

基于机器学习的页岩总有机碳含量评价方法

王宵宇¹, 廖广志^{1*}, 黄文松², 刘海山³, 孔详文², 赵子斌²

1 中国石油大学(北京)油气资源与工程全国重点实验室, 北京 102249

2 中国石油勘探开发研究院, 北京 100083

3 中海油田服务股份有限公司, 三河 065201

* 通信作者, liaoguangzhi@cup.edu.cn

收稿日期: 2024-08-29; 修回日期: 2025-01-07

国家重点研发计划(2019YFA0708301)、国家自然科学基金(42474165)和中国石油天然气集团有限公司-中国石油大学(北京)战略合作科技专项(ZLZX2020-03)联合资助

摘要 总有机碳(TOC)含量是评估烃源岩储层品质和生烃潜力的重要地球化学参数之一, 其准确预测对页岩油气勘探开发具有重要意义。随着人工智能技术的快速发展, 单一机器学习方法常被应用于TOC含量评价。然而, 单一机器学习方法存在过拟合、欠拟合和目标函数局部最优等问题。集成模型被证实通过整合多个智能算法可以提高预测精度和稳定性能, 其中组合策略是优化集成模型的关键之一。算术平均法作为组合策略难以充分发挥最佳模型的预测性能, 而且容易受到预测误差较大的智能算法的影响。加权求和法作为组合策略根据训练数据确定加权系数, 在训练集上表现出色, 却在测试集中表现欠佳。本文提出了一种基于智能匹配技术的集成模型(IMTEM), 采用极限梯度提升、随机森林、支持向量机和极限学习机作为算法模块对输入数据进行初步处理, 提取的特征信息与原始测井响应共同输入到前馈神经网络层中进行非线性转换以及特征学习, 从而对页岩TOC含量进行准确且连续的评价。将本文提出的方法应用于四川盆地龙马溪组页岩TOC含量预测, 测试结果表明, 相比于两种集成模型、5种基础模型和 $\Delta\log R$ 方法, IMTEM的预测结果与岩心实测TOC含量一致性更高, 更适用于页岩TOC含量的预测。

关键词 机器学习; 测井评价; 页岩地层; 总有机碳含量; 集成模型

中图分类号: P618.13; TP18

Evaluation method of total organic carbon content in shale based on machine learning

WANG Xiaoyu¹, LIAO Guangzhi¹, HUANG Wensong², LIU Haishan³, KONG Xiangwen², ZHAO Zibin²

1 State Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum, Beijing 102249, China

2 Research Institute of Petroleum Exploration and Development, Beijing 100083, China

3 China Oilfield Services Limited, Sanhe 065201, China

引用格式: 王宵宇, 廖广志, 黄文松, 刘海山, 孔详文, 赵子斌. 基于机器学习的页岩总有机碳含量评价方法[J]. 石油科学通报, 2025, 10(02): 392–403

WANG Xiaoyu, LIAO Guangzhi, HUANG Wensong, LIU Haishan, KONG Xiangwen, ZHAO Zibin. Evaluation method of total organic carbon content in shale based on machine learning[J]. Petroleum Science Bulletin, 2025, 10(02): 392–403. DOI: 10.3969/j.issn.2096-1693.2025.01.007

Received: 2024-08-29; Revised: 2025-01-07

Abstract Total organic carbon (TOC) content is a crucial geochemical parameter for assessing reservoir quality and hydrocarbon generation potential of source rocks. The accurate prediction of TOC content is important for optimizing the exploration and development processes of shale oil and gas. With the rapid development of artificial intelligence technologies, individual machine learning algorithms have been increasingly applied to evaluate TOC content in shale. Despite the promising results of the individual machine learning algorithms, they are often subject to several challenges including overfitting, underfitting, and getting trapped in local optima of objective function. To address these limitations, the ensemble learning models are developed. Ensemble learning models leverage the strengths of multiple individual intelligent algorithms to enhance prediction accuracy and stability. Among them, combination strategy is one of the key factors in optimizing the ensemble learning models. Arithmetic average method as the simplest combination strategy fails to fully use prediction performance of the best individual intelligent model, and it can be severely affected by the individual intelligent model with a large prediction error, which can interfere with prediction outcome of overall model. In comparison, weighted summation method as a common combination strategy assigns the weights to different individual intelligent models according to their performance on training data. This method will perform excellently on training set, but it tends to have a poor performance when applied to test set. This paper develops an ensemble model based on an intelligent matching technology (IMTEM). The proposed method utilizes a set of robust intelligent algorithms including extreme gradient boosting, random forest, support vector machine, and extreme learning machine as algorithm modules to initially process input data. Then, the processed feature information combined with original log responses is fed to feedforward neural network layer for nonlinear transformation and feature learning, thereby enabling accurate and continuous estimation of TOC content in shale. To validate effectiveness of the IMTEM, the proposed method is applied to the prediction of TOC content in the Longmaxi Formation shale in the Sichuan Basin. Test results indicate that, compared to two ensemble models, five baseline models, and the $\Delta\log R$ method, predictions of the IMTEM exhibit higher consistency with measured TOC content. This demonstrates that the IMTEM is more suitable for predicting TOC content in shale.

Keywords machine learning; well logging evaluation; shale formation; total organic carbon content; ensemble model

DOI: 10.3969/j.issn.2096-1693.2025.01.007

0 引言

总有机碳(TOC)含量是衡量页岩油气地层有机质丰度和生烃潜力的关键地球化学指标之一,可采用地球化学实验方法进行测量^[1]。在地层评价时,为了获得连续性好且准确性高的页岩TOC含量,通常利用测井数据建立经验公式来估算TOC含量^[2-5]。

近年来,机器学习方法被应用于页岩TOC含量的评价中,常用的方法包括全连接神经网络、支持向量机、极限学习机、随机森林和极限梯度提升等,取得了较好的应用效果^[6-11]。在实际应用中,单一机器学习方法对超参数敏感,常会面临过拟合、欠拟合和目标函数局部最优等系列问题,影响了其在应用过程中的稳定性和可靠性^[12]。集成模型通过联合不同类型的智能算法,并制定相应的决策机制能够显著改善预测结果,增强训练模型的稳定性^[13]。同时,根据柯西不等式证实了集成模型的预测精度总是优于算法模块的平均水平^[14]。这些优势使得集成模型在预测孔隙度、渗透率、饱和度和TOC含量等回归任务,以及识别流体性质和地层岩性等分类任务时,均取得了比单一机器学习方法更加准确的预测结果^[15-18]。对于回归问题,集成模型常采用算术平均法或加权求和法作为组合策

略融合每个算法模块的输出^[19]。然而,算术平均法仅是平均所有算法模块的输出,却难以充分发挥最佳模型的预测优势,同时容易受到误差较大的模型影响。加权求和法根据训练数据确定算法模块的加权系数,在训练数据上表现良好,但在测试数据上表现较差。为了更好地学习算法模块与目标参数之间的非线性映射关系,本文开发了一种基于智能匹配技术的集成模型(IMTEM)。首先将原始测井响应与算法模块提取的相关模式特征进行跨层联结,提升数据的质量及信息量。然后利用前馈神经网络出色的非线性拟合能力建立联合特征与预测目标之间的映射关联,实现整体模型的性能提升,从而提高预测TOC含量的准确性。

本文首先阐述了IMTEM的理论基础和网络结构设计。然后将该方法应用于四川盆地龙马溪组页岩TOC含量预测,在分析其应用效果的基础上,与两种集成模型、5种基础模型和 $\Delta\log R$ 方法进行了对比分析,测试结果表明IMTEM取得了较好的应用效果。

1 IMTEM 原理

IMTEM是一种集成了多个智能算法的复合网络,通过整合每个专家模型的预测优势,优化复合网络的

群体智能，以增强整体模型的泛化性能。该方法不依赖于单一模型的预测，确保其能够适应复杂的地层情况。图 1 为用于预测 TOC 含量的 IMTEM，包括了输入层、算法层、前馈神经网络层和输出层。

1.1 输入层

输入层作为数据进入 IMTEM 的门户，负责准备和预处理输入数据，以便随后由算法层的算法模块进行学习和训练。输入层采用最小—最大归一化方法对原始测井响应进行处理。

1.2 算法层

在模型的训练和预测阶段，算法层起着至关重要的作用，负责整合来自不同智能算法的输出。这些智能算法根据其独特的网络结构，专注于处理输入数据的不同方面，利用各自的预测优势产生多样化的预测结果。其本质是建立智能算法与预测目标之间的自适应关系，这种关系会随着应用场景的不同而变化^[13]。为了使模型能够适应复杂的地层情况，算法层采用并行结合策略将不同类型的智能算法进行组合，最大限度地利用了每个智能算法与目标参数之间的自适应关系，即使某个智能算法的预测误差较大，其他智能算法的高性能也能够弥补其对整体模型的影响。在本研究中，算法层用于回归问题的初步处理。

IMTEM 采用极限梯度提升、随机森林、支持向量机和极限学习机作为算法层的算法模块，不同智能算

法的特点及其优势如下。

极限梯度提升是一种基于梯度提升树的 Boosting 集成模型，通过迭代添加决策树拟合前序决策树残差的方式调整数据的权重分布，实现损失函数的持续降低^[20]。因其训练速度快和预测精度高被广泛应用于各类测井评价任务中^[6]。

随机森林是一种基于决策树的 Bagging 集成模型，通过自助采样技术和特征随机选择方式确保基础模型的独立性和多样性，采用算术平均或多数投票整合大量决策树的预测结果来提高整体模型的预测精度^[21]。

支持向量机是一种基于核函数理论的机器学习方法，通过将数据映射到高维特征空间，并构建超平面来最大化不同类别数据点之间的间隔，实现数据的线性可分^[22]。该模型能够有效处理复杂的非线性关系，在解决小样本和高维数据问题时表现出色^[23]。

极限学习机是一种单隐藏层前馈神经网络方法，其输入层与隐藏层之间的权重和偏置随机初始化并保持固定，从而大幅度降低训练时间和简化训练过程^[24]。该模型具有通用逼近能力，在处理大规模数据时仍能保持良好的性能表现^[25]。

1.3 前馈神经网络层

全连接神经网络能够建立特征参数与目标数据之间复杂的非线性映射，常被配置在卷积神经网络、长短期记忆网络或门控循环单元网络之后，形成一种层

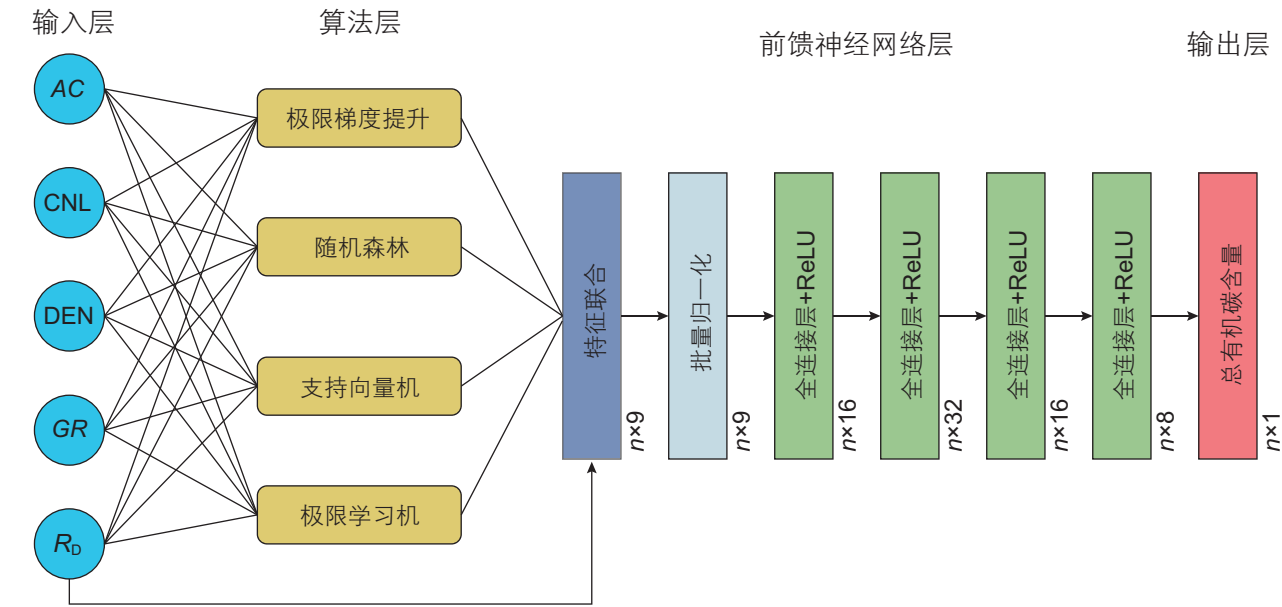


图 1 IMTEM 的模型结构
Fig. 1 Model structure of the IMTEM

序策略,以最大化数据中的信息提取与利用^[26-28]。在本研究中,算法层提取的特征信息与原始测井响应共同输入到前馈神经网络层,通过数据增强技术来提高输入数据的质量和信量^[29]。在前馈神经网络层中构建特征联合矩阵,将这些不同来源的数据按列组合:

$$\mathbf{K} = \begin{bmatrix} k_{1,1} & k_{1,2} & \cdots & k_{1,t} & v_{1,1} & v_{1,2} & \cdots & v_{1,s} \\ k_{2,1} & k_{2,2} & \cdots & k_{2,t} & v_{2,1} & v_{2,2} & \cdots & v_{2,s} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ k_{i,1} & k_{i,2} & \cdots & k_{i,t} & v_{i,1} & v_{i,2} & \cdots & v_{i,s} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ k_{d,1} & k_{d,2} & \cdots & k_{d,t} & v_{d,1} & v_{d,2} & \cdots & v_{d,s} \end{bmatrix} \quad (1)$$

式中, $k_{d,t}$ 表示在深度 d 处的第 t 个智能专家的预测结果; $v_{d,s}$ 表示在深度 d 处的第 s 个测井响应。采用批量归一化方式对特征联合矩阵进行归一化处理,然后输入到全连接层进行非线性转换及目标参数预测。经过多次尝试,前馈神经网络层的隐藏层数量设置为 4 层,神经元数量分别设置为 16、32、16 和 8 个,激活函数选用线性整流函数(ReLU),初始学习率设置为 0.001,使用 Adam 优化器对网络结构进行优化。

1.4 输出层

IMTEM 的最后一层为输出层,负责处理来自前馈神经网络层的数据信息,生成最终的预测结果。

2 实例分析

2.1 地质背景与数据来源

四川盆地的龙马溪组页岩主要分布于志留系地层,以其丰富的有机质储层和天然气资源潜力而闻名^[30]。龙马溪页岩的沉积环境为深水海洋环境,这种低能环境有利于有机质的良好保存,并促进了细粒沉积物的积累,主要是粉砂质泥岩和页岩^[31]。该地层的有机质主要为 II 型干酪根,显示出较好的生烃潜力。龙马溪页岩的热成熟度在盆地内各不相同,最深处达到了天然气生成的最佳成熟窗口^[32]。丰富的有机质含量、适宜的孔隙度和渗透率以及广泛的分布面积,使龙马溪页岩成为了页岩油气勘探和开发的理想目标^[33]。本次工作的研究区域为四川盆地重庆东北方向的两口钻井,井位置如图 2 所示。在该区域,龙马溪页岩的累积厚

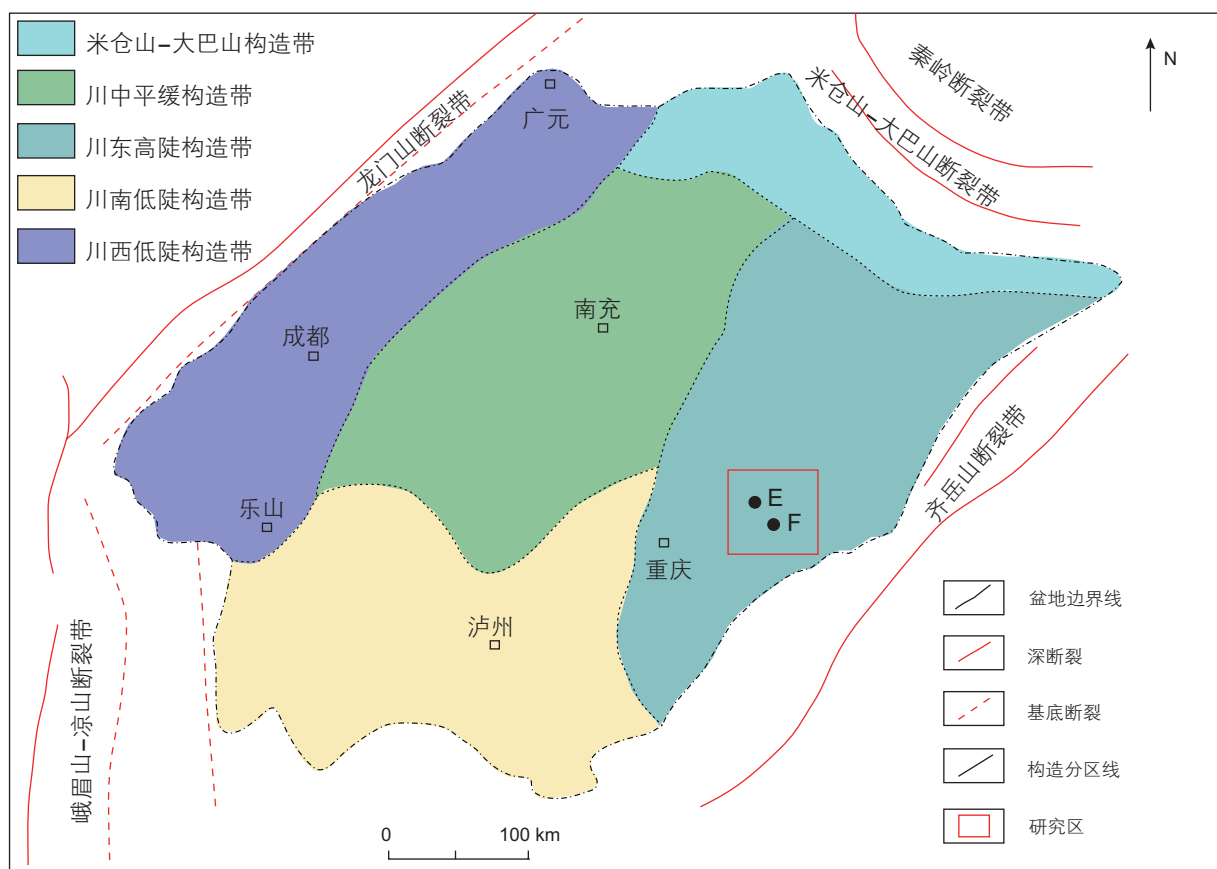


图 2 四川盆地及其构造单元划分图^[34]

Fig. 2 Map of the Sichuan Basin and its structural unit division^[34]

度可达 90 m。

本文采用Liu等^[34]公开的E井和F井的测井数据和岩心实验数据，其中测井数据包括自然伽马(*GR*)、声波时差(*AC*)、体积密度(*DEN*)、补偿中子(*CNL*)和深探测电阻率(*R_D*)，岩心实验数据包括E井 162 块岩心样品和F井 63 块岩心样品实际测量的TOC含量。两口井的岩心样品均来自四川盆地龙马溪组的龙一段、龙二段和龙三段地层，覆盖了龙马溪组的全部层系。表 1 总结了岩心样品的测井响应参数及 TOC 含量的统计结果，包括了最小值、最大值和平均值。可以观察到页岩样品具有明显的非均质性，尤其是在 TOC 含量方面：E 井岩心样品实测的 TOC 含量介于 0.140%~6.210%，平均值为 2.389%；F 井岩心样品实测的 TOC 含量介于 0.820%~5.130%，平均值为 2.257%。225 块岩心样品较好地反映了龙马溪组页岩 TOC 含量的纵向变化，可以用于建立模型完成页岩 TOC 含量的连续预测。

2.2 实验设计与参数优化

为了评估IMTEM的预测性能，采用E井和F井的测井数据和岩心实验数据进行对比分析。E井的 162 组岩心数据作为训练样本，F井的 63 组岩心数据作为测试样本。训练样本用于调整模型的网络参数，测试样本用于评估模型的泛化性能。在本研究中，输入数据为 *GR*、*AC*、*DEN*、*CNL* 和 *R_D*，目标参数为 TOC 含量。

IMTEM的预测结果与两种集成模型、5 种基础模型和 $\Delta\log R$ 方法的预测结果进行了对比，以展示IMTEM在回归建模中的优越性。为了获得模型的最佳性能，对其超参数进行优化。随机网格搜索是一种常用于机器学习和统计建模的超参数优化方法，在参数空间中以随机采样方式选择参数组合进行评估，在保证搜索效率的同时，能够有效降低计算成本^[7]。

本研究利用基于 5 折交叉验证的随机网格搜索方法对模型的超参数进行微调，其超参数配置如下：极

限梯度提升采用分类和回归树作为基础模型，树的数量为 250，树的最大深度为 1，学习率为 0.03；随机森林采用分类和回归树作为基础模型，树的数量为 100，树的最大深度为 12，随机采样技术的样本数据和特征参数的抽取比例为 80%；支持向量机采用径向基函数作为核函数，惩罚系数为 32.5，伽马参数为 0.024；极限学习机的隐藏层包含 6 个神经元，激活函数为 sigmoid 函数。此外，全连接神经网络的超参数配置与前馈神经网络层的参数设置保持一致。对于两种集成模型，分别采用算术平均法和加权求和法作为组合策略。其中采用加权求和法的集成模型利用粒子群优化算法确定算法模块的加权系数，极限梯度提升、随机森林、支持向量机和极限学习机的权重分别为-0.471、1.671、-0.028 和 -0.171。

2.3 模型对比与结果分析

图 3 和图 4 分别比较了两种集成模型和IMTEM预测F井的 TOC 含量(*I_{predicted}*)与岩心实测的 TOC 含量(*I_{measured}*)。两种集成模型预测的数据点相对分散在交会图的中轴线两侧，并显示出大量黑色和灰色数据点，说明预测结果与岩心分析数据之间存在较大误差。而IMTEM预测的数据点更加集中分布在交会图的中轴线上，其预测结果仅有 3 个数据点显示为灰色，体现出IMTEM具有更好的预测效果。

通过对比IMTEM和两种集成模型在预测 TOC 含量方面的性能评价指标(表 2)，可以看到，IMTEM的表现最好，其次是采用算术平均法的集成模型，而采用加权求和法的集成模型表现最差。相比于两种集成模型，IMTEM在决定系数上提高了 25.291%和 79.583%，在均方误差上降低了 55.676%和 73.420%，在平均绝对误差上降低了 33.634%和 49.829%。

通过对 5 种基础模型在预测 TOC 含量方面的性能评价指标的对比(表 3)，可以看到，支持向量机展现出最优的预测性能，其次是极限学习机，而随机

表 1 岩心测井响应及实验分析数据
Table 1 Logging responses and experimental data of core samples

井号		<i>GR</i> /API	<i>AC</i> /μs·ft ⁻¹	<i>DEN</i> /g·cm ⁻³	<i>CNL</i> /%	<i>R_D</i> /Ω·m	TOC/%
E 井	最小值	147.320	64.620	2.435	9.710	21.310	0.140
	最大值	305.980	87.930	2.713	25.940	101.320	6.210
	平均值	192.581	74.302	2.599	17.529	54.201	2.389
F 井	最小值	133.450	69.370	2.502	10.880	27.410	0.820
	最大值	197.290	83.660	2.689	22.770	238.050	5.130
	平均值	159.613	74.985	2.621	16.926	82.998	2.257

森林的预测效果最差。相较于支持向量机，IMTEM在决定系数上提高了 24.387%，在均方误差上降低了 54.945%，在平均绝对误差上降低了 30.824%。相较于

随机森林，IMTEM在决定系数上提高了 47.099%，在均方误差上降低了 66.599%，在平均绝对误差上降低了 43.678%。

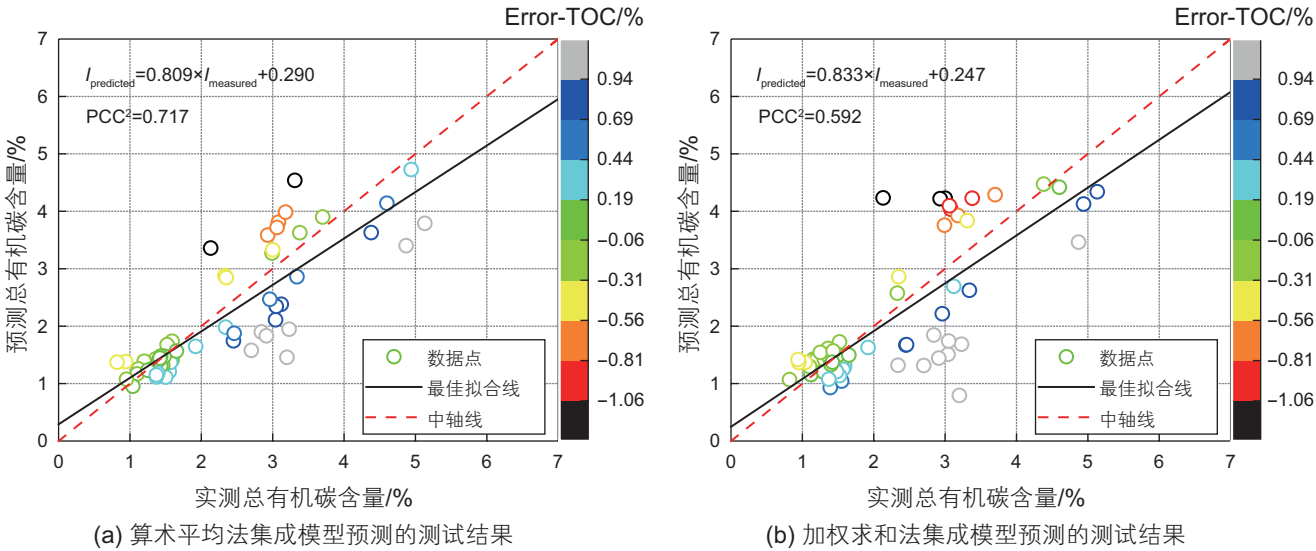


图 3 两种集成模型预测 F 井的 TOC 含量与岩心实测 TOC 含量对比交会图 (其中，Error-TOC 为实测 TOC 含量与预测 TOC 含量间的绝对误差，PCC 为皮尔逊相关系数)

Fig. 3 Comparison cross-plots between predicted TOC content in Well F from two ensemble models and measured core TOC content (where Error-TOC represents absolute error between measured and predicted TOC content and PCC represents the Pearson correlation coefficient)

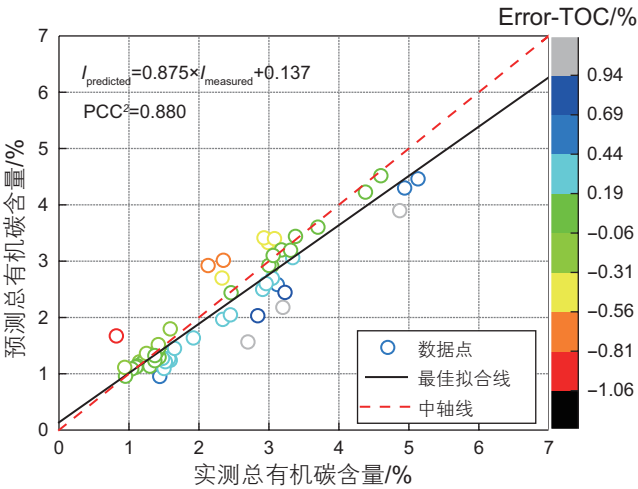


Fig. 4 Comparison cross-plot between the predicted TOC content in Well F from IMTEM model and measured core TOC content

表 2 集成模型预测性能评价表

Table 2 Evaluation of prediction performance from ensemble models

方法	决定系数	均方误差	平均绝对误差
IMTEM	0.862	0.164	0.294
采用算术平均法的集成模型	0.688	0.370	0.443
采用加权求和法的集成模型	0.480	0.617	0.586

采用IMTEM和 $\Delta\log R$ 方法预测F井的TOC含量。
由于缺少最高热解峰温和镜质体反射率等关键数据，
本研究采用胡慧婷等^[4]提出的简化 $\Delta\log R$ 方法：

$$\Delta\log R = \lg(R_D / 27.20) - 2.50 \times (\text{DEN} - 2.68) \quad (2)$$

$$\text{TOC} = 4.91 \times \Delta\log R + 0.02 \quad (3)$$

考虑到两口井数据中TOC含量与DEN之间的相
关系数最高，因此选择了DEN和 R_D 曲线计算TOC含
量。图 5 展示了 $\Delta\log R$ 方法和IMTEM计算F井的TOC

表 3 基础模型预测性能评价表
Table 3 Evaluation of prediction performance of baseline models

方法	决定系数	均方误差	平均绝对误差
极限梯度提升	0.612	0.461	0.487
随机森林	0.586	0.491	0.522
支持向量机	0.693	0.364	0.425
极限学习机	0.655	0.410	0.445
全连接神经网络	0.637	0.430	0.475

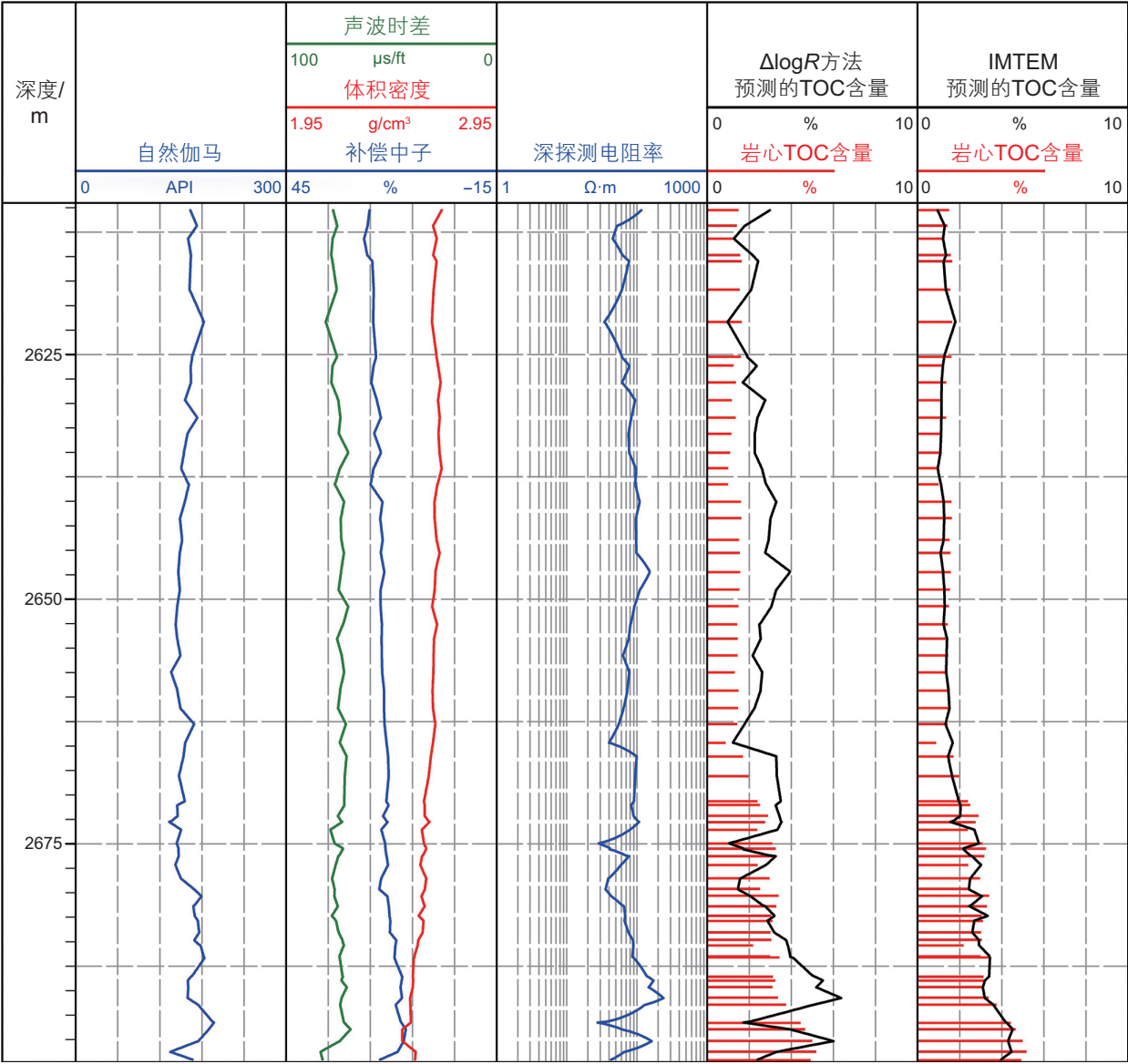


图 5 $\Delta\log R$ 方法和IMTEM模型预测的F井TOC含量
Fig. 5 Predicted TOC content in Well F by $\Delta\log R$ method and IMTEM model

含量结果。评价指标显示， $\Delta\log R$ 方法的决定系数为 -0.550 ，均方误差为 1.839 ，平均绝对误差为 1.181 ，而IMTEM的决定系数为 0.862 ，均方误差为 0.164 ，平均绝对误差为 0.294 。对比结果表明，IMTEM具有更好的泛化性能，能够更加准确的预测F井的TOC含量。此外，在 $2674.2\sim 2682.0\text{ m}$ 的深度区间， $\Delta\log R$ 方法的预测结果显著低于实测TOC含量。而在 $2610.3\sim 2618.7\text{ m}$ 、 $2624.9\sim 2674.0\text{ m}$ 和 $2684.8\sim 2692.1\text{ m}$ 的深度区间， $\Delta\log R$ 方法的预测结果则明显高于实测TOC含量。相比之下，IMTEM的预测结果在目标地层中与实测TOC含量的吻合程度更高，说明了本文提出的方法适用于页岩TOC含量的预测。

为了验证提出方法的有效性，将其应用到另一组实验：E井的162组岩心数据随机划分为129组训练样本和33组测试样本，F井的63组岩心数据随机划分为51组训练样本和12组测试样本，训练样本和测试样本的总数分别为180组和45组。

通过对比IMTEM和两种集成模型在预测TOC含量方面的性能评价指标(表4)，可以看到，IMTEM的表现最佳，其次是采用算术平均法的集成模型，而采用加权求和法的集成模型表现最差。相比于两种集成模型，IMTEM在决定系数上提高了3.712%和8.855%，在均方误差上降低了29.386%和48.889%，在平均绝对误差上降低了12.094%和27.317%。

通过对5种基础模型在预测TOC含量方面的性能评价指标的对比(表5)，可以看到，极限梯度提升展现出最优的预测性能，其次是全连接神经网络

络，而随机森林的预测效果最差。相较于极限梯度提升，IMTEM在决定系数上提高了2.902%，在均方误差上降低了24.766%，在平均绝对误差上降低了11.045%。相较于随机森林，IMTEM在决定系数上提高了7.460%，在均方误差上降低了45.051%，在平均绝对误差上降低了22.396%。

图6展示了 $\Delta\log R$ 方法和IMTEM计算E井的TOC含量结果。评价指标显示， $\Delta\log R$ 方法在测试样本中的决定系数为 0.649 ，均方误差为 0.803 ，平均绝对误差为 0.741 ，而IMTEM在测试样本中的决定系数为 0.926 ，均方误差为 0.171 ，平均绝对误差为 0.308 。对比结果表明，IMTEM具有更好的预测性能。此外，在 $2769\sim 2780\text{ m}$ 和 $2808\sim 2835\text{ m}$ 的深度区间内， $\Delta\log R$ 方法的预测结果显著低于实测TOC含量，而在 $2783\sim 2805\text{ m}$ 的深度区间内，其预测结果则显著高于岩心分析结果。相较于 $\Delta\log R$ 方法，IMTEM的预测结果与实测TOC含量更加吻合。值得注意的是，在 $2744.3\sim 2745.5\text{ m}$ 的深度区间，由于AC偏高，IMTEM高估了TOC含量。 $\Delta\log R$ 方法的计算结果受DEN和 R_D 的影响，较低的 R_D 导致其预测结果出现负值，违反了岩石物理规律，也显示出传统测井评价方法的局限性。这些差异性体现了IMTEM在预测TOC含量方面的准确性和稳定性。

图7展示了 $\Delta\log R$ 方法和IMTEM计算F井的TOC含量结果。评价指标显示， $\Delta\log R$ 方法在测试样本中的决定系数、均方误差和平均绝对误差分别为 0.113 、 1.190 和 0.896 ，而IMTEM在测试样本中的决定系

表 4 集成模型预测性能评价指标对比

Table 4 Comparison of evaluation metrics for predictive performance of ensemble models

方法	决定系数	均方误差	平均绝对误差
IMTEM	0.922	0.161	0.298
采用算术平均法的集成模型	0.889	0.228	0.339
采用加权求和法的集成模型	0.847	0.315	0.410

表 5 基础模型预测性能评价指标对比

Table 5 Comparison of evaluation metrics for predictive performance of baseline models

方法	决定系数	均方误差	平均绝对误差
极限梯度提升	0.896	0.214	0.335
随机森林	0.858	0.293	0.384
支持向量机	0.876	0.255	0.363
极限学习机	0.881	0.244	0.356
全连接神经网络	0.894	0.218	0.346

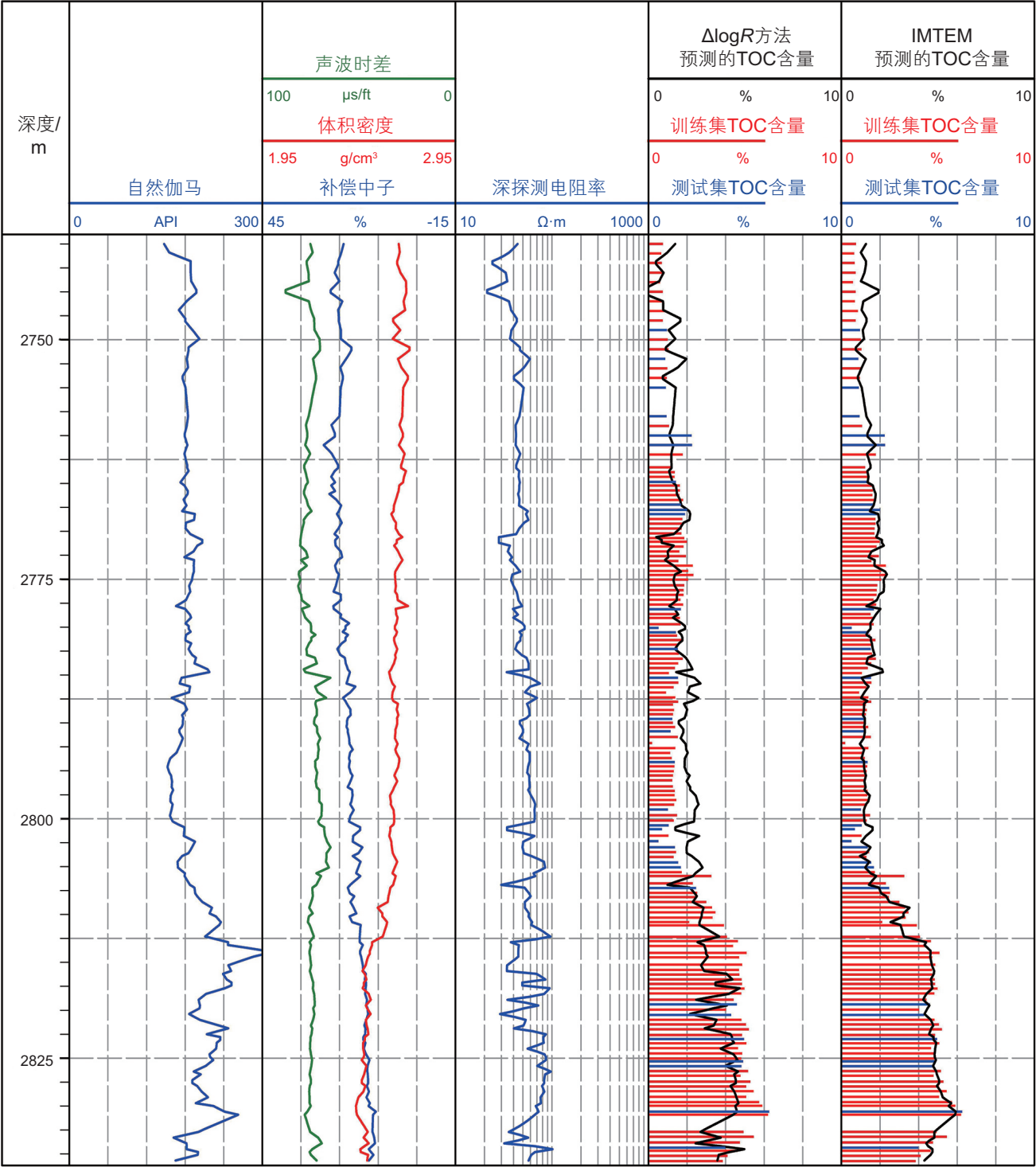


图 6 E 井 TOC 含量预测结果

Fig. 6 Prediction results of TOC content in Well E

数、均方误差和平均绝对误差分别为 0.902、0.132 和 0.269。这些指标表明了 $\Delta\log R$ 方法的预测结果与实测 TOC 含量存在较大偏差，而 IMTEM 的预测结果与岩心分析结果具有较高的一致性。在 2674~2685 m 的深度区间， $\Delta\log R$ 方法的预测结果显著低于实测 TOC 含

量。而在 2629~2670 m 和 2688~2691 m 的深度区间， $\Delta\log R$ 方法的预测结果则明显高于实测 TOC 含量。相比之下，IMTEM 的预测结果在目标地层中与实测 TOC 含量的吻合程度更高，说明了本文提出的方法适用于预测页岩 TOC 含量。

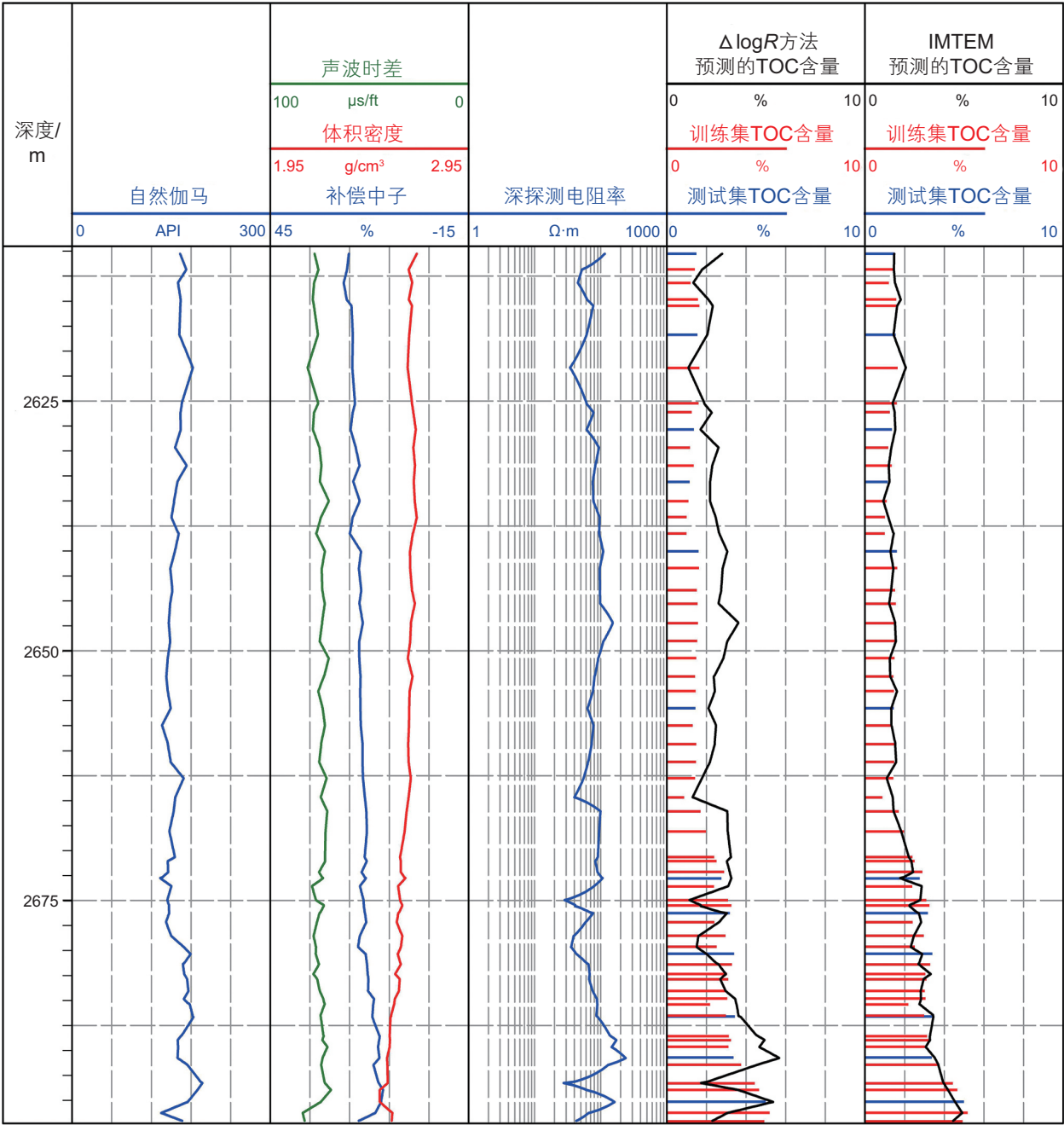


图 7 F 井 TOC 含量预测结果
Fig. 7 Prediction results of TOC content in Well F

3 结论

1) IMTEM 利用算法层整合多种类型的智能算法，并采用数据增强技术和智能匹配技术建立联合特征与目标参数之间的非线性关联，能够显著增强模型的泛化能力。经测试，该方法具有比单一机器学习方法更加准确的预测结果。

2) 智能匹配技术利用前馈神经网络出色的模式拟合能力，能够提取联合特征的深层特征信息。相比于算术平均法和加权求和法，该技术具有更强的特征表达能力。

3) 将 IMTEM 应用于四川盆地龙马溪组页岩 TOC 含量预测，与传统的 $\Delta\log R$ 方法相比，提出方法的预测结果与岩心实测 TOC 含量的吻合程度更高，该方法有助于更加准确地认识和评价地层的生烃潜力。

参考文献

- [1] 赵子斌, 侯读杰, 程熊, 等. 边缘凹陷烃源岩有机质丰度曲面预测法——以渤海湾盆地莱州湾东北洼为例[J]. 石油地球物理勘探, 2018, 53(2): 347–354. [ZHAO Z B, HOU D J, CHENG X, et al. A fitting prediction method of TOC in source rocks in a marginal sag: A case study of northeastern sag in Laizhou Bay, Bohai Bay Basin[J]. Oil Geophysical Prospecting, 2018, 53(2): 347–354.]
- [2] 张博, 姜振学, 原园, 等. 灰色关联度分析法在筛选页岩含气量主控因素中的应用[J]. 石油科学通报, 2018, 3(2): 134–143. [ZHANG B, JIANG Z X, YUAN Y, et al. Grey correlation analysis to elucidate the main controlling factors of shale gas content[J]. Petroleum Science Bulletin, 2018, 3(2): 134–143.]
- [3] PASSEY Q R, CREANEY S, KULLA J B, et al. A practical model for organic richness from porosity and resistivity logs[J]. AAPG Bulletin, 1990, 74(12): 1777–1794.
- [4] 胡慧婷, 卢双舫, 刘超, 等. 测井资料计算源岩有机碳含量模型对比及分析[J]. 沉积学报, 2011, 29(6): 1199–1205. [HU H T, LU S F, LIU C, et al. Models for calculating organic carbon content from logging information: Comparison and analysis[J]. Acta Sedimentologica Sinica, 2011, 29(6): 1199–1205.]
- [5] ZHAO Z B, LITKE R, ZIEGER L, et al. Depositional environment, thermal maturity and shale oil potential of the Cretaceous Qingshankou Formation in the eastern Changling Sag, Songliao Basin, China: An integrated organic and inorganic geochemistry approach[J]. International Journal of Coal Geology, 2020, 232: 103621.
- [6] BIONE F R A, VENANCIO I M, SANTOS T P, et al. Estimating total organic carbon of potential source rocks in the Espírito Santo Basin, SE Brazil, using XGBoost[J]. Marine and Petroleum Geology, 2024, 162: 106765.
- [7] RONG J, ZHENG Z, LUO X, et al. Machine learning method for TOC prediction: Taking Wufeng and Longmaxi shales in the Sichuan Basin, Southwest China as an example[J]. Geofluids, 2021, 2021: 6794213.
- [8] SAPORETTI C M, FONSECA D L, OLIVEIRA L C, et al. Hybrid machine learning models for estimating total organic carbon from mineral constituents in core samples of shale gas fields[J]. Marine and Petroleum Geology, 2022, 143: 105783.
- [9] WANG X, LIU G D, WANG X L, et al. Geophysical prediction of organic matter abundance in source rocks based on geochemical analysis: A case study of southwestern Bozhong Sag, Bohai Sea, China[J]. Petroleum Science, 2024, 21(1): 31–53.
- [10] 魏明强, 周金鑫, 段永刚, 等. 基于机器学习的页岩气总有机碳含量预测模型[J]. 科学技术与工程, 2023, 23(30): 12917–12925. [WEI M Q, ZHOU J X, DUAN Y G, et al. Prediction model of total organic carbon content in shale gas based on machine learning[J]. Science Technology and Engineering, 2023, 23(30): 12917–12925.]
- [11] 李键, 陈桂, DAVID C, 等. 一种基于ConvLSTM神经网络的TOC含量地震预测方法[J]. 地球物理学进展, 2023, 38(3): 1143–1151. [LI J, CHEN G, DAVID C, et al. Seismic prediction method of TOC content based on a ConvLSTM neural network[J]. Progress in Geophysics, 2023, 38(3): 1143–1151.]
- [12] BAI Y, TAN M. Dynamic committee machine with fuzzy-c-means clustering for total organic carbon content prediction from wireline logs[J]. Computers & Geosciences, 2021, 146: 104626.
- [13] 谭茂金, 白洋, 吴静, 等. 多源数据驱动下委员会机器测井解释研究进展[J]. 石油物探, 2022, 61(2): 224–235. [TAN M J, BAI Y, WU J, et al. Progress of research on committee machine logging interpretation methods driven by multi-source data[J]. Geophysical Prospecting for Petroleum, 2022, 61(2): 224–235.]
- [14] 白洋, 谭茂金, 肖承文, 等. 致密砂岩气藏动态分类委员会机器测井流体识别方法[J]. 地球物理学报, 2021, 64(5): 1745–1758. [BAI Y, TAN M J, XIAO C W, et al. Dynamic classification committee machine-based fluid typing method from wireline logs for tight sandstone gas reservoir[J]. Chinese Journal of Geophysics, 2021, 64(5): 1745–1758.]
- [15] BAI Y, TAN M J, SHI Y, et al. Regression committee machine and petrophysical model jointly driven parameters prediction from wireline logs in tight sandstone reservoirs[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 60: 1–9.
- [16] KHOSHNOODKIA M, MOHSENI H, RAHMANI O, et al. TOC determination of Gadvan Formation in South Pars Gas field, using artificial intelligent systems and geochemical data[J]. Journal of Petroleum Science and Engineering, 2011, 78(1): 119–130.
- [17] TAN M J, BAI Y, ZHANG H T, et al. Fluid typing in tight sandstone from wireline logs using classification committee machine[J]. Fuel, 2020, 271: 117601.
- [18] 朱林奇, 张冲. 谱聚类-Adaboost集成数据挖掘算法在岩性识别中的应用[J]. 中国科技论文, 2016, 11(5): 545–550. [ZHU L Q, ZHANG C. Application of spectral clustering-Adaboost integrated data mining algorithm in lithology identification[J]. China Sciencepaper, 2016, 11(5): 545–550.]
- [19] KARIMPOULI S, FATHIANPOUR N, ROOHI J. A new approach to improve neural networks' algorithm in permeability prediction of petroleum reservoirs using supervised committee machine neural network (SCMNN)[J]. Journal of Petroleum Science and Engineering, 2010, 73(3–4): 227–232.

- [20] ZHANG H, WU W, WU H. TOC prediction using a gradient boosting decision tree method: A case study of shale reservoirs in Qinshui Basin[J]. *Geoenergy Science and Engineering*, 2023, 221: 111271.
- [21] ZHAO P, HOU Y, MA F, et al. Genetic mechanisms and identification of low-resistivity pay zones: A case study of Pengyang area, Ordos Basin, China[J]. *Geofluids*, 2022, 2022: 3299768.
- [22] 初勇志, 刘成林, 太万雪, 等. 基于支持向量机(SVM)的不同咸化程度烃源岩总有机碳含量预测模型[J]. *石油实验地质*, 2022, 44(4): 739–746. [CHU Y Z, LIU C L, TAI W X, et al. Prediction model of TOC contents in source rocks with different salinity degrees based on Support Vector Machine (SVM)[J]. *Petroleum Geology & Experiment*, 2022, 44(4): 739–746.]
- [23] RUI J, ZHANG H, ZHANG D, et al. Total organic carbon content prediction based on support-vector-regression machine with particle swarm optimization[J]. *Journal of Petroleum Science and Engineering*, 2019, 180: 699–706.
- [24] HUANG G B, CHEN L, SIEW C K. Universal approximation using incremental constructive feedforward networks with random hidden nodes[J]. *IEEE Transactions on Neural Networks*, 2006, 17(4): 879–892.
- [25] LI M B, HUANG G B, SARATCHANDRAN P, et al. Fully complex extreme learning machine[J]. *Neurocomputing*, 2005, 68: 306–314.
- [26] SHI S Z, SHI G F, PEI J B, et al. Porosity prediction in tight sandstone reservoirs based on a one-dimensional convolutional neural network-gated recurrent unit model[J]. *Applied Geophysics*, 2023, 20(4): 1–13.
- [27] SUN Y, PANG S, ZHANG Y, et al. Application of the dynamic transformer model with well logging data for formation porosity prediction[J]. *Physics of Fluids*, 2024, 36(3): 036620.
- [28] SHAN L, LIU Y, TANG M, et al. CNN-BiLSTM hybrid neural networks with attention mechanism for well log prediction[J]. *Journal of Petroleum Science and Engineering*, 2021, 205: 108838.
- [29] WANG X, LIAO G, XIAO L, et al. Theory-driven neural network for total organic carbon content estimation from well logs[J]. *Journal of Geophysics and Engineering*, 2024, 21(2): 685–697.
- [30] LI H, XIE L, REN L, et al. Influence of CO₂-water-rock interactions on the fracture properties of sandstone from the Triassic Xujiahe Formation, Sichuan Basin[J]. *Acta Geophysica*, 2021, 69: 135–147.
- [31] CHEN Q, DONG Y, TAN X, et al. Applying the tilt-depth and tilt-euler techniques of gravity data to decipher the basement depth in Sichuan Basin, China[J]. *Acta Geophysica*, 2021, 69: 2173–2186.
- [32] CHEN S, ZHU Y, WANG H, et al. Shale gas reservoir characterisation: A typical case in the southern Sichuan Basin of China[J]. *Energy*, 2011, 36(11): 6609–6616.
- [33] 杨鑫翊, 王民, 白雪峰, 等. 川东北地区侏罗系凉高山组页岩储集空间特征及页岩油可动性探索[J]. *石油科学通报*, 2024, 9(2): 196–212. [YANG X Y, WANG M, BAI X F, et al. Reservoir space characteristics and exploration of shale oil mobility of the Jurassic Lianggaoshan Formation shale in the northeastern Sichuan Basin[J]. *Petroleum Science Bulletin*, 2024, 9(2): 196–212.]
- [34] LIU C, ZHAO W, SUN L, et al. An improved $\Delta\log R$ model for evaluating organic matter abundance[J]. *Journal of Petroleum Science and Engineering*, 2021, 206: 109016.

(编辑 付娟娟)

第一作者: 王宵宇(1996年—), 博士研究生, 从事测井评价与资料解释等方面的研究工作, wangxiaouyu_cup@163.com。

通信作者: 廖广志(1981年—), 博士, 教授, 从事地球物理测井、油气人工智能、核磁共振测井等方面的研究工作, liaoguangzhi@cup.edu.cn。