



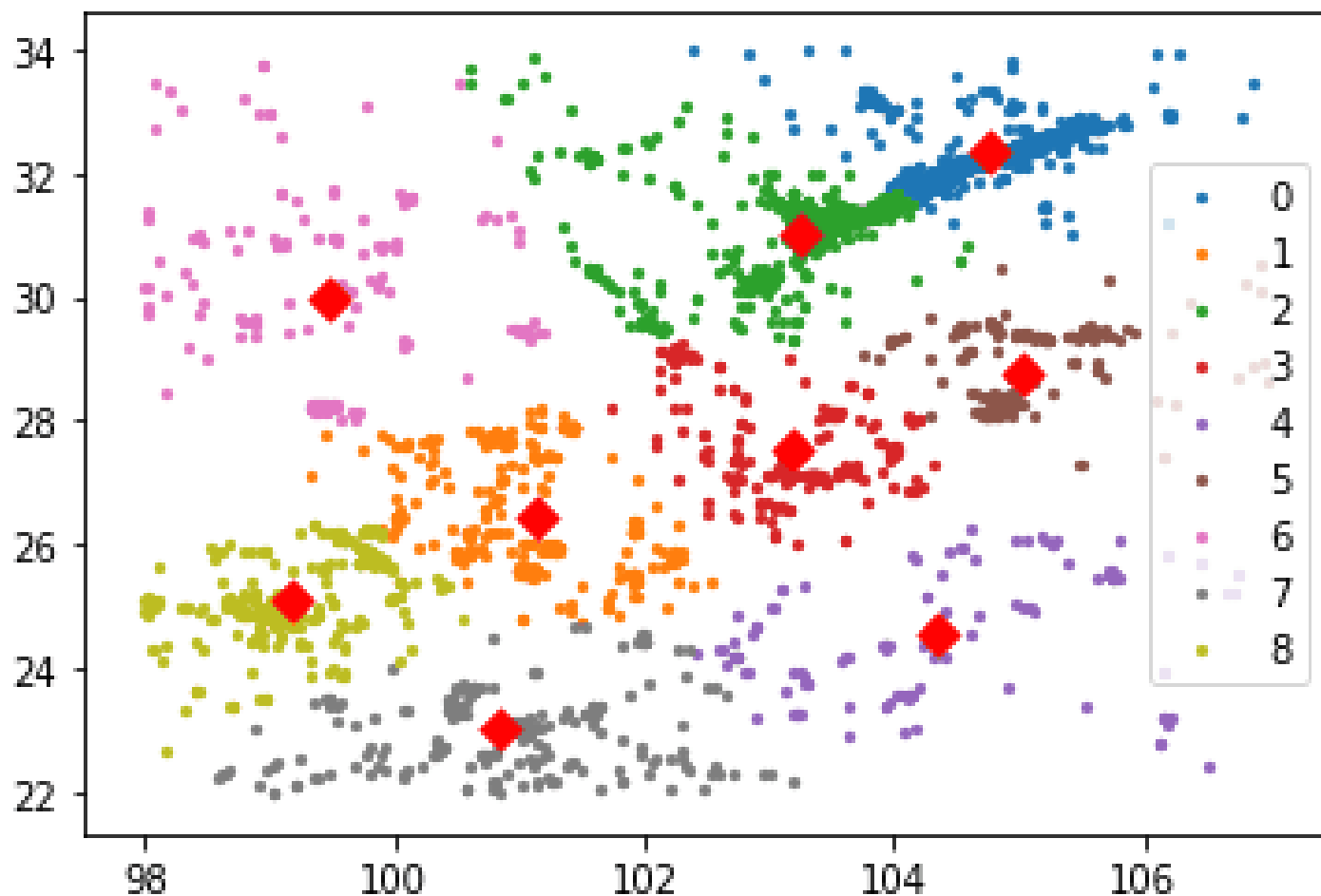
AETA

CIDP-506

简介：

1. 标签样本
2. 数据样本
3. 模型简介
4. 结果产生

2000-2018年所有地震聚类图

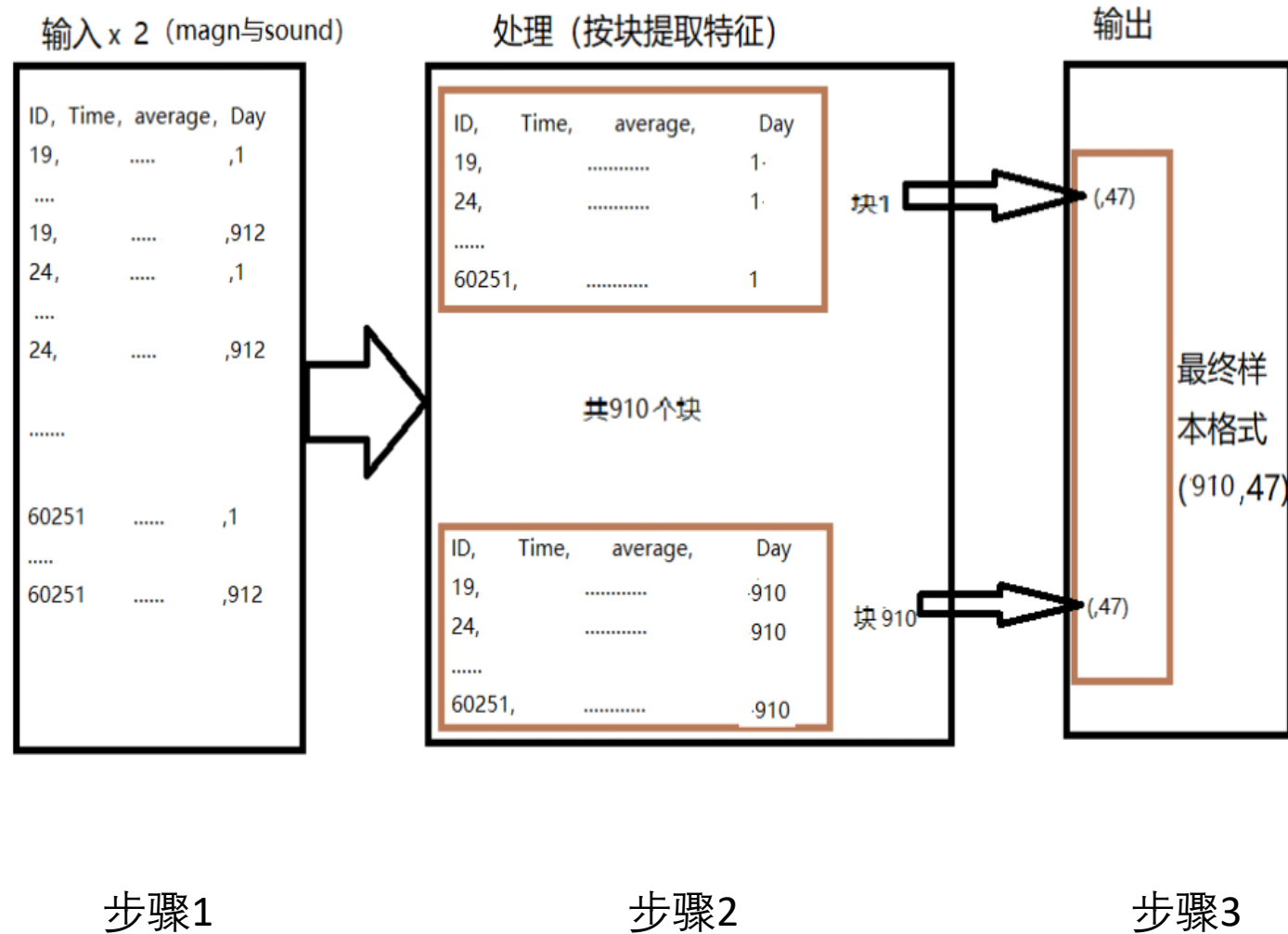


标签样本：

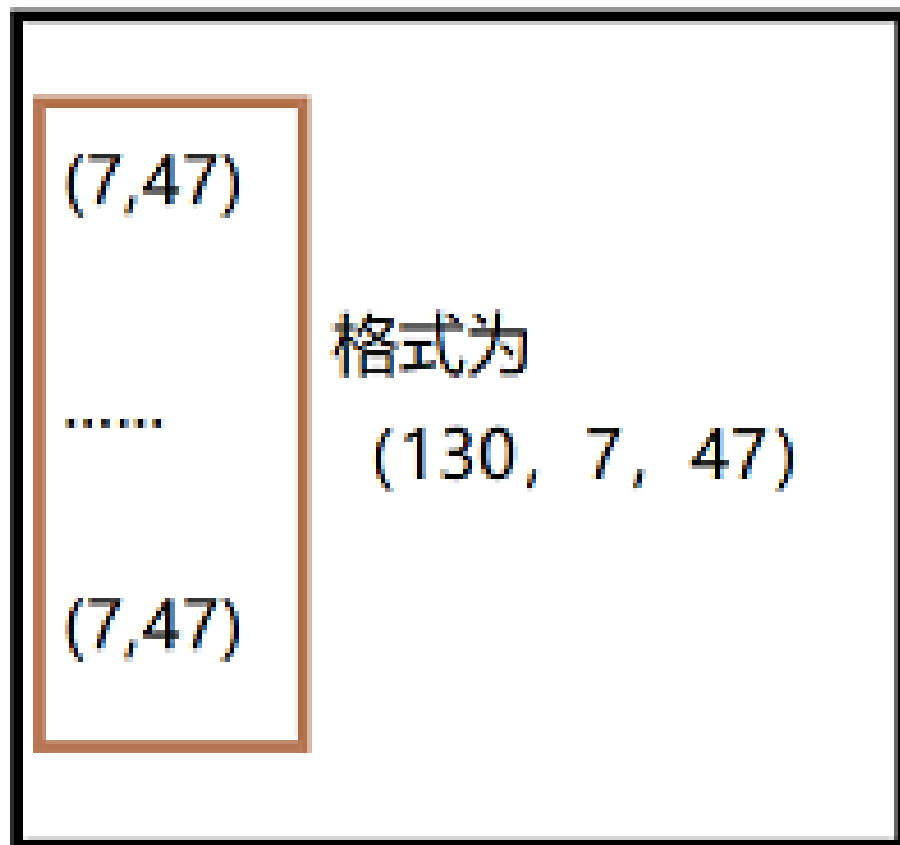
先选取2000年到2018年内规定区域内的所有地震目录（数据来源：中国地震台网），通过K-means++聚类算法将所有地震分为9个区域，然后将题目给定的所有地震目录分属到这九个区域内，同时每个区域内的地震大于3.5级的地震按0.5级为间隔再分类为9个间隔，无地震或震级小于3.5级标为0，据此共将地震目录分类为82类。每个样本标签都选取七天内最大震级的标签作为最终的样本标签。

数据样本：

1. 首先对所有数据以十分钟为间隔进行均值降采样，对降采样之后的数据有缺值时进行线性插值数据补全。
2. 将所有台站的Magn数据，追加到同一个csv文件内。以天为单位进行划分，共划分为910个块。Sound数据做相同他处理。
3. 同样以天为单位，采用官方提供的特征提取方式从两个csv文件内提取每一天的特征数据，选取其中的47个特征作为单天的特征数据。
4. 输出一个二维数据，大小为 (910,47)



格式调整



数据样本：

5.为了适应比赛需要，最终生成的样本如左图所示是，以周为单位进行划分，总共130个样本，每个样本格式为 $(7,47)$



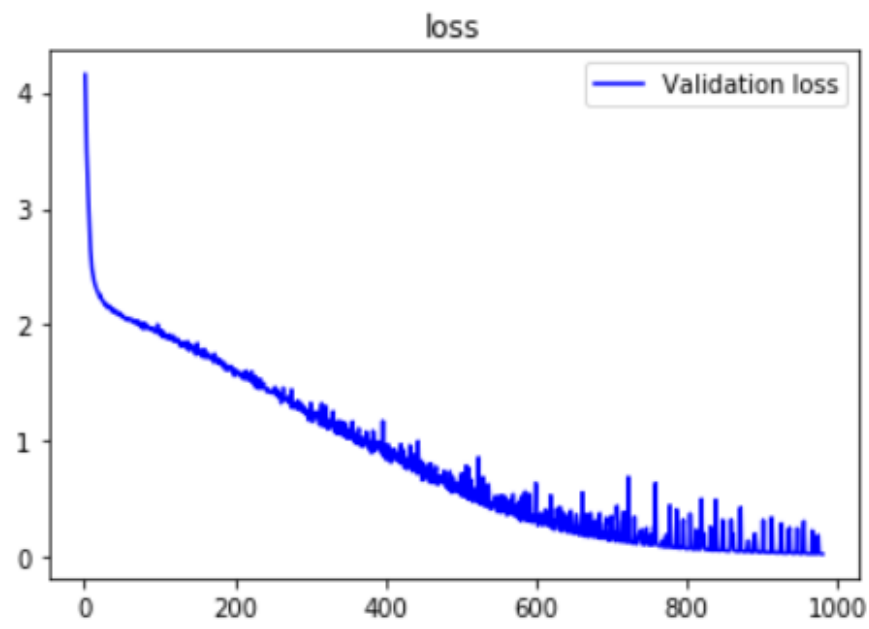
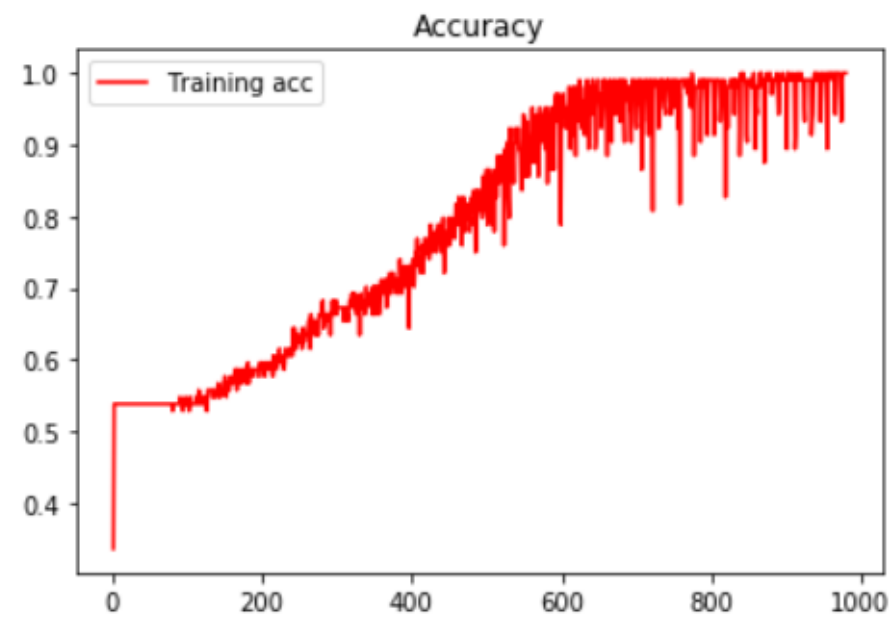
模型简介：

在对比了包括CNN，MALSTM-FCN、LSTM-FCN等模型之后，我们最终使用的是Nested-LSTM模型。

Nested-LSTM模型是对传统LSTM模型的改进，相比于传统的堆栈LSTM，Nested-LSTM可以选择性地访问内部记忆，这使得能在更长的时间规模上操作。

结果产生：

在训练之前为了适应预测的需求，将每一个标签样本后移一个，也就是使用第一周的数据样本对应第二周的标签样本，依此类推。左图为经过900次迭代之后生成的准确率与损失函数图。同时为了防止模型预测的不确定性。我们采用投票法产生最终结果。



第一组

```
{'18': 1, '55': 45, '63': 1}  
#####  
[1, 55]
```

第二组

```
{'2': 2, '55': 45, '63': 2, '74': 1}  
#####  
[1, 55]
```

第三组

```
{'19': 3, '55': 50, '63': 4, '74': 2}  
#####  
[1, 55]
```

第四组

```
{'18': 1, '19': 1, '55': 45, '63': 6, '66': 1, '74': 1}  
#####  
[1, 55]
```

第五组

```
2020-08-08 14:07:13.772788: I tensorflow/core/common_runtime/gpu/g  
{'18': 1, '38': 1, '39': 1, '55': 46, '58': 1, '63': 2, '66': 1}  
#####  
[1, 55]
```

投票法

我们保存了500个model，每100个model为一组，总共5组，500个model进行投票，通过最后投票结果的反馈来决定哪个label的地震发生的概率最大。

如左图所示每一组的模型都是55标签获胜，最终就选取第55个标签作为最终结果。



谢谢！