标轴围成的面积(AUC)值为 ROC 曲线下与坐标轴围成的面积,是衡量算法优劣的一种性能指标,AUC 值越接近 1,则算法越优秀。

2 实际应用

2.1 研究区概况

中国南海东沙海域位于南海东北部陆坡区(图 2),该区域地理位置独特,海域沉积活动和构造运动相对复杂,具有被动大陆边缘特点,是天然气水合物生成的有利富集区,海底存在甲烷渗漏现象,冷泉发育活跃。2013 年,广州海洋地质调查局开展了第二次天然气水合物勘查钻探项目(GMGS2),在该研究区进行了一系列的地质、地球物理、地球化学等方面的调查,并获取了大量的块状、脉状、层状、分散状及结核状等多种不同类型的水合物样品,孔隙型水合物和裂隙型水合物也被证实存在于研究区内,此次勘查在中国海域天然气水合物勘探中具有重要意义。在 GMGS2-05、GMGS2-08 和 GMGS2-16 钻井位置分别采集到了测井数据,因此本文采用这 3 个站位处的数据用来训练和测试水合物智能识别方法。

前期勘探已在 05、08 和 16 站位发现有天然气水合物,根据随钻测井获得了大量的测井数据,图 3 分别给出了 05、08 和 16 站位的纵波速度和密度测

井曲线,蓝色区域为含水合物地层,本论文主要利用 05、08 和 16 站位测井数据作为学习数据集进行实验分析。在 05 站位,在海底以下约 193~206.5 m 发现了大量的孔隙型水合物,并伴随有高速度和高电阻率异常特征;在 08 站位,发现有两层裂隙型水合物分别在海底以下约 9~22 和 65~98 m;在 16 站位,裂隙型水合物出现在海底以下 10~20 m,孔隙型水合物 BSR 在海底以下约 189~198 m 位置^[27]。

2.2 模型调参和训练

在机器学习模型的训练过程中,为了使模型具有更好的泛化能力,将数据集分割为训练集和测试集。本实验选取 05、08 和 16 站位的测井数据,以深度位置计数分别取速度和密度点各 4 247 个建立水合物识别样本数据库(图 3),其中抽取 75% 作为训练样本,然后将剩余 25% 数据点作为测试样本。训练集用于训练模型,测试集用于检验模型的泛化能力。从常规测井系列中优选对岩性变化敏感的纵波速度和密度作为水合物识别模型的输入特征,含水合物地层和非含水合物地层两种类别作为输出特征。

此外,合适的监督模型参数是提高分类器预测精度的关键。为了使上述机器学习监督模型在其它未知的例子上也表现出较好的分类性能,还需要调整模型在训练过程之前所设置的参数(即超参数)获取适当的参数值。但是,训练后的模型不能过于拟合训练集,否则会失去对测试数据的预测能力,导致出现过拟合。可以通过交叉验证结果生成的验证分数选择不引起模型过拟合或欠拟合的最优参数集。网格搜索是模型参数调试的常用方法,

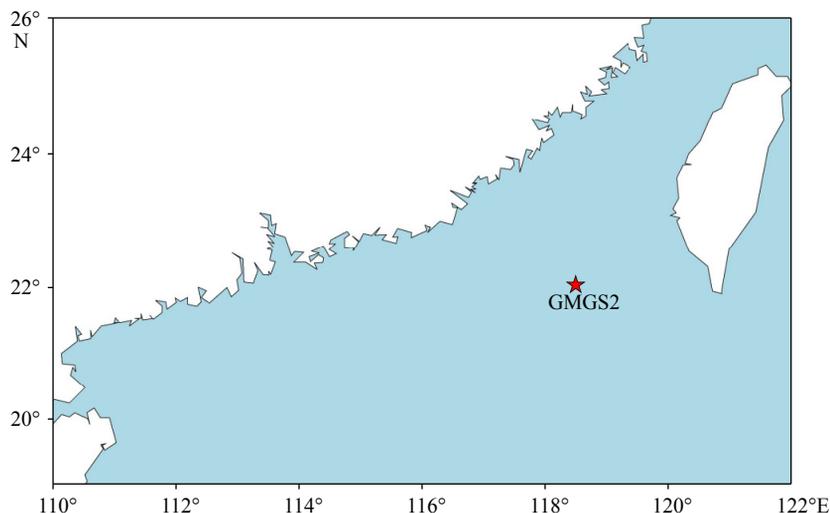


图 2 研究区位置

Fig.2 Location of the study area

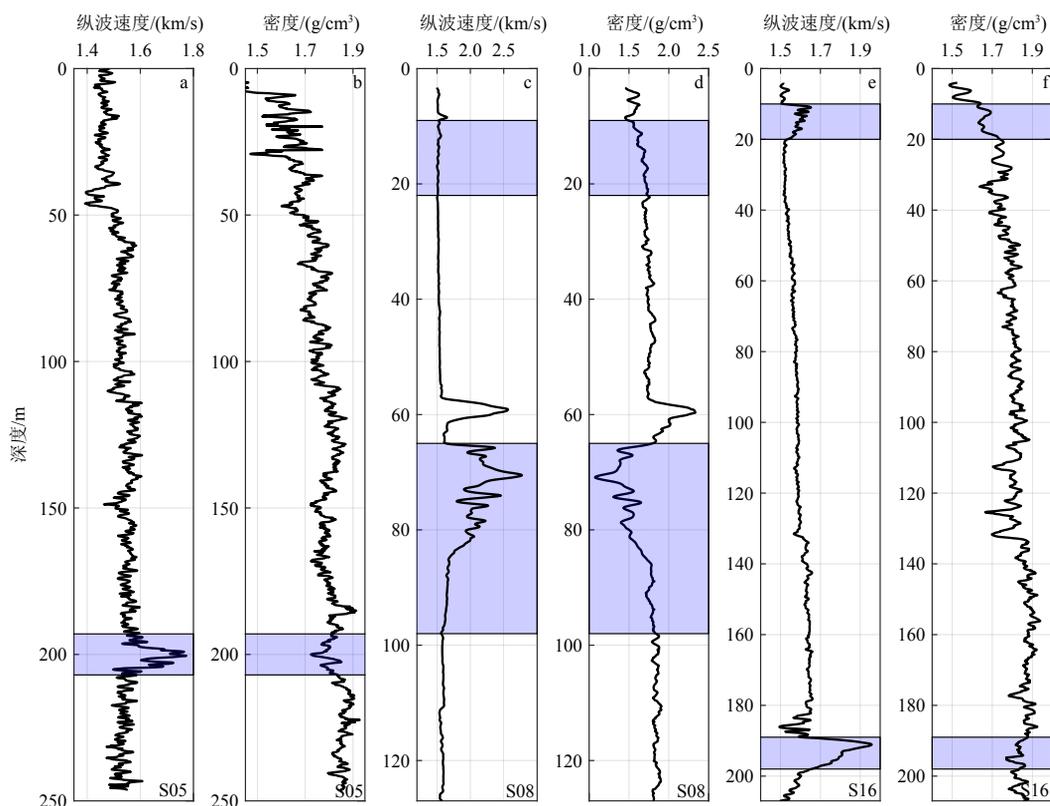


图 3 GMGS2-05、GMGS2-08 和 GMGS2-16 站位纵波速度和密度测井曲线
天然气水合物区域用蓝色标记。

Fig.3 Logging curves of P-wave velocity and density at sites GMGS2-05, GMGS2-08 and GMGS2-16
The gas hydrate areas are marked in blue.

通过选择合理的参数范围网格搜索能够找出模型的最优参数组合。在本案例研究中, 利用网格搜索对模型的超参数进行调试, 试验不同的超参数范围, 使用交叉验证来评估超参数值的所有可能组合, 以获得泛化能力较好的改进模型。本文以准确度作为衡量分类性能的指标。对于每个模型, 需要调优的参数见表 2。

3 结果分析

应用不同机器学习模型对南海东沙海域的 05、08 和 16 站位的测井数据样本进行水合物识别分析, 首先将数据样本中的已知为含水合物地层深度位置的数据标记为 1, 非含水合物地层深度位置的数据标记为 0, 识别结果如下:

(1) AdaBoost 算法的分类效果

在训练 AdaBoost 分类器时, 该算法首先训练一个基础分类器, 然后进行集成, 先选择决策树作为基础分类器, 也就是弱学习器。其中基础分类器决策树算法的最大深度(max depth)为 1, 弱学习器的最大迭代次数(n estimators)为 15, 学习率为 0.5,

表 2 每个模型的超参数调整

Table 2 Hyperparameter tuning for each algorithm

监督模型	超参数	调整范围	最优参数
AdaBoost	N estimators	5~25	15
	Learning rate	0.1~0.9	0.5
RF	max leaf nodes	5~45	10
	n estimators	0~150	30
Bagging	n estimators	10~500	100
	Max samples	50~500	150
KNN	N neighbors	5~200	10
	weights	uniform/distance	uniform

Scikit-Learn 默认使用 SAMME.R, 对样本集分类效果作为弱学习器权重, 由于 SAMME.R 使用了概率度量的连续值, 迭代一般比 SAMME 快。调整参数后的模型对测试数据进行分类, 我们得到准确率、精确率、召回率和 F_1 分数分别是 0.9171、0.8625、0.4726 和 0.6106。此外, AUC 值为 0.9146。用 AdaBoost 算法对 05、08 和 16 站位纵波速度和密度进行分类

训练和测试,结果如图4,其中红色点表示含水合物地层纵波速度和密度交点,蓝色点表示非含水合物地层纵波速度和密度交点,中间的深绿色交线为算法的决策边界,决策边界左边为含水合物地层(绿色区域),决策边界右边为非含水合物地层(黄色区域)。该算法较好地水合物与非水合物地层进行分类,算法稳定可靠。

(2) 随机森林算法分类效果

随机森林算法将多个决策树组合在一起,随机选取不同的数据集,以便输出相似但不相同的模型结果,再将所有决策树结果整合在一起,作为输出结果,而这一训练方式,意味着很难过拟合,并且对噪音不敏感。通过网格搜索调整该模型超参数,确定决策树的个数(n estimators)为30,通过限制最大叶子节点数,防止过拟合,最大叶节点数设置为10。调整参数后的模型对测试数据进行分类,得到准确率、精确率、召回率和 F_1 分数分别为0.9388, 0.8857、0.6370和0.7410,此外,AUC值为0.9278。用随机森林算法对05、08和16站位纵波速度和密度进行分类训练和测试,结果如图5,其中红色点表示含水合物地层纵横波速度交点,蓝色点表示非含水合物地层纵横波速度交点,中间的深绿色交线为算法的决策边界,决策边界左边为含水合物地层(绿色区域),决策边界右边为非含水合物地层(黄色区域),该算法在水合物的识别问题上表现出很好的分类性能,算法稳定可靠,有很强的抗干扰能力、抗过拟合和泛化能力。

(3) Bagging 算法分类性能

在本算法中我们训练多个决策树去组成 bagging 集成算法,经网格搜索调参后,确定该算法包含100个决策树分类器(n estimators=100),每次从训练集中随机采样150个训练实例(\max samples=150)进行训练,然后放回。一旦预测器训练完成,集成就可以通过简单地聚合所有预测器的预测来对新实例做出预测。最后通过该模型对测试数据的预测结果,得到准确率、精确率、召回率和 F_1 分数分别为0.9341、0.8725、0.6096和0.7177,此外,AUC值为0.9449。用 Bagging 算法对05、08和16站位纵波速度和密度进行分类训练和测试,结果如图6,其中红色点表示含水合物地层纵横波速度交点,蓝色点表示非含水合物地层纵横波速度交点,中间的深绿色交线为算法的决策边界,决策边界左边为含水合物地层(绿色区域),决策边界右边为非含水合物地层(黄色区域),该算法性能依赖于基分类器(本模型中使用决策树)的稳定性,能够较好地将水合物与非水合物地层进行区分。

(4) KNN 算法分类性能

在KNN算法中我们主要调整 k 值和选择距离的权重方式, k 值越小越容易过拟合,当 $k=1$ 时,这时只根据单个近邻进行预测,如果离目标点最近的一个点是噪音,就会出错,此时模型复杂度高,稳健性低,决策边界崎岖。但是如果 k 值取的过大,这时与目标点较远的样本点也会对预测起作用,就会导致欠拟合,此时模型变得简单,决策边界变平

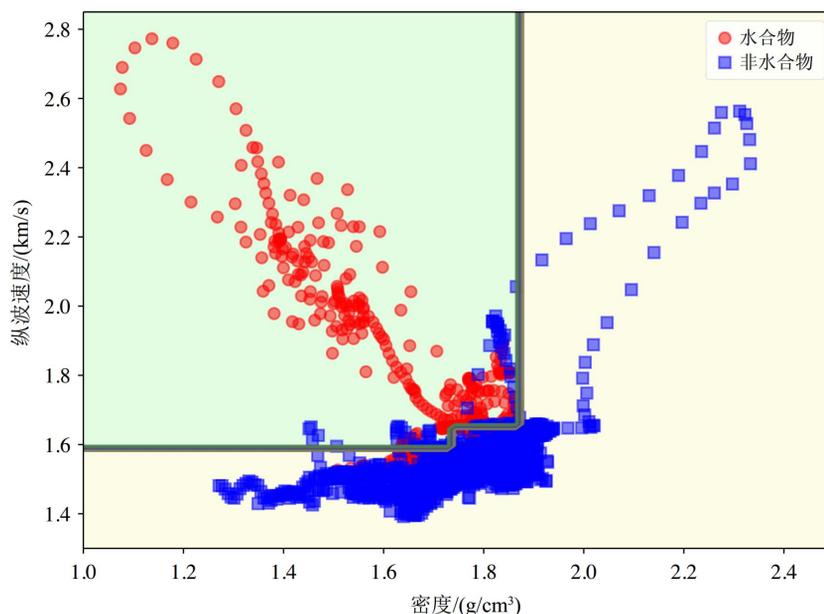


图4 AdaBoost 算法对 GMGS2-05、GMGS2-08 和 GMGS2-16 纵波速度和密度数据训练识别水合物结果

Fig.4 The hydrate identification results using AdaBoost algorithm based on the P-wave velocity and density data at sites GMGS2-05, GMGS2-08, and GMGS2-16

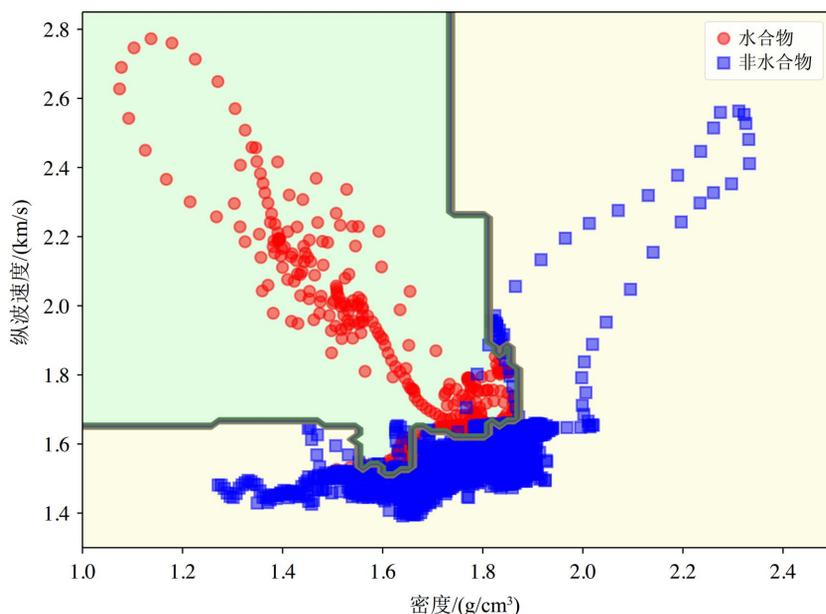


图 5 随机森林算法对 GMGS2-05、GMGS2-08 和 GMGS2-16 站位纵波速度和密度数据训练识别水合物结果
 Fig.5 The hydrate identification results using by random forest algorithm based on the P-wave velocity and density data at sites GMGS2-05, GMGS2-08 and GMGS2-16

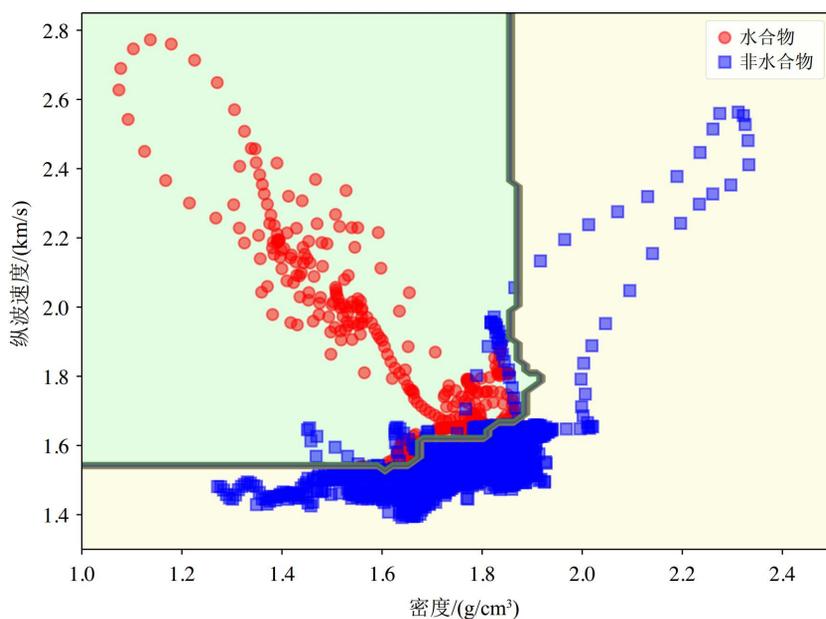


图 6 Bagging 算法对 GMGS2-05、GMGS2-08 和 GMGS2-16 站位纵波速度和密度数据训练识别水合物结果
 Fig.6 The hydrate identification result using Bagging algorithm based on the P-wave velocity and density data at sites GMGS2-05, GMGS2-08 and GMGS2-16

滑。在 sklearn 中 k 近邻算法还有一个 weights 参数, 其默认值为 uniform, 此时是不考虑距离权重的, 当 weights=distance 时, sklearn 就会考虑距离的权重, 权重和距离成反比, 距离预测目标越近具有越高的权重。通过网格搜索最后确定模型的 k 值为 200, weights=distance, 用该模型对测试数据进行分类, 我们得到准确率、精确率、召回率和 F_1 分数分别为 0.9379、0.8636、0.6507 和 0.7422, 此外, AUC

值为 0.9580 用 KNN 算法对 05、08 和 16 站位纵波速度和密度进行分类训练和测试, 结果如图 7, 其中红色点表示含水合物地层纵横波速度交点, 蓝色点表示非含水合物地层纵横波速度交点, 中间的深绿色交线为算法的决策边界, 决策边界左边为含水合物地层(绿色区域), 决策边界右边为非含水合物地层(黄色区域), 该算法能够很好地将水合物与非水合物地层进行分类, 准确率高, AUC 值在这 4 种算

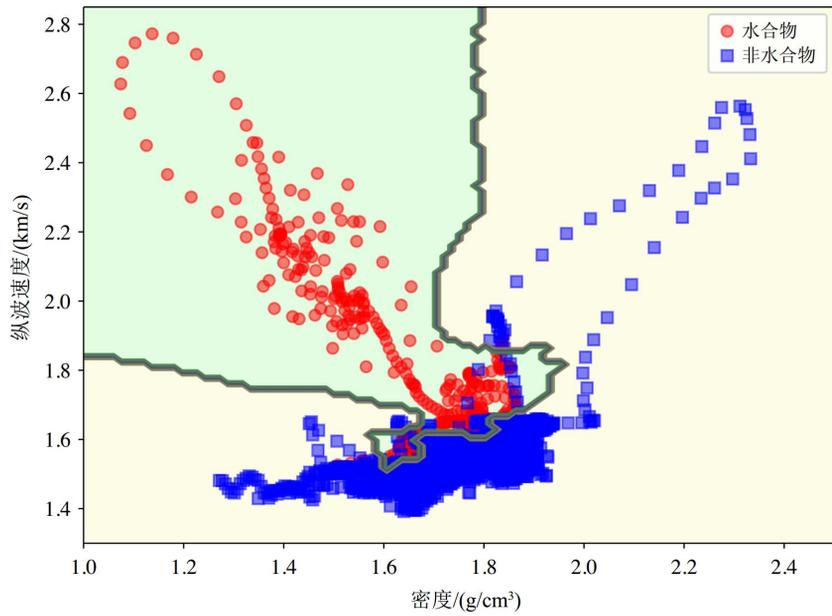


图 7 KNN 算法对 GMGS2-05、GMGS2-08 和 GMGS2-16 站位纵波速度和密度数据训练识别水合物结果
 Fig.7 The hydrate identification result using KNN algorithm based on the P-wave velocity and density data at sites GMGS2-05, GMGS2-08 and GMGS2-16

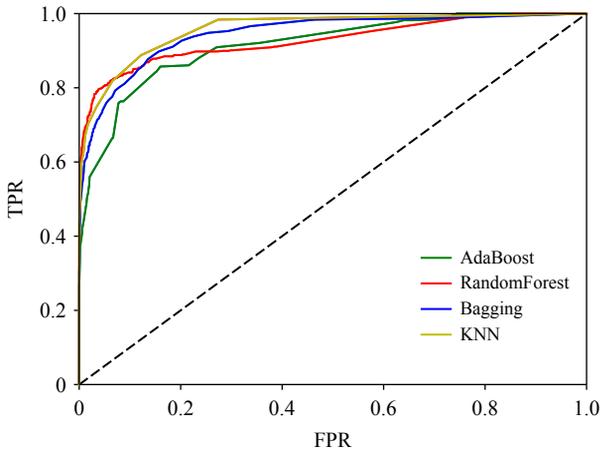


图 8 不同算法模型的 ROC 曲线
 Fig.8 The ROC curves of different algorithms

表 3 各算法评价指标计算结果

Table 3 Calculation results of evaluation indicators of each algorithm

Supervision algorithm	Accuracy	Precision	Recall	F_1 值	AUC
AdaBoost	0.9171	0.8625	0.4726	0.6106	0.9146
Random forest	0.9388	0.8857	0.6370	0.7410	0.9278
Bagging	0.9341	0.8725	0.6096	0.7177	0.9449
KNN	0.9379	0.8636	0.6507	0.7422	0.9580

法中最高(图 8),因此该算法在该地区的水合物分类中所表现的性能最好。

4 结论与建议

本文通过 4 种机器学习算法对南海东沙海域 GMGS2-05、GMGS2-08 和 GMGS2-16 站位的纵波速度和密度数据进行训练,不断调整超参数选值确定最优模型来对水合物和非水合物点分类识别,引入准确率、精确率、召回率和 F_1 分数分别是对识别结果进行评价(表 3),此外用 AUC 值对模型性能进行更直观的评价。识别结果表明,①与其他机器学习算法相比,随机森林算法在测试水合物分类过程中的准确率、精确率和最高,KNN 算法的召回率和 F_1 分数最高。② KNN 算法的 AUC 值是最接近 1 的,也就是说该算法性能相对来说比较优秀。③ 这些算法都能较好地通过纵波速度和密度属性数据来识别水合物,为天然气水合物的高效准确识别提供了技术支撑。利用人工智能技术能够有效解决水合物研究中面临的难点问题,今后仍需在算法和输入地震属性特征方面进行更加深入的研究,不断优化模型算法,当前水合物在自然界中的赋存类型多样,后续值得应用人工智能手段进一步细化研究区分水合物的赋存状态的智能识别方法。

参考文献 (References)

[1] Sloan E D Jr, Koh C A. Clathrate Hydrates of Natural Gases[M]. 3rd

- ed. Boca Raton: CRC Press, 2007.
- [2] Chong Z R, Yang S H B, Babu P, et al. Review of natural gas hydrates as an energy resource: Prospects and challenges[J]. *Applied Energy*, 2016, 162: 1633-1652.
- [3] Collett T S, Johnson A H, Knapp C C, et al. Natural gas hydrates—A review[M]//Collett T, Johnson A, Knapp C, et al. Natural Gas Hydrates—Energy Resource Potential and Associated Geologic Hazards. American Association of Petroleum Geologists, 2009, 89: 146-219.
- [4] Boswell R, Hancock S, Yamamoto K, et al. Natural gas hydrates: Status of potential as an energy resource[M]//Letcher T M. Future Energy: Improved, Sustainable and Clean Options for our Planet. 3rd ed. New York: Elsevier, 2020: 111-135.
- [5] Yang S X, Liang J Q, Lei Y, et al. GMGS4 gas hydrate drilling expedition in the South China Sea[J]. *Fire in the Ice*, 2017, 17(1): 7-11.
- [6] 雷裕红, 宋颖睿, 张立宽, 等. 海洋天然气水合物成藏系统研究进展及发展方向 [J]. *石油学报*, 2021, 42(6): 801-820. [LEI Yuhong, SONG Yingrui, ZHANG Likuan, et al. Research progress and development direction of reservoir-forming system of marine gas hydrates[J]. *Acta Petrolei Sinica*, 2021, 42(6): 801-820.]
- [7] Zhang G X, Liang J Q, Lu J A, et al. Geological features, controlling factors and potential prospects of the gas hydrate occurrence in the east part of the Pearl River Mouth Basin, South China Sea[J]. *Marine and Petroleum Geology*, 2015, 67: 356-367.
- [8] Liang J Q, Zhang W, Lu J A, et al. Geological occurrence and accumulation mechanism of natural gas hydrates in the eastern Qiongdongnan Basin of the South China Sea: Insights from Site GMGS5-W9-2018[J]. *Marine Geology*, 2019, 418: 106042.
- [9] Sha Z B, Liang J Q, Zhang G X, et al. A seepage gas hydrate system in northern South China Sea: Seismic and well log interpretations[J]. *Marine Geology*, 2015, 366: 69-78.
- [10] Milkov A V. Global estimates of hydrate-bound gas in marine sediments: how much is really out there?[J]. *Earth-Science Reviews*, 2004, 66(3-4): 183-197.
- [11] 勾丽敏, 张金华, 王嘉玮. 海洋天然气水合物地震识别方法研究进展 [J]. *地球物理学进展*, 2017, 32(6): 2626-2635. [GOU Limin, ZHANG Jinhua, WANG Jiawei. Progress in seismic identification approach of marine gas hydrate[J]. *Progress in Geophysics*, 2017, 32(6): 2626-2635.]
- [12] 宋海斌, 张岭, 江为为, 等. 海洋天然气水合物的地球物理研究 (III): 似海底反射 [J]. *地球物理学进展*, 2003, 18(2): 182-187. [SONG Haibin, ZHANG Ling, JIANG Weiwei, et al. Geophysical researches on marine gas hydrates (III): bottom simulating reflections[J]. *Progress in Geophysics*, 2003, 18(2): 182-187.]
- [13] White R S. Gas hydrate layers trapping free gas in the Gulf of Oman[J]. *Earth and Planetary Science Letters*, 1979, 42(1): 114-120.
- [14] Stoll R D, Ewing J, Bryan G M. Anomalous wave velocities in sediments containing gas hydrates[J]. *Journal of Geophysical Research*, 1971, 76(8): 2090-2094.
- [15] Liu X W, He J, Sun Q L. Gas hydrate identification from $\Delta V_p/\Delta V_s[C]$ //Beijing 2009 International Geophysical Conference & Exposition. Beijing: Society of Exploration Geophysicists, 2009: 150.
- [16] Tian D M, Liu X W. Identification of gas hydrate based on velocity cross plot analysis[J]. *Marine Geophysical Research*, 2021, 42(2): 11.
- [17] Chen J Z, You J C, Wei J T, et al. Interpreting XGBoost predictions for shear-wave velocity using SHAP: Insights into gas hydrate morphology and saturation[J]. *Fuel*, 2024, 364: 131145.
- [18] Luo Y L, Zhang G L, Liang G W, et al. Limited-label multiscale deep-learning multihorizon tracking[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5918814.
- [19] 王迪, 张益明, 张繁昌, 等. 利用先验信息约束的深度学习定量预测致密砂岩“甜点”[J]. *石油地球物理勘探*, 2023, 58(1): 65-74. [WANG Di, ZHANG Yiming, ZHANG Fanchang, et al. Quantitative prediction of tight sandstone sweet spots based on deep learning method with prior information constraints[J]. *Oil Geophysical Prospecting*, 2023, 58(1): 65-74.]
- [20] Bai Y, Tan M J. Dynamic committee machine with fuzzy-c-means clustering for total organic carbon content prediction from wireline logs[J]. *Computers & Geosciences*, 2021, 146: 104626.
- [21] 陈钢花, 张寓侠, 王军, 等. 双向长短时记忆神经网络在滩坝砂储层岩性识别中的应用 [J]. *测井技术*, 2023, 47(3): 319-325. [CHEN Ganghua, ZHANG Yuxia, WANG Jun, et al. Application of BiLSTM in lithology identification of beach-bar sand reservoir[J]. *Well Logging Technology*, 2023, 47(3): 319-325.]
- [22] Chen Y J, Dunn K J, Liu X W, et al. New method for estimating gas hydrate saturation in the Shenhu area[J]. *Geophysics*, 2014, 79(5): IM11-IM22.
- [23] Zhu X Y, Liu T, Ma S, et al. Morphology identification of gas hydrate based on a machine learning method and its applications on saturation estimation[J]. *Geophysical Journal International*, 2023, 234(2): 1307-1325.
- [24] 杨笑, 王志章, 周子勇, 等. 基于参数优化 AdaBoost 算法的酸性火山岩岩性分类 [J]. *石油学报*, 2019, 40(4): 457-467. [YANG Xiao, WANG Zhizhang, ZHOU Ziyong, et al. Lithology classification of acidic volcanic rocks based on parameter-optimized AdaBoost algorithm[J]. *Acta Petrolei Sinica*, 2019, 40(4): 457-467.]
- [25] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5-23.
- [26] 任涛, 林梦楠, 陈宏峰, 等. 基于 Bagging 集成学习算法的地震事件性质识别分类 [J]. *地球物理学报*, 2019, 62(1): 383-392. [REN Tao, LIN Mengnan, CHEN Hongfeng, et al. Seismic event classification based on Bagging ensemble learning algorithm[J]. *Chinese Journal of Geophysics*, 2019, 62(1): 383-392.]
- [27] Bayar B, Stamm M C. Constrained convolutional neural networks: a new approach towards general purpose image manipulation detection[J]. *IEEE Transactions on Information Forensics and Security*, 2018, 13(11): 2691-2706.
- [28] Nguyen B, Morell C, De Baets B. Large-scale distance metric learning for k-nearest neighbors regression[J]. *Neurocomputing*, 2016, 214: 805-814.