

基于大数据分析的地球物理测井数据处理技术研究

高树林

新疆维吾尔自治区地质局乌鲁木齐地质大队

DOI:10.32629/gmsm.v9i2.2424

[摘要] 文章结合大数据分析开展测井数据处理技术研究。通过优化预处理流程、构建多尺度特征提取体系、适配优化智能算法及搭建多源数据融合框架,形成完整的测井数据处理技术体系。经油气盆地与盐矿实测数据验证,该技术显著提升岩性识别与储层参数反演精度,且在盐穴储库层位优选中应用效果良好。该技术可有效挖掘测井数据价值,为复杂地质场景的测井数据处理提供新路径。

[关键词] 地球物理测井; 大数据分析; 数据处理; 特征提取; 岩性识别

中图分类号: G633.7 **文献标识码:** A

Research on Geophysical Logging Data Processing Technology Based on Big Data Analysis

Shulin Gao

Urumqi Geological Brigade, Geological Bureau of Xinjiang Uygur Autonomous Region

[Abstract] This paper conducts a study on logging data processing technology in combination with big data analysis. A complete technical system for logging data processing is formed by optimizing the preprocessing flow, constructing a multi-scale feature extraction system, adapting and optimizing intelligent algorithms, and establishing a multi-source data fusion framework. Verified by field data from oil and gas basins and salt mines, this technology significantly improves the accuracy of lithology identification and reservoir parameter inversion, and achieves favorable application effects in the horizon optimization of salt cavern reservoirs. It can effectively exploit the value of logging data and provide a new approach for logging data processing in complex geological scenarios.

[Key words] Geophysical logging; Big data analysis; Data processing; Feature extraction; Lithology identification

引言

地球物理测井作为地下地质信息获取的核心手段,在油气勘探、矿产开发及地质灾害监测等领域发挥着不可替代的作用。随着数字化测井仪器的广泛应用,测井数据呈现多源化、高维度、海量性特征,传统基于机理模型的处理方法难以应对数据噪声干扰、分布差异及多尺度适配等挑战,导致数据价值挖掘不充分。大数据分析技术凭借强大的数据处理与模式识别能力,为测井数据处理提供了新路径。当前,如何通过大数据技术破解测井数据预处理繁琐、特征提取不精准、算法适应性差等难题,成为提升地质解释精度的关键。

1 地球物理测井大数据特征与处理挑战

1.1 测井大数据核心特征

地球物理测井大数据源于多种观测仪器与多学科数据融合,呈现鲜明的多维特征。数据来源涵盖电阻率、声波、自然伽马、中子、井径等20余种测井曲线,结合岩心扫描、地震剖面及地质编录等辅助数据,形成多源化数据体系,单口深度3000米的钻井

测井数据量可达500GB以上^[1]。在尺度维度上,数据跨度从微米级岩心孔隙扫描数据到千米级井间地震探测数据,不同尺度数据的关联性分析成为难点,如微米级数据用于孔隙结构表征,而千米级数据用于区域构造分析,二者的耦合关系需通过大数据技术建立。同时,测井数据具有高维度属性,单井测井曲线可达30余条,每条曲线包含10万个以上数据点,且数据分布呈现显著的非线性与非平稳性,如盐岩表现为低自然伽马、高电阻率特征,泥岩则呈现高自然伽马、低电阻率特征,极端差异数据的协同处理难度较大。

1.2 数据处理关键挑战

测井数据处理面临多重技术瓶颈。数据质量方面,受井筒环境、仪器误差及地质条件影响,数据中存在多种干扰因素:缺失值占比可达5%-15%,主要由仪器故障、井眼垮塌等导致;异常值包括仪器漂移引发的突变值、泥浆侵入造成的失真值等,占比约3%-8%;噪声干扰则来自地层干扰与仪器电子噪声,导致曲线平滑性不足。数据分布差异问题突出,不同区块、不同井眼的测井

数据因地质背景与测量条件不同,存在明显分布偏移,如东部某油气盆地与南方某盐矿的自然伽马值范围分别为20-120API与15-150API,传统模型难以直接迁移应用。标签稀缺成为智能处理的重要制约,人工标注测井数据需依赖资深专家,单井标注耗时可达3-5天,成本高昂且主观性强,导致有标签数据占比普遍不足5%。

2 基于大数据的测井数据处理核心技术

2.1 大数据预处理技术优化

数据预处理是提升测井数据质量、保障后续处理效果的基础,主要包括数据净化、数据填补与数据标准化三个关键环节。数据净化聚焦于噪声滤除与异常值处理,针对测井数据中的周期性噪声(如固体潮干扰),采用快速傅里叶变换(FFT)进行频谱分析,划分特定频率区间分离固体潮日潮、半日潮及三分之一潮成分,再通过自适应滤波器实现噪声滤除,该方法可在保留有效地质信号的前提下,显著降低周期性噪声干扰。异常值处理采用统计检验与物理约束相结合的策略,先通过箱线图法识别超出四分位距范围的异常点,再依据测井曲线的物理意义与地质规律进行验证,如自然伽马测井数据若超出对应岩性的理论取值范围,则判定为异常值,通过相邻数据的趋势插值进行修复,避免误删有效地质信号^[2]。

数据填补针对测井数据中的缺失值,采用基于时空相关性的插值算法。对于短时段缺失数据,利用相邻采样点的线性趋势进行插值;对于长时段缺失数据,结合同类地质体的测井响应特征,采用K近邻插值法,选取地质条件相似区域的有效数据作为参考进行填补,确保填补后的数据符合地层地质规律。数据标准化旨在消除不同类型测井数据的量纲差异,考虑到不同岩性地层的测井响应范围存在差异,采用分段归一化策略,根据岩性划分多个区间,对每个区间内的数据进行[0, 1]归一化处理,既保留了数据的相对差异,又避免了量纲对后续算法的影响。

2.2 多尺度特征提取方法

基于测井数据的多尺度特性,构建“宏观-中观-微观”分层级特征提取体系,实现不同维度地质信息的全面捕捉。宏观特征提取聚焦于地层整体趋势与周期性规律,采用快速傅里叶变换对测井数据进行频谱分析,根据固体潮理论划分22-26小时(日潮)、11-13小时(半日潮)、7.5-8.5小时(三分之一潮)等频率区间,分离周期性成分与非周期性地质信号,提取反映地层宏观构造与沉积规律的趋势特征,为区域地质背景分析提供支撑^[3]。

中观特征提取主要捕捉地层岩性突变、夹层分布等局部地质信息,采用小波变换技术对测井曲线进行多尺度分解。选用db4小波函数作为基函数,通过3-5层小波分解,获得不同尺度的近似系数与细节系数,近似系数反映曲线的整体趋势,细节系数反映局部突变特征。通过对细节系数的阈值处理与重构,可精准识别厚度较小的夹层、岩性界面等地质特征。

微观特征提取聚焦于孔隙结构、裂缝发育等细微地质信息,采用深度学习算法构建特征提取模型。构建3层卷积神经网络(CNN),输入层为多参数测井曲线组合(如自然伽马、电阻率、声

波时差),隐藏层通过16个 3×3 卷积核对输入数据进行特征映射,捕捉数据中的细微波动与关联模式,输出层为高维特征向量。该方法能够自动挖掘测井数据中不易被人工识别的微观特征,为精细储层评价提供支持。

2.3 智能算法适配与优化

针对测井数据处理的多样化任务需求,建立算法与任务的精准适配机制,实现岩性识别、储层参数反演等任务的高效处理。岩性识别任务面临标签稀缺与跨域适配难题,采用半监督学习与领域自适应相结合的Log2FT算法。该算法首先在标签充足的源域数据上训练初始模型,利用随机Fourier特征将数据映射到高维空间,增强模型的非线性拟合能力;随后引入样本加权机制,解决目标域样本类别分布不均衡问题;通过流形正则化挖掘数据的几何结构信息,使标签信息有效传播到无标签数据;最后利用目标域少量有标签数据进行模型微调,缩小源域与目标域的分布差异,提升模型的跨域适配能力。

储层参数反演任务需要实现孔隙度、渗透率、饱和度等参数的定量计算,采用集成学习算法融合多种模型的优势。选取随机森林、XGBoost与支持向量机三种算法构建集成模型,随机森林擅长处理高维数据与非线性关系,XGBoost具有较强的特征筛选与过拟合抑制能力,支持向量机在小样本数据中表现出良好的泛化性能。通过网格搜索优化各模型的超参数,采用加权投票方式融合各模型的预测结果,权重系数根据模型在验证集上的性能确定,确保集成模型能够充分利用各单一模型的优势,提升参数反演精度。

盐穴储库层位优选任务中,需精准计算盐岩品位、含矿率等关键参数,采用矿物组分模型法^[4]。根据盐岩地层的矿物组成,建立盐岩、泥质、膏硝(硬石膏与钙芒硝)、孔隙流体四组分体积模型,基于自然伽马、中子、声波时差等测井参数构建测井响应方程,通过最小二乘法求解各组分的体积含量,再根据各组分的密度换算盐岩品位与含矿率。

2.4 多源数据融合技术

构建测井-岩心-地震多源数据融合框架,通过数据互补提升地质解释的可靠性与全面性。测井与岩心数据融合聚焦于参数标定,选取具有代表性的岩心样本,通过实验室分析获得孔隙度、矿物含量等地质参数,建立这些参数与测井曲线的定量关系模型,利用多元线性回归优化模型系数,实现测井曲线对地质参数的精准表征。

测井与地震数据融合主要解决测井数据横向覆盖有限的问题,将地震数据提供的区域地质背景作为约束条件,融入测井数据处理过程。通过地震层位解释结果确定地层界面位置,修正测井曲线的深度对齐偏差;利用地震数据反映的横向岩性变化趋势,约束测井数据的横向插值,减少单一测井数据的多解性。

3 实验验证与应用分析

3.1 实验数据与设计

选取东部油气盆地某坳陷4口实验井及南方某盐矿某勘探孔的测井数据作为实验对象。东部油气盆地某坳陷数据涵盖自

然伽马、电阻率、声波时差等8种测井曲线,深度范围2000-3500米,数据量达120万条,标签类型包括泥岩(Mu)、砂岩(Sa)、页岩(Sh)三类;南方某盐矿某勘探孔数据包含自然伽马、三侧向电阻率等6种测井曲线,深度817-1140米,数据量35万条,涉及盐岩、泥岩、含泥盐岩等四类岩性。实验设置岩性识别、储层参数反演、盐穴层位优选三个任务,分别采用传统方法(多元线性回归、决策树)与本文提出的大数据处理技术进行对比,评价指标包括岩性识别准确率、储层参数反演均方根误差、含矿率计算相对偏差等,每种方法重复实验10次取平均值。

3.2 岩性识别结果

在东部油气盆地某坳陷实验中,本文技术体系的岩性识别平均准确率达94.3%,其中泥岩识别准确率96.7%,砂岩93.5%,页岩89.8%,相较于传统决策树算法(准确率85.7%)提升8.6个百分点。通过混淆矩阵分析,泥岩与页岩的误判率为3.2%,砂岩与含泥砂岩的误判率为4.8%,不同岩性的误判率均低于5%,表明多尺度特征与半监督算法的融合有效提升了识别精度,能够应对复杂岩性的区分需求。可视化结果显示,本文技术识别的岩性剖面与岩心编录结果的吻合度达93%,成功识别出3处厚度小于2米的页岩夹层,而传统方法未能有效区分。

3.3 储层参数反演效果

储层参数反演实验中,孔隙度计算均方根误差为0.85%,渗透率计算均方根误差为1.2mD,相较于传统阿尔奇公式法(孔隙度均方根误差1.25%,渗透率均方根误差1.67mD),误差分别降低32%和28%。在高孔隙度储层(孔隙度大于15%)中,本文技术的反演误差进一步降低至0.62%,表现出更强的适应性。实验结果表明,集成学习算法能够充分利用多参数测井信息,有效挖掘参数间的非线性关系,如电阻率与孔隙度的指数相关性、声波时差与渗透率的对数相关性,提升反演精度,为储层有效性评价提供可靠数据支持^[5]。

3.4 盐穴储库层位优选应用

在南方某盐矿某勘探孔的应用中,通过本文技术反演得到的盐岩平均品位为69.46%,含矿率51.1%,与矿区历史勘探数据

(平均品位67.84%,含矿率50.3%)相对偏差仅0.91%。优选出的906-1095m井段,累计厚度189米,泥岩夹层厚度以2-4米为主,无超过6米的厚夹层,直接顶板为891.3-898.7米的巨厚层泥岩,直接底板为1095.2-1102.2米的巨厚层泥岩,且顶板以上预留一定厚度盐层作为盖层,符合盐穴储库建设对稳定性与密封性的要求,验证了技术在非油气领域的适用性。

4 结束语

本文针对地球物理测井大数据的多源化、高维度等特征及处理中的多重挑战,系统构建了基于大数据的测井数据处理技术体系,涵盖预处理优化、多尺度特征提取、智能算法适配及多源数据融合核心技术。经东部油气盆地与南方某盐矿的实测数据验证,该技术在岩性识别、储层参数反演及盐穴储库层位优选中均表现出优于传统方法的性能,同时验证了其在油气与非油气领域的通用性。该技术有效破解了传统测井数据处理的痛点,实现了测井数据价值的深度挖掘,为地质勘探相关工作提供了切实可行的技术支撑,也为后续测井数据智能处理研究奠定了基础。

[参考文献]

- [1]肖立志.人工智能时代的地球物理测井:实践与展望[J].测井技术,2025,49(3):329-336.
- [2]刘东明,林振洲,刘东彦,等.地球物理测井在盐穴储库层位优选中的应用——以江西清江盐矿ZK01孔为例[J].煤田地质与勘探,2025,53(09):191-203.
- [3]朱剑兵,王继晨.一种改进的地球物理测井资料智能化解释评价方法[J].地球物理学进展,2025,40(6):2629-2641.
- [4]程希,周军,傅海成,等.机器学习算法在地球物理测井中的适用性及应用[J].西北地质,2023,56(04):336-348.
- [5]胡源,朱厚林.基于MATLAB的地球物理数据分析——以六安中心站VP摆为例[J].科技资讯,2025,23(21):161-164.

作者简介:

高树林(1974--)男,汉族,河南商丘人,助理工程师,中专,研究方向:地球物理测量。