

小样本条件下的储层物性参数智能解释方法研究

邬德刚^{1,2,3}, 吴胜和^{2,3*}, 张玉飞⁴, 余季陶^{2,3}

1 中国石油大学(北京)人工智能学院, 北京 102249

2 中国石油大学(北京)地球科学学院, 北京 102249

3 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

4 中海石油(中国)有限公司海南分公司, 海口 570312

* 通信作者, reser@cup.edu.cn

收稿日期: 2024-10-29; 修回日期: 2025-02-26

中国石油天然气集团有限公司—中国石油大学(北京)战略合作科技专题(ZLZX2020-02)资助

摘要 储层物性参数是表征储层储集与渗滤流体能力的重要参数, 测井解释是获取物性参数的重要途径, 是一个复杂非线性回归任务。针对已有物性参数测井解释方法在小样本学习条件下的泛化能力不足这一问题, 本文首先提出了基于聚类分析的样本优选方法, 通过 K 均值聚类划分样本的空间结构, 根据样本在空间结构中的分布优选学习样本, 从而最大化学习样本的多样性; 然后提出了基于层次化残差神经网络的物性参数测井解释方法。方法在全连接神经网络基础上引入 4 种机制: (1) 引入残差连接学习输入与输出间的残差映射, 深化小样本的复杂物性特征提取过程; (2) 引入集成学习整合多种不同机器学习方法, 通过算法多样性降低过拟合风险; (3) 引入多任务学习联系起孔隙度解释和渗透率解释这两个任务, 以提高小数据情况下单一任务的泛化性; (4) 引入二次加权均方根误差损失函数, 降低高渗储层的物性解释误差。在实际研究区中设计的 90 组对照实验的分析结果表明, 基于聚类分析的样本优选方法能够有效提升多种机器学习模型在小样本条件下的泛化能力; 基于本文提出的层次化残差神经网络进行研究区孔隙度与渗透率测井解释, 解释结果的决定系数分别达 88%、94%。与已有的多种方法相比, 本文方法基于分布特征的样本选择及多任务协同等方式的算法优化有效提高了岩石物理数据的特征表征, 方法的物性解释精度更高、泛化能力更强, 在取心盲井上的精度分别领先 12 和 20 个百分点。

关键词 聚类分析; 残差连接; 集成学习; 多任务学习; 储层物性参数测井解释

中图分类号: P618.13; TE311

Research on intelligent interpretation methods for reservoir physical parameters under few-shot conditions

WU Degang^{1,2,3}, WU Shenghe^{2,3}, ZHANG Yufei⁴, YU Jitao^{2,3}

1 College of Artificial Intelligence, China University of Petroleum, Beijing 102249, China

2 College of Geosciences, China University of Petroleum, Beijing 102249, China

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

4 Hainan Branch of CNOOC (China) Co. Ltd., Haikou 570312, China

引用格式: 邬德刚, 吴胜和, 张玉飞, 余季陶. 小样本条件下的储层物性参数智能解释方法研究[J]. 石油科学通报, 2025, 10(02): 378–391
WU Degang, WU Shenghe, ZHANG Yufei, YU Jitao. Research on intelligent interpretation methods for reservoir physical parameters under few-shot conditions[J]. Petroleum Science Bulletin, 2025, 10(02): 378–391. DOI: 10.3969/j.issn.2096-1693.2025.01.008

Received: 2024-10-29; Revised: 2025-02-26

Abstract Reservoir physical parameters serve as fundamental quantitative indices for characterizing the storage capacity and fluid percolation potential of subsurface reservoirs. Well logging interpretation, a critical methodology for accurately estimating these parameters, constitutes a sophisticated nonlinear regression challenge. To address the inherent limitations of existing petrophysical parameter interpretation techniques, particularly their inadequate generalization performance under few-shot learning conditions, this investigation systematically devises a dual-framework analytical approach. This study initially proposes a sample optimization methodology based on cluster analysis. The spatial configuration of samples is partitioned through the implementation of the *K*-means clustering algorithm, followed by selective sample curation according to spatial distribution characteristics to maximize learning sample diversity. Building upon this optimized sample architecture, the study further introduces a hierarchical residual neural network-based interpretation framework for petrophysical parameter estimation. The proposed methodology enhances conventional fully connected neural architecture through four innovative mechanisms: (1) Integration of cross-layer residual connections facilitates progressive refinement of residual mappings between multivariate logging inputs and target petrophysical outputs, thereby enabling hierarchical abstraction of complex petrophysical relationships from limited training instances. (2) The integration of ensemble learning paradigms amalgamates diverse machine learning methodologies, effectively mitigating overfitting risks through algorithmic diversity. (3) The implementation of a multi-task learning framework establishes intrinsic correlations between porosity and permeability interpretation tasks via shared latent representations, thereby enhancing individual task generalizability under data scarcity constraints. (4) The introduction of a quadratically weighted root mean square error loss function preferentially reduces interpretation errors in high-permeability reservoir intervals. Results from 90 rigorously designed comparative experimental configurations in the study area demonstrate that the cluster-based sample optimization methodology effectively enhances generalization performance across multiple machine learning models under few-shot learning constraints. Application of the proposed hierarchical residual neural network framework for well-logging interpretation of reservoir porosity and permeability within the investigated reservoir area achieves coefficients of determination of 88% and 94%, respectively, demonstrating statistically significant superiority over conventional methodologies in both petrophysical interpretation accuracy and generalization capability. Blind testing validation on cored wells reveals 12 and 20 percentage point improvements in predictive precision compared to other various existing methodologies, the proposed approach in this study demonstrates substantial advancements in addressing few-shot learning challenges through algorithm optimization strategies encompassing distribution-based sample selection and multi-task collaborative frameworks. This methodology significantly enhances feature representation fidelity in petrophysical datasets, exhibiting superior petrophysical interpretation accuracy and enhanced generalization capabilities.

Keywords cluster analysis; residual connection; ensemble learning; multi task learning; logging interpretation of reservoir physical parameters

DOI: 10.3969/j.issn.2096-1693.2025.01.008

0 引言

在油气地质领域, 储层物性特指储层的孔隙性和渗透性, 其中孔隙性决定了岩石的储集性能, 渗透性则反映了储层在一定压差下允许油、气、水通过的性能。孔隙度和渗透率分别是衡量孔隙性和渗透性大小的参数, 准确获取这两个物性参数对于油气勘探开发具有重要意义。

虽然储层物性参数可以通过油层物理学方法在油层取心段上直接测定^[1-2], 但鉴于取心成本高昂、测定流程复杂等因素, 应用地球物理测井技术解释物性参数逐渐成为国内外学者的研究热点。

储层的物性参数测井解释与其岩石物理特性密切相关, 自上世纪中期, 学者们依据电、声、核辐射等测井响应机理, 构建了岩石物理线性模型^[3-4]、经验统

计公式^[5-6]等方法定量表征物性参数。然而, 现今储层经历了复杂地质过程演变, 其物性受沉积组构、沉积组分、构型界面等多方面因素影响, 物性参数和岩石物理性质间可能呈现复杂的非线性关系。

为应对这一挑战, 研究学者引入了机器学习领域内的各种智能算法, 涉及多元线性回归、支持向量机、人工神经网络、决策树、贝叶斯算法、*K*近邻算法、高斯混合模型等回归分析方法^[7-13]以及粒子群优化^[14]、自适应差分进化^[15]等超参数优化方法, 岩性特征控制^[16-17]等条件约束方法。这些方法具备成熟的数学理论基础, 在运算效率和结果可解释性方面展现出显著优势。更进一步地, 为充分利用不同算法的优势、弱化某单一算法的偏见性预测, 有学者应用集成学习思想, 通过同质或异质集成等方式提高物性参数预测结果的精度。值得注意的是, 这些方法注重挖掘

单一学习样本的内在特征, 即从每一个样本中寻找影响因素与物性参数的映射关系, 故而方法性能易受噪音、异常值、测井数据的井间差异变化等多方面因素制约。

随着深度学习技术的发展, 逐渐衍生出了挖掘不同样本间关联特征的方法。考虑到测井数据中每一深度处测井值系一定范围内岩石物理性质的平均反映, 有研究学者^[18-19]基于卷积神经网络(CNN)架构中的局部感受野捕捉数据的空间局部特征, 从测井响应及物性特征变化的空间关联性方面提升算法的参数解释性能; 此外, 考虑到地层随时间逐步沉积形成, 不同深度储层的物性特征应当具有一定的时间依赖关系, 有学者^[20-22]基于循环神经网络(RNN)架构中的记忆单元关联不同时刻数据的内部状态特征, 从不同深度储层特征的时间关联性方面提升物性参数解释性能。深度学习领域中的CNN、RNN等方法虽然能够深入挖掘样本间的时空关联性, 刻画样本中隐藏的更加复杂的非线性规律, 但是方法要求输入的学习样本在时间域或空间域上是连续的, 因此难以应用于取心较离散的情况。

近几年, 有学者将关联性聚焦于不同的回归问题上, 利用不同参数预测任务间的相似性或关联性^[23-24]搭建基于迁移学习、多任务学习等思想的深度学习模型, 在降低单一参数预测误差、减缓少标签样本过拟合、提高模型在未取心新工区上的鲁棒性等方面具有重要作用。

尽管国内外学者针对物性参数解释方法开展的大量研究取得了以上众多显著成果, 但部分研究仍然相对薄弱。一方面, 由于取心井取心段数量有限, 在只有少量学习样本(小样本)情况下, 已有研究方法的泛化能力不足, 智能算法在训练集和测试集上的性能差异较大; 另一方面, 鉴于渗透率和孔隙度之间呈现近似的指数关系, 样本间的量级差异大, 虽然常用的对数变换等方法能够降低算法对数据变化的敏感性、提高梯度迭代的稳定性, 但会导致高渗样本的渗透率预测误差较大, 其误差甚至能够达到一个量级。

为解决以上问题, 本文提出了基于聚类分析的样本优选方法和一种基于层次化残差神经网络(Hierarchical residual neural network, HRNN)的储层物性参数测井解释方法, 基于聚类分析技术、集成学习思想、残差连接结构、多任务学习策略、二次加权均方根误差损失实现物性参数的测井解释。这一研究不仅有助于提升小样本条件下储层物性参数解释的准确性, 也为油气勘探开发提供了更为可靠的技术支持。

1 方法与原理

物性参数智能解释属于有监督学习范畴, 基本思路是以取心井岩心测定结果为标签, 基于有监督学习算法建立测井曲线与物性参数之间的复杂非线性映射关系。

小样本学习是物性参数解释等有监督学习过程中遇到的常见问题, 即各类别的标记样本数量较少, 已有样本难以代表总体数据分布特征。度量学习是解决小样本学习问题的有效策略之一, 通过距离函数来捕捉样本之间的相似性, 从而提高模型的迁移能力。参考度量学习思想, 本文提出首先使用聚类分析技术度量不同样本间的相似性, 然后依据样本间的相似性差异优选出具有多样性特征的样本集, 基于样本集设计并训练层次化残差神经网络模型, 从而实现小样本条件下物性参数的测井智能解释。

1.1 基于聚类分析的样本优选方法

物性参数智能解释的核心是建立机器学习模型后, 引导模型学习训练集样本的判别性特征, 并依据模型在验证集上的表现调整参数, 以防止模型过拟合或欠拟合。保持训练集和验证集数据分布的一致性和多样性是实现该过程的关键。

本文提出了使用聚类分析技术优选样本的方法, 即在小样本条件下, 基于聚类分析技术挖掘的数据空间结构特征构建训练集与验证集, 从而最大化数据集的样本多样性, 提高机器学习模型的泛化能力。

聚类分析是一种最基本的智能模式, 通过提取数据分布的潜在结构将对象或模式组织成簇, 以使同一簇内的对象相似, 而不同簇间的对象迥异。 K 均值是聚类分析中最流行的算法之一, 其思想最早由Steinhaus^[25]于1957年提出, 后在1967年被MacQueen^[26]首次命名, 目前广泛应用于图像分割^[27]、文本分析^[28]、能源预测^[29]、信号处理^[30]等领域。考虑到 K 均值算法具有较好的通用性、可解释性等优点, 本文将基于该算法实现样本优选。

K 均值算法是一种无监督学习、基于划分的聚类方法。对于多维空间中由 n 个样本组成的数据集 $X = \{x_1, x_2, \dots, x_n\}$, 算法首先初始化 K 个聚类中心 $\{c_1, c_2, \dots, c_K\}$ (如图1a所示), 并使用各样本 $x_i (i = 1, 2, \dots, n)$ 与各聚类中心 $c_k (k = 1, 2, \dots, K)$ 之间的欧几里得距离 $\|x_i - c_k\| (L_2 \text{ 范数})$ 衡量数据对象间的相似度。然后构造

簇标签标记函数 $\pi(x_i)$ 将样本 x_i 分配到距离最近(相似度最大)的聚类中心所在的簇中,亦即有第 k 个簇中的样本集合表示为 $C_k = \{x_i | \pi(x_i) = k, i = 1, 2, \dots, n\}$ 。随后依据簇中样本的均值更新聚类中心

$$c_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i \quad (1)$$

式中, $|C_k|$ 为集合 C_k 元素总数。

K 均值聚类算法使用距离平方和(Sum of Squared Error, SSE)作为度量聚类质量的目标函数,通过不断重复上述聚类中心及簇集合迭代更新过程(如图 1b 所示)的方式最小化 SSE,直到各聚类中心不再变化或 SSE 收敛时聚类算法达到最优的样本划分(如图 1c 所示)。

$$SSE = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - c_k\|^2 \quad (2)$$

1.2 层次化残差神经网络算法

本文在全连接人工神经网络的基础上,综合应用残差学习、集成学习、多任务学习思想设计了层次化残差神经网络算法,实现孔隙度及渗透率的测井解释。

1.2.1 全连接人工神经网络算法

全连接人工神经网络是由大量简单的处理单元(神经元)彼此按某种方式互连而成的复杂网络系统,从仿生学角度模拟人脑神经系统的运作方式,使机器具有和人脑一样的感知、学习和推理功能。

全连接神经网络以批量化的样本特征 x_1, x_2, \dots, x_m 为输入信息,通过信息的前馈传播解决分类或回归等问题,其基本架构如图 2 所示。神经元是神经网络的基本信息处理单位,由 3 种基本元素构成:突触、累加器及激活函数。其中突触是信号传输的节点,各突触配置一个权重,权值大小代表了信号的重要性,权值的正负分别表征了突触的兴奋与抑制作用;累加器执行神经元上各信号的加权求和操作,实现不同特征的线性组合,并通过偏置对加权信号作仿射变换,调整输出的分布范围;激活函数(Sigmoid、Tanh、ReLU 等)通过非线性变换改变神经元输出,解决线性不可分等问题。神经元的数学内涵可以通过方程阐明

$$y = f\left(\sum_{i=1}^n w_i z_i + b_L\right) \quad (3)$$

式中, f 为激活函数, z_i 为输入信号, w_i 为 z_i 的权重, b_L 为偏置, L 为网络层数的编号, y 为输出信号。

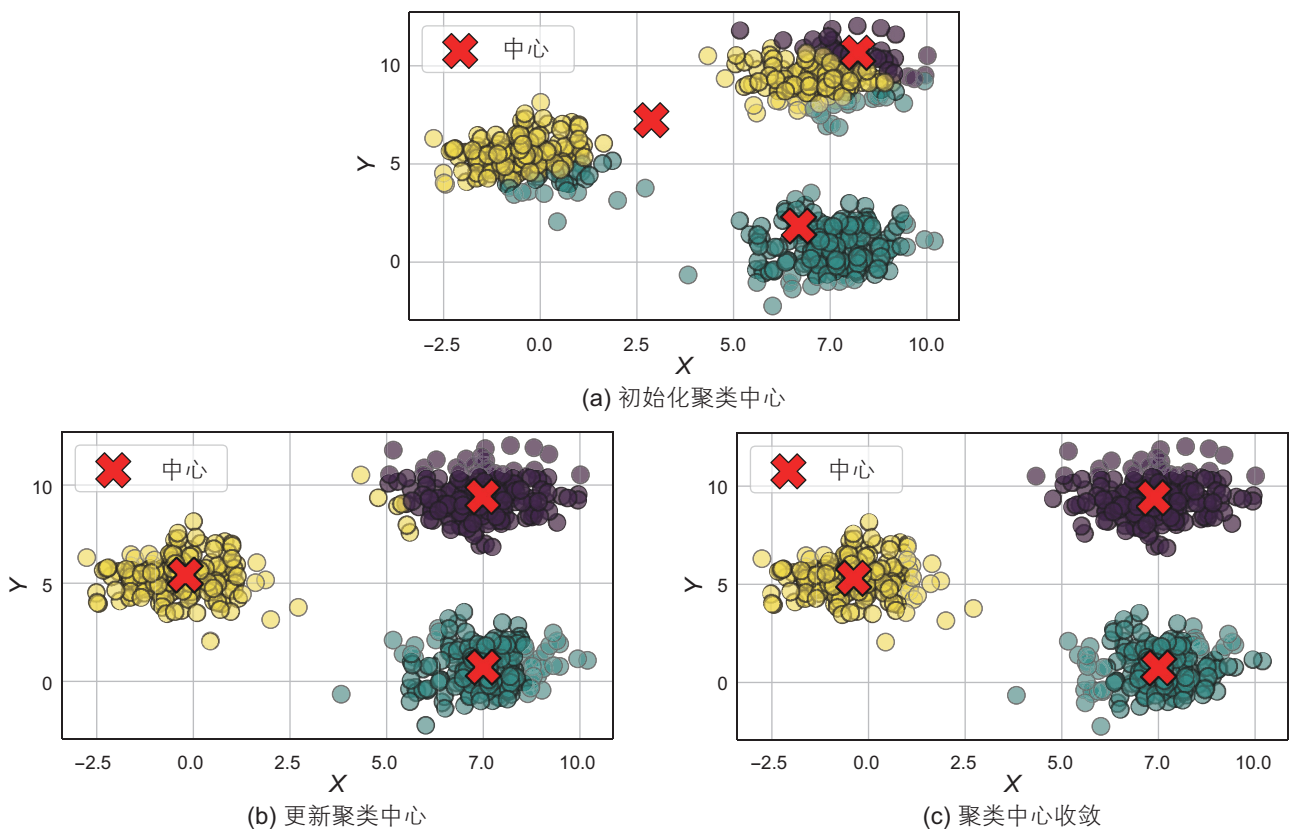


图 1 二维空间的 K 均值聚类算法示意图(不同颜色点表示不同类别样本)

Fig. 1 Schematic of K -mean clustering algorithm in 2D space

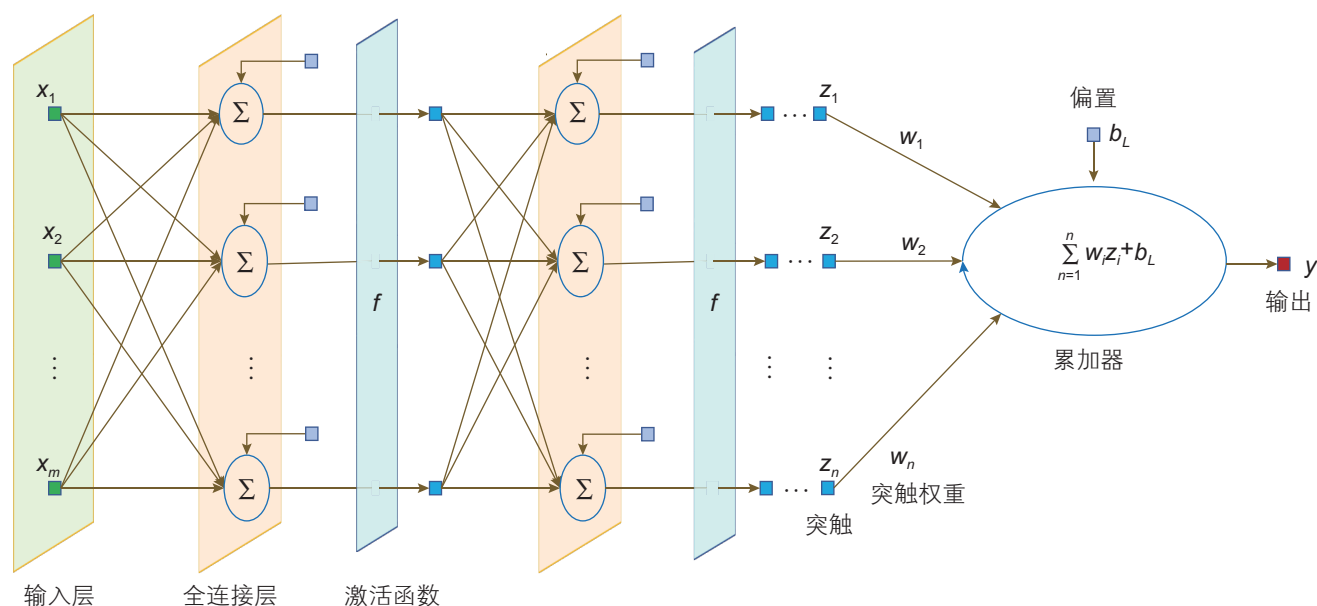


图 2 全连接神经网络的基本架构

Fig. 2 Architecture of fully connected neural network

1.2.2 HRNN 算法架构

HRNN是在人工神经网络的基础上,引入多种深度学习技术形成的储层物性参数测井解释算法。即引入残差连接结构,加速模型的收敛速度;基于集成学习思想,利用多个异质学习器提高算法鲁棒性;使用多任务学习策略先后实现孔隙度和渗透率解释,借助共享特征提高解释精度;应用二次加权均方根误差损失指导算法梯度反向传播过程,降低高渗储层的解释误差。

HRNN算法由3部分构成,分别为残差学习网络、集成学习 α 网络以及集成学习 β 网络。

其中,残差学习网络通过残差连接技术实现测井数据的深层复杂物性特征提取,加速了算法的收敛速度。对于输入的多维特征 x ,期望构造多层神经网络结构实现一个完整映射 $y = H(x)$,为了解决在求解映射 $H(x)$ 过程中深度学习算法存在的梯度消失、学习退化等问题,微软研究院^[31]于2016年提出了深层网络的残差连接方法。残差是指预测值 x 和观测值 y 之间的差异,残差学习的思想是将求解 $H(x)$ 转化为求解残差 $f(x) = y - x$,即令算法学习输入与输出之间的非线性变化。具体而言,残差连接是一种如图3中橙色弧线所示的跳跃连接方法。四层全连接神经网络负责将输入特征 x 映射为 $f(x)$,在跳跃连接作用下,残差学习网络的输出被转变为 $f(x) + x$,亦即 $H(x)$ 。因此残差连接结构通过先求解 $f(x)$,再求解 $H(x)$ 的方式实现了 x 到 y

的深层复杂映射。

α 网络和 β 网络通过引入集成学习和多任务学习技术提高小样本条件下的物性参数解释性能和泛化能力,其具有以下3方面特征:

其一, α 网络和 β 网络分别是实现孔隙度和渗透率解释的集成学习算法,两者结构相似。以图3中 α 网络为例,该网络由物性参数解释及集成决策两个模块组成。物性参数解释模块以残差学习网络提取的复杂物性特征为输入,以孔隙度解释结果为输出,是由输入层、两层隐藏层、单神经元输出层共同构成的全连接神经网络。其中针对隐藏层,为防止数据分布的显著变化导致神经网络梯度迭代等过程不稳定,在全连接层之后、激活函数之前引入一维批量归一化层,通过归一化处理控制数据分布;集成决策模块是由输入层、一层隐藏层、单神经元输出层构成的全连接神经网络。集成是指整合不同学习器输出结果的一种决策方法,考虑到基于不同原理构建的学习器(异质学习器)都是从某一特定角度对数据进行学习,因此可以将多个异质学习器结合起来,充分利用各自的优点,从而得到更好的学习效果。故而,集成模块通过引入多个异质学习器增加学习的多样性,采用全连接神经网络整合各学习器的物性解释成果,从而提高小样本条件下机器学习算法的性能。

其二,使用多任务学习的层次化网络结构组织 α 网络和 β 网络。多任务学习的核心思想是利用关联任务之间的共享信息同时学习多个任务,从而提高单一

任务的学习效果。考虑到岩石的渗透率和孔隙度之间有一定的内在联系，一方面两者都受某些相同的影响因素制约，如当岩石中含有较多的泥质杂基时，孔渗性都会有明显降低；另一方面，孔隙度是渗透率的影响因素之一，两者具有近似的经验统计关系：

$$\lg \kappa = a\varphi + b \tag{4}$$

式中， κ 为渗透率， φ 为孔隙度， a 、 b 为常量。因此，可以构建多任务学习过程，通过共享参数和特征表示捕捉孔隙度与渗透率之间的相关性，从而优化算法整

体性能，特别是在小样本条件下的性能。

基于多任务学习， α 网络和 β 网络的组织结构如图4所示。 β 网络同样由物性参数解释模块及集成决策模块这两部分组成，与 α 网络不同的是， β 网络同时以残差学习网络及 α 网络的输出为输入。

其三， α 网络、 β 网络分别采用均方根误差损失、二次加权均方根误差损失指导算法梯度反向传播过程。 α 网络的均方根误差损失函数 l_α 定义为：

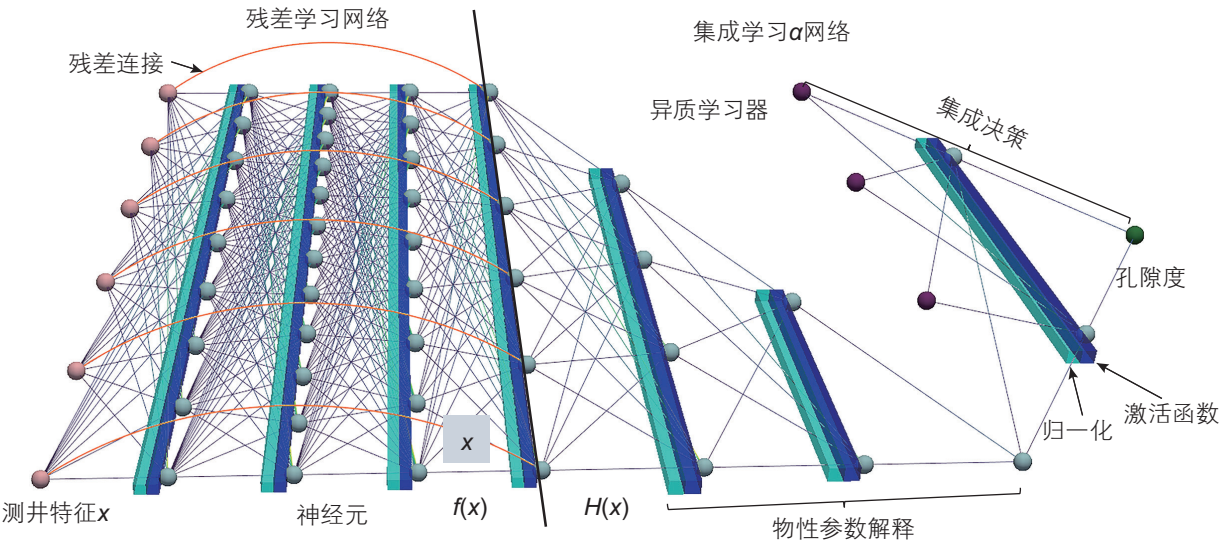


图3 HRNN模型中的残差学习网络及 α 网络
Fig. 3 Residual learning network in HRNN model and α networks

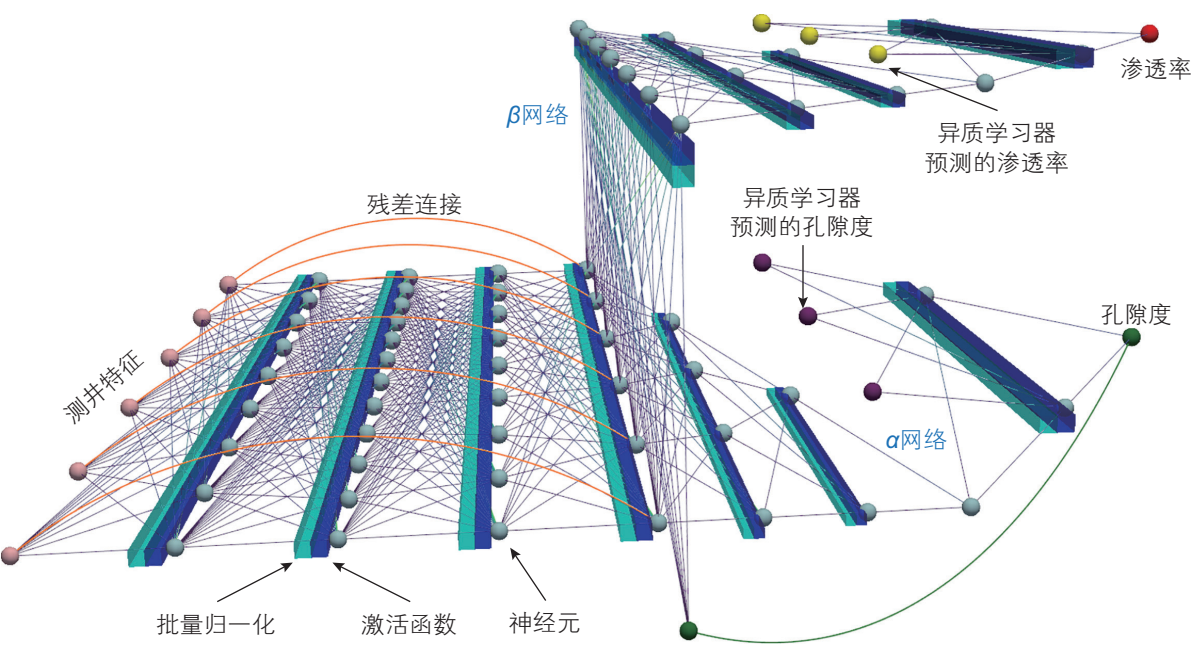


图4 HRNN模型的整体结构
Fig. 4 Overall structure of the HRNN model

$$l_{\phi} = \sqrt{\frac{\sum_{i=1}^m (\hat{\phi}_i - \phi_i)^2}{m}} \quad (5)$$

式中, $\hat{\phi}_i$ 为孔隙度解释值, ϕ_i 为孔隙度岩心测定值。

与 α 网络相比, β 网络通过渗透率对数变换方式提高输出的数值稳定性, 加速算法收敛过程。因此, 对于 β 网络的输出 $\hat{\kappa}_i$ 、对数变换后的渗透率岩心测定值 κ_i , 其真实误差 $e^{\hat{\kappa}_i} - e^{\kappa_i}$ 随 κ_i 的增加呈指数增长。为降低网络对高渗样本的解释误差, 定义加权均方根误差损失函数 l_k 为:

$$l_k = \sqrt{\frac{\sum_{i=1}^m \kappa_i^2 (\hat{\kappa}_i - \kappa_i)^2}{\sum_{i=1}^m \kappa_i^2}} \quad (6)$$

式中, 对于每一个输出误差 $\hat{\kappa}_i - \kappa_i$, 设置其权重为二次权值 κ_i^2 , 从而抑制误差的指数增长趋势。

2 应用分析

应用本文方法解决某油田的储层参数测井解释问题, 并基于解释结果论证方法的优越性。

2.1 研究区概况

研究区为英国北海地区马里福斯盆地海底扇沉积, 目的层段为上侏罗统基默里奇阶-伏尔加阶的沉积地层。粒度和胶结作用是目的层段储层物性的主控因素。研究区砂岩泥质含量总体较低, 依据粒径中值可以将砂岩划分为极粗砂岩、粗砂岩、中砂岩、细砂岩、极细砂岩和粉砂岩 6 类, 各类砂岩数量分布不平衡, 以细砂岩为主, 部分中砂岩、极细砂岩, 少量粉砂岩、粗砂岩、极粗砂岩。研究区整体呈现中高孔-特高渗的物性特征且物性受不同沉积组构控制, 孔渗整体相关性较弱, 相关系数 R^2 约为 0.45。其中, 孔隙度大体分布在 10%~36% 之间, 主要受分选影响; 渗透率集中分布于 100~5000 mD 之间, 主要受泥质含量和粒径影响。储层物性与粒度整体呈正相关关系, 即粒度越粗, 砂岩的平均孔渗越高。同时, 物性还受到石英胶结和钙质胶结作用的影响, 其中石英胶结对储层物性的影响较弱, 而钙质胶结对储层物性的影响较强。在测井响应上, 孔渗的变化与测井响应间的关系较为复杂, 声波时差、密度等测井响应特征与储层物性间不具有同步的变化趋势(如图 5 所示), 难以反映出明显的解释规律, 物性参数解释难度大。

研究区共有 6 口可用的取心井数据, 其中包括 800 个经岩样分析化验获得的孔隙度和渗透率样本, 自然伽马、声波时差、密度、深中浅电阻率、中子等常规测井曲线, 以及高精度的测井岩性解释结果。研究区非取心井数量达百余口, 且在研究区块中较为分散, 取心井上的测井样本难以代表整个研究区块的数据分布特征, 对非取心井的物性参数解释可视为小样本问题。

2.2 物性参数智能测井解释

本文实验使用的数据集构建方法如下: 从研究区的 6 口取心井中随机选取 4 口井构建样本比例为 1:1 的训练数据集及验证数据集, 分别用于机器学习模型训练及优化; 剩余两口井的数据构建测试集, 用于评估最终模型的鲁棒性。

本文实验的基本步骤是: 首先通过敏感性分析优选测井曲线, 然后基于聚类分析方法优选样本, 论证该方法对于提高小样本条件下的物性参数测井解释精度的有效性; 最后构建并训练 HRNN 网络模型, 实现测试集样本的物性参数测井解释, 阐明 HRNN 网络模型在解决小样本条件下的物性参数测井解释问题上的优越性。

2.2.1 测井曲线优选

优选用于解释储层物性的测井曲线是影响物性解释精度的关键步骤。优选测井曲线时需要综合考虑各测井曲线与物性参数间的数据关联性和地质关联性。本文利用 Morris^[32] 方法对研究区的自然伽马等测井曲线和孔渗参数进行全局敏感性分析。Morris 方法是通过一次只改变一个参数的取值方式轮流计算各测井曲线对孔渗的基本效应值, 应用效应值的均值和标准差作为各测井曲线对孔渗的敏感性指数。其中, 均值越大则表明测井曲线对孔渗的影响越大, 亦即该测井曲线的敏感性越强; 标准差则反应了测井曲线之间交互作用的强弱, 其值越高, 说明该测井曲线与其他测井曲线的相互作用越大, 亦同时表明该测井曲线对孔渗的影响是非线性的。据此, 对研究区的 12 种测井数据进行孔渗敏感性分析, 表 1 中所示的敏感性分析结果表明, 岩性、密度、深感应等特征的效应值标准差相对更高, 声波时差、中子、自然伽马等特征的效应值均值相对更高, 结合各测井数据与物性参数在地质上的关联性, 本文选择岩性、自然伽马、声波时差、中子、密度、深感应这 6 种测井特征进行研究区块的物性参数解释。

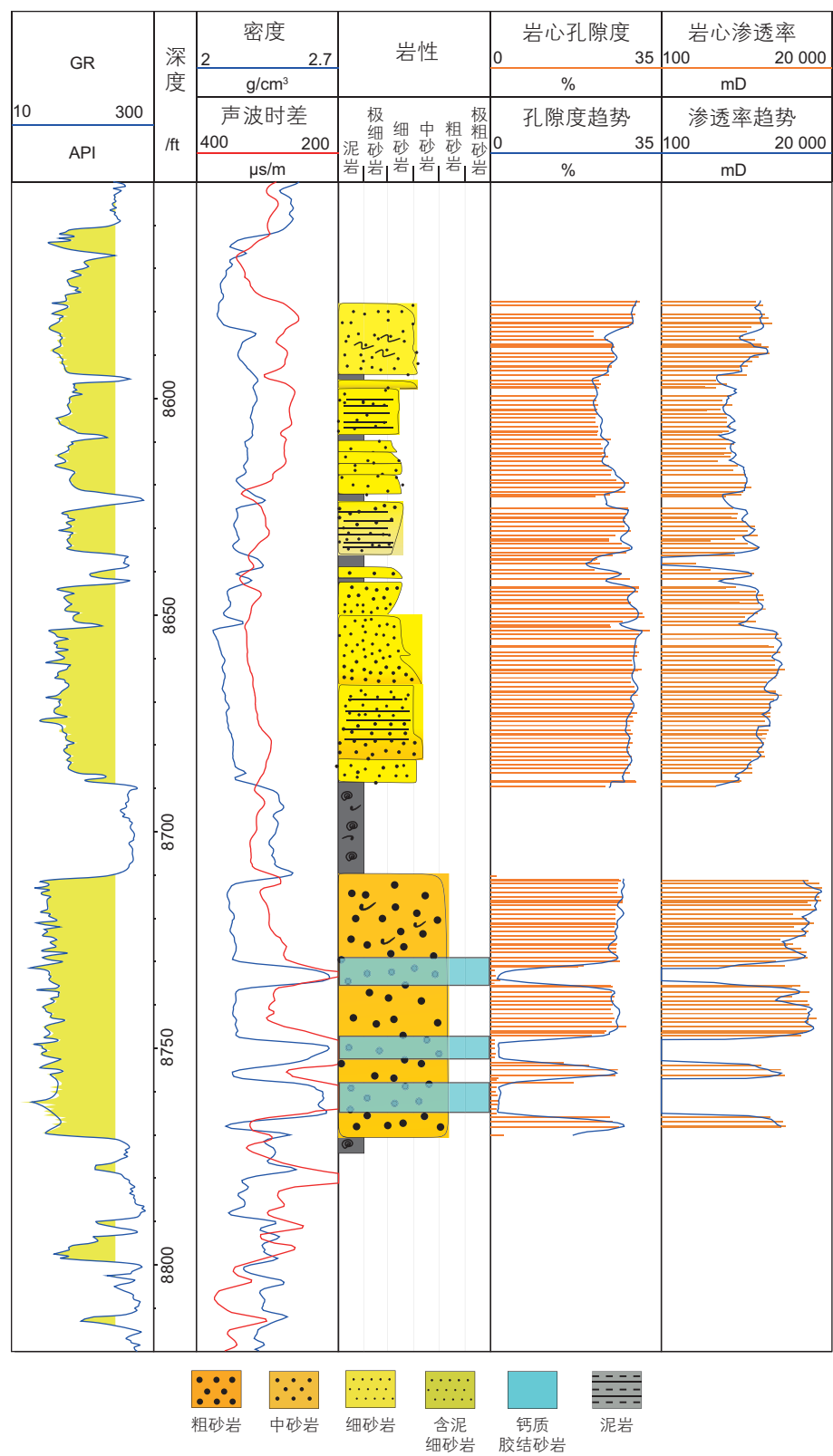


图5 研究区某取心井的测井响应及物性特征

Fig. 5 Logging response and physical characteristics of a coring well in the study area

2.2.2 基于K均值聚类算法的样本优选

本部分实验的基本思路是以是否引入聚类分析方法优选样本为单一变量，使用前人已有研究方法进行

物性参数测井解释，通过对比分析优选样本前后测试集上的物性参数解释效果，论证基于聚类分析技术的样本优选方法的有效性，并进一步探讨样本量对实验

表 1 测井曲线的敏感性分析结果
Table 1 Physical sensitivity analysis of logging curves

测井曲线	孔隙度敏感性指数		渗透率敏感性指数	
	均值	标准差	均值	标准差
光电吸收截面指数	0.007	0.286	-0.243	9.948
钾	0.091	0.217	-1.280	11.303
钍	0.035	0.262	0.227	9.794
铀	0.025	0.267	-0.661	9.962
深感应	0.070	0.318	1.050	10.720
中感应	0.050	0.311	0.761	10.369
浅感应	0.050	0.302	0.830	10.248
声波时差	0.183	0.300	0.916	10.480
中子	0.211	0.298	0.400	11.947
密度	-0.020	0.306	-0.800	11.136
自然伽马	0.072	0.290	1.070	11.072
岩性	0.248	0.467	8.816	17.494

的影响作用。

为增加实验的可信度，从前人研究成果中优选了 5 种不同方法进行多次实验，分别是 K 邻近、多层感知机、决策树算法，基于自助聚合思想的随机森林集成算法以及基于提升思想的梯度提升树集成算法。这些回归学习器在测试集上的性能 R_w^2 使用孔隙度、渗透率的加权决定系数表征：

$$R_w^2 = w_k R_k^2 + w_\phi R_\phi^2 \tag{7}$$

式中， w_k 、 w_ϕ 分别为孔隙度和渗透率的加权系数。考虑到本文研究区块物性参数在测井解释精度上的要求，认为提高孔隙度的解释效果和渗透率的解释效果同等重要，故分别取值 $w_k = 0.5$ 、 $w_\phi = 0.5$ 。 R_k^2 、 R_ϕ^2 分别为孔隙度和渗透率的决定系数，其值越大说明残差越小，即模型输出结果越好。其定义为：

$$R_\phi^2 = 1 - \frac{\sum_{i=1}^m (\phi_i - \hat{\phi}_i)^2}{\sum_{i=1}^m (\phi_i - \bar{\phi})^2} \tag{8}$$

$$R_k^2 = 1 - \frac{\sum_{i=1}^m (\kappa_i - \hat{\kappa}_i)^2}{\sum_{i=1}^m (\kappa_i - \bar{\kappa})^2} \tag{9}$$

式中， $\bar{\phi}$ 、 $\bar{\kappa}$ 分别为孔隙度和渗透率样本期望。

实验设置了 6 种不同的样本优选方法。其中，第一种方法为随机优选方法（已有常规方法），即采用随机方式选取训练集及验证集样本；后 5 种是鉴于不同原理选择的聚类分析方法，分别为层次聚类、谱聚类、混合高斯聚类、密度聚类、 K 均值聚类算法。考虑到

数据集样本量，将聚类类别数设置为 5。值得注意的是，为阐明基于聚类分析的样本优选方法对不同回归学习器算法的性能提升作用，本文中以随机优选方法中各种回归学习器的加权决定系数为基准，计算其它样本优选方法下回归学习器的相对加权决定系数 ΔR_w^2 指示模型的泛化能力变化量。

此外，为进一步探讨样本量对回归学习器泛化能力变化的影响，本文分别按验证集样本占比 θ 取值为 0.1、0.3、0.5 三种条件设计实验。

依据上述条件进行实验（共计 90 组），评估各组实验中回归学习器在测试集上的性能并计算 ΔR_w^2 ，其统计结果如图 6 所示。实验表明基于聚类分析的样本优选方法构建的训练数据集对提升回归学习模型在测试集上的鲁棒性具有积极作用（ $\Delta R_w^2 \geq 0$ ）；对比各种样本优选方法在实验中的综合表现情况， K 均值聚类算法优选的样本具有更丰富的多样性，对回归学习器的性能提升更显著。分别采用核主成分分析（Kernel Principal Components Analysis）、谱嵌入（Spectral Embedding）、等度量映射（Isometric Feature Mapping）方法将 K 均值算法聚类的样本特征映射至二维空间。图 7 的结果表明，聚类得到的各类样本之间具有清晰边界， K 均值聚类算法在抓取数据样本空间结构特征方面具有良好效果；对比分析 θ 从 0.1 到 0.5 条件下的实验结果，回归学习器的泛化能力变化量随着训练数据集样本量的减少呈现显著的上升趋势，表明基于聚类分析的样本优选方法在小样本条件下更能发挥优势，能够显著提升多种机器学习模型在小样本条件下的物性参数测井解释精度。

2.2.3 基于HRNN模型的物性参数测井解释

以K均值聚类算法优选的学习样本为HRNN模型的输入，以K近邻算法、决策树、梯度提升树算法解

释的孔隙度和渗透率分别作为 α 网络、 β 网络部分的输入，搭建HRNN模型并进行16轮次迭代训练。在模型训练过程中，应用自适应学习率算法(Adam)实

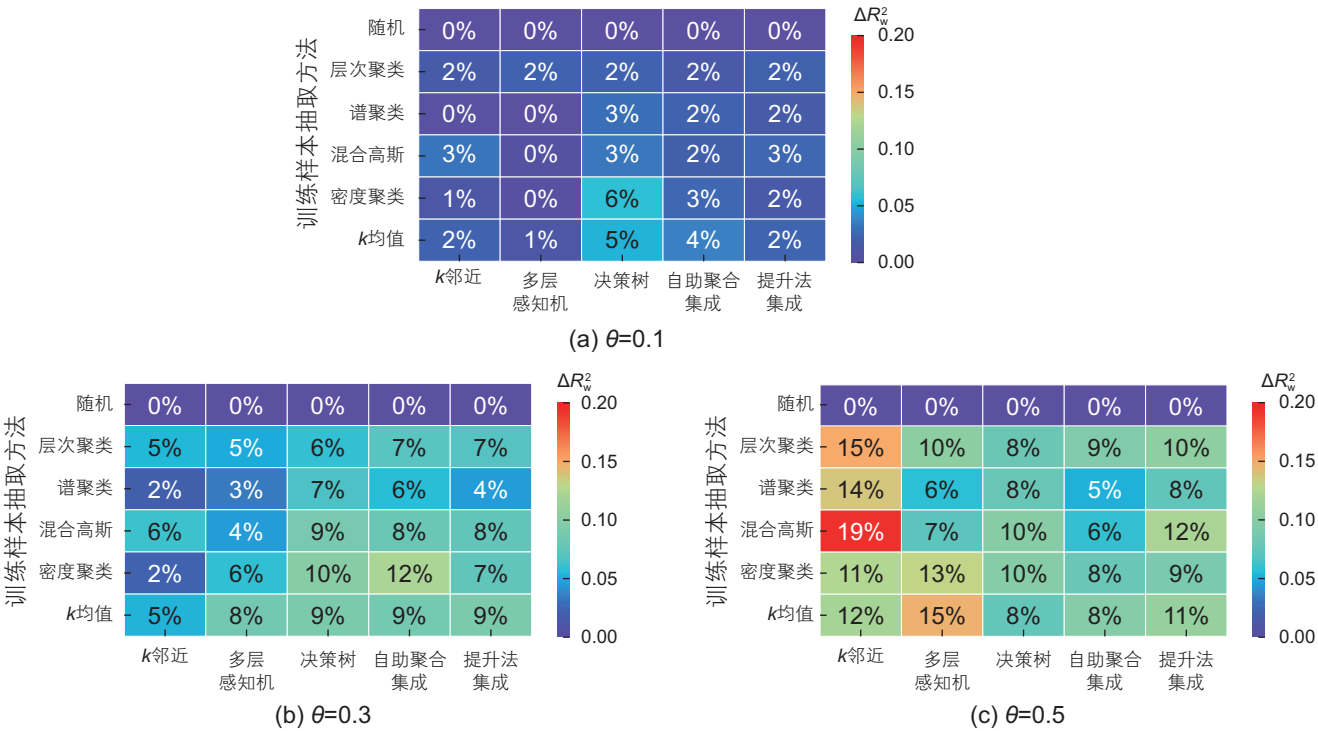


图 6 不同样本量条件下的相对加权决定系数

Fig. 6 Relative weighted coefficients of determination for different sample size conditions

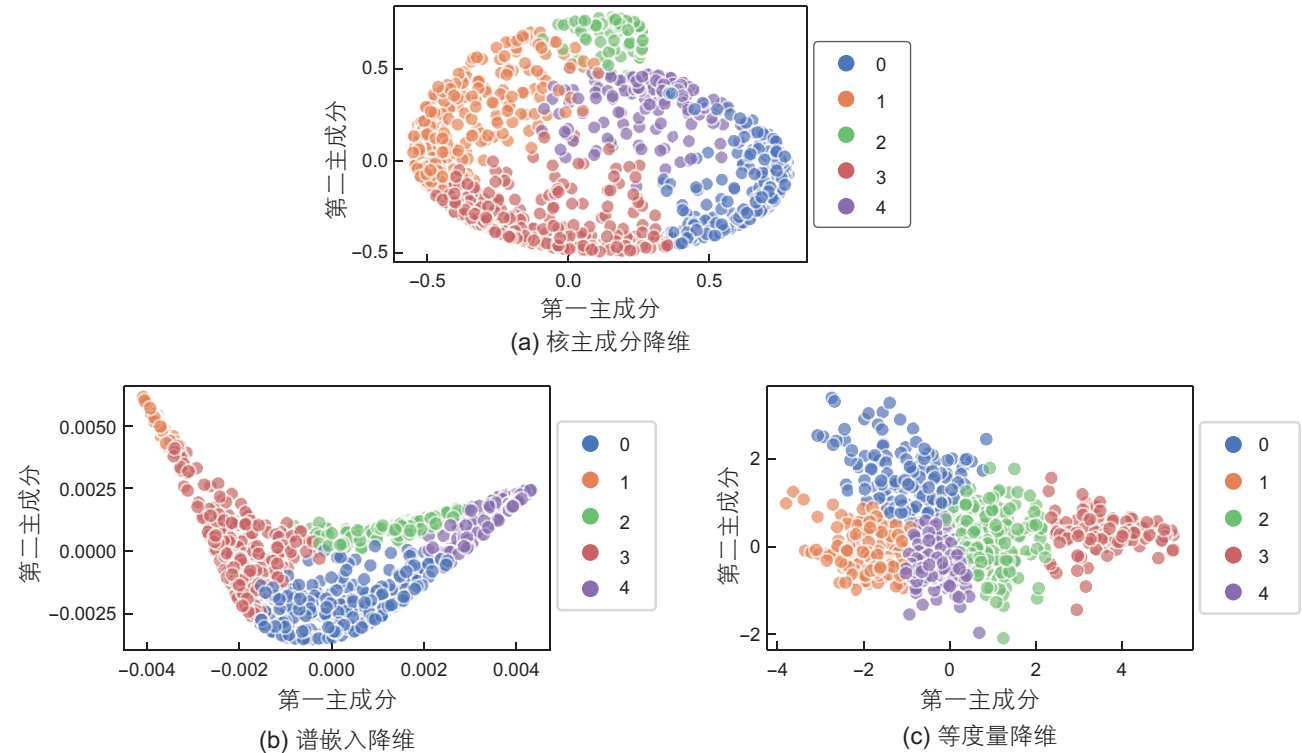


图 7 不同二维特征空间中的K均值聚类结果

Fig. 7 K-mean clustering results in different 2D feature spaces

现神经网络参数更新,采用等间隔的学习率调整策略(StepLR)设置模型每一轮梯度传播的学习率,使得模型在训练集及验证集上的损失函数迭代过程相对稳定,且具有明显的下降趋势和良好的收敛速度,如图 8 所示。

考虑到研究区取心数据较为离散,已有循环神经网络等方法难以良好应用于研究区的物性参数解释,因此从已有研究成果中选择了几种合适方法与本文提出的方法进行对照实验,分别为:多元线性回归(传

统方法)、决策树(单一学习器)、梯度提升树(集成学习方法)、全连接神经网络(深度学习方法)。各方法在验证集及测试集上的孔隙度与渗透率测井解释效果如图 9 所示。结果表明,虽然 HRNN 模型在降低孔隙度解释最大误差方面仍然存在不足,其对部分样本的孔隙度解释结果与岩心测定结果偏差较大(如图 10a 所示),但是大多数样本的解释精度较高,整体样本的平均解释精度高,与其它方法相比具有更低的孔隙度及渗透率解释误差,对验证集上样本的物性参数测井解

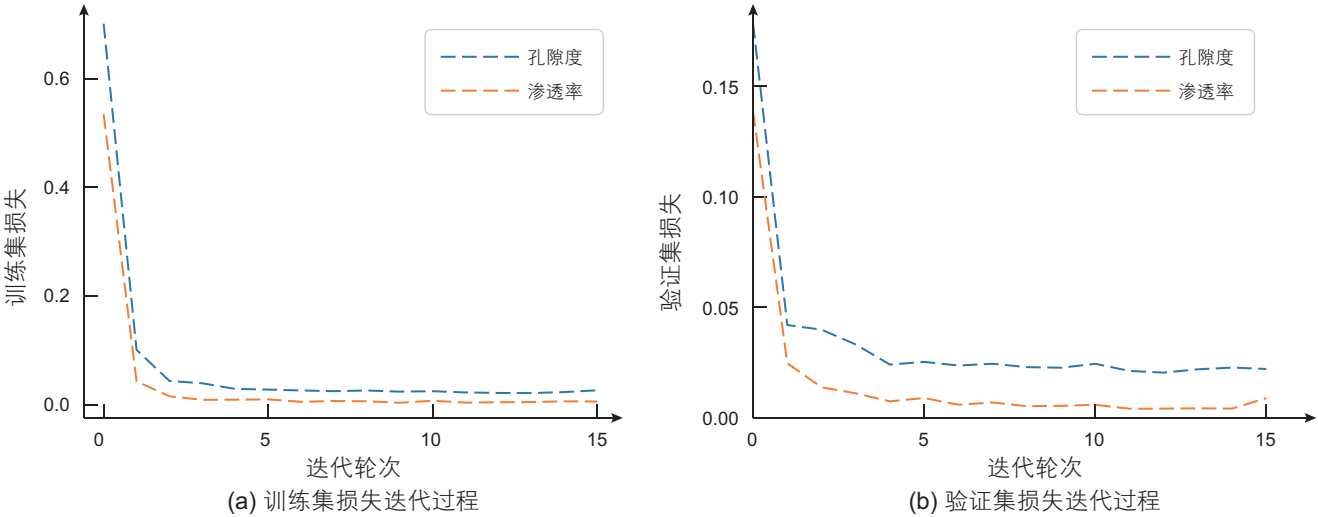


图 8 HRNN 模型训练过程中的损失迭代过程
Fig. 8 Loss iteration process during HRNN model training

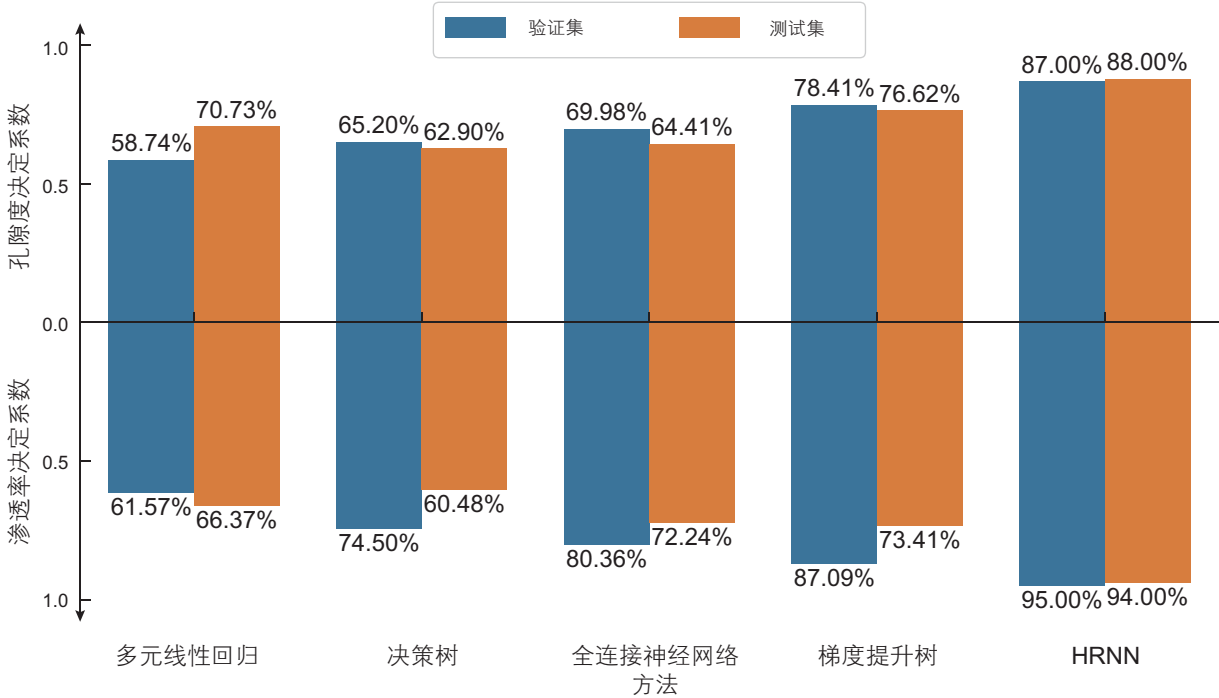


图 9 不同物性参数解释方法的性能对比
Fig. 9 Performance comparison of different physical property parameter interpretation methods

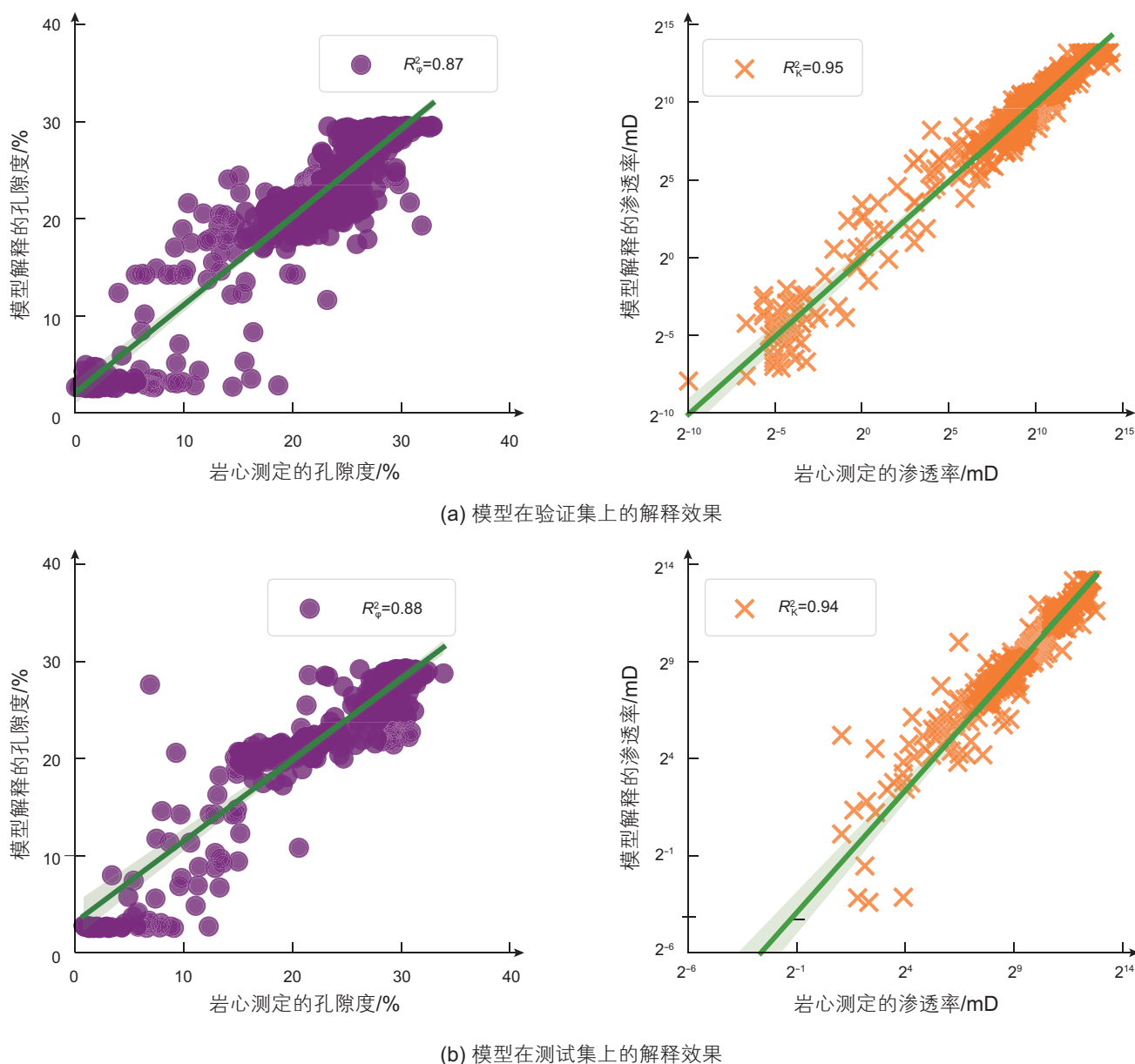


图 10 HRNN 模型在验证集、测试集上的性能表现

Fig. 10 Performance of HRNN model on validation set and test set

释性能分别达到 87%、95%；HRNN 模型具有良好的泛化能力，其在测试集上的性能表现最佳，性能分别达到 88%、94% (如图 10b 所示)；HRNN 模型和多元线性回归算法的鲁棒性相对更强，尤其是 HRNN 模型在测试集、验证集上的性能表现无显著区别。综上所述结果表明，本文提出的 HRNN 方法在解决小样本条件下的物性参数测井解释问题上更具优势。

3 结论

利用测井数据进行物性参数解释是一项极富挑战和重要意义的研究。针对已有方法在小样本条件下储

层物性参数解释泛化能力不足这一难题，本文开展了相关研究并取得了以下成果和认识：

1) 本文设计了基于 K 均值聚类分析技术的样本优选方法构建训练数据集，90 组对照实验结果表明聚类分析能够有效区分不同样本在高维空间中的特征差异，该样本优选方法对于提升多种机器学习模型在小样本条件下的物性参数解释精度及泛化能力具有显著的积极作用。

2) 本文设计了层次化残差神经网络模型 HRNN，该方法引入了残差连接、集成学习、多任务学习在内的多种经典机器学习结构来强化算法对小样本的学习能力，并采用二次加权均方根损失降低高渗样本的解

释误差。实际油田取心盲井样本测试实验结果表明,与已有的多元线性回归、梯度提升树等多种方法相比,HRNN模型能够有效提高小样本条件下孔隙度、渗透率的解释精度和泛化能力,决定系数指标分别领先 12 和 20 个百分点,分别达 88%、94%。

3)虽然本文所提出的方法取得了较好的测井物性解释效果,但是在聚类结果的可解释性、样本物性解释的最大误差等方面还存在一定的不足,还需要在日后的研究中不断进行探索改进。

参考文献

- [1] 韩学辉,张浩,毛新军,等.基于岩心室应力应变和不确定度分析的致密储层氮孔隙度测量方法[J].地球物理学报,2021,64(1): 289–297. [HAN X H, ZHANG H, MAO X J, et al. A method of gas porosity measurement for tight reservoirs based on mechanical analysis of core chamber[J]. Chinese Journal of Geophysics, 2021, 64(1): 289–297.]
- [2] WANG Y, ZHOU B. Recent progress in single and combined porosity-evaluation techniques for porous materials[J]. Materials, 2022, 15(9): 2981–3000.
- [3] LUO Y, HUANG H, JAKOBSEN M, et al. Prediction of porosity and gas saturation for deep-buried sandstone reservoirs from seismic data using an improved rock-physics model[J]. Acta Geophysica, 2019, 67: 557–575.
- [4] WANG P, CHEN X, LI J, et al. Accurate porosity prediction for tight sandstone reservoir: A case study from North China[J]. Geophysics, 2020, 85(2): B35–B47.
- [5] ELLIS D. Formation porosity estimation from density logs[J]. Petrophysics-The SPWLA Journal of Formation Evaluation and Reservoir Description, 2003, 44(05): 306–316.
- [6] 白倩,赵军龙,黄千玲,等. H油田侏罗系延 9 储层物性参数预测方法研究[J]. 录井工程, 2023, 34(3): 81–88. [BAI Q, ZHAO J L, HUANG Q L, et al. Study on prediction method of physical properties of Jurassic Yan 9 reservoir in H Oilfield[J]. Mud Logging Engineering, 2023, 34(3): 81–88.]
- [7] YASIN Q, SOHAIL G M, KHALID P, et al. Application of machine learning tool to predict the porosity of clastic depositional system, Indus Basin, Pakistan[J]. Journal of Petroleum Science and Engineering, 2021, 197: 107975.
- [8] WANG M, FENG D, LI D, et al. Reservoir parameter prediction based on the neural random forest model[J]. Frontiers in Earth Science, 2022, 10: 888933.
- [9] HUO F, CHEN Y, REN W, et al. Prediction of reservoir key parameters in ‘sweet spot’ on the basis of particle swarm optimization to TCN-LSTM network[J]. Journal of Petroleum Science and Engineering, 2022, 214: 110544.
- [10] TIAN J W, QI C, PENG K, et al. Improved permeability prediction of porous media by feature selection and machine learning methods comparison[J]. Journal of Computing in Civil Engineering, 2022, 36(2): 04021040.
- [11] NAZARI H, HAJIZADEH F. Estimation of permeability from a hydrocarbon reservoir located in southwestern Iran using well-logging data and a new intelligent combined method[J]. Carbonates and Evaporites, 2023, 38(1): 20–34.
- [12] FANG Z, BA J, CARCIONE J M, et al. Permeability prediction using logging data from tight reservoirs based on deep neural networks[J]. Journal of Applied Geophysics, 2024, 229: 105501.
- [13] XU P, ZHOU H, LIU X, et al. Permeability prediction using logging data in a heterogeneous carbonate reservoir: A new self-adaptive predictor[J]. Geoenergy Science and Engineering, 2023, 224: 211635.
- [14] LIU J J, LIU J C. Permeability predictions for tight sandstone reservoir using explainable machine learning and particle swarm optimization[J]. Geofluids, 2022, 2022(1): 2263329.
- [15] DAS R, PRASAD D K. Prediction of porosity and thermal diffusivity in a porous fin using differential evolution algorithm[J]. Swarm and Evolutionary Computation, 2015, 23: 27–39.
- [16] 魏国华,韩宏伟,刘浩杰.基于半监督高斯混合模型与梯度提升树的砂岩储层相控孔隙度预测[J].石油地球物理勘探,2023,58(1): 46–55. [WEI G H, HAN H W, LIU H J. Facies-controlled porosity prediction of sandstone reservoirs based on semi-supervised Gaussian mixture model and gradient boosting tree[J]. Oil Geophysical Prospecting, 2023, 58(1): 46–55.]
- [17] 鲁健康,郝彬,李程善,等.基于流动单元分类的致密砂岩储层渗透率预测[J].石油科学通报,2021,6(3): 369–379. [LU J K, HAO B, LI C S, et al. Permeability prediction of tight sandstone reservoirs based on flow unit classification[J]. Petroleum Science Bulletin, 2021, 6(3): 369–379.]
- [18] MASROOR M, NIRI M E, SHARIFINASAB M H. A multiple-input deep residual convolutional neural network for reservoir permeability prediction[J]. Geoenergy Science and Engineering, 2023, 222: 211420.
- [19] 孙正心,金衍,孟翰,等.基于深度学习数据融合的测井数据精细表征[J].石油科学通报,2025,10(1): 75–86. [SUN Z X, JIN Y, MENG H, et al. Fine characterization of logging data based on the deep learning data fusion[J]. Petroleum Science Bulletin, 2025, 10(1):

- 75–86.]
- [20] 宋先知, 姚学喆, 李根生, 等. 基于 LSTM-BP 神经网络的地层孔隙压力计算方法[J]. 石油科学通报, 2022, 7(1): 12–23. [SONG X Z, YAO X Z, LI G S, et al. A novel method to calculate formation pressure based on the LSTM-BP neural network[J]. Petroleum Science Bulletin, 2022, 7(1): 12–23.]
- [21] CHEN W, YANG L, ZHA B, et al. Deep learning reservoir porosity prediction based on multilayer long short-term memory network[J]. Geophysics, 2020, 85(4): WA213–WA225.
- [22] WANG J, CAO J. Deep learning reservoir porosity prediction using integrated neural network[J]. Arabian Journal for Science and Engineering, 2022, 47(9): 11313–11327.
- [23] 汪敏, 郭鑫平, 唐洪明, 等. 深度 Transformer 迁移学习的页岩气储层核心参数预测案例[J]. 地球物理学报, 2023, 66(6): 2592–2610. [WANG M, GUO X P, TANG H M, et al. Prediction case of core parameters of shale gas reservoirs through deep transformer transfer learning[J]. Chinese Journal of Geophysics, 2023, 66(6): 2592–2610.]
- [24] 邵蓉波, 肖立志, 廖广志, 等. 基于迁移学习的地球物理测井储层参数预测方法研究[J]. 地球物理学报, 2022, 65(2): 796–808. [SHAO R B, XIAO L Z, LIAO G Z, et al. A reservoir parameters prediction method for geophysical logs based on transfer learning[J]. Chinese Journal of Geophysics, 2022, 65(2): 796–808.]
- [25] SINAGA K P, YANG M S. Unsupervised *K*-means clustering algorithm[J]. IEEE access, 2020, 8: 80716–80727.
- [26] MACQUEEN J. Some methods for classification and analysis of multivariate observations[C]. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Berkeley, CA, USA, 1967.
- [27] ZHANG H, PENG Q. PSO and *K*-means-based semantic segmentation toward agricultural products[J]. Future Generation Computer Systems, 2022, 126: 82–87.
- [28] LIU W, SUN Y, YU B, et al. Automatic text summarization method based on improved textrank algorithm and *K*-means clustering[J]. Knowledge-Based Systems, 2024, 287: 111447.
- [29] BISEN D, LILHORE U K, MANOHARAN P, et al. A hybrid deep learning model using CNN and *K*-Mean clustering for energy efficient modelling in mobile EdgeIoT[J]. Electronics, 2023, 12(6): 1384–1409.
- [30] FARAGALLAH O S, HOSENY H M, SAYED H S. Efficient brain tumor segmentation using OTSU and *K*-means clustering in homomorphic transform[J]. Biomedical Signal Processing and Control, 2023, 84: 104712.
- [31] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 2016.
- [32] MORRIS M D. Factorial sampling plans for preliminary computational experiments[J]. Technometrics, 1991, 33(2): 161–174.

(编辑 付娟娟)

第一作者: 鄔德刚 (1999 年—), 博士研究生, 主要从事油气资源大数据与智能工程等方面研究, scholarwu@student.cup.edu.cn。

通信作者: 吴胜和 (1963 年—), 博士, 教授, 主要从事油矿地质学、储集层表征与建模方面的教学和科研工作, reser@cup.edu.cn。