

基于 BiLSTM 的地质片段层位预测方法^①

金相臣^{②***} 吴子锐^{*} 石 敏^{③*} 朱登明^{**} 周 军^{***}

(^{*}华北电力大学控制与计算机工程学院 北京 102206)

(^{**}中国科学院计算技术研究所前瞻研究实验室 北京 100190)

(^{***}中国石油集团测井有限公司 西安 710065)

摘要 地质分层是指对某一个地区的地层剖面中的岩层进行划分,可用于指导相应的地质找矿工作。传统的地质分层主要依靠专家根据经验进行人工判断,然而由于地质层位类别繁多,需要消耗大量的时间和人力成本。现有的地质层位自动识别方法,由于没能考虑到测井数据的序列关系以及地质层位分布的特点,导致识别效果较差。基于此,本文提出了一种改进的双向长短期记忆神经网络(BiLSTM)的地质片段层位预测方法,可以根据测井数据自动快速地进行地质分层预测。该方法首先对测井数据进行分段处理,然后基于片段式的数据对 BiLSTM 网络进行相应的修改,其充分利用了地质层位片段式分布的特点,且考虑到了测井数据两个方向上的序列相关性。实验结果表明,本文方法在某油田真实井位数据集上的识别准确率达到了 93%,相较于其他网络有着显著的效果提升。

关键词 双向长短期记忆神经网络(BiLSTM); 测井曲线; 地质分层; 层位预测

0 引言

地质分层是指对一个地区的地层剖面中的岩层进行划分,是研究地球、岩石、岩性的基础,也是研究岩石、岩性的结果。对地质进行分层可以了解当地、当时的地质作用过程,并指导相应地质找矿工作和社会经济建设。岩性是地质分层的主要依据,然而,考虑到岩性包含了岩石的颜色、成分、结构等多种属性,通过人工判断来确定地质层是十分困难的过程,且会带来大量的时间、经济及人力成本。因此,需要一种自动分层技术来辅助对地质层进行快速的划分。

测井数据是利用不同测井仪由测井电缆下入井内,沿着井筒连续等间隔记录随深度变化的各种参数,如 GR(自然伽马)、SP(自然电位)、AC(声波时

差)等。这些参数反映出不同的岩性、层位特征,根据这些参数可以判断出具体的地质层位信息。通过聚类分析法、支持向量机(support vector machine, SVM)、人工神经网络等方法可以借助测井数据实现对地质层位的自动划分。然而,由于测井数据的参数以及地质层位的类别繁多,导致在测井数据与地质层位之间构建出映射关系尤为困难。

综上,本文提出了一种改进的双向长短期记忆神经网络(bidirectional long-short memory neural network, BiLSTM)的地质片段层位预测方法,可以利用测井数据快速地预测出地质层位信息。首先,针对地质层位片段式分布的特点,对测井数据进行分段处理,将一定数量的采样点数据归为一个片段;其次,基于 BiLSTM 建立层位预测模型,并将模型中的部分长短期记忆网络(long short-term memory, LSTM)计算单元合并,合并的计算单元数量与片段中的采

① 国家重大科技专项(2017ZX05019005)资助项目。

② 男,1997 年生,硕士生;研究方向:虚拟现实;E-mail: shzhjxc@163.com

③ 通信作者,E-mail: shi_min@ncepu.edu.cn

(收稿日期:2020-07-06)

样点数量保持一致;最后,对预测结果进行错层消除,使结果更加准确。实验结果表明,本文方法在某油田真实井位数据集上的识别准确率达到了 93%,相较于其他网络有着显著的提升。

1 相关工作

目前,针对地质分层问题的研究思路可分为数理统计法、人工神经网络等。聚类分析法作为一种常见的地质分层算法^[1-3],根据测井曲线的方差或者偏差平方和作为分层指标,并依据上述统计值进行分类。分层后,相同类的方差或者偏差值差异小,不同类别的差异较大。然而,在真实的层位分布中,即便是同一层位中的测井数据也有可能有较大的差异,不同层位间的测井数据也有可能差异较小,这就导致了该方法存在一定的局限性。支持向量机方法在小样本、非线性和高维数据模式识别问题中能够有好的发挥,因此有学者^[4-7]尝试将 SVM 用于测井解释。然而,SVM 对缺失数据敏感,且其在分类类别数较多时无法进行准确地识别,导致该方法并不适合用于地质层细分的任务。

近年来,神经网络的应用越来越广泛,也有很多学者将其用于地质分层与石油分层。反向传播神经网络(back-propagation neural network, BPNN)因其具有自适应能力与泛化能力,其在测井解释中也有着广泛的应用^[8-11]。但该方法应用于地质分层问题时仍存在许多局限性,如学习过程收敛慢、学习率不稳定、容易陷入局部最小值等。卷积神经网络(convolutional neural network, CNN)能够有效地提取数据特征,特别是对于测井这种多属性的问题,CNN 相比之前的神经网络少了许多网络参数,这使其在训练高维数据时有了更大的优势。陈钢花等人^[12]利用 CNN 进行岩性识别,段友祥等人^[13]将 CNN 用于储层识别,张国印等人^[14]在 CNN 的基础上通过地震数据进行储层识别,都取得了较好的成果。然而,这些分类方法都只是探寻测井属性与地质信息之间一对一的映射关系,却忽略了测井数据的序列性。测井数据在纵向上的变化存在连续性,这也符合了地质层在纵向上的分布变化。

BiLSTM 作为一种循环神经网络,可以反映出某采样点测井数据在地质中与其周围采样点数据的相关性,可以充分结合其所在区域的数据特征进行地质层位判断。因此,本文拟将 BiLSTM 用于地质层位预测的任务中。

2 算法描述

2.1 测井数据特征分析

测井数据是利用不同测井仪由测井电缆下入井内,沿着井筒连续等间隔记录随深度变化的各种参数。依据测井专业技术人员的指导信息以及相关实验分析,本文选取了 5 种测井属性作为识别地质层位的依据,如表 1 所示。

表 1 测井属性

测井符号	中文名称
GR/mg	自然伽马
SP/mv	自然电位
AC/ μ s/m	声波时差
AT90/ $\Omega \cdot m$	阵列感应电阻率
Depth/m	深度

测井曲线是指测井属性的各测量值根据深度变化而绘制出来的曲线,图 1 给出了 GR、SP、AC、AT90 4 种测井属性对应的测井曲线。

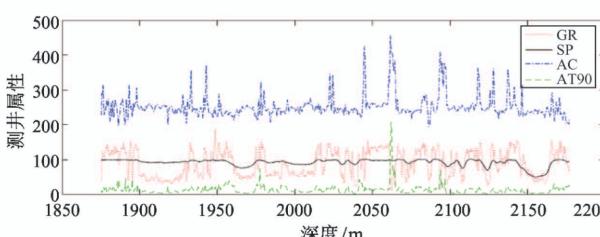


图 1 测井曲线

图 1 中横坐标表示测井数据的深度信息,纵坐标表示各测井属性在对应深度采样点的值。在每个深度的采样点处都会有一组不同参数的信息,同时也会有该采样点所对应的层位信息。从曲线图可以看出,测井属性的变化是连续的,符合序列数据的特点。以往的研究都是通过每一个采样点的测井数据

来判断该采样点所处的地质层位, 这样使得每个采样点之间的数据相互割裂, 只探讨了测井数据间的相似性。而在实际的地质层位中, 即使两个不同的层位之间, 测井数据也可能存在一定的相似性, 如果单从离散的采样点的角度考虑, 很难将其区分。要想准确地判断某个采样点所处的地质层位, 就要同时考虑该采样点周围的数据变化情况。

为此本文提出将多组相邻采样点的数据归为一个片段, 将一口井的数据分为多个片段。假定原始测井数据集为 $P = \{p_i, i = 1, 2, 3, \dots, N\}$, 其中 N 为采样点个数, p_i 为第 i 个样本数据点, 是多个测井属性的取值集合。若分段长度为 L , 那么经过处理后, 测井数据集 P 是分段的集合, 如式(1)所示。

$$P = \{S_1, S_2, \dots, S_k\}, k = \frac{N}{L} \quad (1)$$

式中, L 的具体值需要通过实验确定, 每一个片段 S 都包含了等量的采样点。在预测地质层位的时候, 通过每一个片段来预测该片段所处的地质层位。通过片段预测可以充分利用片段中的所有数据以及这些数据的变化来判断这个片段属于哪一个层位, 相比之前的单个采样点, 片段所包含的信息更加丰富, 精确识别地质层位的可能性也会更大。

2.2 基于 BiLSTM 的地质层位预测模型

经过上述分析, 测井数据属于一种序列数据, 循环神经网络适合用于处理序列数据, 而长短记忆网络则解决了循环神经网络 (recurrent neural network, RNN) 训练神经网络过程中梯度消失和梯度爆炸的问题, 使其能够保留更久以前的信息。LSTM 的网络结构与 RNN 大体接近, 但是隐藏层的结构更为复杂, 如图 2 所示。

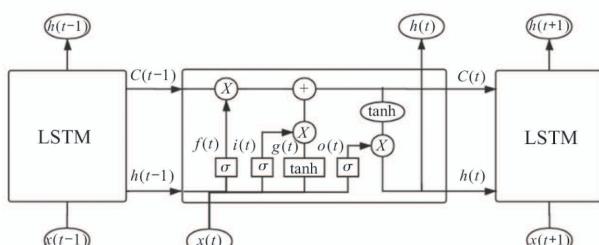


图 2 LSTM 网络结构

图中, $x(t)$ 表示 t 时刻的输入值, $h(t)$ 表示 t 时刻的输出值。 $C(t)$ 表示以 t 时刻累积的历史输入

值。LSTM 计算单元主要由输入门 (用 $i(t)$ 表示)、遗忘门 (用 $f(t)$ 表示)、输出门 (用 $o(t)$ 表示) 组成。LSTM 通过输入门、遗忘门和输出门 3 个门的作用来调控信息的流向以及筛选信息, 解决了网络训练过程中出现的梯度消失和梯度爆炸问题。

LSTM 网络在训练时, 由于前面几个节点所包含的数据相对后面的节点包含的数据较少, 经常会出现前半部分预测结果较差, 后半部分预测结果较好的现象。因此本文对 LSTM 网络进行一定的修改, 将若干个 LSTM 计算单元合并成一个单元, 合并的 LSTM 计算单元数量由测井数据片段的长度动态决定。假设一个测井数据片段包含了 n 个采样点, 则在 LSTM 网络中把每 n 个 LSTM 计算单元合并成一个新的 newLSTM 计算单元, 合并后的神经网络则由这些新的 newLSTM 计算单元组成。合并后的 newLSTM 计算单元如图 3 所示。

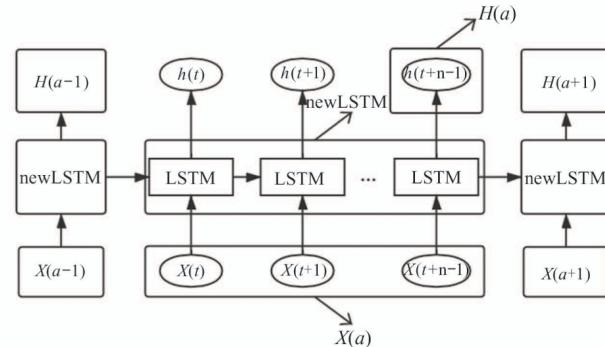


图 3 newLSTM 计算单元

图中的 newLSTM 表示将多个 LSTM 计算单元合并之后的新的计算单元, $x(t)$ 、 $x(t+1)$ 、 $x(t+n-1)$ 表示采样点的测井数据, $X(a-1)$ 、 $X(a)$ 、 $X(a+1)$ 表示片段的测井数据, 即 newLSTM 的输入, $h(t)$ 、 $h(t+1)$ 、 $h(t+n-1)$ 表示采样点的层位信息, $H(a-1)$ 、 $H(a)$ 、 $H(a+1)$ 表示片段的层位信息, 即 newLSTM 的输出。

LSTM 计算单元合并之后仍然保留了 LSTM 网络的特点, 其中每一个 newLSTM 计算单元中输入的是片段数据, 包括了多个采样点的数据, 而 newLSTM 的输出则是片段中最后一个采样点的输出, 也是对这个片段所处层位的预测。该输出包含了片段中所有采样点输入的测井数据信息以及数据间的连

续变化关系,也是最能反映该片段的数据特点的一个输出。如此就相当于抛弃了前半部分预测较差的结果,用较好的预测结果来代替,进一步提高了识别准确率。

然而 LSTM 网络只处理了一个方向上的序列关系,BiLSTM 则在 LSTM 的基础上进行了进一步的改进,使其可以同时考虑两个方向上的序列关系,结合两个方向上的数据变化可以更好地判断其所处的层位。基于 BiLSTM 的层位预测模型结构如图 4 所示。

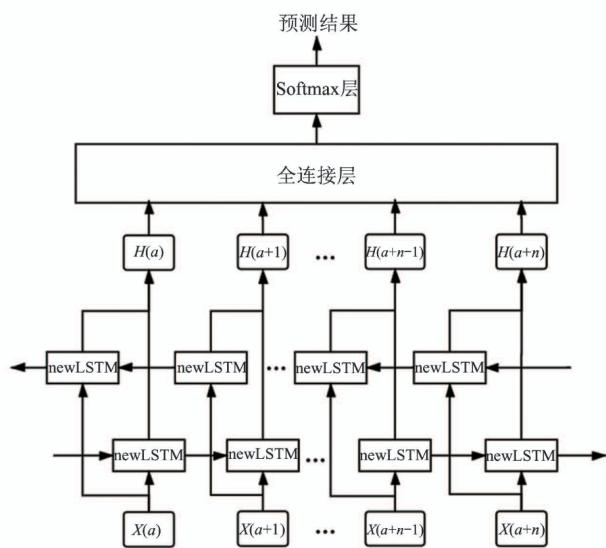


图 4 基于 BiLSTM 的层位预测模型结构

图中 $X(a)$ 、 $X(a+1)$ … $X(a+n-1)$ 、 $X(a+n)$ 是经过切片化处理的测井数据,newLSTM 则由多个 LSTM 计算单元组成。从图中可以看出 BiLSTM 其实由两个 LSTM 结构组成,但是这两个 LSTM 结构中数据的传播方向是相反的。在判断某一片段所处层位时,输出信息 H 包含了该片段前后的地质层位信息。最后再经过全连接层和 Softmax 层得到最终的预测结果。

2.3 损失函数与优化算法

训练模型时需要使用损失函数来判断识别结果是否与真实值贴近,本文选用交叉熵损失函数作为 BiLSTM 网络的损失函数,该函数适用于多分类预测的问题,其计算方法如式(2)所示。

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m y^i \ln(h_\theta(x^i)) + (1 - y^i) \ln(1 - h_\theta(x^i)) \quad (2)$$

式中, m 表示样本大小, x^i 表示第 i 组输入值, y^i 表示第 i 组输入值对应的实际值, h_θ 表示模型中的权值参数, $h_\theta(x^i)$ 表示模型的输出值。

除此之外,还需要使用优化算法来调整网络的权值参数,使得训练过程中损失函数可以不断减小,使模型越来越准确。自适应矩估计 (adaptive moment estimation, Adam) 算法是一种常用的优化算法,适用于大数据和高维度的空间,对内存的需求较小,并且能够根据损失函数动态地更新网络学习速率,因此本文采用 Adam 算法作为优化算法。

2.4 错层消除

错层消除是指对模型的预测结果进行一定的修正。在用训练好的模型进行结果预测时,难免会出现一些错误的预测。而大多数情况下,无法判断预测的结果是否正确,也就无法进行进一步的修正。但在某些特定的情况下,可以明显地发现部分预测结果与事实不符,也就可以进行一定程度的修正。为此,本文通过对地质分层结构的分析,提出了一种错层消除方法,可以进一步提高层位预测的准确率。

经分析发现,地质层位的分布较为规律,某种地质层只会出现在一个区域中,随着深度增加,出现过的地质层不会在后面再次出现。且每种地质层覆盖的范围较大,如果识别出的部分地质层片段长度较小,就很有可能是错误识别。在预测结果中有时会出现个别较小片段的层位与其周围片段的层位不同,这些片段的预测结果明显与事实不符。错层消除的目的就是将这些错层片段与其周围的片段归为同一类。如图 5 所示。

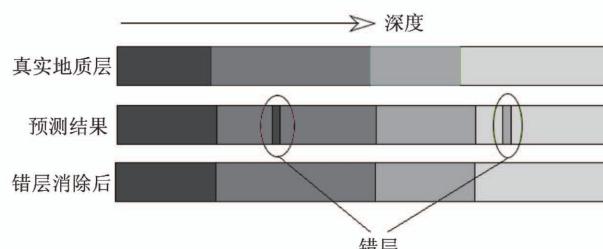


图 5 错层消除

图中灰度不同的部分代表不同的地质层,错层消除步骤如下所述。

(1) 用训练好的模型得到预测结果。预测结果用一个数组 H 表示, $H(i)$ 表示第 i 个采样点的层位序号。

(2) 划分地质层。遍历数组 H , 将连续相同层位的采样点归为一个地质层 $D(i) = [Len, Flag]$, 其中 i 表示该地质层为预测结果中的第 i 个地质层, Len 表示该地质层的长度, $Flag$ 表示该地质层所属的层位序号。

(3) 在地质层中找到疑似错层。遍历地质层 D , 若某一地质层的长度 $D(i).Len$ 小于某固定值 L , 则认为该地质层为疑似错层, L 的值根据具体的实验结果调整。

(4) 确定错层并消除。根据找到的疑似错层 $D(i)$, 若 $D(i-1).Len > D(i).Len$ 且 $D(i+1).Len > D(i).Len$ 且 $D(i-1).Flag = D(i+1).Flag$, 则可以确定该地质层为错层。将 $D(i-1)、D(i)、D(i+1)$ 3 个地质层合并为一个地质层, 新的地质层的长度为 3 个地质层的长度和, 层位序号为 $D(i+1)$ 的层位序号。

(5) 重复步骤(3)和(4)使地质层中不再有错层。

3 实验结果及分析

3.1 样本数据集构造

实验采用的数据是某油田 R3 区块 W1924 ~ W2125 共 125 口井的测井属性数据和层位标定数据。这部分井属于同一区块, 在地质结构上比较接近, 每口井在纵向上的地质层位分布都比较规律, 且层位厚度较大。数据集总共涉及到 17 种不同的地质层, 实验时需要为这些地质层标上不同的标签, 如表 2 所示。

本文实验数据共可划分为 4476 个分段数据, 在保证训练集和测试集数据属于不同的井的情况下, 按照近似 10:1 的比例划分训练集与测试集, 然后对其进行略微调整, 保证训练集和测试集都能包含某口井中完整的数据。最终训练集共有 4059 个分段数据, 测试集有 417 个分段数据, 训练周期为 2000。

表 2 层位类别

层位名	标签号
K1z4	1
K1z3	2
K1z2 + 1	3
J2a	4
J2z	5
J1y	6
chang2	7
chang3	8
chang4 + 5	9
chang6	10
chang7	11
K1z5	12
Q	13
chang8	14
chang9	15
J1f	16
chang1	17

3.2 网络结构

模型的输入是 GR、SP、AC、AT90 属性以及深度信息构造成的片段, 片段长度为 30, 隐藏层节点数设为 128, 即有 128 个 newLSTM 单元。最后一层全连接层输出的是特征值维度为 17 的向量, 对应层位的类别数。

3.3 层位预测实验

本文采用准确率和层位平均偏移距离指标来评价层位预测的结果好坏。准确率计算公式为

$$precision = \frac{T(u)}{F(u) + T(u)} \quad (3)$$

式中, $F(u)$ 表示模型预测错误的片段数量, $T(u)$ 表示模型预测正确的片段数量。准确率指标反映了预测结果与真实结果的差异性, 数值越高表示预测结果与真实结果越接近, 模型的预测能力越强。

平均偏移距离 (average offset distance, AOD) 表示每口井每个层位的预测分界点与真实层位分界点偏差值平均值, 计算公式为

$$AOD = \frac{\sum_{i=1}^n (|\hat{d}_i - d_i|)}{n} \quad (4)$$

式中, n 表示地层分界点的数量, d_i 表示真实的层位分界点深度, \hat{d}_i 表示预测的层位分界点深度。平

均偏移距离反映了预测的地层结果与真实地形的吻合程度,数值约小,吻合程度越好。

3.3.1 实验 1

对于相同的分段长度,分段中采样点的采样间隔大小会影响最终的预测结果。为了选取合适的采样点间隔,本文进行了 12 组对比实验验证分段划分效果,采样点间隔从 0.125 m 到 1.5 m 做了稠密采样,实验结果如表 3 所示。

表 3 网络层位预测实验结果

预测结果 间隔/m	准确率	平均偏移距离/m
0.125	0.8472	13.59
0.250	0.8410	13.125
0.375	0.8849	13.125
0.500	0.8729	11.25
0.625	0.8895	11.71
0.750	0.9084	10.31
0.875	0.9087	9.84
1.000	0.9303	8.75
1.125	0.9261	9.84
1.250	0.9308	10.98
1.375	0.9161	10.31
1.500	0.9231	11.25

从表格中可以看出,采样点间隔为 1 m 左右时,预测结果的准确率最好,平均偏移距离相对较小,当采样间隔再变大时,准确率的变化不太明显,但平均偏移距离会变大,增加层位偏离误差。所以采样间隔为 1 m 时最合适。其中一口井的预测结果如图 6 所示。

图 6 中,左侧第 1 列为真实的地质分层,左侧第 3 列为模型的预测结果,左侧第 2 列为错层消除后的结果,即最终的结果。从图中可以看出,最终结果的层位分布与真实的地质层位分布相似,只有边界处会出现部分误差,而错层消除也能成功地将部分错误的地质分层重新划分为正确的分层。

3.3.2 实验 2

为了验证本文方法相对于其他方法的优势,本文对比了 CNN、LSTM 和 BiLSTM 在实验数据集上的表现,实验结果如表 4 所示。

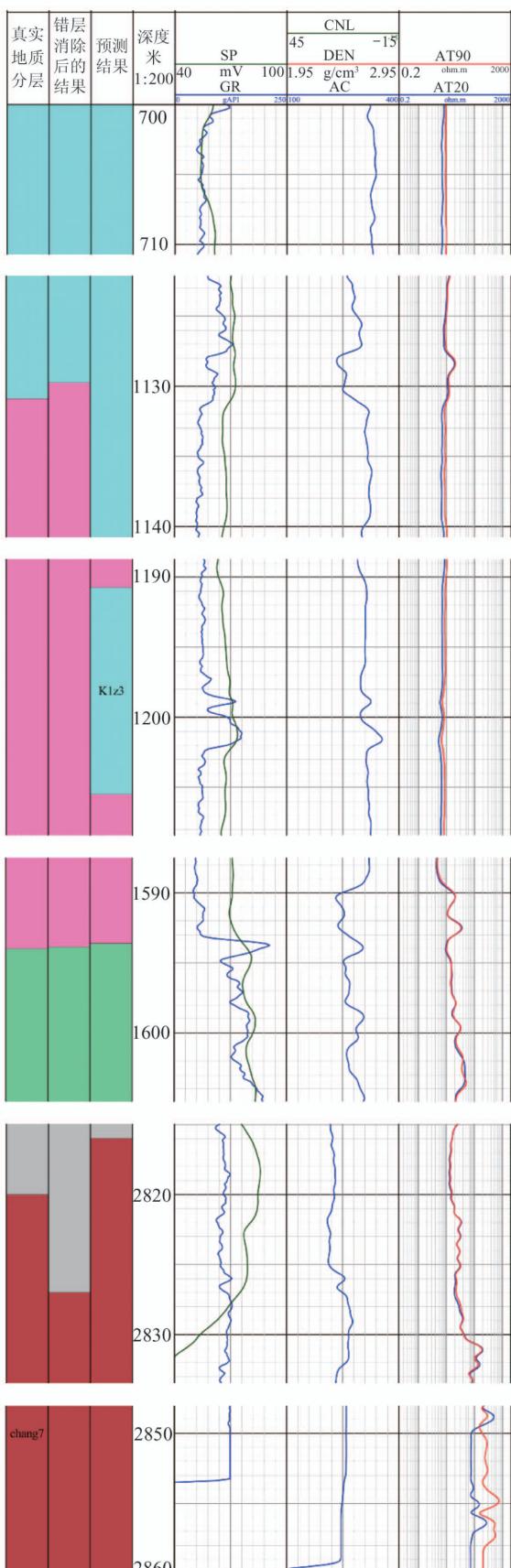


图 6 预测结果图

从表格中可以看出,本文提出的基于 BiLSTM 网络模型的方法在准确率以及平均偏移距离上都要优于 CNN 和 LSTM 网络模型。实验结果证明本文方法对地质层识别有较高的准确率,但是在分界处的识别仍有较大的误差。

表 4 不同网络模型对比实验

模型预测结果 网络	CNN	LSTM	BiLSTM
准确率	0.89	0.93	0.94
平均偏移距离/m	9.3	8.75	6.25

4 结 论

本文根据测井数据的特点,提出了一种基于 BiLSTM 的地质层位预测方法。该方法首先对测井数据进行分段处理,然后再基于 BiLSTM 建立层位预测模型,并将模型中的部分 LSTM 计算单元合并,使其按照切好的片段进行训练,最后再对预测结果进行错层消除,使结果更加贴合实际。实验结果表明,本文方法可以根据测井数据有效识别地质层。

然而本文的方法仍存在一定的局限性,如果某一片段包含了层位分界,则该片段的识别准确率会受到很大的影响。由于地质层位的分布较为规律,本文的方法分出的片段中只有少数会包含层位分界,对整体识别的准确度影响较少,但是却会影响对层位边界的识别。如果用于储层识别,由于其层位分界较多,并不能直接使用。在未来的研究中,需要进一步加强对层位边界的识别,通过对片段进行特征提取来判断该片段是否包含了层位边界。只要可以准确地识别出边界的位置,分段时尽量避开边界,再结合本文的方法就可以实现储层的预测。

参考文献

- [1] 郑庆伟,王祝文,欧希阳,等.聚类分析方法在测井曲线重构中的应用[J].世界地质,2015,34(3):807-812
- [2] 刘昊,朱丹丹,陈冬,等.基于聚类算法的岩性预分类方法研究[C]//2018IPPTC 国际石油石化技术会议论文集,西安,2018:10
- [3] 初颖,吕堂红.基于极值法和聚类分析法的测井曲线自动分层模型——以山东省胜利油井为例[J].长春理工大学学报(自然科学版),2017,40(6):105-110
- [4] Zhang Q H, Shen K L. The research of stratigraphic division method based on support vector machine[C]//International Conference on Frontier of Energy and Environment Engineering, Taiwan, China, 2014: 531-534
- [5] 牟丹,王祝文,黄玉龙,等.基于 SVM 测井数据的火山岩岩性识别——以辽河盆地东部坳陷为例[J].地球物理学报,2015,58(5):1785-1793
- [6] 石锁,余继峰,曹慧涛,等.基于高斯核 SVM 的储层岩性识别——以东濮凹陷上古生界碎屑岩为例[J].中国科技论文,2020,15(1):112-118,136
- [7] 张俊儒. SVM 和 ANN 融合算法在测井岩性识别中的应用研究[D].武汉:武汉理工大学计算机科学与技术学院,2016
- [8] 曹志民,王振涛,韩建,等.基于多特征联合的 BP 神经网络岩性识别[J].化工自动化及仪表,2018,45(5):364-367,395
- [9] 魏杰,杨斌,刘锋,张智南.基于岩性识别的 BP 神经网络孔隙度预测[J].石油化工应用,2020,39(3):105-110
- [10] 倪小威,刘迪仁,艾林,等.基于 BP 神经网络算法的测井资料煤层气储层产水预测[J].煤炭技术,2017,36(9):105-107
- [11] 魏杰,杨斌,刘锋,等.基于岩性识别的 BP 神经网络孔隙度预测[J].石油化工应用,2020,39(3):105-110
- [12] 陈钢花,梁莎莎,王军,等.卷积神经网络在岩性识别中的应用[J].测井技术,2019,43(2):129-134
- [13] 段友祥,李根田,孙歧峰.卷积神经网络在储层预测中的应用研究[J].通信学报,2016,37(S1):1-9
- [14] 张国印,王志章,林承焰,等.基于小波变换和卷积神经网络的地震储层预测方法及应用[J].中国石油大学学报(自然科学版),2020,44(4):83-93

Geological segment horizon prediction method based on BiLSTM

Jin Xiangchen * ** , Wu Zirui * , Shi Min * , Zhu Dengming ** , Zhou Jun ***

(* College of Control and Computer Engineering, North China Electric Power University, Beijing 102206)

(** Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

(*** China Petroleum Logging Co. Ltd. , Xi'an, 710065)

Abstract

Geological stratification aims to divide the rock layers in the stratum section of a certain area, which is of great significant to the problem of geological prospecting. However, considering that there exists a variety of geological horizons, and traditional geological stratification relies heavily on subjective judgment of the expert, performing geological stratification is time-consuming and knowledge-intensive work. Existing automatic geological stratification methods fail to consider the sequence relationship of well logging data and the characteristics of geological horizon distribution, thus making classification accuracy unable to reach the state-of-the-art level. Based on the above background, an improved bidirectional long-short memory neural network (BiLSTM) for geological horizon prediction is proposed. To fully make use of the characteristics of fragmented distribution of geological horizons and the sequence correlation in two directions of well logging data, the BiLSTM network following the structure of the segmented data is modified. The proposed method on a real well data set of an oil field is evaluated, and the experimental results demonstrate that the classification accuracy has reached 93%, which has a significant improvement compared with other methods.

Key words: bidirectional long-short memory neural network (BiLSTM), logging curve, geological stratification, horizon prediction