

基于 RF-Transformer 在气藏开发潜力评价中的应用研究

侯耀晗¹

HOU Yaohan

摘要

评价气藏开发潜力对气藏开发具有重要的指导意义，针对传统方法的不足，文章提出了一种基于 RF-Transformer 的致密气藏产能预测新方法。模型使用 Transformer 进行特征提取，通过随机森林进行气藏产能的分类预测。研究首先引入储层品质因子作为新特征参数，采用 SMOTE-Tomek 算法进行数据增强，解决了类别不平衡问题。并根据数据的特点，设计了串行型的 RF-Transformer 模型。实验结果表明，RF-Transformer 模型在预测致密气藏产能各类别中均为最优值，为气藏开发提供了具有参考价值的信息。

关键词

气藏产能；开发潜力；储层品质因子；数据不平衡；深度学习模型

doi: 10.3969/j.issn.1672-9528.2025.09.039

0 引言

随着能源需求的持续增长，天然气作为一种清洁能源在全球能源结构中的地位日益突出。准确地评价气藏开发潜力能够快速定位到具有开发潜力的气井层位，对合理规划开发方案、优化资源配置、提高气藏开采效率具有重要意义。然而，由于气藏地质条件复杂、渗流机理多样、数据分布不均衡等因素，传统的气藏产能预测方法往往依赖于经验公式和传统机器学习方法^[1-5]。经验公式往往需要大量的测井参数，在实际应用中存在着数据获取困难，计算成本较高等问题。而传统机器学习方法如支持向量机（support vector machine, SVM），随机森林（random forest, RF）等在产能潜力评价中取得了一定的成效，但这些方法在处理高维特征和复杂非线性关系时仍存在一定的局限性。

近十年来深度学习技术不断发展，由于传统机器学习方法存在的局限性，越来越多的研究人员采用深度学习方法来预测气藏产能^[6]，特别是 Transformer 方法凭借其强大的特征提取和长程依赖建模能力，在序列数据处理领域取得了突破性进展。

为克服单一深度学习网络预测时产生的问题，本文提出了一种新型的集成模型——串行型 RF-Transformer 集成模型。模型将 Transformer 算法的特征提取能力与随机森林的分类决策能力相结合，实现了两种方法的优势互补。模型在保留随机森林对特征空间决策分类能力的同时，利用多头自注意力机制捕捉特征间的深层关联，通过对比学习增强特征表达，并将其应用于我国西北部某致密气田的产量预测中，与其他机器学习方法进行对比实验，验证该方法的可行性。

1 数据预处理

1.1 数据异常值处理

孤立森林（Isolation Forest）算法是一种基于集成学习的无监督异常检测方法，其核心思想源于异常样本较易被“孤立”的特性^[7]。在本研究中采用其对数据进行异常分析。算法通过构建多棵隔离树（iTree）来实现异常检测，每棵树通过随机选择特征和分割点，递归地将样本空间进行二分。由于异常点通常分布稀疏且远离正常数据簇，故在隔离树中具有较短的路径长度，本研究设定约 5% 的样本点为异常值，采用 100 棵决策树构建森林以确保模型的稳定性和可靠性。在具体实现中，首先对特征数据进行标准化处理，以消除量纲影响。随后，通过计算每个样本点的异常得分（anomaly score）来量化其异常程度，得分越低表示该数据越可能是异常值。如图 1 所示，浅色圆点是正常值，深色圆点为检测出的异常值，浅色线条为决策边界，检测完成之后将异常值移除数据，便于之后的数据处理。

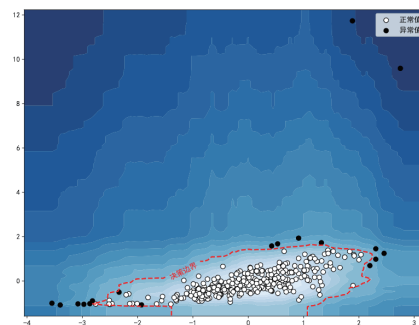


图 1 孤立森林异常值检测结果

1.2 数据集介绍

数据集来自西北地区某气田井段，数据共包含各种测井

参数和产量类别（I、II、III类井）。无阻流量越高则代表该井段的气藏开发潜力越大，经过异常处理之后，数据共计404条，选取其中80%训练集，剩余20%为测试集。表1为部分测试数据。

表1 部分测试数据

井号	层位	无阻流量	类别
D1-1	H8-2	1 502	I
D1-1	H8-3	7 570	I
D3	S1	10 943	II
D38-1	H8-1	108 976	III

1.3 特征参数增强

在气藏工程领域，储层的物理特性，特别是渗透率和孔隙度，是评估储层品质的核心参数，这些参数通常被用来独立评价储层的渗透能力和储集能力。然而，这种独立的评价方法忽视了渗透率和孔隙度之间的相互作用，可能无法全面反映储层的综合物性特征。为了克服这一局限性，本研究引入了储层品质因子（reservoir quality index, RQI）^[8]，作为一个综合性指标，RQI能够更全面地评估储层的渗流特性。

储层品质因子用公式表示为：

$$RQI = 0.0314 \sqrt{\frac{K}{\phi}} \tag{1}$$

式中： K 代表渗透率； ϕ 代表孔隙度；常数0.0314代表一个单位换算系数，使RQI能够与其他微观尺度的地质参数进行直接比较。

根据式（1）计算得到一个能够综合反映储层渗流特性的新特征参数。

1.4 数据类别不平衡处理

在气藏产能分类预测中，由于III类井数量相对较少，数据集存在明显的类别不平衡问题。为了解决这一问题，本研究采用SMOTE-Tomek链接混合重采样方法^[9]。该方法首先使用SMOTE（synthetic minority over-sampling technique）算法对少数类样本进行过采样，然后使用Tomek Links方法清理可能的噪声和重叠样本，从而获得更均衡的数据集。原始训练集和重采样之后的训练集如表2所示。

表2 训练集数据对比

类别	原始训练集数据	重采样后数据
I类井	211	211
II类井	96	210
III类井	16	203

2 模型结构和原理

2.1 Transformer

Transformer是一种基于注意力机制的深度学习模型，最初由Vaswani等人^[10]在2017年的论文《Attention is All You Need》中提出。主要用于处理序列到序列的任务，如机器翻译、文本摘要等。Transformer模型的核心思想是通过自注意力（Self-Attention）机制来捕捉序列内部的依赖关系，这种机制可以直接计算序列中任意两个位置之间的关系，从而有效地解决了传统循环神经网络在处理长序列时效率低下和梯度消失的问题^[11]。Transformer中最大的亮点是多头注意力机制。如图2所示，单个注意力头的计算基于查询（Query, Q ）、键（Key, K ）和值（Value, V ）3个矩阵，其中 Q 为查询矩阵，表示当前位置的特征； K 是键矩阵，用于与查询计算相似度； V 是值矩阵用于在计算出查询 Q 与键 K 的相似度权重后，依据式（2），通过加权求和运算，输出最终包含关键信息的注意力结果。

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

为防止点积结果过大导致softmax梯度消失问题，引入缩放因子 $\sqrt{d_k}$ （ d_k 为 K 的维度），将 Q 和 K 点乘的结果除以 $\sqrt{d_k}$ 之后再行归一化处理，最后再与 V 相乘得到最终结果。

多头注意力的计算根据式（3）将输入投影到 h 个不同的子空间，并行计算 h 个注意力头，然后将结果拼接并投影回原始维度：

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^O \tag{3}$$

基于此，完整的Transformer特征提取器采用层叠式结构由多个Transformer块串联组成。如图2所示，每个Transformer块都包含多头自注意力层、层归一化和前馈神经网络，并通过残差连接形成信息传递。

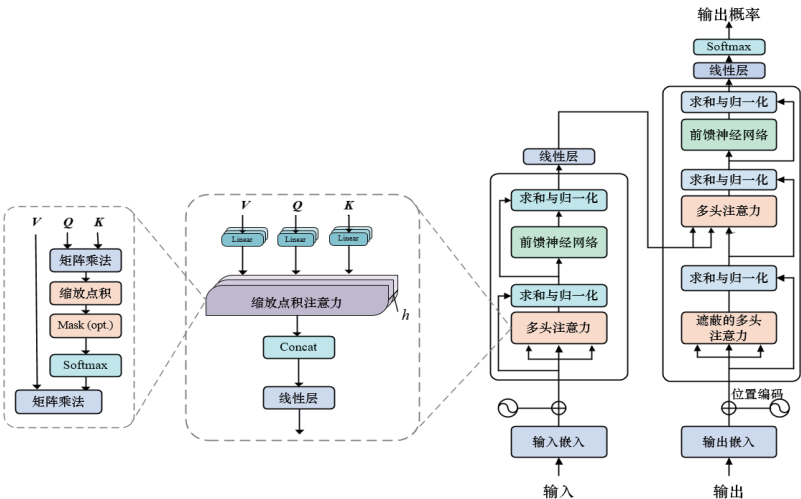


图2 Transformer模型结构图

2.2 RF-Transformer 模型

针对单一模型在特征提取和分类决策上的局限性，本研究提出了一种新颖的 RF-Transformer 串行模型，通过特征变换与决策分离的机制，将深度学习的特征提取能力与传统机器学习的分类优势有机整合。模型结构如图 3 所示，该串行模型首先利用 Transformer 的多头自注意力机制和对比学习对数据集进行深度特征提取，再通过随机森林模型实现分类，形成了“特征提取—分类决策”的串行架构。

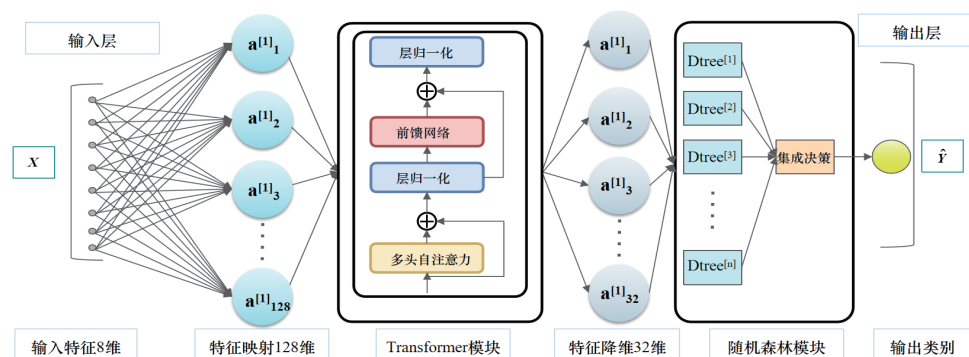


图 3 RF-Transformer 模型结构图

算法的具体实施步骤如表 3 所示，这种设计结合了 Transformer 模型和随机森林模型，形成了一个“Transformer 提取特征到随机森林分类”的串行处理流程。并且由于充分利用了 Transformer 优秀的捕捉特征间复杂关系的能力，将提取后并变换的特征输入随机森林，从而在保持决策树模型可解释性的同时提升了整体分类性能。

表 3 RF-Transfomer 产能分类预测算法

RF-Transformer 产能分类预测算法：	
输入：致密气藏测井数据；	
输出：产量类别，评价指标：预测准确率；	
1. 致密气藏测井数据进行数据预处理；	
2. 将数据预处理后进行特征筛选，选出关键的 8 个特征；	
3. 将处理过的数据划分成训练集和测试集，以便模型的训练和测试；	
4. 构建 RF-Transfomer 模型并训练：	
4.1 输入特征经特征映射层扩展到 128 维；	
4.2 经过 3 层 Transformer 块处理，每层包含 4 头注意力机制、特征扩展和前馈降维层，将特征以 128 维的形式输出；	
4.3 对提取的特征进行降维，从 128 维降至 32 维，以 32 维输出；	
4.4 随机森林模型输入 32 维特征，并根据“sqrt”特征选择策略，结合多棵决策树的投票结果，得出最终分类预测和概率；	
4.5 根据模型每一轮的训练结果，优化网络参数；	
5. 预测完成之后，根据训练好的模型参数，将测试集代入模型，得到测试结果，并且计算预测准确率和混淆矩阵；	
6. 返回最终的预测结果。	

3 实验过程

3.1 气藏产能影响特征筛选

气藏数据具有显著的非线性特征、多尺度分布和强专业相关性等特点：储层参数（如渗透率、孔隙度等）之间存在复杂的非线性关系，且不同参数的量纲和数值范围差异较大；同时，由于地质条件的非均质性，数据往往呈现出高度的空间异质性和多模态分布特征；基于这些特点，本研究选择了基于 F -score 的特征选择方法，该方法的优势在于它不对数据

分布做任何假设，能够有效处理非正态分布和非线性关系评分机制，能够直观地反映不同储层参数对产能的影响程度，为气藏开发方案的优化提供了可靠的理论依据。实践表明，该方法筛选出的特征集提高了模型的预测准确率，充分验证了该方法在气藏产能开发潜力评价中的适用性和有效性。

从图 4 可以看出，渗透率和储层厚度的相关性最强，是影响气井产能的最关键因素，这是由于在致密气藏中，渗透率直接决定了气体在储层中的通过能力。气层厚度直接决定了气井的可动用储量，气层厚度越大往往表明该砂体位于更好的储藏位置。而孔隙度大小则直接关系到储层的储气能力，是评价储层品质的基础参数。储层品质因子作为渗透率和孔隙度的综合表征参数，影响同样重要。在测井参数方面，钾含量、自然伽马、密度反映了储层的岩性特征和矿物组成，与气井产能相关性也很强。

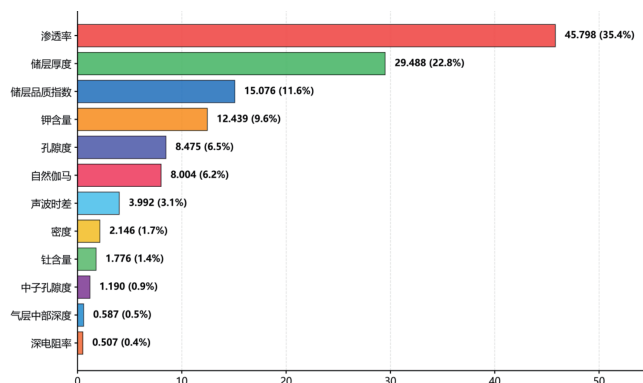


图 4 各特征 F_1 -score 排名图

结合上述分析结果，最后选择渗透率、储层厚度、储层品质因子、钾含量、孔隙度、自然伽马、声波时差、密度作为最终筛选参数。

3.2 模型超参数寻优

模型超参数的设置往往对模型性能有决定性的作用，因

此，合理地设置超参数是十分有必要的。预设超参数寻优的范围如表 4 所示。模型对该范围的超参数进行网格寻优，在每一次训练完成后保存训练结果和超参数结果。

表 4 参数寻优列表

参数名称	搜索范围
Transformer 层数	[2,3,4]
Transformer 特征维度	[32, 64, 128]
学习率	[0.1, 0.01, 0.001]
批次大小	[16, 32, 64, 128]
决策树数量	[64, 128, 256]
丢弃率	[0.1, 0.2, 0.3, 0.4]
优化器	AdamW
激活函数	GELU

经过网格寻优之后得到最佳超参数，优化器使用 AdamW，该优化器可以自适应调节学习率，使得模型更快的拟合。经网格寻优之后得到的最优超参数如表 5 所示。

表 5 模型最优超参数

参数名称	参数取值
Transformer 层数	3
Transformer 特征维度	128
学习率	0.001
批次大小	32
决策树数量	200
优化器	AdamW
激活函数	GELU

4 实验结果与分析

将参数应用模型并进行训练 20 次后，取各类准确率的均值，与各项模型均训练 20 次做对比，运行结果如表 6 所示。

表 6 模型运行结果

模型	单位：%			
	I 类井	II 类井	III 类井	综合
DecisionTree	74.62	74.58	43.75	73.09
LightGBM	81.60	78.54	50	79.14
RandomForest	77.55	79.16	50	76.54
XGBoost	83.11	81.04	50	80.86
Transfomer	71.69	79.16	75	74.07
本文模型	83.01	83.33	75	82.72

基于 20 次训练评估的统计结果，可以看出本文模型展现出了较好的预测效果，具体来说，本文模型以 82.72% 的综合准确率优于其他模型，在 I 类井和 II 类井的预测上分别达到了 83.01% 和 83.33% 的高准确率，同时在 III 类井中达

到了 75% 的预测准确率，相比 XGBoost、RandomForest 等传统模型有显著提升。特别是在 III 类井的预测上，本文模型展现出了明显优势，准确率达到 75%，高于其他模型的 43.75%~50% 的准确率水平。

综上所述，本文的 RF-Transformer 串行模型，不仅提升了各类别预测的准确率，而且在保持高准确率的同时实现了类别识别的均衡性，这都说明了 RF-Transformer 在致密气藏产量潜力预测方面的准确性，更能帮助相关开发人员定位到具有一定开发潜力的 II 类井和具有相当开发潜力的 III 类井，能够为气井开发提供一定的帮助。

5 结论

为了克服致密气藏产量潜力评价过度依赖于经验公式以及传统机器学习算法的问题，本文提出一种基于 RF-Transformer 的串行性网络结构，与传统机器学习和深度学习方法相比，RF-Transformer 综合表现最为优异，能够快速定位到具有开发意义的 III 类井，更具实用价值。虽然本文方法在处理复杂数据时表现出色，但由于气井的实际情况未能收集到更多的储层参数数据，考虑在后续的研究中，扩充致密气藏数据集，进一步优化模型，提高预测准确率。

参考文献:

[1] 王雯娟, 雷霄, 鲁瑞彬, 等. 南海西部异常高温高压气藏区城产能预测技术 [J]. 地球科学, 2019,44(8):2636-2642.

[2] 王成旺, 李曙光, 王玉斌, 等. 气藏产能测井预测方法研究综述 [J]. 科学技术与工程, 2023, 23(28):11927-11936.

[3] HE J, LIU Z W, ZHANG H, et al. Study on development methods of different types of gas wells in tightsandstone gas reservoirs[EB/OL].(2023-09-29)[2025-04-25].<https://www.nature.com/articles/s41598-023-43640-7>.

[4] 曹朋亮, 杨小松, 张一果, 等. 致密低渗气藏水平井新的产能评价“一点法” [J]. 油气井测试, 2019,28(6):1-8.

[5] 刘成川, 陈俊, 黎华继, 等. 中江气田沙溪庙组气藏致密砂岩储层测井评价 [J]. 石油物探, 2020,59(1):131-140.

[6] 钱玉贵. 机器深度学习技术在致密砂岩储层预测中的应用: 以川西坳陷新场须家河组为例 [J]. 油气藏评价与开发, 2023, 13(5): 600-607.

[7] ZHANG L, LIU L Y.Data anomaly detection based on isolation forest algorithm[C/OL]//2022 International Conference on Computation, Big-Data and Engineering (ICCBE).Piscataway:IEEE,2022[2025-03-13].<https://ieeexplore.ieee.org/document/9888169>.DOI:10.1109/ICCBE56101.2022.9888169.

深度学习下计算机通信网络动态跟踪异常数据流辨识

卜莹雪¹
BU Yingxue

摘 要

计算机通信网络入侵攻击往往混合多种协议,攻击者会依据网络防御策略的调整而改变攻击方式,使得异常数据流的特征随之发生变化。依赖固定阈值的辨识方法,其特征提取能力有限,无法充分挖掘数据在不同尺度和频率上的特征信息,导致辨识准确度不高,且在辨识过程中产生脏数据。为此,文章提出了一种深度学习下计算机通信网络动态跟踪异常数据流辨识方法。通过数据预处理环节标准化处理原始数据,利用降噪自编码器进行降维处理。在特征提取阶段,构建了包含卷积层、池化层和全连接层的深度学习模型,结合小波包分析和 Soft 分类器,实现高区分度的特征提取。设计多状态特征辨识矩阵,通过动态调整矩阵结构来适应不同网络环境,基于该矩阵构建异常数据流辨识模型,引入数据包最大丢失率作为关键阈值实现首次辨识;运用动态跟踪方法将辨识结果进行验证,实现对计算机通信网络动态跟踪异常数据流的精准辨识。实验结果表明,所提方法的数据流辨识准确度高达 90%,在辨识过程中产生的脏数据区间为 0 个,能够达到良好的异常数据辨识效果。

关键词

深度学习;计算机;网络入侵;数据流;辨识

doi: 10.3969/j.issn.1672-9528.2025.09.040

0 引言

计算机通信网络已成为社会运转和经济发展的关键基础设施。随着网络规模的不断扩大、网络应用的日益复杂以及网络攻击手段的持续升级,计算机通信网络面临着前所未有的安全挑战。其中,异常数据流的产生和传播对网络的安全性和稳定性构成了严重威胁,可能导致数据泄露、服务中断、系统瘫痪等严重后果。因此,如何有效地动态跟踪计算机通信网络中的异常数据流并进行准确辨识,成为当前网络安全领域亟待解决的重要问题。

文献[1]通过采样和差分隐私技术,为数据流中的最近

数据集快速创建可以安全发布的差分隐私直方图。根据用户的隐私保护强度,通过隐私分配的方式自适应生成最终可发布直方图。文献[2]通过主动选择最具信息量的样本进行标注和学习,提高模型对非平衡数据流的分类性能。将多个分类器集成在一起,通过投票方式提高算法对非平衡数据流的适应性。以上两种方法没有充分考虑异常数据流特征随网络环境变化的动态性。当攻击模式发生变化时,无法提取高区分度的特征,而这些特征对于准确辨识异常数据流非常重要,从而影响整体的辨识效果。基于此,本研究以计算机通信网络动态跟踪异常数据流辨识作为研究对象,结合深度学习技术,开展新的研究分析。该方法的创新点如下:

(1)在数据预处理环节,对原始数据进行标准化处理,

1. 江西飞行学院 江西南昌 330088

[8] 刘诗琼,石强,汤继周,等.基于球缝模型优化反演的裂缝储层评价新方法[J].地球物理学报,2022,65(4):1451-1460.

[9] 李艳霞,柴毅,胡友强,等.不平衡数据分类方法综述[J].控制与决策,2019,34(4):673-688.

[10] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.

[11] 张曦,杨颖,陈超君,等.增强双流 Transformer 的柴油发动机剩余寿命预测模型[J].汽车工程,2025,47(2):292-300.

【作者简介】

侯耀晗(2001—),男,河南新乡人,硕士研究生,研究方向:机器学习、图像处理等。

(收稿日期:2025-05-08 修回日期:2025-09-15)