



# 大数据技术与城市计算

## 轨迹大数据和人群移动模式探究及预测

---

姚尧 博士，副教授，高级工程师

地理与信息工程学院，地图制图学与地理信息工程

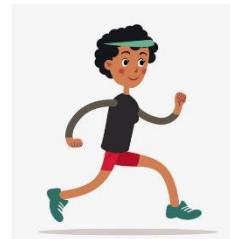
阿里巴巴集团，访问学者

Email: [yaoy@cug.edu.cn](mailto:yaoy@cug.edu.cn)

办公地点：未来城校区地信楼522办公室



# 如何去发现人群的移动模式？



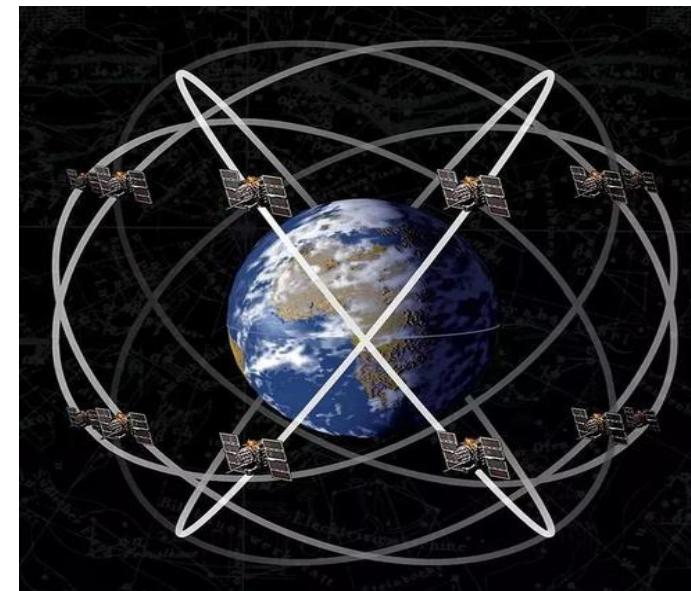


# 主要内容



- **1 轨迹大数据简介**
- **2 轨迹大数据处理与预测**
- **3 轨迹大数据与人群移动模式探究**
- **4 轨迹大数据其它应用**

- 轨迹数据：是具有**时空特征**的，通过对一个或多个移动对象运动过程的**采样**所形成的数据信息。

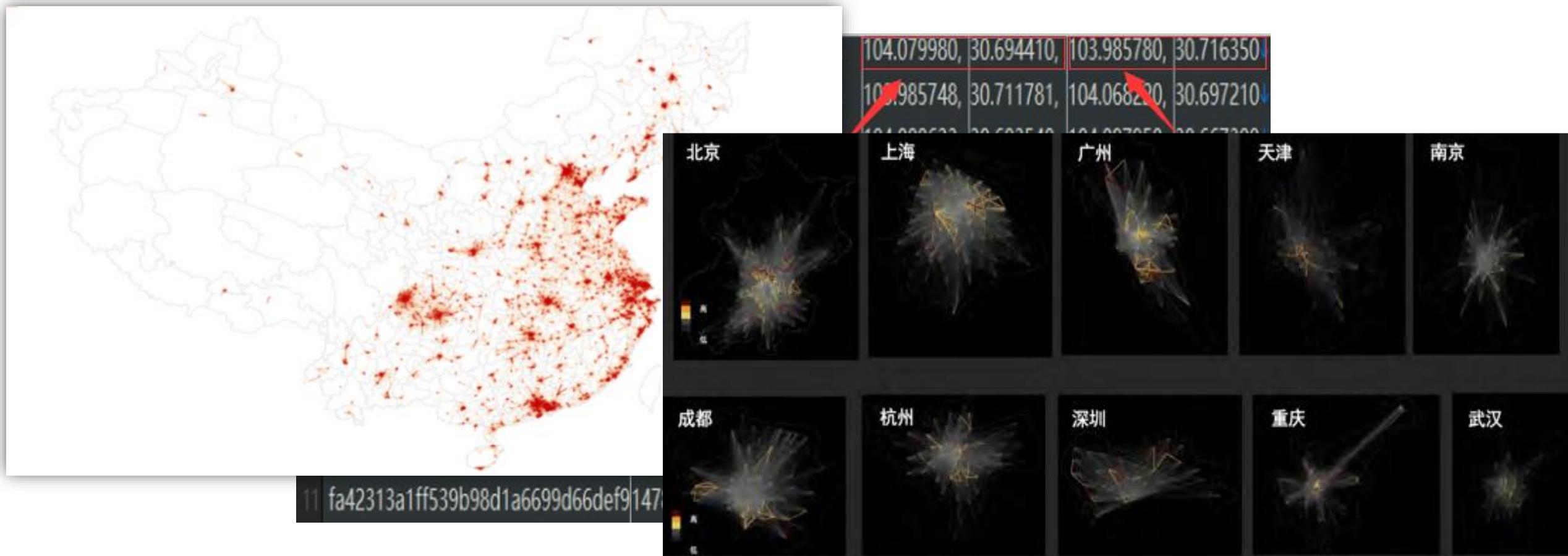




## ➤ 出租车GPS轨迹数据

1	TID,Time,Longitude,Latitude,Speed,Status,Head,AsynFlag↓
2	MMC8000GPSANDASYN051113-17097-00000000,2013-03-21 02:39:35,114.3590,30.5309,12.682777,0,199.340000,GPS↓
3	MMC8000GPSANDASYN051113-16636-00000000,2013-03-21 02:39:35,114.2634,30.5556,15.992222,0,264.530000,GPS↓
4	MMC8000GPSANDASYN051113-38261-00000000,2013-03-21 02:39:34,114.3490,30.5415,4.217486,262144,293.280000,GPS↓
5	MMC8000GPSANDASYN051113-16908-00000000,2013-03-21 02:39:35,114.3507,30.6051,17.122778,0,50.540000,GPS↓
6	MMC8000GPSANDASYN051113-18975-00000000,2013-03-21 02:39:34,114.2631,30.5459,12.513194,262144,70.090000,GPS↓
7	MMC8000GPSANDASYN051113-18436-00000000,2013-03-21 02:39:34,114.3029,30.6160,14.758889,0,303.530000,GPS↓
8	MMC8000GPSANDASYN051113-21131-00000000,2013-03-21 02:39:34,114.2475,30.5855,0.000000,0,0.000000,GPS↓
9	MMC8000GPSANDASYN051113-18388-00000000,2013-03-21 02:39:34,114.3939,30.6404,5.812083,262144,36.660000,GPS↓
10	MMC8000GPSANDASYN051113-20787-00000000,2013-03-21 02:39:34,114.2918,30.6205,7.271528,0,46.670000,GPS↓
11	MMC8000GPSANDASYN051113-20648-00000000,2013-03-21 02:39:33,114.3635,30.6777,10.837916,262144,246.270000,GPS↓
12	MMC8000GPSANDASYN051113-18638-00000000,2013-03-21 02:39:33,114.4676,30.5108,17.544167,262144,320.180000,GPS↓
13	MMC8000GPSANDASYN051113-19763-00000000,2013-03-21 02:39:33,114.2900,30.5307,7.924167,262144,0.440000,GPS↓
14	MMC8000GPSANDASYN051113-19565-00000000,2013-03-21 02:39:33,114.3959,30.5001,9.085556,0,272.430000,GPS↓
15	MMC8000GPSANDASYN051113-18557-00000000,2013-03-21 02:39:33,114.2957,30.5855,10.231527,262144,25.470000,GPS↓

## ➤ 网约车OD轨迹数据



➤ 共享单车骑行轨迹



- 具有GPS模块的共享单车（如所有的摩拜和部分OFO）骑行过程中产生的定位轨迹数据、开关锁记录等
- <https://mobike.com>、<http://www.ofo.so>
- 以及其他共享单车公司推测的辅助数据如用户居住地、就业地、使用习惯、年龄阶段 等，以及故障报告记录
- 2017年共享单车与城市发展白皮书：<https://zhuanlan.zhihu.com/p/26443639>

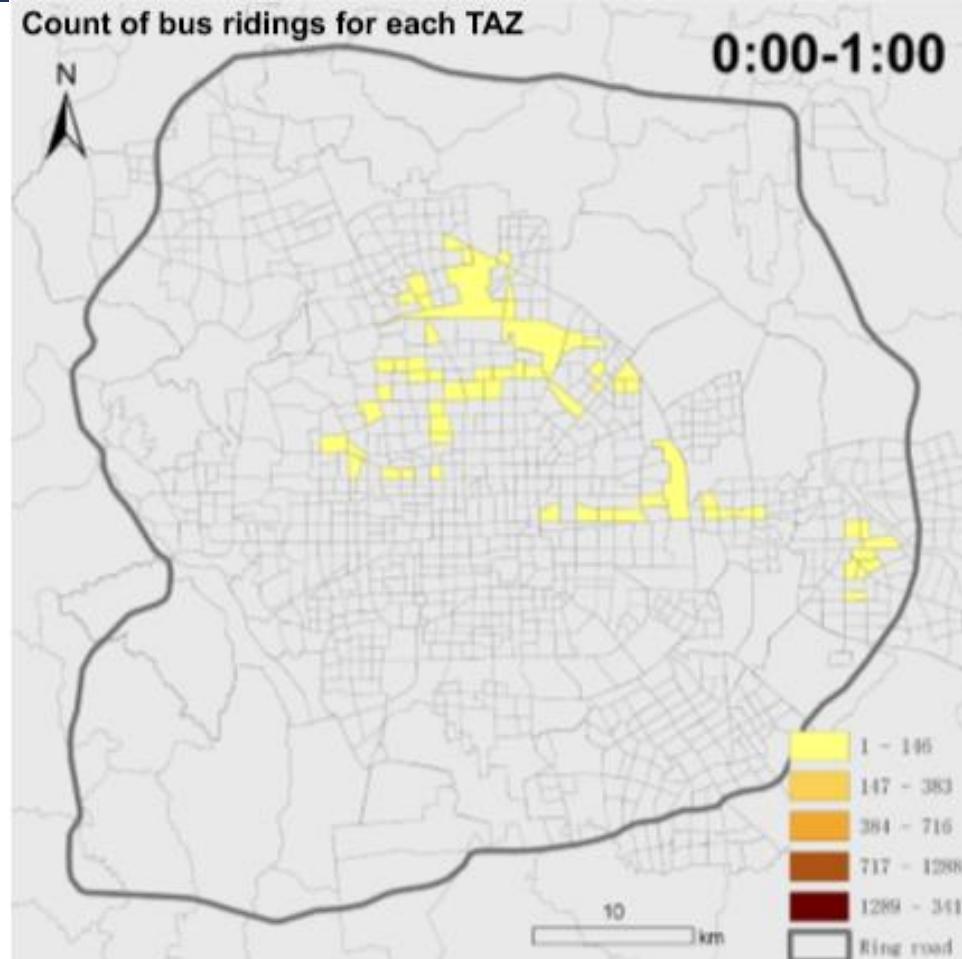
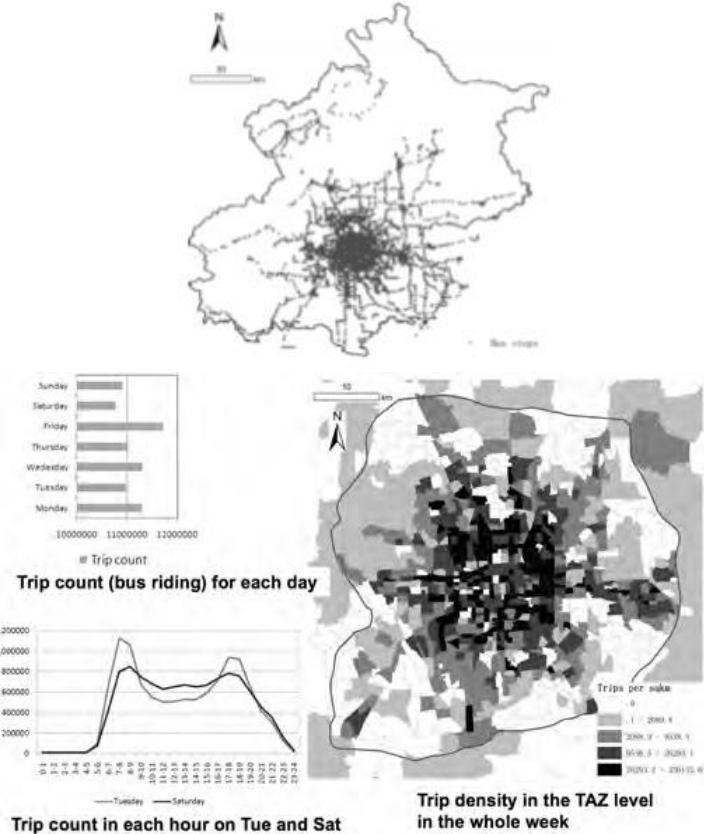
## ➤ 公共交通刷卡记录

- 大量城市的公共交通系统采用智能卡作为交通收费手段 (AFC)
- 公共交通刷卡记录是AFC的副产品
- 空间分辨率为站点metro station/bus stop, 时间分辨率精确到秒
- 轨道交通、分段计价公交、一票制公交
- 辅助数据：公交车GPS、居民家庭出行调查数据、线路、站点、交通分析小区 (TAZ)

Variable	Exemplified Values
Card ID	“10007510038259911”, “10007510150830716”
Card Type	1, 2, 3, 4
Line ID	602, 40, 102
Line Type	0, 1
Driver ID	11032, 332
Vehicle ID	111223, 89763
Departure Date	2008-04-08
Departure Time	“06-22-30”, “11-12-09”
Departure Stop	11, 5, 14
Arrival Time	“09-52-05”, “19-07-20”
Arrival Stop	3, 14, 9



## ➤ 公共交通刷卡记录



- 潜在应用领域：职住平衡、城市贫困、极端出行、乘客画像、线路调整、规划实施评价、群体出行、学生出行、灰色人群、城市功能识别（北京研究：上千万持卡人连续一周近亿次出行）

## ➤ 手机服务



数据收集APP



人的行为轨迹



- 1 轨迹大数据简介
- 2 轨迹大数据处理与预测
- 3 轨迹大数据与人群移动模式探究
- 4 轨迹大数据其它应用



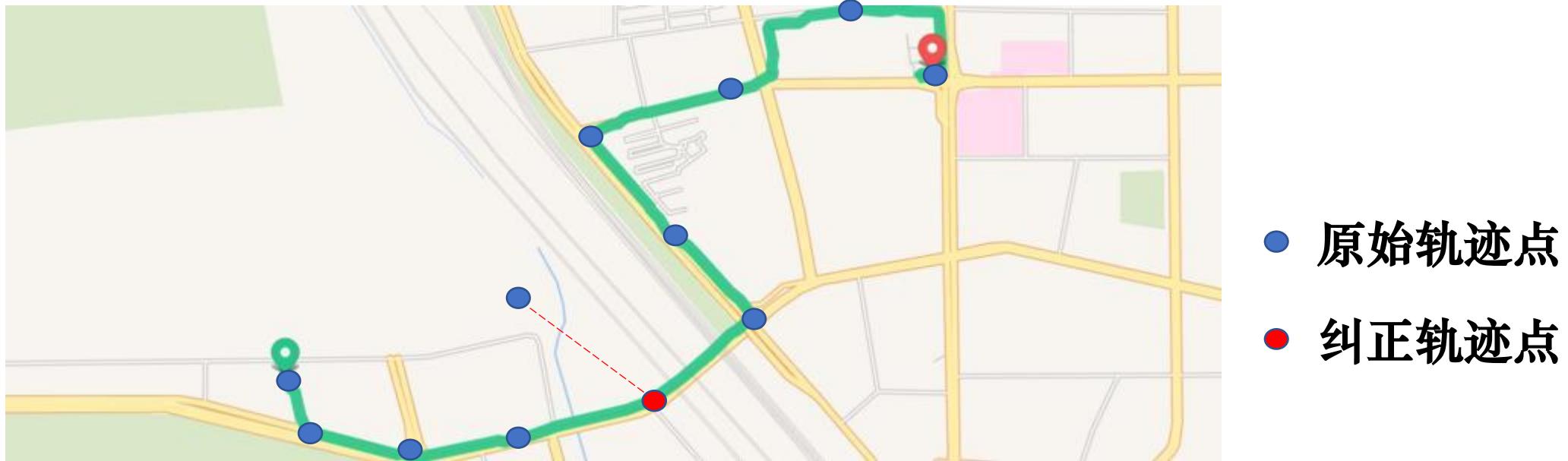
- ✓ 轨迹数据采集的过程中由于硬件设备信号的不稳定，产生异常轨迹，容易丢失部分轨迹点，产生**无效**数据。
- ✓ 连续性的运动轨迹被离散化表示，易受到采样精度、位置的不确定影响，记录的轨迹点存在**数据质量**问题。
- ✓ 轨迹大数据普遍存在的问题还包括**数据冗余、信息重复**等。

## 2.1 | 轨迹大数据的处理



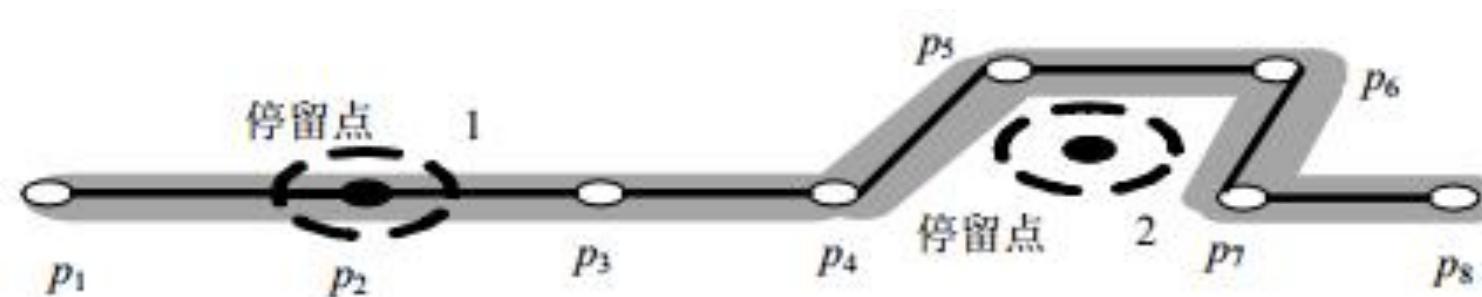
### ➤ 轨迹清洗

数据进行重新审查和校验的过程，目的在于删除重复信息、纠正存在错误，并提供数据一致性。



### ➤ 停留点

时空轨迹中存在的轨迹点，其重要性并不相同，往往轨迹中某些点反映了人们一段时间的行为，例如购物、观光某个旅游景点等。在轨迹数据中，在某一时间区域或空间区域内产生某种行为的轨迹数据定义为停留点。



### ➤ 停留点监测方法

停留点分为两类：静止停留点、轨迹环绕停留点

静止型停留点：具体表现在轨迹数据中相邻的多个时刻停留在同一位置。可通过**距离阈值**、**速度**等监测识别。

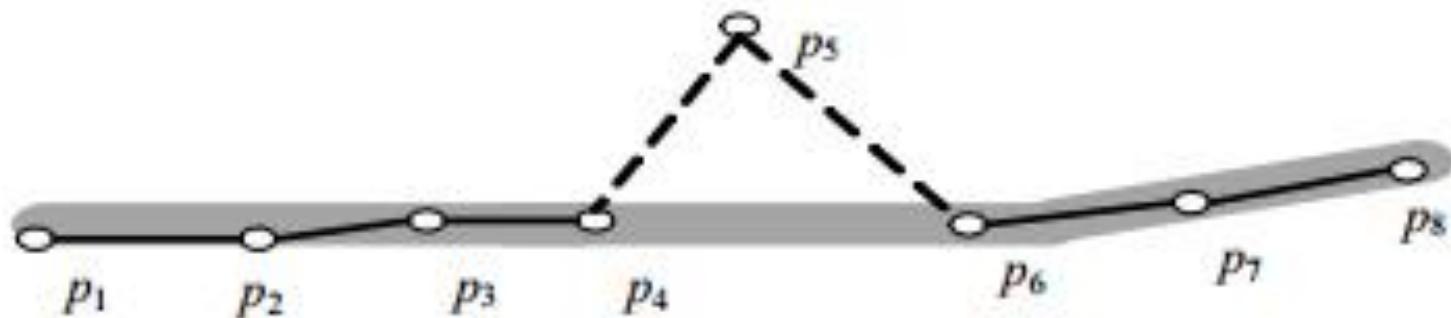
徘徊型停留点：具体表现在轨迹中部分相邻的点在一定**距离阈值**和**时间阈值**内。该类型停留点可通过找出距离阈值外的一点，检验时间跨度是否超过阈值，若超过则将中间的点识别为停留点。

## 2.1 | 轨迹大数据的处理



### ➤ 噪音点

由于GPS设备定位异常所输出的点。

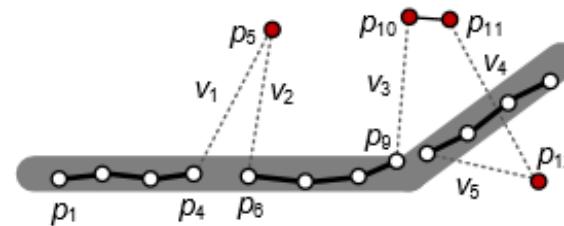


## ➤ 噪音点过滤方法

均值、中值滤波

优点是：均值（中值）滤波器可以看作是一个滑动窗口，覆盖n个时间上相邻的值。

缺点是：处理多个连续噪音点时，误差较大不能取得很好的效果。



卡尔曼滤波

优点是：卡尔曼滤波是通过假设线性模型修正噪音点，可以包含更多的状态向量，例如速度。

缺点是：存在一定的局限性，对于移动对象受限于预定义的路径。

### 轨迹大数据处理步骤

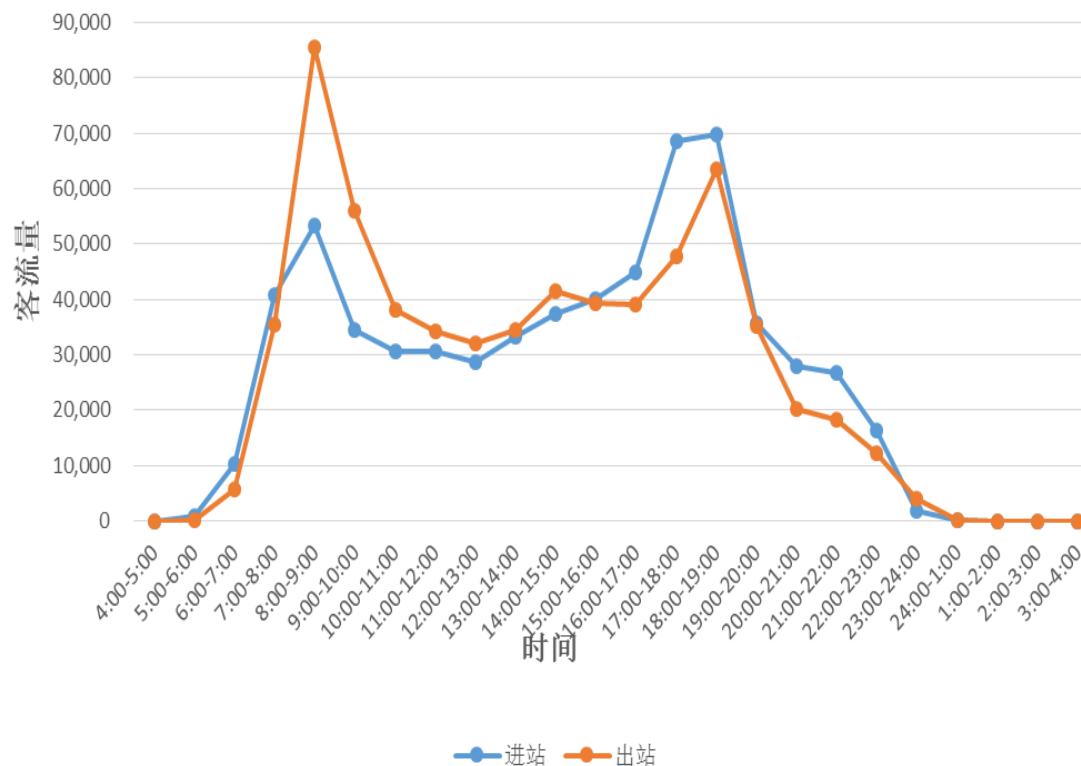
1. 对轨迹数据进行数据清洗，主要剔除掉**重复**的轨迹数据和**无效**的轨迹数据；
2. 设定停留时间阈值与停留距离阈值，对轨迹数据进行停留点监测，去除**冗余**轨迹点数据；
3. 计算相邻点的距离，设置最大可容忍距离，去除**漂移**数据；
4. 此时形成了初步**有效的轨迹序列数据集**，按照分析方法的输入要求组织成适合数据分析的样式。

## 2.2 | 轨迹大数据的分析

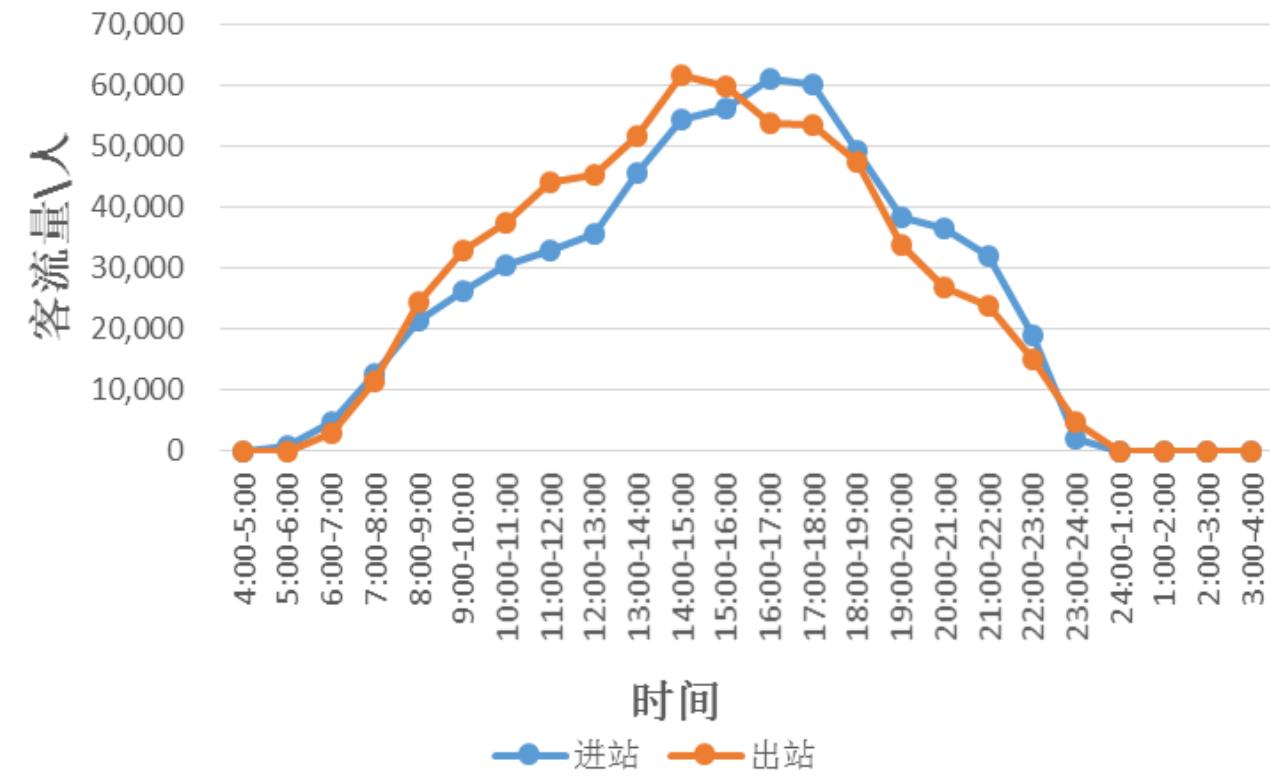


### ➤ 轨迹分析

通过构建模型、指标等方法可以挖掘数据所隐含的信息或规律。



2014年3月3日（星期一）广州市地铁1号线  
进出站客流量时间曲线图



2014年3月9日（星期日）广州市地铁1号线  
进出站客流量时间曲线图

### ➤ 轨迹预测

对事件的全面预测，不仅要能够指出事件发生的各种可能结果，而且还必须给出每一种结果出现的概率。这就叫做**事件发生概率预测**。

马尔科夫（Markov）预测方法，就是一种预测事件发生概率的方法。它是基于马尔科夫链，根据事件的目前状况预测其将来各个时刻（或时期）变动状况的一种预测方法。马尔科夫预测法是对地理事件进行预测的基本方法，它是地理预测中常用的重要方法之一。

### ■ 状态

指某一事件在某个时刻（或时期）出现的某种结果。比如：商品销售“畅销”、“一般”、“滞销”等状态。农业收成预测中，有“丰收”、“平收”、“歉收”等状态。

### ■ 状态转移过程

事件的发展，从一种状态转变为另一种状态，称为状态转移。

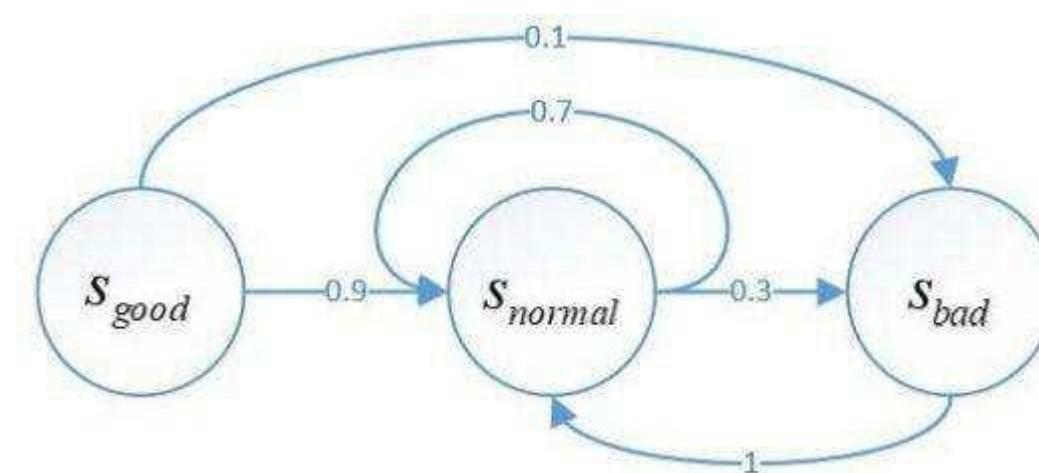
### ■ 马尔科夫过程

在事件的发展过程中，若每次状态的转移都仅与前一时刻的状态有关，而与过去的状态无关，或者说状态转移过程是无后效性的，则这样的状态转移过程就称为马尔科夫过程。

## ■ 状态转移概率

在事件的发展变化过程中，从某一种状态出发，下一时刻转移到其他状态的可能性，称为状态转移概率。由状态 $E_i$ 转为状态 $E_j$ 的状态转移概率  $P(E_i \rightarrow E_j)$  就是条件概率  $P(E_j / E_i)$  即

$$P(E_i \rightarrow E_j) = P(E_j / E_i) = P_{ij} \quad (1)$$





## ■ 状态转移概率矩阵

假定某一个事件的发展过程有 $n$ 个可能的状态，即 $E_1, E_2, \dots, E_n$ 。记为从状态 $E_i$ 转变为状态 $E_j$ 的状态转移概率，则矩阵 (2) 称为状态转移概率矩阵。

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix} \quad (2)$$

## ■ 概率矩阵

$$\begin{cases} 0 \leq P_{ij} \leq 1 & (i, j = 1, 2, \dots, n) \\ \sum_{j=1}^n P_{ij} = 1 & (i = 1, 2, \dots, n) \end{cases} \quad (3)$$

一般地，将满足条件（3）的任何矩阵都称为随机矩阵，或概率矩阵。

不难证明，如果 $P$ 为概率矩阵，则对于任何整数 $m > 0$ ， $P^m$ 矩阵都是概率矩阵。

## ■ 标准概率矩阵、平衡向量

如果  $P$  为概率矩阵，而且存在整数  $m > 0$ ，使得概率矩阵  $P^m$  中诸元素皆非零，则称  $P$  为标准概率矩阵。可以证明，如果  $P$  为标准概率矩阵，则存在非零向量

$$\alpha = [x_1, x_2, \dots, x_n], \text{ 而且 } x_i \text{ 满足 } 0 \leq x_i \leq 1, \sum_{i=1}^n x_i = 1$$

使得  $\alpha P = \alpha \quad (4)$

这样的向量  $\alpha$  称为平衡向量，或终极向量。这就是说，标准概率矩阵一定存在平衡向量。



## ■ 状态转移概率矩阵的计算

计算状态转移概率矩阵 $P$ ，就是求从每个状态转移到其他任何一个状态的状态转移概率  $P_{ij}$  ( $i,j = 1,2, \dots, n$ )

为了求出每一个，一般采用**频率近似概率**的思想进行计算。



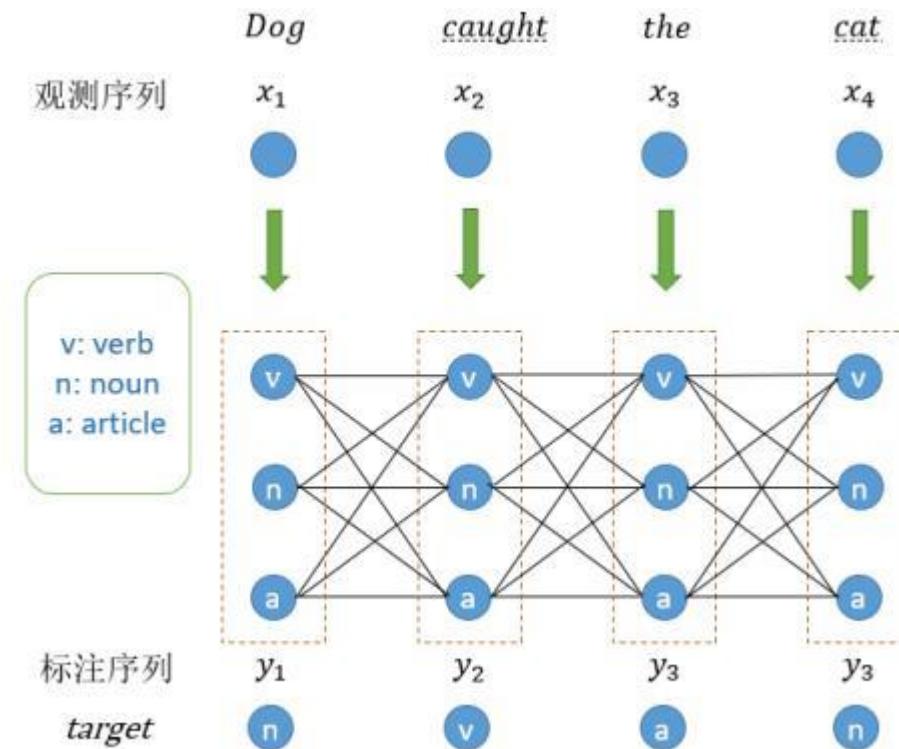
马尔科夫预测方法的基本要求是状态转移概率矩阵必须具有一定的稳定性。因此，必须具有足够的统计数据，才能保证预测的精度与准确性。

换句话说，马尔科夫预测模型必须建立在大量的统计数据的基础之上。这一点也是运用马尔科夫预测方法预测地理事件的一个最为基本的条件。

## 2.3 | 轨迹大数据预测



条件随机场 (conditional random fields, 简称 CRF) 由Lafferty等人于2001年提出，它是一种概率无向图的判别模型。该模型具有表达**长距离依赖性**的能力，能够较好地标注（分类）偏置等问题。



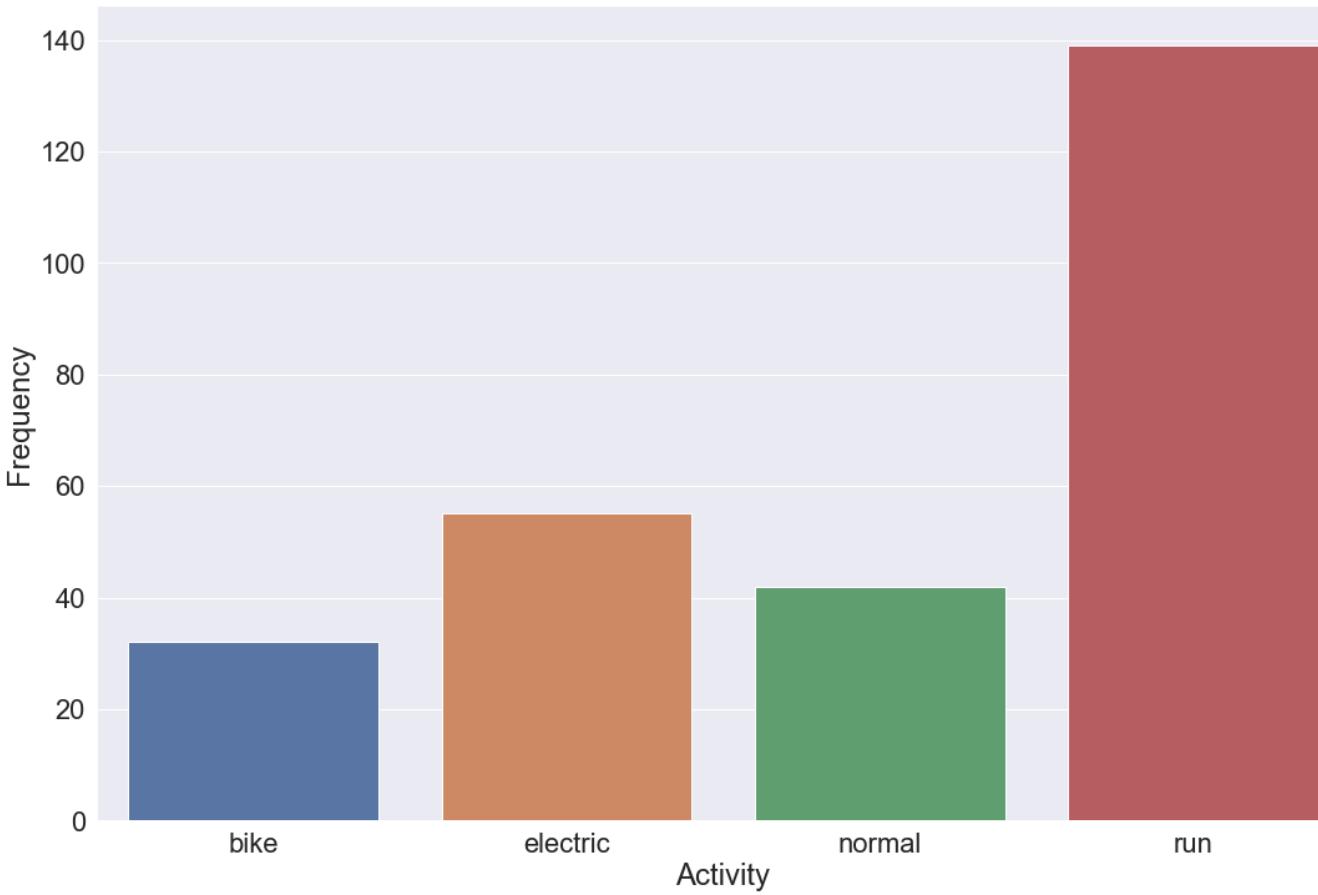


# 主要内容

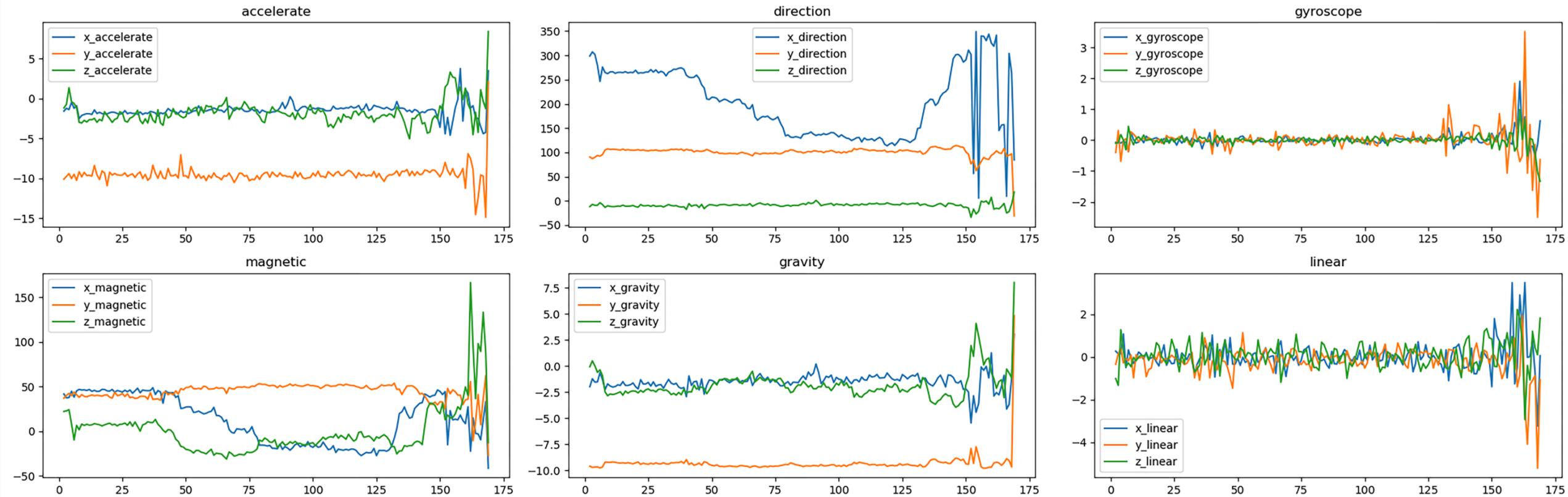


- 1 轨迹大数据简介
- 2 轨迹大数据处理与预测
- 3 轨迹大数据与人群移动模式探究
- 4 轨迹大数据其它应用

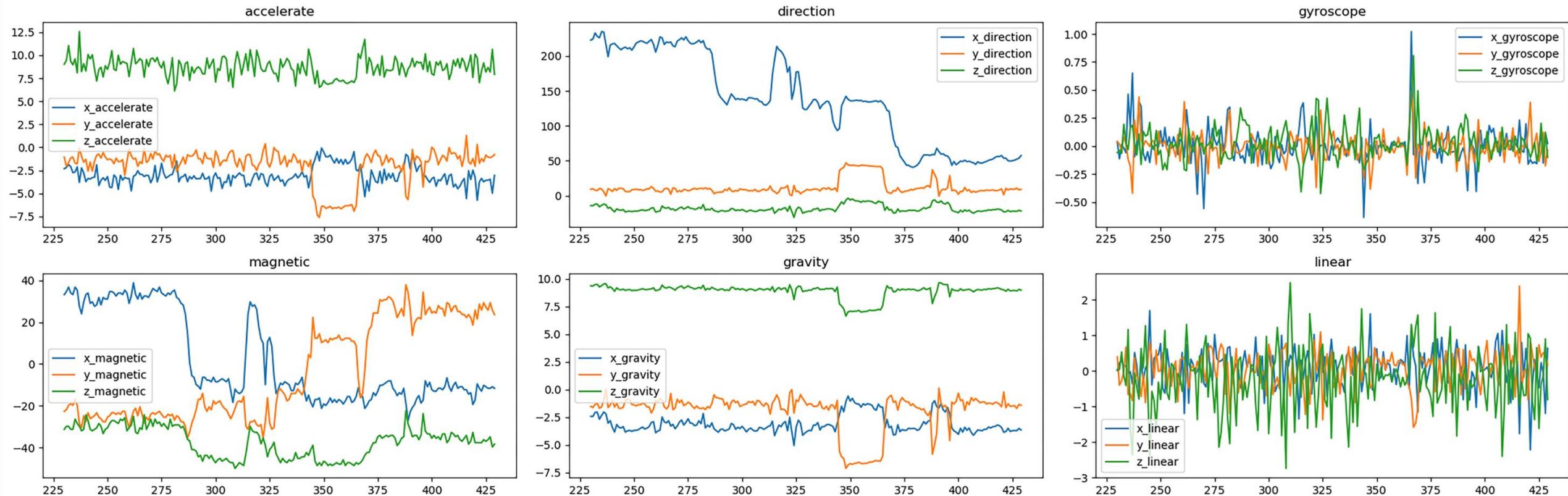
## ■ 研究数据



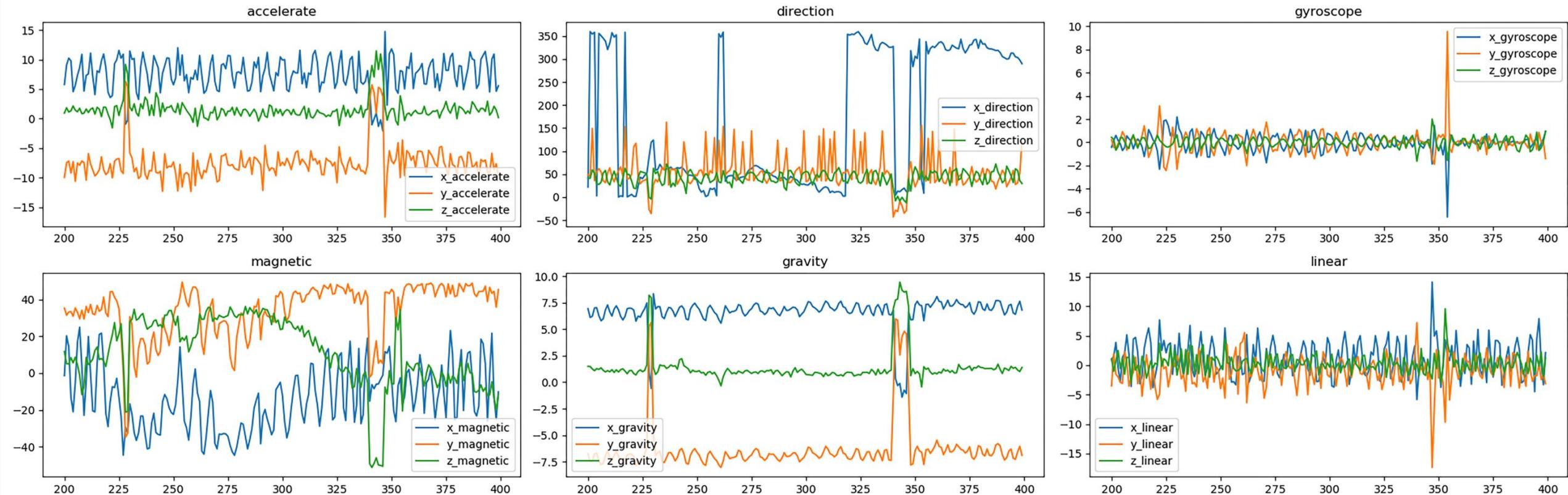
- 志愿者人数: XXX人
- 采样频率: 60 Hz
- 测试场景: 手机放在裤子口袋
- 运动模式类型: 4 种
  - 骑自行车
  - 骑电动车
  - 步行
  - 正常跑步
- 传感器类型:
  - 经纬度坐标
  - 三轴方向传感器
  - 三轴线性加速度传感器
  - 重力加速度传感器
  - 三轴陀螺仪传感器
  - 三轴磁场传感器



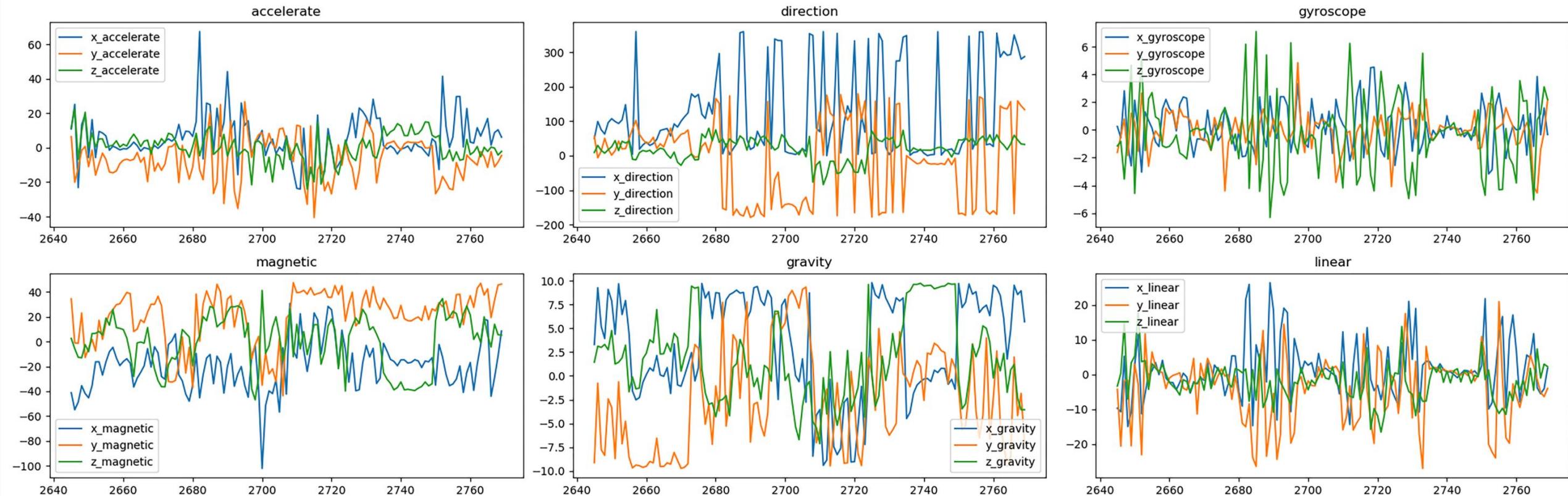
手机传感器数据—自行车



手机传感器数据—电动车

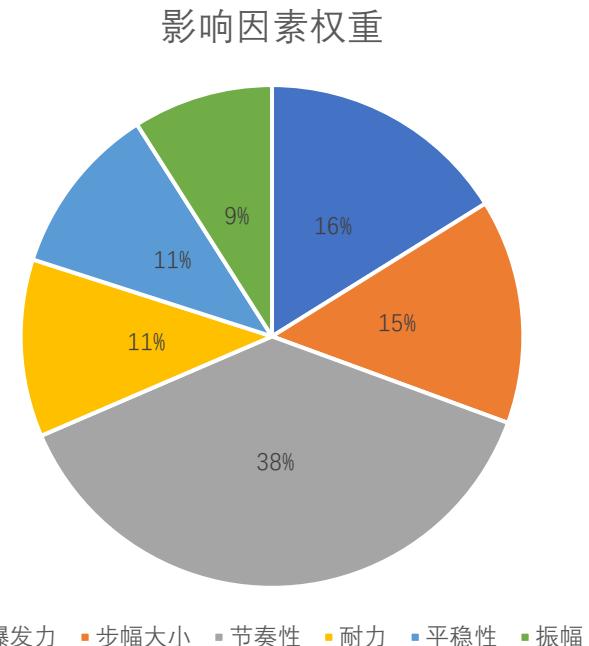


手机传感器数据—步行



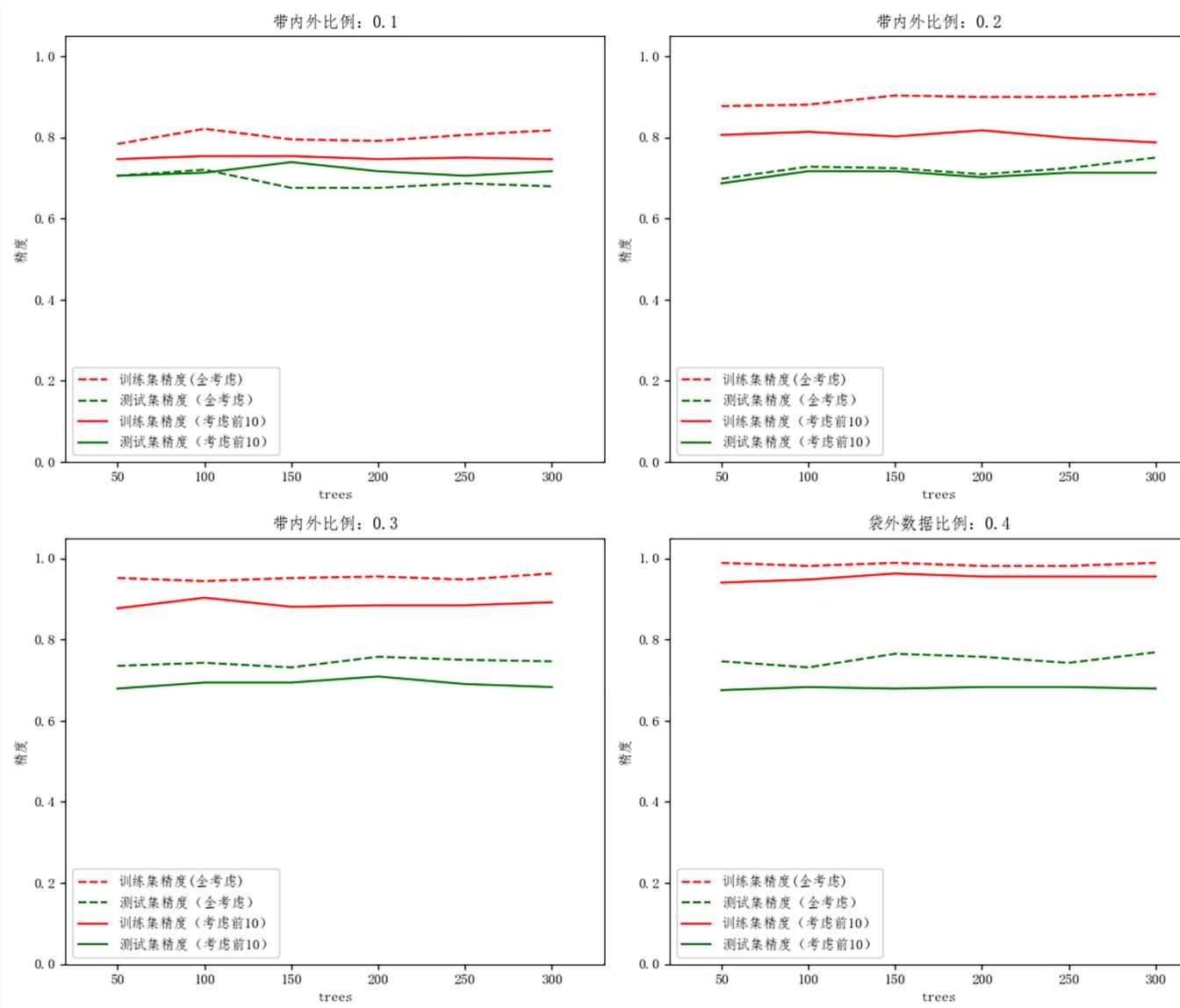
手机传感器数据—跑步

## ■ Runner DNA构建



受到自身因素和环境因素等多方面的影响，人们的跑步风格是混合因素的组合，没有明确的定义和独特的量化指标。但是可以结合数据特征的意义和运动时人体姿态变化、实际运动场所等多方面因素，对运动者跑步风格进行分析。

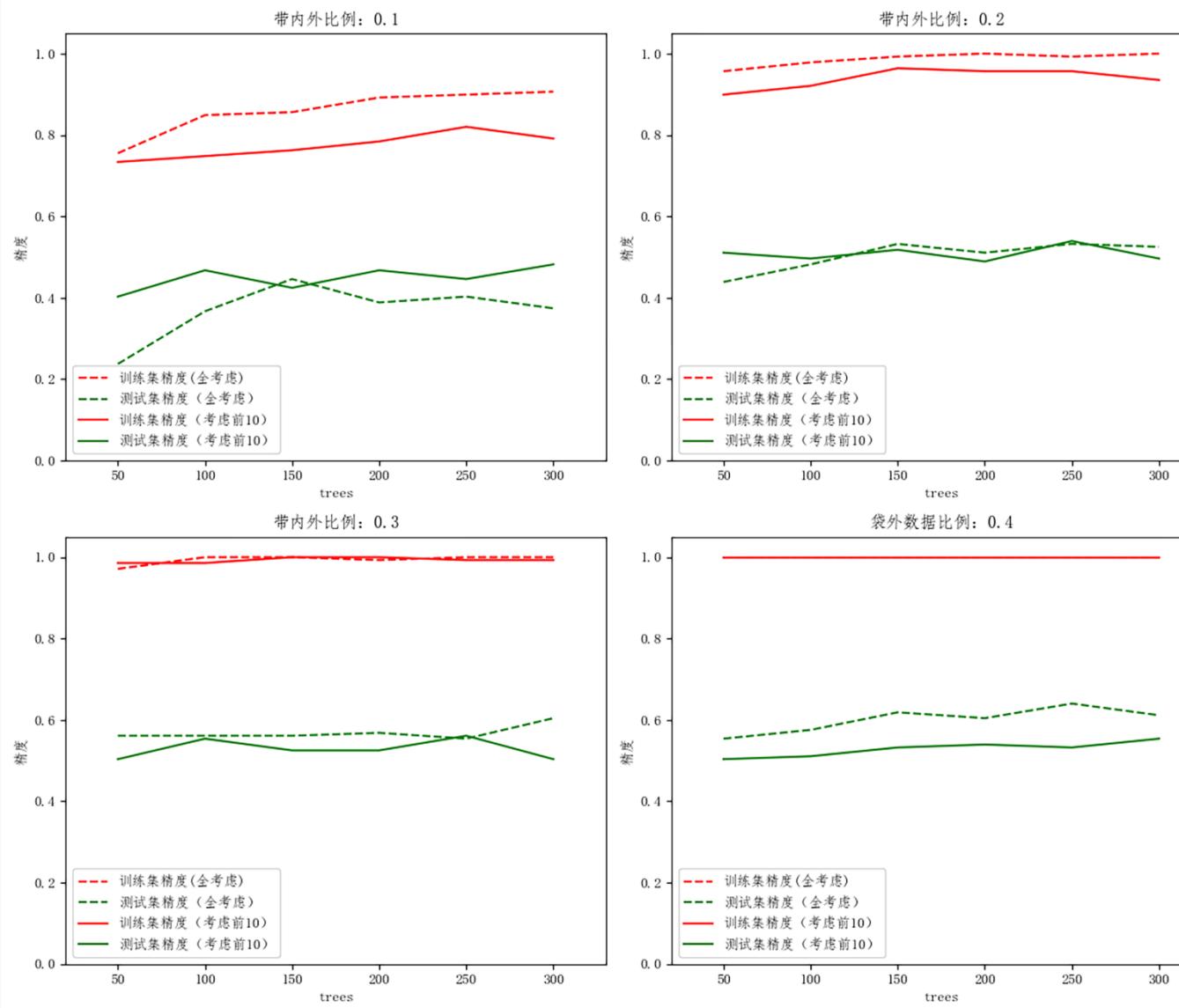
本研究提出“Runner DNA”的概念来描述人的复杂运动行为。将随机森林模型中权重最大的10个特征进行合理分析与解释，将10个特征解释为6个和跑步风格相关的维度，并以此为依据构建Runner DNA。



在运动方式识别方面（跑步、走路、骑自行车、骑电瓶车）：

随机森林模型在考虑所有特征数据时：  
训练集精度：0.989； 测试集精度：0.769

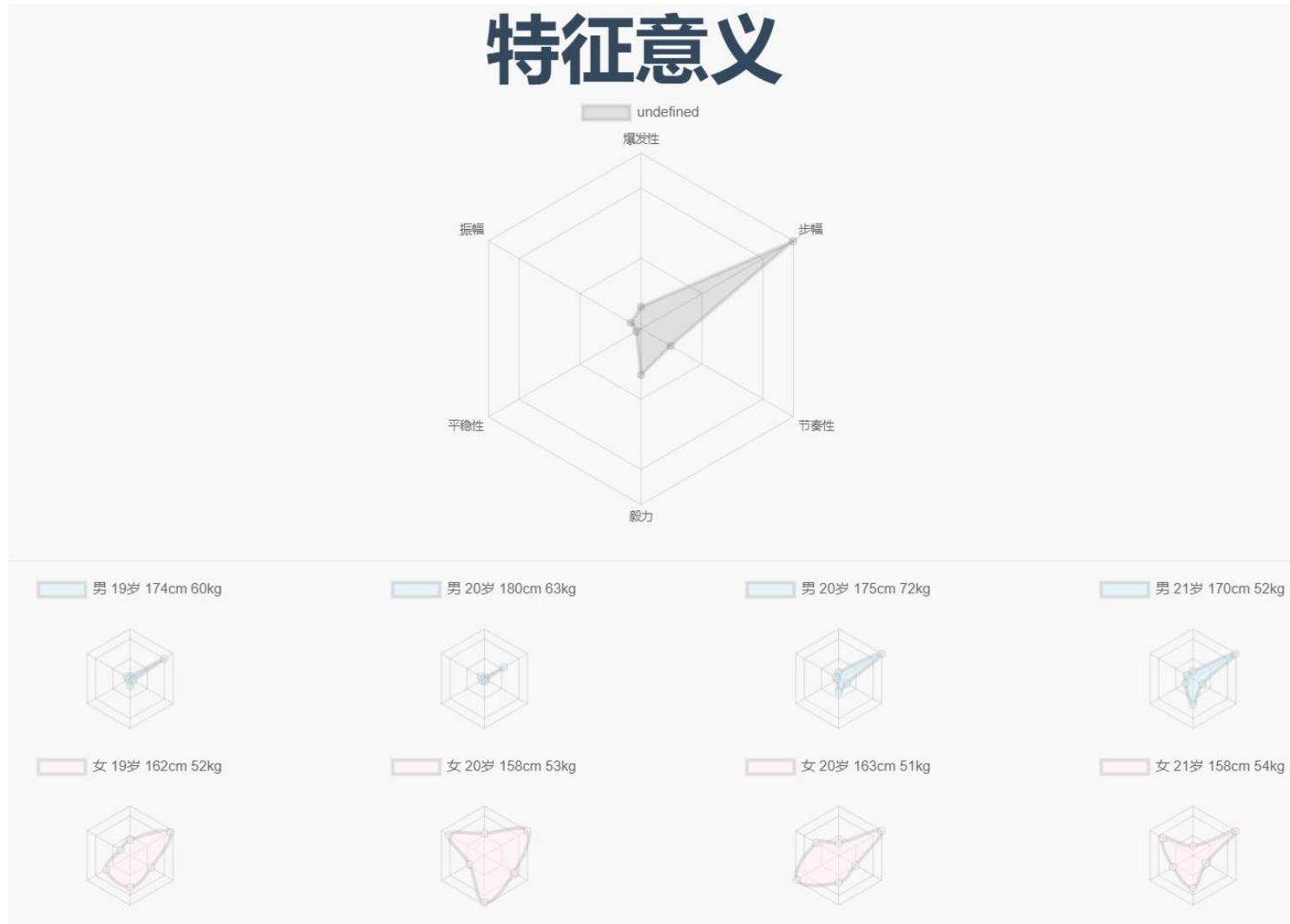
随机森林模型考虑最重要的10个特征数据时：  
训练集精度：0.963； 测试集精度：0.739

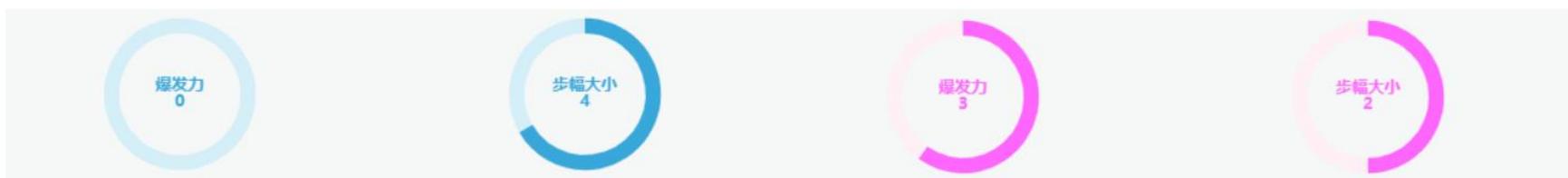
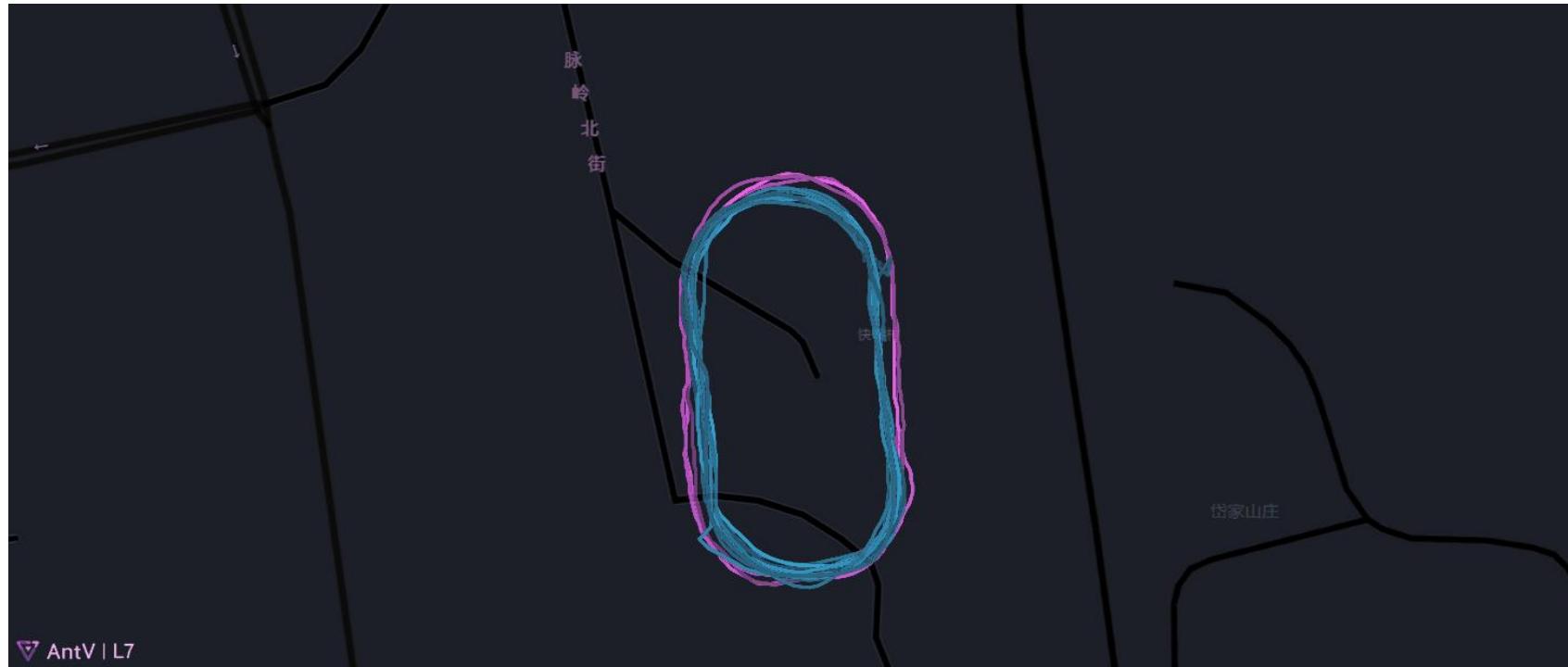


在跑步者识别方面（识别20个跑步者）：

随机森林模型在考虑所有特征数据时：  
训练集精度：1.0； 测试集精度：0.640

随机森林模型考虑重最重要的10个特征数据时：  
训练集精度：1.0； 测试集精度：0.561





— 男性  
— 女性



# 主要内容



- 1 轨迹大数据简介
- 2 轨迹大数据处理与预测
- 3 轨迹大数据与人群移动模式探究
- 4 轨迹大数据其它应用

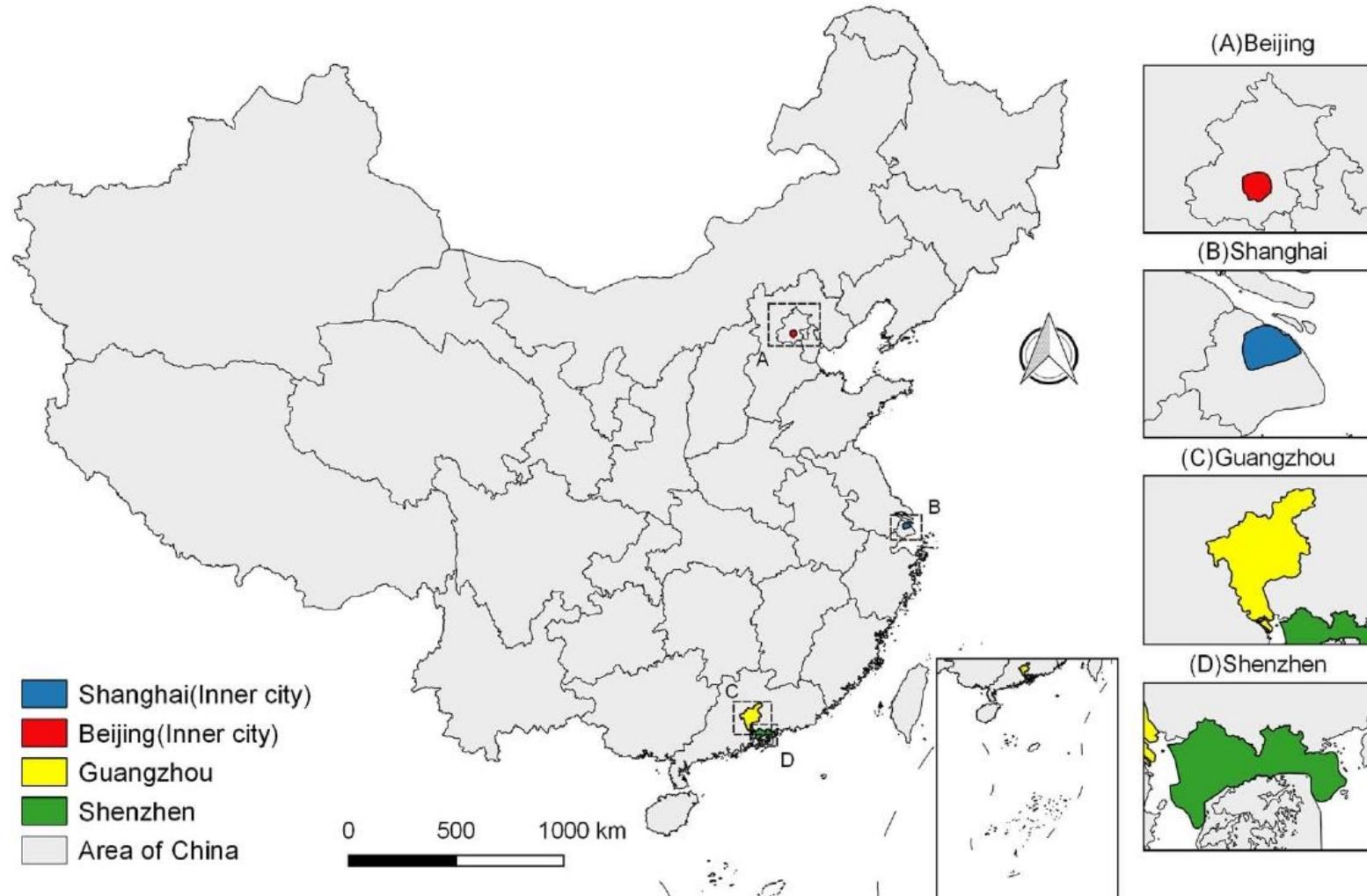
## ■ 研究背景

- 2016年初，中国中央政府发布了城市规划指导方针，称为社区开放政策(新华社2016年);
- 新建社区不再封闭，旧封闭式社区实行公路网开放;
- 有评论认为社区开放将改善路网结构，从而缓解城市交通拥堵的情况(中国日报2016)。

# 4.1 | 小区开放政策研究



## ■ 研究区域和数据描述



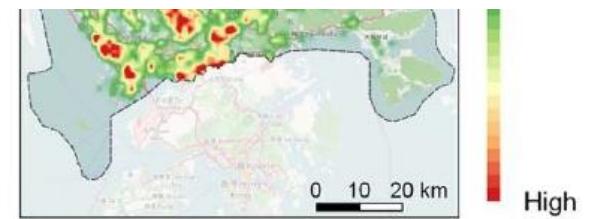
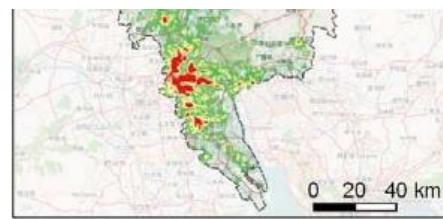
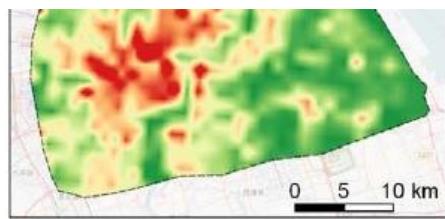
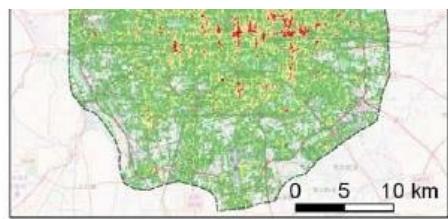
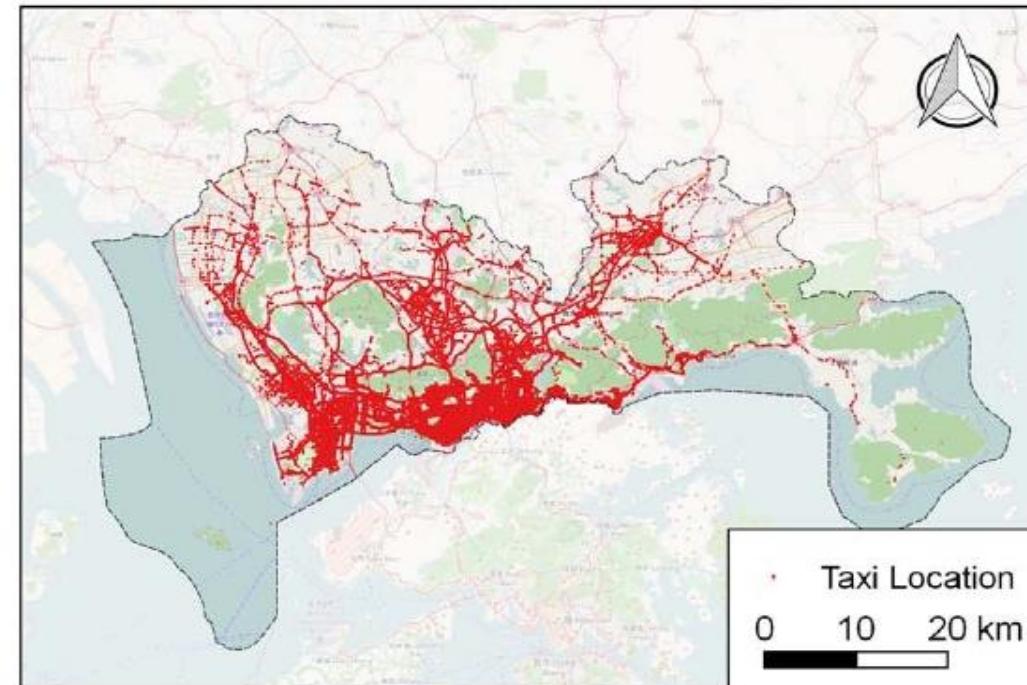
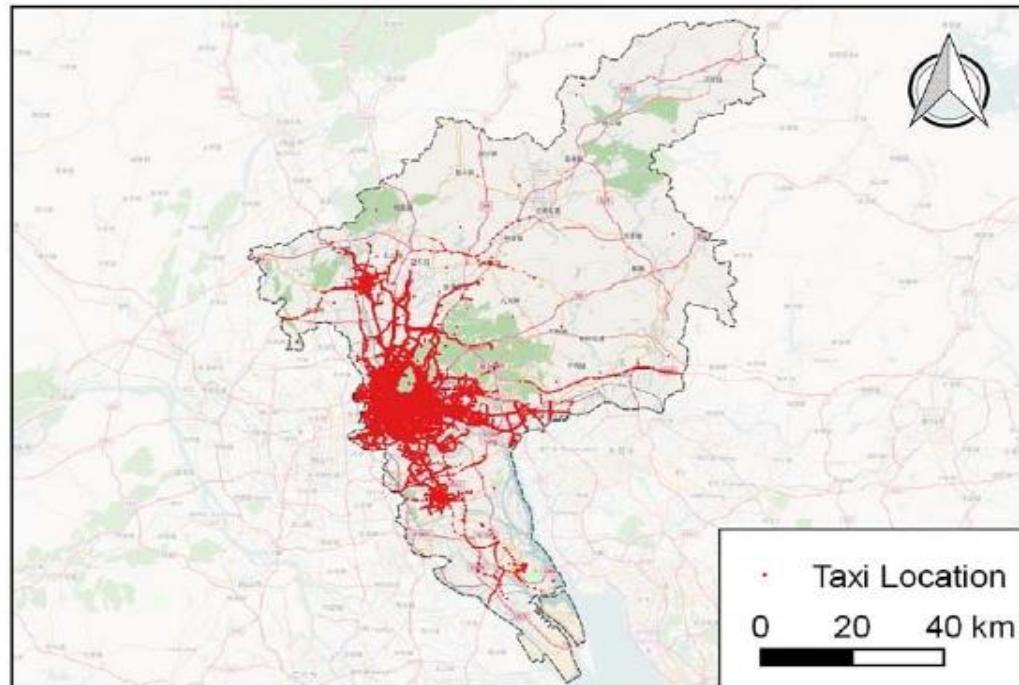
- 北京、上海、广州、深圳
- 研究区面积超过1900Km<sup>2</sup>
- 常住人口超过6500万

# 4.1 | 小区开放政策研究



## ■ 研究区域和数据描述

GPS浮动车数据



(E)

(F)

(G)

(H)

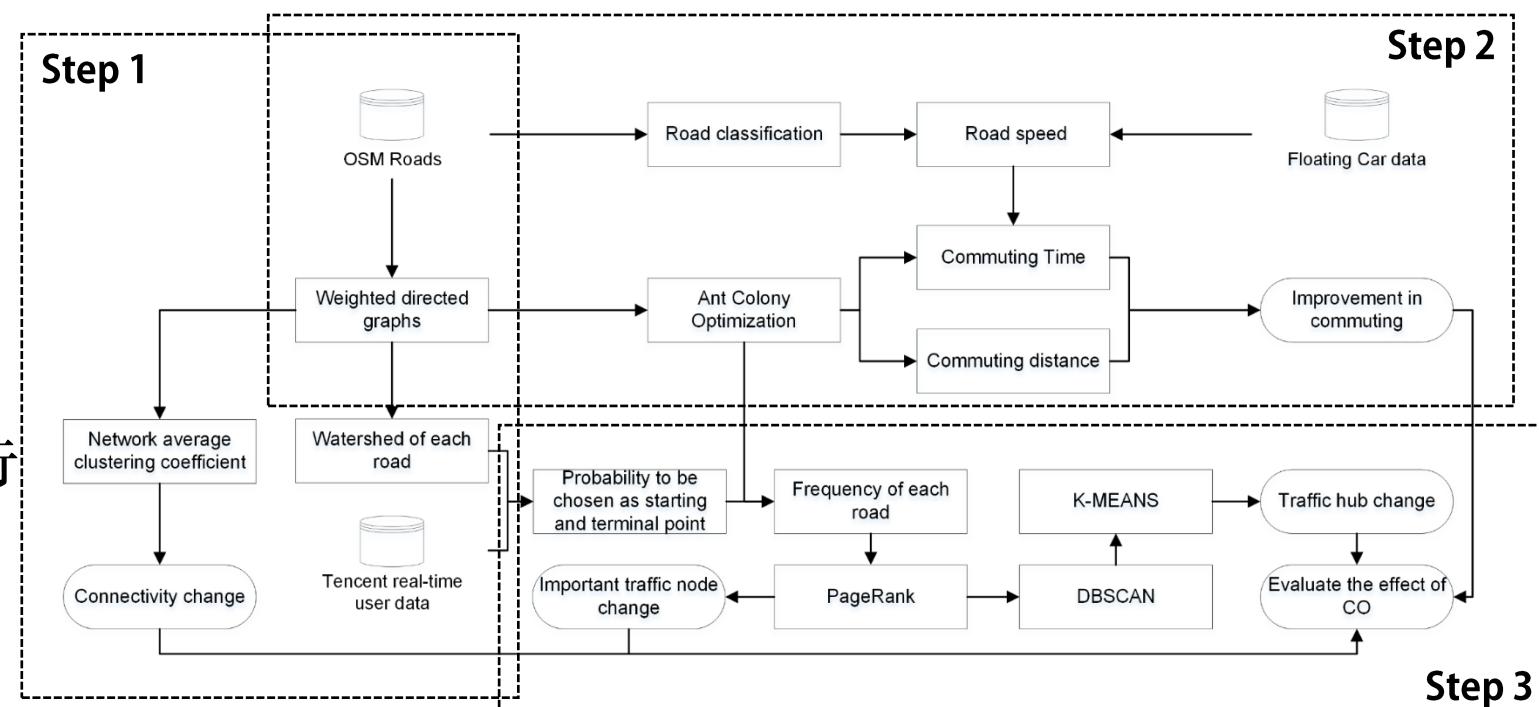
# 4.1 | 小区开放政策研究



Step1: 基于OSM路网数据和小区开放政策，建立城市路网模型

Step2: 利用出租轨迹数据，计算居民出行时间和出行距离

Step3: 分析交通拥堵区域的空间位置和变化



Workflow

# 4.1 小区开放政策研究

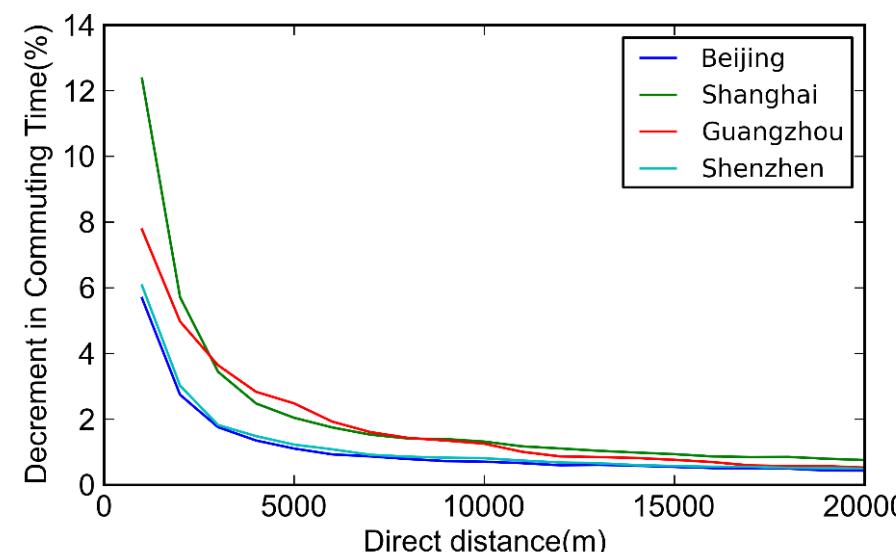
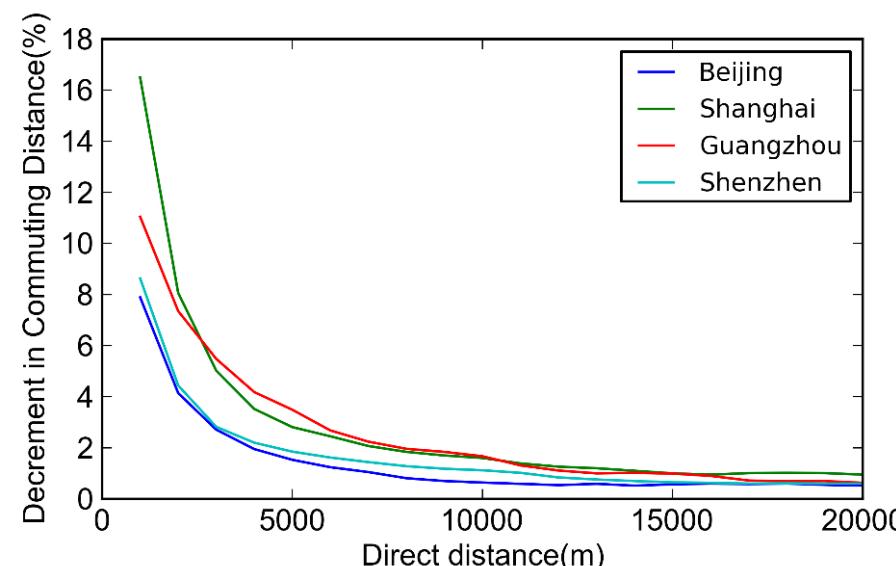


## ■ 网络连接

Table 1 ACC of the study area

City	Before CO	After CO	Improvement
Beijing	0.437	0.507	16.15%
Shanghai	0.481	0.526	9.43%
Guangzhou	0.398	0.516	29.80%
Shenzhen	0.455	0.539	18.69%

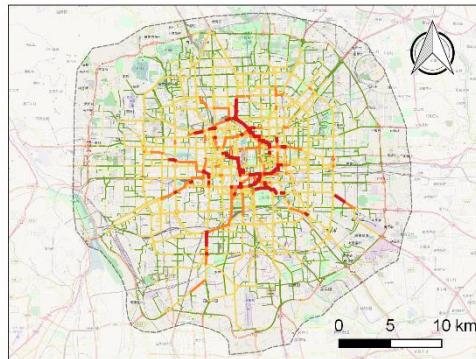
## ■ 通勤效率



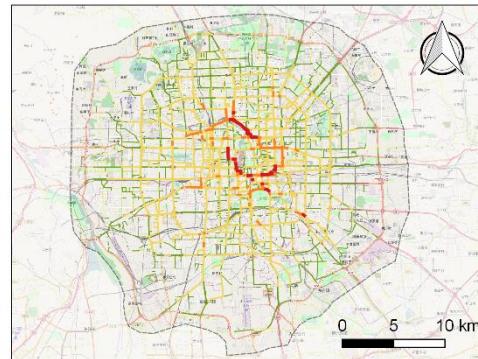
# 4.1 | 小区开放政策研究



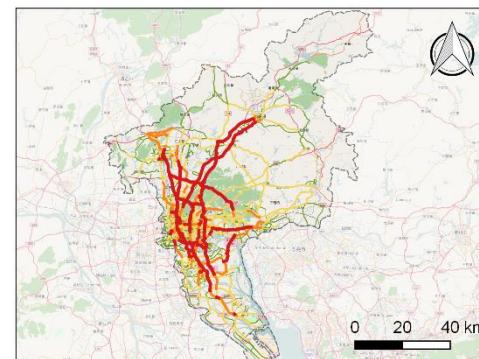
## ■ 交通流变化



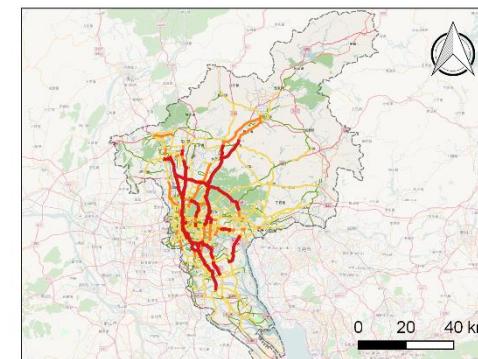
(A1)



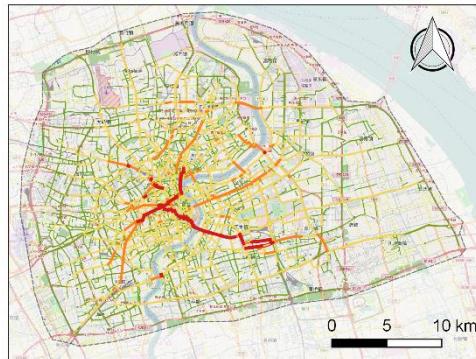
(A2)



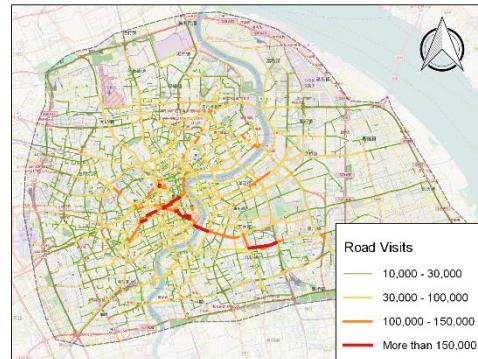
(A1)



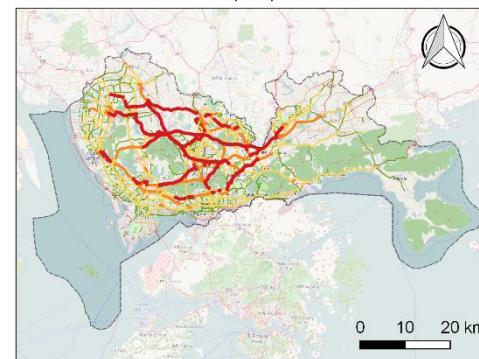
(A2)



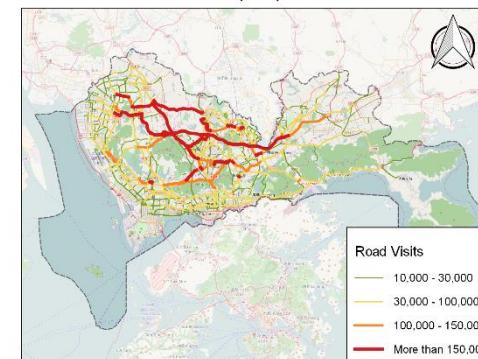
(B1)



(B2)



(B1)



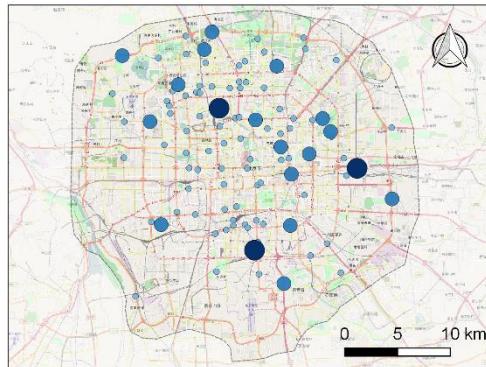
(B2)

Road visits of before and after the implementation of Community opening.

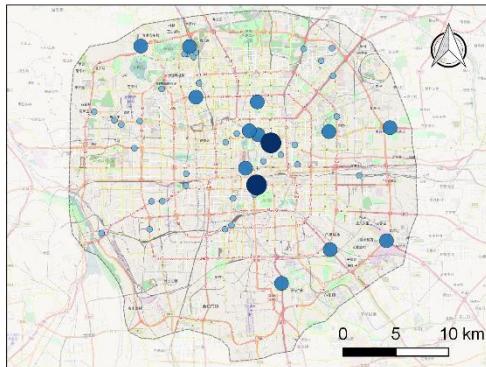
# 4.1 | 小区开放政策研究



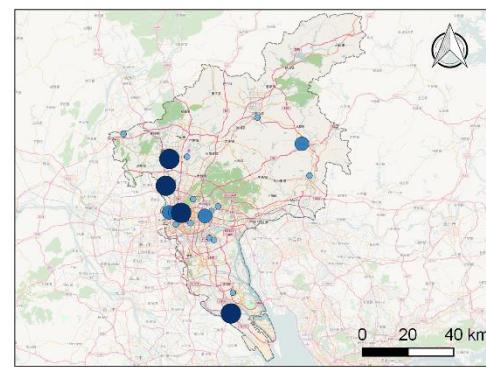
## ■ 交通拥堵点



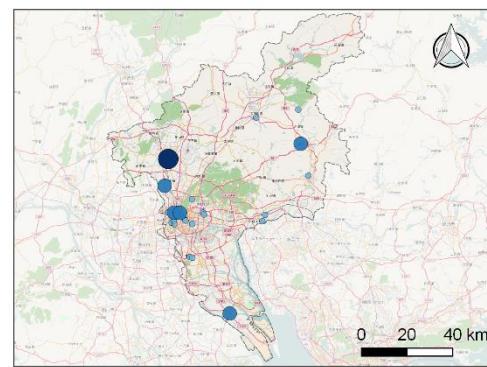
(A1)



(A2)



(A1)



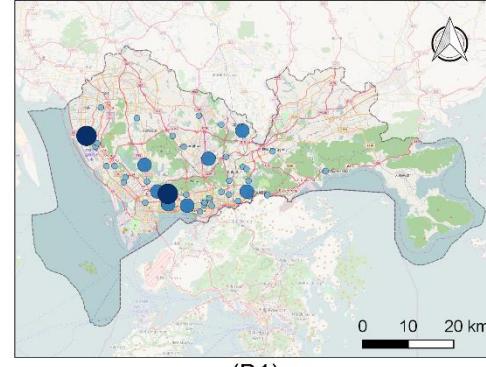
(A2)



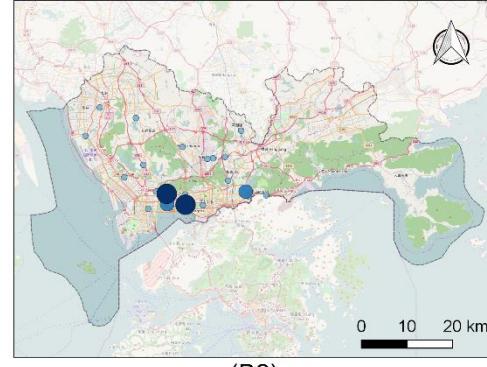
(B1)



(B2)



(B1)



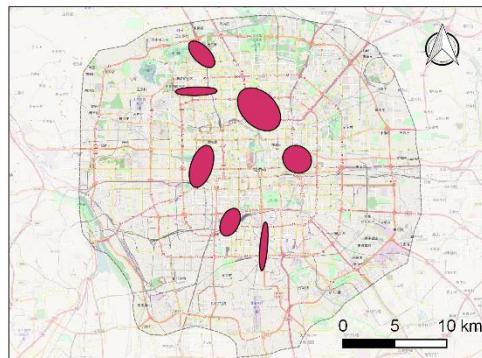
(B2)

High PageRank vertexes before and after the implementation of Community Opening

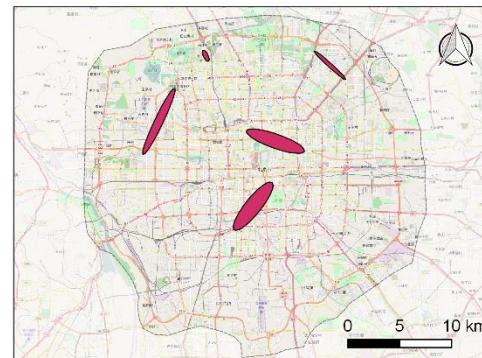
# 4.1 小区开放政策研究



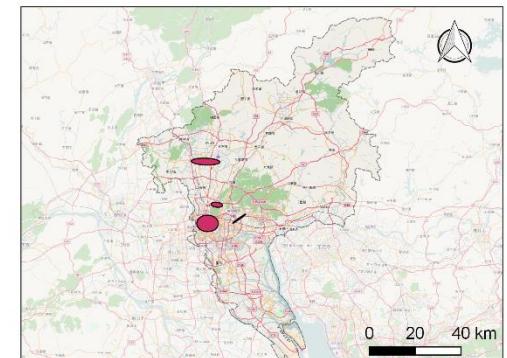
## ■ 交通拥堵区域



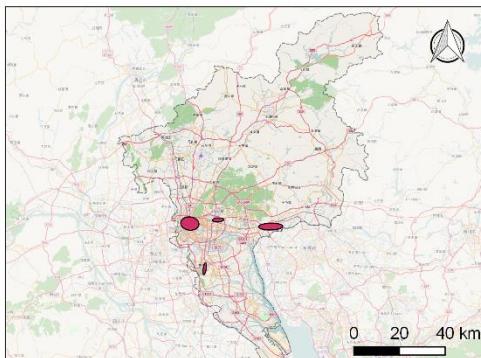
(A1)



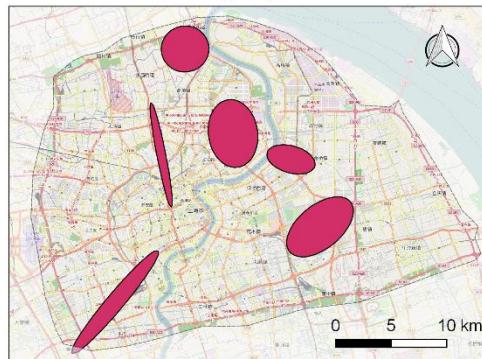
(A2)



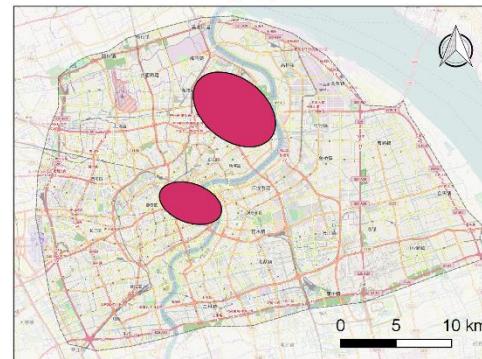
(A1)



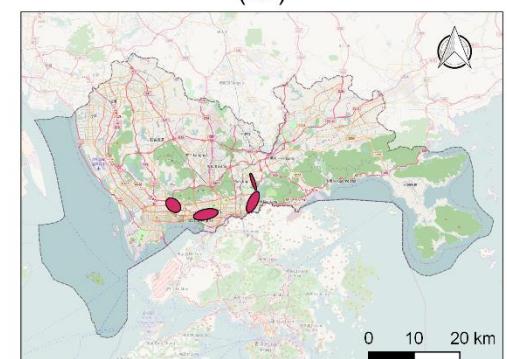
(A2)



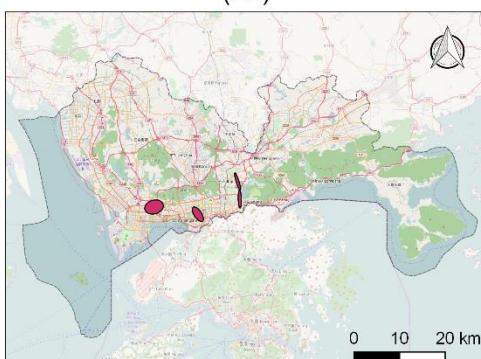
(B1)



(B2)



(B1)



(B2)

Traffic congestion area before and after the implementation of Community Opening

## ■ 研究背景

- 城市是衡量一个**国家经济发展水平**的重要标志，而伴随着中国城市化进程的快速发展，城市中发生的**交通拥堵问题**对于人们的日常生活产生了一定的负面影响（郭继孚等 2011）。
- 交通拥堵问题受限于各种各样的因素，其中包括城市的管理水平，人口密度以及**天气状况**等等（戴东昌等 2004）。在不同的天气状况下，**交通流**会受到一定的影响，导致城市的**交通拥堵中心**发生一定的变化（张金满等 2016）。
- **大数据**的应用和发展，为研究不同天气下的交通拥堵状况提供了便利（Manyika et al. 2011, 陈美 2012）。
- 对深圳市民的出行规律以及交通拥堵中心的变化进行探究和分析对于**解决城市的交通拥堵问题**有着重要的作用，也为之后的**城市建设**提供了新的思路。

# 4.2 | 降雨天气对城市交通拥堵中心变化分析



## ■ 研究区和数据

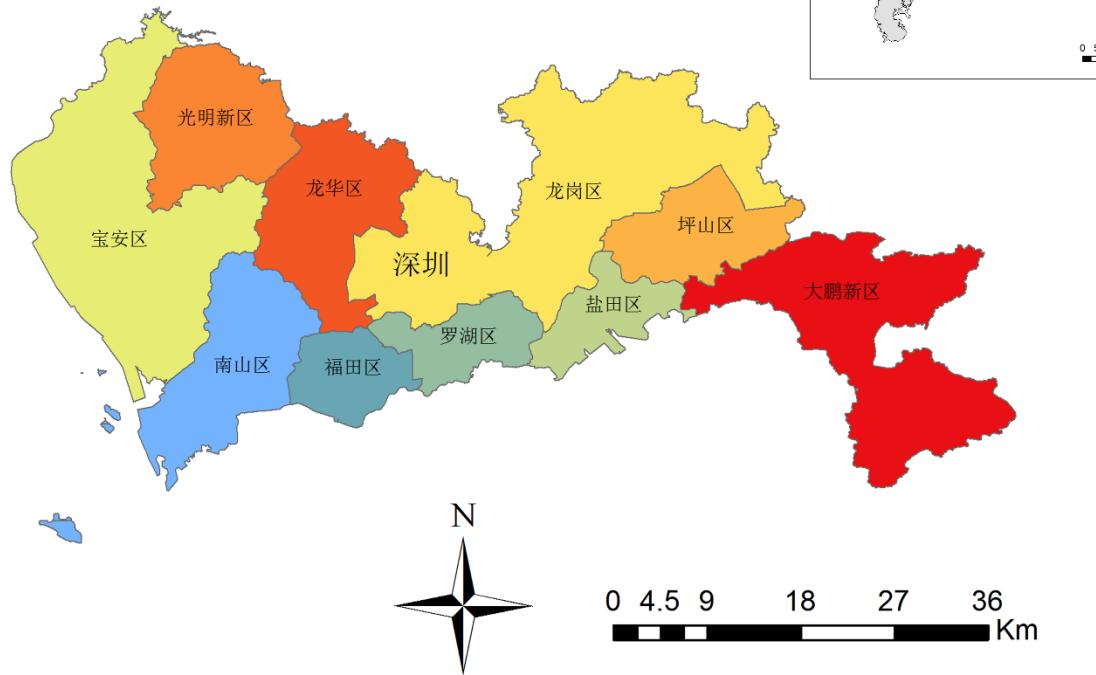


图1 研究区-深圳市

本文的研究区为广东省深圳市（如图1所示），深圳市是中国四大**一线城市**之一，也是中国**三大全国性金融中心**之一(徐薇 2015)。下辖**8个行政区和2个新区**，分别为福田区、罗湖区、南山区、盐田区、宝安区、龙岗区、坪山区、龙华区、光明新区、大鹏新区。

# 4.2 | 降雨天气对城市交通拥堵中心变化分析



## ■ 研究区和数据

表1 浮动车GPS轨迹数据

gpstime	plateno	lon	lat	speed	direction	iseffective	carrystate
426	B0HT07	114.1597	22.63167	67	50	1	1
427	BW0P59	114.1247	22.56213	0	180	1	1
427	BL4Z16	113.9185	22.54731	0	0	1	1
425	BV7N86	114.1583	22.56026	40	254	1	1
407	B4T3A0	114.0497	22.53139	74	45	1	1
426	B4WK56	113.8470	22.63117	81	308	1	1
426	B300VX	113.8824	22.71257	95	342	1	1
413	B4T3A0	114.0509	22.53142	73	45	1	1
426	B1D7X1	113.9451	22.53912	0	353	1	1
420	B4T3A0	114.0523	22.53144	73	45	1	1
421	B4T3A0	114.0525	22.53144	73	90	1	1
426	BV6N47	114.0444	22.53528	77	179	1	1
419	BU06V3	114.0685	22.54467	75	330	1	1
427	B0V2N5	114.1079	22.61139	0	0	1	1
425	BW0A42	114.0996	22.55061	59	359	1	1
413	BK47K9	114.1121	22.53796	76	81	1	1
427	BN3A22	114.1280	22.58650	26	135	1	1
359	B2BS70	114.0143	22.53969	40	45	1	1
427	BW2P95	114.0442	22.54549	76	135	1	1
426	BW7Q53	113.9522	22.57157	55	315	1	1
425	BS14U5	114.0691	22.53183	78	267	1	1

**浮动车GPS轨迹数据**是本实验最重要的数据之一，数据来源于深圳市交通委员会。每条数据包含GPS时间（gpstime）、车牌号（plateno）、经度（lon）、纬度（lat）、速度（speed）、方向（direction）、运营状态（iseffective）和载客状态（carrystate）等信息，具体的数据格式如表1所示。

# 4.2 | 降雨天气对城市交通拥堵中心变化分析



## ■ 研究区和数据

表2 降雨量数据

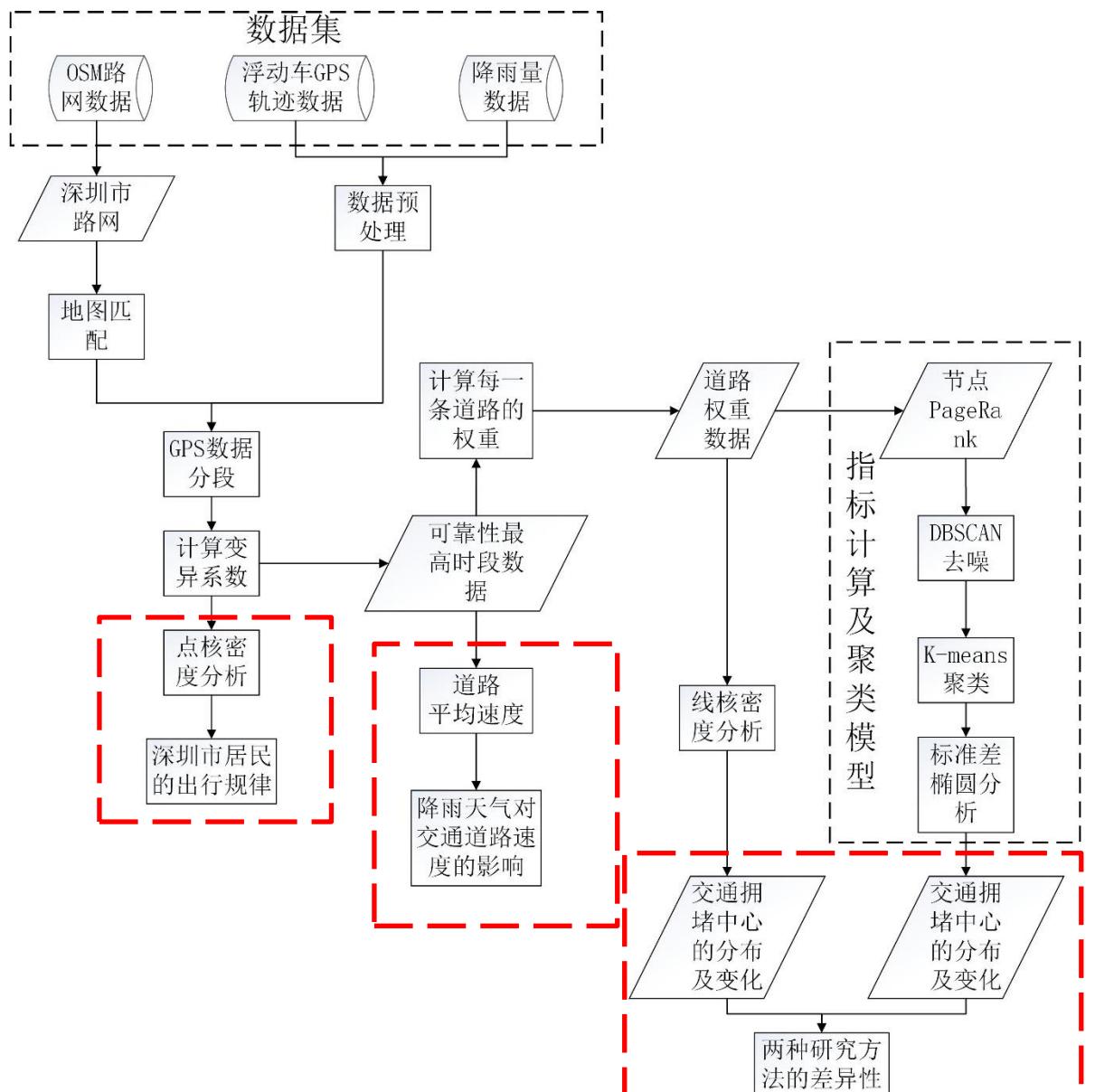
area	monitor	rh	r24h	lon	lat
宝安	西乡泵站	0	0	113.8662	22.82337
宝安	燕山泵站	0	0	113.9300	22.70743
龙华	福城	0	0	114.0536	22.70088
宝安	共和	0	0	114.0462	22.65676
龙岗	双龙	0	0	114.2841	22.73472
宝安	西乡	0	0	113.8691	22.58174
宝安	宝安公园	0	0	113.9071	22.59316
大鹏	葵新	0	0	114.4298	22.63442
福田	笔架山	0	0	114.0870	22.56848
龙岗	荷坳	0	0	114.2376	22.68513
宝安	新安	0	0	113.9061	22.57158
龙岗	六约	0	0	114.1958	22.64259
南山	桃源	0	0	113.9869	22.56417
宝安	燕川	0	0	113.8726	22.81564
坪山	竹坑	0	0	114.3909	22.71724
坪山	沙湖	0	0	114.3315	22.68414
龙岗	坂田	0	0	114.0676	22.63900
坪山	江岭	0	0	114.3605	22.69013
龙岗	坪西	0	0	114.2886	22.76997
福田	南园	0	0	114.1014	22.54445
宝安	和平	0	0	113.8030	22.69990

降雨量数据来源于深圳气象数据网 (<https://data.szmb.gov.cn/>)。每条数据包括地区 (area)、观测站 (monitor)、每小时降雨量 (rh)、24小时降雨量 (r24h)、经度 (lon) 和纬度 (lat) 等信息。数据格式如表2所示。

## 4.2 | 降雨天气对城市交通拥堵中心变化分析



- 深圳市居民出行规律探索
- 降雨对城市道路速度的影响
- 交通拥堵中心的变化分析



## ■ 研究方法

- 本文采取的可靠性分析的指标是**变异系数**（Coefficient of Variation, CV）(Pu et al. 2011)。变异系数是用来比较不同组数据的**离散程度大小的指标**，它是一个**无量纲量**，相比于标准差，变异系数更能体现数据的可靠性。变异系数具体的计算公式如下：

$$CV = \left( \frac{SD}{Mean} \right) * 100\%$$

其中： CV为变异系数； SD为数据的标准差； Mean为数据的平均值。

- 本文将浮动车GPS轨迹数据**按时段**进行分组后通过计算每个时段的**变异系数**来判断对应时间的**可靠性**，以便于从中选取可靠性较高的时段进行研究。

## ■ 研究方法

- 本文提出了基于PageRank的指标计算及聚类模型。PageRank是谷歌在网页排名上使用的算法，用于对网页的检索结果进行排序。它能够很好的避免特征向量中心性和Katz中心性两个指标存在的缺陷，不受到无出度导致的中心性为0和Katz完全传递性的影响(Wang et al. 2010)，具体的节点PageRank计算公式为：

$$C_p(v_i) = \alpha \sum_{j=1}^n A_{j,i} \frac{C_p(v_j)}{d_j^{out}} + \beta$$

其中： $C_p(v_i)$ 和 $C_p(v_j)$ 分别为 $v_i$ 和 $v_j$ 点的PageRank值， $n$ 是 $i$ 点的入度， $d_j^{out}$ 是 $j$ 点的出度， $A_{j,i}$ 表示的是整个网络的邻接矩阵， $\alpha$ 是用来调整中心性程度的一个参数， $\beta$ 是中心性的偏差项。

- 本文将路网节点的PageRank指标计算出来并分级，以此来判断路网节点的拥堵程度(李树彬等 2011)。

## 4.2 | 降雨天气对城市交通拥堵中心变化分析



### ■ 交深圳市居民出行规律

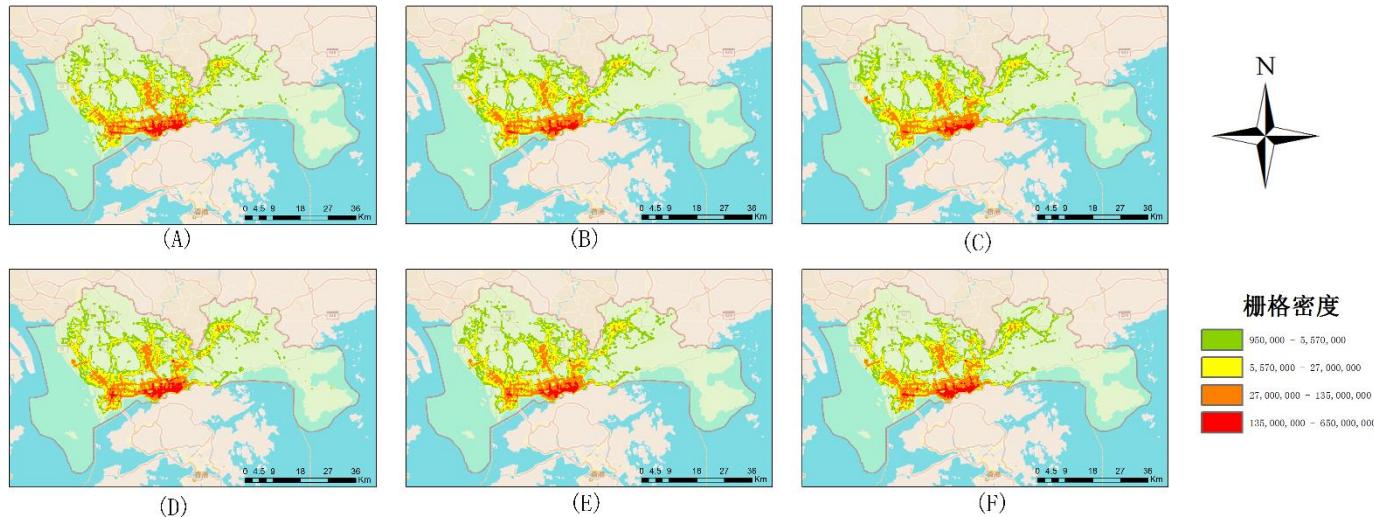


图4 08:00-20:00GPS点数据密度分布图: (A) 8:00-10:00, (B) 10:00-12:00, (C)  
12:00-14:00, (D) 14:00-16:00, (E) 16:00-18:00, (F) 18:00-20:00

- 三个出行人数**高峰时段**;
- 两个地区（罗湖区，福田区）出行人数明显高出其他区域；

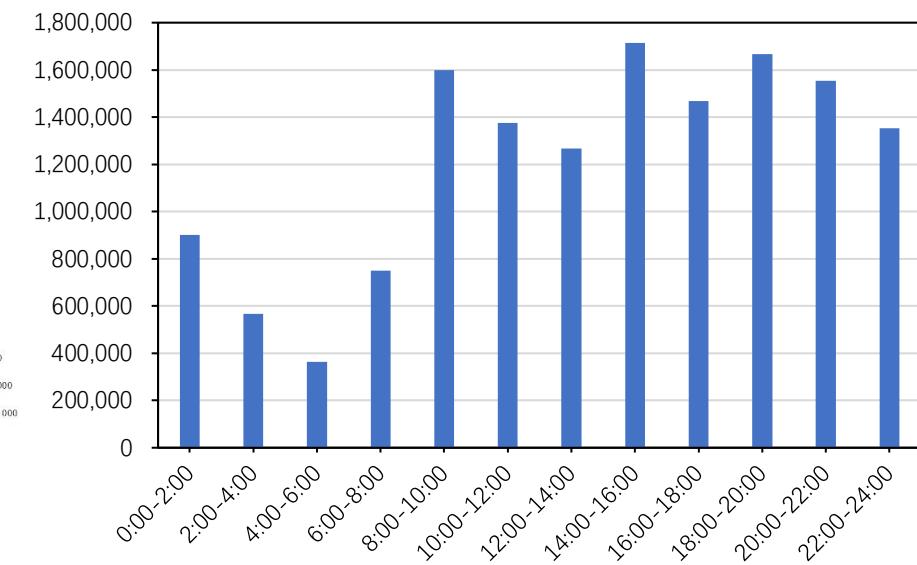


图5 1月7日GPS数据点统计柱状图

## 4.2 | 降雨天气对城市交通拥堵中心变化分析



### ■ 降雨天气对道路速度的影响

表3 不同时段的变异系数表

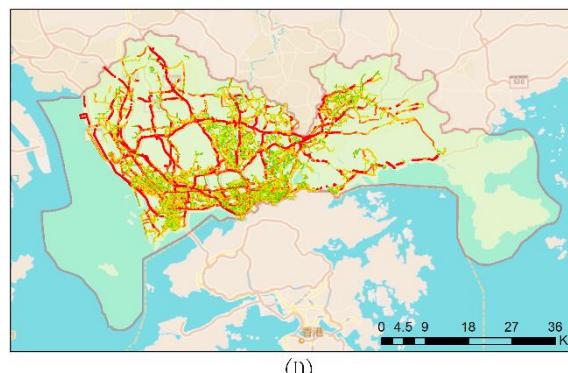
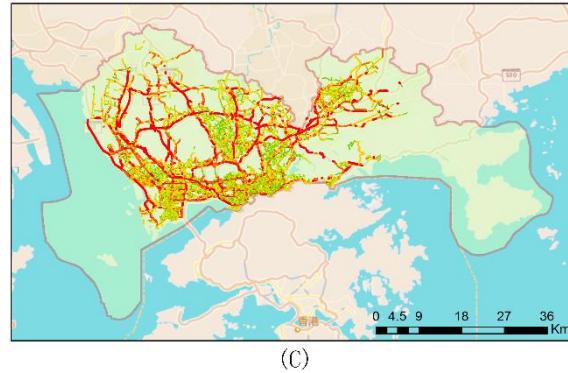
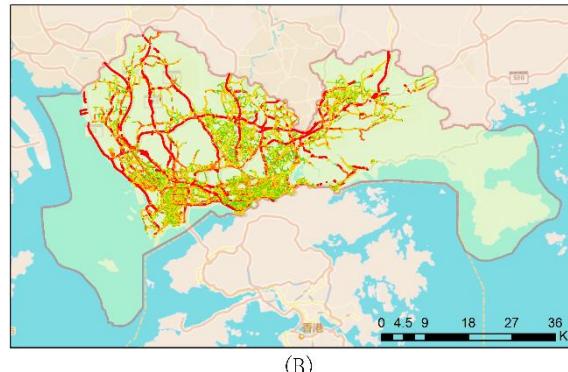
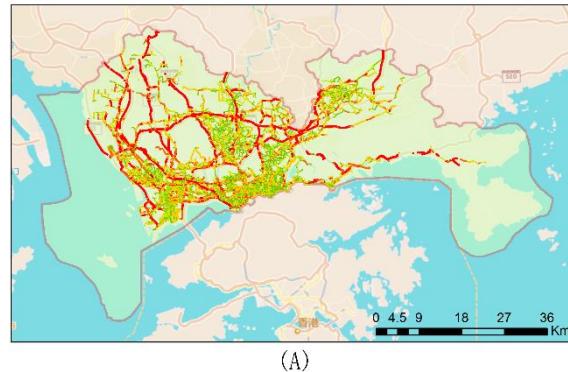
time	07_CV	09_CV	17_CV	28_CV
0:00-2:00	0.45796	0.44010	0.43997	0.43457
2:00-4:00	0.51680	0.46513	0.46690	0.49568
4:00-6:00	0.57861	0.52848	0.52015	0.52929
6:00-8:00	0.51626	0.47435	0.48000	0.48589
8:00-10:00	0.55026	0.47591	0.46311	0.50669
10:00-12:00	0.49855	0.50532	0.48528	0.50456
12:00-14:00	0.47704	0.50448	0.49853	0.49284
14:00-16:00	0.51274	0.53797	0.55621	0.53739
16:00-18:00	0.54089	0.53902	0.55632	0.56382
18:00-20:00	0.57104	0.55447	0.53318	0.61256
20:00-22:00	0.48838	0.51442	0.50597	0.51322
22:00-24:00	0.44702	0.46626	0.46333	0.46376

- 整体上**变异系数**的值在0.5左右波动；
- 0:00-2:00,6:00-8:00,12:00-14:00以及22:00-24:00这**四个时间段**的**变异系数**是**相对较低**；
- 结合实际生活的时间考虑，本文最终选取了**6:00-8:00**以及**12:00-14:00**两个时间段的**浮动车GPS轨迹数据**进行研究。

## 4.2 | 降雨天气对城市交通拥堵中心变化分析



### ■ 降雨天气对道路速度的影响



道路速度  
— 7 - 20  
— 20 - 34  
— 34 - 50  
— 50 - 68  
— 68 - 116

图6 路网速度图：(A) 1月7日（工作日晴天），(B) 1月28日（工作日雨天），  
(C) 1月9日（休息日晴天），(D) 1月17日（休息日雨天）

表4 道路数量数据表

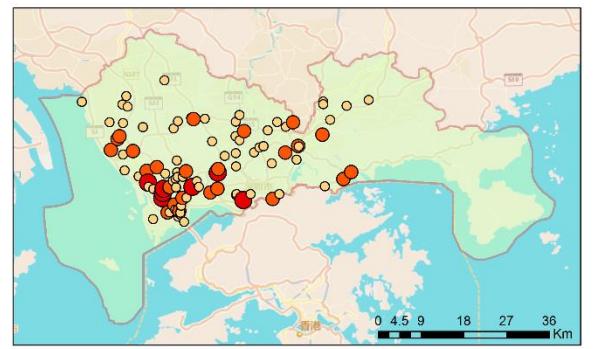
日期	速度	
	20-50Km/h	50-116Km/h
1月7日	12,971	6,424
1月9日	11,319	7,465
1月17日	12,041	7,283
1月28日	11,105	5,694

- 整体道路行驶速度下降
- 不同类道路下降百分比不同
- 工作日和休息日的影响有差异

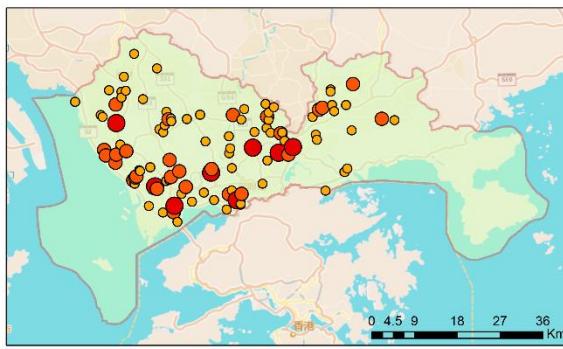
## 4.2 | 降雨天气对城市交通拥堵中心变化分析



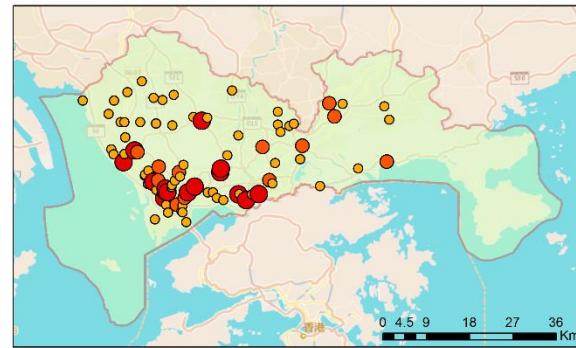
### ■ 降雨天气对道路交通拥堵中心的变化影响



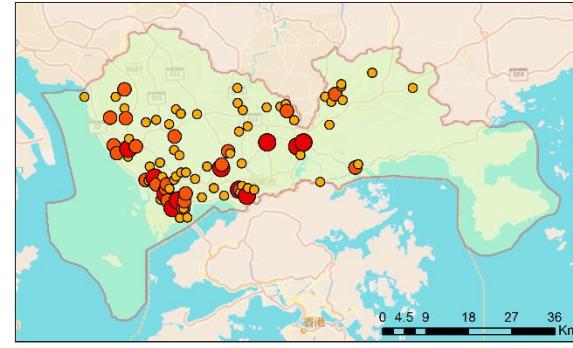
(A)



(B)



(C)



(D)



图7 PageRank点值分布图：(A) 1月7日（工作日晴天），(B) 1月28日（工作日雨天），(C) 1月9日（休息日晴天），(D) 1月17日（休息日雨天）

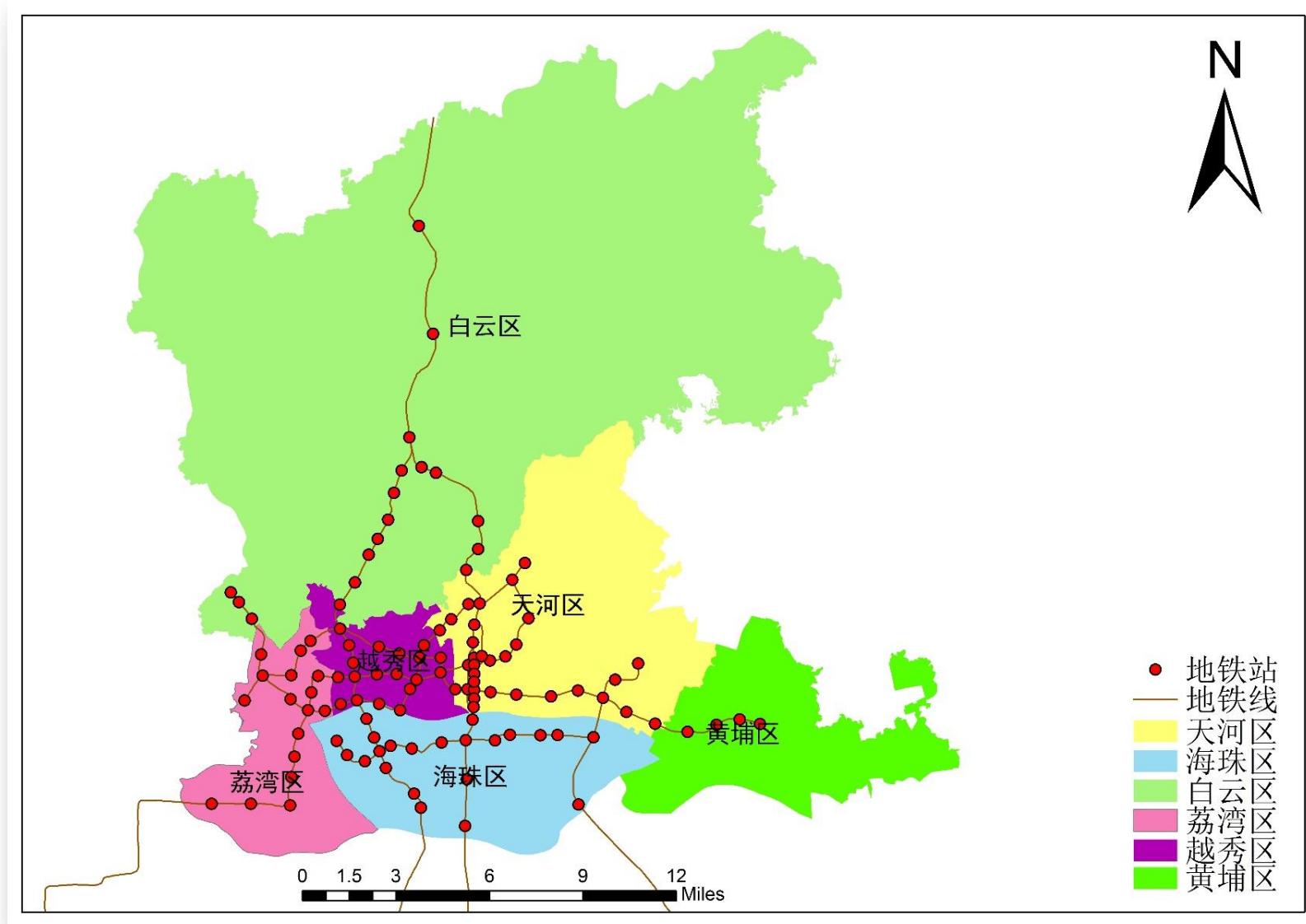
表5 拥堵点分类数量表

日期 \ 点类	一级拥堵点	二级拥堵点	三级拥堵点
1月7日	7	26	69
1月28日	8	28	77
1月9日	14	13	61
1月17日	11	21	61

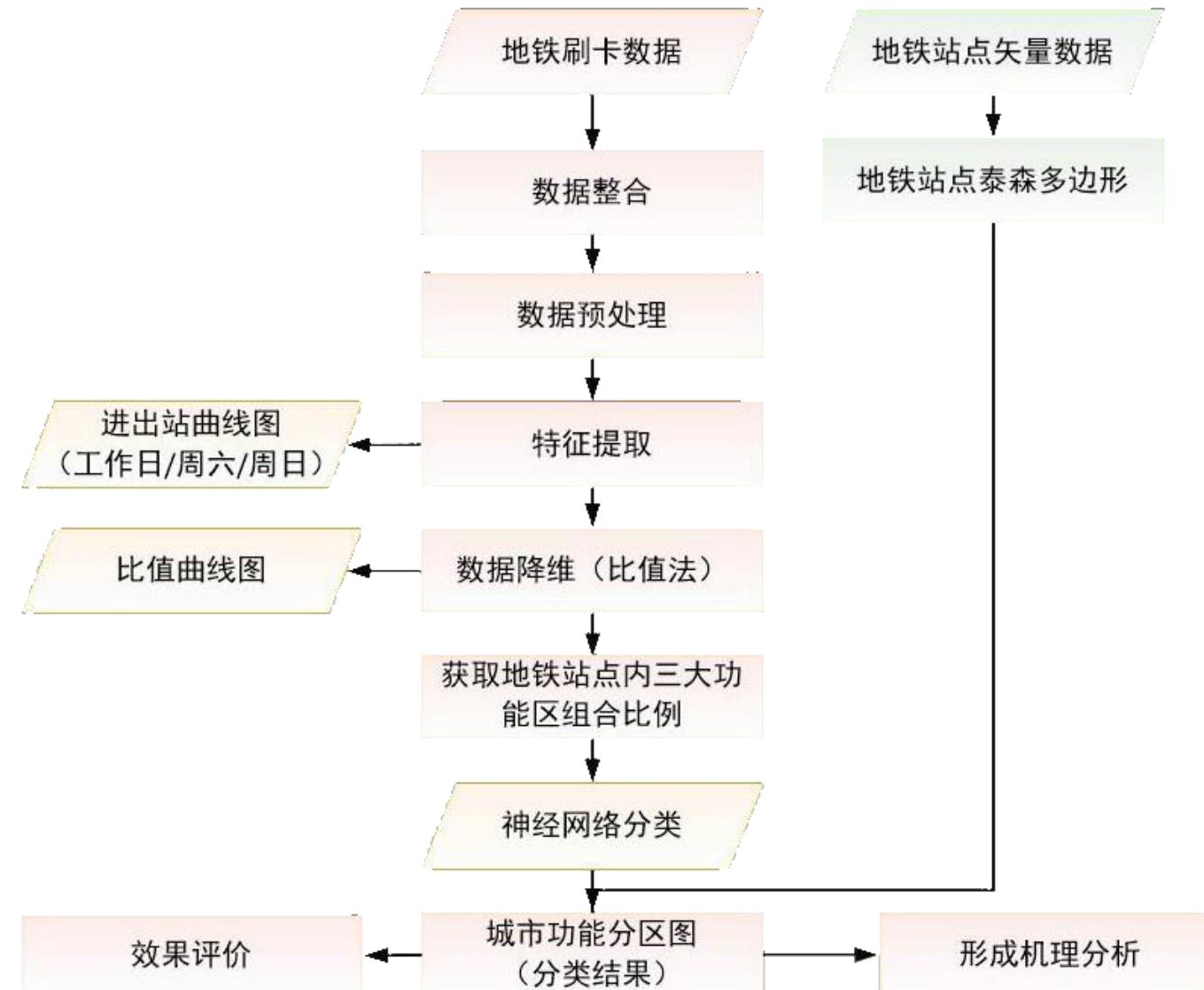
- 交通拥堵点增加
- 工作日和休息日拥堵点增加的类别有差异

## ■ 数据介绍

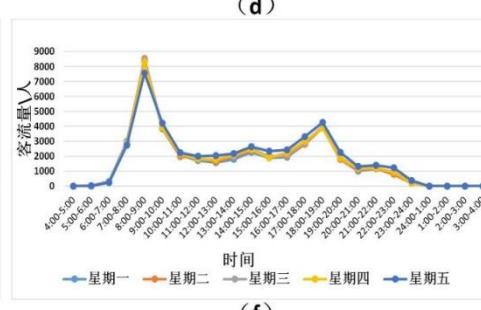
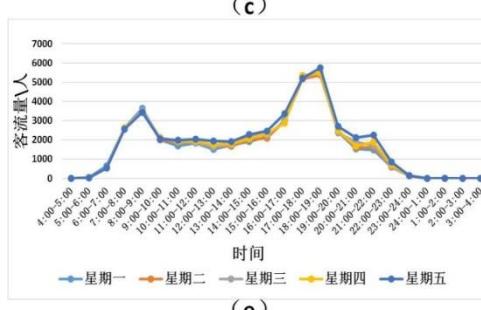
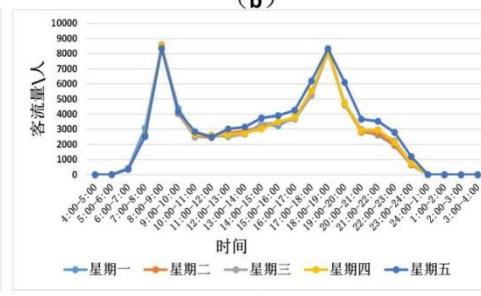
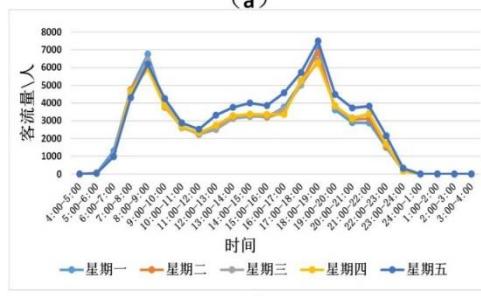
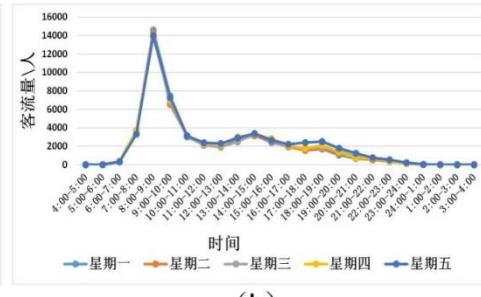
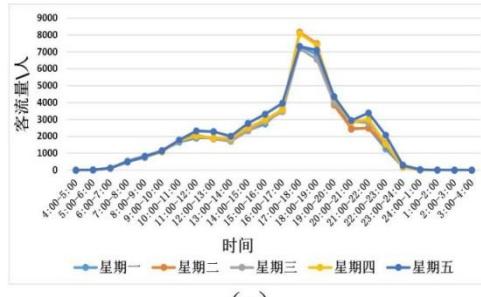
- 2014年3月3日至2014年3月9日广州市地铁刷卡数据
- 广州市地铁站点分布图
- 广州市市区行政区划图
- 广州市地铁线路分布图



## 研究路线



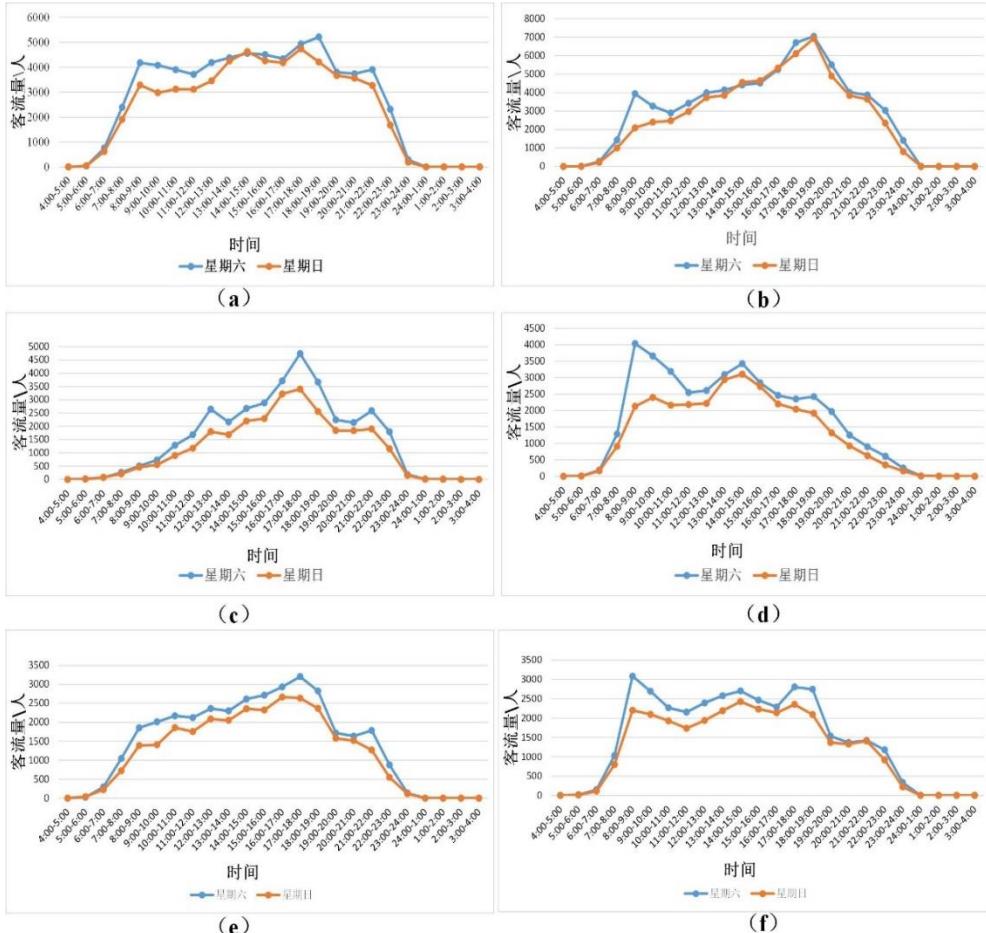
## ■ 数据预处理



周一至周五出行规律的高度相似性---工作日出行规律

$$Aver\_WD_i^k = \frac{1}{5} \sum_{j=1}^5 flow_{ji}^k$$

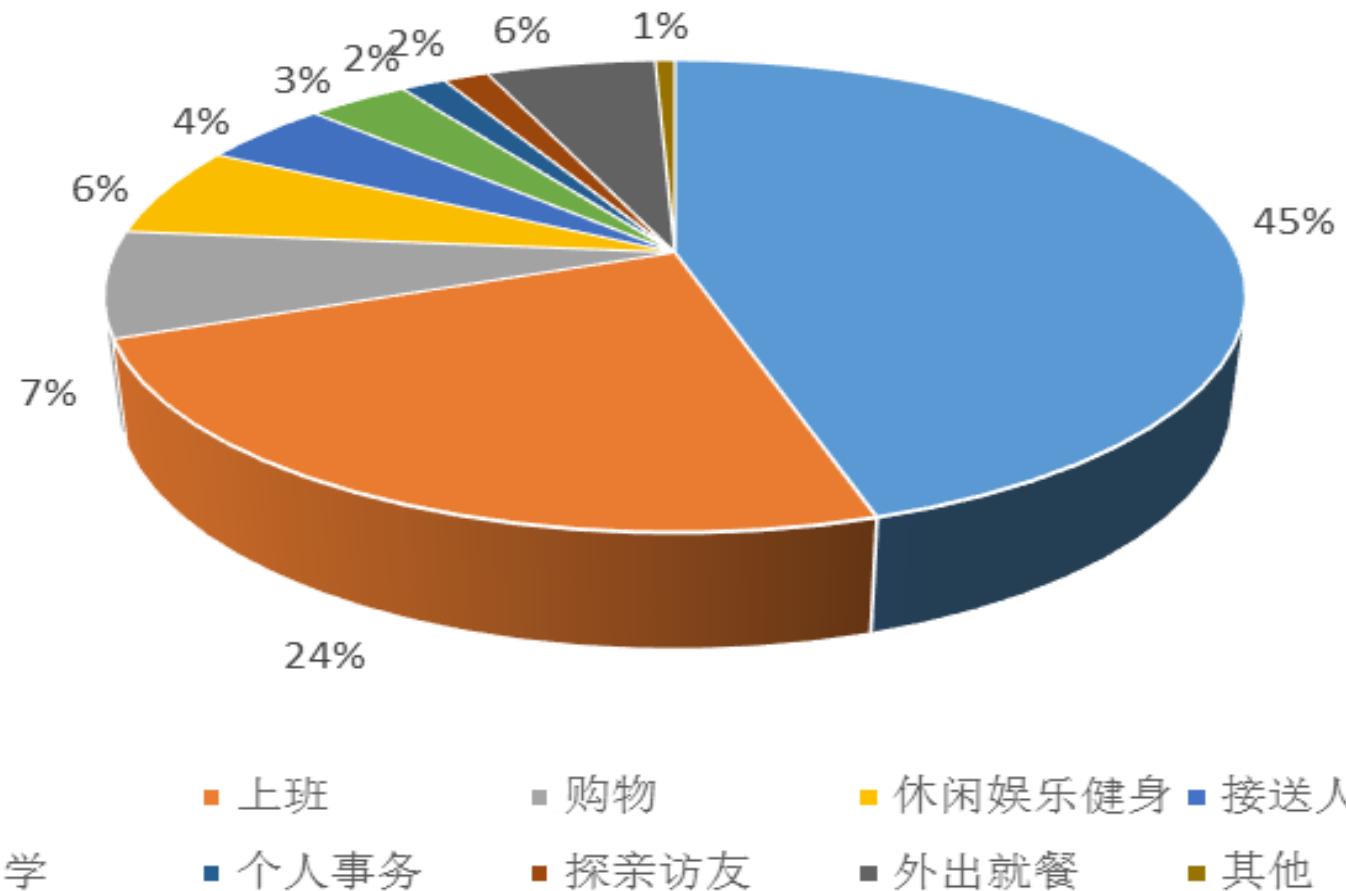
## ■ 数据预处理



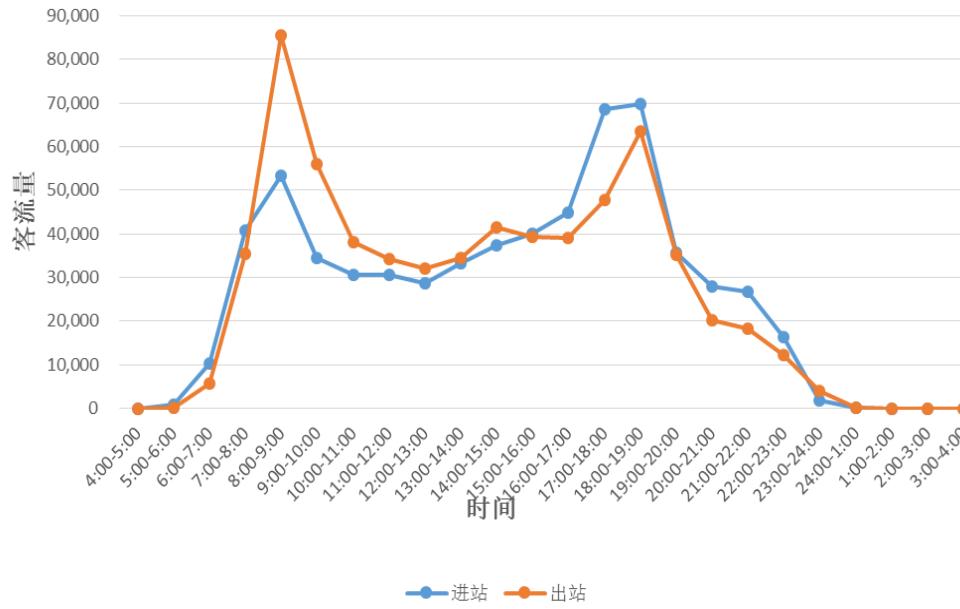
**周六日出行规律的高度相似性  
---休息日出行规律**

$$Aver\_RD_i^k = \frac{1}{2} \sum_{j=1}^2 flow_{ji}^k$$

## ■ 特征提取



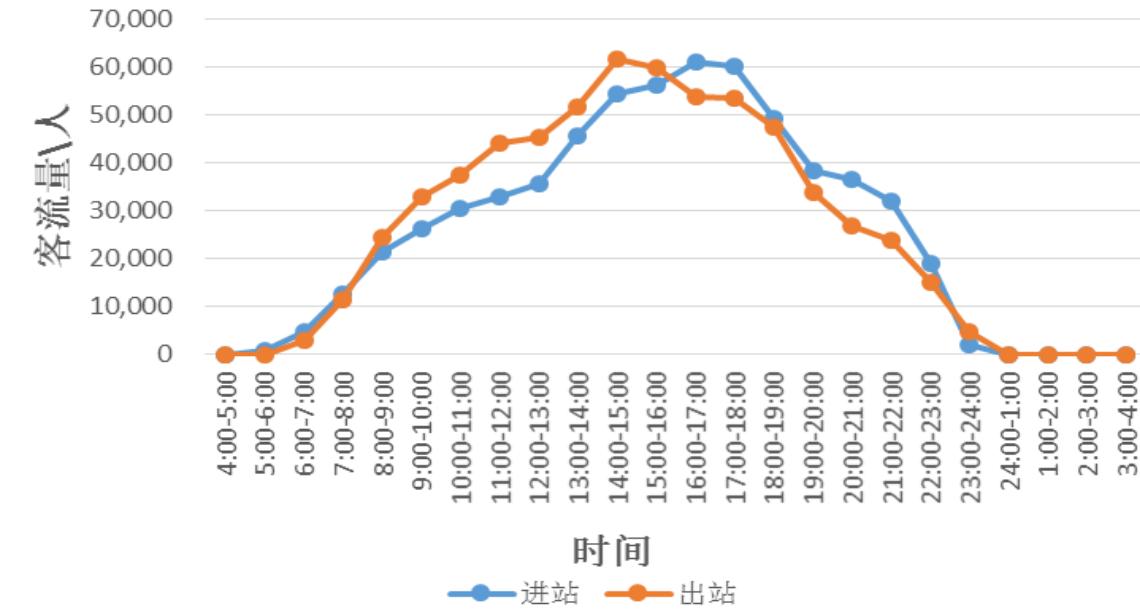
## ■ 特征提取



2014年3月3日（星期一）广州市地铁1号线  
进出站客流量时间曲线图

1

工作日早上7点至11点进出  
站客流量，下午14点至20点  
进出站客流量；



2014年3月9日（星期日）广州市地铁1号线  
进出站客流量时间曲线图

2

休息日早上10点至晚上19  
点进出站客流量；

## ■ 数据降维

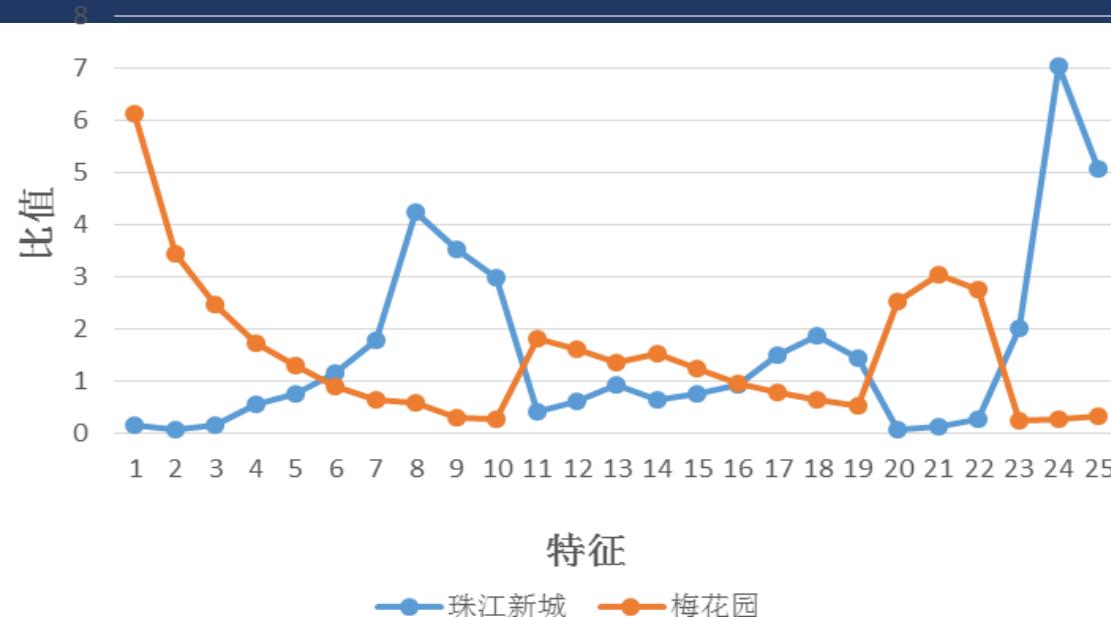
降维原理:  $Z_{ij}^k = \frac{X_{ij}^k}{Y_{ij}^k}$

$X_{ij}^k$ : 编号为k的地铁站第i天第j个小时的进站人数

$Y_{ij}^k$ : 编号为k的地铁站第i天，第j个小时的出站人数

$$IN_{ql}^k = \frac{M_q^k}{M_l^k}$$

$$OUT_{ql}^k = \frac{o_q^k}{o_l^k}$$



$IN_{ql}^k$ : 编号为k的地铁站早上q点进站客流量与下午l点进站客流量比值;

$M_q^k$ : 编号为k的地铁站早上q点进站客流量;

$M_l^k$ : 编号为k的地铁站下午l点进站客流量;

$OUT_{ql}^k$ : 编号为k的地铁站早上q点出站客流量与下午l点出站客流量比值;

$O_q^k$ : 编号为k的地铁站早上q点出站客流量;

$O_l^k$ : 编号为k的地铁站下午l点出站客流量;

k, l为有序数对, 且 $(q, l) \in \{(7, 17), (8, 18), (9, 19)\}$ 。<sup>66</sup>

## ■ BPNN原理

### 1. 正向传播

$$\mathbf{X} = (x_1, x_2, \dots, x_n)^T$$

$$o_p = f(\dots(f(f(X_p W_{(1)})W_{(2)})\dots)W_{(L)})$$

$$f = \frac{1}{1 + e^{-x}}$$

### 2. 反向传播

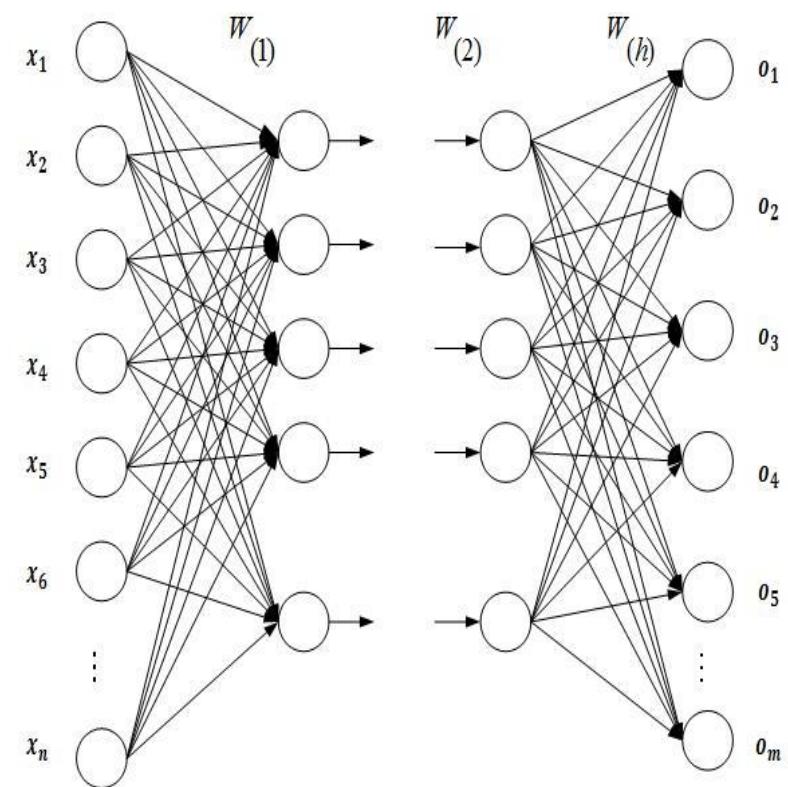
输出层权重修改:

$$w_{pq} = w_{pq} + \Delta w_{pq} \quad \Delta w_{pq} = -\alpha \delta_q o_p$$

隐含层权重修改:

$$w_{ij} = w_{ij} + \Delta w_{ij} \quad \Delta w_{ij} = -\alpha d_i^k x_j^{k-1}$$

$$d_i^k = x_i^k (1 - x_i^k) \sum_l w_{li} d_l^k$$

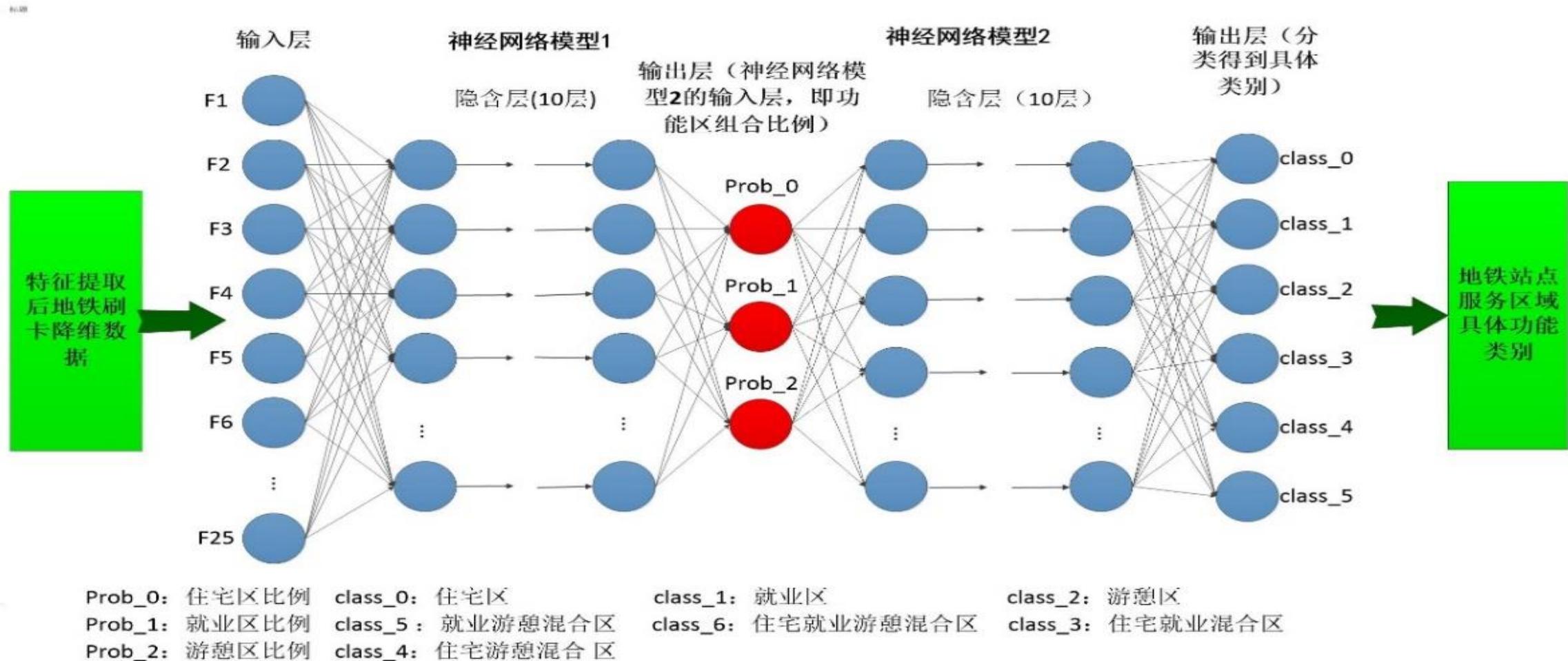


输入层

隐含层

输出层

## ■ 城市功能区识别



## ■ 城市功能区识别



住宅区分解比例



就业区分解比例



游憩区分解比例

## ■ 地铁站点服务区三大功能区分解比例



住宅区分解比例

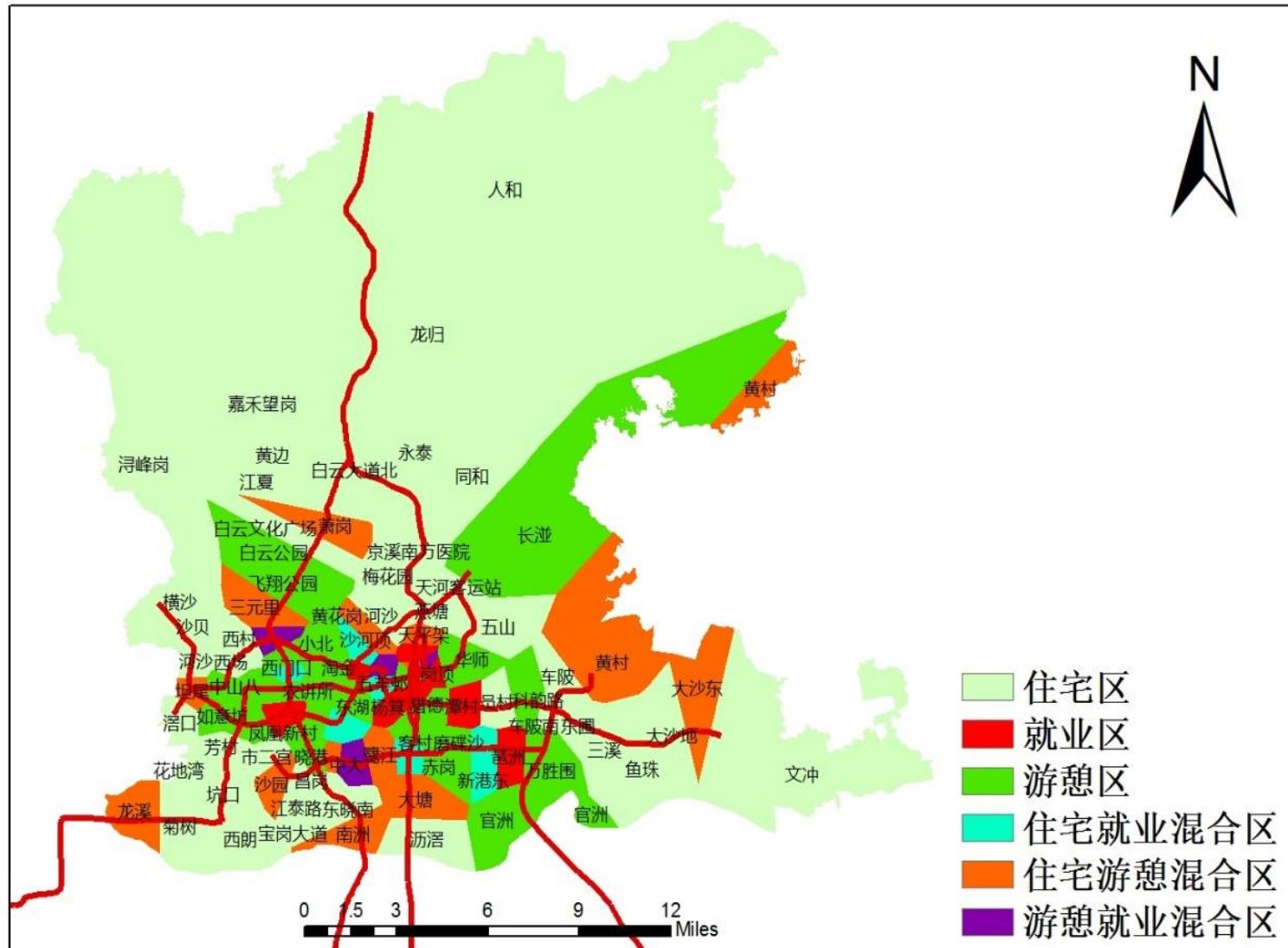


就业区分解比例



就业区分解比例

## ■ 城市功能区域识别结果及分析



- 就业区主要分布在城市的中心区域，在就业区的周边地区则成为了主要的居住空间和休憩空间，就业、游憩、居住的层状结构
- 复合型功能区主要分布在单一功能区的交界地带。

# 本章总结



本章介绍了轨迹大数据的数据来源、数据预处理过程以及常用的轨迹大数据分析方法以及其在人群移动模式、降雨和车速影响、小区开放政策研究、职-住-休功能区识别和分析等方面的应用。

轨迹大数据主要形式为浮动车数据、公共交通刷卡数据、手机信息服务数据等。

数据预处理轨迹清洗的过程中主要是对轨迹中的漂移点与停留点，还包括一些无效数据与重复数据进行清除。还介绍了漂移点与停留点的监测方法。

马尔可夫、条件随机场等方法是常用的轨迹预测和分析方法，对轨迹的精准预测有利于了解城市的空间结构，掌握城市运行的内在规律。

从轨迹大数据中挖掘移动对象的出行规律和出行轨迹模式，为城市道路规划、交通应急管理、城市广告投放等方面提供决策支持。



# Discussion!

---

姚尧 博士, 副教授, 高级工程师

地理与信息工程学院, 地图制图学与地理信息工程

阿里巴巴集团, 访问学者

Email: [yaoy@cug.edu.cn](mailto:yaoy@cug.edu.cn)

办公地点: 未来城校区地信楼522办公室

