摩拜单车数据分析项目

比赛要求

比赛训练数据已发布。地理位置通过Geohash加密，选手可以通过开源的方法获得经纬度数据。但是选手不能使用任何其他的外部数据，包括地图、天气等数据。比赛希望鼓励选手在地理位置画像，用户画像和车辆画像三个方面对数据进行挖掘建模。

目前，摩拜单车在北京的单车投放量已经超过40万。用户可以直接在人行道上找到停放的单车，用手机解锁，然后骑到目的地后再把单车停好并锁上。因此，为了更好地调配和管理这40万辆单车，需要准确地预测每个用户的骑行目的地。

参赛者需要根据摩拜提供的数据，预测骑行的目的地所在区块。

比赛地训练集取北京某一区域的一段时间内的部分数据，测试集为同一区域未来一段时间的数据。训练集的将划分为public set和private set划分。比赛过程中，选手将只能看到public set的分数，但比赛的最终排名将由private set的分数决定。

标注数据中包含300万条出行记录数据，覆盖超过30万用户和40万摩拜单车。数据包括骑行起始时间和地点、车辆ID、车辆类型和用户ID等信息。参赛选手需要预测骑行目的地的区块位置。

1. 下载比赛数据
2. 规整比赛数据
3. 发现比赛数据特点
4. 抽象出模型
5. 选择方法进行模型设计
6. 根据最简方案提交结果
7. 根据规则调优代码和模型