



Block-City:基于出租车数据的城市区块功能探索

## 项目建议书

作 者 庞潇 王翔坤 闫建荣 Leo

提 交 日 期 2018.10.23

评 审 人

评 审 日 期

---

## 目录

1 选题背景.....	2
2 项目成员与分工.....	2
2.1 项目成员 .....	2
2.2 成员分工 .....	3
2.3 项目进度 .....	3
3 数据说明.....	4
3.1 数据来源 .....	4
3.2 数据描述 .....	4
4 研究内容描述.....	6
5 研究方法与技术路线.....	7
5.1 研究方法 .....	7
5.2 技术路线 .....	7
6 预期效果.....	8

## 1 选题背景

可视分析是遵循人类认知一般过程并用恰当的视觉界面来促进分析推理的学科，目前这是一个处于高速发展中尖端研究领域，给其他领域的不同学科注入了新的活力。近年来，城市问题不断涌现，发展滞缓，缺乏活力。因此城市通过大面积调整原有规划，使得城市快速扩张，因此城市不再是单核城市活力带动城市的发展，而逐渐发展城市两极带动发展甚至多核的发展模式。研究城市活力中心的移动具有重要的现实意义。

城市从形成的历史来看，城市是人与人之间活动的空间集合。城市活力的诞生是与城市各个方面的数据一起影响的结果。城市的数据是由城市活动的人产生，因此研究城市的人群空间分布和空间流动特征是研究城市的活力中心发现、功能布局调控、城市安全等空间研究具有重要意义。

出租车作为城市居民交通的一种重要交通工具，研究出租车数据结合城市行政区域的关系能一定程度上反映出城市人流集散数据分布和人群流动的特征。然而出租车数据和行政区域数据结合使其分析工作十分艰难，其中一个特性就是无法预测。因此结合可视化的展示可以通过可视分析来加强人对于整个分析过程中的核心作用，并且通过数据可视分析方法处理这些海量信息。

## 2 项目成员与分工

### 2.1 项目成员

庞潇 2018223045180

王翔坤 2018223045138

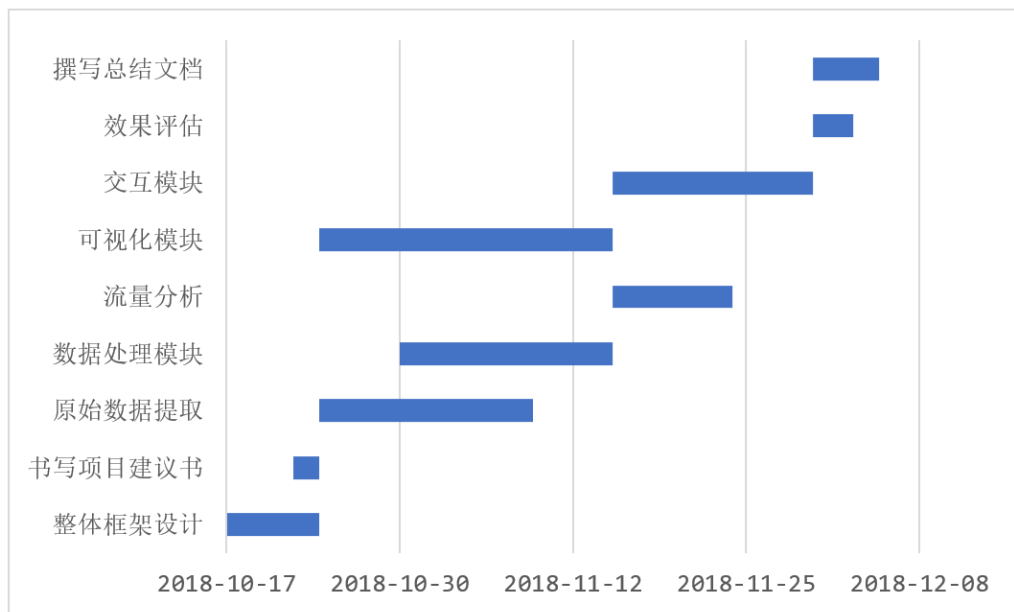
闫建荣 2018223045146

Leo 2018521461006

## 2.2 成员分工

整体框架设计	庞潇、王翔坤、闫建荣
书写项目建议书	庞潇、王翔坤、闫建荣
原始数据提取	王翔坤、闫建荣
数据处理模块	王翔坤、闫建荣
流量分析	王翔坤、闫建荣
可视化模块	庞潇
交互模块	庞潇、王翔坤
效果评估	庞潇、王翔坤、闫建荣、Leo
撰写总结文档	庞潇、王翔坤、闫建荣

## 2.3 项目进度



## 3 数据说明

### 3.1 数据来源

本项目数据有三部分组成：其中使用的成都市出租车 GPS 数据和成都行政区域数据均为实验室已有数据，成都区域的 POI 数据从百度地图获得。

### 3.2 数据描述

数据有三部分组成：

1、出租车 GPS 数据的数据项包括：数据 ID、车牌号、经度、维度、载客状态、数据时间、记录时间，如图 3-1 所示。

数据编号	车牌号	经度	维度	载客状态	数据时间	记录时间
1	12327	30.65773	104.0908	1	2014年8月24日	20:00:00
1	12327	30.66066	104.0919	1	2014年8月24日	20:00:00
1	12327	30.66387	104.0933	1	2014年8月24日	20:00:00
1	12327	30.66639	104.0939	1	2014年8月24日	20:00:00
1	12327	30.66859	104.0908	1	2014年8月24日	20:00:00
1	12327	30.67051	104.0875	1	2014年8月24日	20:00:00
1	12327	30.67256	104.085	1	2014年8月24日	20:00:00
1	12327	30.67386	104.0812	1	2014年8月24日	20:00:00
1	12327	30.67559	104.0775	1	2014年8月24日	20:00:00
1	12327	30.67655	104.0746	1	2014年8月24日	20:00:00
1	12327	30.67747	104.0717	1	2014年8月24日	20:00:00
1	12327	30.67968	104.0686	1	2014年8月24日	20:00:00
1	12327	30.68071	104.0658	1	2014年8月24日	20:00:00
1	12327	30.6815	104.0622	1	2014年8月24日	20:00:00
1	12327	30.68043	104.059	1	2014年8月24日	20:00:00
1	12327	30.67925	104.0564	1	2014年8月24日	20:00:00
1	12327	30.67793	104.0545	1	2014年8月24日	20:00:00
1	12327	30.67785	104.0541	0	2014年8月24日	20:00:00
1	12327	30.67625	104.0516	0	2014年8月24日	20:00:00
1	12327	30.67466	104.0492	0	2014年8月24日	20:00:00
1	12327	30.672	104.0481	0	2014年8月24日	20:00:00
1	12327	30.66919	104.0484	0	2014年8月24日	20:00:00
1	12327	30.66593	104.0475	0	2014年8月24日	20:00:00
1	12327	30.66371	104.0457	0	2014年8月24日	20:00:00
1	12327	30.66441	104.0442	0	2014年8月24日	20:00:00
2	7318	30.69061	104.1059	1	2014年8月24日	10:00:00

图 3-1 出租车 GPS 数据属性

2、成都的行政区域的 shp 文件，分为如图 3-2 所示。

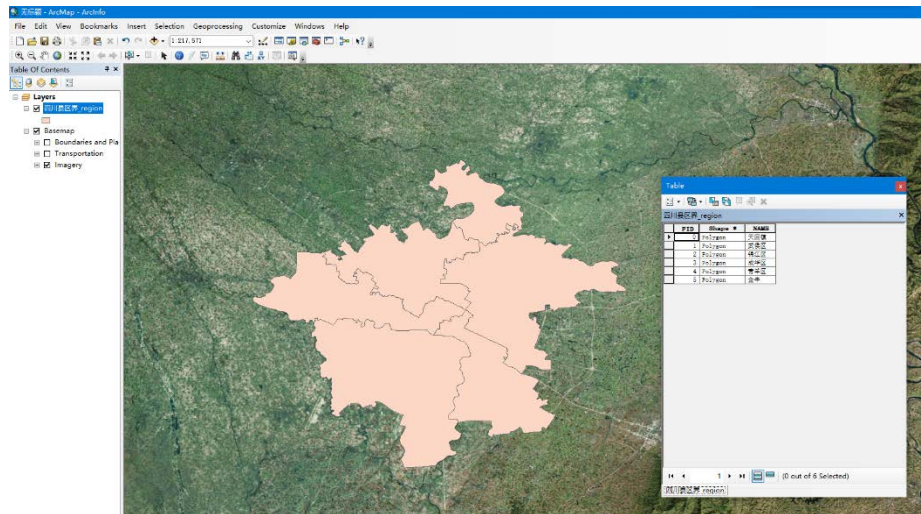


图 3-2 成都行政区域数据

### 3、百度 POI 数据

从百度地图爬取，包含三个属性：名称、坐标、地址，先对数据进行预处理：1)对数据异常点，如经纬度不在成都行政区域范围内的，认为是异常数据，删除此记录；2)对数据重复点，两条记录完全相同的点，仅保留第一条记录，删除其余相同的记录；3)对于数据缺失的记录，由于采集设备和存储设备的异常，会使某些属性缺失，如缺失的是名称、经纬度这类重要属性，则删除此条记录。

百度 POI 数据内容如图 3-3 所示。



图 3-3 百度 POI 数据

## 4 研究内容描述

城市是人以及人活动聚集的空间，研究人群的空间分布和人群流动特征，对研究城市的功能布局调控、城市安全等空间研究具有重要意义。

本项目研究的内容：

- 1、通过出租车的 GPS 数据的 OD 属性判断出出租车上下车载客位置，通过可视化系统分析出人群集散区域，从而比较城市不同时间段城市各个行政区域的城市活跃度的变化。
- 2、基于可视化系统展示出租车的行程路线的 GPS 数据，从动态分析的研究角度挖掘城市每天人流移动的情况以及每月城市活力中心的移动情况，结合区域的 POI 数据，分析各个区域中 POI 活力的变化。

## 5 研究方法与技术路线

### 5.1 研究方法

#### 4.1.1 区域流量分析

区域的出入量可以反映区域当前的交通热度，目前很多研究只关注某个区域本身的流入流出量，而没有关注区域与区域之间的流量关系及变化，而这一点能够反映某一时段内城市的移动趋势，对此进行分析可以帮助使用者更好地判断接下来的城市热点转移，理解变化的模式和过程。直接再地图上叠加图层进行流动模型的展示可以让用户更加直观地研究区域间地流动情况，对感兴趣地区域可以单独观察其流入流出情况，研究人们的出行模式。

#### 4.1.2 区域活跃度分析

城市中某一时段哪些区域较为活跃，区域的活跃度随时间的变化情况等问题对研究人员了解城市整体的交通状况，区域活跃度的变化模式都有很大的帮助。使用流图和排名相结合可以很方便的快速了解不同区域的变化情况，还可以聚焦自己感兴趣的区域进行深入研究。

#### 4.1.3 POI 分布

区域内的 POI 分布在城市规划中是很重要的问题，使用地图展示区域内 POI 可以帮助规划者直观地感受区域内的 POI 分布，再辅以基于饼图的新型图表，系统地对区域的 POI 结构进行分析，并且在此之上将 POI 与区域移动相结合，了解出入区域的 POI 情况，可以对人们的出行目的进行更加细致深入的分析。

### 5.2 技术路线

可视化流程如下图所示，其中数据变换过程我们使用 Python 和必要的框架对数据进行整理和清洗，得到理想的数据结构。可视化映射、视图变换、人机



交互使用 html/css/javascript 来构造 web 对数据进行展示和交互，计划采用 d3.js 可视化库进行联动交互展示。

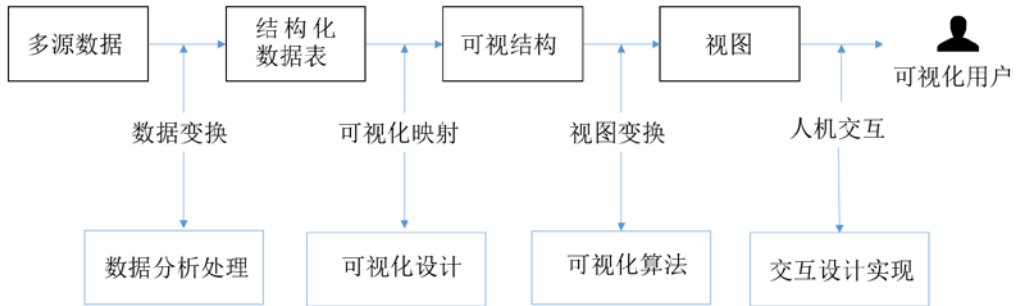


图 5-1 项目技术路线

## 6 预期效果

项目预期实现的系统界面如图 6-1 所示，系统分为 6 个部分，每部分的视图之间互相联动。

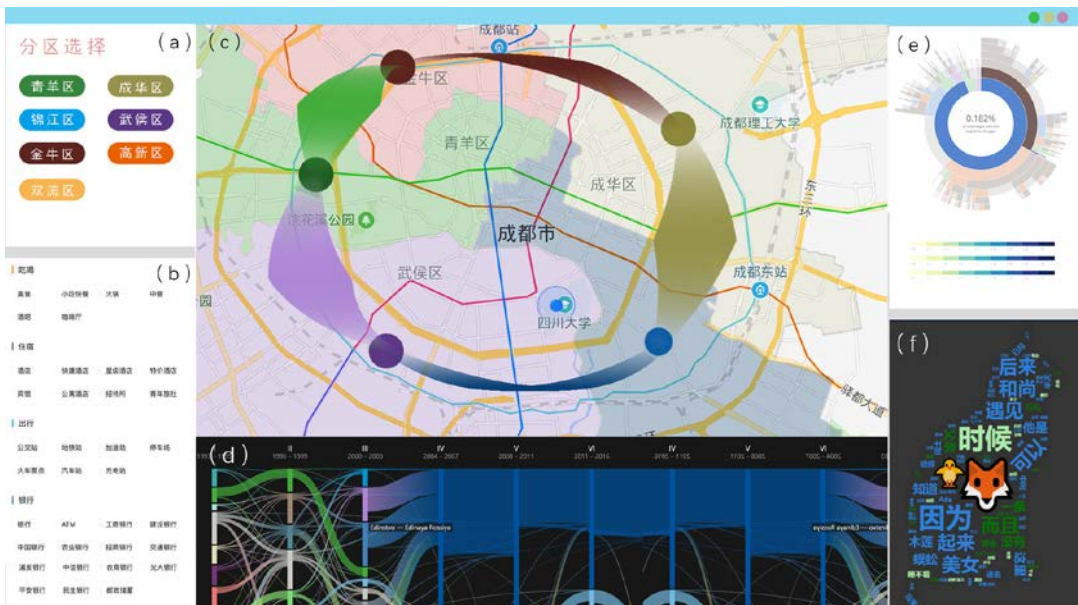


图 6-1 预期系统界面

(a)是城市划分后的各个区域，我们采用数据划分好的行政区域，可以从这里选择自己感兴趣的区域，进行深入分析(b)是地图的 POI，我们可以限定某一个 POI，重点研究，也可以结合(a)研究某一特定区域内的 POI 分布。图(a)和图(b)的数据过滤结果最终会在图(c)上进行直观的呈现，并且图 c 还可以直观的表示各区域之间的迁移情况，发生迁移的时段是由(d)进行控制的，从(d)中可以看出不同区域的活跃度排名在不同时间尺度的变化情况。(e)是所选区域的 POI 占比情况，以及所去区域的 POI 占比情况，将二者结合分析可以了解移动的目的，下图的柱状变化显示了该区域不同时间点的去向热点，(f)则是词云，让用户对区域内的某一 POI 分类下进行概览，了解这个 POI 在此区域有哪些特点。