

基于 LSTM 循环神经网络模型的冲击地压危险预测

摘要

煤炭是中国主要能源和工业原料，但随着开采深度增加，地应力增大，导致井下煤岩动力灾害风险上升，尤其是冲击地压事故频发，严重威胁煤矿安全。尽管中国煤矿安全形势稳步改善，冲击地压事故仍时有发生，其监测预警和有效防控是当前煤矿安全生产中亟待解决的关键科技问题。

对于问题一，要求分析电磁辐射和声发射信号中的干扰信号数据的特征，并识别其发生区间。本文通过**计算统计特征**如均值、中位数等，并进行**数据可视化**，进一步分析干扰信号的数据特征。利用**双向长短期记忆 (Long Short-Term Memory, LSTM) 循环神经网络模型**，识别干扰信号、传感器断线和工作面休息数据，并确定了电磁辐射和声发射中各自最早出现的五个干扰信号区间，结果见表 5 和表 6。

对于问题二，要求分析电磁辐射和声发射信号中前兆特征信号数据的趋势特征并识别其发生区间。本文基于问题一所建立模型，首先对不同类数据进行可视化处理，**绘制时间序列图**，以辨别前兆数据的趋势特征，而后通过**双向 LSTM 循环神经网络模型**，分别识别电磁辐射和声发射信号中的前兆特征信号，并求出其最早发生的 5 个时间区间，结果见表 10 和表 11。

对于问题三，旨在早期识别并预警前兆信号，要求分析附件 3 中非连续时间段的电磁辐射与声发射数据，预测时间片段最后时刻前兆特征的出现概率。本文采用 LSTM 作为预测工具，对数据进行清洗，去除缺失与异常值。随后，使用**滑动窗口技术**将数据集转化为 LSTM 模型输入格式。通过优化超参数，构建了高效预测模型，最终求出非连续时间段采集的电磁辐射和声发射信号数据中每个时间段最后时刻出现前兆特征数据的概率表 12。

本文基于双向 LSTM 循环神经网络模型，对电磁辐射和声发射信号中的干扰信号数据进行深入分析，以明确其特征并识别出具体发生区间。在此基础上，进一步探讨了电磁辐射和声发射信号中前兆特征信号数据的趋势特性，并精确定位了这些前兆特征的发生区间。此外，本文还成功预测了电磁辐射和声发射信号在特定时刻出现前兆特征信号的概率，为煤矿深部开采冲击地压危险预测的研究与应用提供了有力支持。

关键字：传感器数据 双向 LSTM-RNN 模型 LSTM 预测模型 指数加权移动均值

一、问题重述

1.1 问题背景

煤炭为中国主要能源和工业原料，但随开采加深，地应力增大，煤岩动力灾害风险骤升，危及煤矿安全高效开采。冲击地压尤为严重，常致人伤亡、财损。近年研究深入，防控措施多样，煤矿安全形势稳步向好。但是冲击地压事故仍存，安全形势严峻，监测预警与有效防控仍为紧迫科技难题。

1.2 问题要求

问题 1：分析电磁辐射和声发射信号中干扰信号数据的特征并识别其发生区间。

问题 2：分析电磁辐射和声发射信号中前兆特征信号数据的趋势特征并识别其发生区间。

问题 3：预测电磁辐射和声发射信号所在时刻出现前兆特征的概率。

二、问题分析

2.1 问题一分析

对于问题一，要求分析电磁辐射和声发射信号中干扰信号数据的特征并识别其发生区间。本文首先计算了电磁辐射和声发射信号中各类数据的均值、中位数、众数、方差、变异系数等统计特征，而后对不同类数据进行可视化处理，进一步分析干扰信号数据的特征，关于对干扰信号所在时间区间的识别，本文通过建立双向 LSTM 循环神经网络模型，识别干扰信号数据、传感器断线数据与工作面休息数据，并得出电磁辐射和声发射最早发生的 5 个干扰信号所在的区间。

2.2 问题二分析

对于问题二，需要深入分析电磁辐射与声发射信号中前兆特征信号数据的趋势特性，并精确定位这些特征信号的发生区间。本文依托问题一中所构建的模型框架，首先采取了数据可视化的手段，对不同类型的信号数据进行处理，并绘制出相应的时间序列图，以研究前兆特征数据的趋势特征。随后，借助双向 LSTM 循环神经网络模型，我们针对 2020 年 4 月 8 日-2020 年 6 月 8 日及 2021 年 11 月 20 日-2021 年 12 月 20 日的电磁辐射和 2021 年 11 月 1 日-2022 年 1 月 15 日声发射信号分别进行了细致的分析，成功识

别出其中所蕴含的前兆特征信号，得出电磁辐射和声发射信号最早发生的 5 个前兆特征信号所在的时间区间。

2.3 问题三分析

对于问题三，需要对非连续时间段采集的电磁辐射和声发射信号数据进行分析，预测每个时间段最后时刻出现前兆特征数据的概率，从而尽早的识别前兆特征信号，在前兆特征信号出现的第一时间发出预警。本文采用 LSTM 进行预测，首先对附件 3 中的数据进行清洗，去除数据中的缺失值和异常值，然后将其作为训练集，使用滑动窗口的方法将训练集转化为适合 LSTM 输入的格式，对 LSTM 模型进行训练，同时调整超参数，优化预测模型，最终得到每个时间段最后时刻出现前兆特征数据的概率。

三、模型假设

为简化问题，本文做出以下假设：

- 假设 1：传感器数据准确，具有可信度。
- 假设 2：采集数据时出现的极端异常情况不在考虑范围内
- 假设 3：煤矿开采中的电磁辐射数据与声发射数据能够表征波动规律，且该波动存在普适性。

四、符号说明

符号	含义
X_t	输入
I_t	输入门
F_t	遗忘门
O_t	输出门
H_t	隐藏状态
C_t	记忆细胞
C'_t	候选记忆细胞
α	平滑因子
S_i	平滑值
x_i	观察值
n	样本数
y	表示真实值
\hat{y}	预测值
M	预测结果数
m	识别出的前兆特征信号数
a	前兆特征数据出现概率

五、问题一的模型的建立和求解

5.1 模型简介

对于问题一，要求得出电磁辐射与声发射信号中干扰信号数据特征及其发生区间。本文计算了干扰信号数据的均值、中位数、众数、方差、变异系数等统计特征，并进行可视化分析，而后采用双向 LSTM 循环神经网络模型识别干扰信号、断线及休息数据，确定电磁辐射与声发射最早发生的 5 个干扰信号所在的区间，流程图如图 1。

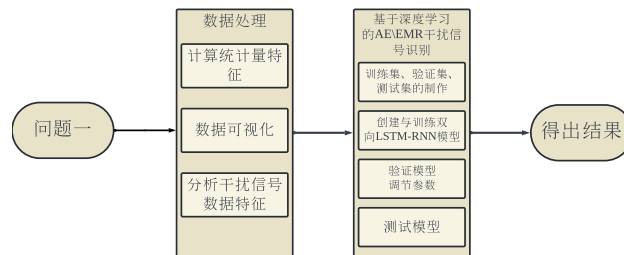


图 1 问题一求解流程图

5.2 模型建立

5.2.1 数据预处理

Step1: 计算数据的统计特征

为识别干扰信号数据的特征，需计算附件 1 中已知的各种数据的均值、标准差、峰峰值、变异系数、中位数、众数等统计特征，对比分析得到干扰信号数据的统计特征。其中，峰峰值为一个周期内信号最高值和最低值之间差的值，可以描述信号值的变化范围的大小，变异系数为又称离散系数，是概率分布离散程度的一个归一化量度，一般来说，变量值平均水平高，其离散程度的测度值越大，反之越小。

Step2: 数据可视化

时间序列图可以观察变量变化发展的趋势及偏差的统计图，本文首先对附件 1 中数据进行处理，按照既有分类，分别制了电磁辐射和声发射信号的时间序列图如图 3和图 2:

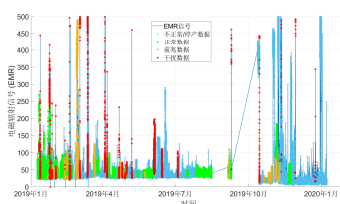


图 2 电磁辐射信号时间序列图

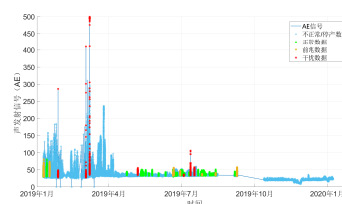


图 3 声发射信号时间序列图

由图 2和图 3较难看出干扰信号数据的特征，可以通过附件 1 绘制相关数据的时间序列图，初步对比分析得到干扰信号数据的特征，同时在此基础上通过循环神经网络处理，从而实现干扰信号的自动识别。本文在此以电磁辐射（EMR）为例，绘制各类数据的时间序列图。

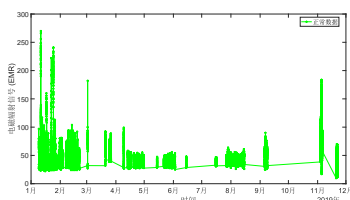


图 4 EMR 正常工作信号数据时间序列图

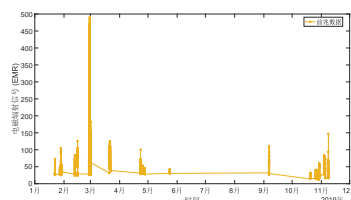


图 5 EMR 前兆特征信号数据时间序列图

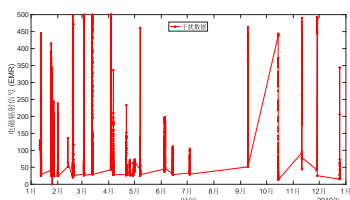


图 6 EMR 干扰信号数据时间序列图

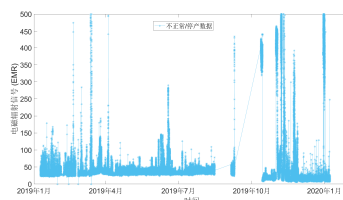


图 7 EMR 不正常/停产信号数据时间序列图

分析图 6可知，相较于其它数据，干扰信号数据在相同时间轴上具有更高频率的大振幅的表现形式，同时电磁辐射信号中干扰信号的强度值都较高，多为 400mv 以上，且其余干扰信号强度值都处于 190 260mv 范围内，而前兆特征信号数据、正常工作信号数据少有强度值处于 190 260mv 内或超过 400mv 的情况出现，监测系统不正常时的信号数据和停产期间的信号数据几乎没有强度值位于 190 260mv 范围内的情况出现。

5.2.2 循环神经网络模型简介

循环神经网络（Recurrent Neural Network, RNN），常用于处理序列数据，拥有自环结构，能记住前序状态，捕捉时序关系的模型，在时间序列和文本数据上表现出色。该模型的主要特点如下：

1. 含隐藏状态的 RNN：处理序列数据，利用隐藏状态存储前序信息，建模时间依赖性。
2. 长短期记忆单元（LSTM）：为应对循环神经网络经常出现的梯度衰减或梯度爆炸问题，捕捉大跨度时间步距离的数据间的联系，本文引用了门控循环神经网络模型^{[4][5]}，引入门控机制（输入门、遗忘门、输出门、记忆细胞），记录并传递数据^[6]。对于长短期记忆中隐藏状态的计算如图 8所示：

其中，长短期记忆的门的输入是输入 X_t 与隐藏状态 H_{t-1} ，输出由激活函数为 *sigmoid* 函数的全连接层计算得到。时间步 t 的输入门 I_t 、遗忘门 F_t 和输出门 O_t 分

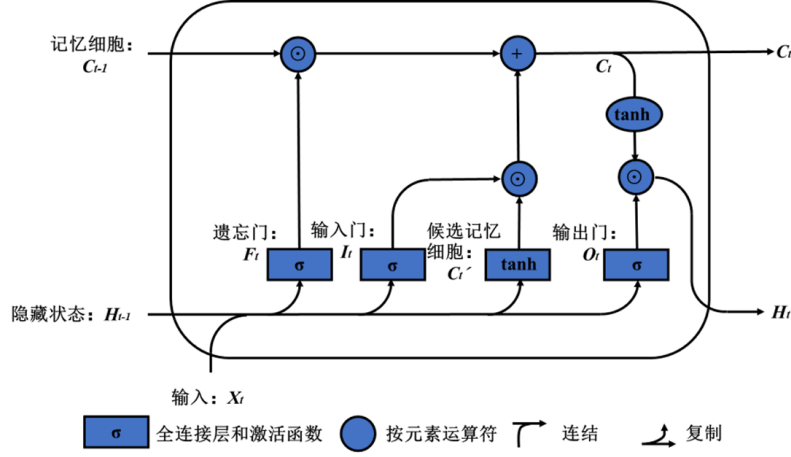


图 8 长短期记忆中隐藏状态的计算

别计算如下：

$$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i)$$

$$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f)$$

$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o)$$

其中的 W_{xi} 、 W_{xf} 、 W_{xo} 和 W_{hi} 、 W_{hf} 、 W_{ho} 都是权重参数， b_i, b_f, b_o 是偏差参数。

此外，长短期记忆需要候选记忆细胞 C'_t 。它使用 \tanh 函数作为激活函数。时间步 t 的候选记忆细胞 C'_t 的计算为：

$$C'_t = \tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c)$$

其中的 W_{xc}/W_{hc} 是权重参数， b_c 是偏差参数。

记忆细胞 C_t 的计算如下：

$$C_t = F_t C_{t-1} + I_t C'_t$$

隐藏状态 H_t 的计算如下：

$$H_t = O_t \tanh(C_t)$$

记忆细胞 C_t 结合了记忆细胞 C_{t-1} 和候选记忆细胞 C'_t 的数据，通过输入门 I_t 和遗忘门 F_t 来控制数据的流动，输入门 I_t 控制输入 X_t 通过候选记忆细胞 C'_t ，如何进入记忆细胞 C_t ，而遗忘门 F_t 则控制记忆细胞 C_{t-1} 中的数据如何进入到时间步 t 。如果遗忘门 F_t 一直近似 1 且输入门 I_t 一直近似 0，过去的记忆细胞将一直保存并传递至当前时间步^[6]。

1. 门控循环单元：同为解决 RNN 梯度问题提出的另一种门控循环神经网络^[7]。
2. 双向循环神经网络：不同于单向 RNN，双向循环神经网络通过增加反向流动的隐藏层来处理数据。本文所建立含单隐藏层的双向循环神经网络的架构如图 9^{[8][9]}：

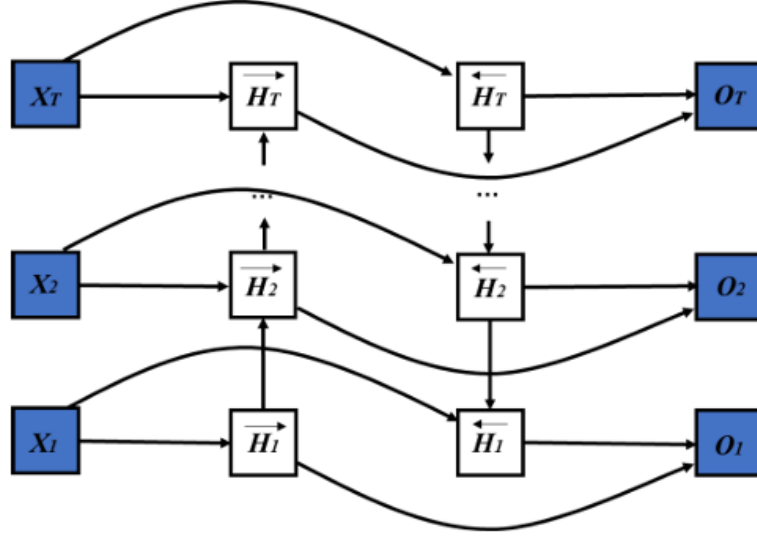


图 9 含单隐藏层的双向循环神经网络的架构

给定输入 X_t 和隐藏层激活函数 ϕ 。在双向循环神经网络的架构中，设时间步 t 正向隐藏状态为 \vec{H}_t ，反向隐藏状态为

$$\overleftarrow{H}_t$$

正向隐藏状态和反向隐藏状态的计算如下：

$$\begin{aligned}\vec{H}_t &= \phi(X_t W_{xh}^{(f)} + \vec{H}_{t-1} W_{hh}^{(f)} + b_h^{(f)}) \\ \overleftarrow{H}_t &= \phi(X_t W_{xh}^{(b)} + \overleftarrow{H}_{t+1} W_{hh}^{(b)} + b_h^{(b)})\end{aligned}$$

5.2.3 建立双向 LSTM-RNN 模型

对于干扰信号数据的识别，本文依据双向长短期记忆循环神经网络模型（模型如图 10所示），分析大量的电磁辐射和声发射信号中的干扰信号数据与其他数据，快速准确的对干扰信号数据进行识别。

Step 1：数据集的制作

由于数据量较小，本文采用附件 1 中全部的干扰信号数据作为训练集。

Step 2：创建、训练模型

通过读取训练集、创建数据迭代器、使用双向长短期记忆单元和输出分类结果等手段，定义完整的双向长短期记忆循环神经网络模型，然后训练模型、定义预测函数。训练好模型后，读取验证集数据，调节循环神经网络模型超参数，优化循环神经网络模型。

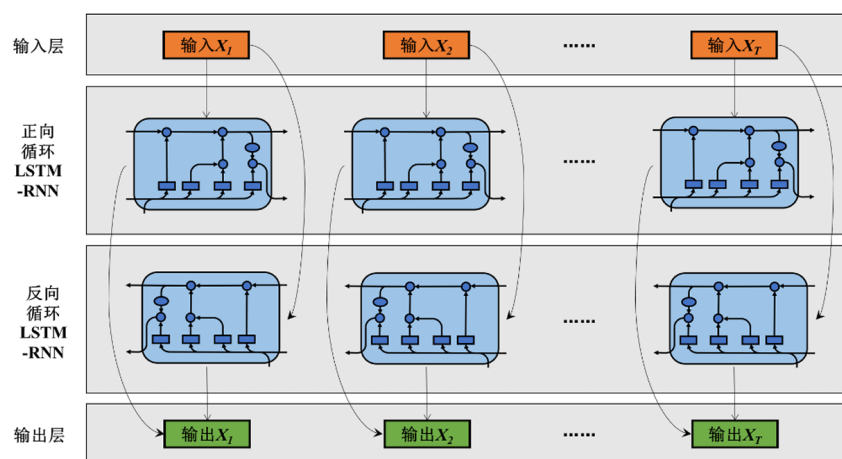


图 10 双向长短期记忆循环神经网络模型

Step 3: 识别干扰信号数据

对于优化好的循环神经网络模型，读入附件 2 中需识别的时间区间内的电磁辐射和声发射信号数据，得出识别出的电磁辐射和声发射最早发生的 5 个干扰信号所在的区间。

5.3 模型求解

5.3.1 分析干扰信号数据的特征

Step1: 计算数据的统计特征

计算附件 1 中已知的各种数据的均值、标准差、峰峰值、变异系数、中位数、众数等统计特征，由题设可知 A 代表正常工作数据，B 代表前兆特征数据，C 代表干扰信号数据，D 代表传感器断线数据，E 工作面休息数据，所得结果如表 1 和表 2 所示：

表 1 电磁辐射 (EMR) 信号数据的统计特征

AE	C	A	B	D/E
峰峰值	473.000	51.000	52.000	310.000
均值	59.969	37.433	41.669	32.901
标准差	80.422	3.727	8.304	13.187
中位数	42.376	36.661	39.105	33.419
众数	34.000	48.000	47.000	34.000
变异系数	134.105	9.957	19.928	40.080

简单分析表 1 可知，对于电磁辐射 (MER) 信号的各类数据，在集中趋势上，C 的均值高于 A 和 D/E，但明显低于 B，表明 C 数据集整体水平较高的同时也存在低值。相

表 2 声发射 (AE) 信号数据的统计特征

EMR	C	A	B	D\E
峰峰值	500.000	260.390	479.330	500.000
均值	77.959	49.815	71.055	42.089
标准差	90.716	18.279	91.900	62.663
中位数	45.000	46.490	41.000	31.000
众数	28.000	45.000	37.000	29.000
变异系数	116.364	36.694	129.337	148.884

比之下，A 的均值最低，这显示出不同数据集的整体分布不同。在分散程度上，C 和 B 都显示出较高的波动性（标准差和变异系数），但 B 的中位数和众数相对较高，表明 B 的整体分布集中在较高的数值上，而 C 则分布较为分散，说明其包含了较多的低值。在极端值影响上，C 的高峰峰值和与中位数、众数的差距显示出可能存在显著的极端值。这种特征在很多统计分析中可以引起注意，可能会影响整体结果。在不对称性上，C 的中位数和众数均远低于其均值，表明 C 的数据可能受到少数高值的影响，且数据可能是右偏的。相比之下，其他数据集如 A 和 D/E 的数据分布更为对称。

综上所述，C 的均值和标准差较高，且峰峰值显示了数据的较大变异性，然而，其较低的中位数和众数表明，C 中确实存在较多低值，这些低值主导了数据的分布形态，同时，相较于 A、B 和 D/E 的数据特征，C 的波动性大和且具有较强的不对称性。

表 3 声发射 (AE) 信号数据的统计特征

	峰峰值	均值	标准差	中位数	众数	变异系数
A	51	37.43299	3.727368	36.661	48	9.957442
B	52	41.66866	8.303728	39.105	47	19.928
C	473	59.9691	80.4218	34.000	34	134.1054
D\E	310	32.90133	13.18683	33.419	34	40.07994

简单分析表 3 可知，对于声发射信号的各类数据，在集中趋势上，C 的均值虽然较高，但受到高峰峰值的影响，可能不代表数据的典型水平。相反，A 和 D/E 的均值较低，更为集中，而 B 的均值较高，显示出更为偏向于中高值区间。在分散程度上，C 的标准差和变异系数最高，表明数据点的波动性非常大。尤其是在比较整体时，A 的标准差和变异系数非常低，显示出其数据点相对均匀。在极端值影响上，C 的峰峰值高达 473，说明存在数据极端值的可能性。中位数和众数都较低，表明尽管有一些高值拉动均值上升，但整体数据中仍包含大量低值。在不对称性上，C 的中位数和众数都明显低于均值，这表明数据的分布存在右偏现象，即低值较为集中，而部分高值主导了均值的计算。

综上所述，数据集 C 显示出独特的特征，整体均值高但中位数和众数相对较低，反映出数据的不对称性，高峰峰值和标准差表明数据的离散性大，而变异系数显示出数据的相对变异性非常高。这样的特征说明 C 中较低的值在整体数据中占据较大比例，但偶尔会出现较高的异常值。

Step 2: 分析干扰信号数据变化特征

在图 6(EMR 干扰信号数据时间序列图) 基础上，选取较短的时间间隔，观察干扰信号数据的变化特征，绘制时间序列图如图 11：

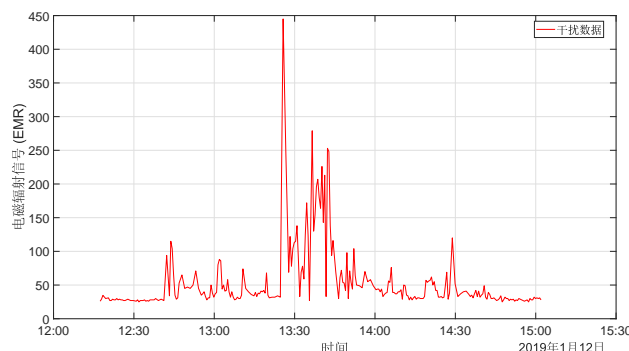


图 11 电磁辐射干扰信号数据时间序列图 (局部)

分析图 11 可知，干扰信号数据中，较低的值在整体数据中占据较大比例，说明干扰信号在短时间内出现较大的突增与突降，会出现极密集的尖刺状突增变化，同时，在出现干扰信号的时间区间内，干扰信号数据的强度值呈现偏态分布，存在右偏现象。

Step 3: 总结综合上文对于干扰信号的统计特征和变化特征的分析，可以得出电磁辐射和声发射中的干扰信号数据的特征如表 4：

	干扰信号数据的特征			
电磁辐射 (EMR)	离散程度大	突变趋势	偏态分布 (右偏)
声发射 (AE)	离散程度大	突变趋势	偏态分布 (右偏)

表 4 电磁辐射和声发射中的干扰信号数据特征

5.3.2 双向 LSTM-RNN 模型识别干扰信号

通过上文所建立模型，导入 2022 年 5 月 1 日-2022 年 5 月 30 日的电磁辐射和 2022 年 4 月 1 日-2022 年 5 月 30 日及 2022 年 10 月 10 日-2022 年 11 月 10 日声发射信号的全部未知数据，对干扰信号所在的时间区间进行识别，得出电磁辐射和声发射最早发生的 5 个干扰信号所在的区间为：

表 5 电磁辐射干扰信号时间区间

序号	时间区间起点	时间区间终点
1	'2022-05-01 10:57:50'	'2022-05-01 11:08:48'
2	'2022-05-03 08:23:06'	'2022-05-03 08:23:06'
3	'2022-05-07 18:04:53'	'2022-05-07 18:04:53'
4	'2022-05-10 06:27:57'	2022-05-10 08:30:29'
5	2022-05-13 06:45:57'	'2022-05-13 10:34:42'

表 6 声发射前兆特征时间区间

序号 (AE)	时间区间起点	时间区间终点
1	2021-11-1 0:01	'2021-11-01 18:53:59'
2	'2021-11-26 05:18:04'	'2021-11-26 05:28:56'
3	'2021-12-07 03:13:21'	'2021-12-07 13:40:11'
4	'2021-12-07 13:40:11'	'2021-12-07 13:40:11'
5	'2022-01-04 03:50:02'	'2022-01-04 08:31:41'

六、问题二的模型的建立和求解

6.1 模型简介

本文通过时间序列图分析出前兆数据的趋势特征，然后通过双向 LSTM 循环神经网络模型，对 2020 年 4 月 8 日-2020 年 6 月 8 日及 2021 年 11 月 20 日-2021 年 12 月 20 日的电磁辐射和 2021 年 11 月 1 日-2022 年 1 月 15 日声发射信号数据进行分析，成功识别出电磁辐射和声发射信号的前兆特征信号，并确定了这些前兆特征信号最早发生的时间区间。具体流程图如图 12：

6.2 模型建立

6.2.1 数据预处理

Step1：计算数据的统计特征

为识别前兆特征信号的趋势特征，需计算附件 1 中前兆特征信号的均值、标准差、峰峰值、变异系数、中位数、众数等统计特征，借以分析其趋势特征。

Step2：数据可视化

绘制电磁辐射和声发射信号中前兆特征数据的时间序列图，进一步前兆特征信号的趋势特征。

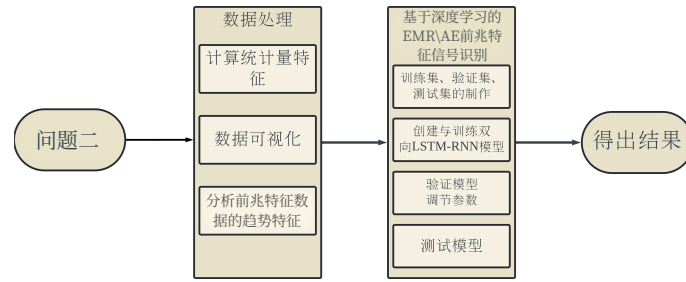


图 12 问题二流程图

6.2.2 建立双向 LSTM-RNN 模型

此处模型建立基于问题一的模型基础，基本原理一致，在此不做重复赘述。

Step 1: 数据集的制作

由于数据量较小，本文采用附件 1 中全部的前兆特征信号数据作为训练集。

Step 2: 创建、训练模型

通过读取训练集、创建数据迭代器、使用双向长短期记忆单元和输出分类结果等手段，定义完整的双向长短期记忆循环神经网络模型，然后训练模型、定义预测函数。训练好模型后，读取验证集数据，调节循环神经网络模型超参数，优化循环神经网络模型。

Step 3: 识别干扰信号数据

对于优化好的循环神经网络模型，读入附件 2 中需识别的时间区间内的电磁辐射和声发射信号数据，得出识别出的电磁辐射和声发射最早发生的 5 个干扰信号所在的区间。

6.3 模型求解

6.3.1 分析前兆特征信号的趋势特征

Step1: 计算前兆特征信号的统计特征

计算附件 1 中电磁辐射和声发射信号的前兆特征数据，得出下表：

表 7 : 电磁辐射（EMR）前兆特征信号统计特征

峰峰值	均值	标准差	中位数	众数	变异系数
479.33	71.05481	91.89984	41	37	129.3366

分析表 7 可得，EMR 的前兆特征信号呈现不对称分布、高变异性和潜在的极端信号

等趋势特征。其中，**偏态分布**体现在数据集可能呈现右偏分布，大多数数据集中在低值区域，但也存在显著的较高值。**高变异性**体现在数据的波动性很大，说明信号在不同时间或条件下可能有很大的变化，这需要在进一步的分析和处理中考虑异常值对结果的影响。**潜在的极端信号**体现在峰峰值较高和较低的均值、中位数及众数的对比，表明信号中可能包含一些重要的极端特征。此特征若与事件发生的时间或条件相联系，可能指向潜在的预警。

表 8 声发射信号（AE）前兆特征信号统计特征

峰峰值	均值	标准差	中位数	众数	变异系数
52	41.66866	8.303728	39.105	47	19.928

分析表 8可知，声发射信号中的前兆特征信号呈现稳定性高、集中性高、正态分布趋势等趋势特征。其中，**稳定性高**体现在数据的标准差和变异系数较小，说明信号的变化不大，适合在影像处理、模型训练等场景中使用，且结果相对可预测。**集中性高**体现在数据分布较为集中，众数和均值的接近程度指向了数据点在较高值区域的频繁出现，表明有可能的趋势和模式。正态分布趋势体现在数据显示出一定的**正态分布趋势**，分布的对称性较好，且极端值几乎不存在。

Step 2: 分析前兆特征信号的变化趋势特征

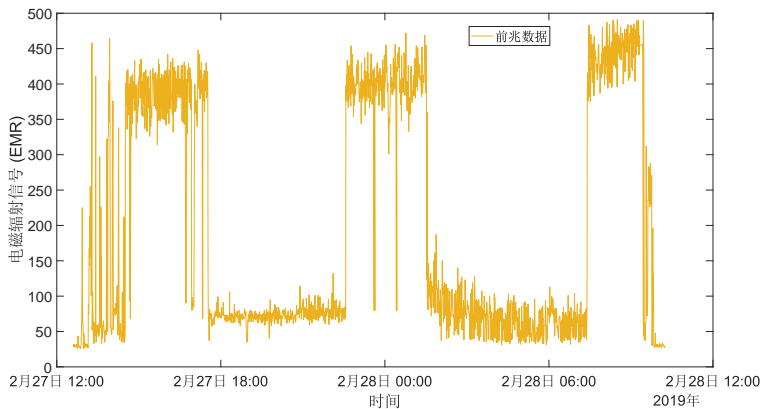


图 13 电磁辐射（EMR）前兆特征信号时间序列图（局部）

分析图 13可得，电磁辐射中的前兆特征信号，较低的值与较高的值分布比较均衡，具有一定的低—高一低—高的周期性，同时前兆特征信号的强度值都会随时序推移呈现上升—下降—上升的变化趋势，分布并不对称，具有极端信号的存在。

图声发射信号前兆特征信号时间序列图（局部）分析图 ○ 可得，声发射信号中的前兆特征信号数据较为平稳，偶尔有剧烈的上下波动，其余大部分数据集集中于 40 50 的范围区间，说明声发射特征信号数据具有集中性，且稳定性高。

6.3.2 双向 LSTM-RNN 模型识别前兆特征信号

通过上文所建立模型，导入 2020 年 4 月 8 日-2020 年 6 月 8 日及 2021 年 11 月 20 日-2021 年 12 月 20 日的电磁辐射和 2021 年 11 月 1 日-2022 年 1 月 15 日声发射信号的全部未知数据，对前兆特征信号所在的时间区间进行识别，得到各自前兆特征信号最早出现的五个区间。

6.4 求解结果

6.4.1 前兆特征信号趋势

综合上文对于前兆特征信号的统计特征和变化特征的分析，可以得出电磁辐射和声发射中前兆特征信号数据的趋势特征如表 9所示：

表 9 干扰信号数据的特征

	干扰信号数据的特征			
电磁辐射（EMR）	高变异性	潜在的极端信号	偏态分布（右偏）
声发射（AE）	稳定性高	集中性高	正态分布趋势

6.4.2 前兆特征信号识别结果

分析图（前兆特征信号识别（EMR））与图（前兆特征信号识别（AE））可得 2020 年 4 月 8 日-2020 年 6 月 8 日及 2021 年 11 月 20 日-2021 年 12 月 20 日的电磁辐射和 2021 年 11 月 1 日-2022 年 1 月 15 日声发射信号中前兆特征信号最早发生的 5 个区间如表 10所示：

表 10 电磁辐射前兆特征时间区间

序号	时间区间起点	时间区间终点
1	2020.4.12	2020.4.15
2	2020.5.1	2020.5.3
3	2020.5.15	2020.5.20
4	2021.11.25	2021.11.28
5	2021.12.01	2021.12.03

表 11 声发射前兆特征时间区间

序号	时间区间起点	时间区间终点
1	2021.11.5	2021.11.07
2	2021.11.15	2021.11.20
3	2021.12.01	2021.12.05
4	2021.12.10	2021.12.14
5	2021.12.20	2021.12.23

七、问题三的模型的建立和求解

7.1 模型简介

本文针对非连续时间段的电磁辐射和声发射信号数据，采用 LSTM 模型进行前兆特征信号数据预测。首先对数据进行清洗，去除缺失和异常值，然后通过滑动窗口方法转换为适合 LSTM 输入的格式，并进行模型训练和超参数调整，以优化预测结果，最终预测每个时间段末尾出现前兆特征数据的概率，并实现及时预警，具体流程如图 14：

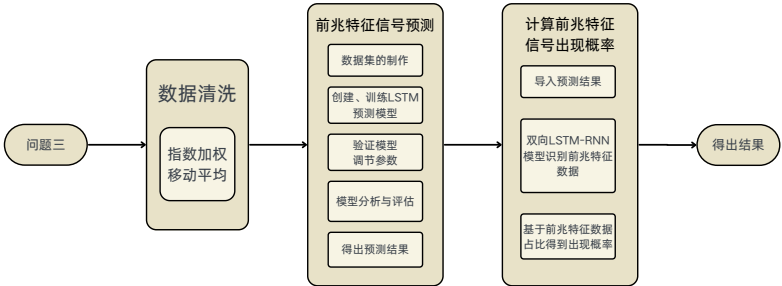


图 14 问题三流程图

7.2 模型建立

7.2.1 数据预处理

由于 C 类和 D、E 类数据均为干扰类数据，属于异常值，对于附件 3 中数据，首先需要进行数据清洗，通过指数加权移动平均法进行去噪，具体步骤如下：

Step 1: 选择平滑因子 α

α 值越大，模型对最新数据点的敏感度越高； α 值越小，模型的反应相对较慢，更加关注历史数据，本文选择平滑因子 $\alpha = 0.05$ 。

Step 2: 初始化第一个平滑值

对于第一个平滑值 S_1 ，本文使用前几个值的平均值进行计算，计算公式为：

$$S_1 = \frac{1}{n} \sum_{i=1}^n x_i$$

其中， n 是初始样本的数量。

Step 3: 迭代计算后续平滑值

$$S_i = \alpha x_i + (1 - \alpha) S_{i-1}$$

其中， S_i 是当前时刻的平滑值， x_i 是当前观察值， S_{i-1} 是上一个时刻的平滑值

Step 4: 重复计算

对于每一个新的观察值，使用上述公式计算新的平滑值，直到处理完所有数据点。

7.2.2 LSTM 预测模型

Step 1: 数据集的制作

由于数据量较小，本文将附件 3 中数据的 80% 为训练集，20% 作为验证集优化模型参数。

Step 2: 创建、训练 LSTM 预测模型通过读取训练集、创建数据迭代器、使用长短期记忆单元和输出分类结果等手段，定义完整的长短期记忆循环神经网络模型，然后训练模型、定义预测函数。训练好模型后，读取验证集数据，调节模型超参数，优化 LSTM 预测模型。

7.2.3 模型分析与评估

对于预测结果，本文在这里进行定性分析和定量分析，以确定每个时间段最后时刻出现前兆特征数据的概率。

Step 1: 定性分析

使用图像法，即在一张图中绘制出测试数据的曲线与相应时间段的各种模型的预测数据的曲线，对比测试数据与预测数据的曲线发展趋势和相关性。若曲线发展趋势相同，且具有较高的相关性，甚至有大量的重合曲线，则表明模型的预测结果与测试数据的拟合效果良好，模型的预测效果较好。反之，则表明模型预测效果较差。

Step 2: 定量分析

为了定量地对模型的预测结果进行评价，假设 y 表示真实值， \bar{y} 表示真实值的平均值， \hat{y} 表示预测值， n 为样本数。评价指标如下：

1. 评价绝对误差：表示预测值与真实值之间的绝对误差的平均值。

$$MAE = \frac{1}{n} \sum_{j=1}^n |\hat{y}_j - y_j|$$

2. 均方根误差：表示预测值与真实值误差的平方与样本数比值的平方根。

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (\hat{y}_j - y_j)^2}$$

3. 平均绝对百分误差：表示预测值与真实值之间的绝对百分比误差的平均值。

$$MAPE = \frac{1}{n} \sum_{j=1}^n \left| \frac{\hat{y}_j - y_j}{y_j} \right|$$

4. 决定系数：表示预测值与真实值之间的拟合程度。

$$R^2 = 1 - \frac{\sum_{j=1}^n (\hat{y}_j - y_j)^2}{\sum_{j=1}^n (y_j - \bar{y})^2}$$

由以上各式可以得出，平均绝对误差和均方根误差从预测值与真实值之间直接误差的方面对预测结果进行评价，平均绝对误差和均方根误差的值越小，表明模型预测效果越好。平均绝对百分误差则从预测值与真实值的误差，以及误差与真实值的比值方面对预测结果进行评价，平均绝对百分误差解释了预测值与真实值之间的相对误差，其值越小，表明模型预测效果越好，平均绝对百分误差经常用于评价不同模型的表现。决定系数则反映了预测结果的可靠程度，其值越接近 1，表明预测值与真实值越接近，模型拟合效果越好。

7.2.4 双向 LSTM-RNN 模型识别

对于附件 3 中的每个时间段最后时刻出现前兆特征数据的概率，本文基于问题二所建立的模型，通过双向 LSTM-RNN 模型识别得到的预测结果中的前兆特征信号，然后计算识别出的前兆特征信号在所有预测结果中的占比即为每个时间段最后时刻出现前兆特征数据的概率 a ，计算公式如下：

$$a = \frac{m}{M} \cdot 100\%$$

其中， M 为预测结果数， m 为识别出的前兆特征信号数。

7.3 模型求解

7.3.1 LSTM 预测模型预测前兆特征信号

对于附件 3 中的数据，首先按照各自时间段进行分类，然后通过指数加权移动平均法对数据中的异常值进行去噪，提高数据质量。由于数据量较小，本文将附件 3 中分类

处理后的数据全部作为训练集进行 LSTM 预测模型的训练，预测每个时间段电磁辐射和声发射信号各自最后时刻往后的 170 个数据，预测结果如下：由电磁辐射时序预测的

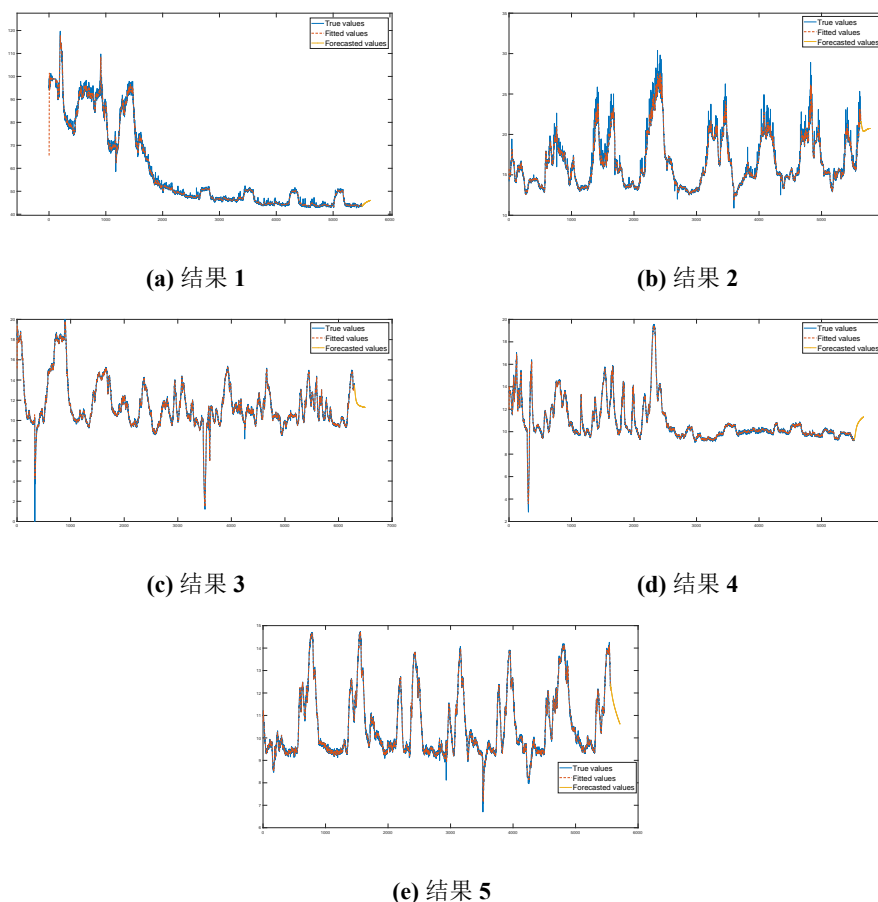


图 15 电磁辐射时序预测结果

结果图可以初步推断，图 15b、图 15c、图 15e 中的时间段最后时刻出现前兆特征信号的概率较大，图 15a、图 15d 中的时间段最后时刻出现前兆特征信号的概率较小。

由声发射时序预测的结果图可以初步推断，图 16b、图 16c、图 16e 中的时间段最后时刻出现前兆特征信号的概率较大，图 16a、图 16d 中的时间段最后时刻出现前兆特征信号的概率较小。

7.3.2 计算前兆特征信号出现概率

基于预测结果与问题二所建立的双向 LSTM-RNN 模型，识别所得 170 个预测结果中的前兆特征数据，本文以声发射时序预测结果 2 为例，通过问题二所训练的双向 LSTM-RNN 模型，得到识别结果，如图 17：

由图 17 所得结果可计算出声发射数据所在时刻为 2023-2-11 23:59:20 时，前兆特征信号出现的概率 $a = 0.43$ ，以此类推，可以计算出附件 3 中的每个时间段最后时刻出现前兆特征数据的概率。

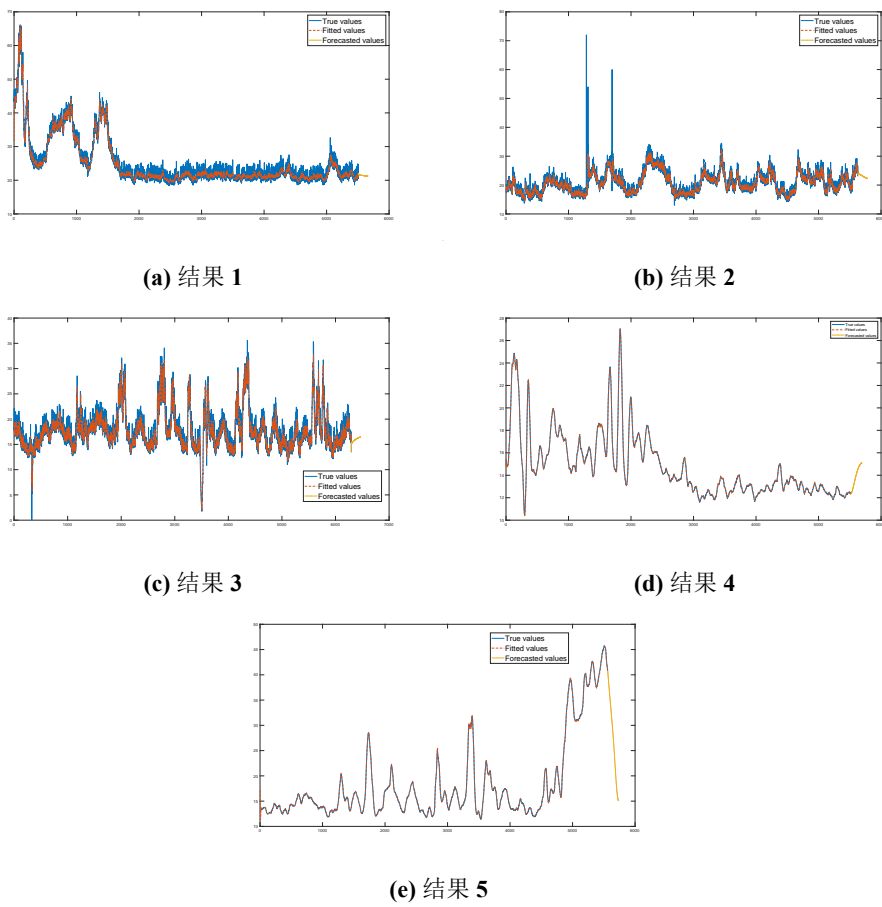


图 16 声发射时序预测结果

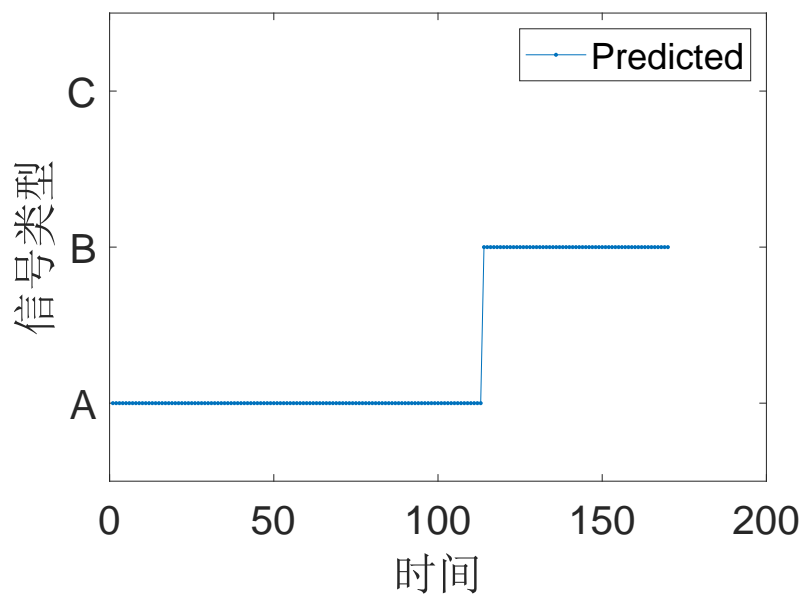


图 17 声发射时序预测结果 2 识别效果图

7.4 求解结果

综合上述模型，得出问题三所求结果如表 12：

表 12 问题三所求结果

电磁辐射数据所在时刻	前兆特征的概率	声发射数据所在时刻	前兆特征的概率
2023-1-24 23:58:36	0.85	2023-1-24 23:58:36	0.05
2023-2-11 23:59:20	0.98	2023-2-11 23:59:20	0.43
2023-2-26 23:59:27	0.55	2023-2-26 23:59:27	0.87
2023-3-10 23:58:14	0.23	2023-3-10 23:58:14	0.13
2023-3-30 23:58:13	0.95	2023-3-30 23:58:13	0.86

八、模型的评价

8.1 模型的优点

- LSTM 模型具有良好的**灵活性**，可以处理可变长度的输入序列，非常适合时间序列预测和分类任务。
- LSTM 模型具有良好的**泛化能力**：在合适的设置下，LSTM 可以适应复杂的模式和噪声，表现出较好的泛化能力。
- 使用移动均值方法平滑数据**简单直观**，易于实现，通过计算一定窗口内的数据均值，可以**有效抑制随机噪声**，**突出数据的趋势**，平滑效果好，且能实时更新。

8.2 模型的缺点

- LSTM 模型有时会存在**过拟合风险**，如果数据样本不足或模型过于复杂，LSTM 可能容易发生过拟合，导致在测试数据上表现欠佳。
- LSTM 模型的**超参数调整复杂**，多个超参数需要调优，如时间步长、隐藏层单元数、学习率等，可能会耗费较多时间进行实验。
- LSTM 模型的**计算复杂度高**，LSTM 模型通常参数较多，计算成本高，训练时间较长，尤其在大规模数据集上。
- 使用移动均值方法平滑数据会引入**滞后**，导致平滑后的数据相较于原始数据出现延迟，可能影响实时决策，同时具有**窗口选择问题**，选择合适的窗口大小至关重要，窗口过小可能无法有效平滑，窗口过大则可能导致过度平滑。

参考文献

- [1] 司守奎, 孙玺菁. 数学建模算法与应用[M]. 北京: 国防工业出版社, 2011.
- [2] 卓金武. MATLAB 在数学建模中的应用[M]. 北京: 北京航空航天大学出版社, 2011.
- [3] 狄洋阳. 基于深度学习的冲击地压震声电综合预警研究[D]. [出版地不详]: 中国矿业大学, 2022.
- [4] GRAVES A, MOHAMED A R, HINTON G. Speech recognition with deep recurrent neural networks[C]//2013 IEEE international conference on acoustics, speech and signal processing. [S.l.]: Ieee, 2013: 6645-6649.
- [5] SAK H, SENIOR A, BEAUFAYS F. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition[J]. arXiv preprint arXiv:1402.1128, 2014.
- [6] SUTSKEVER I. Sequence to sequence learning with neural networks[J]. arXiv preprint arXiv:1409.3215, 2014.
- [7] GERS F A, SCHMIDHUBER J, CUMMINS F. Learning to forget: Continual prediction with lstm[J]. Neural computation, 2000, 12(10):2451-2471.
- [8] GRAVES A, SCHMIDHUBER J. Framewise phoneme classification with bidirectional lstm and other neural network architectures[J]. Neural networks, 2005, 18(5-6):602-610.
- [9] LUKOŠEVIČIUS M, JAEGER H. Reservoir computing approaches to recurrent neural network training[J]. Computer science review, 2009, 3(3):127-149.

附录 A 文件列表

文件名	功能描述
pro1.m	问题一程序代码
pro2.m	问题二程序代码
pro3.m	问题三程序代码

附录 B 代码

pro1.m

```
1 data1_EMR = readtable("Attachment1.xlsx", 'Sheet', 'EMR');
2 data1_AE = readtable("Attachment1.xlsx", 'Sheet', 'AE');
3
4 %%
5 % 分离 EMR, time 和 class 列
6 EMR = data1_EMR.EMR;
7 time_EMR = datetime(data1_EMR.time, 'InputFormat', 'yyyy-MM-dd
    HH:mm:ss');
8 class_EMR = data1_EMR.class;
9
10 % 创建图形
11 figure;
12 hold on; % 保持当前图
13
14 % 使用逻辑索引绘制不同类别的数据
15 C_idx = class_EMR == "C"; % 找到C类数据的索引
16 normal_idx = (class_EMR == "A"); % 找到正常数据的索引
17 predict_idx = (class_EMR == "B");
18 %DE
19 DE_idx = (class_EMR == "D/E");
20 plot(time_EMR, EMR, 'LineWidth', 1); % 蓝色线
21
22
23 plot(time_EMR(DE_idx), EMR(DE_idx), '*', 'Color', [0.3010 0.7450
    0.9330], 'LineWidth', 0.5)
```

```

24 % 绘制 C 类 的电磁辐射信号
25
26
27 % 绘制正常类 （A, B, D）的电磁辐射信号
28 plot(time_EMR(normal_idx), EMR(normal_idx), 'g*', 'LineWidth',
      1); % 蓝色线
29
30 plot(time_EMR(predict_idx), EMR(predict_idx), '*', 'Color'
      , [0.9290 0.6940 0.1250], 'LineWidth', 1); % 紫色线
31
32 plot(time_EMR(C_idx), EMR(C_idx), 'r*', 'LineWidth', 1); % 红
      色线
33
34 % 设置图形的属性
35 xlabel('时间');
36 ylabel('电磁辐射信号 (EMR)');
37 %title('电磁辐射信号时间序列图');
38 legend('EMR信号', '不正常/停产数据', '正常数据', '前兆数据', '干
      扰数据', 'Location', 'Best');
39 grid on;
40
41 hold off; % 停止保持图形
42 set(gca, 'fontsize', 20)
43 % 在这里你可以继续添加分析 C 类数据的其他代码
44 % 例如:
45 % 绘制 C 类的统计特征（均值、标准差等）
46 range_C = max(EMR(C_idx)) - min(EMR(C_idx));
47 mean_C = mean(EMR(C_idx));
48 std_C = std(EMR(C_idx));
49 med_C = median(EMR(C_idx));
50 mod_C = mode(EMR(C_idx));
51 cv_C = std_C/mean_C*100;
52
53 disp(['C类数据的峰峰值: ', num2str(range_C)]);
54 disp(['C类数据的均值: ', num2str(mean_C)]);

```



```

55 disp(['C类数据的标准差: ', num2str(std_C)]);
56 disp(['C类数据的中位数: ', num2str(mod_C)]);
57 disp(['C类数据的众数: ', num2str(mod_C)]);
58 disp(['C类数据的变异系数: ', num2str(cv_C)]);
59 %%
60
61 % 创建图形
62 figure;
63 % hold on; % 保持当前图
64
65 plot(time_EMR, EMR, 'LineWidth', 1); % 蓝色线
66 % 设置图形的属性
67 xlabel('时间');
68 ylabel('电磁辐射信号 (EMR)');
69 title('电磁辐射信号时间序列图');
70 legend('EMR信号','Location', 'Best');
71 set(gca,'fontsize',20)
72
73
74
75 figure
76 plot(time_EMR(DE_idx), EMR(DE_idx),'-*','Color',[0.3010 0.7450
    0.9330] , 'LineWidth', 0.5)
77 % 绘制 C 类 的电磁辐射信号
78 % 设置图形的属性
79 xlabel('时间');
80 ylabel('电磁辐射信号 (EMR)');
81 % title('电磁辐射信号时间序列图');
82 legend('不正常/停产数据','Location', 'Best');
83 set(gca,'fontsize',20)
84
85 figure
86 % 绘制正常类 (A, B, D) 的电磁辐射信号
87 plot(time_EMR(normal_idx), EMR(normal_idx), 'g-*', 'LineWidth'
    , 1); % 蓝色线

```

```

88 % 设置图形的属性
89 xlabel('时间');
90 ylabel('电磁辐射信号 (EMR)');
91 % title('电磁辐射信号时间序列图');
92 legend('正常数据','Location', 'Best');
93 set(gca,'fontsize',20)
94
95
96
97
98 EMR_C_time = time_EMR(C_idx);
99 EMR_C_data = EMR(C_idx);
100
101
102 EMR_B_time = time_EMR(predict_idx);
103 EMR_B_data = EMR(predict_idx);
104
105 figure
106 % plot(EMR_B_time(59:736), EMR_B_data(59:736) , '-', 'Color
    ',[0.9290 0.6940 0.1250], 'LineWidth', 1); % 紫色线
107 plot(EMR_B_time(1187:1767), EMR_B_data(1187:1767) , '-', '
    Color',[0.9290 0.6940 0.1250], 'LineWidth', 1);
108 % plot(EMR_B_time(2544:4665), EMR_B_data(2544:4665) , '-', '
    Color',[0.9290 0.6940 0.1250], 'LineWidth', 1);
109 % 设置图形的属性
110 xlabel('时间');
111 ylabel('电磁辐射信号 (EMR)');
112 % title('电磁辐射信号时间序列图');
113 legend('前兆数据','Location', 'Best');
114 set(gca,'fontsize',20)
115
116
117 figure
118 plot(EMR_C_time(40:308), EMR_C_data(40:308), 'r-', 'LineWidth'
    , 1); % 红色线

```

```

119
120 % 设置图形的属性
121 xlabel('时间');
122 ylabel('电磁辐射信号 (EMR)');
123 %title('电磁辐射信号时间序列图');
124 legend('干扰数据','Location', 'Best');
125 grid on;
126
127 % hold off; % 停止保持图形
128 set(gca,'fontsize',20)
129
130
131 %%
132 % 分离 EMR, time 和 class 列
133 AE = data1_AE.AE;
134 time_AE = datetime(data1_AE.time, 'InputFormat', 'yyyy-MM-dd
    HH:mm:ss');
135 class_AE = data1_AE.class;
136
137 % 创建图形
138 figure;
139 hold on; % 保持当前图
140
141 % 使用逻辑索引绘制不同类别的数据
142 C_idx = (class_AE == "C"); % 找到C类数据的索引
143 normal_idx = (class_AE == "A"); % 找到正常数据的索引
144 predict_idx= (class_AE == "B");
145 %DE
146 DE_idx= (class_AE == "D/E");
147 plot(time_AE, AE, 'LineWidth', 1); % 蓝色线
148
149
150 plot(time_AE(DE_idx), AE(DE_idx), '*', 'Color', [0.3010 0.7450
    0.9330], 'LineWidth', 0.5)
151 % 绘制 C 类 的电磁辐射信号

```

```

152
153
154 % 绘制正常类（A, B, D）的电磁辐射信号
155 plot(time_AE(normal_idx), AE(normal_idx), 'g*', 'LineWidth',
156       1); % 蓝色线
157
158
159 plot(time_AE(predict_idx), AE(predict_idx), '*', 'Color'
160       , [0.9290 0.6940 0.1250], 'LineWidth', 1); % 紫色线
161
162
163 plot(time_AE(C_idx), AE(C_idx), 'r*', 'LineWidth', 1); % 红色
164       线
165
166 % 设置图形的属性
167 xlabel('时间');
168 ylabel('声发射信号（AE）');
169 %title('电磁辐射信号时间序列图');
170 legend('AE信号', '不正常/停产数据', '正常数据', '前兆数据', '干
171       扰数据', 'Location', 'Best');
172
173 grid on;
174
175 hold off; % 停止保持图形
176 set(gca, 'fontsize', 20)
177 % 在这里你可以继续添加分析 C 类数据的其他代码
178 % 例如：
179 % 绘制 C 类的统计特征（均值、标准差等）
180
181
182 range_C = max(AE(C_idx)) - min(AE(C_idx));
183 mean_C = mean(AE(C_idx));
184 std_C = std(AE(C_idx));
185 med_C = median(AE(C_idx));
186 mod_C = mode(AE(C_idx));
187 cv_C = std_C/mean_C*100;
188
189 disp(['C类数据的峰峰值：', num2str(range_C)]);

```

```

183 disp(['C类数据的均值: ', num2str(mean_C)]);
184 disp(['C类数据的标准差: ', num2str(std_C)]);
185 disp(['C类数据的中位数: ', num2str(mod_C)]);
186 disp(['C类数据的众数: ', num2str(med_C)]);
187 disp(['C类数据的变异系数: ', num2str(cv_C)]);
188 %%
189
190 AE_B_time = time_AE(predict_idx);
191 AE_B_data = AE(predict_idx);
192
193 figure
194 % plot(EMR_B_time(59:736), EMR_B_data(59:736) , '-', 'Color
    ', [0.9290 0.6940 0.1250], 'LineWidth', 1); % 紫色线
195 plot(AE_B_time(1:547), AE_B_data(1:547) , '-', 'Color'
    , [0.9290 0.6940 0.1250], 'LineWidth', 1);
196 % plot(AE_B_time(548:845), AE_B_data(548:845) , '-', 'Color
    ', [0.9290 0.6940 0.1250], 'LineWidth', 1);
197 % plot(EMR_B_time(2544:4665), EMR_B_data(2544:4665) , '-', '
    Color', [0.9290 0.6940 0.1250], 'LineWidth', 1);
198 % 设置图形的属性
199 xlabel('时间');
200 ylabel('声发射信号 (AE)');
201 % title('电磁辐射信号时间序列图');
202 legend('前兆数据', 'Location', 'Best');
203 set(gca, 'fontsize', 20)
204
205 %%
206 AE_feature = zeros(6,4);
207 AE_feature(:,1) = data_feature(AE(C_idx));
208 AE_feature(:,2) = data_feature(AE(normal_idx));
209 AE_feature(:,3) = data_feature(AE(predict_idx));
210 AE_feature(:,4) = data_feature(AE(DE_idx));
211 %%
212 EMR_feature = zeros(6,4);
213 EMR_feature(:,1) = data_feature(EMR(C_idx));

```

```

214 EMR_feature(:,2) = data_feature(EMR(normal_idx));
215 EMR_feature(:,3) = data_feature(EMR(predict_idx));
216 EMR_feature(:,4) = data_feature(EMR(DE_idx));
217 %%
218 %深度学习数据划分
219 EMR_train_X = EMR';
220 EMR_train_Y =categorical(class_EMR)';
221 %%
222 %深度学习数据划分
223 EMR_train_X_ABC = EMR_train_X;
224 EMR_train_X_ABC(DE_idx) = [];
225 EMR_train_Y_ABC = class_EMR;
226 EMR_train_Y_ABC(DE_idx) = [];
227 EMR_train_Y_ABC = categorical(EMR_train_Y_ABC)';
228
229 %%
230 %深度学习数据划分
231
232 AE_train_X_ABC = data1_AE1.AE';
233 AE_train_X_ABC(DE_idx) = [];
234 AE_train_Y_ABC = data1_AE1.class;
235 AE_train_Y_ABC(DE_idx) = [];
236 AE_train_Y_ABC = categorical(AE_train_Y_ABC)';
237 %%
238 EMR_Test = readtable("Attachment2.xlsx",'Sheet','EMR');
239 EMR_Test_X1 = EMR_Test.EMR1;
240
241 %%
242 AE_Test = readtable("Attachment2.xlsx",'Sheet','AE');
243 AE_Test_X =AE_Test.AE3;
244
245
246
247 %%
248 % 加载数据，假设 EMR_Test 已经定义并包含相关数据

```

```

249 % EMR_Test = ... ;
250
251 time_EMR1 = datetime(EMR_Test.time1, 'InputFormat', 'yyyy-MM-
    dd HH:mm:ss');
252 % class_EMR = data1_EMR.class;
253 % 检测离群值
254 outlierMask = isoutlier(EMR_Test.EMR1, 'movmedian',16000,"
    ThresholdFactor",6);
255
256 % 提取所有离群值
257 % outliers_time = EMR_Test(outlierMask, :);
258
259 figure
260 hold on
261 % 绘制正常类 (A, B, D) 的电磁辐射信号
262 plot(time_EMR1, EMR_Test.EMR1, '-', 'LineWidth', 1); % 蓝色线
263 plot(time_EMR1(outlierMask), EMR_Test.EMR1(outlierMask), '*',
    'LineWidth', 1); % 蓝色线
264 % 设置图形的属性
265 xlabel('时间');
266 ylabel('电磁辐射信号 (EMR)');
267 % title('电磁辐射信号时间序列图');
268 legend('电磁辐射信号','干扰数据','Location', 'Best');
269 set(gca,'fontsize',20)
270 hold off
271
272 %%
273 AE_TEST = readtable("Attachment2.xlsx",'Sheet','AE');
274 time_AE1 = datetime(AE_TEST.time3, 'InputFormat', 'yyyy-MM-dd
    HH:mm:ss');
275 %%
276 % class_EMR = data1_EMR.class;
277 % 检测离群值
278 outlierMask = isoutlier(AE_TEST.AE3, 'movmedian',16000,"
    ThresholdFactor",4);

```

```

279
280 % 提取所有离群值
281 % outliers_time = EMR_Test(outlierMask, :);
282
283 figure
284 hold on
285 % 绘制正常类 (A, B, D) 的电磁辐射信号
286 plot(time_AE1, AE_TEST.AE3, '-', 'LineWidth', 1); % 蓝色线
287 plot(time_AE1(outlierMask), AE_TEST.AE3(outlierMask), '*', '
    LineWidth', 1); % 蓝色线
288 % 设置图形的属性
289 xlabel('时间');
290 ylabel('声发射信号 (AE)');
291 % title('电磁辐射信号时间序列图');
292 legend('声发射信号','干扰数据','Location', 'Best');
293 set(gca,'fontsize',20)
294 hold off
295
296
297 AE_OUT_TIME = time_AE1(outlierMask);
298 %%
299 % 求导
300 dt = seconds(diff(time_EMR)); % 计算时间间隔
301 dEMR = diff(EMR); % 计算 EMR 的差分
302
303 % 计算导数 (相对于时间)
304 dEMR_per_time = dEMR ./ dt; % 按时间间隔进行归一化
305
306 % 绘图
307 figure;
308 hold on;
309
310 % 绘制 EMR 信号
311 % plot(time_EMR, EMR, 'b-', 'DisplayName', '原始EMR信号');
312 % 绘制导数信号

```



```

313 % 由于求导后数据长度减小1，需要调整时间向量
314 % plot(time_EMR(2:end), dEMR_per_time, '-', 'DisplayName', '其
    他导数');
315 plot(time_EMR(normal_idx), dEMR_per_time(normal_idx), '*', '
    DisplayName', '正常导数');
316
317 % plot(time_EMR(C_idx), dEMR_per_time(C_idx), '*', '
    DisplayName', '干扰导数');
318 plot(time_EMR(predict_idx), dEMR_per_time(predict_idx), '*', '
    DisplayName', '前兆导数');
319
320 % 设置图形的属性
321 xlabel('时间');
322 ylabel('信号与导数');
323 title('电磁辐射信号及其导数');
324 legend('Location', 'Best');
325 grid on;
326
327 hold off;
328 set(gca, 'fontsize', 20)
329
330 %%
331 data3_EMR = readtable("Attachment3.xlsx", 'Sheet', 'EMR');
332 data3_AE = readtable("Attachment3.xlsx", 'Sheet', 'AE');

```

pro2.m

```

1
2 YPred_EMR = classify(net_ABC, Lstm_Forecast_Value');
3
4 pro2_C_idx = (YPred_EMR == "C"); % 找到C类数据的索引
5 pro2_normal_idx = (YPred_EMR == "A"); % 找到正常数据的索引
6 pro2_predict_idx = (YPred_EMR == "B");
7 figure
8 plot(YPred_EMR, '.-')
9 % plot(EMR_Test.time2, YPred_EMR, '.-')

```

```

10 % plot(EMR_Test.time2,EMR_Test.EMR2,'.r-')
11 hold on
12 % plot(EMR_Test.time2(pro2_predict_idx), EMR_Test.EMR2(
    pro2_predict_idx), '*', 'LineWidth', 1); % 紫色线
13 % plot(YTest{1})
14 hold off
15 xlabel("时间")
16 ylabel("信号类型")
17 % title("Predicted Activities")
18 legend("Predicted" )
19 set(gca, 'fontsize', 20)
20 %
21 %
22 % figure
23 % plot(EMR_Test.EMR2)
24 %
25 % xlabel("时间")
26 % ylabel('电磁辐射信号 (EMR)');
27 %%
28
29 YPred_AE = classify(net_AE_ABC, AE_Test_X');
30
31 pro2_C_idx = (YPred_AE == "C"); % 找到C类数据的索引
32 pro2_normal_idx = (YPred_AE == "A"); % 找到正常数据的索引
33 pro2_predict_idx = (YPred_AE == "B");
34 figure
35 plot(YPred_AE, '.-')
36 % plot(EMR_Test.time2, YPred_EM, '.-')
37 % plot(EMR_Test.time2, EMR_Test.EMR2, '.r-')
38 hold on
39 % plot(EMR_Test.time2(pro2_predict_idx), EMR_Test.EMR2(
    pro2_predict_idx), '*', 'LineWidth', 1); % 紫色线
40 % plot(YTest{1})
41 hold off
42 xlabel("时间")

```

```

43 ylabel("信号类型")
44 % title("Predicted Activities")
45 legend("Predicted" )
46 set(gca,'fontsize',20)

```

pro3.m

```

1 numFeatures = 1;
2 numHiddenUnits = 100;
3 numClasses = 3;
4
5 layers = [ ...
6     sequenceInputLayer(numFeatures)
7     lstmLayer(numHiddenUnits,'OutputMode','sequence')
8     fullyConnectedLayer(numClasses)
9     softmaxLayer
10    classificationLayer];
11 Specify the training options. Set the solver to 'adam'. Train
    for 60 epochs. To prevent the gradients from exploding, set
    the gradient threshold to 2.
12 options = trainingOptions('adam', ...
13     'MaxEpochs',200, ...
14     'InitialLearnRate',0.001 , ...
15     'GradientThreshold',2, ...
16     'Verbose',0, ...
17     'Plots','training-progress');
18 Train the LSTM network with the specified training options
    using trainNetwork. Each mini-batch contains the whole
    training set, so the plot is updated once per epoch. The
    sequences are very long, so it might take some time to
    process each mini-batch and update the plot.
19 % net = trainNetwork(EMR_train_X,EMR_train_Y,layers,options);
20 % net_ABC = trainNetwork(EMR_train_X_ABC,EMR_train_Y_ABC,
    layers,options);
21 net_AE_ABC = trainNetwork(AE_train_X_ABC,AE_train_Y_ABC,layers
    ,options);

```

```

22 %%
23 %深度学习数据划分
24 EMR_train_X = EMR';
25 EMR_train_Y =categorical(class_EMR)';
26 %%
27 %深度学习数据划分
28 EMR_train_X_ABC = EMR_train_X;
29 EMR_train_X_ABC(DE_idx) = [];
30 EMR_train_Y_ABC = class_EMR;
31 EMR_train_Y_ABC(DE_idx) = [];
32 EMR_train_Y_ABC = categorical(EMR_train_Y_ABC)';
33
34 %%
35 %深度学习数据划分
36
37 AE_train_X_ABC = data1_AE1.AE';
38 AE_train_X_ABC(DE_idx) = [];
39 AE_train_Y_ABC = data1_AE1.class;
40 AE_train_Y_ABC(DE_idx) = [];
41 AE_train_Y_ABC = categorical(AE_train_Y_ABC)';
42 %%
43 EMR_Test = readtable("Attachment2.xlsx", 'Sheet', 'EMR');
44 EMR_Test_X1 = EMR_Test.EMR1;
45
46 %%
47 AE_Test = readtable("Attachment2.xlsx", 'Sheet', 'AE');
48 AE_Test_X =AE_Test.AE3;
49
50
51
52 %%
53 % 加载数据，假设 EMR_Test 已经定义并包含相关数据
54 % EMR_Test = ... ;
55
56 time_EMR1 = datetime(EMR_Test.time1, 'InputFormat', 'yyyy-MM-
```

```

        dd HH:mm:ss');
57 % class_EMR = data1_EMR.class;
58 % 检测离群值
59 outlierMask = isoutlier(EMR_Test.EMR1, 'movmedian',16000,"
        ThresholdFactor",6);
60
61 % 提取所有离群值
62 % outliers_time = EMR_Test(outlierMask, :);
63
64 figure
65 hold on
66 % 绘制正常类（A, B, D）的电磁辐射信号
67 plot(time_EMR1, EMR_Test.EMR1, '-', 'LineWidth', 1); % 蓝色线
68 plot(time_EMR1(outlierMask), EMR_Test.EMR1(outlierMask), '*',
        'LineWidth', 1); % 蓝色线
69 % 设置图形的属性
70 xlabel('时间');
71 ylabel('电磁辐射信号 (EMR)');
72 % title('电磁辐射信号时间序列图');
73 legend('电磁辐射信号','干扰数据','Location', 'Best');
74 set(gca,'fontsize',20)
75 hold off
76
77 %%
78 AE_TEST = readtable("Attachment2.xlsx",'Sheet','AE');
79 time_AE1 = datetime(AE_TEST.time3, 'InputFormat', 'yyyy-MM-dd
        HH:mm:ss');
80 %%
81 % class_EMR = data1_EMR.class;
82 % 检测离群值
83 outlierMask = isoutlier(AE_TEST.AE3, 'movmedian',16000,"
        ThresholdFactor",4);
84
85 % 提取所有离群值
86 % outliers_time = EMR_Test(outlierMask, :);

```

```

87
88 figure
89 hold on
90 % 绘制正常类 (A, B, D) 的电磁辐射信号
91 plot(time_AE1, AE_TEST.AE3, '-', 'LineWidth', 1); % 蓝色线
92 plot(time_AE1(outlierMask), AE_TEST.AE3(outlierMask), '*', '
    LineWidth', 1); % 蓝色线
93 % 设置图形的属性
94 xlabel('时间');
95 ylabel('声发射信号 (AE)');
96 % title('电磁辐射信号时间序列图');
97 legend('声发射信号','干扰数据','Location', 'Best');
98 set(gca,'fontsize',20)
99 hold off
100
101
102 AE_OUT_TIME = time_AE1(outlierMask);
103 %%
104 % 求导
105 dt = seconds(diff(time_EMR)); % 计算时间间隔
106 dEMR = diff(EMR); % 计算 EMR 的差分
107
108 % 计算导数 (相对于时间)
109 dEMR_per_time = dEMR ./ dt; % 按时间间隔进行归一化
110
111 % 绘图
112 figure;
113 hold on;
114
115 % 绘制 EMR 信号
116 % plot(time_EMR, EMR, 'b-', 'DisplayName', '原始EMR信号');
117 % 绘制导数信号
118 % 由于求导后数据长度减小1, 需要调整时间向量
119 % plot(time_EMR(2:end), dEMR_per_time, '-', 'DisplayName', '其
    他导数');

```