

基于弱监督学习的大地电磁噪声检测和分类

张培林, 雷万里

桂林理工大学地球科学学院, 广西 桂林

收稿日期: 2025年7月16日; 录用日期: 2025年8月5日; 发布日期: 2025年8月18日

摘要

在大地电磁法中, 提高数据质量一直是大地电磁信号采集与处理的核心问题。本文提出使用Anomaly Transformer异常检测模型对大地电磁数据进行噪声识别和CNN-BiLSTM噪声分类的弱监督学习联合算法, 首先, 将处理好的数据集放入Anomaly Transformer异常检测模型训练, 采用两个并行的分支计算, 在先验关联分支输出为先验关联矩阵, 在序列关联分支输出序列关联矩阵, 之后用KL散度计算关联差异, 对测试集数据进行无监督异常点识别并标记, 然后将标记好的测试集数据切分好放入CNN-BiLSTM模型进行训练, 利用CNN提取信号的时域和频域特征, 并将其融合, 最后利用BiLSTM网络学习信号的序列特征, 完成噪声分类任务。实验结果表明: 在噪声检测部分, 与LSTM对比Anomaly Transformer无监督异常检测模型在噪声识别方面表现出明显的优势, 在噪声分类部分, CNN-BiLSTM与CNN、LSTM、BiLSTM、CNN-LSTM对比该模型仍具有卓越的噪声分类诊断性能、出色的泛化能力和快速的计算速度。这一结果充分验证了该模型在弱监督大地电磁噪声识别和分类任务上的有效性, 为后续信号去噪、反演和地质解释提供了有力保障和重要的实际应用价值。

关键词

大地电磁法, 异常检测, 噪声分类, 弱监督学习

Magnetotelluric Noise Detection and Classification Based on Weakly Supervised Learning

Peilin Zhang, Wanli Lei

College of Earth Sciences, Guilin University of Technology, Guilin Guangxi

Received: Jul. 16th, 2025; accepted: Aug. 5th, 2025; published: Aug. 18th, 2025

Abstract

In magnetotelluric (MT) methods, improving data quality has consistently been a central challenge

文章引用: 张培林, 雷万里. 基于弱监督学习的大地电磁噪声检测和分类[J]. 地球科学前沿, 2025, 15(8): 1172-1186.
DOI: 10.12677/ag.2025.158109

in signal acquisition and processing. This paper proposes a joint algorithm based on weakly supervised learning, integrating the Anomaly Transformer for noise detection and a CNN-BiLSTM network for noise classification in MT data. First, the preprocessed dataset is input into the Anomaly Transformer model for training. The model employs two parallel branches: the prior association branch outputs a prior association matrix, while the series association branch produces a sequence association matrix. The Kullback-Leibler (KL) divergence is then used to compute the difference between the two, enabling unsupervised identification and labeling of anomalies in the test dataset. The labeled test data are then segmented and fed into the CNN-BiLSTM model. The CNN is responsible for extracting and fusing time-domain and frequency-domain features of the signal, while the BiLSTM captures the sequential characteristics to complete the noise classification task. Experimental results show that the Anomaly Transformer significantly outperforms traditional LSTM models in noise detection. In the classification phase, the CNN-BiLSTM model also demonstrates superior performance, better generalization, and faster computation compared with CNN, LSTM, BiLSTM, and CNN-LSTM models. These results fully validate the effectiveness of the proposed model for weakly supervised MT noise detection and classification, offering strong support and practical value for subsequent denoising, inversion, and geological interpretation.

Keywords

Magnetotelluric, Anomaly Detection, Noise Classification, Weakly Supervised Learning

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

大地电磁测深法因其施工轻便、高阻层屏蔽影响小、勘探范围广、探测深度大等特点,广泛应用于地球深部结构探测、工程勘察和矿产普查等领域[1][2]。其施工流程主要包括数据采集、资料处理和反演解释,而数据质量是影响该方法实际应用效果的关键因素。提升数据质量一直是大地电磁信号采集与处理的核心问题。因此,准确高效地进行噪声的异常检测、识别与分类,对后续的信号去噪、反演以及地质解释具有重要意义。

大地电磁法以天然场为场源,天然电磁场具有能量弱、幅度变化大、频带范围宽等特点,野外实测信号常常容易受到各类噪声的干扰[3][4]。其噪声类型主要分为四类:场源噪声、地质噪声、随机噪声和人文噪声。其中,人文干扰噪声是 MT 测量中最主要的噪声来源,主要包括阶跃波噪声、脉冲噪声、方波噪声、充放电三角波噪声及工频干扰噪声等[5]。在复杂环境下的人文电磁噪声中,有用信号与噪声在时间域上通常表现出一定的特征,通过对噪声类型的定量分析及特征规律的定性研究,可为大地电磁数据的有效识别与分类提供重要依据。

近年来,Robust 估计法[6][7]、远参考道估计法[8][9]、小波变换[10]、VMD [11]、PCA [12] 2 机器学习[13][14]等信号处理方法均被引入到 MT 信号噪声压制领域,Robust 估计法通过残差驱动的加权机制、远参考道估计法通过“空间参考”、小波变换和 VMD 通过频带分解和能量分布、PCA 通过空间投影与能量分布来识别噪声。在机器学习方面,比如李博[15]提出使用 VMD-LSTM 异常检测模型构建无噪声信号数据集进行训练识别噪声,史维[16]等提出将模糊熵和短期能量作为模糊 C-means (FCM)聚类算法的特征参数来识别噪声,但是都没有对噪声进行分类。

上述方法大多聚焦于异常信号的检测或去除,尚缺乏系统性地对噪声类型进行分类的研究。噪声类

型的进一步识别与分类，不仅有助于提高去噪策略的选择性，也对地质解释具有重要意义。鉴于此，笔者尝试使用弱监督学习对大地电磁异常数据进行识别和分类，以期更好地改善数据质量，产生实际应用价值。

2. 理论基础

2.1. Transformer

Transformer 模型结构如图 1 所示，编码器和解码器是模型的基础构件，编码器由多个相同层堆叠构建，主要包含两个子层：一是自注意力层，用于捕捉序列中各单元间的复杂依赖关系，二是前馈神经网络层，用于增强模型的表达能力。解码器结构与编码器相似：但额外包含一个“掩蔽自注意力”层，该层通过设置掩蔽矩阵，确保在计算注意力权重时，模型仅关注当前输出位置之前的序列信息，从而保证生成的顺序性和自回归特性，解码器中的第 2 个自注意力机制层则负责整合来自编码器的上下文信息，以生成更精准的输出[17]。

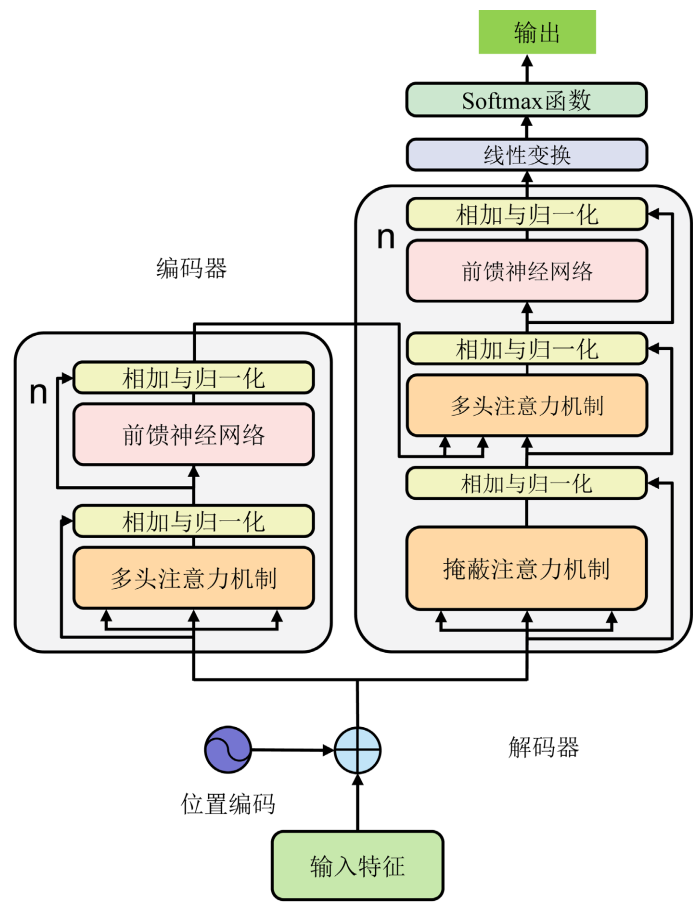


Figure 1. Transformer model architecture
图 1. Transformer 模型结构

Transformer 的核心在于自注意力机制，使得模型在处理序列中的某个位置时，能够动态地关注序列中其他所有位置的信息，从而捕捉长距离依赖关系。自注意力机制的计算如下所示：

$$Q = XW^Q, K = XW^K, V = XW^V \tag{1}$$

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

式中: $\frac{1}{\sqrt{d_k}}$ 是缩放因子, 作用是防止点积值过大。

为了解决单头自注意力机制在捕捉不同空间的信息存在局限性问题, Transformer 引入了多头自注意力机制。多头注意力是由多个并行运算的注意力层组成, 在不同位置能够同时关注来自不同子空间的信息。其计算公式如下所示:

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (3)$$

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O \quad (4)$$

式中: W^O 是输出层的投影矩阵。

本文用的噪声异常检测模型如图 2 所示, 主要由输入嵌入层、异常注意力、前馈神经网络与层归一化等几个模块组合在一起, 使用堆叠式网络结构, 核心思想是: 用两种注意力(先验和系列)建立关联并比较差异, 借助“关联差异”来识别异常点[18]。

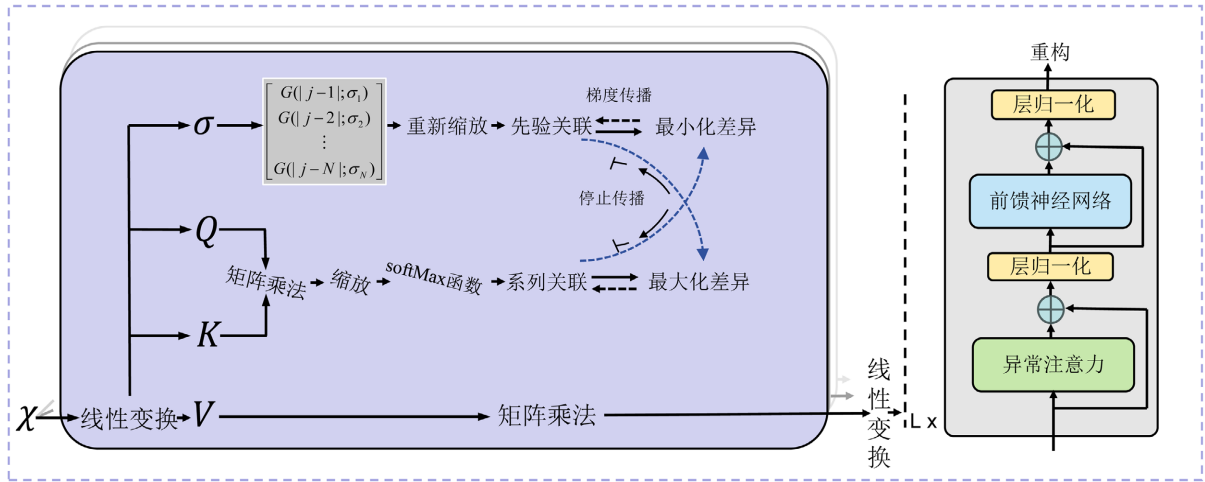


Figure 2. Anomaly transformer time series anomaly detection model

图 2. Anomaly Transformer 时间序列异常检测模型

异常注意力机制是基于 Transformer 中的自注意力机制改进而来的, 并采用一个双分支的结构来对先验关联、序列关联进行建模。计算公式如下:

假设有 L 层, 一层模型计算步骤为:

$$Z^l = \text{Layer-Norm}\left(\text{Anomaly-Attention}\left(X^{l-1}\right) + X^{l-1}\right) \quad (5)$$

$$X^l = \text{Layer-Norm}\left(\text{Feed-Forward}\left(Z^l\right) + Z^l\right) \quad (6)$$

其中, 在异常注意力机制模块中, 首先, 对上一层输出 X 的进行线性变换, 生成 Q 、 K 、 V 以及可学习的尺度参数(σ), 先验关联分支利用可学习的尺度参数 σ 与高斯核函数构造先验分布, 最后输出为先验关联矩阵, 序列关联分支直接采用标准自注意力机制计算全局关联, 输出包含全局的序列关联信息。

$$Q, K, V, \sigma = X^{l-1}W_Q^l, X^{l-1}W_K^l, X^{l-1}W_V^l, X^{l-1}W_\sigma^l \quad (7)$$

$$P^l = \text{Rescale} \left(\left[\frac{1}{\sqrt{2\pi}\sigma_i} \exp \left(-\frac{|j-i|^2}{2\sigma_i^2} \right) \right]_{i,j \in \{1, \dots, N\}} \right) \quad (8)$$

$$S^l = \text{Softmax} \left(\frac{QK^T}{\sqrt{d}} \right) \quad (9)$$

$$\hat{Z}^l = S^l V \quad (10)$$

式中, P^l 表示先验关联矩阵; S^l 表示序列关联矩阵; $\hat{Z}^l$ 表示当前层的重构矩阵。

随后, 通过 KL 散度计算关联差异, 对序列关联输出与先验关联输出进行比较, 以捕捉二者之间的差异, 从而实现无监督异常点的识别。优化过程中引入“差异损失”以引导序列关联分支寻找最具信息量的关联关系: 在先验关联中构建最小化差异的目标, 而在序列关联中构建最大化差异的目标。即在最小化阶段, 使先验关联逼近序列关联; 在最大化阶段, 优化序列关联以扩大其与先验关联的差异, 从而使其更加关注非相邻的时间依赖关系, 使异常点更难形成远程关联。为实现该机制, 模型引入了停止梯度策略来控制不同阶段中各分支参数的更新: 最小化阶段仅更新先验关联分支的参数, 最大化阶段则仅更新序列关联分支的参数。具体计算公式如下:

$$\text{AssDis}(P, S; X) = \left[\frac{1}{L} \sum \left(KL(P_{i,:}^l \| S_{i,:}^l) + KL(S_{i,:}^l \| P_{i,:}^l) \right) \right]_{i=1, \dots, N} \quad (11)$$

$$L_{\text{Total}}(\hat{X}, P, S, \lambda, X) = \|X - \hat{X}\|_F^2 - \lambda \times \|\text{AssDis}(P, S, X)\|_1 \quad (12)$$

$$L_{\text{Total}}(\hat{X}, P, S_{\text{detach}}, -\lambda, X) \quad (13)$$

$$L_{\text{Total}}(\hat{X}, P_{\text{detach}}, S, \lambda, X) \quad (14)$$

式中, $\text{AssDis}(P, S; X)$ 表示多个层的先验关联 P 和序列关联 S 的逐点关联差异; $KL(\bullet \| \bullet)$ 表示 P^l 和 S^l 每一行的两个离散分布之间的 KL 散度; $\|\bullet\|_F$ 与 $\|\bullet\|_1$ 分别为 F 范数与 L_1 范数; λ 为平衡系数。

在重构标准中, 添加了正则化的关联差异, 这样可以同时利用时间表征和可区分的关联差异的优势。得出的评判标准如下:

$$\text{AnomalyScore}(X) = \text{Softmax}(-\text{AssDis}(P, S; X)) \odot \left[\|X_{i,:} - \hat{X}_{i,:}\|_2^2 \right]_{i=1, \dots, N} \quad (15)$$

式中, $\odot$ 表示元素乘法。

关联差异越小, 异常分数越大; 重构误差越大, 异常分数越大。为了更好地重构, 异常点通常会减小关联差异, 导致异常分数增大。这样的设计可以让重构误差和关联差异协同作用, 提高检测性能。

2.2. CNN-BiLSTM 模型

2.2.1. CNN 网络

CNN 模型结构如图 3 所示, CNN 的基本结构包括卷积层、池化层和全连接层。其中, 卷积层作为最关键的环节, 主要以卷积核对输入的多特征复杂数据进行卷积运算, 从而提取数据潜在特征; 池化层作用为降采样输入数据, 降低计算量, 同时保留重要特征; 全连接层将卷积层和池化层的输出进行拼接, 最终输出分类结果[19]。

CNN 虽然具有从大数据中自动提取多维度空间特征的能力, 但对于时间依赖性较强的时序数据处理能力较差, 相比之下 BiLSTM 基于 LSTM 的门控机制, 通过遗忘门、输入门和输出门的协同作用, 有效

地解决长期依赖性问题。通过将二者结合, 就能够增强空间、时序特征的提取能力并相对减少计算时间。

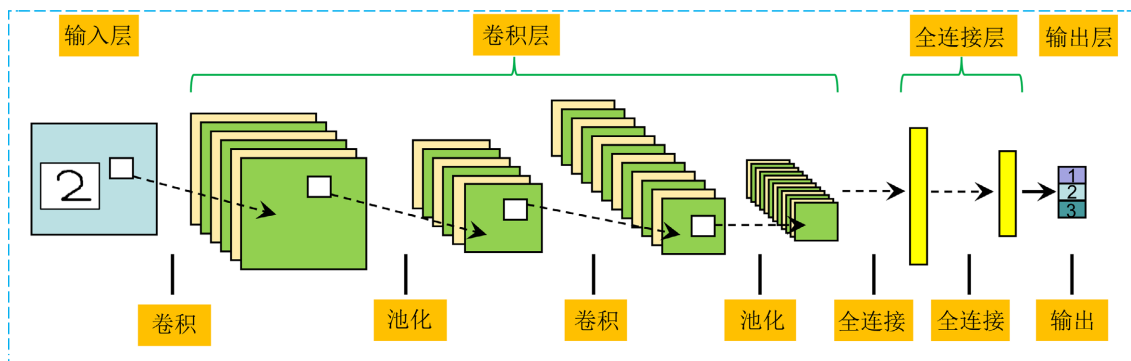


Figure 3. CNN model architecture

图 3. CNN 模型结构

2.2.2. BiLSTM 网络

BiLSTM 是由前向 LSTM 和反向 LSTM 组成的, 网络结构如图 4 所示, 正向 LSTM 网络负责学习从序列开始到结束的前向信息; 反向 LSTM 网络负责学习从序列结束到开始的后向信息。这样的 BiLSTM 不仅保留了 LSTM 对长依赖问题的有效处理能力, 还能充分利用数据序列的整体上下文信息, 提高了模型预测的准确性与鲁棒性[20]。

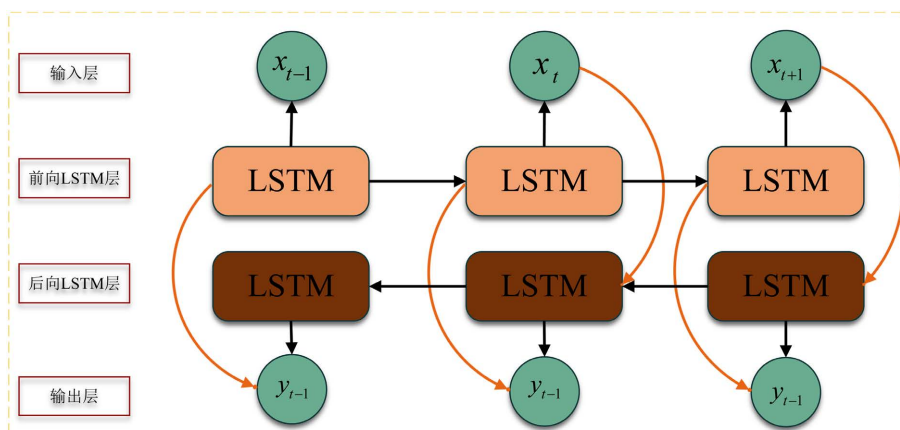


Figure 4. BiLSTM model architecture

图 4. BiLSTM 模型结构

大地电磁信号具有数据量大、时序长的特点, 传统神经网络难以有效建模其长期依赖关系, 且易出现梯度爆炸问题。BiLSTM 既能规避梯度消失或爆炸问题, 又能稳定捕获长时序依赖特征。其双向结构可同步提取序列的前后向信息, 充分利用全局上下文特征, 通过选择性遗忘噪声干扰、保留有效信号特征, 实现对时序数据的精准建模。与 Transformer 类似, BiLSTM 支持端到端训练, 能自动学习时序特征模式, 这些特性使其特别适用于大地电磁强噪声识别和分类任务。

3. 基于 Anomaly Transformer 噪声识别和 CNN-BiLSTM 噪声分类

为了对大地电磁噪声数据进行检测和分类, 本文提出了 Anomaly Transformer 噪声识别和 CNN-BiLSTM 噪声分类的弱监督学习联合算法, 如图 5 所示。

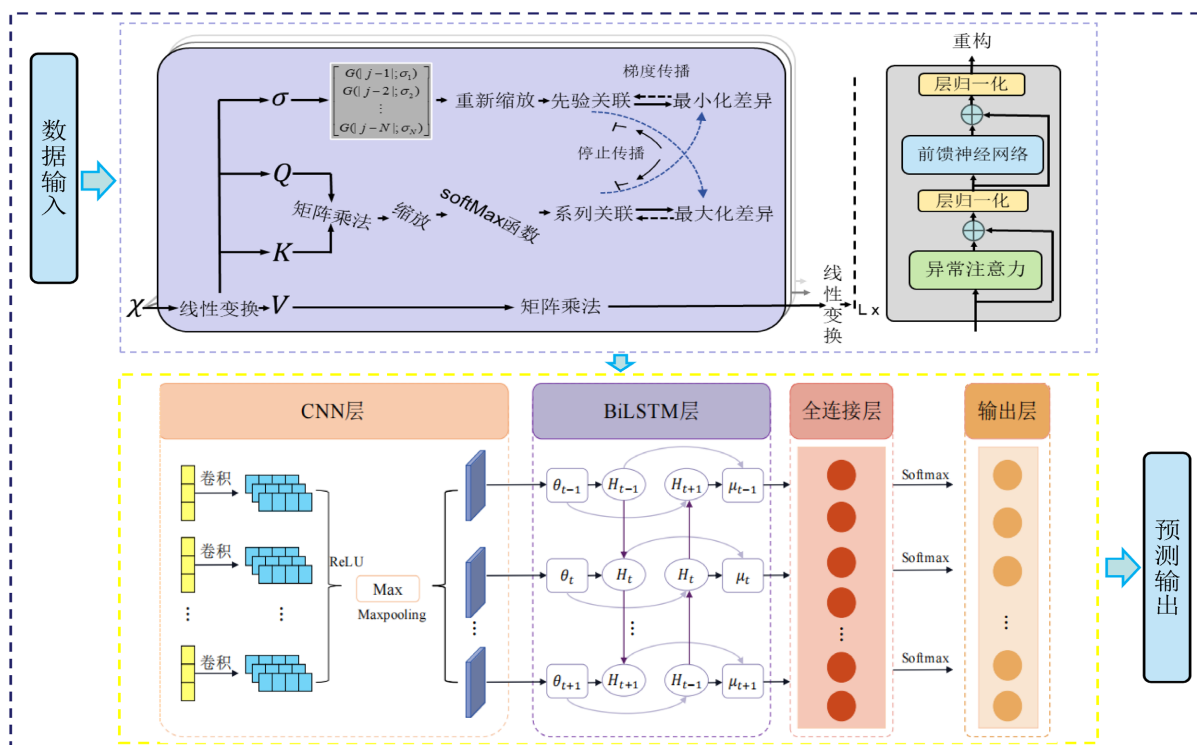


Figure 5. Flowchart of the Anomaly Transformer-based anomaly detection and CNN-BiLSTM-based anomaly classification model

图 5. Anomaly Transformer 异常检测 CNN-BiLSTM 异常分类模型流程图

该模型实现对噪声进行识别和分类的具体步骤如下：

(1) 构建大地电磁时间序列异常检测数据集，异常检测模型使用正常的大地电磁信号数据进行训练，经过大量正常信号数据的学习，在遇到强干扰噪声数据时，能分辨二者的不同，从而识别出异常点数据。验证集和测试集则无需考虑数据中是否含有强噪声干扰，测试集需要单独做一个测试标签，方便检测到异常点时进行标记；

(2) 将数据处理成均值为 0、标准差为 1 的标准正态分布，消除各特征之间因为量纲不同造成的差异，利用滑动窗口对数据进行切分，允许数据在窗口之间有重叠，有利于异常检测模型训练时学习局部和全局的时序关联；

(3) 将切分后的数据输入异常检测模型，模型通过双分支网络分别计算先验关联矩阵和序列关联矩阵。利用两矩阵间的对称 KL 散度量关联差异，并结合输入与重构数据的误差，综合计算异常分数。最终，基于异常分数对测试集实现无监督异常点识别与标注；

(4) 将标记好的测试集数据切分好放入 CNN-BiLSTM 模型进行训练，利用 CNN 提取时频域特征，并将其融合，利用 BiLSTM 网络学习序列特征，完成噪声分类任务。

4. 预测实验和分析

4.1. 实验数据

本文的实验数据来源于陕西省宁陕县丰富镇东沟矿区实测的大地电磁数据，选用测点 A1008AGH 进行数据处理和分析，如图 6 所示。测点采样率为 15，约 88 万个数据点，该测点的电场通道数据前面数据受噪声影响小，只有后面受到噪声影响，而磁场通道则频繁受到大量尖脉冲噪声干扰。

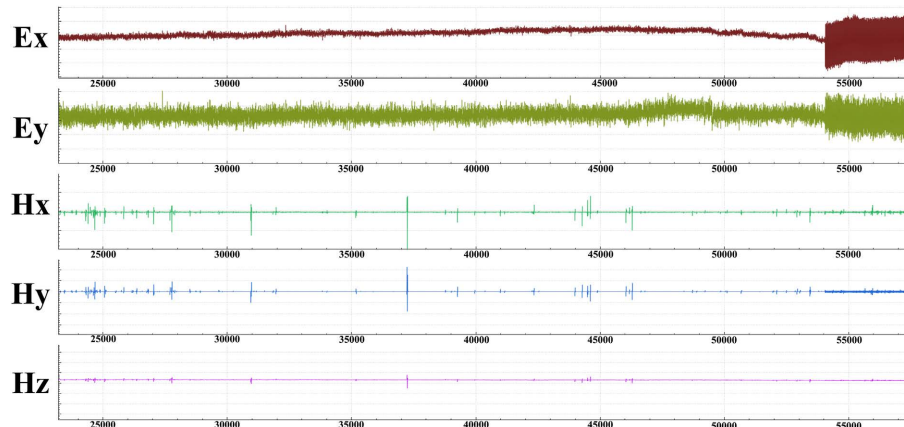


Figure 6. Time series of measured magnetotelluric signals with noise

图 6. 实测含噪大地电磁信号时序图

本文选用 Ex 通道未受到明显噪声干扰的数据, 作为无噪声数据, 在此基础上叠加人工合成噪声信号来模拟大地电磁在实测环境中遇到的常见强干扰噪声, 由于该测点 Ex 分量已涵盖工频噪声, 只需模拟方波噪声、阶跃噪声, 脉冲噪声, 三角波噪声, 如图 7 所示。

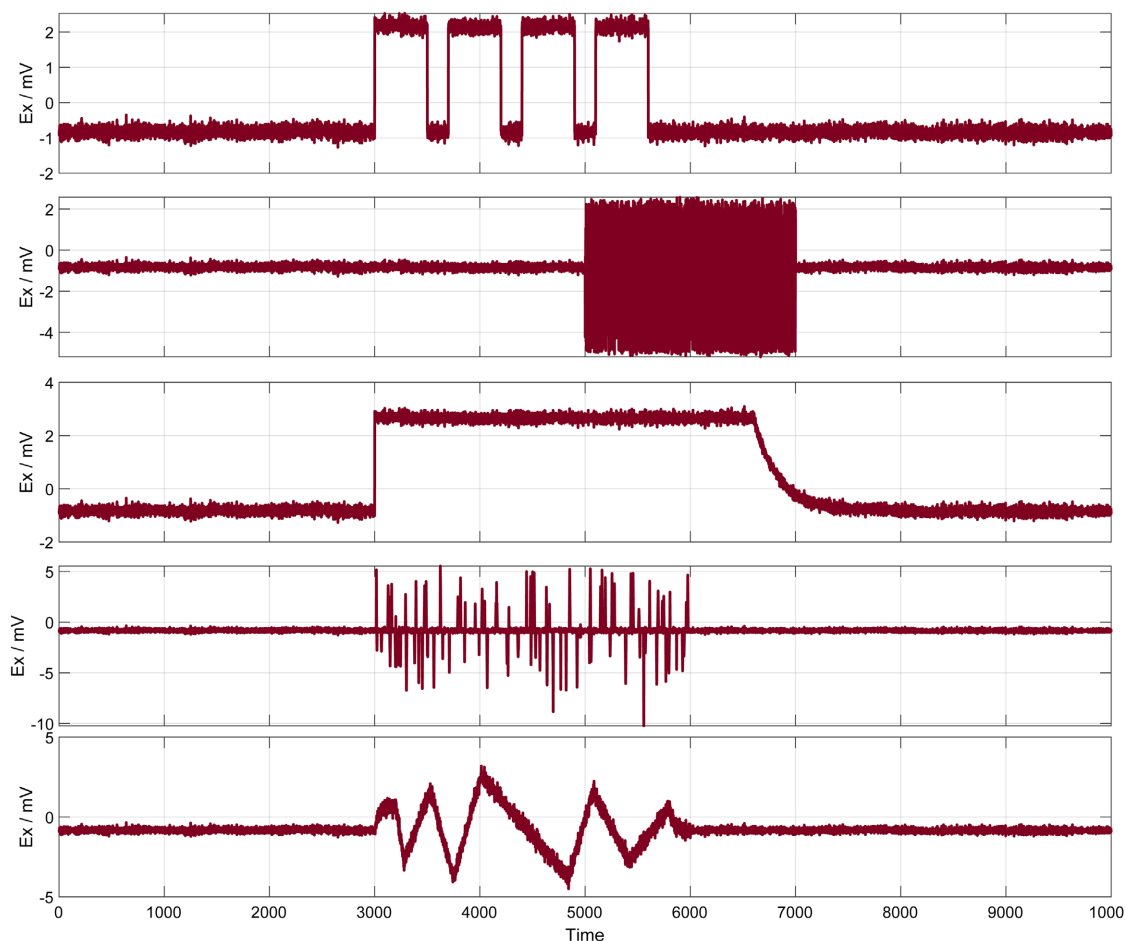


Figure 7. Time series of noisy magnetotelluric signals

图 7. 含噪大地电磁信号时序图

方波噪声在时间序列上表现为一个突然的抬升或下降的信号, 往往伴随着周期性变化; 工频噪声主要指的是由供电电源的频率(通常是 50 Hz)及其谐波成分引起的噪声; 和方波噪声一样, 阶跃波噪声在时间序列也表现为突然的抬升或下降, 往往长时间后才回到正常信号的幅值水平, 并且没有周期性变化; 脉冲噪声的特点是非连续的, 由持续时间短和幅度大的不规则脉冲或噪声尖峰组成, 振幅远远大于纯净电磁信号[21]。这几种噪声同时出现在电道和磁道, 一般来说, 对电道影响比对磁道影响大。三角波噪声在时间序列上呈现为非正弦曲线的锯齿波形, 幅值较大, 往往以正负成对的形式出现, 一般来说, 对磁道影响比对电道影响大。

4.2. 结果与分析

在异常检测任务阶段, 流程如图 8 所示, 本文选取测点 A1008AGH 中 Ex 通道的 50w 正常大地电磁信号数据作为异常检测模型的训练集, 并采用 1w 模拟噪声信号数据构建测试集。异常检测模型基于 Anomaly Transformer 构建, 通过双分支注意力机制分别建模先验关联与序列关联, 并计算其关联差异用于识别潜在异常点。模型输出的异常分数结合重构误差和 KL 散度共同反映每个时间点的异常程度, 随后通过设定阈值及滑动窗口策略, 将连续异常分数转化为离散标注样本, 最终为后续的 CNN-BiLSTM 分类网络提供可靠的训练数据基础。

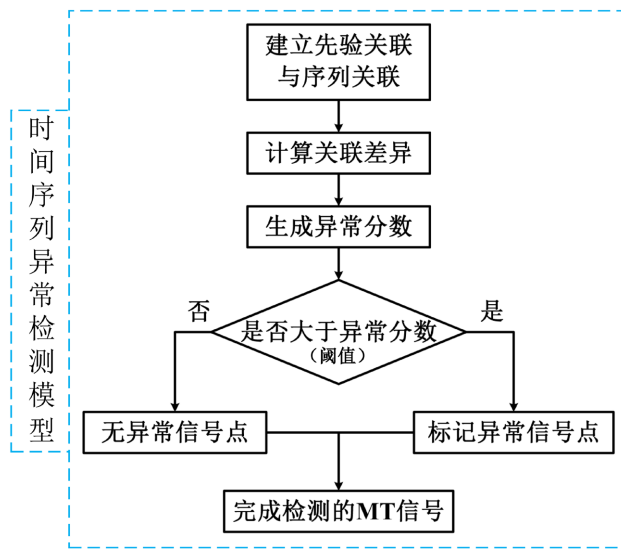


Figure 8. Flowchart of time series anomaly detection
图 8. 时间序列异常检测流程图

该阶段的检测性能评估采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F1 分数(F-score)作为评价指标, 为验证效果, 实验将基于 LSTM 的异常检测模型与 Anomaly Transformer 异常检测模型进行了对比。在参数设置保持一致的前提下, 测试结果如表 1 和表 2 所示。

根据表 1 和表 2 的对比结果可知, Anomaly Transformer 异常检测模型在准确率、精确率、召回率和 F1 分数等各项指标上均优于 LSTM 模型, 整体性能表现更加优秀。该模型在各类噪声信号的检测中均展现出更稳定、全面的检测能力, 充分验证了其在无监督异常检测任务中的优越性和适用性。

在噪声分类阶段, 先对异常识别阶段的一共 6w 测试集拼接, 再进行滑动窗口为 100, 步长为 1 数据切分, 得到 $60,000 \times 100$ 的数据集, 为了验证模型的准确性, 标签数据先用真实数据标签, 其中 1 代表正常信号; 2 代表含方波噪声的信号; 3 代表含工频噪声的信号; 4 代表含阶跃波噪声的信号; 5 代表含

脉冲噪声的信号; 6 代表含三角波噪声的信号。笔者将 LSTM、BiLSTM、CNN、CNN-LSTM、CNN-BiLSTM 模型进行了对比实验, 按每 2000 数据集取 1800 进行训练, 剩下 200 进行预测分类, 实验结果如图 9 所示。

Table 1. Anomaly detection results using the Anomaly Transformer

表 1. Anomaly Transformer 异常检测结果

噪声信号	Accuracy	Precision	Recall	F1
方波噪声	0.9740	0.8874	0.9965	0.9388
工频噪声	0.9821	0.9202	0.9970	0.9570
阶跃噪声	0.9830	0.9680	0.9990	0.9833
脉冲噪声	0.9594	0.5881	0.8514	0.6957
三角波噪声	0.9755	0.9258	0.9983	0.9607

Table 2. Anomaly detection results using the LSTM

表 2. LSTM 异常检测结果

噪声信号	Accuracy	Precision	Recall	F1
方波噪声	0.9622	0.8410	1.0000	0.9137
工频噪声	0.9645	0.8803	0.9520	0.9147
阶跃噪声	0.8692	0.9666	0.7648	0.8540
脉冲噪声	0.8023	0.2097	0.9486	0.3434
三角波噪声	0.9348	0.9193	0.8580	0.8876

根据各个分类模型测试结果对比图可知 CNN-BiLSTM 模型在分类结果上的准确率最高, 可达 98.62%, 明显优于 LSTM、BiLSTM、CNN 及 CNN-LSTM 等对比模型, 可见该模型能解决噪声的多级分类的问题。

为进一步检验 CNN-BiLSTM 模型在噪声分类上的有效性及其泛化能力, 笔者尝试使用每种信号的 5w 测试集, 滑动窗口为 100, 步长为 1 对数据进行切分, 得到 $300,000 \times 100$ 的数据集再输入模型进行噪声分类, 每 2000 取 1800 数据集进行训练, 剩下 200 数据集进行预测分类, 实验结果如图 10 所示。

根据各个分类模型测试结果对比图可知 CNN-BiLSTM 模型在分类结果上的准确率最高, 可达 99.27%, 可见该模型的泛化能力强, 具备更强的特征提取与时间序列建模能力。也进一步验证了该方法在处理大规模 MT 勘探数据时的可行性与扩展性, 相较于传统基线模型, CNN-BiLSTM 不仅提升了分类准确率与鲁棒性, 也在整体性能平衡方面展现出较好的应用潜力。

为了实现全程无监督学习噪声识别和分类, 笔者将异常识别阶段中标记好的测试集数据切分好放入 CNN-BiLSTM 模型, 按每 2000 取 1800 进行训练, 剩下 200 进行预测分类, 从而达到无监督学习的目的, 并将其与真实数据进行对比, 实验结果如图 11 所示。

从图 11(a), 将 Anomaly Transformer 异常检测阶段标记的预测标签作为噪声分类的真实标签的情况下, CNN-BiLSTM 模型依然能较好地噪声进行分类, 图 11(b)可以看出, 模型还是能对方波噪声, 工频噪声, 阶跃波噪声进行准确的分类; 从图 11(c)和图 11(d)可知, 测试集上的预测结果与真实标签没有发生明显偏移, 准确性也达到了 98.37%, 可见该模型能解决噪声的多级分类的问题。

为全面衡量模型在不同置信度下的判别能力, 根据模型输出对应于各类别的概率, 采用“一对多”的策略, 将每个类别依次作为正类, 其余类别合并为负类, 绘制各类别的一对多 ROC 曲线, 并计算其 AUC 值, 并将其与真实数据进行对比, 实验结果如图 12 所示。

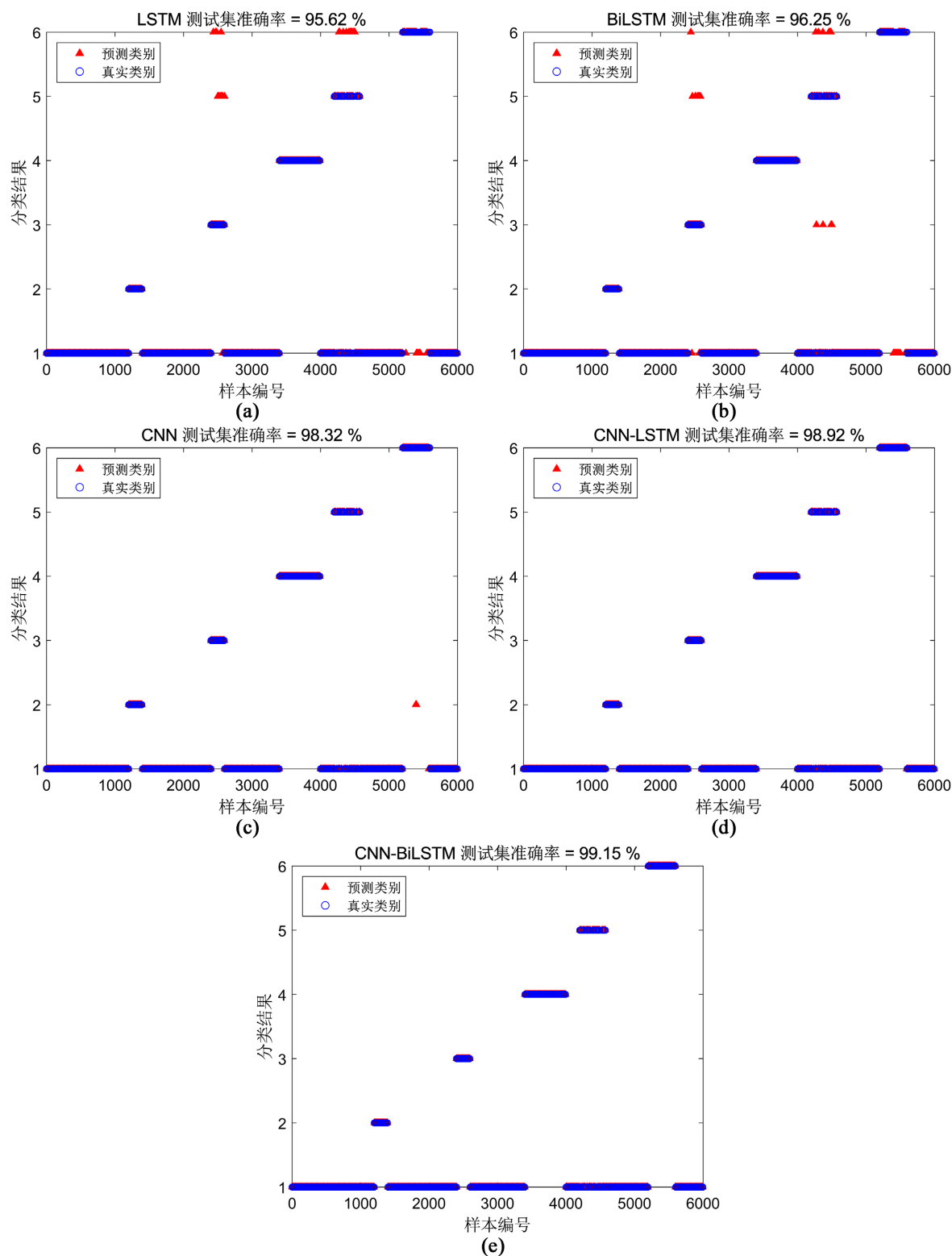


Figure 9. Comparison of test results of different models on the 60,000-sample dataset. (a) LSTM; (b) BiLSTM; (c) CNN; (d) CNN-LSTM; (e) CNN-BiLSTM

图 9. 6w 数据集下不同模型测试结果对比图。(a) LSTM 测试结果图; (b) BiLSTM 测试结果图; (c) CNN 测试结果图; (d) CNN-LSTM 测试结果图; (e) CNN-BiLSTM 测试结果图

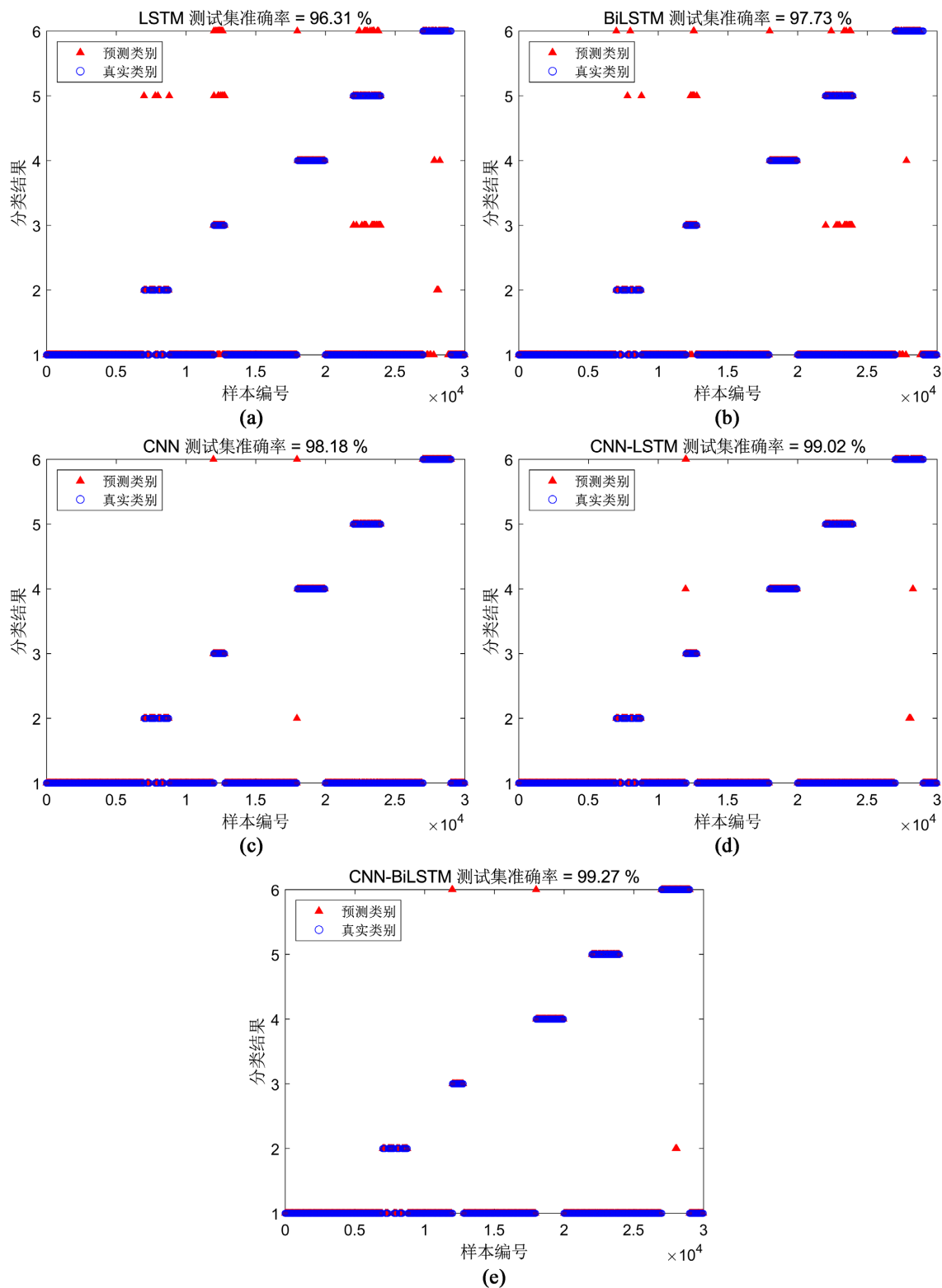


Figure 10. Comparison of test results of different models on the 30,000-sample dataset. (a) LSTM; (b) BiLSTM; (c) CNN; (d) CNN-LSTM; (e) CNN-BiLSTM

图 10. 30w 数据集下不同模型测试结果对比图。(a) LSTM 测试结果图; (b) BiLSTM 测试结果图; (c) CNN 测试结果图; (d) CNN-LSTM 测试结果图; (e) CNN-BiLSTM 测试结果图

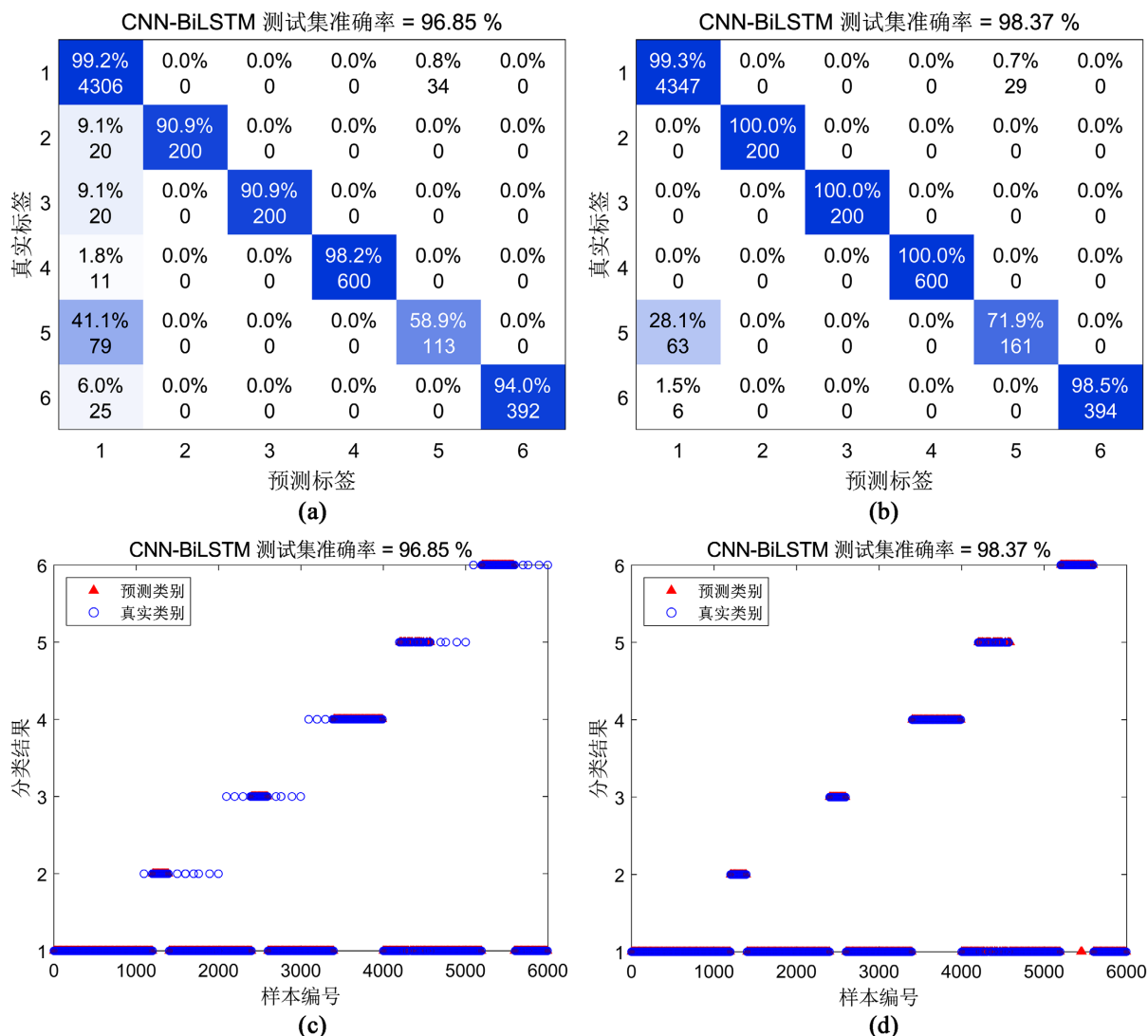


Figure 11. Comparison of model outputs under different ground-truth labeling strategies. (a) Confusion matrix with anomaly detection labels as ground truth; (b) Confusion matrix with true data labels as ground truth; (c) Classification scatter plot using anomaly detection labels as ground truth; (d) Classification scatter plot using true data labels as ground truth

图 11. 不同分类标签设置下模型输出结果对比图。(a) 异常识别标签作为分类真实标签的混淆矩阵图；(b) 真实数据标签作为分类真实标签的混淆矩阵图；(c) 异常识别标签作为分类真实标签的分类结果散点图；(d) 真实数据标签作为分类真实标签的分类结果散点图

从图 12(a)可以看出, 将 Anomaly Transformer 异常检测阶段标记的预测标签作为噪声分类的真实标签的情况下, CNN-BiLSTM 模型模型依然对各类别的识别性能表现良好, AUC 值均高于 0.97, 图 12(b)可以看出, 模型对方波噪声, 工频噪声, 阶跃波噪声, 三角波噪声进行完全准确的分类, 同时对正常信号和脉冲噪声分类性能理想, 表现出模型具备良好的分类准确性和鲁棒性。

5. 结论

数据质量问题一直是大地电磁信号采集与处理的核心问题, 为了准确、高效地进行噪声进行异常检测识别和分类, 本文提出了 Anomaly Transformer 噪声识别和 CNN-BiLSTM 噪声分类的无监督学习联合算法, 该算法融合了集成学习和深度学习的优点, 实现了从端到端的噪声预测和分类, 并通过实验证明

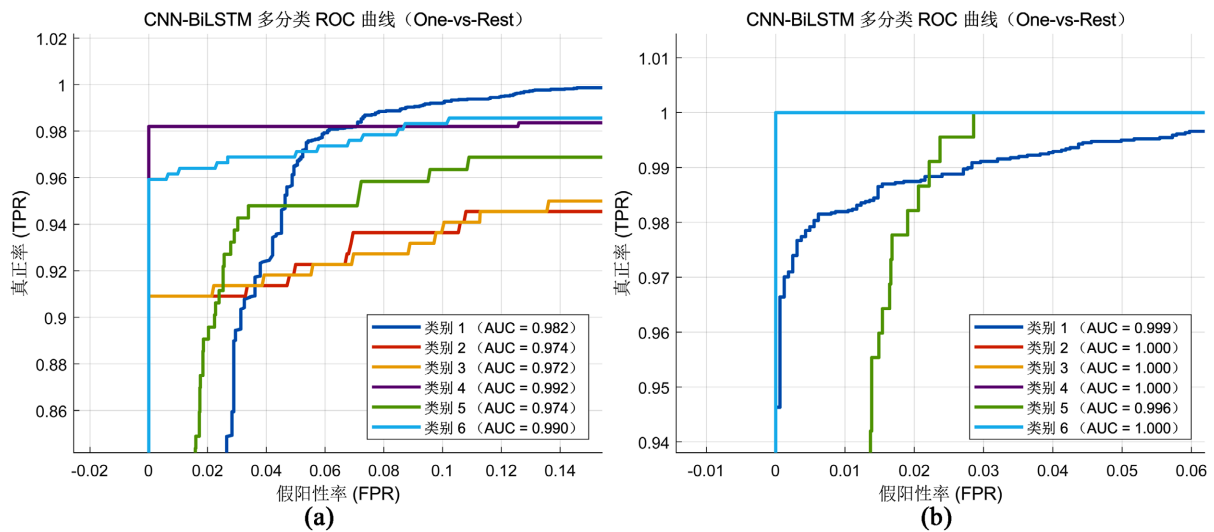


Figure 12. ROC curves under different ground-truth labeling strategies. (a) ROC curve using anomaly detection labels as ground truth; (b) ROC curve using true data labels as ground truth

图 12. 不同分类标签设置下的 ROC 曲线图。(a) 异常识别标签作为分类真实标签的 ROC 曲线图; (b) 真实数据标签作为分类真实标签的 ROC 曲线图

了其有效性以及较为优秀的性能。

实验和研究结果表明:

(1) **Anomaly Transformer** 异常检测模型中改进的异常注意力机制, 能有效捕捉信号中的局部先验关联和全局序列关联, 实现了对强干扰噪声的精准识别和高效检测;

(2) **CNN-BiLSTM** 模型通过 CNN 特征提取和 BiLSTM 有效处理长依赖能力, 还能充分利用数据序列的整体上下文信息, 提高了模型预测的准确性与鲁棒性;

(3) **Anomaly Transformer** 噪声识别和 CNN-BiLSTM 噪声分类的无监督学习联合算法相较于传统机器学习模型、单一深度学习模型和简单组合模型, 具有更好的预测精度和分类性能;

在后续的学术研究中, 笔者将进一步优化自身方法与改进数据处理流程, 提高大地电磁数据处理的自动化程度与识别精准度, 增强操作的便捷性与实用性, 从而为大地电磁数据的高质量处理与地质解释提供更加坚实的技术支持。值得注意的是, 当前方法的训练仍依赖于“具备一定噪声模式标签的样本数据”, 因此在信噪比较低、噪声类型极为复杂的实际野外测区中, 其泛化能力与鲁棒性仍面临挑战。未来仍需依托更广泛的实测数据采集与跨区域验证, 不断提升模型在复杂环境下的适应能力与稳定性。

参考文献

- [1] 王培杰, 陈小斌, 张赞昀, 等. 大地电磁电场远参考道法适用条件及应用效果[J]. 石油地球物理勘探, 2023, 58(6): 1489-1498.
- [2] Li, J., Liu, Y., Tang, J. and Ma, F. (2023) Magnetotelluric Noise Suppression via Convolutional Neural Network. *Geophysics*, **88**, WA361-WA375. <https://doi.org/10.1190/geo2022-0258.1>
- [3] 张昆, 刘磊, 马兴知, 等. 大地电磁测深数据处理方法技术进展[J]. 中国地质调查, 2024, 11(5): 129-138.
- [4] Law, J., Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986) Robust Statistics-The Approach Based on Influence Functions. *The Statistician*, **35**, 565-566. <https://doi.org/10.2307/2987975>
- [5] 张贤. 基于特征聚类与优化学习的电磁法信噪分离方法及应用[D]: [博士学位论文]. 长沙: 中南大学, 2022.
- [6] 严家斌, 刘贵忠. 基于各向异性扩散的 ROBUST 阻抗估计方法[J]. 地球物理学进展, 2007, 22(5): 1403-1407.
- [7] Wang, P.J., Chen, X.B., Han, P., et al. (2024) Strong Interference Magnetotelluric Data Processing Method Based on

- Robust Estimation, Data Screening and Rhoplus Constraint. *Chinese Journal of Geophysics*, **67**, 4325-4342.
- [8] Wang, G., Wang, D., Meng, Y., Li, Y., Wang, W., Zhu, W., *et al.* (2023) High-Speed Railways Interference Signal Characteristics and Multiple Remote References Denoising of Magnetotelluric Data in Jizhong Depression, China. *Applied Sciences*, **13**, Article 4304. <https://doi.org/10.3390/app13074304>
- [9] 徐志敏, 辛会翠, 谭新平, 等. 强电磁干扰区大地电磁远参考技术试验效果分析[J]. 物探与化探, 2018, 42(3): 560-568.
- [10] Alexandrescu, M., Gibert, D., Hulot, G., Le Mouél, J. and Saracco, G. (1995) Detection of Geomagnetic Jerks Using Wavelet Analysis. *Journal of Geophysical Research: Solid Earth*, **100**, 12557-12572. <https://doi.org/10.1029/95jb00314>
- [11] 郭振宇. 短时观测大地电磁信号稳健阻抗估计方法研究[D]: [博士学位论文]. 长春: 吉林大学, 2023.
- [12] 何云玲, 刘琳. 基于核主成分分析的 EMD 去噪算法[J]. 数字技术与应用, 2014(1): 120.
- [13] Zhang, P., Pan, X., Guo, Z., Liu, J. and Hou, Q. (2023) Marine Controlled Source Electromagnetic Data Denoising While Weak Signal Preserving Based on Jointly Sparse Model and Dictionary Learning. *Journal of Applied Geophysics*, **215**, Article ID: 105122. <https://doi.org/10.1016/j.jappgeo.2023.105122>
- [14] Li, G., He, Z., Tang, J., Deng, J., Liu, X. and Zhu, H. (2021) Dictionary Learning and Shift-Invariant Sparse Coding Denoising for Controlled-Source Electromagnetic Data Combined with Complementary Ensemble Empirical Mode Decomposition. *Geophysics*, **86**, E185-E198. <https://doi.org/10.1190/geo2020-0246.1>
- [15] 李博. 基于 VMD-LSTM 的大地电磁强干扰噪声处理[D]: [硕士学位论文]. 桂林: 桂林理工大学, 2024.
- [16] 史维, 严良俊, 谢兴兵, 等. 基于 CEEMDAN-DFA 与 FCM 聚类算法的大地电磁强噪声识别与抑制[J]. 长江大学学报(自然科学版), 2021, 18(5): 13-22.
- [17] 徐猛, 谢凯. 基于分层 Transformer 的相同时间戳错误修复[J/OL]. 计算机系统应用: 1-10. <https://doi.org/10.15888/j.cnki.csa.009911>, 2025-07-04.
- [18] Xu, J., Wu, H., Wang, J., *et al.* (2021) Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy. arXiv: 2110.02642.
- [19] 王祎颜, 王衍学, 姚家驰. 基于 VMD-CNN-BiLSTM 的轴承故障多级分类识别[J]. 机电工程, 2024, 41(9): 1554-1564.
- [20] 张鲁一航, 杨彦明, 陈永展, 等. 基于 VMD-CNN-BiLSTM 的变工况涡扇发动机剩余寿命预测[J/OL]. 北京航空航天大学学报: 1-15. <https://doi.org/10.13700/j.bh.1001-5965.2021.0051>, 2025-07-11.
- [21] 朱威, 范翠松, 姚大为, 等. 矿集区大地电磁噪声场源分析及噪声特点[J]. 物探与化探, 2011, 35(5): 658-662.