

# QG 工作室周记

姓名：王伟杭

组别：数据挖掘组 年级：2020 级

周次：第 5 周

生活随记	还是那句话，麦当劳 yyds！最近也在做一些比较基础的锻炼，想把身体锻炼得好一点 hh。
学习 开发 比赛 情况	1.出租车大数据分析系统； 2.数学建模算法学习； 3.LSTM 算法
一周总结	<p>这周主要的任务就是对最终考核的需求进行处理。</p> <p>先是一二天刚拿到数据的时候，主要在学习怎么利用这个数据，再加上一开始无法连接上数据库，整个的工作进度就变得比较缓慢，在成功连接上数据之后，便着手开始研究数据的字段，最终确定了一些字段来进行统计和预测。最开始的工作是进行司机收入的预测，一开始的思路是对每个司机都分别进行预测，朝着这个方向做了一天，发现和预期的方向比较相悖，于是又和几个伙伴开了几个会确定下来要的数据样式和工作安排，然后就开始着手做了起来，统计的方向做得比较顺利，很成功的把一百多万条数据进行了统计，接下来一个比较棘手的任务就是做预测，这个方向在网上找了比较多的博客，最后还是确定用 LSTM 模型来对时间序列进行预测，又花了一些时间学习怎么使用 TensorFlow 的 LSTM 模型，看了一些博客之后，终于成功的实现了一个比较简单的模型，并应用到预测中，不知道是什么原因，预测的效果并不是非常理想，只能说是勉强能用，后面尽量还是希望能对这个模型加深一些理解，看看能不能调出预测效果更好一些模型。</p> <p>然后就是对利用率的一个统计和预测，思路大体上也是通过统计 59 天的数据之后进行预测。主要的一个瓶颈还是在处理数据上，因为数据量比较大，而且评估的指标也需要自己来进行确定，所以也花了不少时间。最后就是对车流量的统计预测，这一块与伙伴们也是讨论得最多的板块，对于他的指标比较模糊，所以花了一些时间来进行确定，最后决定通过聚类来进行区域划分，然后采用一样的方法来进行预测，因为车流量的特性，我们把最小时间粒度划分到小时进行统计，然后传给模型进行评估，因为数据量比较大的原因，模型学习得比较彻底，预测效果比之前要好一些，但是取而代之的是时间</p>

	复杂度的提高，训练了比较久的时间，我也是趁着模型学习的时间对 LSTM 进行了一些学习，以及为了之后的数学建模比赛学习了一些模型，阅读了一篇去年国一的论文，对这次的数学建模我还是比较看重的，毕竟是自己参加的第一个大赛事，希望可以有比较好的结果收尾。
存在问题 未来规划	1.对深度学习了解欠缺，不能调出比较好的模型来进行预测； 2.对数学建模比赛的准备欠缺；
导师评价	