

szpianpian团队AETA比赛分享

一、处理方法

- 1、由于数据是精确到分钟的检测值，数据量庞大，因为自然环境、人为因素等作用导致监测结果中存在少量缺失值。
- 2、删除数据缺失严重的部分station数据（如station48等）
用经纬度临近station数据代替

二、特征提取（地声数据）

特征选取标准：

1. 可解释性
2. 可泛化
3. 训练时长

三、特征提取

- 1、'magnitude' (震级)
- 2、'depth' (震源深度)
- 3、'time' (发生时间)

(经反复测试, 震级和震源深度的p值显著, 受制于train dataset的天数限制, 如果数据更多, 结果的统计显著性会更好)

(中国地震网数据 <http://www.ceic.ac.cn/speedsearch>)

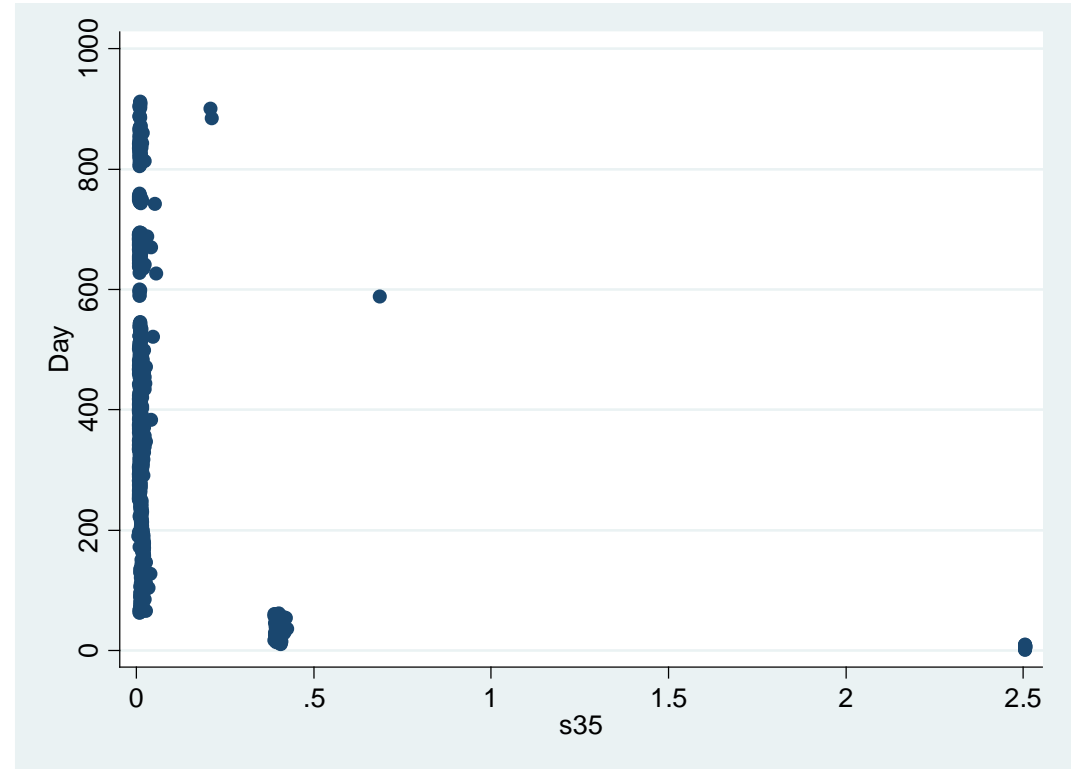
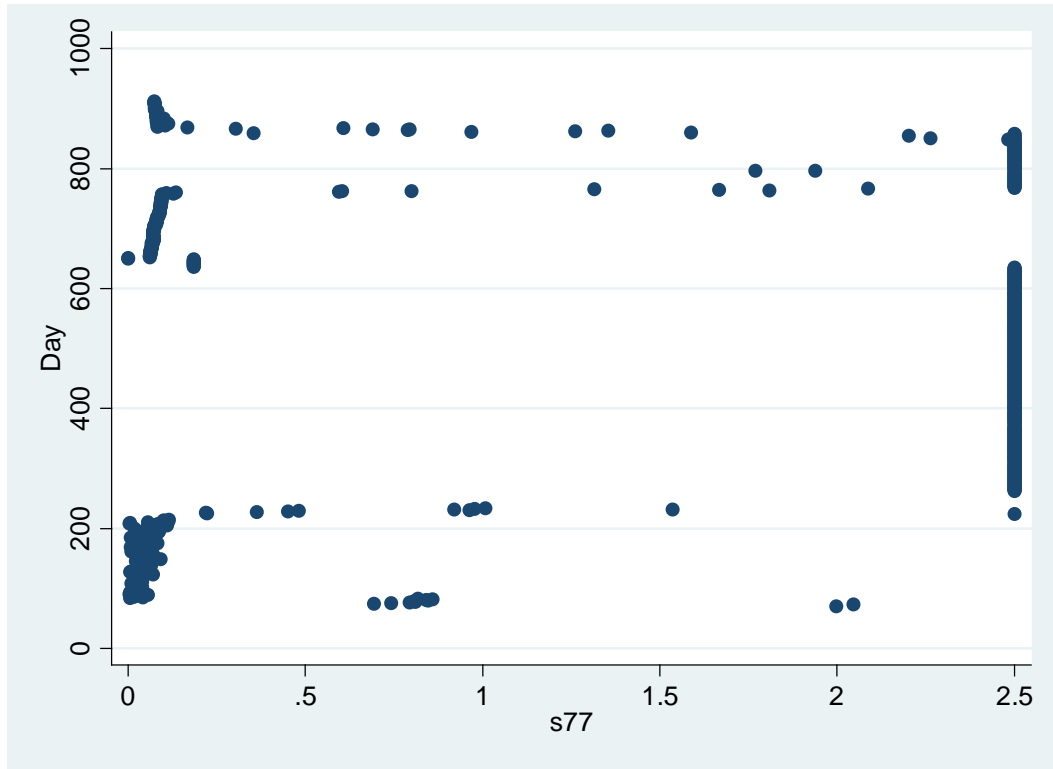
四、样本集

- 选取地声数据集
- 降噪（侧重夜间数据）
- 替代地声值没有变化或缺失及干扰严重的station数据
- 结合中国地震网数据 <http://www.ceic.ac.cn/speedsearch>

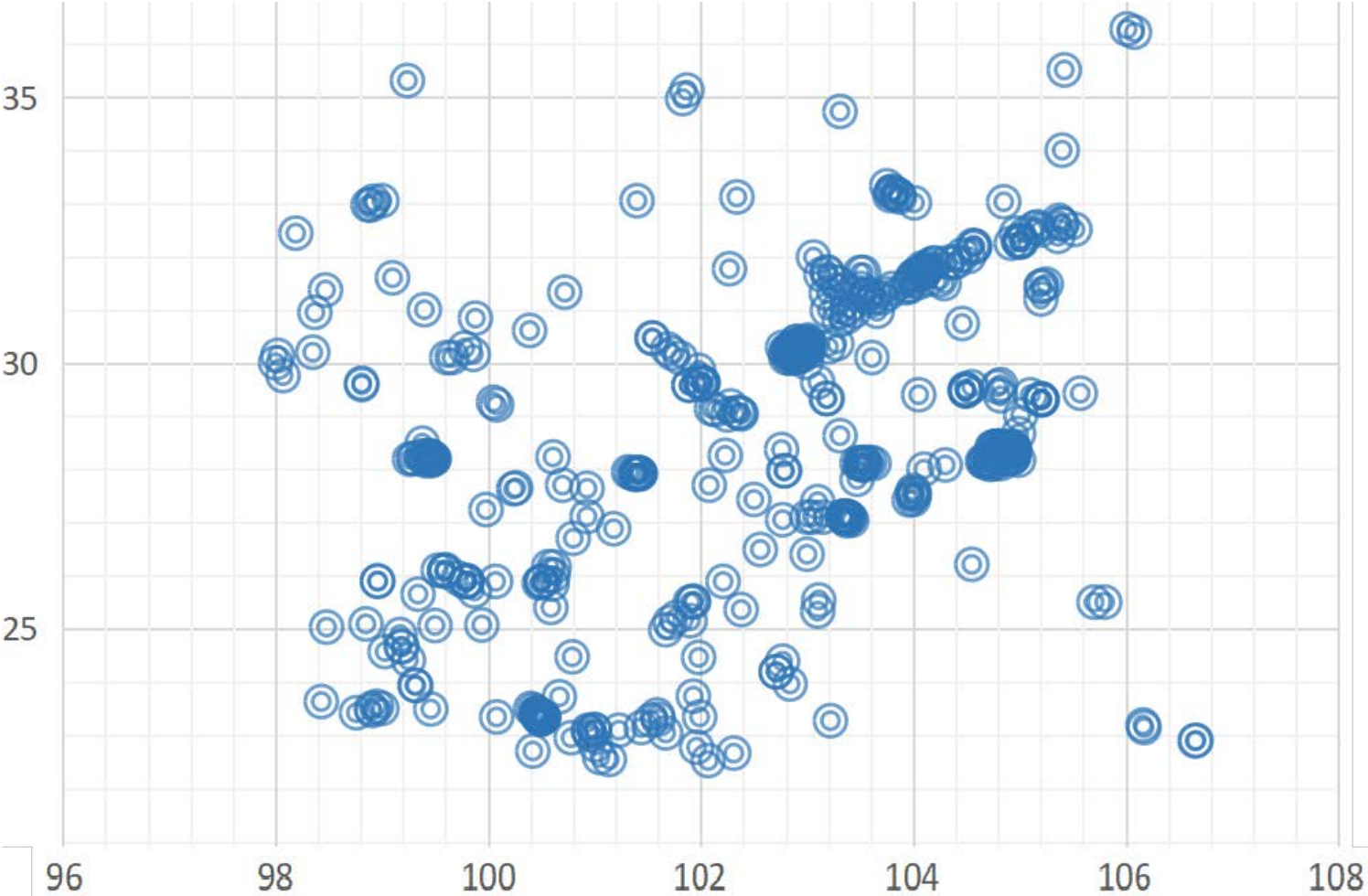
李帛珊，李金 译．2020．基于回归算法和大数据云基础设施的美国加利福尼亚州地震预测．*世界地震译丛*．51(1):33 - 53．

1970~2017 美国加利福尼亚州的地震事件: ANSS复合地震目录

五、样本数据干扰程度比较：以station 77 和 35为例



六、震中分布

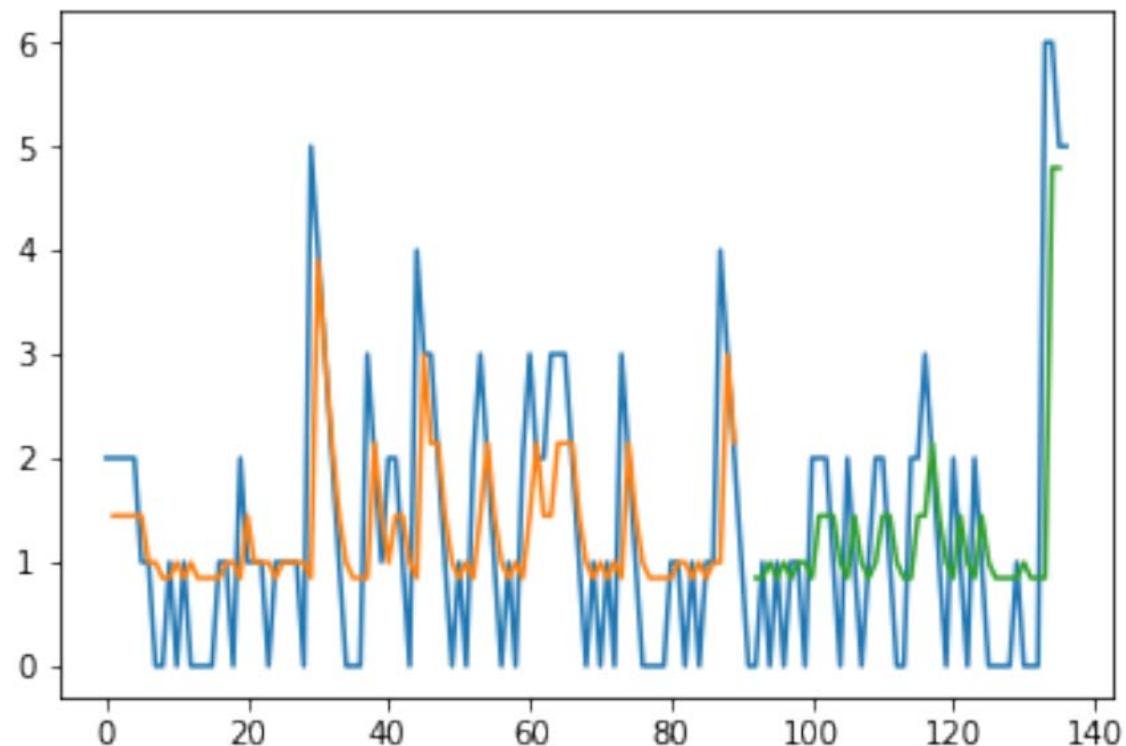
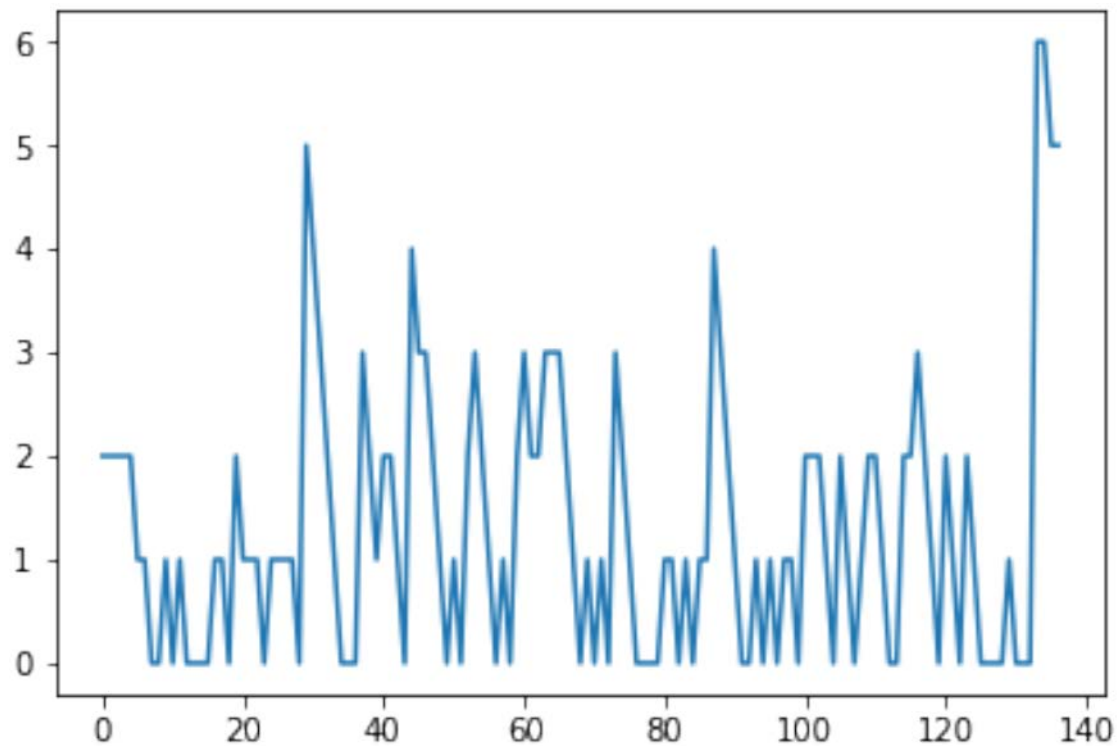


七、预测模型-算法和模型说明：

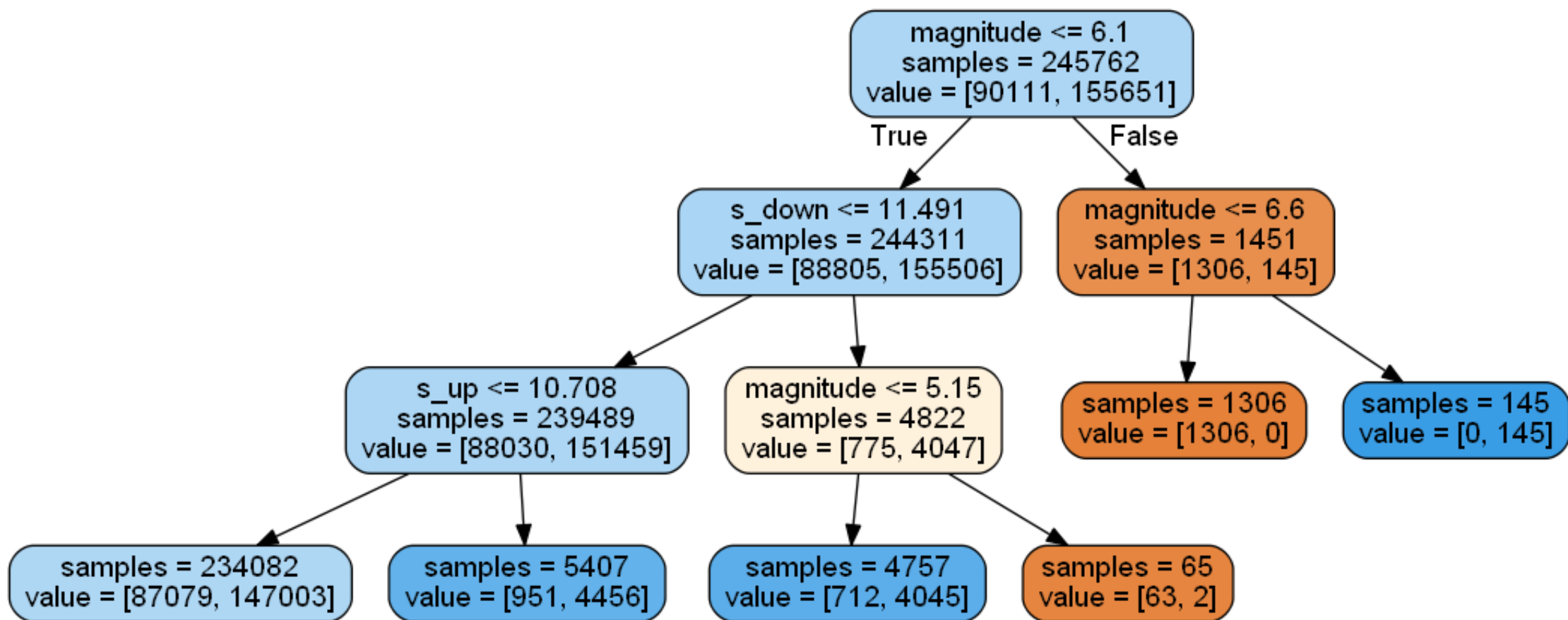
混合模型：回归模型，长短记忆神经网络（Long Short -Term Memory, LSTM），决策树

- LSTM 是一种循环神经网络，具有极强的非线性拟合能力，在训练间缓解反向传播过程中的梯度消减。
- 回归和决策树模型有很好的解释性。

八、LSTM模型验证



九、决策树预测（以部分监测站点为例）



十、检验结果

8次预测震中较小偏离（1-16周共13次>3.5级震）

		偏离震中 距离(km)	震中	震级
1	2020/05/18-2020/05/24	125.9	云南昭通市巧家县	5
2	2020/05/25-2020/05/31	132.5	四川乐山马边县	3.8
4	2020/06/08-2020/06/14	21.9	四川宜宾市珙县	4
7	2020/06/29-2020/07/05	121.103	贵州毕节市赫章县	4.5
8	2020/07/06-2020/07/12	153.47	云南昆明市东川区	4.2
10	2020/07/20-2020/07/26	116.525	云南普洱市澜沧县	4.1
12	2020/08/03-2020/08/09	31.98	四川宜宾市长宁县	3.9
13	2020/08/10-2020/08/16	127.489	四川宜宾市兴文县	3.8

谢谢！

(gezhan@gmail.com)