

三王一张

2020/10/16

# 目录

- 背景介绍
- 算法和模型
- 结果分析

# 背景

- 我国是地震多发的国家，断裂带分布广泛。地震，尤其是大地震，一旦在人们毫无察觉的情况下发生在人口密集区，将对人们的生命财产造成难以估量的损失。围绕地震预测预报问题的求解而开展的前兆观测、前兆相关性分析、前兆机理研究、地震三要素预测模型等研究工作是非常具有挑战性，同时也是非常有科学价值和社会意义的。
- “AETA地震预测AI算法大赛”，旨在通过创新算法挖掘前兆观测数据与地震三要素的相关性，发现与临震相关的异常信号和特征，并基于历史观测数据和地震目录，构建地震预测模型，期望推动地震预测预报科学问题的求解。同时也希望通过本次大赛，引起更多社会各界人士的关注和参与，将更多的新技术新方法应用于地震预测预报中。
- 本次大赛提供的数据包括AETA地震监测预测系统在川滇实验场进行近3年的电磁和地声观测数据，以及地震目录。
- 本次大赛重点评价特征提取和样本构建方法的创新性和先进性，以及地震预测模型的适用性和准确性，鼓励通过创新算法解决地震预报的科学问题。
- 通过一定时间区间内的观测数据的分析和挖掘，结合对应时间区间的全球地震目录，构建地震预测模型。然后输入特定时间范围内的观测数据，对对应时间区间内的目标区域（**北纬22~34，东经98~107**）的 $\geq 3.5$ 级的地震事件进行地震三要素的预测，并根据实际地震三要素进行准确度检验。
- 以测试数据集开始的那天起，**每7天给出未来7天是否有 $\geq 3.5$ 级以上的地震的预测（Y/N）**，预测的地理范围为（ $22.00^{\circ}$  N~ $34.00^{\circ}$  N， $98.00^{\circ}$  E ~ $107.00^{\circ}$  E），是否有震为是否有 $\geq 3.5$ 级以上的地震，震中给出具体的经纬度（**XX.XX, XX.XX**），震级为**Ms**。

# 算法和模型

- Light GBM

预测是否有地震，以及地震震级的大小范围；

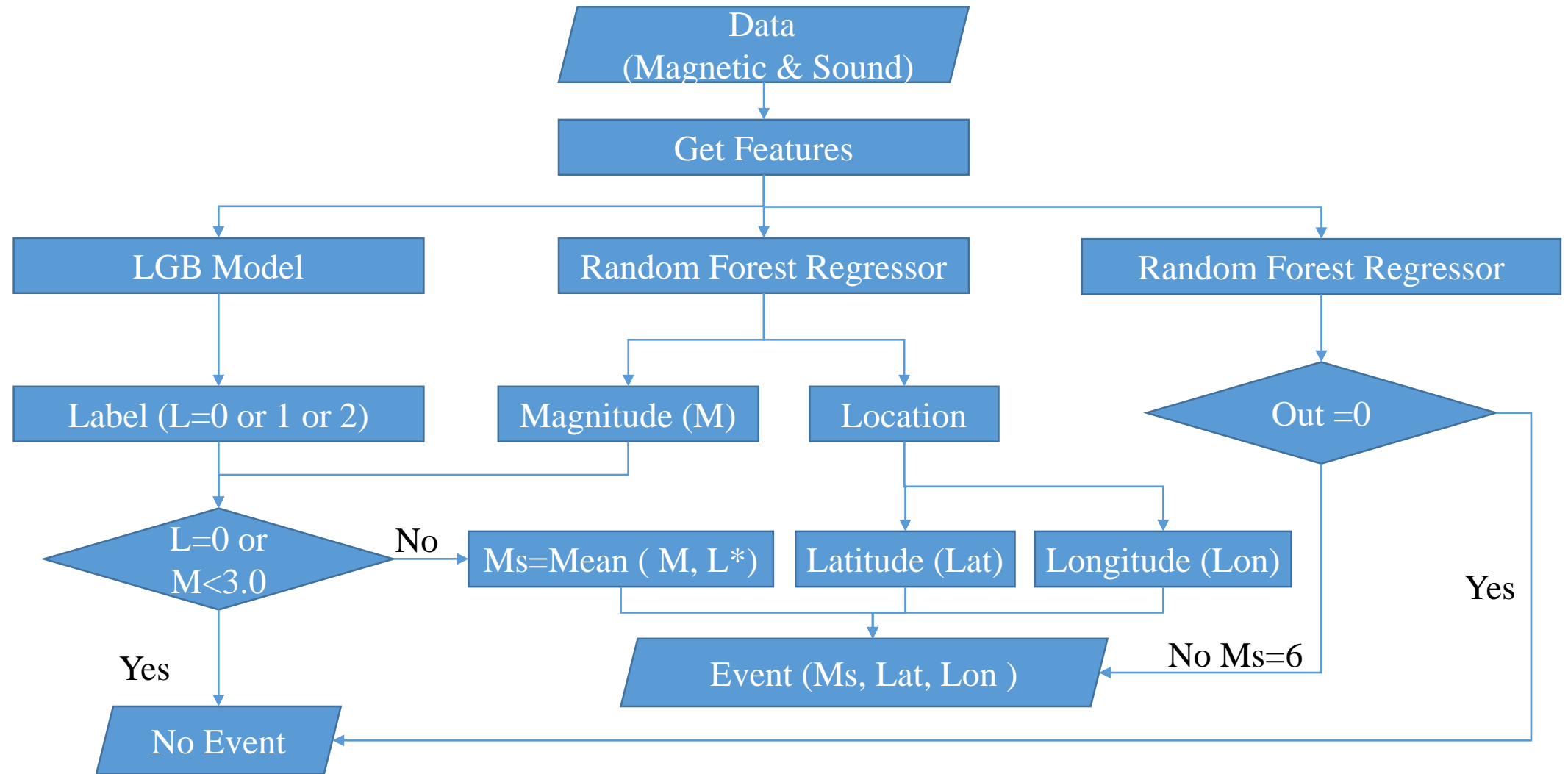
- Random forest

预测震级的大小，地震的位置；

- Random forest

判断是否为大震或者无震；

# Workflow



# 特征提取

特征	电磁数据	地声数据
均值	Magn_average	Sound_average
差值	Magn_diff	Sound_diff
每天的最大值	Magn_day_max	Sound_day_max
每天的最小值	Magn_day_min	Sound_day_min
每天的平均值	Magn_day_mean	Sound_day_mean
每天的最大值与最小值的差值	Magn_day_max_min	Sound_day_max_min
一周内每天的最大值的均值	Magn_day_max_mean	Sound_day_max_mean
一周内每天的最小值的均值	Magn_day_min_mean	Sound_day_min_mean
一周内每天的平均值的最大值	Magn_day_mean_max	Sound_day_mean_max
一周内每天的平均值的最小值	Magn_day_mean_min	Sound_day_mean_min
偏度	Magn_skew	Sound_skew
方差	Magn_var	Sound_var
差值的偏度	Magn_diff_skew	Sound_diff_skew
差值的方差	Magn_diff_var	Sound_diff_var

# LGB

Parameters:

- 'num\_leaves': 48,
- 'learning\_rate': 0.05,
- "boosting": "rf",
- 'objective': 'multiclass',
- "feature\_fraction": 0.6,
- "bagging\_fraction": 0.6,
- "bagging\_freq": 2,
- 'num\_class': 3,
- "lambda\_l1": 0.05,
- "lambda\_l2": 0.05,
- "nthread": -1,
- 'min\_child\_samples': 10,
- 'max\_bin': 200

Label = { 0 , 1 , 2 }

0:  $M_s < 3.5$

1:  $M_s < 6.0$

2:  $M_s \geq 6.0$

Weight = { 1 , 4.5 , 10 }

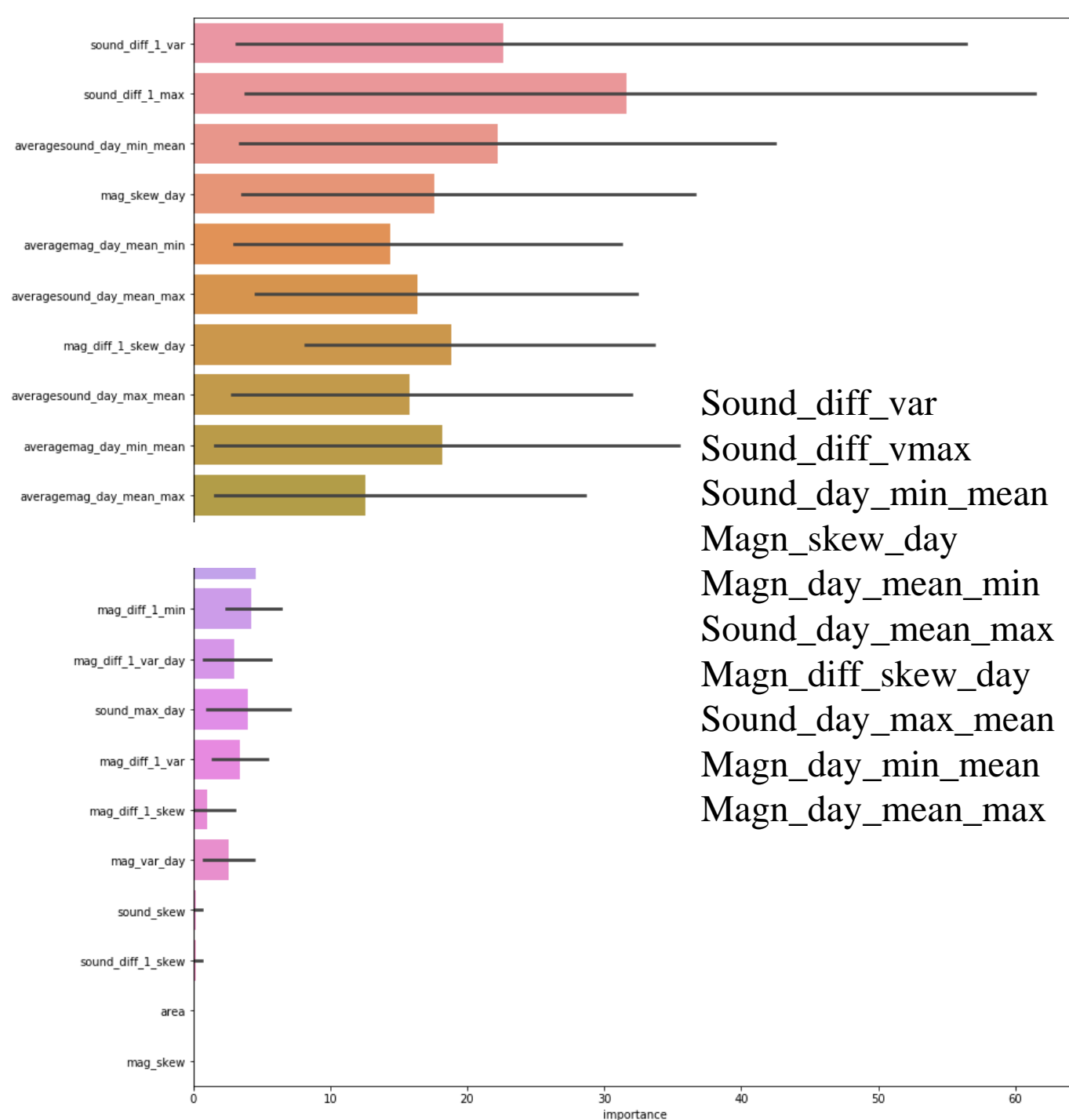
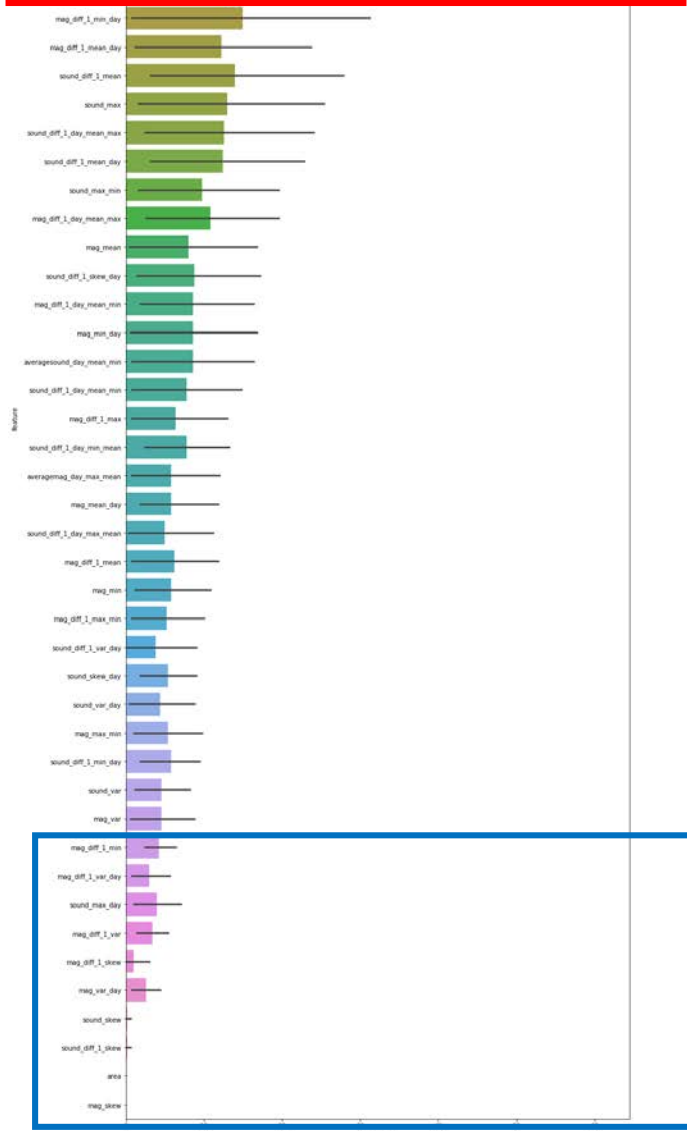
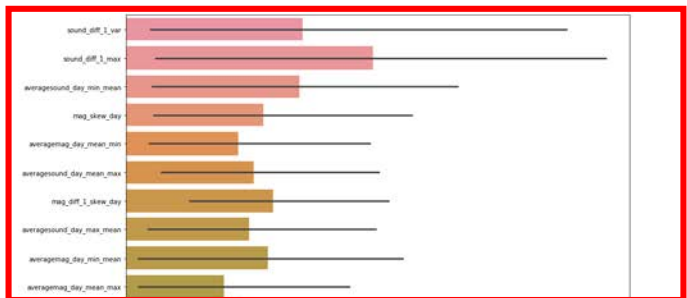
0: 1

1: 4.5

2: 10

StratifiedKFold

Set n\_split=5



Sound\_diff\_var  
 Sound\_diff\_vmax  
 Sound\_day\_min\_mean  
 Magn\_skew\_day  
 Magn\_day\_mean\_min  
 Sound\_day\_mean\_max  
 Magn\_diff\_skew\_day  
 Sound\_day\_max\_mean  
 Magn\_day\_min\_mean  
 Magn\_day\_mean\_max



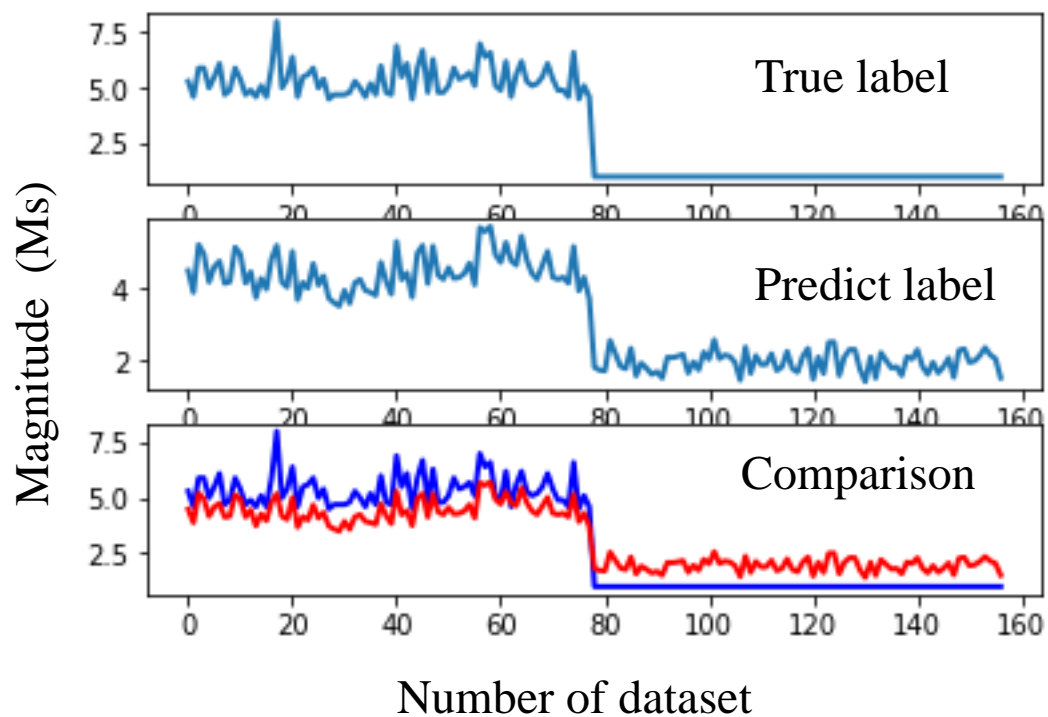
# Random Forest Regressor

Parameters:

- `max_depth=100,`
- `random_state=0,`
- `n_estimators=1000,`
- `min_samples_split=2,`
- `min_samples_leaf=2;`

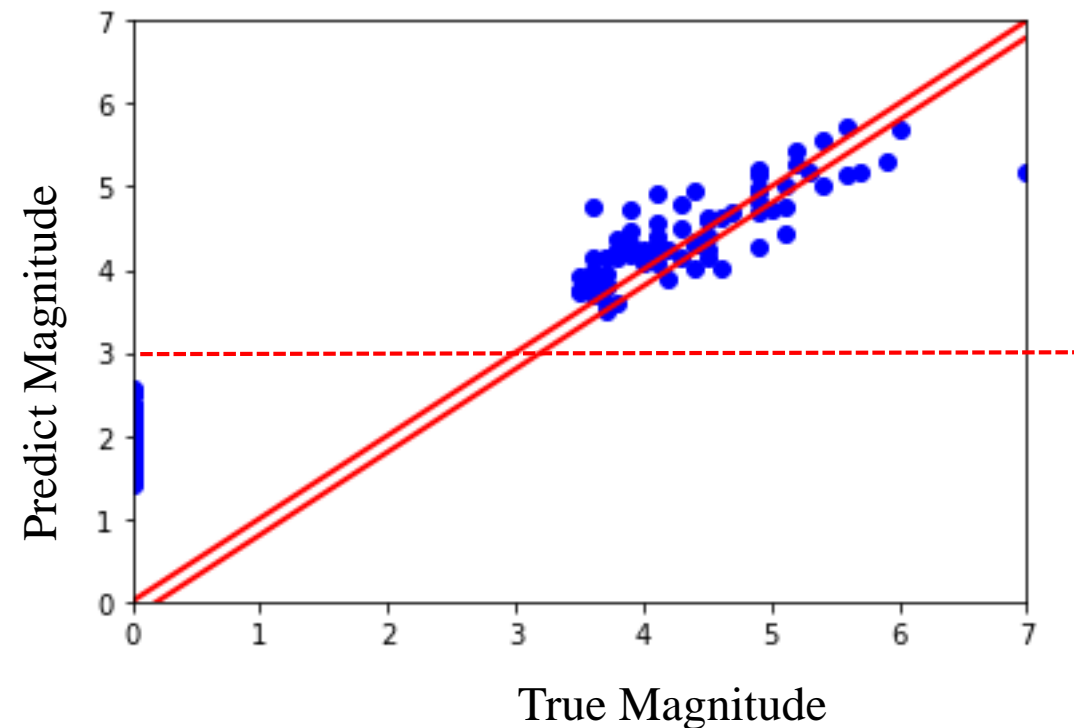
利用随机森林的方法来实现对地震位置和震级的预测。

# Magnitude



标签的预处理

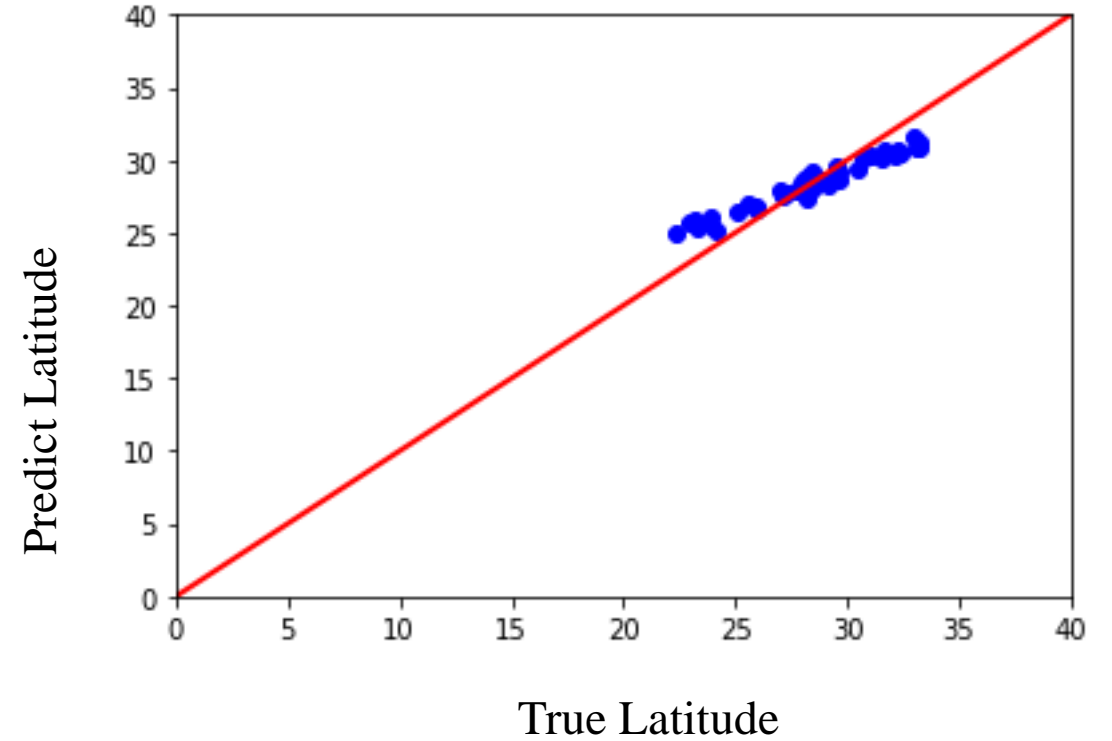
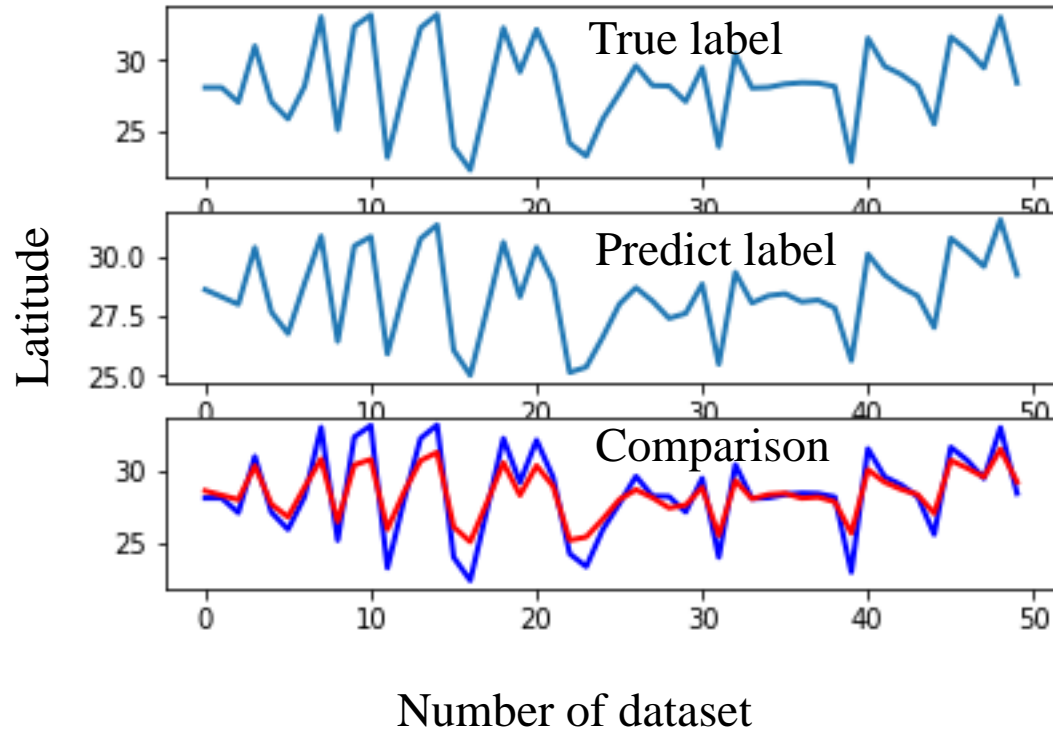
$$\text{Label} = \begin{cases} Ms & \text{if } Ms \geq 3.5 \\ 0 & \text{if } Ms < 3.5 \end{cases}$$



预测的震级如果小于3，则不进行预报；

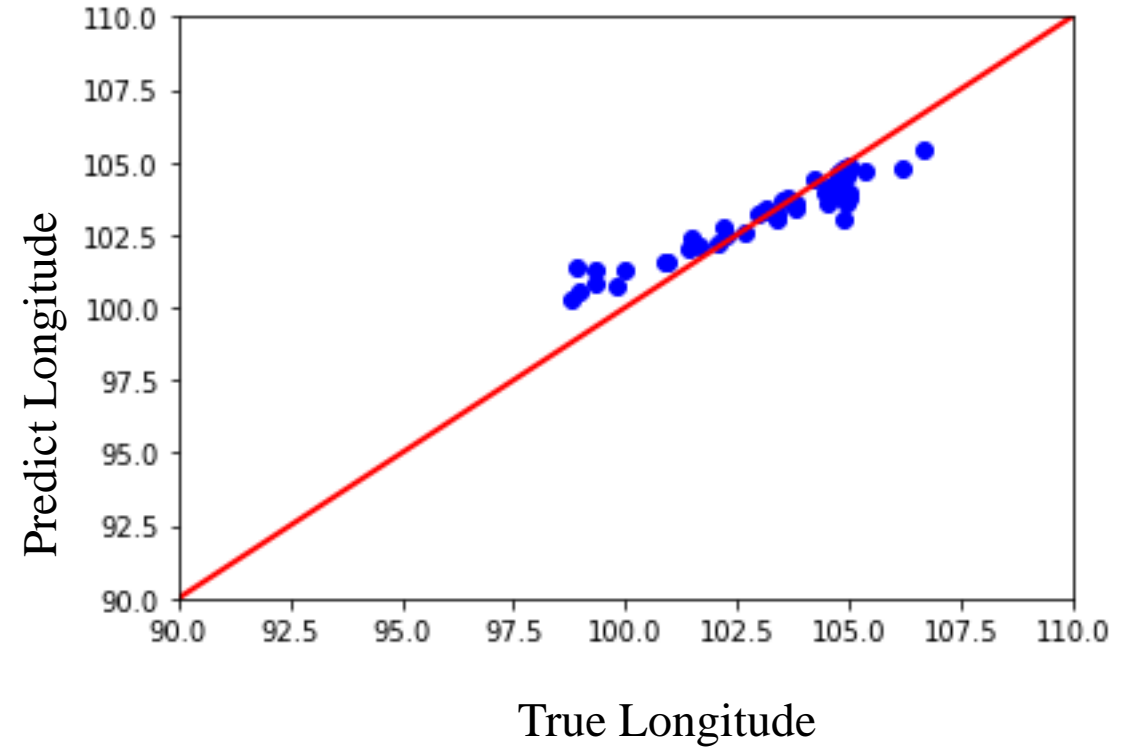
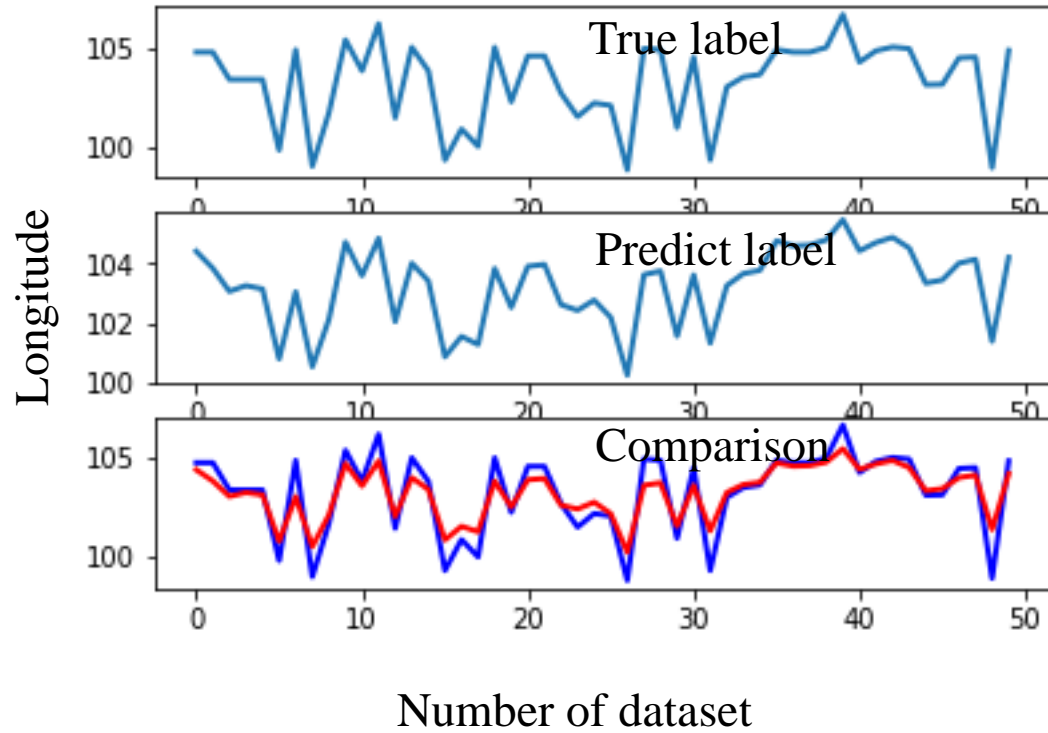
$$\text{Predict} = \begin{cases} 3.5 & \text{if } Ms \geq 3.0 \text{ and } Ms < 3.5 \\ Ms & \text{if } Ms \geq 3.5 \\ 0 & \text{if } Ms < 3.0 \end{cases}$$

# Latitude



Mean absolute error: 1.0  
Variance: 1.6

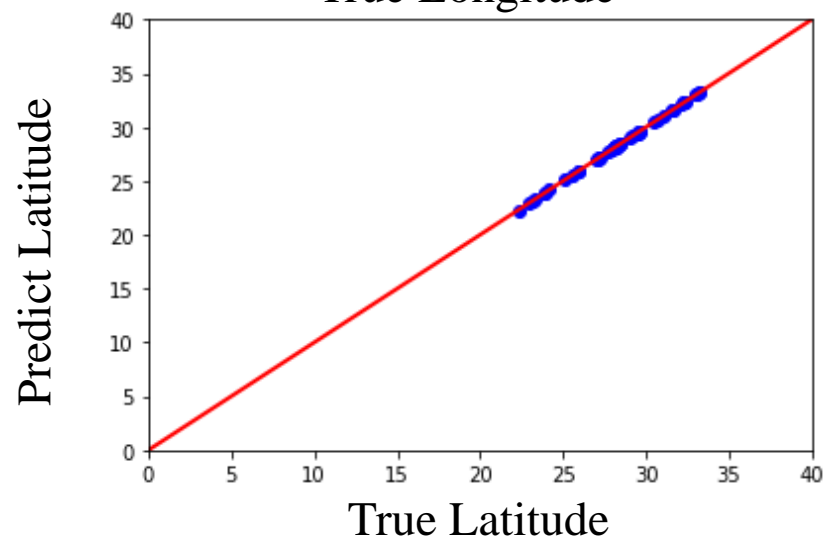
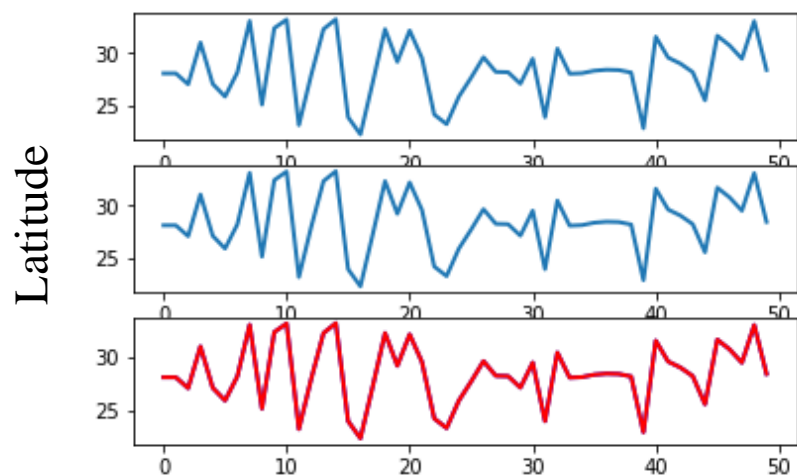
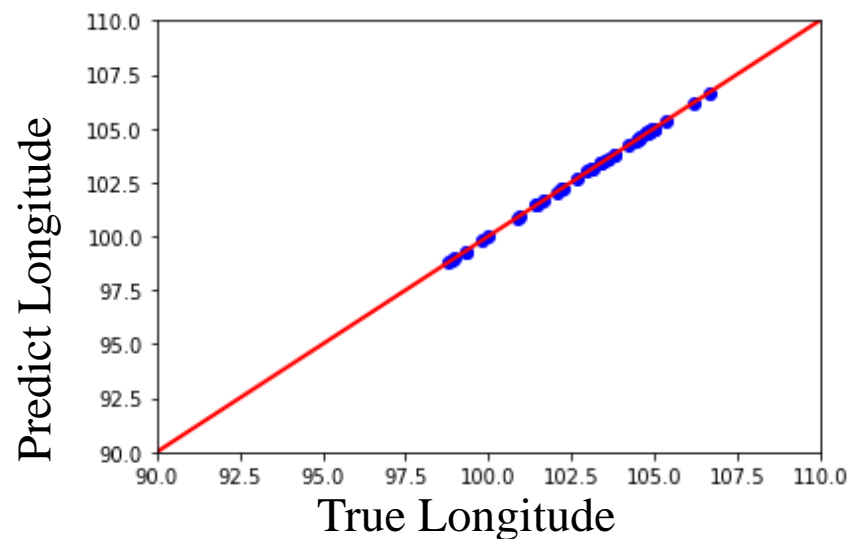
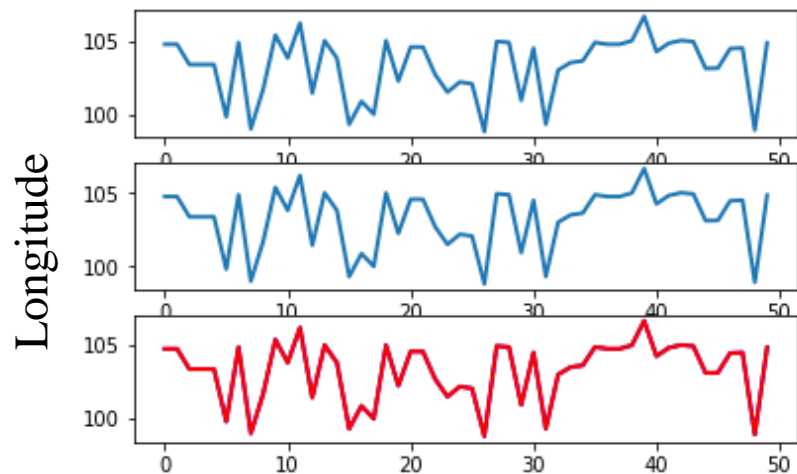
# Longitude



Mean absolute error: 0.67

Variance: 0.77

# Gaussian Process Regressor



# $M_s > 6$

Parameters:

- `max_depth=200,`
- `random_state=0,`
- `n_estimators=1000,`
- `min_samples_split=2,`
- `min_samples_leaf=2;`

利用随机森林的方法来实现对大震进行判断。

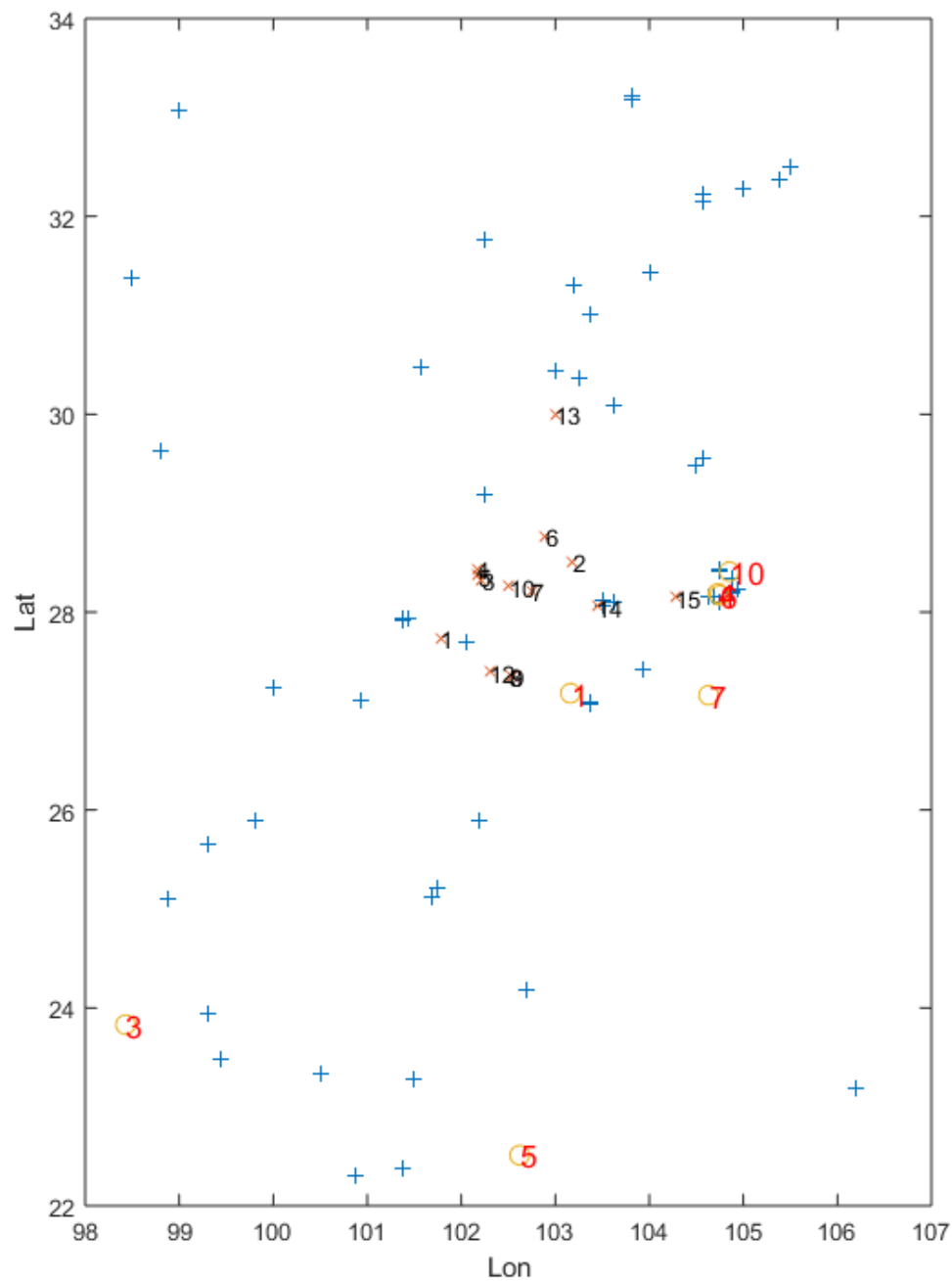
# 结果分析

- 如果预测的震级小于3.0，则判断无地震；
- 如果LGB预测的标签为0，则判断无地震；
- 位置加权分析，在原来预测位置的基础上，加上地震多发区位置的权重；
- 震级加权分析，LGB预测对应的震级与RF预测的震级进行平均；
- 对于大震利用RF进行判定，标签为大震和无震；

Date duration	Y/N	Latitude	Longitude	Magnitude
2020/09/04-2020/09/10	Y	26.87	103.38	4.1

# 结果

+ : 训练样本;  
o : 真实的大震;  
x : 预测的地震;





# 代码说明

- 1\_read.py: 从原始记录中提取数据;
- 2\_label.py: 生成对应数据的标签, 有无地震, 地震的震级和位置;
- 3\_stationlist.py: 提取出台站的坐标信息;
- 4\_feature.py: 从数据中计算对应的特征;
- 5\_lgb.py: 利用LGB和RF对特征进行预测分析;
- 6\_rf.py: 利用RF对是否存在地震进行判断;

Thanks !