

引用本文: 杨子杰, 陈冬霞, 王翹楚, 等. 基于人工神经网络方法预测油气资源丰度——以渤海湾盆地东濮凹陷文留地区古近系沙河街组三段为例[J]. 石油实验地质, 2024, 46(2): 428-440. DOI: 10.11781/sysydz202402428.

YANG Zijie, CHEN Dongxia, WANG Qiaochu, et al. Prediction of petroleum resource abundance based on artificial neural network method: a case study of third member of Paleogene Shahejie Formation in Wenliu area of Dongpu Sag, Bohai Bay Basin[J]. Petroleum Geology & Experiment, 2024, 46(2): 428-440. DOI: 10.11781/sysydz202402428.

基于人工神经网络方法预测油气资源丰度

——以渤海湾盆地东濮凹陷文留地区古近系沙河街组三段为例

杨子杰^{1,2}, 陈冬霞^{1,2}, 王翹楚^{1,2}, 王福伟^{1,2}, 李莎^{1,2},
田梓葉^{1,2}, 陈淑敏^{1,2}, 张婉蓉^{1,2}, 姚东升^{1,2}, 王昱超^{1,2}

1. 中国石油大学(北京)地球科学学院, 北京 102249;

2. 油气资源与工程全国重点实验室 中国石油大学(北京), 北京 102249

摘要: 油气资源丰度通常受多个因素控制, 其相关参数信息种类繁杂、数据量庞大, 应用传统的地质统计学方法定量预测准确度不高。为了快速预测油气资源量丰度并明确其主控因素, 以渤海湾盆地东濮凹陷文留地区古近系沙河街组三段为例, 采用基于多层感知器神经网络(MLP)方法对油气资源丰度进行定量预测, 同时采用 Boosting 集成学习算法优化预测模型, 分别对 66 组样本油气资源丰度数据进行预测。结果表明, 训练集数据实测值与预测值相关系数分别达 0.789 和 0.989, 验证集数据实测值与预测值相关系数分别达 0.618 和 0.825, 测试数据中实测值和预测值相关系数分别达 0.689 和 0.845; 有效厚度、平均渗透率、有效孔隙度是影响油气资源丰度最主要的 3 个地质因素, 重要性系数分别为 33.93%、20.12% 和 19.53%, 圈闭面积、地面原油密度、生烃中心贡献等参数为次要影响因素。采用 Boosting 集成学习算法优化之后的多层感知器模型预测准确性得到了很大的提升, 能为有利目标优选及勘探开发方案调整提供可靠依据, 为凹陷内其他区块油气资源评价提供较好的参考和借鉴。

关键词: 机器学习; 神经网络; 预测模型; 资源丰度; 东濮凹陷; 渤海湾盆地

中图分类号: TE122.3

文献标识码: A

DOI: 10.11781/sysydz202402428

Prediction of petroleum resource abundance based on artificial neural network method:

a case study of third member of Paleogene Shahejie Formation in Wenliu area of Dongpu Sag, Bohai Bay Basin

YANG Zijie^{1,2}, CHEN Dongxia^{1,2}, WANG Qiaochu^{1,2}, WANG Fuwei^{1,2}, LI Sha^{1,2},

TIAN Ziyi^{1,2}, CHEN Shumin^{1,2}, ZHANG Wanrong^{1,2}, YAO Dongsheng^{1,2}, WANG Yuchao^{1,2}

1. College of Geosciences, China University of Petroleum (Beijing), Beijing 102249, China;

2. National Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing), Beijing 102249, China

Abstract: The abundance of petroleum resource is influenced by various factors and involves complex parameters and extensive data. Consequently, traditional geostatistical methods often lack precision in quantitative prediction. To address this issue, this study focuses on the third member of Paleogene Shahejie Formation (member Es³) in the Wenliu area of the Dongpu Sag and utilizes a multi-layer perceptron neural network (MLP) for predicting petroleum resource abundance and employed the Boosting ensemble learning algorithm to optimize the prediction model. The MLP and MLP-Boosting algorithm models were test on 66 sample groups, yielding correlation coefficients of 0.789 and 0.989 for the training set, 0.618 and 0.825 for the validation set and 0.689 and

收稿日期 (Received): 2023-04-26; 修订日期 (Revised): 2024-02-06; 出版日期 (Published): 2024-03-28.

作者简介: 杨子杰 (1998—), 男, 博士生, 从事油气藏形成机理与分布规律研究。E-mail: yangzj2834@163.com.

通信作者: 陈冬霞 (1974—), 女, 博士, 教授, 从事油气藏形成机理与分布规律研究。E-mail: Lindachen@cup.edu.cn.

基金项目: 国家自然科学基金面上项目 (41972124) 资助。

© Editorial Office of Petroleum Geology & Experiment. This is an open access article under the CC BY-NC-ND license.

0.845 for the test set. The analysis identified effective thickness, average permeability and effective porosity are the most significant geological factors influencing petroleum resource abundance, with importance coefficients of 33.93%, 20.12% and 19.53%, respectively. Other factors such as trap area, surface crude oil density and sedimentary facies assignment were found to be less influential. Overall, the Boosting ensemble learning algorithm significantly enhanced the prediction accuracy of the multi-layer perceptron model, offering valuable insights for target optimization, exploration planning and petroleum resource evaluation in other blocks in the sag.

Key words: machine learning; neural network; prediction model; resource abundance; Dongpu Sag; Bohai Bay Basin

油气储量丰度是指某一地质单元内油气资源量与其单元面积或体积的比值,即单位面积或体积内含油量或含气量,可分为面积资源丰度和体积资源丰度^[1],这里主要指在油气资源评价中广泛应用的面积资源丰度^[2]。常规储层的油气储量丰度影响因素众多^[2-4]。由于这些因素从不同方面影响着油气储量丰度,不能单一用某个因素或简单地用多个因素混合的方法来评价,需要找到各因素之间的关系来进行综合评价。目前,油气储量丰度计算多用油气田的静态资料和参数,首先计算出油气储量,通过叠合含油面积间接计算得出^[5-6]。还有基于资源评价刻度区数据库,类比得到评价单元的7种主要影响因素,进行权值分配之后计算出油气藏的储量丰度^[7-9]。还有学者利用油气资源丰度及其相关地质参数的统计数据,采用多元线性回归技术,建立了油气资源丰度的数学模型^[3,6-10]。此外,有的学者对线性回归模型进行优化处理,优选出对油气资源丰度等资源评价关键参数具有主控作用的地质参数,采用曲线回归分析法建立单参数地质模型。在此基础上,采用多元非线性回归的方法,建立基于主控地质参数的油气资源评价关键参数预测模型^[2,11]。有学者采用改进的分形模拟方法来预测资源丰度,改进了条件模拟的目标函数,提高了预测资源丰度的精度^[12]。这些方法优点在于定量预测盆地油气资源丰度,可以减少类比法计算资源量过程中人为因素的影响,使评价结果更加客观。然而现有的预测评价方法只考虑了地质参数之间的线性关系,忽略了各参数之间存在的复杂非线性关系。

姚纪明等^[13]利用GM(1,1)和BP人工神经网络建立的组合预测模型,对塔里木盆地石油产量发展趋势进行短期预测,预测结果具有一定的可靠性;ZHANG等^[14]对致密气储层渗透率预测时发现非线性多层感知器神经网络和SVR算法的性能优于传统线性回归方法。多层感知器神经网络实现几种简单的函数组合变换和极其复杂的各种非线性拟合,适用于解决复杂的非线性问题。

渤海湾盆地东濮凹陷是一个在中—古生界基础上发育起来的新生代断陷湖盆,蕴藏着丰富的油气资源^[15-16]。东濮凹陷古近纪构造活动强烈,湖盆高频震荡,形成大量砂泥薄互层,且发育多层盐岩层,断层发育,存在多套生储盖组合,具有多类型油气藏特征^[17-18]。由于实际地质情况复杂,受到原始资料质量、预测技术方法等多种因素限制,常规的资源丰度类比法、线性回归法、曲线回归法在预测油气资源丰度时精度大大降低,降低了油气资源评价成果的可信度和可比性。

本文首先通过量化各油气资源丰度影响参数,在明确油气藏基本特征的基础上,采用多层感知器法建立了各影响参数与油气资源丰度之间的非线性关系模型,在此基础上采用 Boosting 集成学习算法优化该预测模型,优选出适合研究区的预测方法,较为准确地预测出油气资源丰度的大小,以期预测凹陷其他区块油气资源丰度的评价提供借鉴。

1 区域地质背景

东濮凹陷是一个位于渤海湾盆地西南缘,具有多套烃源岩、多个生烃洼陷的箕状凹陷。研究区文留地区位于东濮凹陷北中段,呈北东走向,南北长约40 km,东西宽约50 km,面积约2 000 km²;东濮凹陷古近系油气藏主要分布在北部中央隆起区和西部斜坡带(图1a,b)。位于中央隆起带中部的文留地区是东濮凹陷最主要的油气富集区之一^[19]。文留地区含油层系主要为古近系沙河街组,根据沉积的旋回性和岩电组合特征,沙河街组由老到新分为沙四段、沙三段、沙二段和沙一段。其中,沙三段进一步划分为沙三上亚段($E_{s_3}^{\pm}$)、沙三中亚段($E_{s_3}^{\text{中}}$)和沙三下亚段($E_{s_3}^{\text{下}}$)(图2)。据前人研究发现文留地区发育了多套生储盖组合,具备良好的成藏条件;沙三段发育了两套巨厚的盐岩与泥岩组成的韵律层,砂泥岩段是本区的主要含油气层段^[20]。东濮凹陷生烃洼陷数量多,其中前梨园洼陷、海通集洼陷、柳屯洼陷在构造位置上距离文留地区最近,其供烃强度复杂多源^[21]。

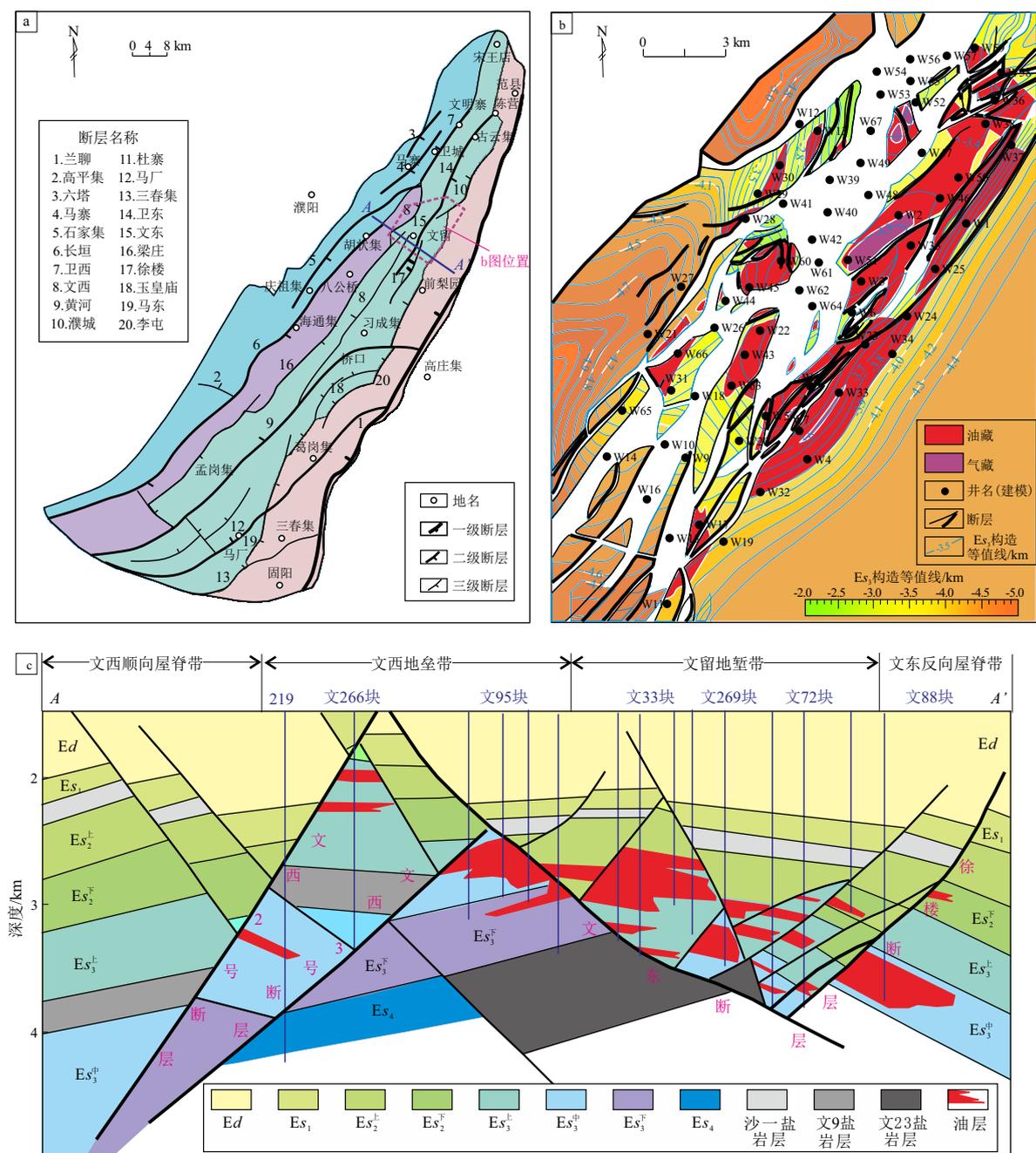


图 1 渤海湾盆地东濮凹陷研究区位置(a)、沙三段构造等值线(b)和典型油藏剖面(c)
据中国石化中原油田分公司。

Fig.1 Location of study area (a), isolines of third member of Shahejie Formation (b) and typical reservoir sections (c) in Dongpu Sag, Bohai Bay Basin

受海西期—印支期、燕山期、喜马拉雅期等多期构造运动的影响,古近系沉积地层经历了复杂的埋藏演化历史和多期生烃演化(图 1c),文留地区派生的次级断层极为发育,这也使得文留地区的构造更为复杂^[18]。这种具有多套生储盖组合、多源供烃和复杂的构造特征会降低线性方法预测资源丰度的准确度。相比较而言,采用神经网络机器学习的方法,对文留地区沙三段油气资源丰度进行预测,会取得较为理想的结果。

2 研究方法

影响文留地区 Es₃ 段油气资源丰度的地质参数极为复杂,涉及到源岩、储层、圈闭、运聚等四大成藏条件,各成藏条件均受众多地质参数综合影响,需要从中优选出与资源丰度相关性较强、具有代表性的主控地质参数。据地质理论、勘探实践及前人研究成果^[17,22],初步筛选出可能与油气资源丰度具有较密切关系的7项可量化的地质参数:

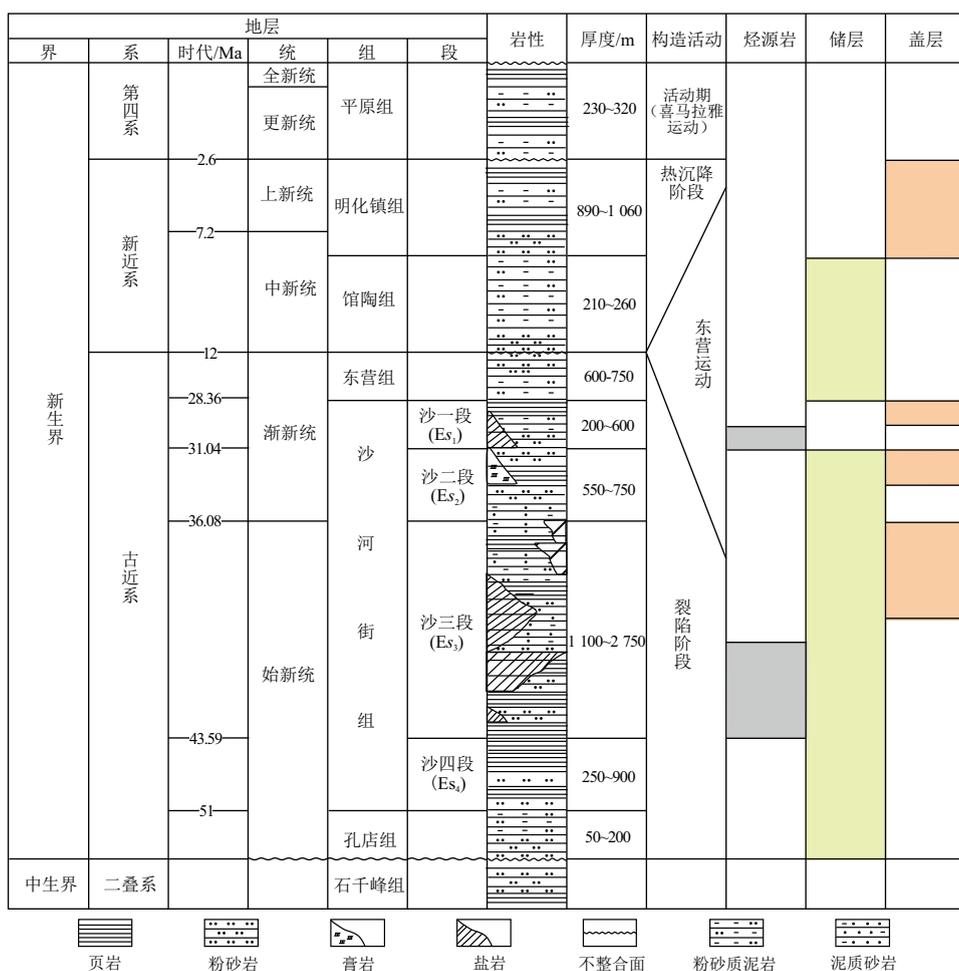


图 2 渤海湾盆地东濮凹陷新生界地层柱状图

Fig.2 Stratigraphic histogram of Cenozoic in Dongpu Sag, Bohai Bay Basin

有效厚度、有效孔隙度、平均渗透率、圈闭面积、地面原油密度、压力系数和离断层距离;2 项定性的地质参数;沉积相和生烃中心。采用 SPSS 23.0 软件对东濮凹陷文留地区 Es₃ 段地质参数原始数据进行预处理。

2.1 数据预处理

生烃中心是影响烃源岩的重要因素,沉积相是评价储集性能优劣的另一重要参数。该定性数据不可直接作为建模的原始资料,需要进行定性数据量化处理。对于上述参数统计整理之后,进行数据预处理,具体操作步骤如下。

2.1.1 标准化

由于建立模型涉及各个影响参数的量纲、数量级等存在较大差别,不同量纲和数量级的数据会对建模结果产生影响,因此需先对原始数据进行无量纲化以减小因量纲差异或数量级差异带来的影响,使其具有一定的可比性^[23]。本次原始数据采取“平移伸缩法”中的标准差进行处理,标准化后

样本 (X'_i) 具体公式如下:

$$X'_i = \frac{X_i - \bar{X}}{\sqrt{S}} \quad (i = 1, 2, 3, \dots, k) \quad (1)$$

式中: X_i 为第 i 个样本; \bar{X} 为样本的均值; S 为样本的标准差。

2.1.2 定性数据量化

(1) 烃源岩生排烃贡献

烃源岩对油气成藏的贡献体现在 2 个方面:一是烃源岩生烃能力对油气成藏及资源丰度的影响;二是油气藏离烃源岩排烃中心的距离。郭秋麟等^[24]采用随机抽样方法,将供烃条件转化成定量评价图;姜福杰等^[25]在“源控论”基础上,排除其他地质因素的影响,选取烃源岩最大排烃强度、油气藏至排烃中心距离以及油气藏至排烃边界距离共 3 个参数建立油气成藏概率 (F_c) 的定量评价模式,用来定量分析生烃中心对油藏或圈闭的生烃贡献能力(图3)。将生烃中心对各个油气藏的贡献进

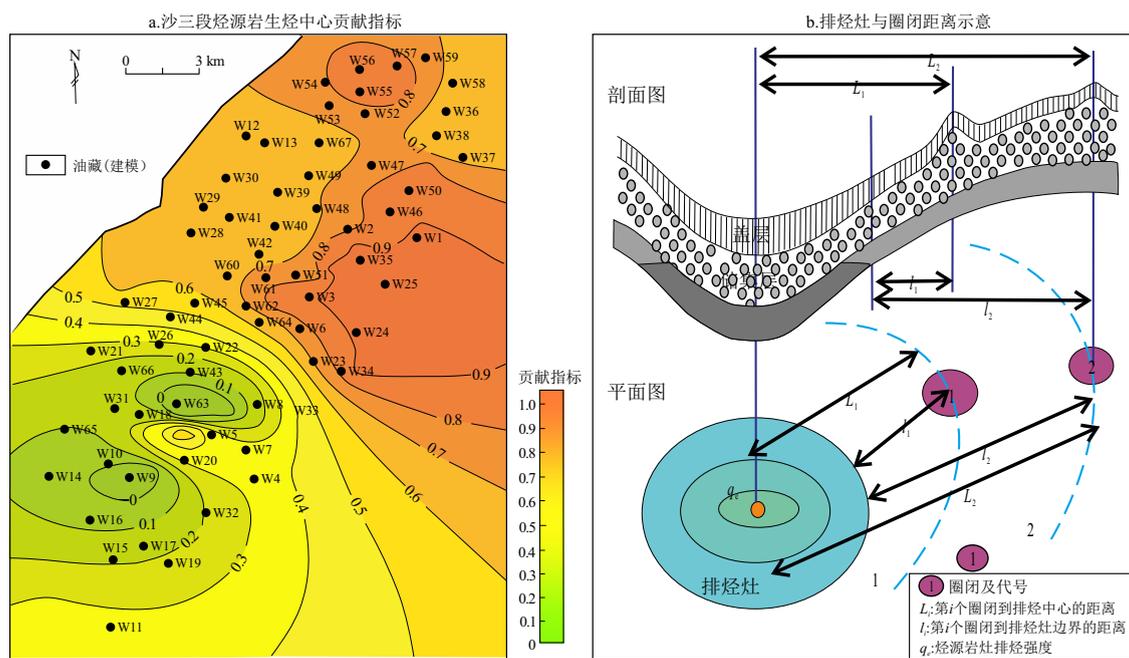


图 3 渤海湾盆地东濮凹陷研究区排烃强度及烃源岩贡献赋值综合示意
据参考文献[25]修改。

Fig.3 Comprehensive schematic diagram of hydrocarbon expulsion intensity and contribution assignments of hydrocarbon source rocks in study area in Dongpu Sag, Bohai Bay Basin

行定量化处理($R^2=0.8825$):

$$F_e = 0.046e^{0.12q_e} - 0.16\ln L + 0.65e^{-8.2357(1+0.1)^2} + 0.1345 \quad (2)$$

式中: q_e 为烃源岩灶排烃强度,单位 10^8 t/km^2 ; L 为标准化的油藏至排烃中心的距离,无量纲; l 为标准化的油藏至排烃边界的距离,无量纲。

(2) 沉积相赋值

通过对前人文献调研与实际研究分析,本次主要针对沉积亚相进行量化赋值。通过不同油藏储量与其沉积亚相做统计分析(图 4),三角洲前缘对探明地质储量的贡献最大,占比 46.6%;其次为半深湖—深湖亚相。对沉积亚相类型进行了优选排序,对油气成藏最有利三角洲前缘亚相赋值为 1,依照不同沉积亚相与油气富集表征参数之间的关系进行 0~1 之间赋值(表 1)。

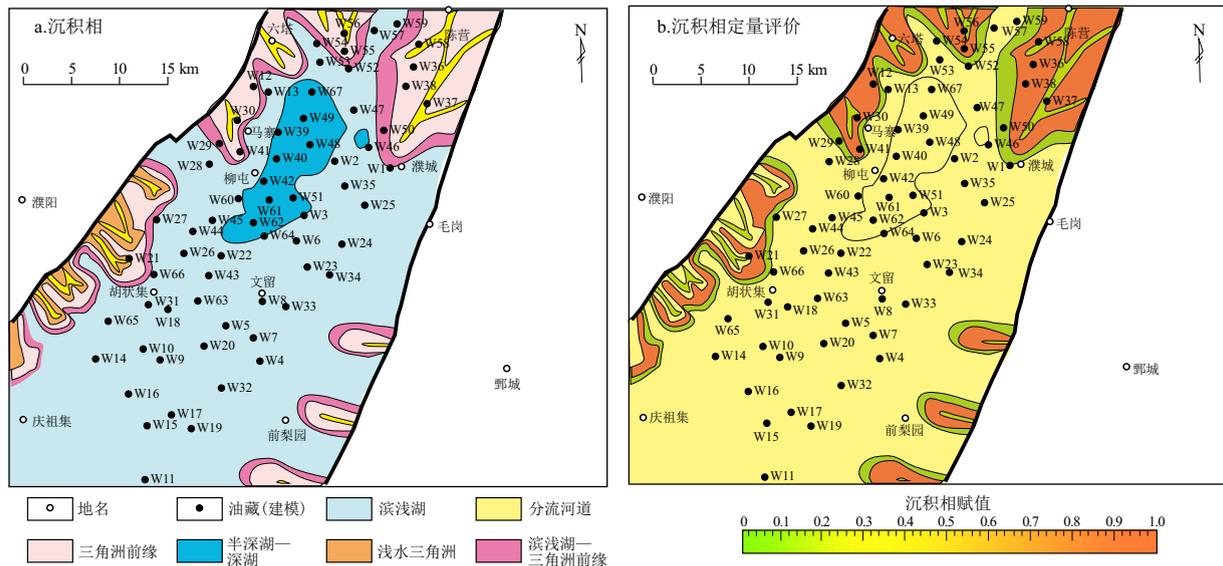


图 4 渤海湾盆地东濮凹陷文留地区沙三段沉积相赋值综合示意
据中国石化中原油田分公司。

Fig.4 Comprehensive schematic diagram of sedimentary facies assignment in third member of Shahejie Formation in Wenliu area of Dongpu Sag, Bohai Bay Basin

表 1 渤海湾盆地东濮凹陷文留地区沉积相赋值统计

Table 1 Statistics of sedimentary facies assignments of Wenliu area of Dongpu Sag, Bohai Bay Basin

参数	滨浅湖	滨浅湖— 三角洲前缘	三角洲 前缘	半深湖— 深湖	分流河道	浅水三角洲
地质储量/10 ⁴ t	319	118	1302	484.87	124	449
地质储量占比	0.11	0.04	0.47	0.17	0.04	0.16
沉积相赋值	0.25	0.09	1.00	0.37	0.10	0.34

沉积相赋值:

$$B = \begin{cases} 1 & \text{三角洲前缘亚相} \\ \frac{F_{\text{沉积相}}}{F_{\text{总}}} & \text{其他沉积亚相} \end{cases} \quad (3)$$

式中: B 为沉积相赋值; $F_{\text{沉积相}}$ 为单个沉积相赋存油气资源量; $F_{\text{总}}$ 为所有沉积相赋存油气资源量。

2.2 多层感知器神经网络

神经网络是一种模拟人脑的神经网络系统, 基于生物学中的神经网络的基本原理而建立, 以期能够实现类人工智能的机器学习技术。神经元模型是一个包含输入、输出与计算功能的模型(图 5)。多层感知器神经网络前一层信号映射被定义好

的传播函数与核函数转换为低维信号, 然后输出层的映射函数被激活。通过激活后的映射函数, 每一个输入的信号都会有一个独自的输出映射^[26], 如下所示:

$$Y_i = f_i \left(\sum_{j=1}^n Q_{ij} X_j + b_i \right) \quad (4)$$

式中: X_j 为输入节点; Y_i 为输出节点; Q_{ij} 为输入节点和输出节点之间的连接权; b_i 为节点的偏斜点; f_i 是节点传递函数。

2.3 Boosting 集成学习算法

Boosting 集成学习算法是由 KEARNS 等^[27]提出的一种用来提高学习算法准确度的方法。这种方法是通过构造一个预测函数系列, 以一定的方式将它们组合成一个预测函数, 达到把弱学习算法提升为强学习算法的目的, 进而提高预测样本的预测精度^[28](图 6)。因此, Boosting 算法可被视为一种通用的增强基础算法性能的回归分析算法。

3 实例分析

3.1 多层感知器神经网络预测油气资源丰度模型

多层感知器神经网络法(MLP)多用于发现数据间极为复杂的关系。用于油气资源丰度预测的模型一般采用三层 BP 神经网络模型(图 5), 本研究构建的网络结构模型如图 7 所示。

在实测点中选择东濮凹陷文留地区已勘探数据 86 个代表性样本, 选择 66 组数据, 将其中 47 组(71.2%) 作为多层感知器模型训练样本, 19 组

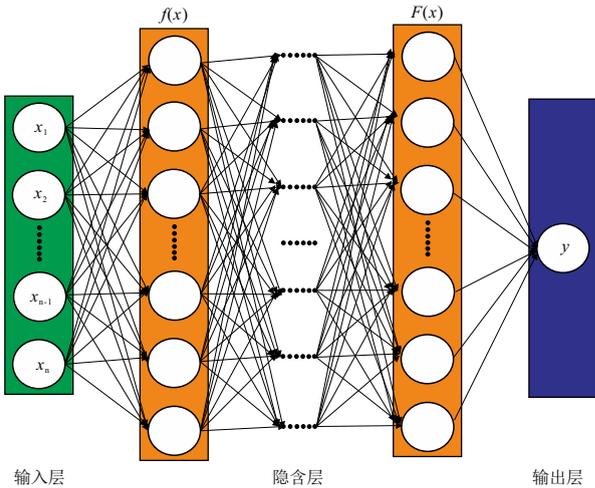


图 5 多层感知器学习网络结构示意图

Fig.5 Structure of multi-layer perceptron (MLP) learning network

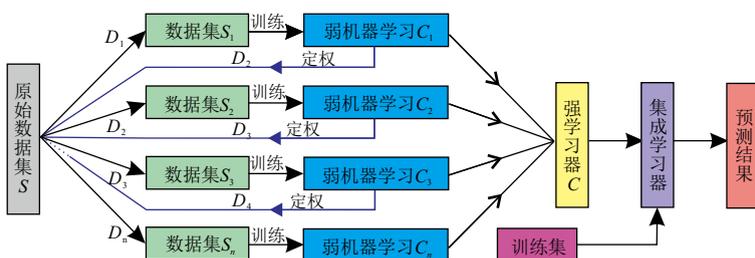


图 6 Boosting 集成学习算法示意

Fig.6 Diagram of Boosting ensemble learning algorithm

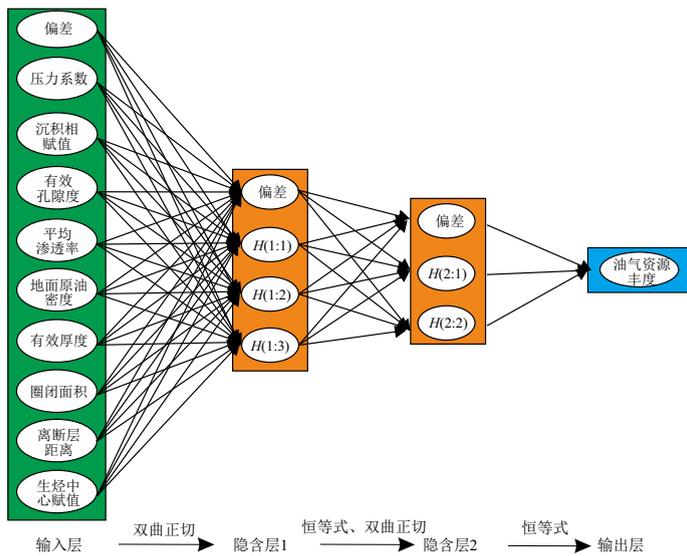


图 7 多层感知器网络结构
Fig.7 MLP network structure

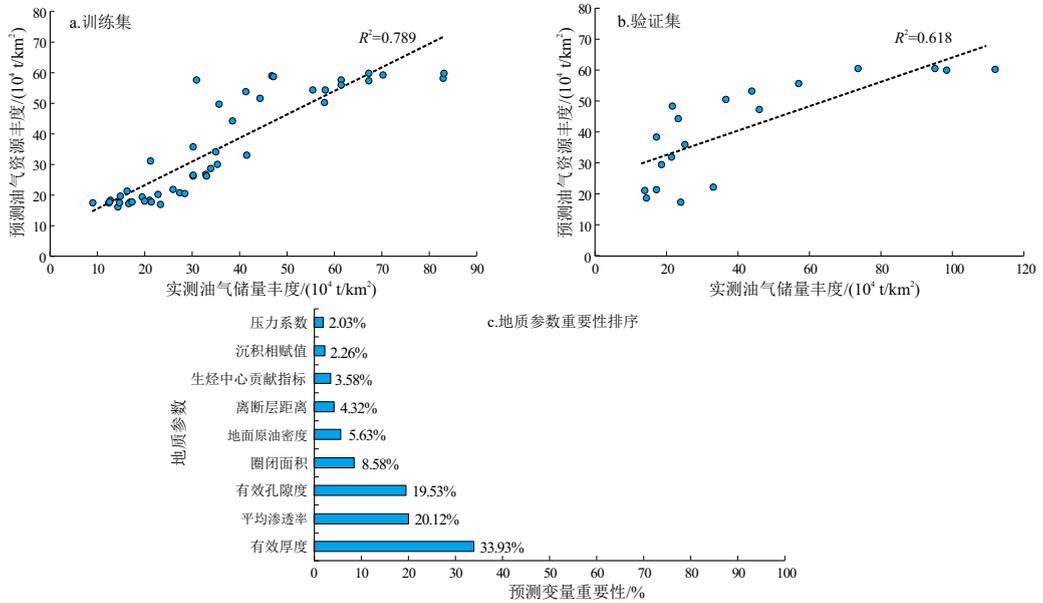


图 8 基于多层感知器神经网络油气资源丰度建模结果

Fig.8 Results of petroleum resource abundance modeling based on MLP neural network

(28.8%)作为验证样本,建立自学习多层感知器神经网络模型,其余 20 组数据用于对 MLP 神经网络模型评价效果的检验。然后用已建立的多层感知器神经网络模型预测油气资源丰度(图 8a,b)。

油气勘探实践表明,油气资源较丰富的区域,孔隙度、渗透率和有效厚度与油气存在一定的正相关性^[29-30]。因此,有效厚度、孔隙度、渗透率对油气资源丰度高低具有重要的影响,与多层感知器神经网络模型分析结果相符(图 8c)。在训练集样本中,预测数据与实测数据相关系数为 0.789,在验证集样本中相关系数为 0.618(表 2),预测结果相对较理想,相关性系数均高于 0.6;但从误差参数分析

表 2 基于多层感知器神经网络模型
油气资源丰度预测结果
Table 2 Prediction results of petroleum resource
abundance based on MLP neural network modeling

参数	训练集	验证集
最小误差/(10 ⁴ t/km ²)	-26.961	-26.606
最大误差/(10 ⁴ t/km ²)	24.535	51.742
平均误差/(10 ⁴ t/km ²)	0.346	0.7
平均绝对误差/(10 ⁴ t/km ²)	6.773	15.736
标准差	9.01	21.081
相关系数	0.789	0.618
样品数	47	19

注:最小误差和最大误差:真实值和预测值之间的差值;平均误差:显示所有样本的误差的平均值;平均绝对误差:显示所有样本的误差绝对值的平均值(不考虑正负)。表 3,表 4 同。

得知,基于普通的多层感知器模型,其预测值和实测值之间的差距较大,预测效果还有待提升。

3.2 Boosting 集成学习算法优化预测模型

在一般的机器学习中,会存在一定的误差,Boosting 集成学习算法可以减小这种偏差。MLP-Boosting 集成学习算法模型所建立的单个弱学习器准确性普遍高于 50%,最高可达 97.78%的准确率(图 9c)。在训练集样本中预测数据与实测数据相关系数为 0.989,验证集样本中相关系数为 0.825(表 3);不难发现通过 MLP-Boosting 集成学习算法建立的模型预测油气资源丰度与实测油气储量丰度具有很好的一致性,二者相关性明显高于 MLP 模型。

3.3 测试模型

利用文留地区 Es₃段未参与建模的 20 个油藏数据,测试这两种模型准确性。MLP 模型测试样本的未参与建模数据与实测值相关性为 0.689,MLP-

表 3 基于 MLP-Boosting 算法模型
油气资源丰度预测结果
Table 3 Prediction results of petroleum resource
abundance based on MLP-Boosting algorithm model

参数	训练集	验证集
最小误差/(10 ⁴ t/km ²)	-6.627	-21.478
最大误差/(10 ⁴ t/km ²)	5.963	44.258
平均误差/(10 ⁴ t/km ²)	-0.205	1.754
平均绝对误差/(10 ⁴ t/km ²)	1.287	11.399
标准差	2.068	16.545
相关系数	0.989	0.825
样品数	47	19

Boosting 模型相关性为 0.845(图 10)。为了更加准确地评价这两种预测方法的优劣,采用最小误差、最大误差、平均绝对误差和标准差指标进行对比(表 4)。根据综合对比结果显示 MLP-Boosting 预测模型各误差指标参数均小于 MLP 神经网络法,经过 Boosting 集成学习算法优化的预测模型表现

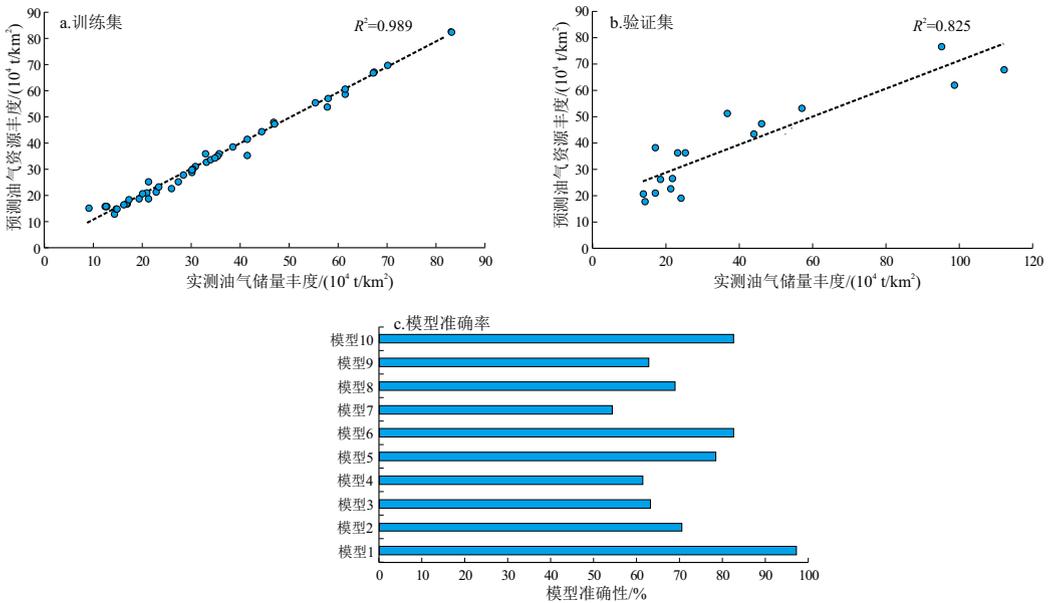


图 9 基于 MLP-Boosting 算法油气资源丰度建模结果

Fig.9 Results of petroleum resource abundance modeling based on MLP-Boosting algorithm modeling

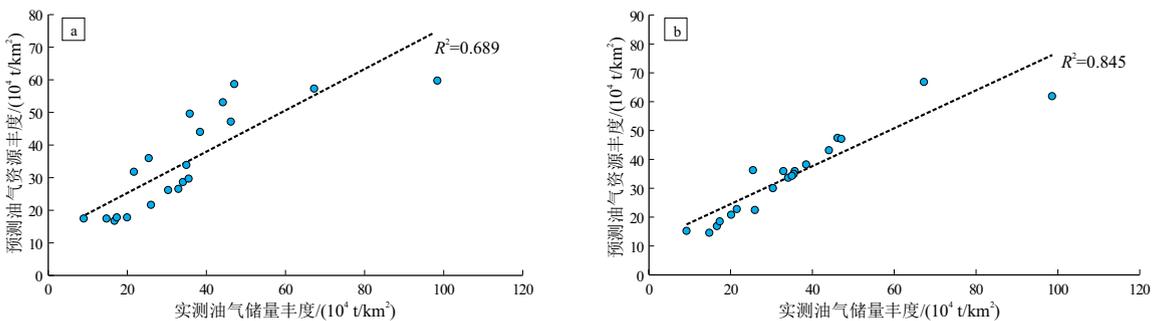


图 10 MLP 模型(a)和 MLP-Boosting 集成算法模型(b)检验数据相关性

Fig.10 Correlation coefficients of checking data of MLP model (a) and MLP-Boosting ensemble algorithm model (b)

表 4 MLP 模型和 MLP-Boosting 集成算法模型检验结果

Table 4 Test results of MLP and MLP-Boosting ensemble algorithm models

参数	检验数据	
	MLP	MLP-Boosting
最小误差/(10 ⁴ t/km ²)	-14.362	-11.288
最大误差/(10 ⁴ t/km ²)	38.262	36.32
平均误差/(10 ⁴ t/km ²)	7.6	3.507
平均绝对误差/(10 ⁴ t/km ²)	11.408	8.936
相关系数	0.689	0.845
样品数	20	20

出更高的准确度。尽管 MLP-Boosting 模型验证数据相关系数未达 0.9 以上,但最大程度改善了 MLP 方法预测精度较低的问题,在面对复杂的地质因素时选择集成学习模型预测更符合实际地质特征(图 11)。

3.4 实例分析

以东濮凹陷 1 号文留地区北部沙三中亚段地区、2 号庆祖集东南部沙三下亚段地区和 3 号葛岗集中部沙三中亚段地区共 3 个预选有利区带作为研究实例,采用基于人工神经网络算法的评价模型对油气资源丰度进行了预测(图 12)。在具体评价

过程中,首先对每个预选有利区进行了 4×4 的等间距取点,并利用均值法、线性回归法、区域外推法对缺失值进行了补齐,确保了机器学习模型预测数据库的完整性。在实际预测区带资源丰度时,由于预测区处于低勘探或未勘探区块,不易获取实际的有效厚度,在此可将储层平均厚度作为实际有效厚度。最终,采用前文所建的多层感知器神经网络模型和 Boosting 算法集成学习模型计算出 3 个区带内共计 48 个数据点的油气资源丰度,并将每个区带内 16 个数据点进行加权平均得出 3 个区带油气资源丰度值(表 5)。模型预测结果与东濮凹陷实际勘探开发认识基本一致,在预测高油气资源丰度的地区发现了油气显示^[16-17],如文留地区已发现多个岩性油气藏,展现出了人工神经网络算法在油气资源丰度预测中的有效性。

4 讨论

运用两种方法预测油气资源丰度,取得了较好的实验结果,反映出机器学习的非线性模型适用于地质构造复杂多样、多套储盖组合和生烃中心地区的资源预测,但这几种方法模型对油气资源丰度预测的准确性依然有待提高。

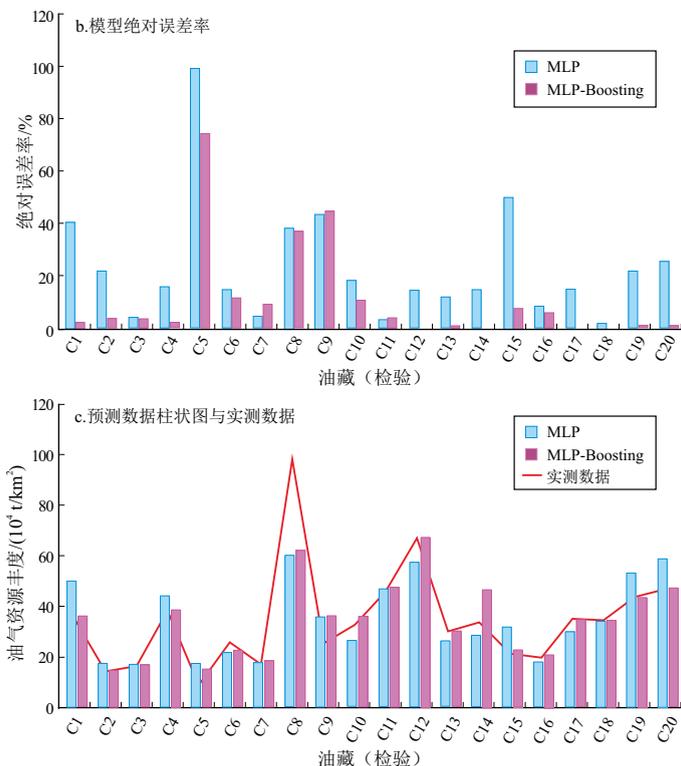
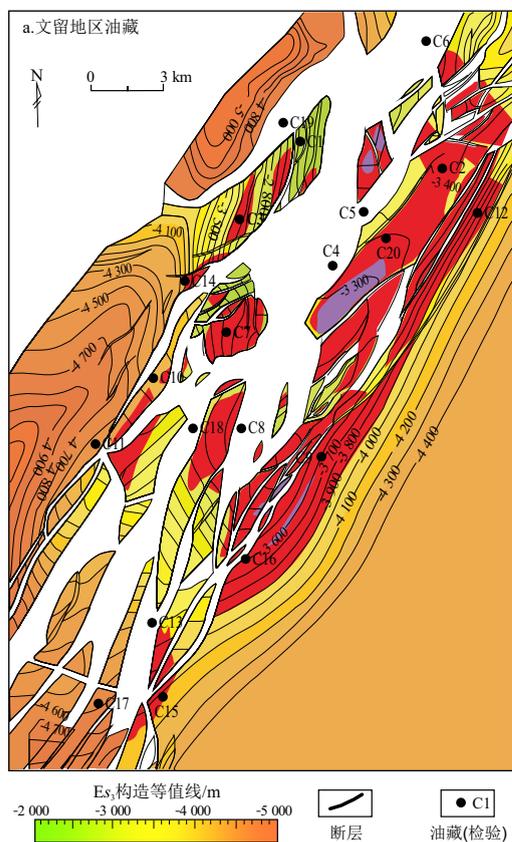


图 11 MLP 模型和 MLP-Boosting 集成算法模型可靠性分析

Fig.11 Reliability of MLP and MLP-Boosting ensemble algorithm models

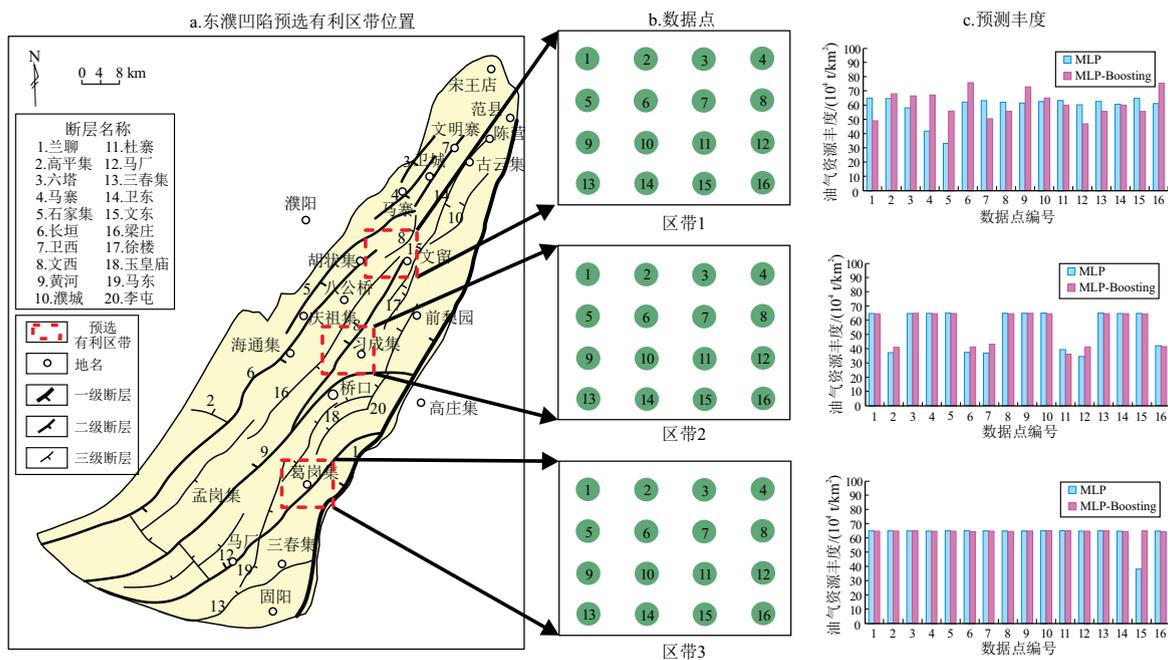


图 12 渤海湾盆地东濮凹陷古近系沙河街组三段预选有利区分布
Fig.12 Distribution of preselected favorable areas of third member of Paleogene Shahejie Formation in Dongpu Sag, Bohai Bay Basin

表 5 渤海湾盆地东濮凹陷古近系沙河街组三段 3 个预选有利区带的资源丰度预测结果

Table 5 Resource abundance evaluation results of three pre-selected favorable zones in third member of Paleogene Shahejie Formation of Dongpu Sag, Bohai Bay Basin

预选有利区带	有利区带资源丰度/(10 ⁴ t/km ²)					
	MLP-Boosting 算法模型			MLP 模型		
	最大值	最小值	平均值	最大值	最小值	平均值
区带 1	75.79	47.11	61.37	64.84	33.20	59.13
区带 2	65.27	36.28	56.16	65.43	35.03	55.23
区带 3	65.32	65.12	65.21	65.43	38.56	63.54

(1) 本文中所选的两方法的预测结果可能存在较大的偶然性和随机性,需要进一步深入研究或进一步优化模型。未来可尝试使用更加科学智能的方法对模型的参数进行寻优,同时改善其神经元之间的激活函数,使其权重值分配更加科学合理。

(2) 为了包含更多的地质信息参数提高油气资源预测的准确性,采用 9 个地质参数进行机器学习建模,通过计算各参数的相关系数矩阵热图(图 13)发现,其中相关系数越靠近 1,地质参数间的相关性越好。这从侧面也反映出弊端,参数之间的相关性过高,就会产生自变量之间的共线性问题和过拟合现象,因此,在后续研究分析中将尝试采用主成分分析、岭回归法等方法进行数据压缩、降维处理。

(3) 数据量不足会影响数据的质量和数据的

分布规律,当数据分布不符合正态分布时,所建立的模型不能囊括全部的信息资料(图 14)。

同时,该模型对其他凹陷地区适用性还有待更新,在对其他凹陷地区进行资源丰度评价预测时,需用该地区钻井、录井等实测资料重新对模型进行拟合,以求拟合到适应本凹陷的预测模型。

5 结论

(1) 通过强学习器 MLP-Boosting 模型预测油气资源丰度,得到该模型训练集精度为 0.989,验证集精度为 0.825;同时采用未参与建模的数据进行模型测试,尽管实测数据与预测数据相关性系数降低,但 MLP-Boosting 模型精度仍高于 MLP 模型且在 0.8 以上,因此,采用 Boosting 集成学习算法优化的神经网络模型能够较好地进行油气资源丰度预测。

(2) 采用多层感知器神经网络模型得到各参

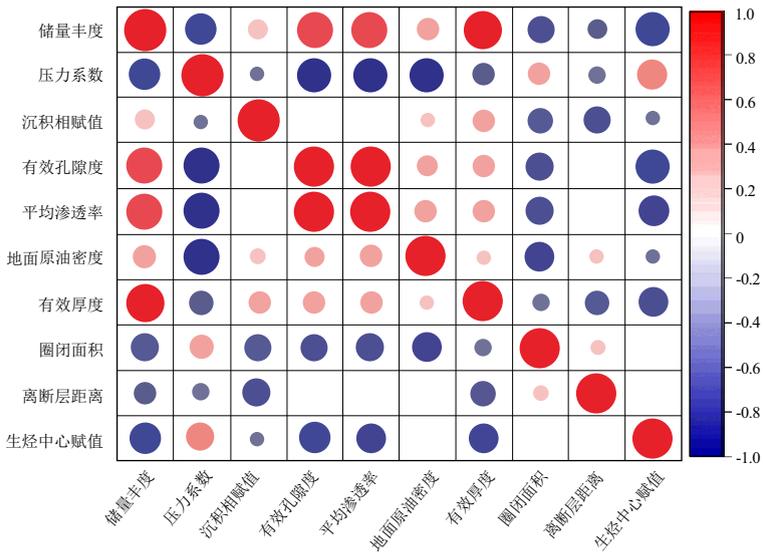


图 13 渤海湾盆地东濮凹陷古近系沙河街组三段地质参数相关系数矩阵热图

Fig.13 Matrix heat map of correlation coefficients of geological parameters in third member of Paleogene Shahejie Formation, Dongpu Sag, Bohai Bay Basin

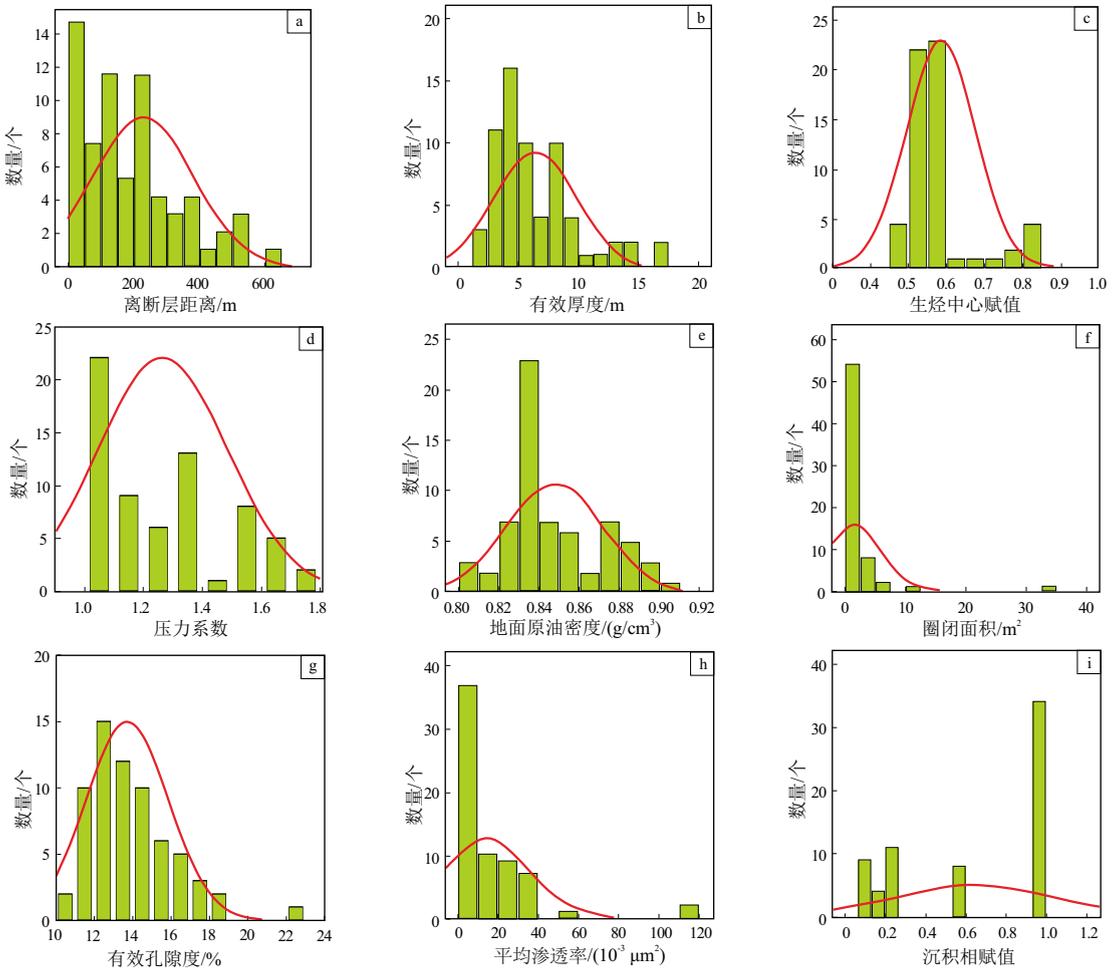


图 14 渤海湾盆地东濮凹陷古近系沙河街组三段地质参数数据

Fig.14 Geological parameters of third member of Paleogene Shahejie Formation, Dongpu Sag, Bohai Bay Basin

数对油气资源丰度高低的影响程度排队,发现有效厚度、有效孔隙度、平均渗透率是影响油气资源丰

度最主要的 3 个地质因素,而圈闭面积、地面原油密度、生烃中心、压力系数、离断层距离、沉积相的

控制作用则相对次之,与传统地质对油气资源丰度影响的认识一致。

(3)本文虽然采用 Boosting 集成算法优化的多层感知器神经网络方法预测油气资源丰度,取得了较好的实验结果,但此方法的准确性依然有待提高,如数据量、数据质量优劣、算法的选择等方面仍需改进。

致谢:感谢期刊编辑与审稿人对论文提出的宝贵意见,同时感谢中国石化中原油田分公司在资料收集等方面提供的大力支持与帮助,在此致以衷心感谢!

利益冲突声明/Conflict of Interests

所有作者声明不存在利益冲突。

All authors disclose no relevant conflict of interests.

作者贡献/Authors' Contributions

杨子杰、陈冬霞、王翹楚、王福伟、李莎、田梓棣参与文章的构思与设计;杨子杰、陈淑敏、张婉蓉、姚东升、王昱超完成模型数据的预测;杨子杰、陈冬霞、王翹楚、王福伟参与论文写作和修改。所有作者均阅读并同意最终稿件的提交。

The study was designed by YANG Zijie, CHEN Dongxia, WANG Qiaochu, WANG Fuwei, LI Sha and TIAN Ziyue. The prediction of model data was made by YANG Zijie, CHEN Shumin, ZHANG Wanrong, YAO Dongsheng and WANG Yuchao. The manuscript was drafted and revised by YANG Zijie, CHEN Dongxia, WANG Qiaochu and WANG Fuwei. All the authors have read the last version of paper and consented for submission.

参考文献:

- [1] 柳广弟,刘成林,郭秋麟.油气资源评价[M].北京:石油工业出版社,2018.

LIU Guangdi, LIU Chenglin, GUO Qiulin. Oil and gas resources evaluation[M]. Beijing: Petroleum Industry Press, 2018.

- [2] 张蔚,刘成林,吴晓智,等.中国不同类型盆地油气资源丰度统计特征及预测模型[J].地质与勘探,2019,55(6):1518-1527.

ZHANG Wei, LIU Chenglin, WU Xiaozhi, et al. Statistical characteristics and prediction models for oil and gas resources abundance in different types of Chinese basins[J]. Geology and Exploration, 2019, 55(6): 1518-1527.

- [3] 鄢琦,周总瑛.中国东部断陷盆地石油资源丰度统计模型的建立[J].石油实验地质,2009,31(3):292-295.

YAN Qi, ZHOU Zongying. Establishment on oil resources abundance statistical model in East China rift basins[J]. Petroleum Geology & Experiment, 2009, 31(3): 292-295.

- [4] 赵文智,胡素云,王红军,等.中国中低丰度油气资源大型化成藏与分布[J].石油勘探与开发,2013,40(1):1-13.

ZHAO Wenzhi, HU Suyun, WANG Hongjun, et al. Large-scale accumulation and distribution of medium-low abundance hydrocarbon resources in China[J]. Petroleum Exploration and Development, 2013, 40(1): 1-13.

- [5] 张金川,金之钧,郑茂浚.深盆气资源量—储量评价方法[J].天然气工业,2001,21(4):32-35.

ZHANG Jinchuan, JIN Zhijun, ZHENG Junmao. Deep basin gas resource-reserve evaluation method[J]. Natural Gas Industry, 2001, 21(4): 32-35.

- [6] 崔宝文,赵莹,张革,等.松辽盆地古龙页岩油地质储量估算方法及其应用[J].大庆石油地质与开发,2022,41(3):14-23.

CUI Baowen, ZHAO Ying, ZHANG Ge, et al. Estimation method and application for OOIP of Gulong shale oil in Songliao Basin[J]. Petroleum Geology & Oilfield Development in Daqing, 2022, 41(3): 14-23.

- [7] 郭秋麟,武娜,闫伟,等.深层天然气资源评价方法[J].石油学报,2019,40(4):383-394.

GUO Qiulin, WU Na, YAN Wei, et al. An assessment method for deep gas resources[J]. Acta Petrolei Sinica, 2019, 40(4): 383-394.

- [8] 徐旭辉,周卓明,宋振响,等.油气资源评价方法关键参数研究和资源分布特征:以中国石化探区“十三五”资源评价为例[J].石油实验地质,2023,45(5):832-843.

XU Xuhui, ZHOU Zhuoming, SONG Zhenxiang, et al. Methods and key parameters for oil and gas resource assessment and distribution characteristics of oil and gas resource: a case study of resource assessment of SINOPEC during the 13th Five-Year Plan period[J]. Petroleum Geology & Experiment, 2023, 45(5): 832-843.

- [9] 凡玉梅.未开发油气储量不确定性潜力评价方法[J].石油实验地质,2022,44(6):1100-1104.

FAN Yumei. Evaluation method for uncertain potential of undeveloped reserves[J]. Petroleum Geology & Experiment, 2022, 44(6): 1100-1104.

- [10] 柳广弟,胡素云,赵文智.中国主要含油气盆地运聚单元石油资源丰度及其预测模型[J].石油勘探与开发,2006,33(6):759-761.

LIU Guangdi, HU Suyun, ZHAO Wenzhi. Oil resource abundance of petroleum plays in Chinese basins and its prediction model[J]. Petroleum Exploration and Development, 2006, 33(6): 759-761.

- [11] 白琨琳,赵迎冬.油气资源评价中成因法分析与运聚系数取值模型研究[J].地质与勘探,2021,57(3):656-666.

BAI Kunlin, ZHAO Yingdong. Valuation model of the migration-accumulation coefficient in the genetic method for assessment of oil and gas resources[J]. Geology and Exploration, 2021, 57(3): 656-666.

- [12] GUO Qiulin, REN Hongjia, WU Xiaozhi, et al. A fractal simulation method for simulating the resource abundance of oil and gas and its application[J]. Mathematical Geosciences, 2022, 54(5): 873-901.

- [13] 姚纪明,于炳松,车长波,等.组合法在塔里木盆地石油产量预测中的应用[J].自然资源学报,2009,24(5):907-914.

YAO Jiming, YU Bingsong, CHE Changbo, et al. Application of a combination forecast model in the trend forecast of oil production in the Tarim Basin[J]. Journal of Natural Resources, 2009, 24(5): 907-914.

- [14] ZHANG Guoyin, WANG Zhizhang, LI Huaji, et al. Permeability prediction of isolated channel sands using machine learning[J]. Journal of Applied Geophysics, 2018, 159: 605-615.

- [15] 刘铁桩,张清正,李亚萍.东濮凹陷天然气资源潜力、勘探方向及前景[J].西部探矿工程,2001,13(5):45-46.
LIU Tiezhu,ZHANG Qingzheng,LI Yaping.The natural gas resource potential and exploration direction and looking forward of Dongpu Sag[J].West-China Exploration Engineering,2001,13(5):45-46.
- [16] 董满阳.东濮凹陷油气资源潜力与分布评价[D].北京:中国石油大学(北京),2018.
DONG Xiaoyang.Potential and distribution of oil and gas resources in Dongpu Sag[D].Beijing:China University of Petroleum (Beijing),2018.
- [17] 谈玉明,李红磊,张云献,等.东濮凹陷古近系优质烃源岩特征与剩余资源潜力分析[J].断块油气田,2020,27(5):551-555.
TAN Yuming,LI Honglei,ZHANG Yunxian,et al.Analysis to high quality source rock characteristics and residual resource potential in Dongpu Sag in Paleogene[J].Fault-Block Oil & Gas Field,2020,27(5):551-555.
- [18] 余海波,程秀申,徐田武,等.东濮凹陷古近系构造特征及其对油气成藏的控制作用[J].油气地质与采收率,2021,28(3):42-52.
YU Haibo,CHENG Xiushen,XU Tianwu,et al.Paleogene tectonic characteristics and their controlling effect on hydrocarbon accumulation in Dongpu Sag[J].Petroleum Geology and Recovery Efficiency,2021,28(3):42-52.
- [19] 胡斌,陈传浩,王长征,等.东濮凹陷文留地区沙三中($E_{3中}$) 遗迹化石与沉积环境[J].河南理工大学学报(自然科学版),2017,36(3):40-46.
HU Bin,CHEN Chuanhao,WANG Changzheng,et al.Trace fossils and sedimentary environments in the middle part of third member Shahejie Formation in Wenliu area, Dongpu Sag[J].Journal of Henan Polytechnic University (Natural Science),2017,36(3):40-46.
- [20] 慕小水.东濮凹陷文留地区含盐层系油气成藏机理与模式[D].北京:中国地质大学(北京),2011.
MU Xiaoshui.Hydrocarbon reservoir formation mechanism and pattern for saline series in Wenliu area, Dongpu Depression[D].Beijing:China University of Geosciences (Beijing),2011.
- [21] 刘宣威,王学军,李红磊,等.东濮凹陷古近系烃源岩特征及其形成环境分析[J].断块油气田,2021,28(4):452-455.
LIU Xuanwei,WANG Xuejun,LI Honglei,et al.Characteristics and formation environment analysis of Paleogene source rocks in Dongpu Depression[J].Fault-Block Oil & Gas Field,2021,28(4):452-455.
- [22] 李浩,王保华,陆建林,等.东濮凹陷古近系页岩油富集地质条件与勘探前景[J].中国石油大学学报(自然科学版),2021,45(3):33-41.
LI Hao,WANG Baohua,LU Jianlin,et al.Geological characteristics and exploration prospects of Paleogene continental shale oil accumulation in Dongpu Sag, Bohai Bay Basin[J].Journal of China University of Petroleum (Edition of Natural Science),2021,45(3):33-41.
- [23] 高晓红,李兴奇.主成分分析中线性无量纲化方法的比较研究[J].统计与决策,2020,36(3):33-36.
GAO Xiaohong,LI Xingqi.Comparative study on linear dimensionless methods in principal component analysis[J].Statistics & Decision,2020,36(3):33-36.
- [24] 郭秋麟,任洪佳,于京都,等.基于贝叶斯网络的油气勘探风险预测方法:以准噶尔盆地腹部侏罗系三工河组为例[J].中国石油勘探,2023,28(1):108-119.
GUO Qiulin,REN Hongjia,YU Jingdu,et al.Prediction method of petroleum exploration risks based on Bayesian network:a case study of the Jurassic Sangonghe Formation in the hinterland of Junggar Basin[J].China Petroleum Exploration,2023,28(1):108-119.
- [25] 姜福杰,姜振学,庞雄奇.东营凹陷油气成藏体系的划分及定量评价[J].地球科学(中国地质大学学报),2008,33(5):651-660.
JIANG Fujie,JIANG Zhenxue,PANG Xiongqi.Division and quantitative evaluation of petroleum accumulation system in Dongying Sag[J].Earth Science(Journal of China University of Geosciences),2008,33(5):651-660.
- [26] ASSI K J,NAHIDUZZAMAN K M,RATROUT N T,et al.Mode choice behavior of high school goers:evaluating logistic regression and MLP neural networks[J].Case Studies on Transport Policy,2018,6(2):225-230.
- [27] KEARNS M J,VALIANT L G.Learning Boolean formulae or finite automata is as hard as factoring[R].Cambridge:Harvard University,Center for Research in Computing Technology,Aiken Computation Laboratory,1988.
- [28] 陈凯,朱钰.机器学习及其相关算法综述[J].统计与信息论坛,2007,22(5):105-112.
CHEN Kai,ZHU Yu.A summary of machine learning and related algorithms[J].Statistics & Information Forum,2007,22(5):105-112.
- [29] 吕传炳,庞雄奇,马奎友,等.渤海湾盆地束鹿凹陷“牙刷状”油藏成藏特征与模式[J].石油与天然气地质,2022,43(3):566-581.
LÜ Chuanbing,PANG Xiongqi,MA Kuiyou,et al.Characteristics and reservoiring patterns of “teeth-brush-shaped” oil pools in the Shulu Sag, Bohai Bay Basin[J].Oil & Gas Geology,2022,43(3):566-581.
- [30] 张晓磊,唐颖,李卓奕,等.鄂尔多斯盆地西南缘延长组长 8 段低充注油藏成藏模式:以环西—彭阳地区为例[J].西安石油大学学报(自然科学版),2023,38(1):31-44.
ZHANG Xiaolei,TANG Ying,LI Zhuoyi,et al.Hydrocarbon accumulation patterns of Chang 8 low-charging oil reservoirs in southwestern margin of Ordos Basin:taking Huanxi-Pengyang area as an example[J].Journal of Xi'an Shiyou University (Natural Science Edition),2023,38(1):31-44.