

基于机器学习的致密储层流体识别方法研究

罗刚, 肖立志*, 史燕青, 邵蓉波

中国石油大学(北京)人工智能学院, 北京 102249

* 通信作者, xiaolizhi@cup.edu.cn

收稿日期: 2021-09-23

国家自然科学基金项目(42102118)、中国石油天然气集团有限公司项目(ZLZX2020-03)资助

摘要 机器学习算法已经成为工程领域建模的有力工具, 这些方法通过使用复杂结构或多重非线性变换从更高的维度拟合多个变量之间的非线性关系, 适用于解决工程中由于变量关系复杂, 传统物理模型或经验模型无法有效解决的问题。由于传统的测井资料解释方法是基于岩石物理机理和模型为基础, 需要许多假设条件, 在实际应用中存在一定偏差。因此, 利用机器学习算法对测井资料进行处理和解释, 尤其是对储层流体进行识别具有重要意义。现有的储层流体识别方法没有充分挖掘测井数据多维空间的关联性。此外, 还存在储层类别分布严重不均衡、相似物性储层的识别容易混淆的问题。本文针对储集层流体识别任务提出了基于机器学习算法的高效解决方案。采用长短期记忆网络(LSTM)和卷积神经网络(CNN)分别表征测井曲线时序特征以及多条测井曲线之间的相互关联关系; 考虑到油气储集层识别任务的类别分布不均衡性问题以及不同储层的价值排序有所差异, 本文采用加权交叉熵损失函数, 在模型训练中更注重学习少样本类别的特征, 进一步提升含油储层的识别准确度。依据储层物性差异和相似度, 设计了多层级储层流体识别方法, 将LSTM和CNN的模型结构应用于层级II(含油储层、含水储层和干层)和层级III(油层、油水同层、差油层和水层、含油水层)的识别。本方法在真实油田测井数据上加以验证, 其中数据分布极不均衡, 高价值的含油类储层占比9%, 符合真实工业场景。通过大量对比实验证明, LSTM和CNN的并行网络结构能够充分挖掘测井数据多维空间的相关特征; 加权交叉熵损失函数显著提高了高价值含油储层的识别准确度, 多层级储层流体识别方法对于物性相近的易混淆储层, 如油水同层和含油水层的区分更加准确。实验结果表明, 该方法能有效克服储层流体识别任务中的各种问题, 对于帮助地质专家和工程师寻找地下储层, 完成储层评价具有一定的实用价值。

关键词 机器学习; 损失函数; 测井资料; 油气储层; 流体识别

Machine learning for reservoir fluid identification with logs

LUO Gang, XIAO Lizhi, SHI Yanqing, SHAO Rongbo

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

Abstract Machine learning algorithms have become powerful tools for modeling in the engineering field. These methods fit the nonlinear relationships among multiple variables from a higher dimension by using complex structures or multiple nonlinear transformations. They are suitable for solving problems that cannot be effectively solved by traditional physical models or empirical models due to the complex relationship of variables in engineering. Since the traditional interpretation approaches of logging data are based on petrophysical mechanisms and models, many assumptions are needed, and there may be deviations in practical

引用格式: 罗刚, 肖立志, 史燕青, 邵蓉波. 基于机器学习的致密储层流体识别方法研究. 石油科学通报, 2022, 01: 24-33

LUO Gang, XIAO Lizhi, SHI Yanqing, SHAO Rongbo. Machine learning for reservoir fluid identification with logs. Petroleum Science Bulletin, 2022, 01: 24-33. doi: 10.3969/j.issn.2096-1693.2022.01.003

application. Therefore, when using machine learning for logging data processing and interpretation, reservoir fluid identification is of great significance. The existing reservoir fluid identification methods have not thoroughly mined the multi-dimensional correlation of logging data. Moreover, the distribution of reservoir types is seriously unbalanced. Reservoirs with similar physical properties may be easily confused. We present an efficient method using machine learning to identify reservoir fluids with logs. A long and short-term memory network (LSTM) is used to characterize the time series characteristics of logs varying with depth domain. The convolution kernel of the convolutional neural network (CNN) is used to examine multiple logging curves to characterize the correlation between them. Considering the unbalanced distribution of categories and the different value ranking of reservoirs, this paper uses the weighted cross entropy loss function to improve the weight of small sample categories in model training, which further improves the identification accuracy of oil-bearing reservoirs. According to the difference and similarity of reservoir physical properties, a multi-layer reservoir fluid identification method is designed. The LSTM + CNN model structure is applied to the prediction of layer level II (oil-bearing reservoirs, water-bearing reservoirs, and dry layer) and layer level III (oil layer, oil-water layer, poor oil layer, and water layer, oily water layer). This method is verified on the logging data of natural oil fields, in which the data categories distribution is highly unbalance. Moreover, the oil-bearing reservoirs account for 9%, which aligns with the actual industrial scene. A series of comparative experiments proved that the parallel network structure of LSTM and CNN can fully capture the correlation characteristics of the multi-dimensional space of logging data. The weighted cross-entropy loss function significantly improves the identification accuracy of high-development-value oil-bearing reservoirs. Moreover, the multi-layer reservoir fluid identification method is more accurate in avoiding confusing reservoirs with similar physical properties, such as oil-water layer and oily water layer. The experimental results demonstrate that this method can effectively overcome many of the problems in reservoir fluid identification. It has specific practical value to help geological experts and engineers find underground reservoirs and complete reservoir evaluation.

Keywords machine learning; loss function; logging data; oil and gas reservoir; fluid identification

doi: 10.3969/j.issn.2096-1693.2022.01.003

0 引言

测井是研究地下储集层岩性及物性、识别储层流体性质的有效途径。通过对测井资料的分析解释可以极大加深对地下情况的认知,测井资料是辅助专家寻找和评价储集层的重要资源。但是,由于测井资料的处理及解释以岩石物理机理模型为基础,需要许多先验条件,在实际应用中会出现不适应性而引起误差。测井数据和地层模型的不确定性及不适应性,使资料处理解释具有多解性。随着大数据时代的到来,学者相继尝试通过数据驱动的方法实现测井资料的处理与解释,提高生产效率^[1]。

源于20世纪50年代初的机器学习依次经历了推理期、知识期和学习期3个历史阶段,1943年,神经科学家麦卡洛克和数学逻辑学家皮兹提出了MCP模型^[2],使人工神经网络迎来发展时期,在夯实了神经网络模型理论基础的同时也标志着深度学习的开端。21世纪以来,面对信息化、大数据的时代,传统的机器学习算法在语音识别、图像识别、目标检测等领域的发展遇到了瓶颈,而计算能力的提升使深度学习突破了机器学习所面临的局限,与此同时也扩充了人工智能所涉及的领域,将机器学习相关理论应用到测井资料处理和解释,进一步提高解释结果的精度和可靠性,将具有重要的实际意义和研究价值。越来

越多的学者提出使用KNN、委员会机器、支持向量机、分布式梯度提升(XGBoost)等传统机器学习方法完成测井数据的处理与解释。Bestagini等人(2017)应用XGBoost模型根据测井数据预测岩相^[3];Ao等人(2018)提出用随机森林法解决岩相分类问题^[4];Zhou等人(2020)提出用合成少数过采样技术(SMOTE)平衡已标记的测井数据,并使用梯度提升决策树法识别岩相^[5]。白洋等人(2021)选用分类委员会机器预测致密砂岩气藏流体类别,提升预测精度和泛化能力^[6]。这些传统机器学习方法可解释性强,训练速度快,但模型结构相对简单、参数量少,难以表征复杂非均质情况下的地层环境。

1986年Rumelhart等人提出了BP神经网络,解决了单层感知器网络模型无法处理线性不可分的问题,从而掀起了人工神经网络研究的热潮^[7]。Hinton和Salakhutdinov提出利用深度自动编码网络学习得到初始权值的方法,有效地实现深层网络的数据降维,开启了深度学习广泛研究的序幕^[8]。深度神经网络通过使用包含复杂结构或多重非线性变换从更高维度拟合多变量之间的非线性映射关系,有利于表征复杂非均质的地层环境。近年来已提出一系列网络模型,其中卷积神经网络在图形图像处理领域已取得巨大成功,也是目前应用较为广泛的网络模型^[9]。最近,卷积神经网络也被应用于测井资料的处

理与解释。廖广志(2020)等人结合智能聚类、主成分分析等预处理方法对毛管压力数据进行处理,通过卷积神经网络(CNN)实现储层微观孔隙结构的预测并应用于储集层的划分^[10]。传统的机器学习方法和卷积神经网络在处理测井数据时将每一个采样点视作独立的样本,每个隐藏层的多个神经元互不相连,而多个隐藏层的神经元可以相互传递参数,因此无法获取数据前后采样点的内在联系,难以表征测井数据的序列特征。

循环神经网络(RNN)通过构建循环体结构使隐藏神经单元的数据与当前输入和过去输出均有关,从而学习序列数据前后采样点的关联特征。Hochreiter和Schmidhuber在常规循环神经网络的基础上进行优化改进,提出了长短期记忆(LSTM)神经网络,并由Graves(2012)进行了改良和推广^[11]。LSTM网络改进了常规循环神经网络简单的循环体结构,增加了3个门限(输入、遗忘和输出),使模型的权重可变,训练过程中避免梯度消失和爆炸的问题。安鹏等人(2019)将多条测井数据作为输入,应用LSTM网络进行泥质含量和孔隙度预测,结果表明基于长短期记忆网络模型的稳定性和准确度要明显优于传统全连接神经网络^[12];张东晓等人(2018)证明LSTM能够在序列数据中充分提取信息,并在合成测井曲线时优于其他模型^[13]。

随着油气田勘探开发的不断深入,机器学习算法在测井资料处理和解释方面已有了进一步结合,也取得了相应成果^[14]。但是,大部分研究都将算法应用于储层参数预测,在此基础上对测井曲线进行分析得到分层解释结论。对于储层流体识别任务而言,依旧没有规避解释精度不高、储层物性相近容易混淆的问题。本文目的在于充分利用机器学习方法的最新理论和成果,用测井数据特征工程及试油解释结论实现储层流体的自动识别。本文结合卷积神经网络和长短期记忆神经网络表征测井数据随深度域变化的时序特征以及多条测井曲线之间的相关性。针对目前存在的问题,采用加权交叉熵损失函数克服实际应用场景中类别分布不均衡的问题;基于储层物性差异和相似度设计了多层级储层流体识别方法,提升易混淆储层的识别准确度,最终用真实油田的测井数据验证本文提出方法的有效性。

1 原理与方法

测井数据是典型的序列式数据,随着井的深度增

加,相邻采样点测井响应存在相关性,后续采样点受到前序采样点的影响^[15]。由于地质条件复杂多变,地质运动导致褶皱、断层等构造,地下层位局部重叠,其对应的物理、化学性质也会相应改变,反映在测井曲线上的响应值也会随之变化,这就导致局部深度空间中的测井曲线呈现剧烈变化。地下储层厚度不一,不同深度区间测井响应值对单一深度尺度下地质特征具有明显易变性^[16];同时测井曲线特征相互依赖关联,例如自然电位与声波特征对地下岩层特征描述的共同作用等。本节将讨论用以表征测井数据时序特征以及多条测井曲线相互关联关系的机器学习算法。

1.1 长短期记忆网络

长短期记忆网络(LSTM)是一种适用于处理序列数据 $x^{(1)}, \dots, x^{(n)}$,表征数据时序关系的神经网络,其可以处理变长度的序列,因此长短期记忆网络适用于处理测井数据,可以对蕴含的层位信息进行更充分的挖掘。

LSTM网络主要由遗忘门(f)、输入门(i)、输出门(o),以及与隐藏状态形状相同的记忆细胞(c)构成,从而记录额外的信息,主要结构如图1所示。

LSTM神经元的输入向量为 x_t ;遗忘门的激活向量为 f_t ;输入门的激活向量为 i_t ;输出门的激活向量为 o_t ;LSTM神经元的输出向量为 h_t ;神经元细胞状态向量为 C_t ;权重矩阵为 W ;偏置项为 b ; σ 表示sigmoid函数; \tanh 表示双曲正切函数;下标 t 表示当前时刻, $t-1$ 表示上一时刻。

遗忘门查看上一输入和当前输入,选择性丢弃部分信息,表达式为

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

输入门将新信息存放到单元状态中,表达式为

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

单元状态包含小的线性相互作用使信息稳定向下传输,这种链式结构是长短期记忆网络的关键,表达式为

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (4)$$

输出门决定本时刻需要输出的信息,表达式为

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (6)$$

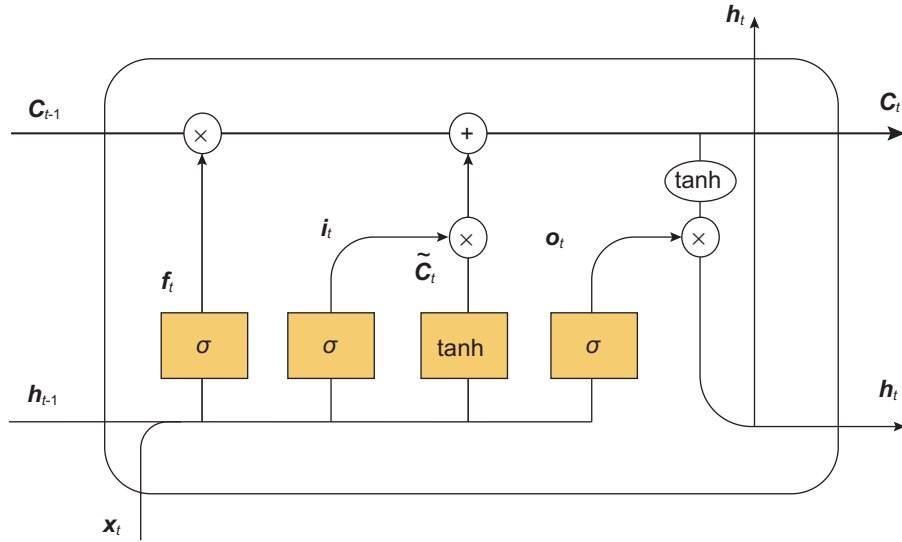


图 1 LSTM 网络的循环结构
Fig. 1 LSTM recurrent structure

1.2 卷积神经网络

输入层、卷积层、池化层、全连接层和输出层构成了卷积神经网络(CNN)的基本结构，卷积池化层一般为多组且交替设置，多个特征平面(Feature Map)组成卷积层，卷积核用于卷积操作提取特征，将上一层特征面的局部区域与每个神经元相连，在卷积层中每个神经元通过局部加权传递给饱和非线性 ReLU 函数输出，解决梯度膨胀或梯度消失的问题，加快收敛速度。

在 CNN 中，经过多个卷积池化层组后往往连接 1 个或多个全连接层。上一层所有神经元与全连接层进行全连接，整合卷积池化层组中具有类别区分性的关键局部信息，输出层接受最后一个全连接层的输出值，通过激活函数(softmax)使得输出结果在 [0,1] 之间，其中 w 为网络权值， M 表示储层类别数，最终输出为不同类别的概率向量 P ，如公式 7 所示：

$$P(i) = \frac{\exp(w_m^T x)}{\sum_{m=1}^M \exp(w_m^T x)} \quad (7)$$

测井曲线的卷积过程如图 2 所示，卷积核横向滑动学习多条测井曲线关联性，同时纵向滑动学习测井曲线局部深度响应特征。因此卷积神经网络适用于处理测井数据，在表征多条曲线关联性和深度局部变化上具有显著优势。

2 多层级储层流体识别方法

2.1 网络结构

网络输入为多段测井数据序列 $x_t = [x_t^1, x_t^2, \dots, x_t^n]$ ，每段由 N 条测井曲线组成，每条测井曲线包含 M 个采样点，每段储层的测井曲线采用二次插值法，使得 $M = 60$ ，因此网络输入数据为 $N \times 60 \times K$ ，其中 K 表示储层数量。

设置 CNN 和 LSTM 并行网络，其中卷积神经网络设置为 8 通道卷积，卷积核大小为 3×3 ，步长为 2，使用 SAME 卷积方式，应用 ReLU 激活函数输出 8 个 8×60 的矩阵，将 8 个矩阵展开成一维向量并连接，送入隐藏层，通过 Dropout 随机丢弃一些神经元，不更新其网络权值，再输入一层隐藏层，通过全连接后得到长度为 60 的向量。LSTM 层中每个单元中隐藏层数量设置为 128，数据通过 LSTM 输出为长度 128 的一维向量，再输入一层具有 64 个神经元的隐藏层，输出为长度 64 的一维向量。

将 LSTM 与 CNN 输出的一维向量拼接，通过全连接层后用 softmax 函数输出储层类型标签向量 P ，具体流程如图 3 所示。

2.2 加权交叉熵损失函数改进

传统分类任务中，通过最小化交叉熵损失函数 L_{ori} 来优化模型，提升模型预测准确度。

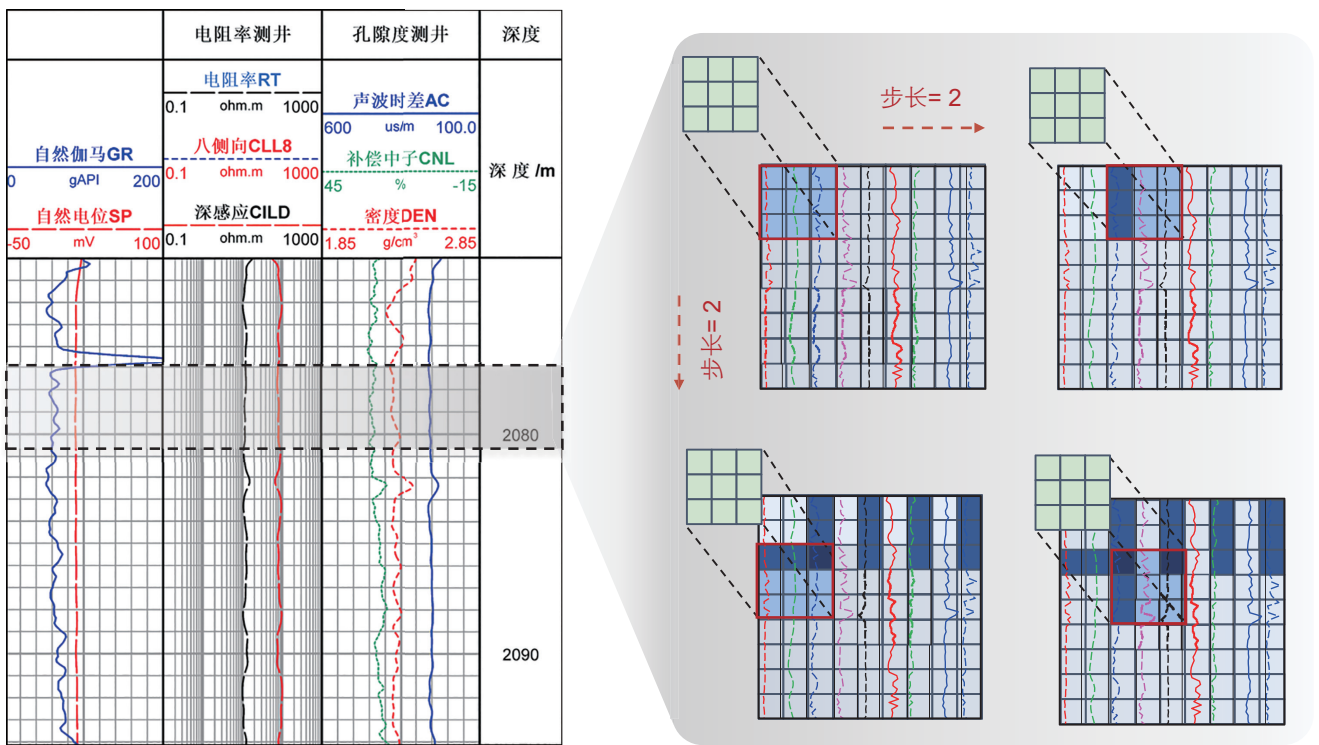


图2 卷积神经网络在多条测井曲线上的卷积过程

Fig. 2 Convolution process of convolutional neural network on multiple logging curves

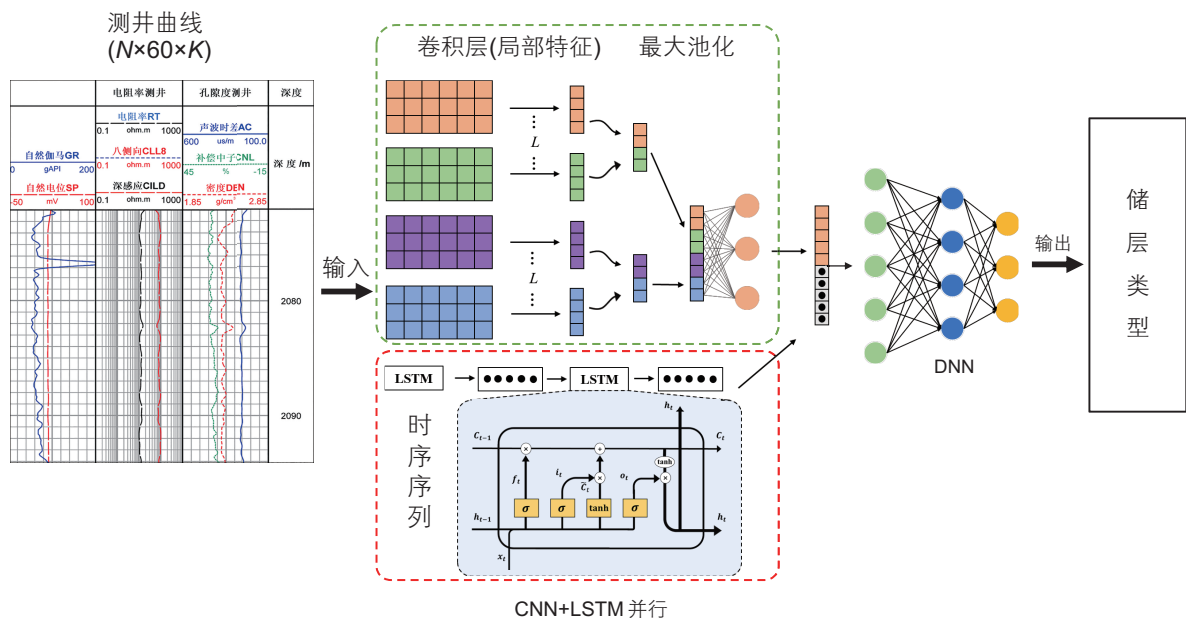


图3 基于LSTM+CNN储层预测流程图

Fig. 3 Flow chart of reservoir identification based on LSTM + CNN

$$L_{ori} = -\frac{1}{N} \sum_{m=1}^M y_m \log(y'_m) \quad (8)$$

其中，训练集总样本数以及油气储集层类别数用M和N表示，类别真实值和模型对该类别的预测值用 y_m 和

y'_m 表示。

对于油气储集层识别任务而言，水层和干层占比较大，而油层、油水同层、含水水层的占比较小，类别分布不均衡性十分突出；另一方面，由于不同储层

经济效益有所差异，其准确度的要求也不同，例如对于油层具有开采价值的储集层准确度要求就会高于对于水层、干层等开采价值较低的储集层。基于此，本文将采用基于类别加权的优化函数，在模型训练中提高少样本类别的权重。

本文设计的加权交叉熵优化函数 L 如下所示：

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{m=1}^M \alpha_m y_m \log(y'_m) \quad (9)$$

其中， α_m 表示 m 类别的权重，默认将权重设置为 m 类别占总体样本比例的倒数。

2.3 多层次储层流体识别方法

在储层流体识别任务中，部分储层的油水性质难以根据其流体饱和度界定，例如水层、油水同层和含油水层的含油饱和度差异较小，难以用数值准确区分。为此本文提出多层次储层流体识别方法，以克服部分储层流体物性相似，反应在测井曲线上的响应差异较小，导致流体识别过程中容易混淆的问题。

储层具有连通孔隙，油水在其中储存，而非储层相反，因此储层与非储层差异较大，层级 I 对两类进行识别。根据储层含水和含油量差异可将其分为 3 大类：含水储层（水层、含油水层）、干层、含油储层（油层、差油层、油水同层），因此层级 II 对 3 类进行划分；最后对含水储层和含油储层进行精细划分，由此构成了层级 III。基于此提出的多层次储层流体识别方法如图 4 所示。

实现方法与主要步骤如下：

层级 I：以多条测井曲线作为输入，用 XGBoost 建立储层、非储层的分类模型，确定储层位置和厚度。其中，XGBoost 是一种基于回归树的传统机器学习算法，该算法具有运算速度快、效果好等特点。

层级 II：建立含水储层、含油储层和干层的 LSTM

+ CNN 分类模型，将层级 I 中预测为储层的采样点作为输入，对储层进行分类。

层级 III：

(1) 建立油层、差油层、油水同层的 LSTM + CNN 分类模型，将层级 II 中预测为含油的储层作为输入，进一步划分储层类型。

(2) 建立水层、含油水层的 LSTM + CNN 分类模型，将层级 II 中预测为含水的储层作为输入，进一步划分储层类型。

3 实验评估

实验选取某致密砂岩油田 173 口井作为数据集。致密砂岩储层存在低孔低渗、物性较差、孔隙结构及油水关系复杂、束缚水饱和度高、流体性质及饱和度难以确定等问题。因此，准确识别致密砂岩储层的流体类型，尤其是找到地层中的含油储层存在一定的难度。此外，在该区域数据集中，水层和干层占比 91%，含油储层占比 9%，数据分布不均衡，在模型训练时往往容易忽略少样本类别的特征，导致储层预测时难以识别含油储层。基于以上问题，分别从网络结构和多层次储层识别流程验证本文提出方法的有效性。

3.1 实验数据

选取声波时差 (AC)、自然伽马 (GR)、密度 (DEN)、自然电位 (SP)、补偿中子 (CNL)、电阻率 (RT)、感应电导率 (CILD)、侧向电导率 (CLL8) 等八条测井曲线，删除含有异常值的采样点并对每口井进行归一化，建立多层次标签数据集 (表 1)。实验选取 3 口井为测试井，剩余井按照 3:1 的比例划分训练集和验证集。

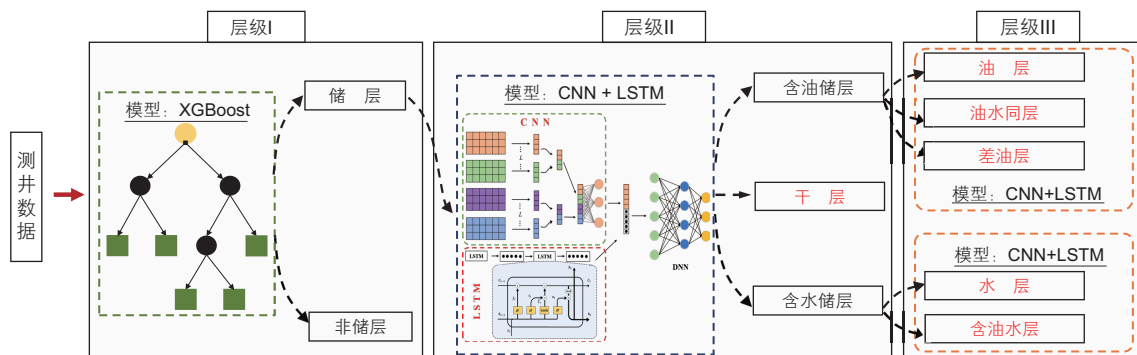


图 4 多层次储层流体识别方法流程图

Fig. 4 Flow chart of multi-layer reservoir fluid identification method

表1 多层次标签数据集

Table 1 Multi scale label data set

	标签 1	标签 2	标签 3
层级 I	储层	非储层	
层级 II	含油储层	含水储层	干层
层级 III	油水同层	油层	差油层
层级 III	水层	含油水层	

3.2 实验设置

本实验均使用python和Tensorflow实现,为提升神经网络的训练效率,采用批训练法进行模型训练,批次(Batch size)设置为512,即每次训练随机抽取512组训练数据,学习率设置为0.001,选用Adam优化器。

本实验根据混淆矩阵计算出精确率 P_r 、召回率 R_e 和 F_1 。混淆矩阵也称为误差矩阵,是精度评价最常用的方法。需要指出的是在数据极度不均衡场景任务中,高召回率 R_e 比高精度 P_r 更有意义,即可以更多地找到所有样本中的含油储层,但相对会将部分其他储层(如水层等)误分为油层。这是由于油层具备更高的开采价值,相比较其他类储层被误分为油层,遗漏地层中的油层往往损失更大,而前者可以通过取岩心试油来进一步排除非油层。因此,在本文中,召回率 R_e 将作为的主要评价指标。

3.3 实验结果分析

3.3.1 基于CNN和LSTM的单层级储层流体识别方法

本实验将验证基于长短期记忆网络(LSTM)和卷积神经网络(CNN)在储层流体识别任务中的有效性,对比以下3种单层级储层分类方法:

(1)基于长短期记忆网络的单层级储层分类方法:模型输入为多段测井数据序列 $x_t = [x_t^1, x_t^2, \dots, x_t^n]$,数

据通过LSTM输出为长度128的一维向量,再输入两层隐藏层,最后输出预测的储层类型。

(2)基于长短期记忆与卷积神经网络并行架构的单层级储层流体识别方法:多段测井曲线作为输入,CNN和LSTM分别提取特征,将输出的一维向量拼接后再经过两层全连接,最终输出储层类型。

(3)在CNN和LSTM并行网络架构的基础上引入加权交叉熵损失函数。

以上3种实验方案在验证集上的预测结果(召回率)如表2所示。

从表中可以看到基于LSTM的储层流体识别方法对于水层和干层的预测效果较好,召回率 R_e 相对较高,油层、差油层、油水同层、含油水层的预测效果较差,这是由于在储层流体类别中,水层和干层的占比过大,样本的不均衡性导致模型在训练时忽略了含油储层的特征;引入CNN后,油层、含油水层、油水同层和差油层的召回率均有提升,其中油层提升23%,差油层提升5%。由此可以证明基于CNN和LSTM并行的网络模型架构可以有效表征测井数据的时序特征以及多条测井曲线之间的相互关系。

对损失函数改进后,虽然水层和干层的识别效果降低,但对于高价值储层的识别效果有所增强,油层召回率提升29%,含油水层提升22%,油水同层提升34%,差油层提升26%。由此可以证明改进模型的优

表2 CNN+LSTM网络结构有效性证明实验结果

Table 2 Experimental results of reservoir classification method after introducing weighted cross entropy optimization function

	LSTM	LSTM+CNN	LSTM+CNN+Weight_loss
油层	0.07	0.30	0.59
差油层	0.15	0.20	0.46
油水同层	0.05	0.06	0.40
含油水层	0	0.01	0.23
水层	0.95	0.94	0.63
干层	0.85	0.86	0.69

化函数可以有效克服储层流体识别任务中类别不均衡性的问题，进一步提升含油储层的识别效果。

3.3.2 多层次储层流体识别方法

本实验将基于已有结论验证多层次储层流体识别方法的有效性。基于最优的LSTM和CNN并行网络架构，采用加权交叉熵损失函数训练3个模型：

- (1) 含油储层、含水储层、干层分类模型；
- (2) 油层、油水同层、差油层分类模型；
- (3) 水层、含油水层分类模型。

将3个模型集成在一起实现层级II和层级III的多层级储层流体划分，并与CNN+LSTM单层级直接判别六种储层流体类型的方法进行比较，在验证集上的预测效果(召回率)如表3所示。

多层次储层流体识别方法相比较单层级直接判别的方法，干层召回率提升13%，水层提升2%，含油水层提升12%，对于物性相近、较易混淆的储层(水层和含油水层)的识别效果较好，由此验证了多层次储层流体识别方法可以充分考虑到不同储层由于物性不同，反应在测井曲线上的响应差异，从而提升整体的识别准确度。

3.3.3 测试井实验结果与分析

层级I训练划分储层和非储层的XGBoost模型，在此基础上将3个LSTM+CNN模型集成在一起，完成层级II和层级III的划分，实现图4完整的多层级储层流体识别流程。选取测试集中的井验证预测效果，并与多种基于机器学习的储层流体识别方法对比。实验结果如表4和图5所示。

可以看出，XGBoost算法和单层级CNN+LSTM网络模型对于含油储层的预测效果较差，将大部分含油储层识别为水层。使用加权交叉熵损失函数可以有效识别出含油储层，但是对于含油饱和度相近的储层(油层、含油水层、油水同层)难以划分准确。本文提出的多层次储层流体识别方法比单层级直接判别具有更高的准确度，可以找到地层中的含油储层，对于油层、油水同层和含油水层的区分效果有进一步提升。然而，针对部分储层流体的识别仍会出现混淆的情况，因此在未来的研究中将对测井数据做进一步分析，通过特征工程提升储层的可分离性，对易混淆储层做更准确的划分。

表3 多层次储层流体识别方法效果对比

Table 3 Results of multi-layer reservoir fluid identification method

	CNN+LSTM+Weight_Loss 单层级直接判别	CNN+LSTM+Weight_Loss 多层次储层流体识别方法
油层	0.59	0.59
差油层	0.46	0.28
油水同层	0.40	0.36
含油水层	0.23	0.35
水层	0.63	0.65
干层	0.69	0.82

表4 多方案在测试井的储层预测效果对比

Table 4 Comparison of reservoir identification results of multiple schemes in test wells

测试井	真实储层深度段/m		预测储层深度段/m		真实储层	XGBoost	CNN+LSTM	CNN+LSTM	CNN+LSTM
	起始深度	结束深度	起始深度	结束深度				weight loss 单层级方法	weight loss 多层次方法
I	2953.75	2957.38	2954.00	2960.13	差油层	油层	差油层	差油层	差油层
	3004.75	3016.50	3005.00	3016.88	油水同层	水层	油层	油层	油水同层
	2428.00	2433.25	2427.50	2433.25	油层	水层	水层	油水同层	油水同层
II	2447.50	2449.88	2445.80	2450.00	含油水层	水层	水层	含油水层	水层
	2450.62	2455.88	2450.63	2455.38	含油水层	水层	水层	油水同层	含油水层
	2352.00	2356.00	2351.63	2365.50	含油水层	含油水层	水层	含油水层	含油水层
III	2476.80	2478.30	2476.50	2478.88	油层	水层	水层	含油水层	油层
	2771.00	2775.80	2771.88	2773.75	水层	水层	水层	含油水层	水层

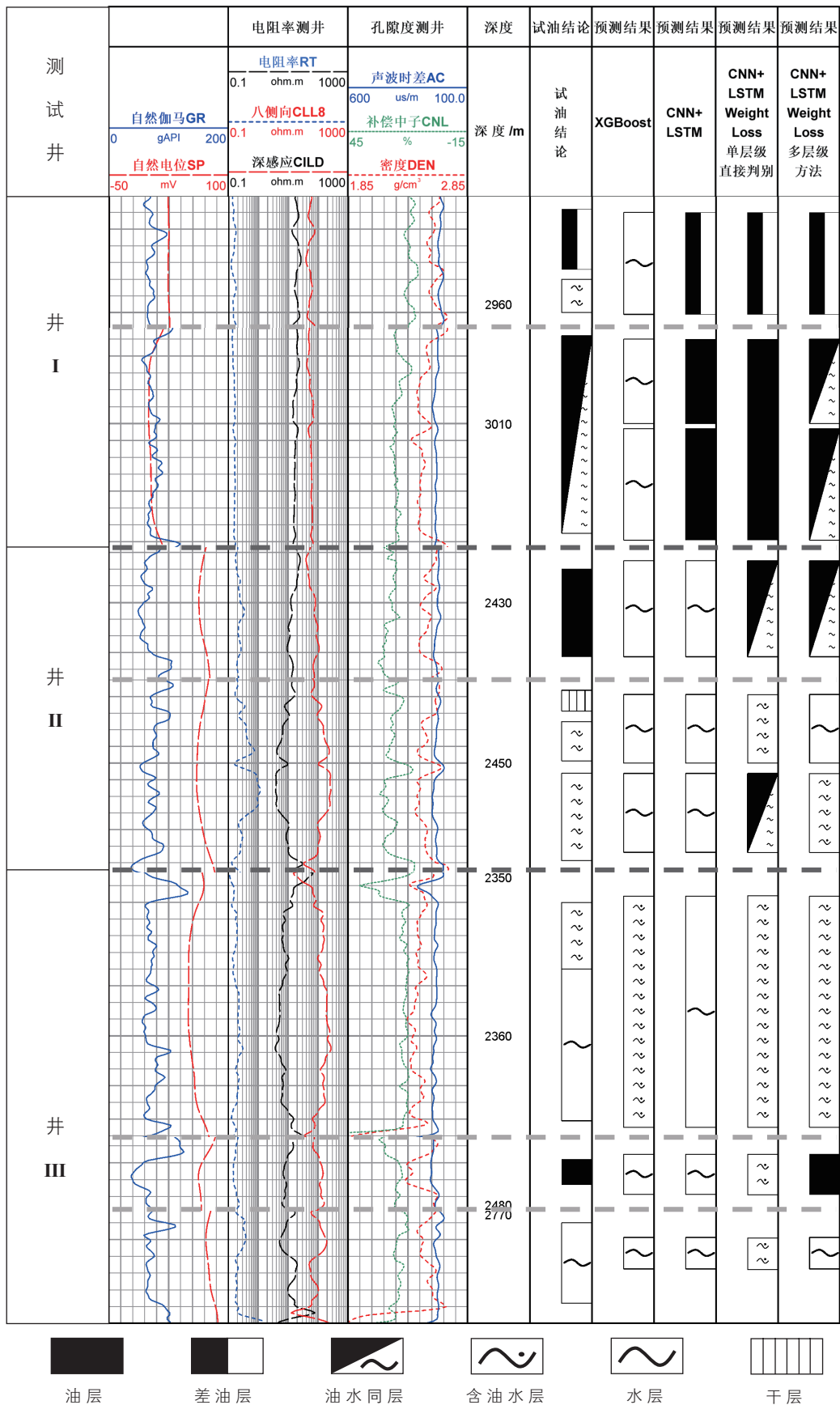


图 5 多方案在测试井的解释成果图

Fig. 5 Interpretation results of multiple schemes in test wells

4 结论

本文利用机器学习对基于测井数据的储层流体识别进行了研究,针对不同储层物性差异,提出多层次储层流体识别方法;采用长短期记忆网络学习测井数据的序列特征;引入卷积神经网络表征多条测井曲线间的相互关系以及测井曲线随深度的局部变化特征;考虑到油气储集层识别任务的类别不均衡性以及价值排序差异,引入加权交叉熵损失函数,在模型训练中

增加含油储层的权重。提出的多层次储层流体识别方法提升了油层、油水同层和含油水层的识别准确度,证明该网络模型结构可以适应测井数据和地层条件特性,在储集层预测任务中表现优异,可以帮助地质学家和工程师利用测井资料开展沉积岩石学的研究及寻找地下储层和实现储层评价,具有理论意义和实用价值。

致谢 郭云龙、陈梦玲、罗嗣慧、廖广志教授等给予多方面帮助,评审专家给出建设性修改意见和建议,在此一并致谢。

参考文献

- [1] 李宁,徐彬森,武宏亮,等.人工智能在测井地层评价中的应用现状及前景[J].石油学报,2021,42(4):508-522.[LI N, XU B S, WU H L, et al. Application status and prospects of artificial intelligence in well logging and formation evaluation[J]. Acta Petrolei Sinica, 2021, 42(4): 508-522.]
- [2] WARREN S M, WALTER P. A logical calculus of the ideas immanent in nervous activity[J]. The Bulletin of Mathematical Biophysics, 2006, 5(4).
- [3] BESTAGINI P, LIPARI V, TUBARO S. A machine learning approach to facies classification using well logs[C]// SEG Technical Program Expanded Abstracts 2017: Society of Exploration Geophysicists, 2017, 2137-2142.
- [4] AO Y, LI H, et al. Logging lithology discrimination in the prototype similarity space with random forest[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 16: 687-691.
- [5] ZHOU K, ZHANG J, REN Y, et al. A gradient boosting decision tree algorithm combining synthetic minority oversampling technique for lithology identification. Geophysics, 2020, 85: WA147-WA158.
- [6] 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(05):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(05): 1745-1758.]
- [7] RUMELHART D E, HINTON G E, Williams R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533-536.
- [8] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [9] 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(CVPR). Las Vegas, 2016.
- [10] 廖广志,李远征,肖立志,等.利用卷积神经网络模型预测致密储层微观孔隙结构[J].石油科学通报,2020,5(01):26-38.[LIAO G Z, LI Y Z, XIAO L Z, et al. Prediction of microscopic pore structure of tight reservoirs using convolutional neural network model[J]. Petroleum Science Bulletin, 2020, 5(01): 26-38.]
- [11] GRAVES A. Supervised sequence labelling with recurrent neural networks[M]. Springer Berlin Heidelberg, 2012.
- [12] 安鹏,曹丹平,赵宝银,等.基于LSTM循环神经网络的储层物性参数预测方法研究[J].地球物理学进展,2019,34(05):1849-1858.[AN P, CAO D P, ZHAO B Y, et al. Reservoir physical parameters prediction based on LSTM recurrent neural network[J]. Progress in Geophysics, 2019, 34(5): 1849-1858.]
- [13] ZHANG D, CHEN Y, MENG J. Synthetic well logs generation via recurrent neural networks[J]. Petroleum Exploration and Development, 2018, 45: 629-639.
- [14] JIANG C, ZHANG D, CHEN S. Lithology identification from well log curves via neural networks with additional geological constraint[J]. Geophysics, 2021: 1-77.
- [15] 胡家琦,孙连山,石敏,等.基于RNN-FCNN的多尺度油水层识别方法[J].高技术通讯,2020,351(03):97-105.[HU J Q, SUN L S, SHI M, et al. Multiscale oil-water layer recognition method based on RNN-FCNN[J]. Chinese High Technology Letters. 2020, 351(03): 97-105.]
- [16] 李微,刘淇,王喆锋,等.基于地质知识蒸馏学习的油气储集层识别方法[J].中国科学:信息科学,2021,51:40-55.[LI Z, LIU Q, WANG Z F, et al. Method of oil and gas reservoir detection based on geological knowledge distillation learning[J]. Sci Sin Inform, 2021, 51: 40]