

城市公交站点设置的优化分析 2017/10/29 姜鹏辉

目录



◆ 城市交通情况对于城市规划,居民城市归属感,城市品牌有着至关重要的影响。大城市的可持续 发展,应该立足当前、着眼长远,倡导绿色环保出行,大力优先发展城市公共交通,构建性能优 良的交通系统工程,是解决城市交通拥堵的有效手段。

而"智能交通"是解决城市

交通拥堵的制胜法宝之一!!



- ◆ 在自动采集技术日益发达的今天,有什么例子可以展现"智能交通"的应用呢?
 - 1. 高德地图如何实现"实时躲避拥堵"
 - 2. 酷米客实时公交-上班不迟到



高德地图APP躲避拥堵





◆ 高德地图如何实现"实时躲避拥堵"

- 1. 高德地图为用户提供实时躲避拥堵服务,背后是高德强大的实时交通服务能力。
- 一方面是通过在全国主要城市的出租车、物流车等行业运营车辆上,装载GPS和无线通信设备,将车辆行驶时间、时速、方向、坐标等参数信息实时传送到浮动车信息中心,汇总、处理后生成反应实时道路路况的交通信息。由于运营车辆长期上路行驶,且行驶路线多为主要干道,据此生成的实时路况信息,足以展现真实路况,以此预判拥堵,精确度极高。

2.另一方面,每天近千万使用高德地图在线导航的用户也给高德实时交通贡献了大量的用户出行服务数据。通过对多种来源的交通大数据进行融合和分析,实时交通预测水准大幅度提高。高德地图实时躲避拥堵功能可实现为用户计算出最佳行车路线,并按路况和行车位置变化实时调整,为用户提供实时躲避拥堵的解决策略



◆ 酷米客实时公交-上班不迟到APP





◆ 酷米客实时公交-上班不迟到

1. 内容提要

【公交查询利器——候车神器】

北京、上海、西安、东莞、惠州、广西来宾部分实时线路试运营。实时公交覆盖深圳主要区域

◆ 酷米客实时公交-上班不迟到APP

1.到站信息实时报

深圳公交车实时线路超过500条可查询。我不等车,车等我。查看离你最近的公交车,还有几站、 还有几分钟到,误差不超过1分钟。

2.换乘查询

输入起点与终点即可获取最便捷的出行方案。还可以根据不同需求选择"少换乘、少步行、不做地铁"等方案。面面俱到的出行方案,让你到哪都不用担心乘车的问题。

3.地图上查看公交

地图上查看动态的移动公交,还原行驶情况。得知公交车在线路上的位置、到站时间、行驶方向, 还可以了解到公交车运行时间段。信息详实齐全,地图上查看更加直观明了。

4.常用线路、站点查询

智能公交自动收藏你常用的线路和站点,无需再次输入,使用更简便。人性化服务,为你量身定做属于自己的酷米客公交



◆ 酷米客实时公交-上班不迟到APP

5.站点查询

输入站点, 查看途经该站所有公交线路, 哪辆车最先来一目了然, 多种选择也不犹豫。

6.附近站点

智能定位你的当前位置,查看附近站点详情,提供全面信息。

7.公告牌

官方发布最新深圳公交线路变更公告、失物招领,还有酷米客公交最新活动。第一时间掌握最新资讯。

8.候车提醒

自由选择提前1-10分钟或1-5个站公交车到站提醒,闹钟响起再出发,来得早不如来得巧



本案例研究某一座城市?它的交通发展概况如何呢?





该城市的交通发展概况



■ 某城市地处南海沿海地区,有独特的地理位置,是珠江三角洲区域的核心城市之一。随着社会经济迅速发展和城市规模不断扩大,全国各地的从业人员不断涌入,城市人口也随之不断增加,然而城市交通却赶不上人口和经济的发展,因此城市交通也逐渐成为阻碍该城市市发展的重要因素。常规公交是城市公共交通的主体,地面公交作为城市公交的一部分,是城市居民日常出行的重要交通工具,关系到城市经济的发展。

思考

◆该城市的交通系统现能提供哪些数据??



思考

现能提供的数据有:

- ✓ 出租车GPS监控数据
- ✓ 地面公交车GPS监控数据
- ✓ 地面公交车刷卡数据
- ✓ 地铁站刷卡交易数据



原始数据情况

◆出租车GPS监控数据样本

流水号	车牌号	数据时间	经度	纬度	速度(Km/h)	空重车状态	是否空载
1902891463	粤0BV65	2014/6/9 0:07	113. 925761	22. 515592	0	0	0
1902892812	粤4775D	2014/6/9 0:05	114. 040518	22. 618699	70	0	0
1902892817	粤4775D	2014/6/9 0:05	114. 040518	22. 618699	0	0	0
1902893320	粤J66E7	2014/6/9 0:06	114. 108046	22. 56201	0	0	0

◆地面公交车GPS监控数据样本

经度	纬度	海拔(米)	业务时间	记录时间	到离站 (1, 到;2, 离)	GPS速度	车牌号
114. 18	22. 638299	0.00	2014/6/9 0:00	2014/6/8 7:14		5. 55	粤N4646
114. 18	22. 638399	0.00	2014/6/9 0:00	2014/6/8 7:14	2	7. 5	粤N4646
114. 181999	22. 639499	0.00	2014/6/9 0:00	2014/6/8 7:14		6.11	粤N4646
114. 182998	22. 639999	0.00	2014/6/9 0:01	2014/6/8 7:14	1	8.05	粤N4646

原始数据情况

◆地面公交车刷卡数据样本

卡片记录编 码	交易金额 (打折前) 元	实际交易金额 (打折后)	交易时间	公司名称	线路名称	车牌号
292915914	2	1.2	2014/6/9 0:00	东部公共交通	866	75528
30084633	2	1.6	2014/6/9 0:08	东部公共交通	M415	F3337
326479819	2	1.6	2014/6/9 0:00	金华南巴士	367(福永)	粤W2839
322657049	3	2	2014/6/9 0:25	东部公共交通	M310	92350
293194948	2	1. 6	2014/6/9 0:22	东部公共交通	866	80683

◆地铁站刷卡交易数据样本

卡片记录编 码	交易金额 (打折前)	实际交易金额 (打折后)	交易时间	公司名称	线路名称	车牌号	进出站
251386615	2	1	2014/6/9 0:00	地铁五号线	黄贝岭	AT-142	出站
280228491	3	2.85	2014/6/9 0:00	地铁五号线	翻身	OT-102	出站
290392057	7	6.65	2014/6/9 0:00	地铁五号线	黄贝岭	AT-115	出站
293865243	5	4.35	2014/6/9 0:00	地铁五号线	翻身	OT-114	出站



挖掘目标

(1)利用公交车载GPS数据与公交刷卡数据,构建模型,分析居民出行规律。

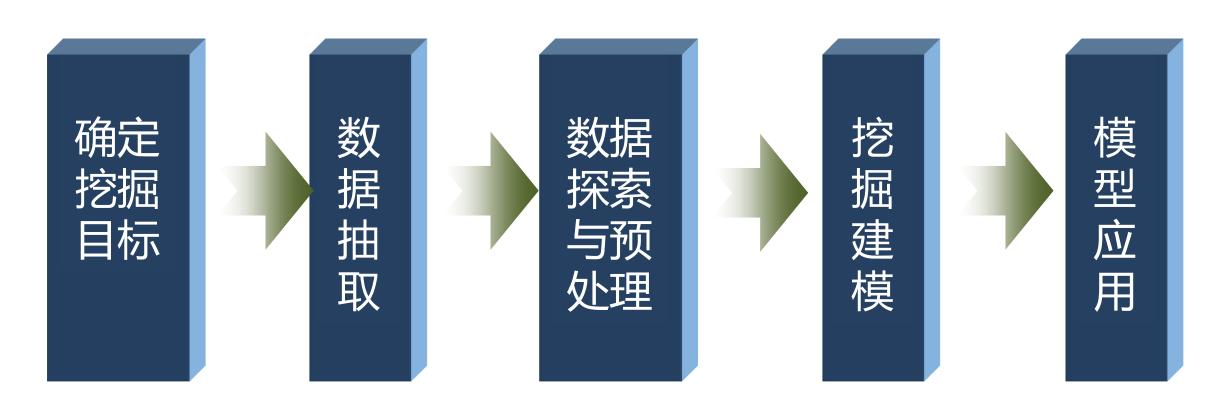
(2)结合目标1的因素分析,最终提出城市公交站点设置的优化建议。



目录



◆ 总体流程图:



第1步:数据抽取

◆ 与城市公交站点相关的原始数据主要有地面公交车GPS监控

数据以及地面公交车刷卡数据等

	属性名称	属性说明
	卡片记录编码	
	交易金额	打折前,单位为元
	实际交易金额	打折后,单位为元
公交刷卡数据	交易时间	
	公司名称	
	线路名称	
	车牌号	
	经度	单位:度
	纬度	单位: 度
	海拔	单位:米
	业务时间	
	记录时间	
	数据类型	3:GPS;4:到离站;55:违规;47:DSRC检到离场;71:GPS 到离场;53:开关门
公交GPS数据	线路号	
公文GP3数编	子线号	
	到离站	1:到;2:离
	GPS速度	
	传感器速度	
	方向角	
	违规标准	
	车牌号	

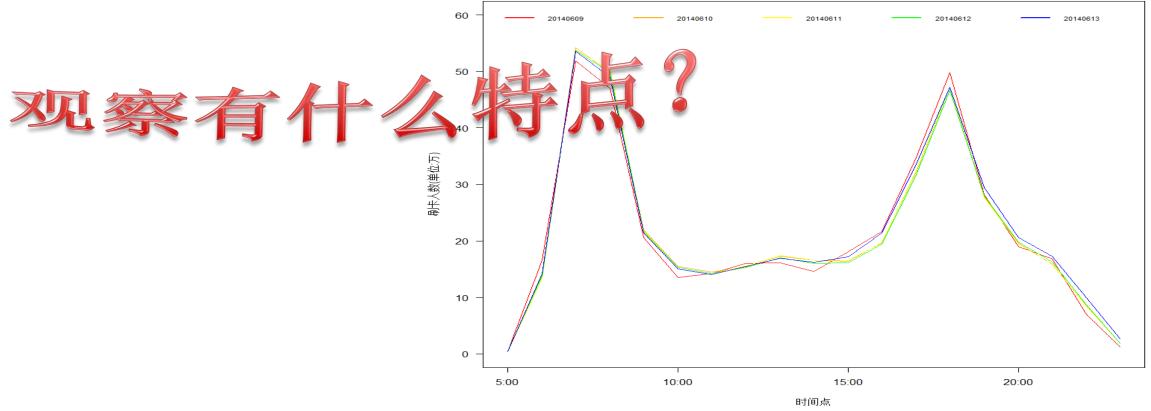


属性规约: 属性选择后的数据集

	公交刷-	卡数据集			公交GP	S数据集	
公交卡 卡号	交易时间	线路名 称	车牌号	经度 (度)	纬度 (度)	业务时间	车牌号

第2步:探索分析—在地面公交车刷卡数据表的基础上,画出2014年6月10日和

2014年6月13日每个时间段刷卡的总人数的折线图





第3步:数据预处理

1. 数据清洗:从业务以及建模的相关需要方面考虑,筛选出需要的数据

第3步:数据预处理

2. 缺失值处理:在原始计量数据,特别是经纬度数据抽取过程中,发现存在缺失的现象。

经度	纬度	业务时间	卡片记录编码	线路名称	车牌号
114.0205	22.73463	2014/6/10 9:41	330357953	M285路	粵B63317
114.0205	22.73463	2014/6/10 9:41	330357953	M285路	粵B63317
114.0205	22.73463	2014/6/10 9:41	330357953	M285路	粵B63317
114.0149	22.74483	2014/6/10 14:12	326064411	M285路	粵B63567
114.0504	22.71837	2014/6/10 8:10	323169447	M285路	粵B64759
114.0504	22.71837	2014/6/10 8:10	323169447	M285路	粵B64759
114.0149	22.74471	2014/6/10 10:07	281036907	M285路	粵B64759
		2014/6/10 10:07	281036907	M285路	粵B64759
114.0494	22.71788	2014/6/10 14:22	291080364	M285路	粵B64759
114.0494	22.71788	2014/6/10 14:22	291080364	M285路	粵B64759
	22.6675	2014/6/10 7:30	323973256	M424路	粵B64849
114.0196	22.6675	2014/6/10 7:30	323973256	M424路	粵B64849
	22.66323	2014/6/10 7:10	293918874	M424路	粵B64860
114.0186	22.66323	2014/6/10 7:10	293918874	M424路	粵B64860
114.021	22.65992	2014/6/10 7:13	325748940	M424路	粵B64860



第3步:数据预处理

本案例缺失值处理

通过数据探索分析,发现数据中存在缺失值,到离站为空,经度为空、纬度为空、由于原始数据量大,这类数据所占比例较小,对于问题影响不大,因此对其进行丢弃处理。具体处理方法如下:

- a)丢弃经度为空的记录
- b)丢弃纬度为空的记录
- c)丢弃到离站为空的记录

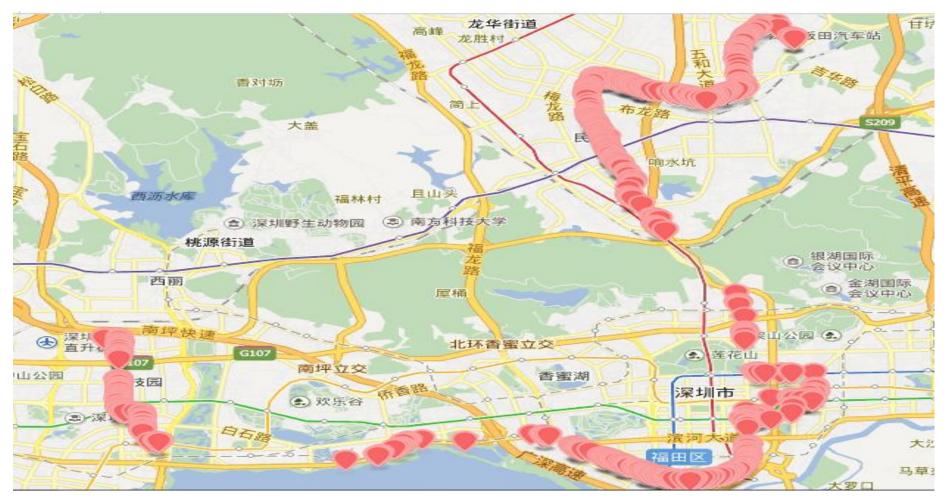


68路实际站点的路线图

仅选取68路作为本案列的研究目标

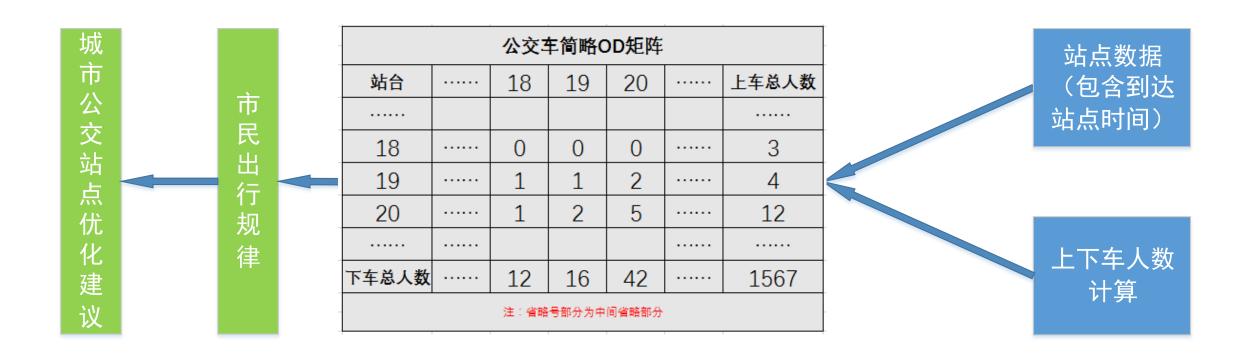


68路GPS经纬度数据的可视化图

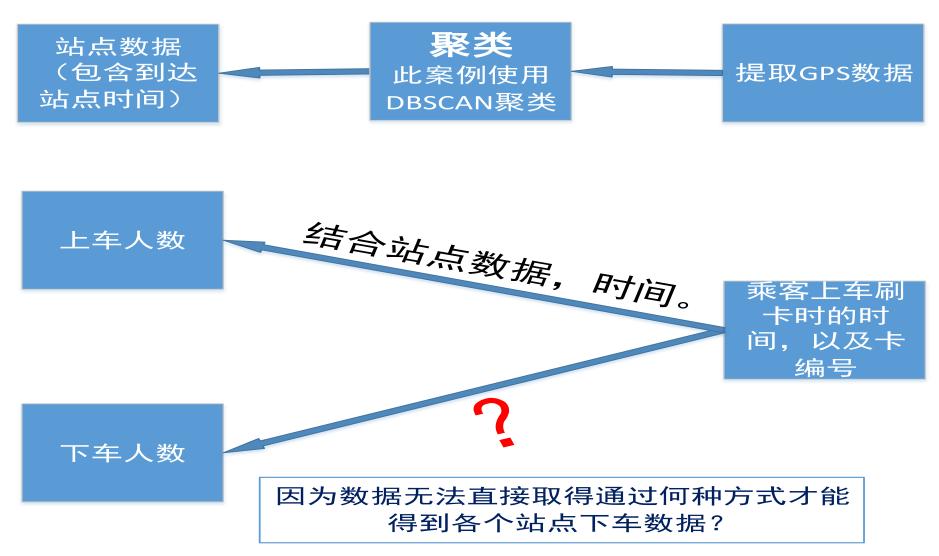




本案例总体思路



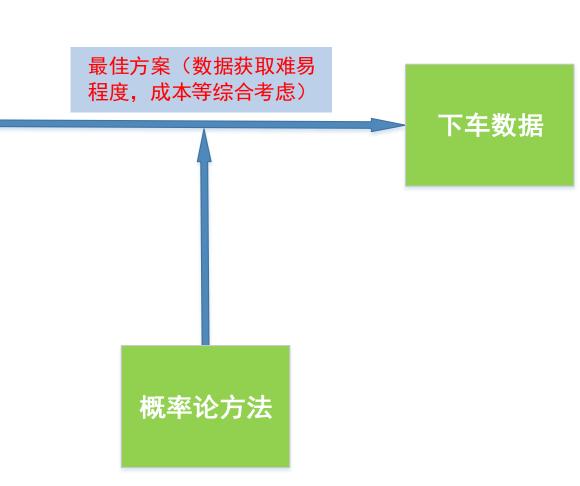






下车人数数据获取

	已有数据							
	数据名称	数据示例						
	经度	123.123456						
GPS数据	纬度	12.123456						
	业务时间	2014/7/17 00:00:00						
	车牌号	粤A12345						
	交易时间	2014/7/17 00:00:00						
刷卡数据	线路名称	339						
	车牌号	粤A12345						





第4步:构建模型

1. 城市公交站点设置优化的相关模型

- a) 基于密度的DBSCAN聚类算法
- b) 计算下车人数
- c) OD矩阵

第4步:构建模型

城市公交站点设置优化的相关模型

1. 基于密度的DBSCAN聚类算法

DBSCAN算法主要是依赖两个主要的参数来进行聚类的,即对象点的区域半径 Eps和区域内点的个数的阈值MinPts。DBSCAN算法通过查找数据集中任意一个点的距离在Eps区域来进行聚类,如果这个区域内的点数大于MinPts,则将这些点放在同一个簇中,形成新的一类。DBSCAN算法中用到的定义如下:

第4步:构建模型

- 1. 点的Eps区域:以空间中任意一点p为圆心,Eps为半径的区域中的点的集合,记为集合D。Eps是由用户指定的。
- 2. 点的密度:集合D中点的个数即为点P的密度。
- 3. 阈值MinPts: 在集合D中使点p成为核心点的限定值。
- 4. 核心点:如果点p的密度等于或者大于阈值MinPts,则P为核心点。
- 5. 边界点:如果点p不是核心点,却又落在其他核心点的区域内,那么p点为边界点。
- 6. 噪声点:如果点p既不是核心点,也不是边界点,则p点为噪声点。
- 7. 密度直接可达:存在空间任意一点q在集合D中,且p是核心点,则称点q从点p直接密度可达。
- 8. 密度可达:空间存在点 在集合D中,且,如果从直接密度可达,那么点p从点q密度可达。



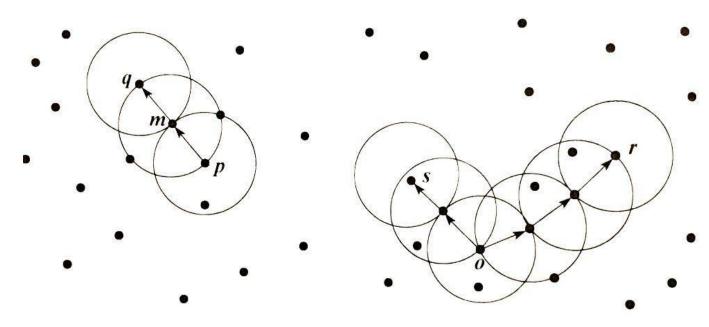
第4步:构建模型

DBSCAN算法聚类过程如下:

- ◆ 将所有的点分别标记为核心点、边界点或噪声点。
- ◆ 删除识别出的噪声点。
- ◆ 将在Eps之内的所有核心点连通,并形成一个簇。
- ◆ 将每个边界点归属到一个与之相关联的核心点的簇中。

第4步:构建模型

DBSCAN算法原理图



- 1.对象q是从m直接密度可达的。对象m从p直接密度可达的。
- 2.对象q是从p (间接) 密度可达的,因为q从m直接密度可达, m从p直接密度可达。
- 3.r和s是从o密度可达的,而o是从r密度可达的,所有o,r和s都是密度相连的。



第4步:构建模型

• 采用DBSCAN聚类法对**68路线所有站点**进行聚类,取半径为0.0011, 阈值为3聚类得到的站点数为39,实际的站点数为56,由于有几个连 续的站点相离较近,而且上车人数也比较多,所以聚类得到的站点 数比实际的少是属于正常情况;

• 选取68路一部分经纬度进行聚类,得到68路部分密度聚类图

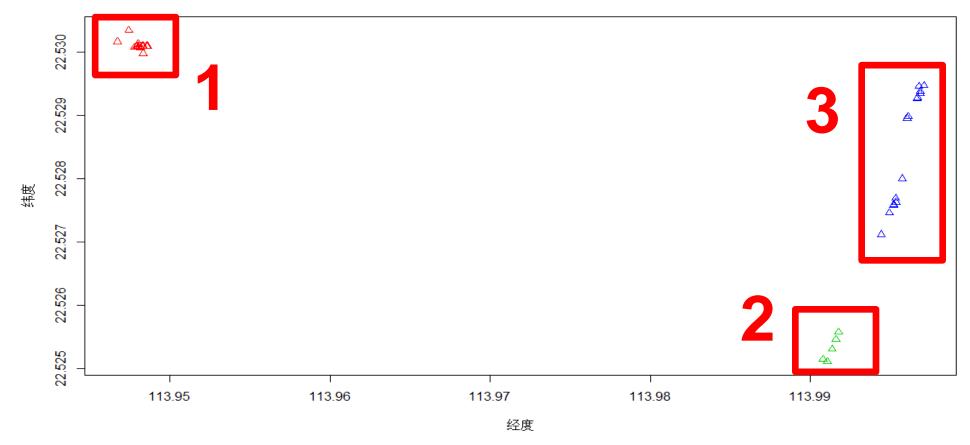
第4步:构建模型

68路部分经纬度数据

经度	纬度	业务时间	线路名称	车牌号
113.9483	22.5301	2014/6/9 10:42	68路	粤7316
113.9483	22.5301	2014/6/9 10:42	68路	粤7316
113.9918	22.52558	2014/6/9 11:21	68路	粤7388
113.9969	22.52938	2014/6/9 8:44	68路	粤7368
113.9969	22.52938	2014/6/9 8:44	68路	粤7368
113.9969	22.52938	2014/6/9 8:44	68路	粤7368
113.9969	22.52938	2014/6/9 8:44	68路	粤7368
113.9969	22.52938	2014/6/9 8:44	68路	粤7368
113.9953	22.52763	2014/6/10 7:17	68路	粤7067
113.9486	22.5301	2014/6/10 13:08	68路	粤7067
113.9944	22.52712	2014/6/10 11:48	68路	粤7226
113.9479	22.5301	2014/6/10 14:04	68路	粤7335
113.948	22.53013	2014/6/10 19:11	68路	粤7335
113.948	22.53012	2014/6/11 12:30	68路	粤7067
113.9911	22.52512	2014/6/11 12:35	68路	粤7067
113.9483	22.52998	2014/6/11 7:50	68路	粤7131
113.9916	22.52547	2014/6/11 17:50	68路	粤7197

第4步:构建模型

68路部分密度聚类图





第4步:构建模型

(68路部分密度聚类图)对应的实际公交站点图



(68路部分密度聚类图)对应的GPS经纬度数据可视图



第4步:构建模型

城市公交站点设置优化的相关模型

2. 计算下车人数

- 一、模型假设和符号说明
- 假设1:提供的地面公交车GPS监控数据和地面公交车刷卡数据都是真实的, 而且在一定程度上可以代表某市城市居民出行的情况;
- 假设2:城市居民出行的距离可以近似看作是服从正态分布的。因为如果出行的距离太短,居民一般会选择相对方便的交通工具,如步行或者骑车;如果出行的距离太长,居民一般会选择私家车或者汽车出行。所以可以推断,城市居民出行的距离更集中在一定范围内。
- 假设3:城市居民公交出行的出行站数服从泊松分布。



第4步:构建模型

我们以一条路线单向行驶的数据进行分析,很明显,起点站是没有乘客下车的,所以:在第二个站点下车的乘客是来自起点站的,

以此类推,所以之后每个站的下车人数为:

$$D_{j} = \sum_{k=1}^{j-1} S_{k} \times p_{kj}, j = 1, 2, ..., m$$

其中 D_j 为j站下车人数; S_k 为k站上车人数; p_{kj} 为k站上车,途经j-k站 点下车的概率。

第4步:构建模型

计算下车概率:

确定下车概率有两个重要的因素,一是乘客出行的距离,二是站点对乘客的吸引力。基于这两个因素,如果一条线路有n个站点(包括起点站和终点站),那么我们可以得到概率矩阵为

$$P = (p_{i,j})_{n \times n}$$



第4步:构建模型

引入站点对乘客的吸引权重,即在某条线路上的某个站点乘客上车人数占某条路线的总上车人数的比例,其公式为:

$$w_{j} = \frac{S_{j}}{\sum_{k=1}^{m} S_{k}}$$

其中 W_j 为j站权重; S_k 为k站上车人数。

第4步:构建模型

当某条路线途经的站点数较少时,下车的人数也会比较少。相反,某条线路如果途经的站点数较多时,乘客下车的人数会较多。所以可以推断,居民公交出行的出行站数概率服从泊松分布

其公式为:
$$\begin{cases} F_{ij} = \frac{e^{-\lambda} \lambda^{(j-i)}}{(j-i)!} \\ 0, \quad i \geq j \end{cases}$$

其中 λ 为某条公交路线出行途经的站点数的数学期望; W_j 为j站权重; F_{ij} 为i站上车,途经j-i个站点下车的概率。

第4步:构建模型

综上,最终居民公交出行 的出行站数概率为:

$$p_{ij} = \begin{cases} \frac{F_{ij} \times W_j}{\sum_{j=1}^{m} F_{ij} \times W_j}, i < j \\ 0, i \ge j \end{cases}$$

第4步:构建模型

城市公交站点设置优化的相关模型

3. OD矩阵

OD调查即交通起止点调查又称OD交通量调查,OD交通量就是指起终点间的交通出行量。"O"来源于英文ORIGIN,指出行的出发地点,"D"来源于英文DESTINATION,指出行的目的地。OD调查结果通常用一个二维表格表示,称为OD表,也叫OD矩阵

第4步:构建模型

1、分时段分析: 68路线时段2上班高峰期的OD矩阵,如表2所示:

	2	3	4	7	10	11	12表	2:	时.	段2	的O	D矩	阵	28	30	\$ 31	32	34	36	37	÷ 39	上车总人数
4	o	0	0	0	0	0	0	0	0	0	0	0	0	0	0	О	0	0	0	0	О	О
7	o	0	0	0	0	1	1	0	1	0	0	0	0	О	0	О	0	0	0	0	О	3
10	0	0	0	0	0	1	2	0	1	0	0	0	0	0	1	О	0	0	0	0	О	5
11	0	0	0	0	0	0	3	0	2	1	0	0	0	О	1	0	0	0	0	0	О	7
12	0	0	0	0	0	0	0	1	4	1	0	0	0	o	2	1	0	0	0	0	О	9
13	0	0	0	0	0	0	0	0	0	0	0	0	0	О	0	0	0	0	0	0	0	О
14	0	0	0	0	0	0	0	0	0	1	1	0	1	0	3	1	0	0	0	0	0	7
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
17	0	0	0	0	0	0	0	0	0	0	0	8	0	0	1	0	0	0	0	0	0	1
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	=
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	=
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	1	0	2	0	0	34
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	2	12	3	0	22
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	1	1:
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	7
39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
下车总人数	0	0	0	0	0	3	8	1	10	3	1	0	1	0	15	35	6	2	16	15	8	124





第4步:构建模型

1、分时段分析: 68路线时段4下班高峰期的部分OD矩阵,如表3所示:

	1	2	3	4	7	9	10	‡ 11	表	‡ 1 4	时	没4	部分	O	D为	5阵	‡ 26	28	\$30	‡ 31	‡ 32	÷ 33	\$ 34	÷ 36	[‡]	÷ 39	上车总人数
11	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
12	0	0	0	0	0	0	0	0	0	4	5	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	14
14	0	0	0	0	0	0	0	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
15	0	0	0	0	0	0	0	0	0	0	0	3	1	1	1	0	0	0	0	0	0	0	0	0	0	0	6
17	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
20	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	8	1	0	0	0	0	0	0	0	0	0	0	0	1
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	C
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	2
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	3
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0	4
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	10
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Ō	5	2	5	1	1	0	14
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	0	5
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	7
39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
下车总人数	0	0	0	0	0	0	0	0	10	7	13	6	2	3	3	0	1	1	4	13	5	2	7	3	6	7	93

第4步:构建模型

模型应用-居民出行规律&城市公交站点优化建议

由OD矩阵结果可以看出,上班高峰期上车的人数主要集中在第30、31和36站,第30和31站是横岭工业区站到水牛新村站之间,这些站点居民住宅和商场较多;第36站是万科城站,其附近居民住宅较多。下班高峰期上车的人数主要集中在第12、30和31站,第12站实际站点有三个,即妇幼保健院站到益田花园站之间,

第4步:构建模型

在益田花园站附近有一个公园,这也是一个人群较多的地方。整条线路中,有些站点几乎没有人上下车,如福田中学站。针对上车人数较多的站点,建议增加一些安全措施,如安排一些相关人员维持秩序。在上下班高峰期的时候可以多安排一些车次或者增加区间车,也可以对现有的一部分车在一些上下车人数较少的站点可以不设置停靠站点,这样可以有效利用公交车,也可以缓解上班高峰期拥挤的情况。

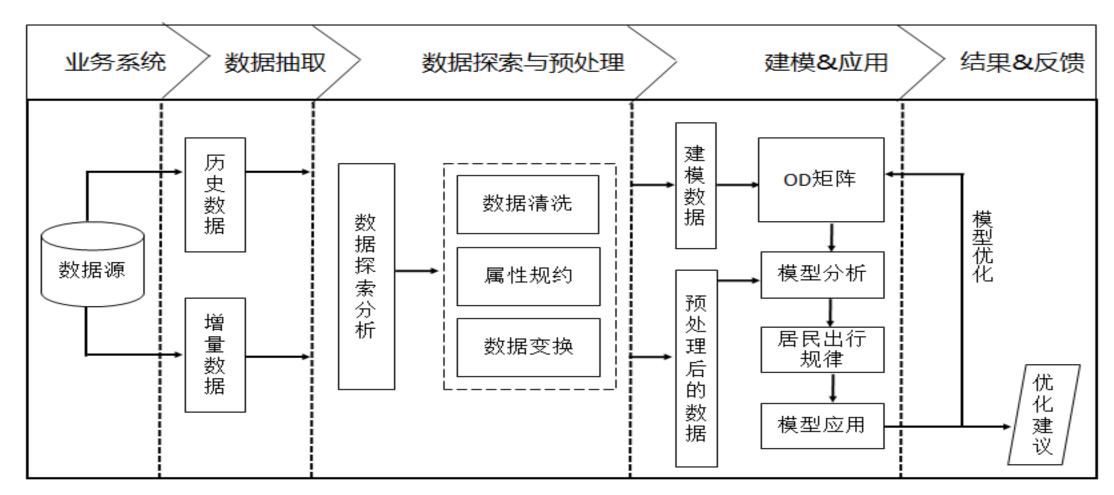


第4步:构建模型

本案例只是针对一条公交线路推导其OD矩阵,为了结合实际应用,不妨推导多些公交路线的OD矩阵,与人工调研OD矩阵相对比,全面分析这些公交路线所在地理区域的规律特征,作出更准确的优化分析和优化建议。

总结

◆ 总体流程:





目录





目录





Thank you!

泰迪科技:www.tipdm.com

热线电话: 40068-40020

