# 交接文档

## 数据介绍

原始数据为json格式。每一个时刻的全部数据存为一个json，文件名存在两种格式，new/中文件名为时间戳，old/中文件名为时间字符串，由于这一部分数据由海外服务器抓取，所以在转换时应注意时区的变换。每一个区域的数据作为json的键值对。键为区域的编码，格式为“0\_1”，代表在经度上第0个区域，纬度上第1个区域，坐标原点为城市区域左上角。

每一个区域的数据格式详见<https://lbs.amap.com/api/webservice/guide/api/trafficstatus>

## 脚本介绍

1. 爬虫：

我在重构爬虫代码，现在的写的太差了。等重构完会教会刘育杉。

1. 解码：

在Decoder.py中，将原始的json数据解析为csv、js等不同格式，以方便后续的处理。在traffic\_predict/utils/Mygraph.py将csv文件解析为每个时刻的图，图的数据结构如毕业论文中定义。

1. 统计：

在analysis/中，为各种统计的函数，每个函数有不同的功能，调用即可。

1. 训练

模型使用traffic\_predict/tools/rain.py进行训练，traffic\_predict/tools/evaluate.py对模型进行测试，traffic\_predict/tools/seq2seq\_evaluate.py对baseline的模型进行测试。

模型训练的训练集和验证集和测试集由traffic\_predict/utils/generate\_dataset.py中的DataBuilder类完成。将数据集的csv文件分别放在文件夹里，并将文件夹的名字添加到DataBuilder类的初始化函数中的for mode in [] 一句的列表中。