

车速网络数据集构建

数据：

1. 杭州 2012 年 11 月中 30 天出租车轨迹数据一共 920,000,000+条

MESSAGE_ID	VEHICLE_ID	VEHICLE_NUM	LONGI	LATI	PX	PY	SPEED	DIRECTION	STATE	CARSTATE	SPEED_TIME	DB_TIME	NOTE
1093241264	22029	浙ATB888	120.094000	30.283750	0.000000	0.000000	0.00	90.00	0	0	2011/11/1 0:02:56	2011/11/1 0:00:58	*
1093241266	17177	浙ATB151	120.124183	30.309950	0.000000	0.000000	0.00	90.00	0	0	2011/11/1 0:02:57	2011/11/1 0:00:58	*
1093241269	17126	浙ATB183	120.116050	30.311683	0.000000	0.000000	0.00	90.00	0	0	2011/11/1 0:02:56	2011/11/1 0:00:58	*
1093241271	18623	测飞田4	120.167050	30.278133	0.000000	0.000000	0.00	0.00	1	0	2011/11/1 0:02:55	2011/11/1 0:00:58	*

2. 杭州市道路网络 shpfile 文件 ——来自 openStreetMap



数据库：

Oracle 11g (内置 Oracle spatial)

车速网络数据集构建：

基本思路：

方案一：通过出租车轨迹记录中的 speed 数据，作为该车该时刻的车速，根据 GPS 位置信息，在设定的阈值范围内使用 Oracle 空间函数与道路网络判断包含关系，匹配到最近的道路，统计该条道路上所有车的车速，平均之后作为该条道路的平均车速，分别计算之后可以根据需要得出每条道路网络以给定时间间隔的平均车速信息

方案二：通过出租车在时间上相邻两点间的 GPS 位置信息，通过 Oracle 空间函数计算在该时刻内汽车通过的距离，除以时间得到汽车的车速，作为该车该时刻的车速，之后的做法与方法一相同。

可视化：

通过 JS 可以与数据库相连，通过 SQL 语言得到包含道路几何以及道路的平均车速信息的 GeoJSON 文件可直接用于 WEB 端的展示。

存在的问题：

1. 目前 OpenStreetMap 只下载到杭州市最新的道路网络 shp 文件或是几天前的，与 2012 年的出租车 GPS 信息匹配可能有出入（）

2. 由于 GPS 信息量较大，如何提高数据集的构建效率（空间索引之类的）
3. 学长给的 GPS 数据导入之后直接建表，其在建立的时候并没有将 GPS 数据中经纬度作为 geometry 类型（能否直接使用 Oracle 空间函数还存在疑问，可能需要重新整理学长的数据）