



# AETA地震预测AI算法大赛



Helloworld 队

武汉大学

2021/1/28

# 目录

contents

**PART 01 / 背景简介**

**PART 02 / 地震预测方法研究**

**PART 03 / 总结与展望**

01

PART ONE

背景简介

# 01/ 背景简介

## 赛题简介

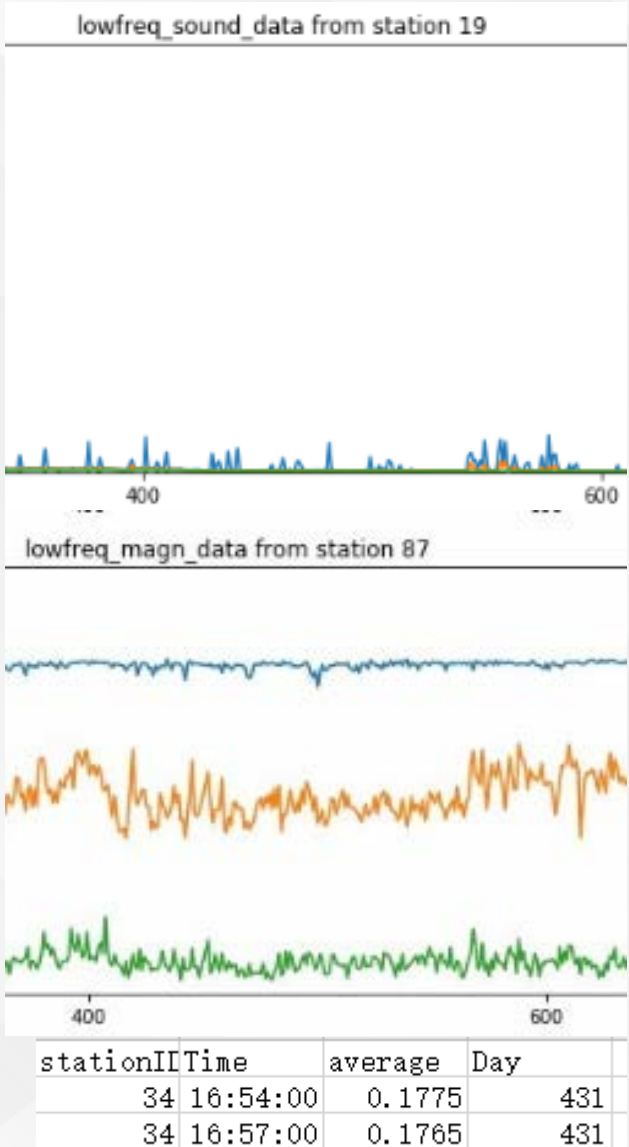
选手通过分析电磁和地声数据，发现其中与临震相关的异常信号和特征，挖掘前兆观测数据与地震三要素的相关性，并基于历史观测数据和地震目录，构建地震预测模型。

比赛要求：使用四川云南地区，某一周的电磁和地声数据预测下一周该地区的地震事件（时间，经纬度，震级）。

## 赛题分析

**数据特征构建：**原数据是典型的时序数据，其大量的信息藏在它的结构中，不仅仅体现在其数值上。构建新特征，并寻找其中的强特征是比赛的难点之一。

**地震机制复杂：**地震发生前，当地的电磁和地声会出现某些现象，但是不同地区不同的地震前出现的现象没有明确的相通性，如何细分这些现象，并建立这些现象与地震的关联关系是比赛的难点。

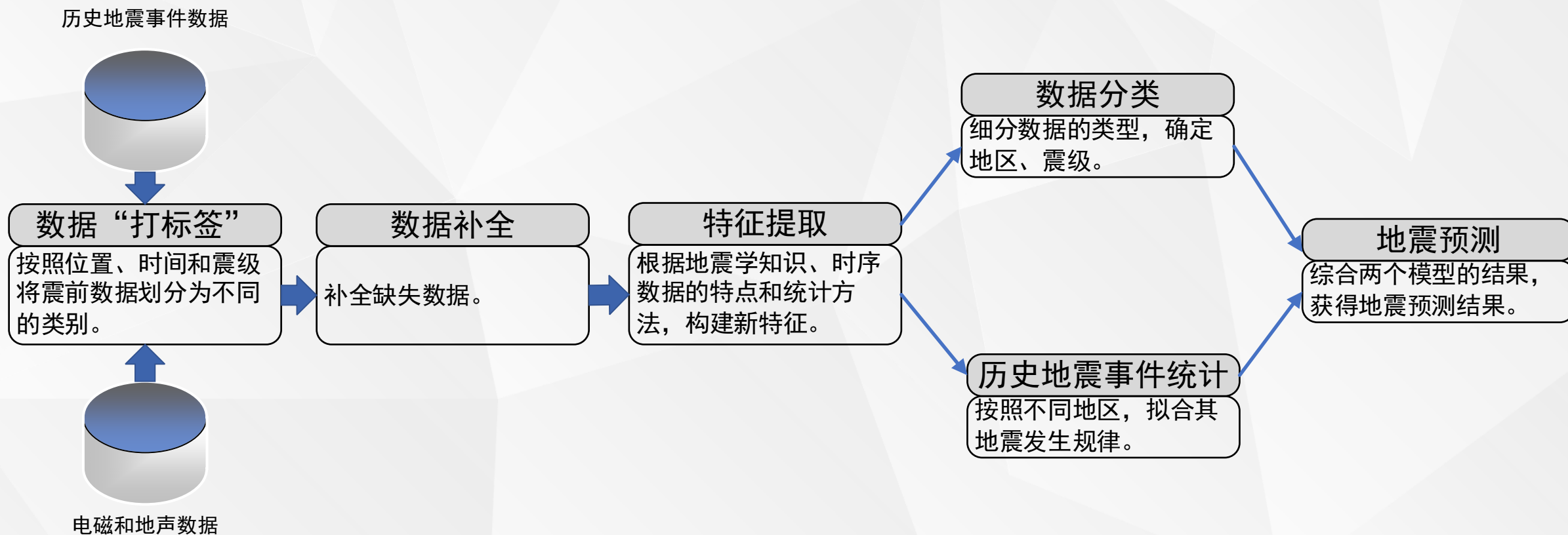


02

PART TWO

# 地震预测方法研究

## 02 地震预测方法研究——整体思路与挖掘方案



# 02/ 地震预测方法研究——数据 “打标签”

按照地区划分：鲜水河地震带、安宁河-则木河地震带、金沙江地震带、松潘-较场地震带、龙门山地震带、理塘地震带、木里-盐源地震区、名山-马边-昭通地震带、小江地震带、中甸—南涧地震带、大关—马边地震带、澜沧—耿马地震带、泸水—腾冲地震带、普洱—宁洱地震带、通海—石屏地震带。

按时间划分：地震事件发生前一周

按震级划分：经过统计以上区域的三年内历史地震多发在3~6之间，所以按照震级影响范围和频次划分为四档。

地区	时间范围	震级	label	
鲜水河地震带 (区域1)	地震事件发生前一周	0~3.4	1_0	0
		3.5~4.5	1_1	1
		4.6~5.5	1_2	2
		5.6~8.0	1_3	3

# 02/ 地震预测方法研究——数据补全

## 线性插值

以 1 天数据的 10%作为  
阈值，如果连续缺失 10%  
以内的数据，将采用线性  
插值法来对缺失的数据进  
行补全，如果连续缺失  
10%以上的数据，插值补  
全会引入较大误差，则删  
掉当天数据。





# 02/ 地震预测方法研究——特征提取

## 衍生特征

衍生特征：计算电磁和地声数据的临近差值、每天SRSS波存在与否、每天SRSS波幅值等。

## 统计特征

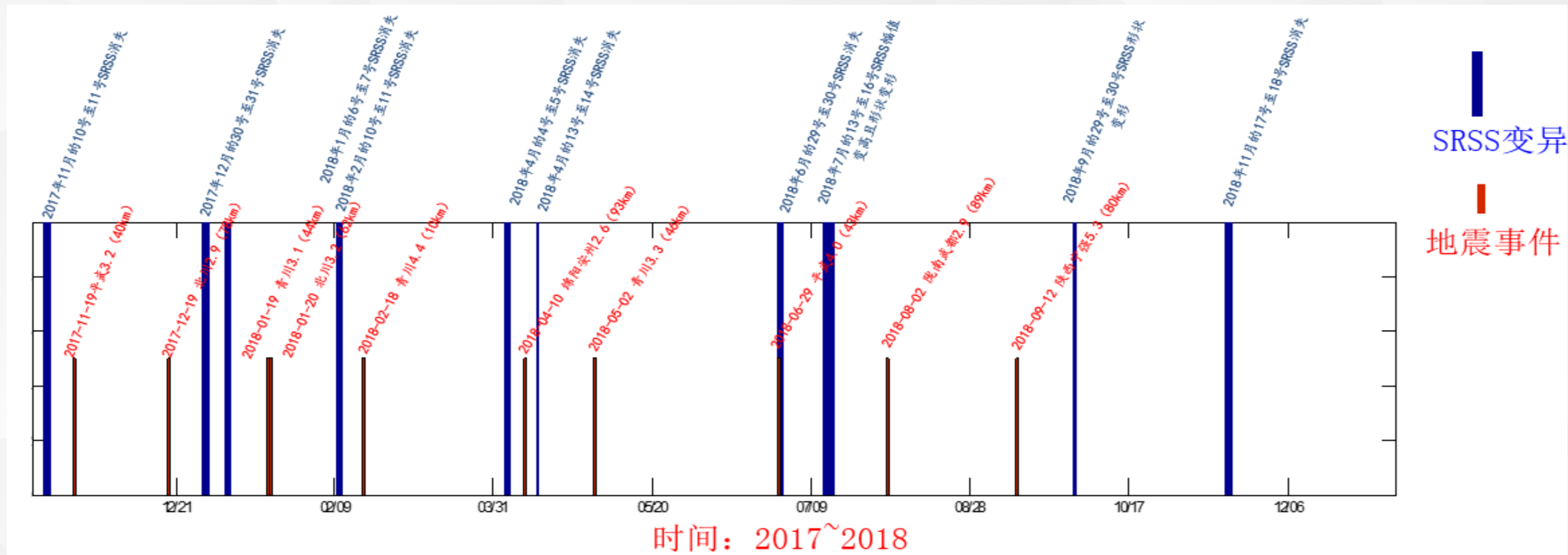
统计特征：

统计电磁和地声数据的衍生特征及基础特征一天内的最大值，最小值，均值，方差，偏度，峰度，均方根，波形因子，峰值因子，脉冲因子，裕度因子，最大值与最小值的差、滑动四分位数，以及变异系数。

## 额外衍生特征

额外衍生出来的特征：通过用随机森林的特征重要性筛选方法选出特征重要性排前50的特征，两两交叉相乘得出新特征。

# 02 地震预测方法研究——特征提取



## SRSS波

SRSS 波形与地震有一定的相关性，SRSS波及其变异是重要的临震信号。

图为青川房石镇 AETA 台站 SRSS 波变异与地震对应关系的部分示例。

# 02 地震预测方法研究——特征提取

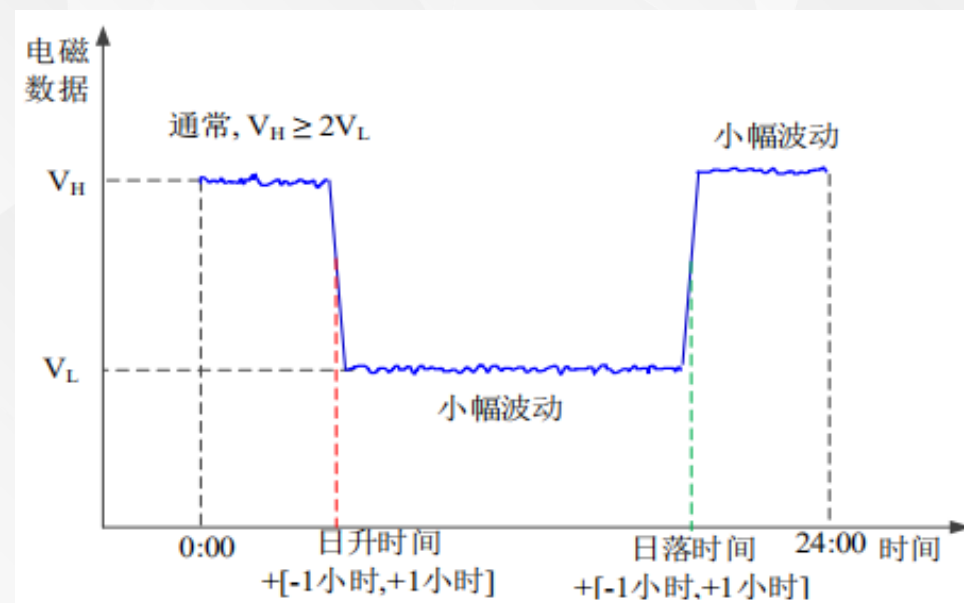
## SRSS波

### SRSS波定义：

每日数据按照小时统计其最大值，若前一个小时的最大值（ $V_H$ ）明显高于后一个小时的最大值（ $V_L$ ）2 倍或以上，则认为当天存在SRSS波。

### SRSS波-特征提取：

- ① 以天为滑窗的范围，数据标准化。
- ② 筛选符合SRSS波特点的数据，定义为SRSS波“模板”
- ③ 检测SRSS波

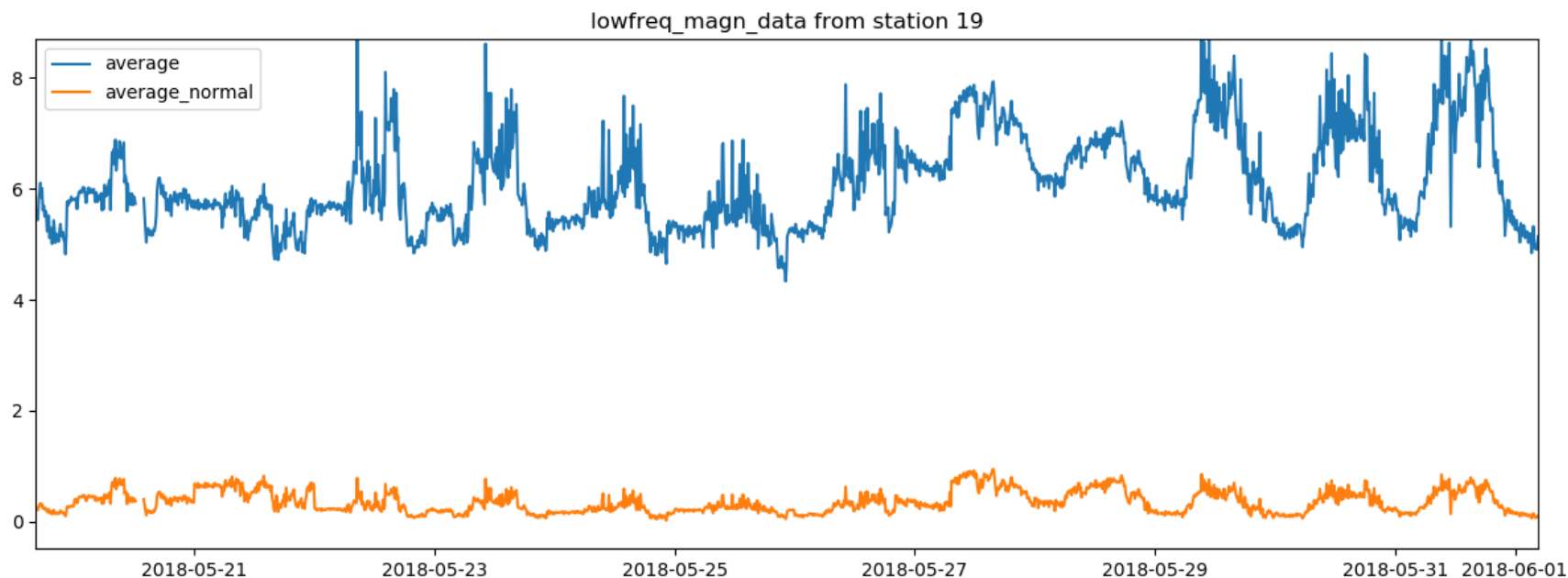


# 02/ 地震预测方法研究——特征提取

## SRSS波-预处理

加窗Max-Min标准化:

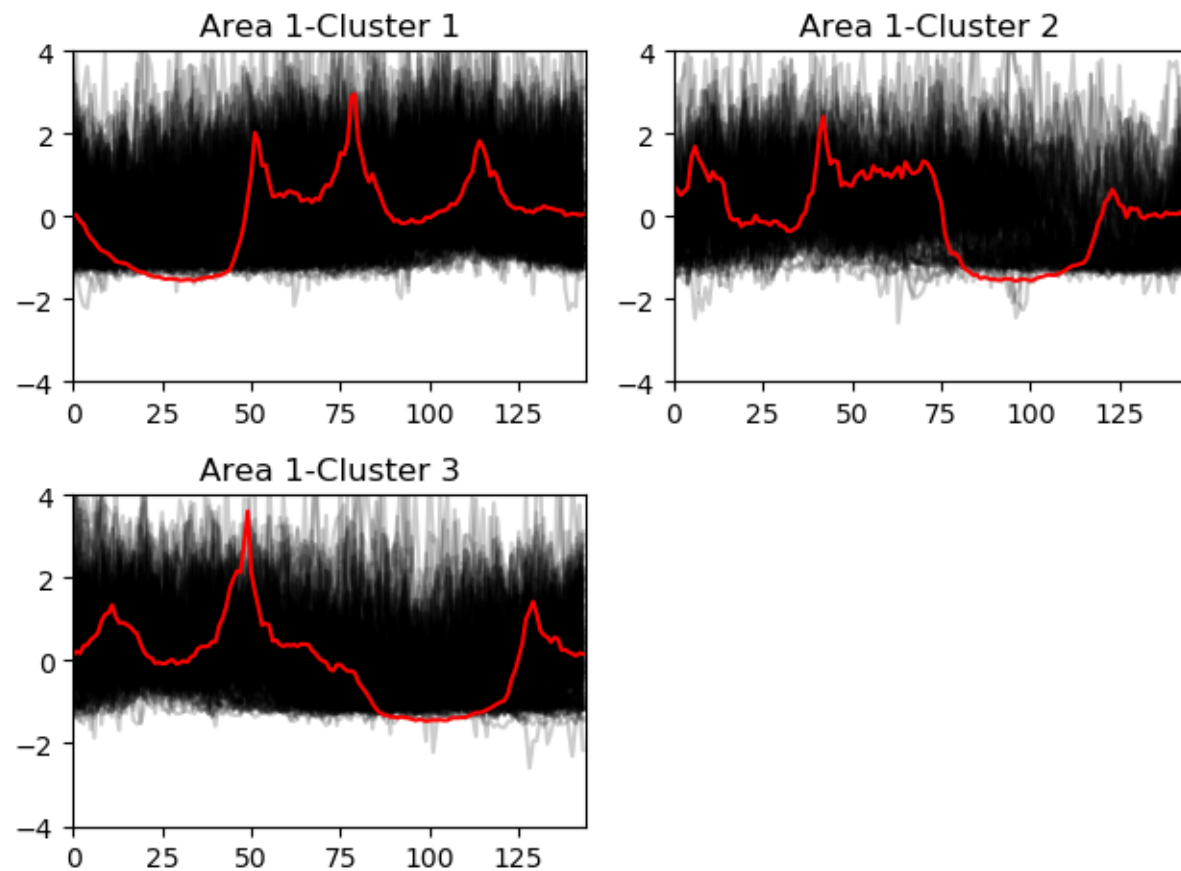
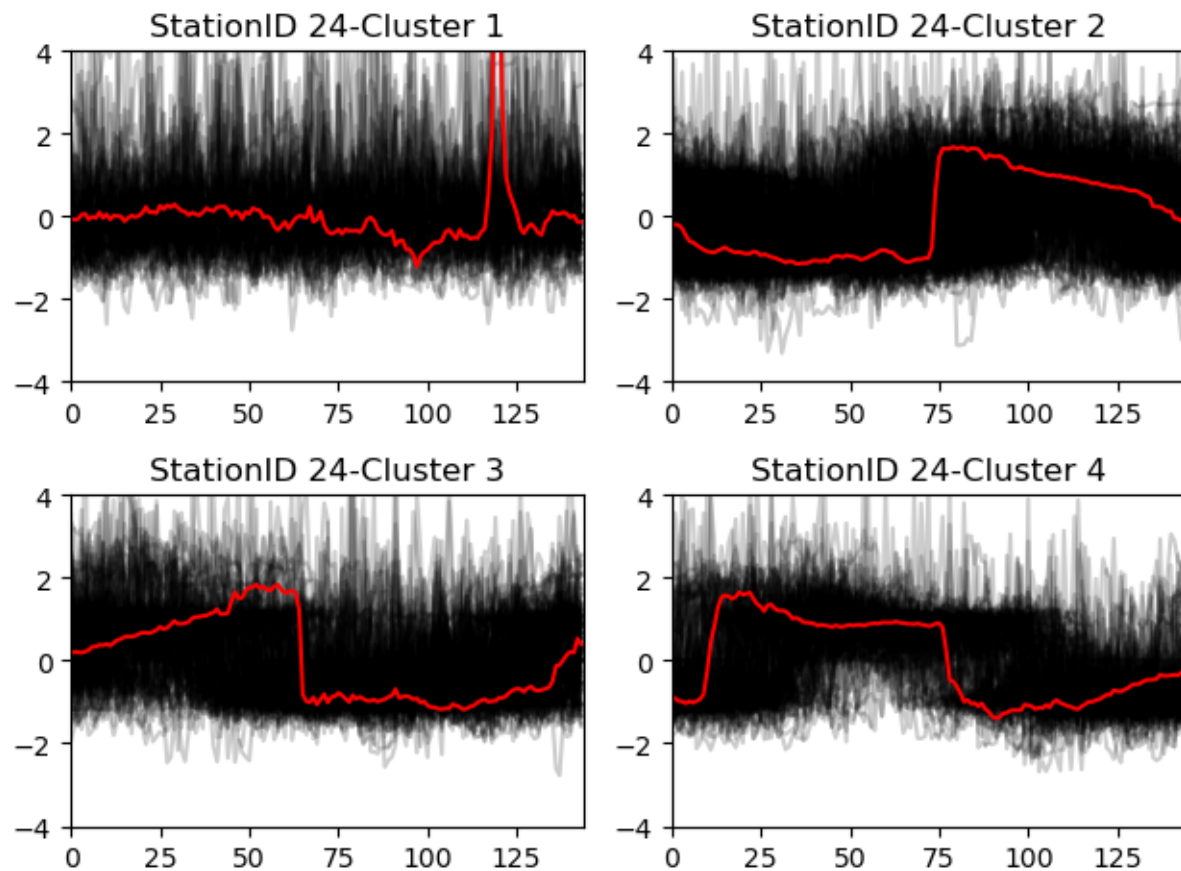
以天为窗口单位，利用每天数据中的极大值、极小值对该窗口内的时间序列进行标准化处理。



# 02 地震预测方法研究——特征提取

## 定义SRSS波

分析台站数据 → 分析区域内数据 → 定义SRSS波模板

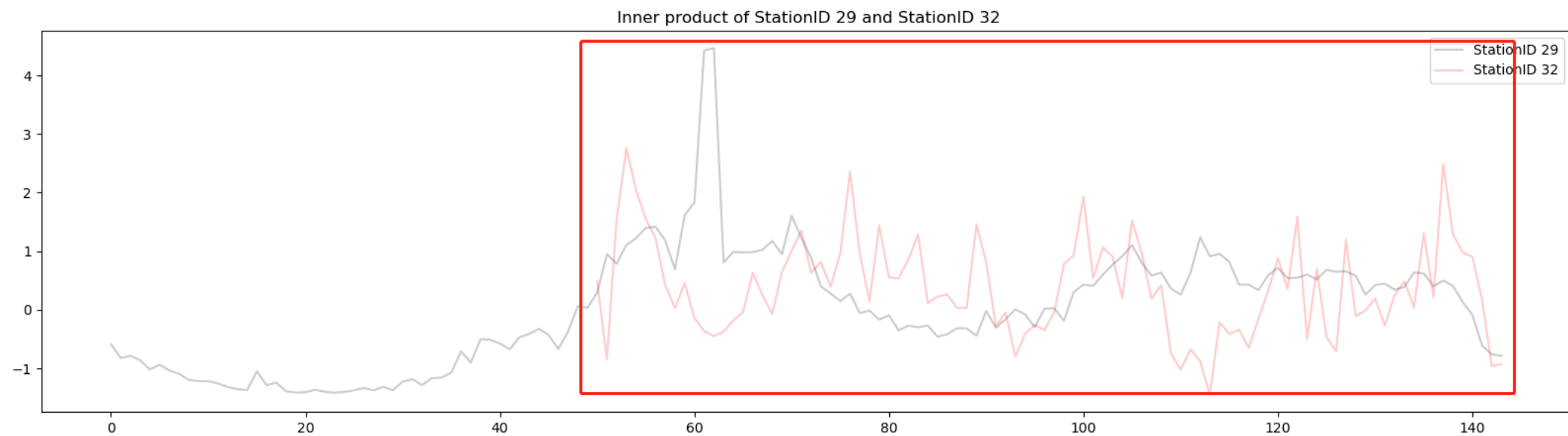


# 02 地震预测方法研究——特征提取

## 检测SRSS波

互相关测度：

按天为单位，计算每一个序列与SRSS波模板之间滑动内积，以其最大内积为两条序列的相关性度量。



# 02/ 地震预测方法研究——数据分类

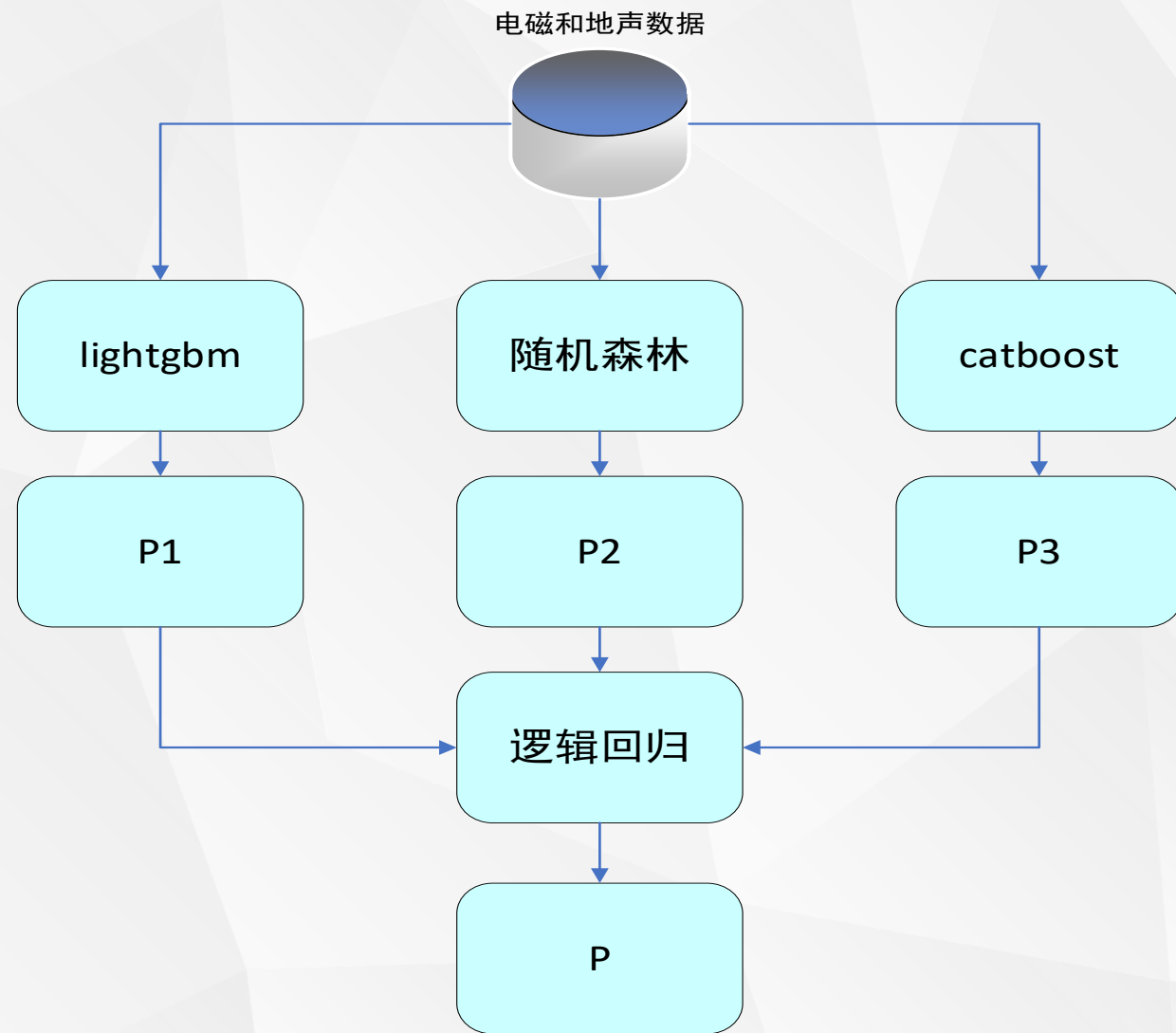
## Stacking

使用Stacking方法进行建模，第一层基模型为lightgbm、随机森林和catboost，第二层使用逻辑回归。

**Classification Model**

**Prediction**

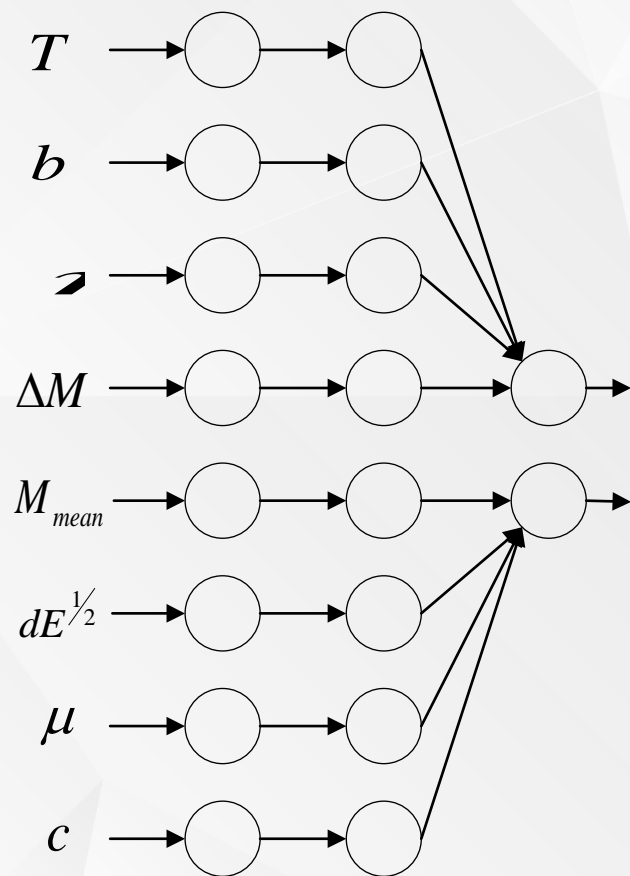
**Final Prediction**



## 02 地震预测方法研究——历史地震事件统计

# 古登堡-理查德定律

$$\lg N = a - bM \quad \rightarrow \quad F = \{b, \eta, \Delta M, T, \mu, c, dE^{1/2}, M_{mean}, Type\}$$

[illegible]





PART FOUR

# 总结与展望

# 05/ 总结

## ➤ 基于时序特点和统计方法，构建新特征，精细描述数据

基于统计和时序数据的处理方法，构建新特征，并在此基础上对特征进行融合构建额外的衍生特征，丰富特征空间。

## ➤ 充分补全缺失数据

利用数据的物理特性，精细的补全了缺失的数据。

## ➤ 使用集成学习构建预测模型

有层次的融合lightgbm、随机森林和catboost，在其上在训练一个次学习器（逻辑回归），用于组织利用基学习器的答案。在数据特征与数据标签之间关系不明确的情况下，使用Stacking可以有效地挖掘数据与类别之间的关系。

# 05/ 总结——下一步工作

## ➤ 基于时序数据的特点，采用信号处理方式和时间序列分割方法

信号处理的方法，希尔伯特变换，小波变换等方式处理时序数据，获取时序数据在频域的特点；

采用时序序列分割的方式，将数据分段，精细刻画数据，使数据成为更易操作的向量，从而挖掘其中的波形的分布规律。

## ➤ 针对不同的数据缺失情况，采用不同的数据补全方式

邻近值填充、回归法、拉格朗日插值法、分段插值法、样条插值法等

**感谢各位老师的指导！**