



数据结构与算法 (Python) -00/引子

刘云淮 Yunhuai.liu@pku.edu.cn

<http://www.yunhuai.net/DSA2023/CoursePage/DSA2023.html>

北京大学计算机学院

目录

- › 自我介绍
- › 这是一门什么课?
- › 课程内容与目标
- › 往年我们做什么?
- › 我们的教材
- › 联系方式
- › 有用的软件和网站



北京大学大数据科学研究中心
CENTER FOR DATA SCIENCE OF PEKING UNIVERSITY

请输入关键字



首页

中心概况

师资队伍

人才培养

新闻动态

学术交流

招贤纳士



中心介绍

更多 >>

北京大学依托数学、统计学、计算机科学、软件、公共卫生与预防医学、心理学、社会学、金融学等学科于2015设置了数据科学与大数据技术交叉学科专业，并依托前沿交叉学科研究院成立了大数据科学研究中心。北京大数据科学研究中心将更好地整合校内数据科学的研究力量，依托北大优势学科资源，建设数据科学交叉学科，探索数据科学高层次人才培养模式。

2014年9月在校领导的领导下，依托前沿交叉学科研究院，联合多个院系，完成通过了自主设置大数据科学交叉学科的论证方案。北大成为国内首个具有特色交叉学科“数据科学”博士和硕士授予点单位，2015年元培学院、数学学院、信息与计算机科学学院开始设置数据科学本科专业，同年正式启动了数据科学交叉学科研究生的招生。学校专门成立了北京大学大数据科学委员会，高松副校长任委员会主任，鄂维南院士任委员会副主任，以及各院系相关负责人共计20位专家组成，为北京大学大数据学科的发展制定规划。

大数据科学研究中心，负责统筹数据科学交叉学科的人才培养项目，组织校内各单位的大数据相关研究，同时作为“北京大数据研究院”在北大内部的协调中心。在北京大学，与中关村管委会、海淀区政府、北京工业大学共同支持下，北京大数据研究院于2015年8月27日正式成立，目标是吸引国际一流大数据研究人员来京发展，建设国际化的大数据产学研协同创新平台。

北京大学大数据科学研究中心目前已成立深度学习实验室、自然语言处理与认知智能实验室、大数据安全实验室、数据管理实验室以及生物医学影像分析实验室。

2017年3月，由北京大学牵头成立大数据分析与应用技术国家工程实验室，旨在建设大数据分析技术研发与应用试验平台，形成国内一流的科研环境，培养大数据分析技术研发与应用高端人才，形成可持续的产学研协同创新机制，为推动我国大数据分析与应用的技术进步和产业发展提供技术支撑。

大数据科学研究中心一直极力致力于吸引国内外一流的大数据研究人员，助推创新人才培养。希望有志于在大数据领域深耕的各类优秀人才，加入北京大学大数据科学研究中心。大数据科学研究中心会尽力为你打造一个充分发挥你潜能的平台，我们也会尽力给你提供一个合理的待遇。你会发现，这是实现你人生价值最好的平台。



自我介绍

> 刘云淮

研究员，北京大学大数据科学研究中心

- 清华大学计算机 学士
- 香港科技大学计算机 博士
- 中国科学院深圳先进技术研究院 副研究员
- 公安部第三研究所物联网中心 副主任、副研究员、二级警督
- 2016年加入北京大学
- 2019年获得国家杰出青年称号

> 研究兴趣

智慧城市

群智感知网络

工业物联网



这是一门什么课？

- › 首先，这不是一门关于编程的课
- › 本课为你展示：

如何把数据组织起来
进行有效的处理
以解决问题

每一天，有.....

2940亿封电子邮件发送

平均每个地球人每天发送42封

6288个新移动应用可被下载

日均下载量已达3500万

57万6000小时视频

上传到YouTube

3亿5000万张照片

上传到Facebook

如果把它们都印出来，叠起来能有
80个埃菲尔铁塔那么高

500TB数据上传到Facebook

如果用2TB的硬盘储存
每年Facebook要新购65吨硬盘

20亿小时电视与电影

在Netflix上观看

整个因特网的流量信息可以装满

24万亿张DVD光盘

需要6万艘10万吨的油轮运送

2亿3000万条tweets

在Twitter上发布

2020年

全年互联网总数据为1.93ZB

即1,930,000PB

每年数据量增长60%

其中非结构化数据增长80%

大数据在中国 – 新一轮“信息革命”

3.5ZB in 2011

1天的数据量
> 文明起始到2003年



2090亿

2021年RFID标签销售量
2011年是1200万



\$8000亿

10年个人位置信息服务创造的价值



10.77亿

移动互联网用户
中国 2016年12月



200PB/季度

智慧城市数据
中国某一线城市



5PB/年

健康档案数据
中国某一线城市



“数据日益成为商业的新源材料：一种与资本和劳动力并列的新经济元素。”

— *The Economist*, 2010

“信息将成为21世纪的石油” .

— *Gartner*, 2010

衡量数据量的方法

- › 计算机基础存储单位：**字节**
- › $1 \text{ KB} = 2^{10} = 1,024$ 字节
- › $1 \text{ MB} = 2^{20} = 1,048,576$ 字节
- › $1 \text{ GB} = 2^{30} = 1,073,741,824$ 字节
- › $1 \text{ TB} = 2^{40} = 1,099,511,627,776$ 字节
- › $1 \text{ PB} = 2^{50} = 1,125,899,906,842,624$ 字节
- › $1 \text{ EB} = 2^{60} = 1,152,921,504,606,846,976$ 字节
- › $1 \text{ ZB} = 2^{70} = 1,180,591,620,717,411,303,424$ 字节



640 KB 足够所有人用了！
——比尔·盖茨·1981

Gartner技术发展趋势图(2014)



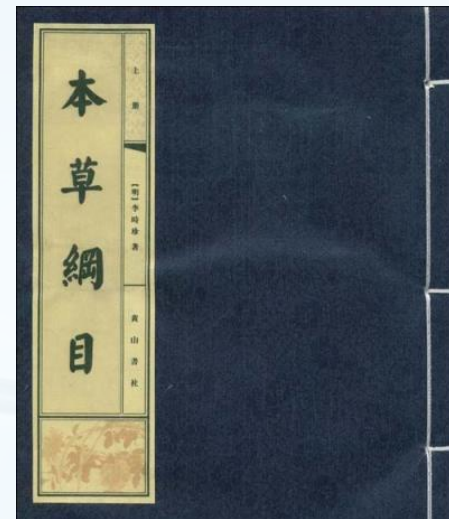
数据的前世今生

中国传统药学著作



神农本草经

上古，先秦，秦汉时期
多位医家集结整理
上中下三卷，载药 365 种

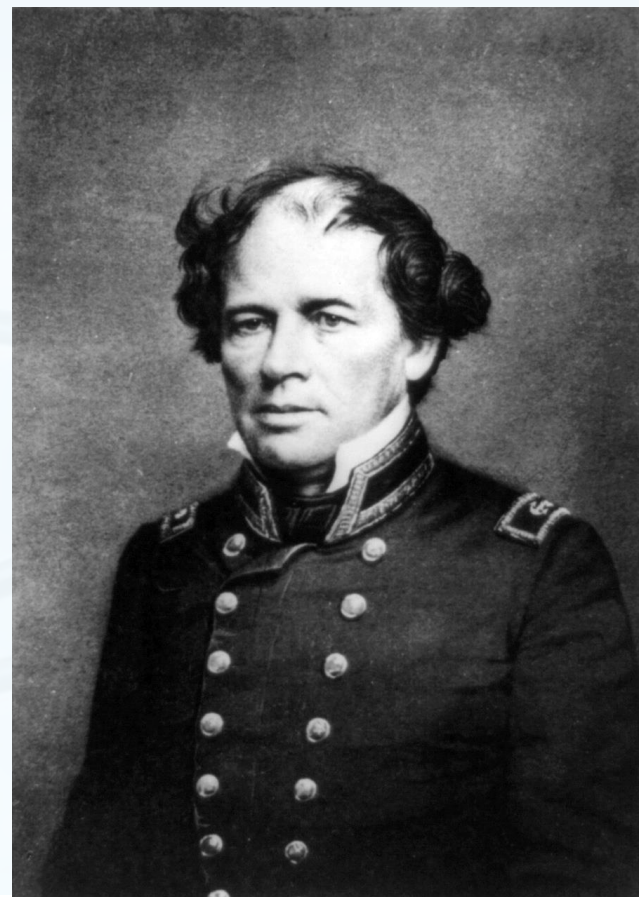


本草纲目

明朝李时珍
历时 27 年编纂，1590 年出版
共 52 卷，载药 1892 种
方剂 11096 个

美国海军上尉和他的大数据实践

- › 马修·方丹·莫里
- › Matthew Fontaine Maury
- › 1806 年出生于美国弗吉尼亚
- › 1824 年刚刚达到入伍年龄便进入了美国海军学校
- › 1839 年，已经晋升为海军上尉的莫里在一次事故中不幸腿部致残
- › 不适合于服役远航的莫里在 1842 年被任命为主管海图和仪器库的负责人



大航海时代的海图



六分仪



经典的书籍、教材



指南针



1459年毛罗地图



郑和航海图



哥伦布航海图

莫里的目标



大量快发霉的航海日志

前人视为垃圾

变废
为宝

莫里的目标



精确的图表

与商船交换信息，在自愿基础上以互利互惠的合作方式开创了国际气象界公开交换环境资料的传统

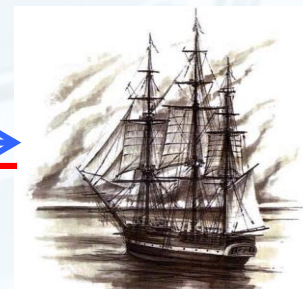


莫里

更新后的图表

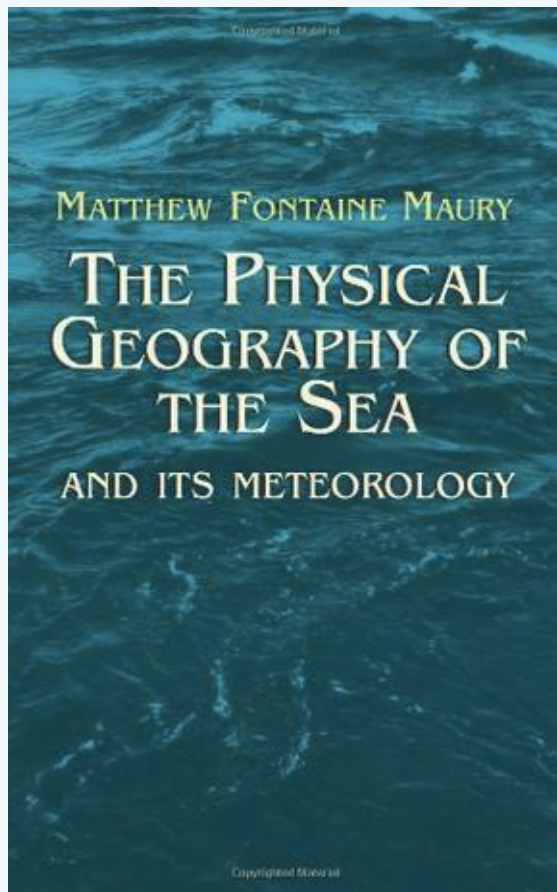


航海日志

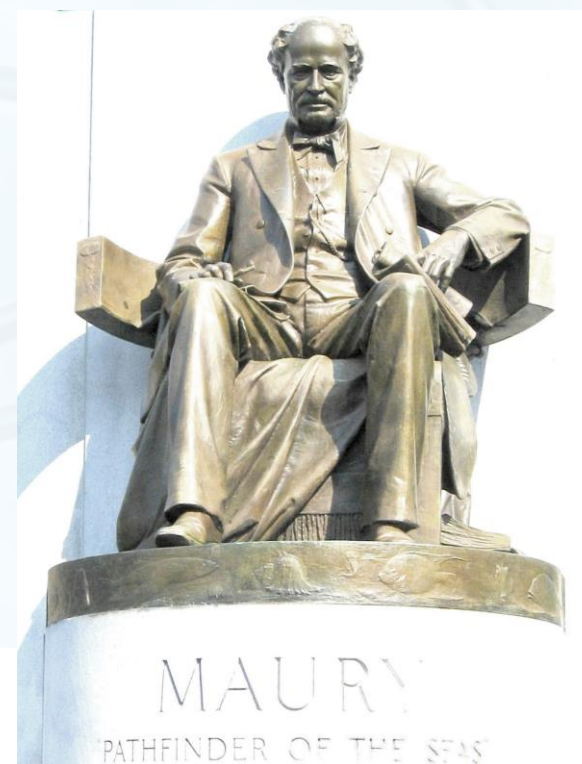


商船

名垂千古：海洋学的奠基人



- 1855 年,莫里出版权威著作《海洋物理地理学和气象学》, 被誉为**海洋学的奠基人**
- 当时, 他已经绘制了 120 万个数据点
- 四个国家授予了他爵士爵位, 包括梵蒂冈在内的其他八个国家还颁给了他金牌奖章
- 即使到今天,美国海军颁布的导航图上仍然有他的名字



信息技术令人类获取数据的能力大幅提高

数据无处不在



交通数据



金融数据



物联网数据



零售数据



社交网络数据



科学数据

无处不在的数据

政务数据

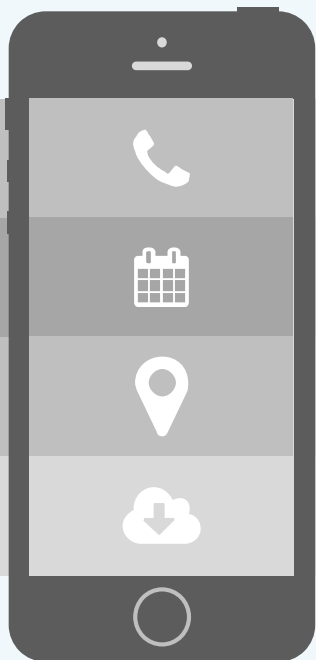
数据结构「算法」(Python)



商业数据



互联网数据



新闻信息

- ✓ 百度新闻
- 今日头条
- 搜狐新闻
- 网易新闻

...

社交媒体信息

- ✓ 新浪微博
- 网易云音乐
- QQ空间

...

专利信息

- ✓ 佰腾专利
- 中国专利局官网
- 中国专利信息网

...

房地产信息

- ✓ 链家
- 58同城网
- 贝壳二手房网
- 我爱我家租房网

...

公共资源招投标信息

- ✓ 各城市政府招投标监管网官网
- 各城市政府采购官网
- 各城市规划和国土资源局官网
- 各城市联合产业交易所官网

...

招聘信息

- ✓ 51job招聘
- 智联招聘
- 前程无忧

...

空气质量信息

- ✓ Aqistudy
- 各省环境保护厅官网

...

餐饮店铺点评信息

- ✓ 大众点评
- 美团网
- 猫途鹰网
- 百度糯米网

...

如果汽车工业发展也像信息技术一样快...

	容量 (内存)	速度 (CPU)	价格
IBM7030 (Stretch)1961	128KB	1.2 MIPS	RMB: 1亿元
Core i9-900K 2000	16G (typical)	81666 MIPS	RMB: 5000块

	容量 (乘客)	速度 (公里/小时)	价格
一辆1960年的车	5	100	20万
一辆今天的车, 应该能达到的性能 (如果按照信息技术发展)	64万人	80万公里/小时	1块

每一个智能手机...



典型和非典型物联网设备



智能设备：每时每刻收集数据



智能手机

地理位置数据
运动数据
环境亮度数据
图像数据
语音数据
...



空气质量数据
温湿度数据
气压数据
...



智能家居设备



身体状况数据
运动习惯数据
实时图像数据
...

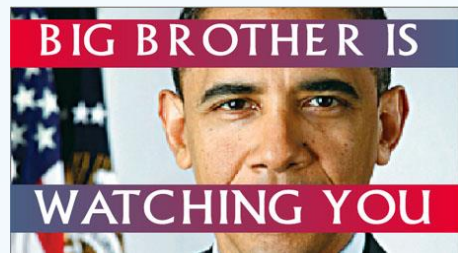
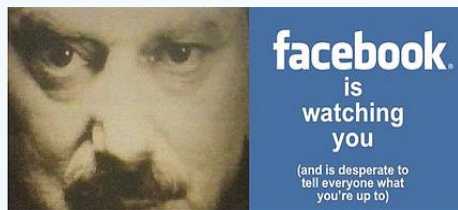
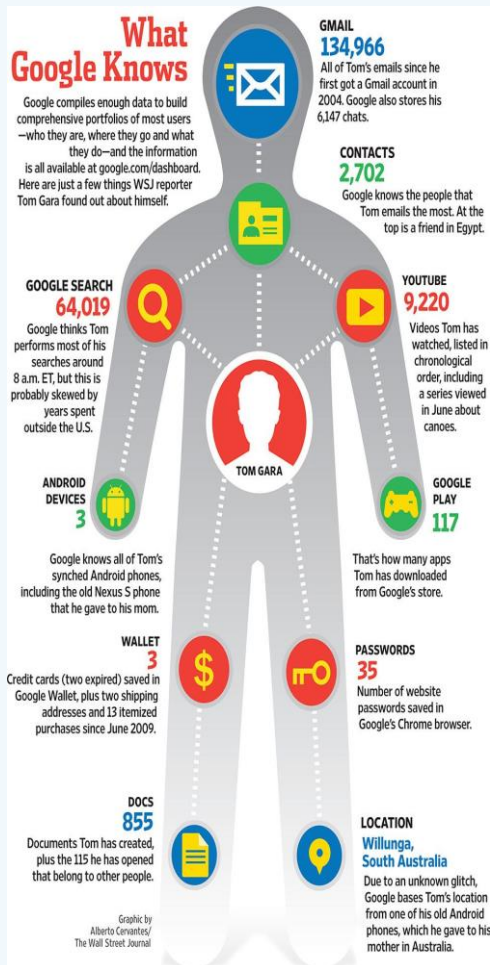


可穿戴设备

信息时代的数字脚印



互联网企业比你更了解你



发现·好货 **淘宝网 Taobao.com**

“宝贝不错，就是不知道买来不”
124 收藏

无糖糕点代餐食品
“宝贝包装十分精美，里面有独”
548 收藏

苹果便携式钥匙扣
“宝贝收到了，虽然说不是我真”
172 收藏

从你的数据中了解你!

从NBA总决赛说起



VS



NBA总决赛，美国观众人数超过2700万

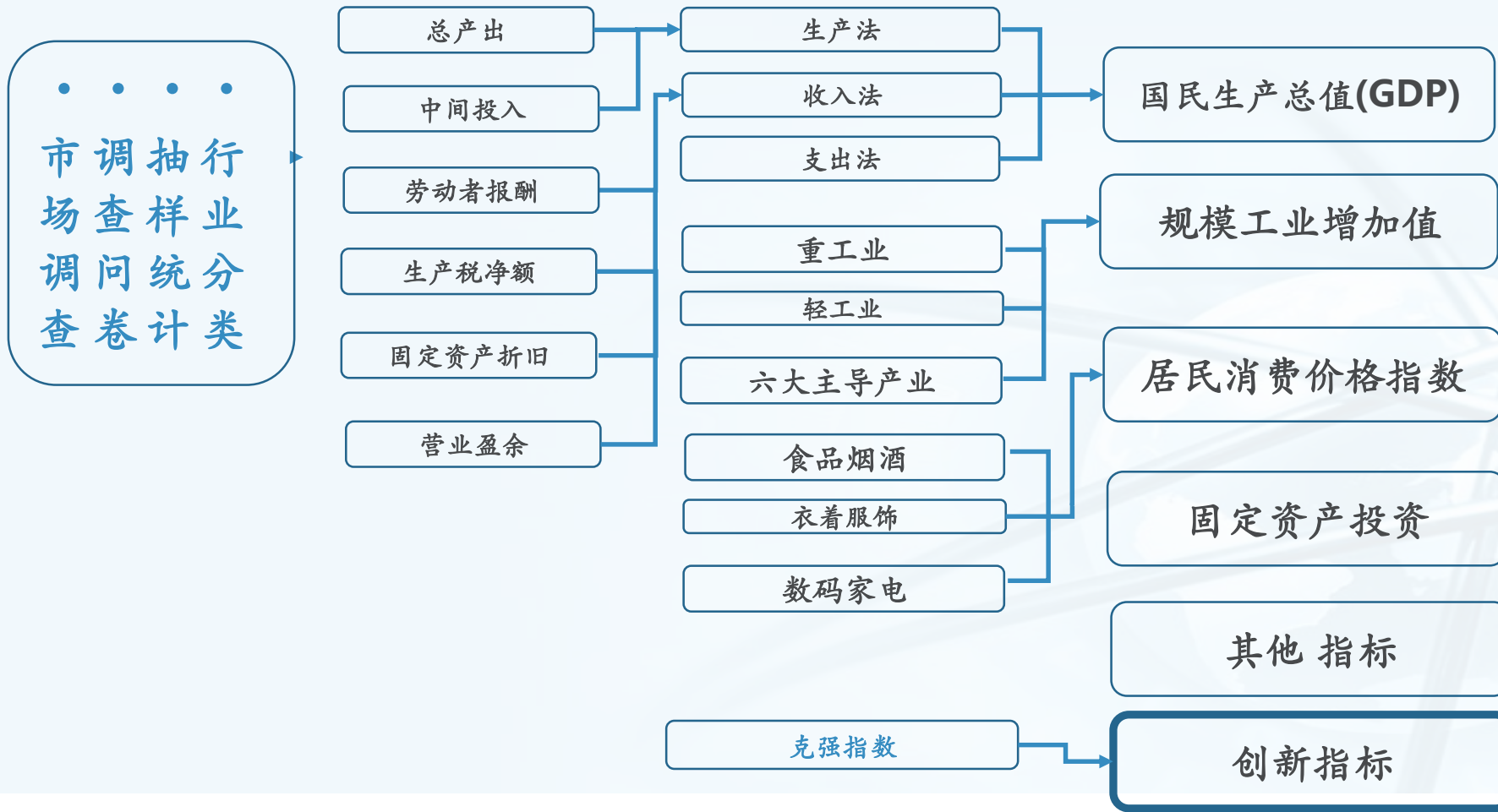
宏观经济大数据典型案例

大数据宏观经济监测预警意义

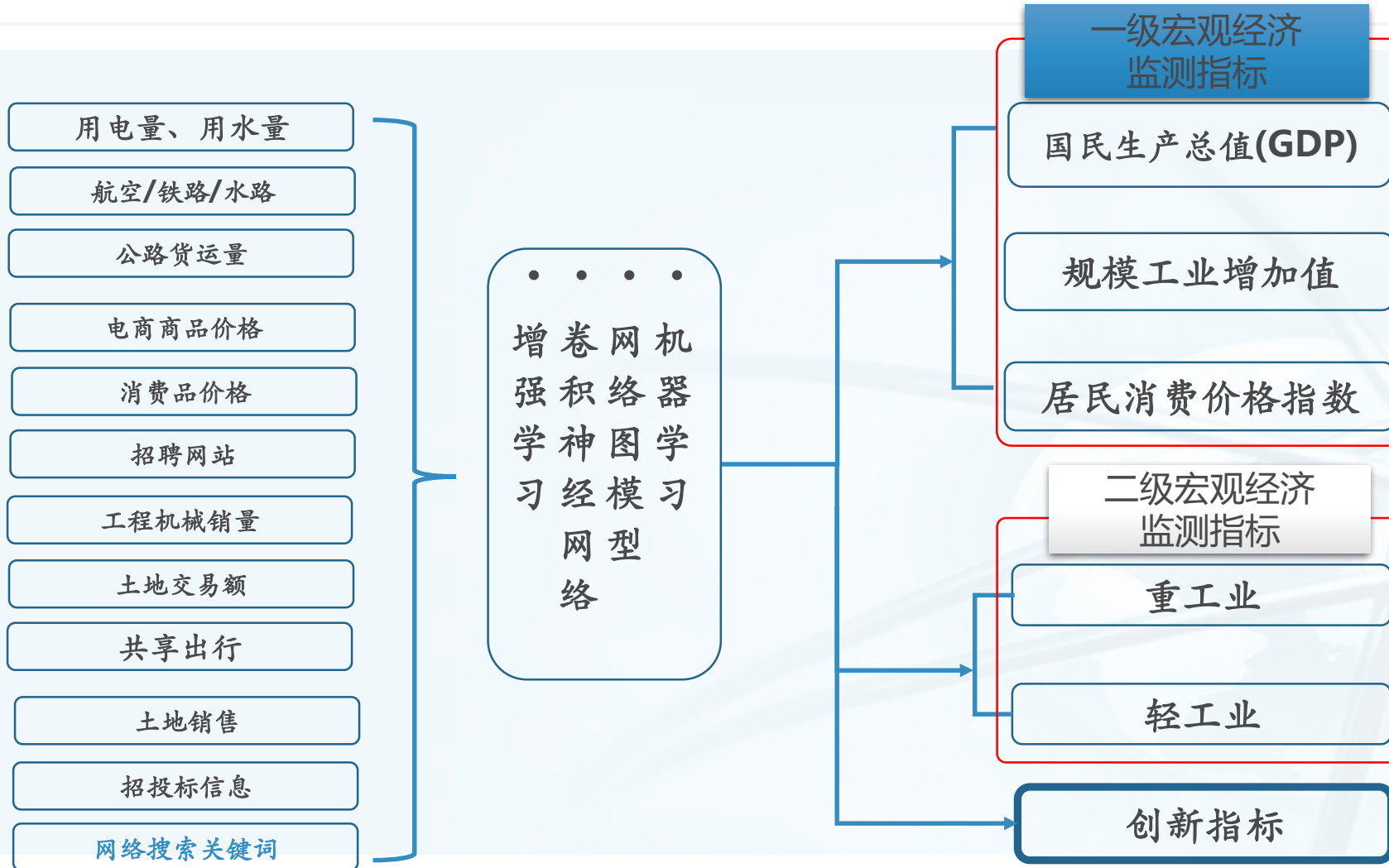
- › **提高经济运行及时性、全面性和准确性**
《十三五规划纲要》
- › **提高决策的针对性、科学性和时效性**
《促进大数据发展行动纲要》
- › **建立基于大数据的宏观经济监测体系**
《促进大数据发展2017年工作要点》
- › **百家大数据机构描绘中国经济**



传统宏观经济统计监测



大数据宏观经济监测预警



失业率计算 (现有方式)

登记失业率

城镇登记失业人员情况

表号：人社统202号
制表机关：人力资源和社会保障部
批准机关：国家统计局
技术文号：国统制[2018]302号
有效期至：2019年10月
单位：人

填报单位：无锡市		2018年 一季度																	
项目	序号	登记失业人员			就业人员			登记失业人员			就业人员			登记失业率					
		失业人数	失业人数	失业人数	失业人数	失业人数	失业人数	失业人数	失业人数	失业人数	失业人数	失业人数							
甲	乙	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
合计	1																		

补充资料：符合规定条件的农村转移就业人员和城镇非本地户籍人员。

单位负责人签章： 统计负责人签章： 填表人签章： 报告日期： 20 年 月 日

说明：
1.《城镇登记失业人员情况》表中登记失业率为2位小数。
2.指标关系：列关系：(3) < (2)、(4) < (2)、(5) < (2)、(6) < (2)、(7) < (2)、(8) < (2)、(9) < (2)、(10) < (2)、(11) < (2)、(12) < (2)、(13) < (2)、(14) < (2)、(15) < (2)、(16) < (2)、(17) < (2)、(18) < (2)、(19) < (2)、(20) < (2)。



手续繁琐，主动登记的人较少

调查失业率



劳动力调查表

20 年 月

应在本户登记的人：
调查时点居住在本户的人；
本户人口中，居住不满半年的人。

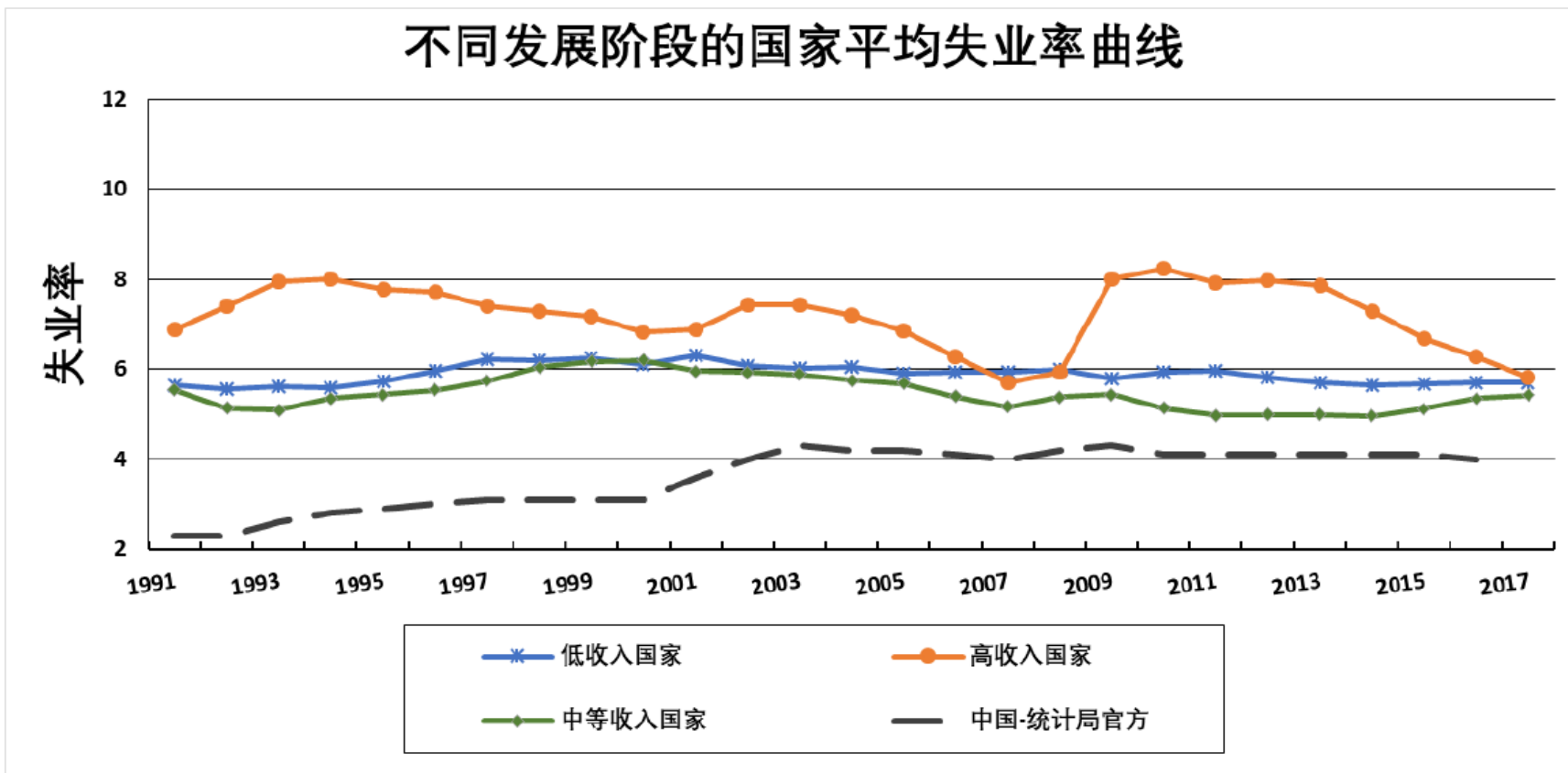
本户地址：_____ 区（市、县）_____ 乡（镇、街道）_____ 社区居委会（村委会）_____ 住户组_____

组别	1. 家庭户	2. 集体户	3. 调查时点居住在本户的人口数	4. 本户人口中，已外出但不满半年的人口数	5. 居住来源
合计	共_____人	共_____人	共_____人	共_____人	1. 自有
	其中：_____人	其中：_____人			2. 租赁住房
	其中：_____人	其中：_____人			3. 其他来源
	其中：_____人	其中：_____人			4. 单位宿舍
	其中：_____人	其中：_____人			5. 单位家属院
					6. 其他

一个拥有400万人的地市级市，每月调查样本仅800户，共2100多人

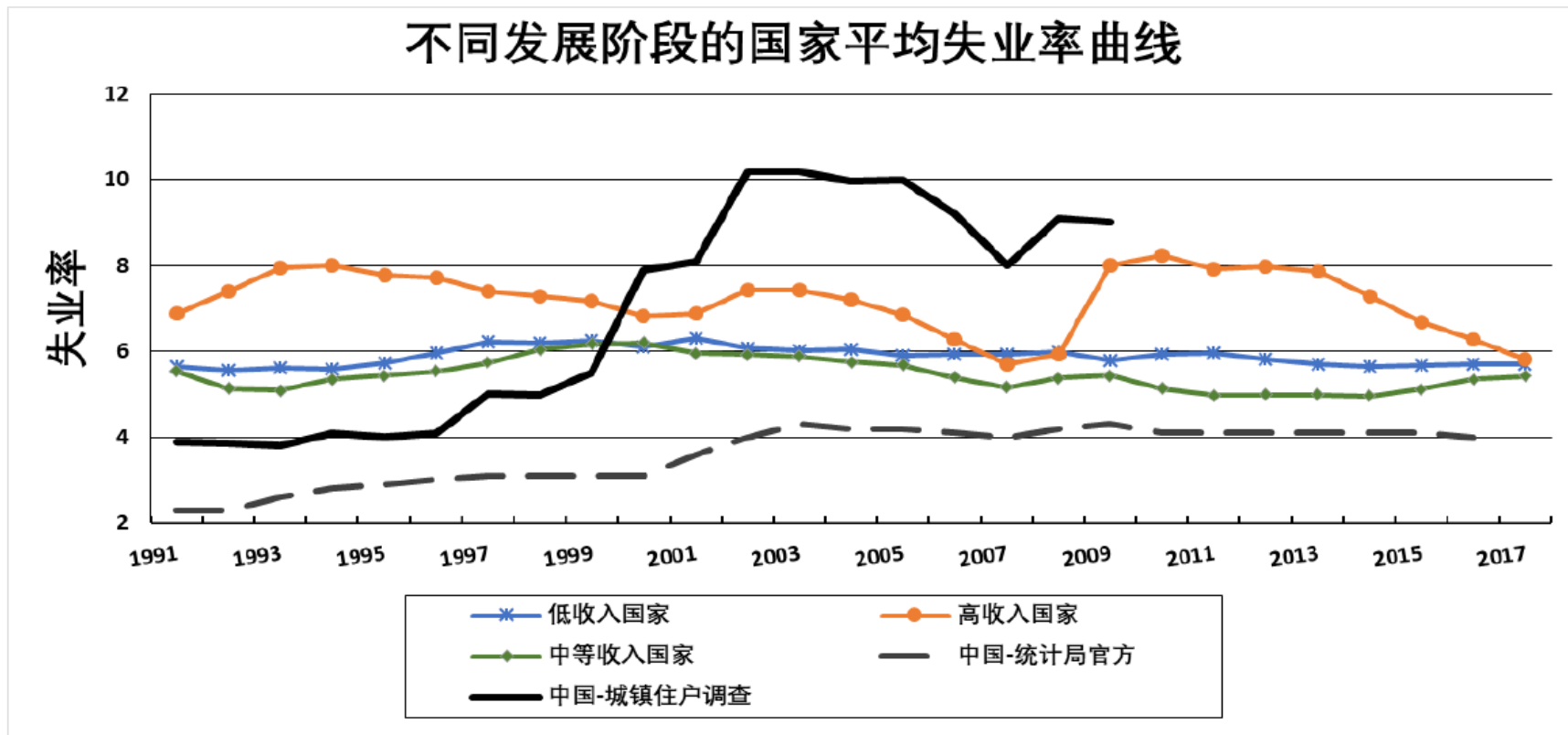
失业统计

右图为不同发展阶段国家的平均失业率曲线
最下方虚线为中国官方数据



失业统计

黑色实线为使用中国城镇住户调查数据计算的失业率



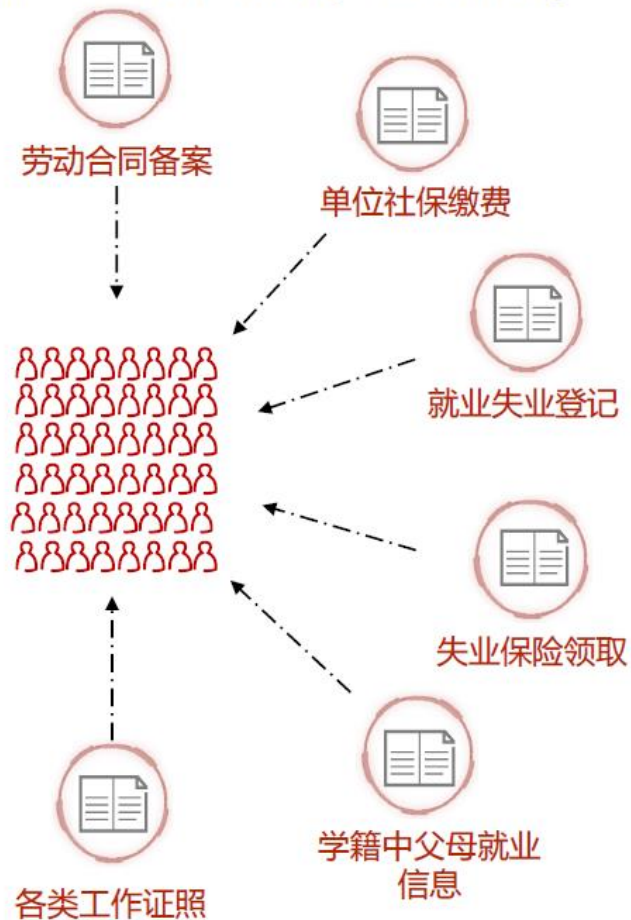
数据来源：世界银行、UHS调查、Feng et al. (2017) 与北京大数据研究院

大数据失业统计

人口画像 (400万人)



部分样本的就业记录数据 (50万人)



模型结果

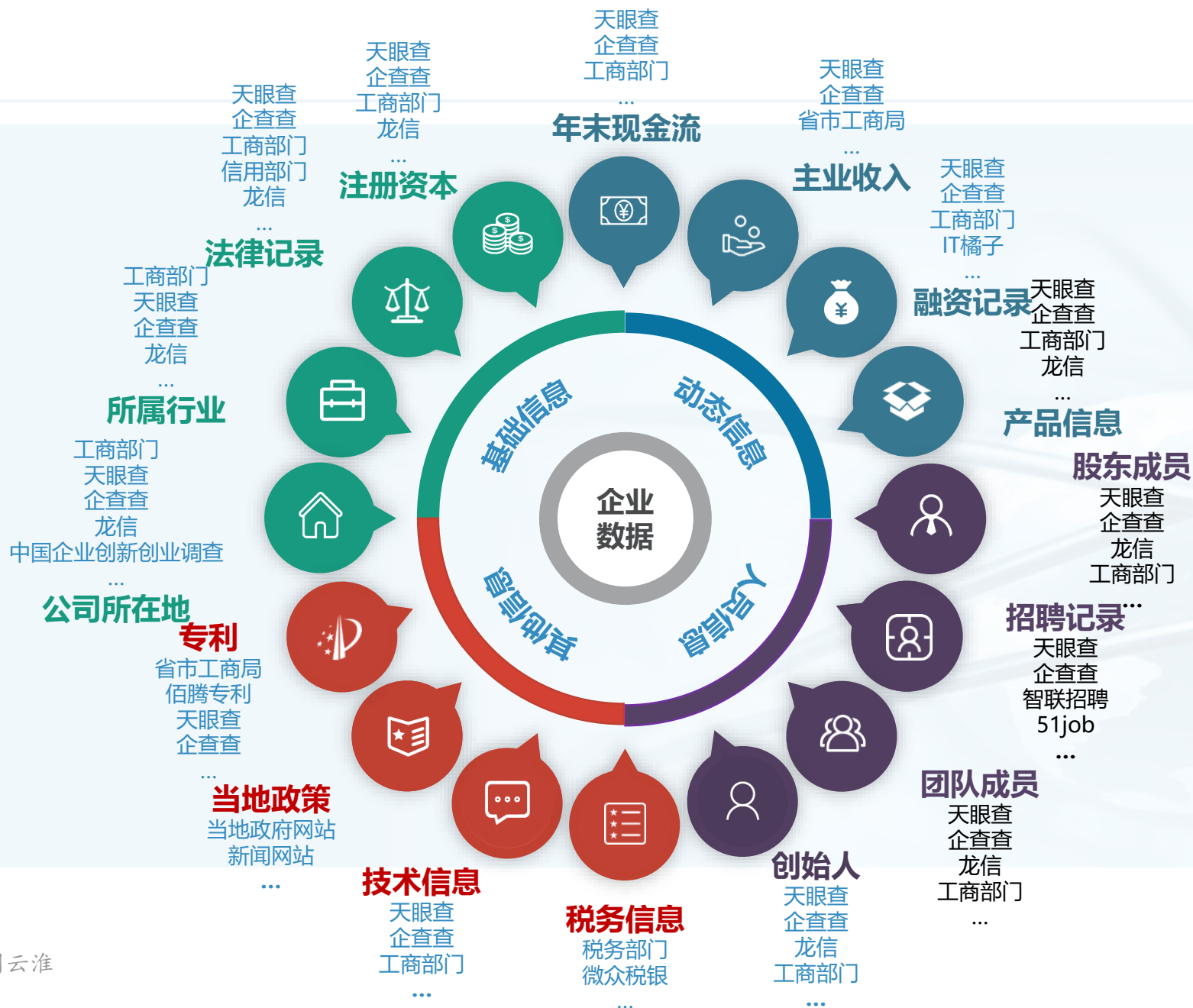


模型集成结果

交叉验证结果	预测为失业	预测为就业
真实失业样本	24624 😊	24 ☹️
真实就业样本	1164 ☹️	409925 😊

交叉验证准确率：99.72%，失业标签召回率：99.90%

企业





企业层面：基于大数据的企业活跃度分析

数据结构与算法 (Python)

中国工商企业数据

- 注册数据：注册时间、地点、行业、所有制类型、经营范围、股东总数、股东认缴资本
- 年报数据：资产总额、营业总收入、主营业务收入、纳税总额、从业人数、利润总额、净收入

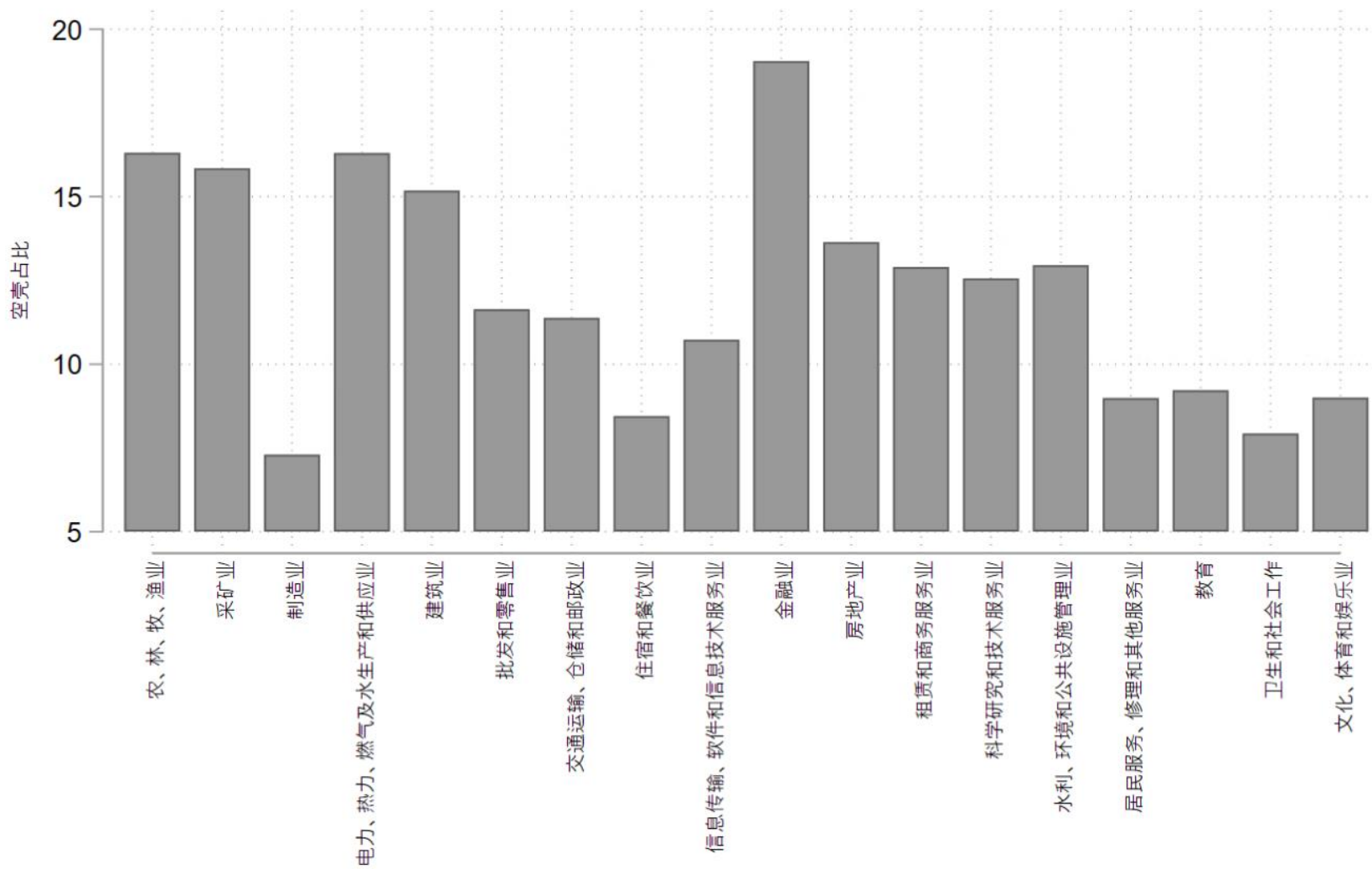
企业外部数据

- 税务数据：税务信息
- 司法风险：司法信息
- 经营风险：司法执行信息
- 公司发展：融资历史、核心团队、企业业务等
- 经营状况：行政许可、资质证书、招投标信息等
- 知识产权：商标、专利、软件著作权等

2018年中国企业创新创业调查 (ESIEC 2018)

- 辽宁、上海、浙江、河南、广东、甘肃6个省117个县
- 58500个样本，11700家个体户，46800家公司制企业
- 添加企业存续状态标签 (391个空壳)

企业层面：基于大数据的企业活跃度分析



空壳企业概率行业分布

- 金融业
- 农、林、牧、渔业
- 采矿业
- 电力、热力等供应业
- 卫生和社会工作
- 制造业
-

企业财务报表

营业收入

利润总额

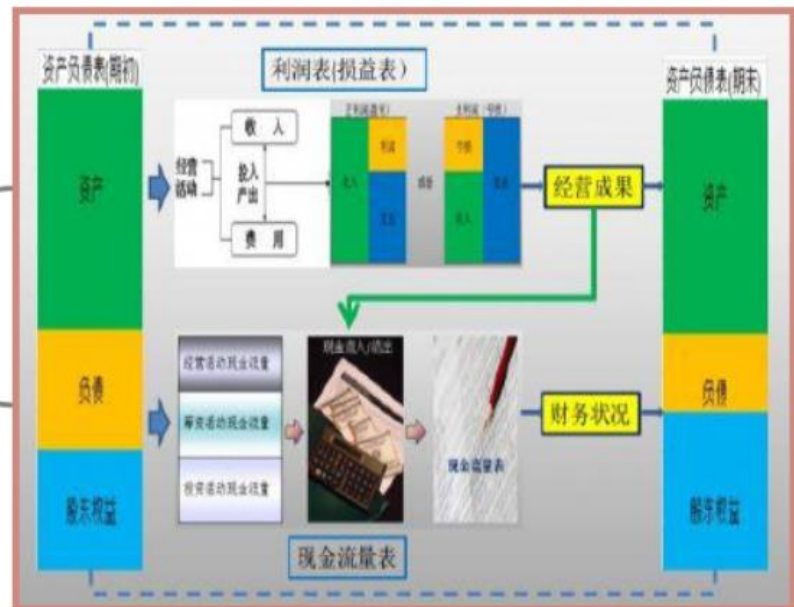
经营成本

从业人员数量

收集与计算

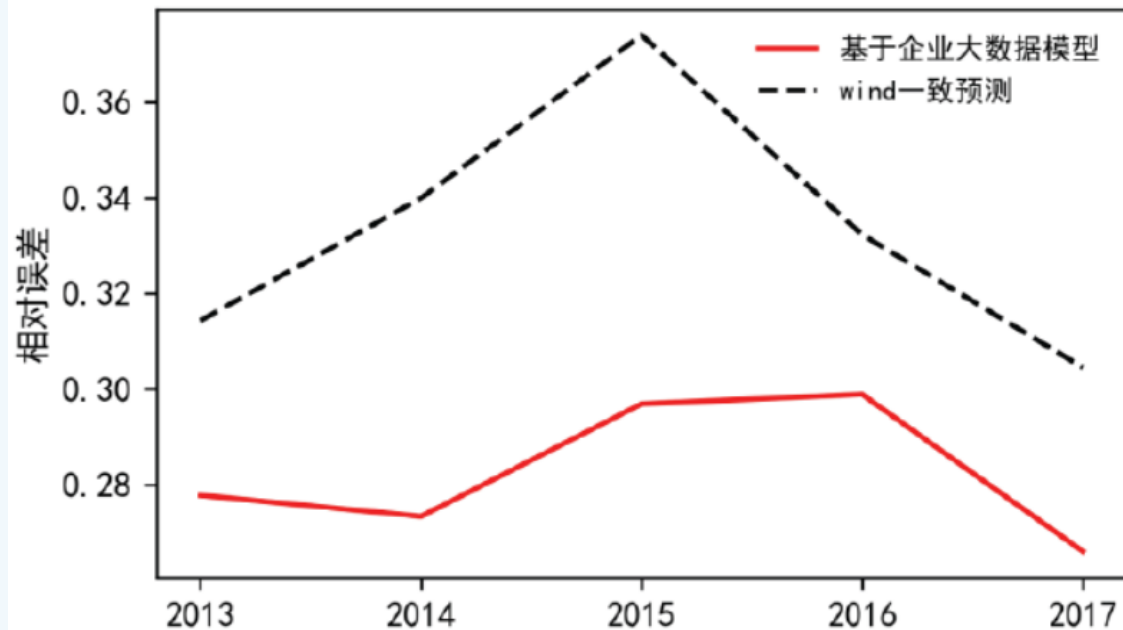
企业财报

利润表	2017-12-31	资产负债表	2017-12-31	现金流量表	2017-12-31
报告期	年报	报告类型	年报	报告类型	合并报表
期间跨度	12个月	报表类型	合并报表	报表类型	合并报表
数据来源	合并报表	负债资产:		将净利润调整为经营活动的现金流量:	
利润表摘要		现金及现金等价物	1,108,400	净利润	7,151,000
营业收入	23,801,400.00	交易性金融资产		加: 折旧与摊销	2,261,100
同比(%)	56.23	其他短期投资	13,166,900	减: 资产变动	2,266,200
营业收入	17,153,800.00	应收账款合计	641,700	其他非现金调整	-1,294,400
同比(%)	6,647,600.00	应收利息及票据	457,100	经营活动产生的现金流量净额	10,614,000
同比(%)	32.13	其他应收款	184,600	投资活动:	
税金附加	6,821,500.00	存货		出售固定资产收到的现金	2,800
同比(%)	70.82	其他流动资产	189,900	减: 资本性支出	3,200,400
净利润	7,151,000.00	递延资产合计	15,116,900	投资减少	9,230,400
同比(%)	74.01	非流动资产:		减: 投资增加	16,223,400
非经常性损益	1,666,900.00	固定净资产	1,247,500	其他投资活动产生的现金流量净额	551,400
扣非后归属于母公司股东的净利润	5,484,100.00	权益性投资		投资活动产生的现金流量净额	-9,639,200
同比(%)	59.80	持有至到期投资		筹资活动:	
研发费用	1,745,600.00	可供出售投资		债务增加	5,019,300
EBIT	6,647,600.00	其他长期投资	5,975,000	减: 债务减少	2,118,100
EBITDA	9,008,700.00	商誉及无形资产	2,100,200	股本增加	17,100
利润表摘要(NON-GAAP)		土地使用费	27,100	减: 股本减少	
净利润(NON-GAAP)		其他非流动资产	706,100	支付的股利合计	606,200
稀释每股收益(NON-GAAP)		非流动资产合计	10,055,900	其他筹资活动产生的现金流量净额	246,700
		总资产	25,172,800	筹资活动产生的现金流量净额	2,609,800
资产负债表摘要		流动负债:		现金净流量:	
流动资产	17,844,800.00	应付账款及票据		汇率变动影响	-255,100
固定资产		应交税金	127,100	其他现金流量调整	
权益性投资	14,458,100.00	交易性金融负债		现金及现金等价物净增加额	3,379,500
资产总计	55,467,200.00	短期借款及长期借款当期	775,400	现金及现金等价物期初余额	7,190,200
流动负债	15,174,000.00	其他流动负债	7,203,200	现金及现金等价物期末余额	10,569,700
非流动负债	12,563,900.00	流动负债合计	8,205,700	货币币种	CNY
负债总计	27,737,900.00	非流动负债:		原始币种	CNY
股东权益	27,709,300.00	长期借款	3,581,200	记账汇率	1
归属于母公司股东权益	25,607,400.00				

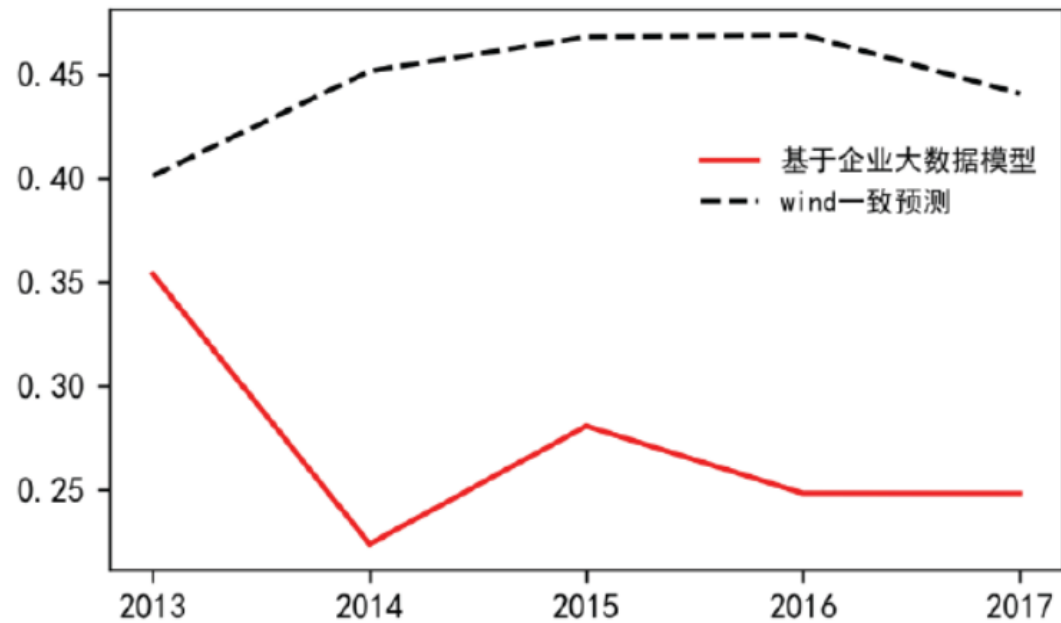


企业大数据建模结果

国有企业营业收入预测相对误差



民营企业营业收入预测相对误差



主要发现

- 基于大数据方法的营业收入预测误差明显低于专业分析师的一致预测
- 使用及时可得的企业大数据可以显著提高预测效果

地区



交通大数据的探索

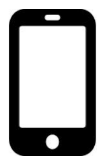
城市数据融合分析

数



Transportation

500 K
10 TB



Telecom

10 M
1 TB



Finance

16 M
1 TB

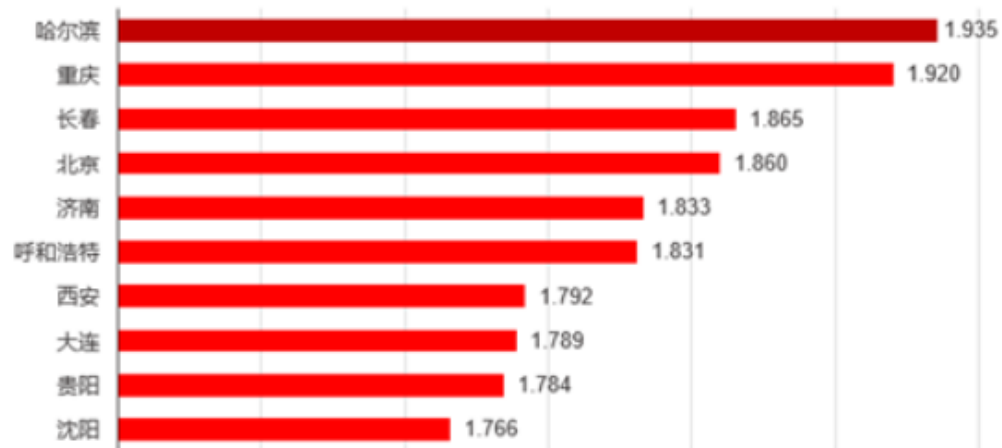


Geography

70 K
100 K

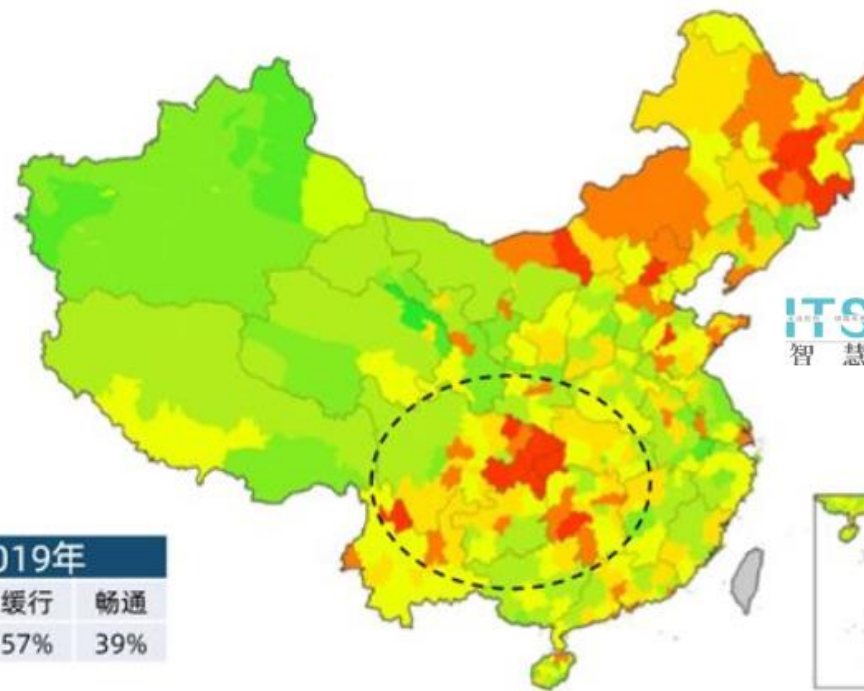
交通拥堵，人民心中的痛

年度高峰拥堵延时指数TOP10



- 1/3城市通勤拥堵
- 32城市拥堵指数 > 1.8
- 线城市、省会城市拥堵严重
- 发达城市富贵病

2019年度全国拥堵地图

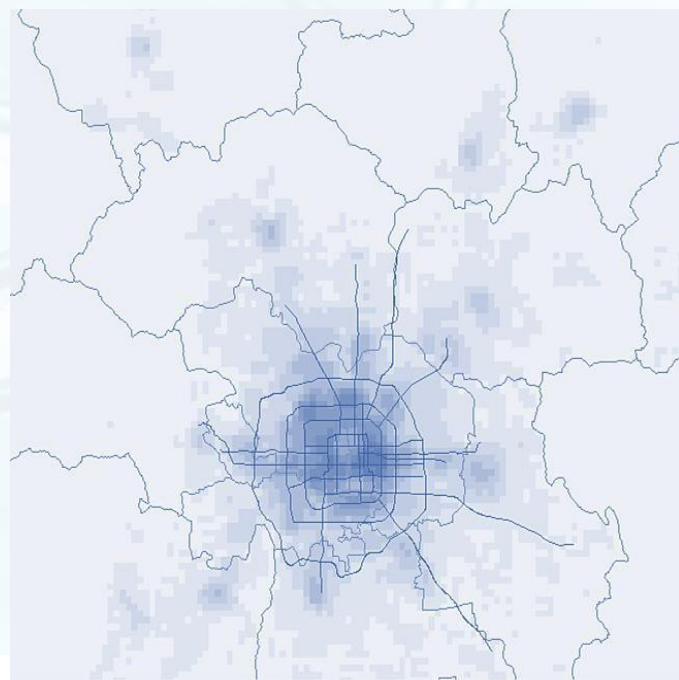
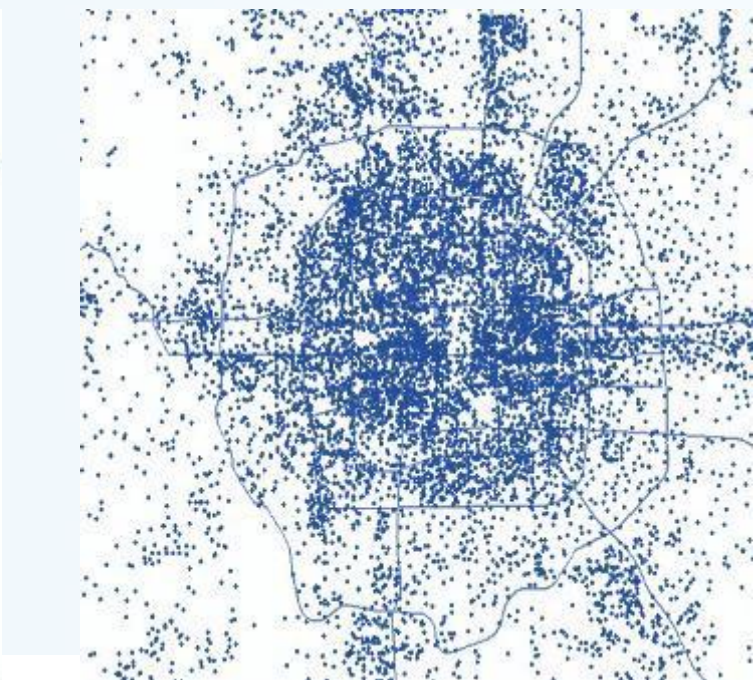
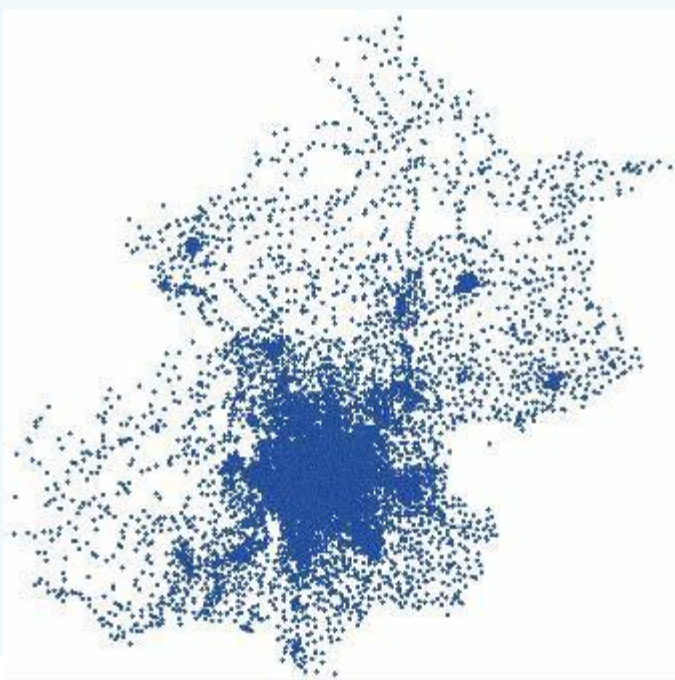


2019年

拥堵	缓行	畅通
4%	57%	39%

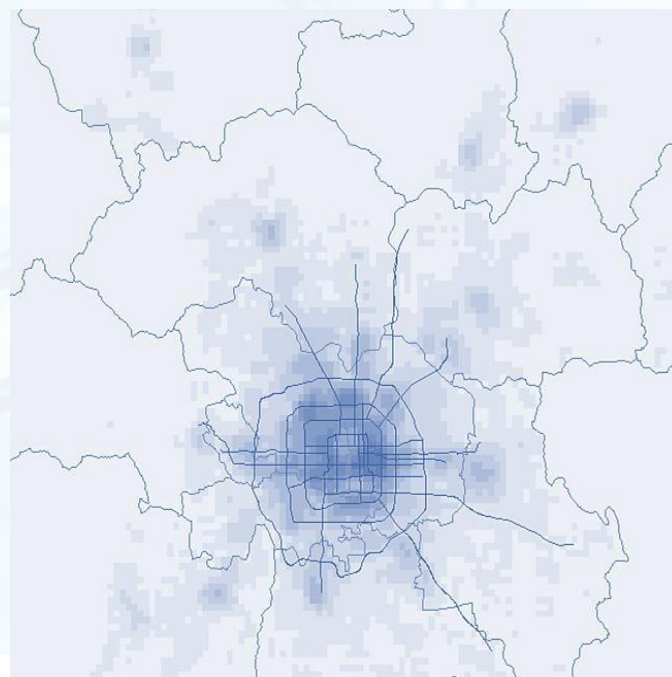
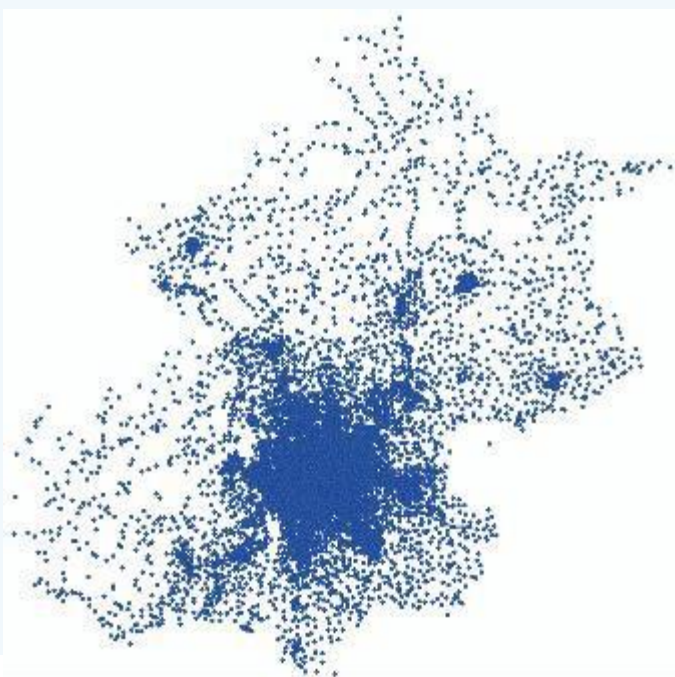
基于手机的交通出行感知

- 目前北京市中国移动手机用户数量高达1700万，约占总常住人口数的75%。
- 目前，北京市中国移动基站扇区总数超过38,000个,市区基站密度高达44cells/km²。
- 每日数据总量高达12亿条，占用空间约70~110GB。

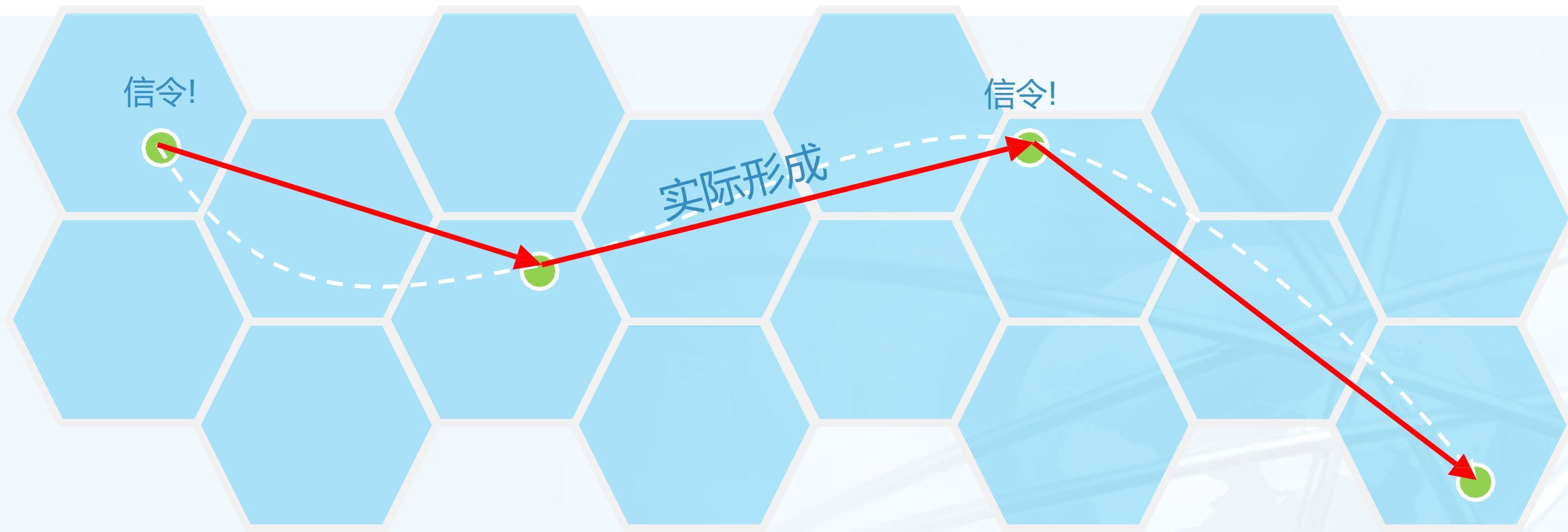


基于手机的交通出行感知

- 目前北京市中国移动手机用户数量高达1700万，约占总常住人口数的75%。
- 目前，北京市中国移动基站扇区总数超过38,000个,市区基站密度高达44cells/km²。
- 每日数据总量高达12亿条，占用空间约70~110GB。

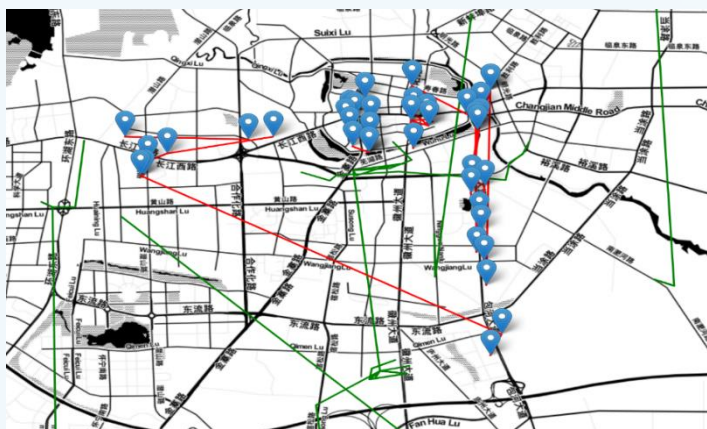
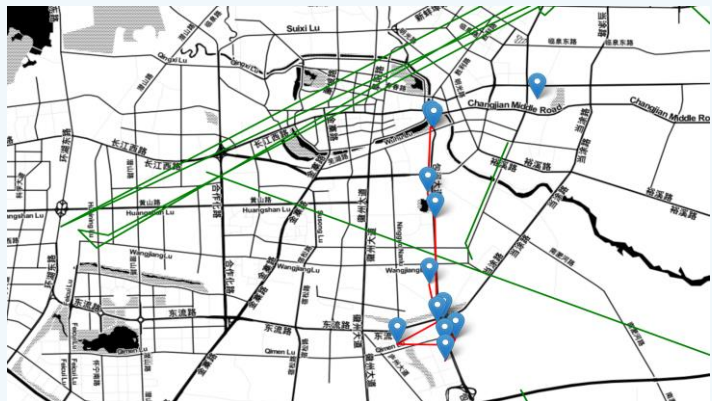


基于手机的技术原理



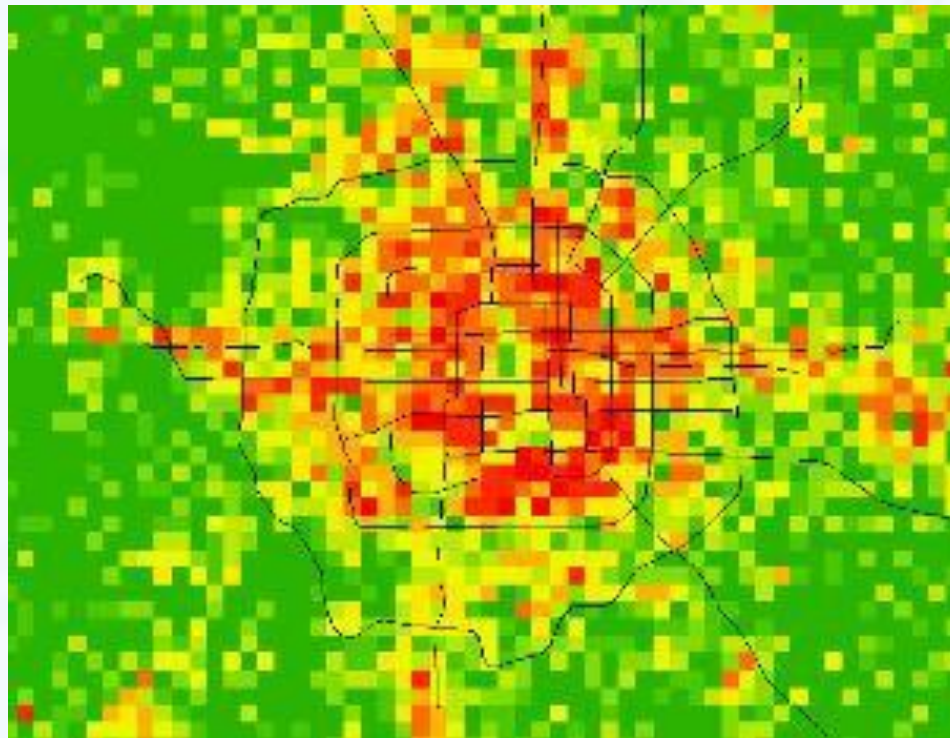
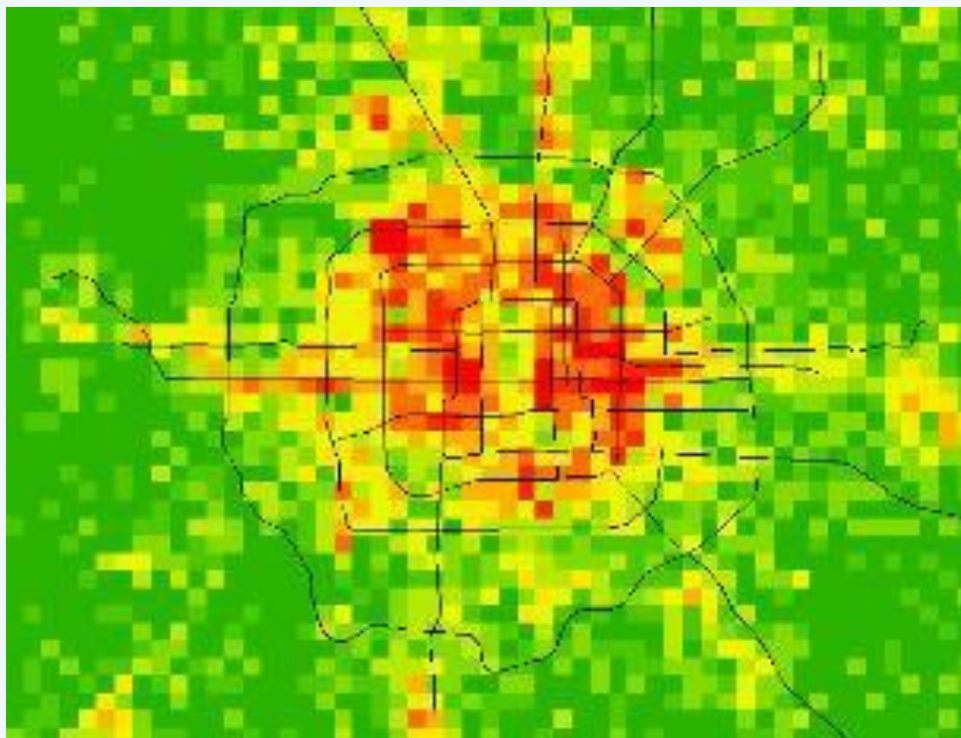
- 手机在日常使用中会与基站不断交互，即信令
- 主要的信令有7种，打电话、发短信、开机、关机、上网、2G/3G/4G切换、基站切换
- 即使没有任何行为，也会定期发送信令

出行类别分类 (公交、地铁、私家车、出租车)

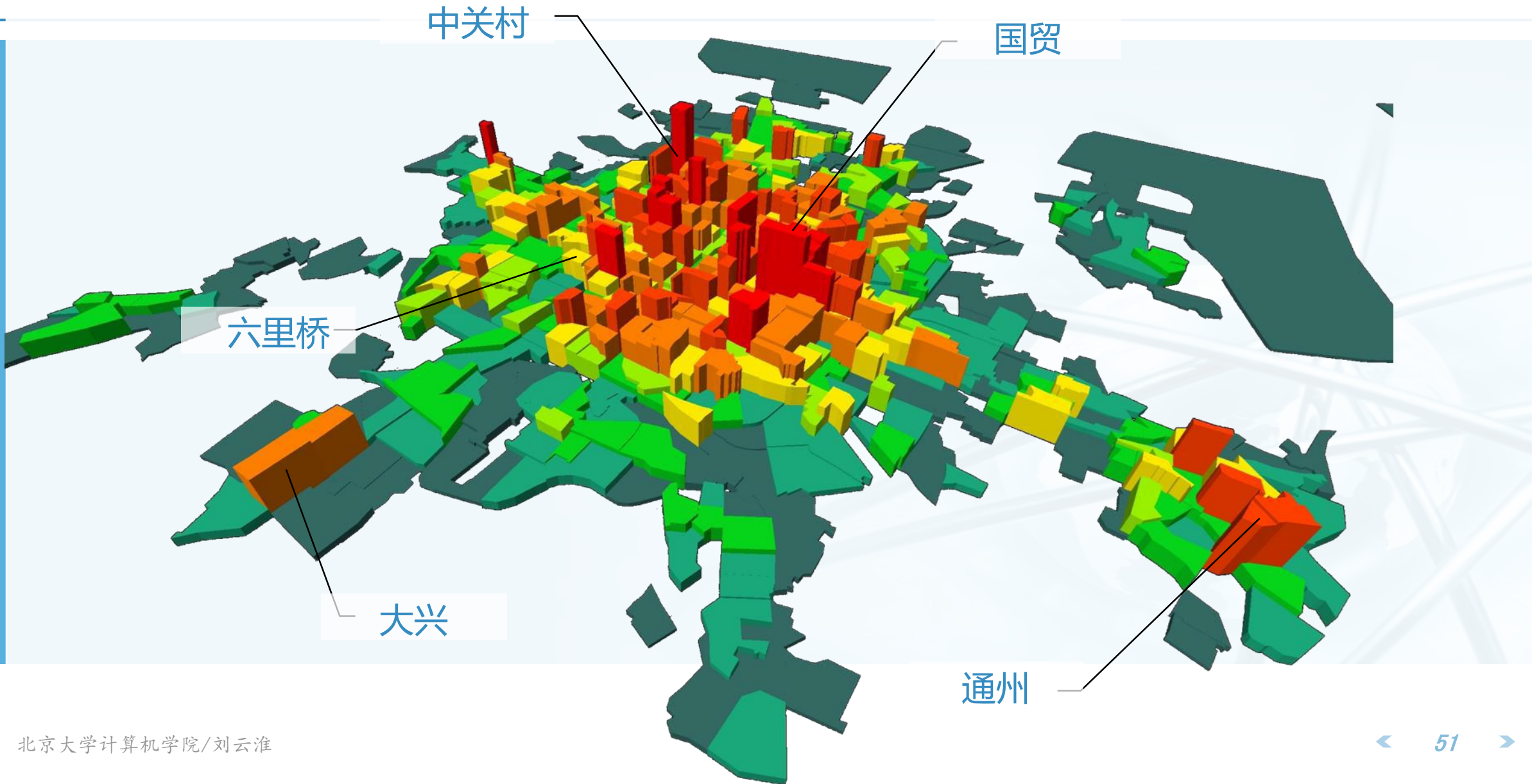


北京就业和居住分布

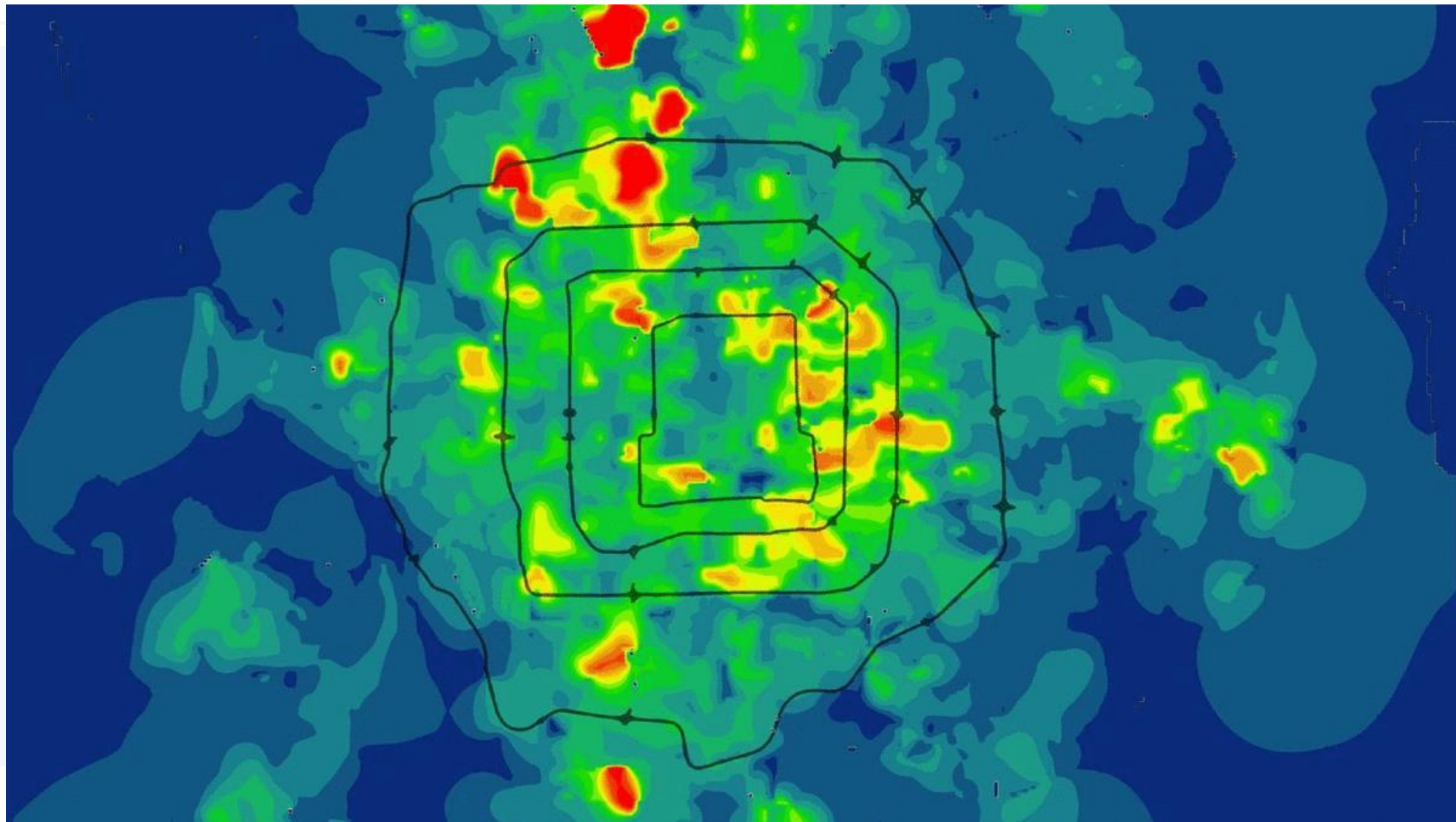
- › 根据信令行为，分析就业和居住的分布
- › 夜晚所在地为居住地，白天所在地为就业地



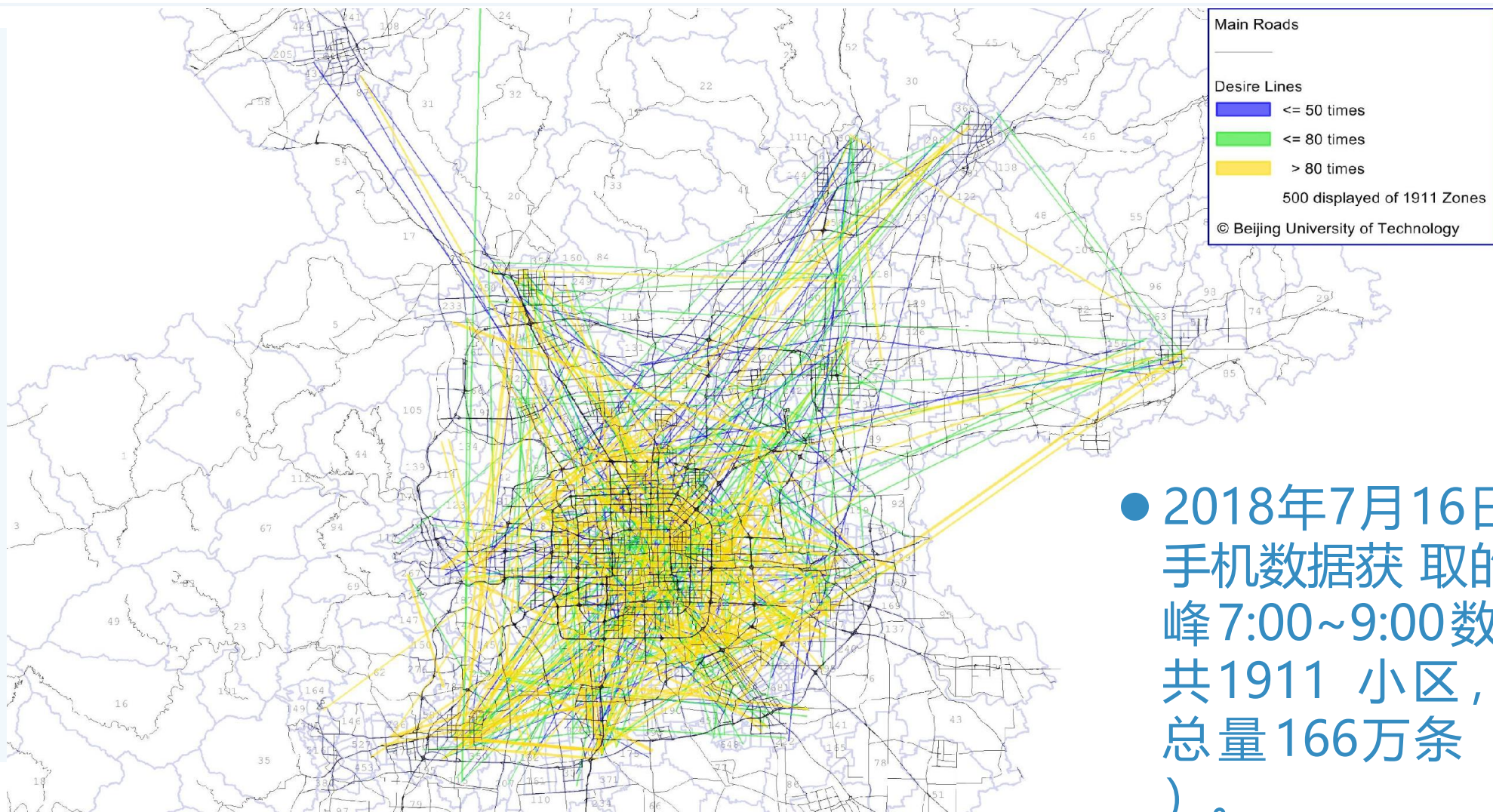
北京不同地区人口密度



北京地区不同时间人口分布



区域间出行

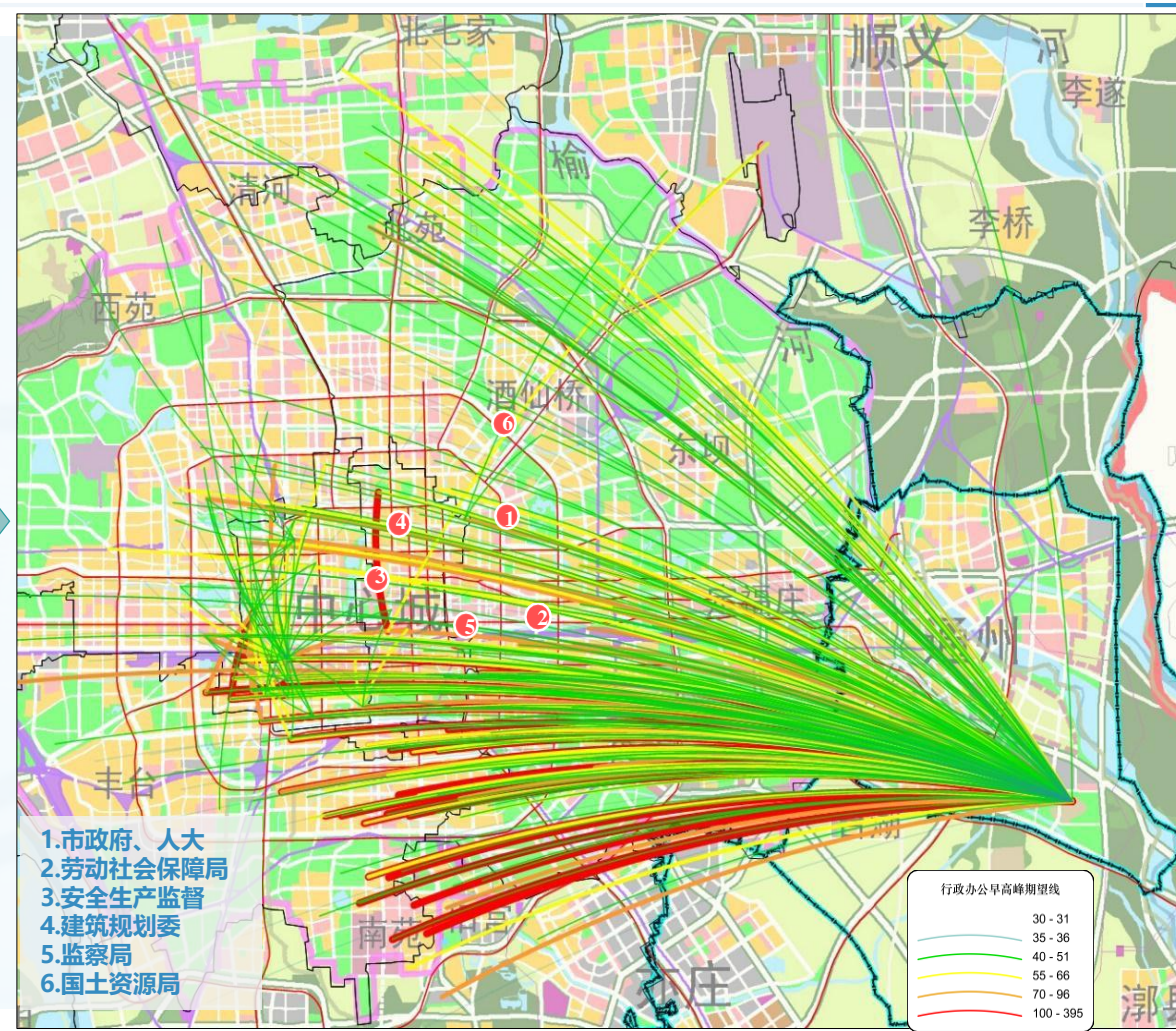
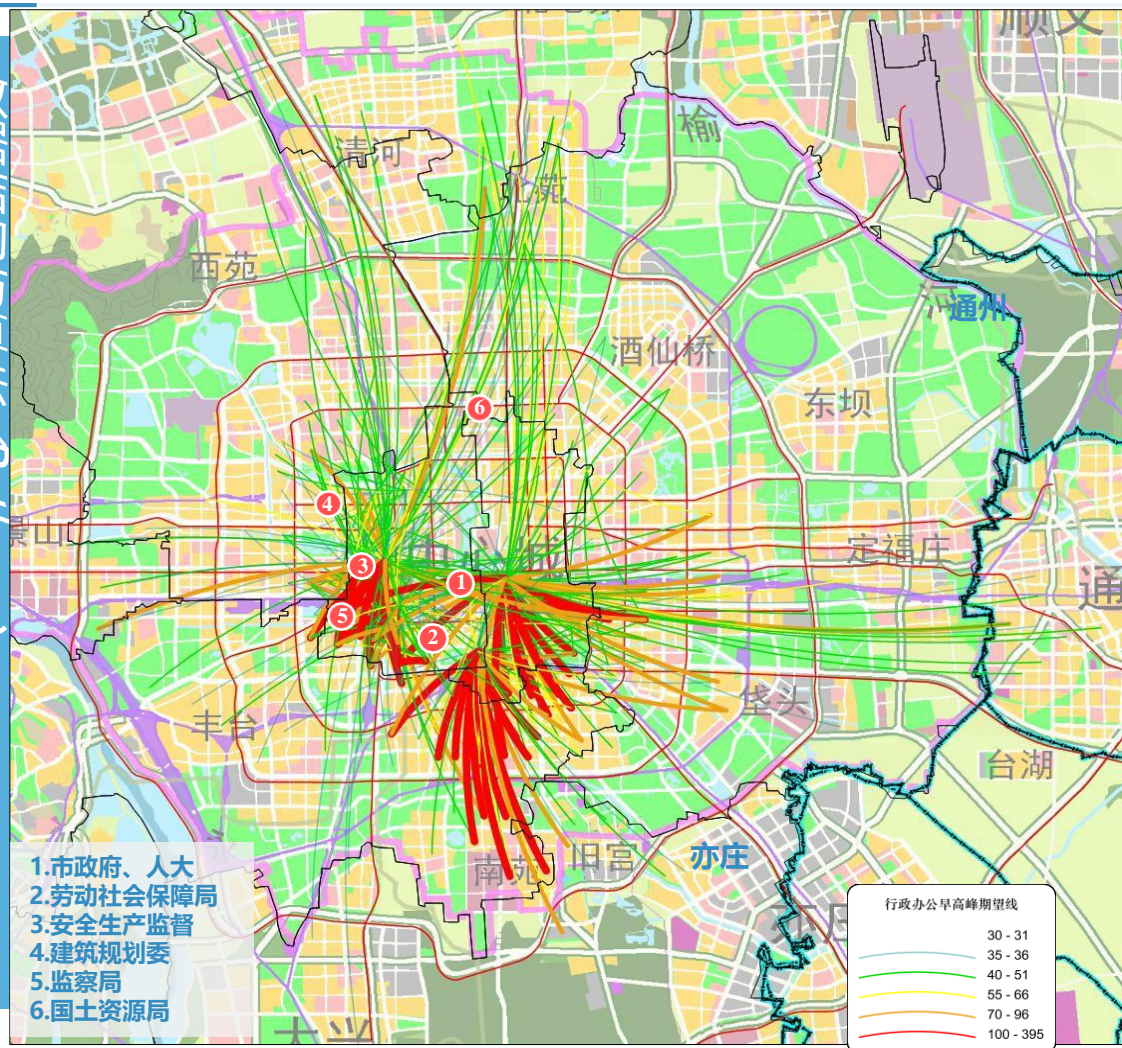


- 2018年7月16日根据手机数据获取的早高峰7:00~9:00数据。共1911小区，OD总量166万条（人次）。

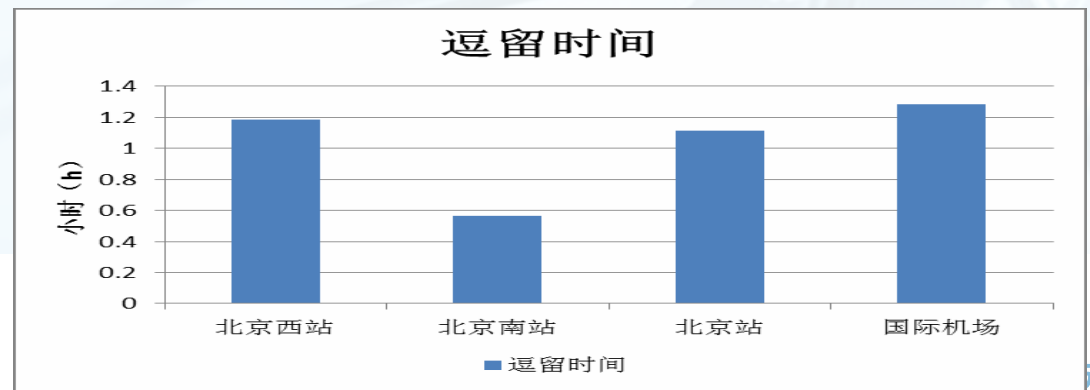
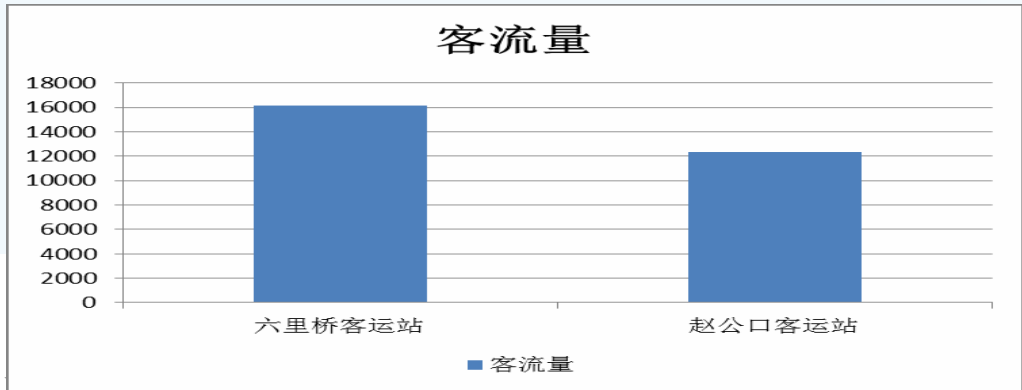
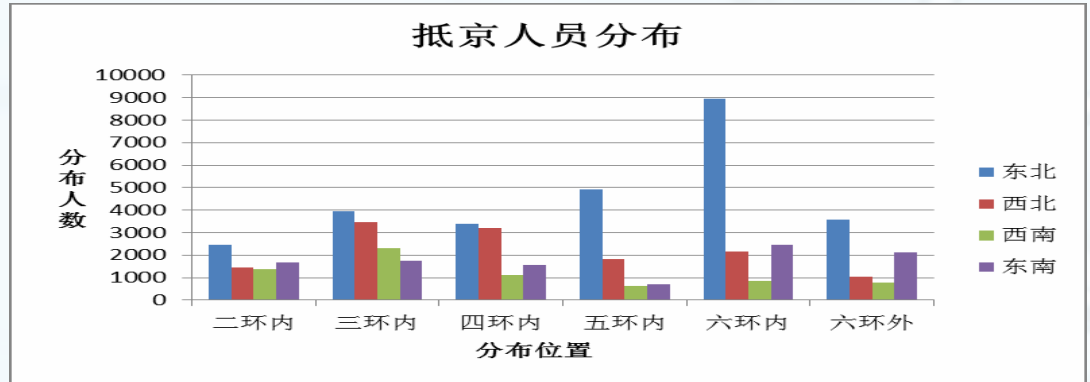
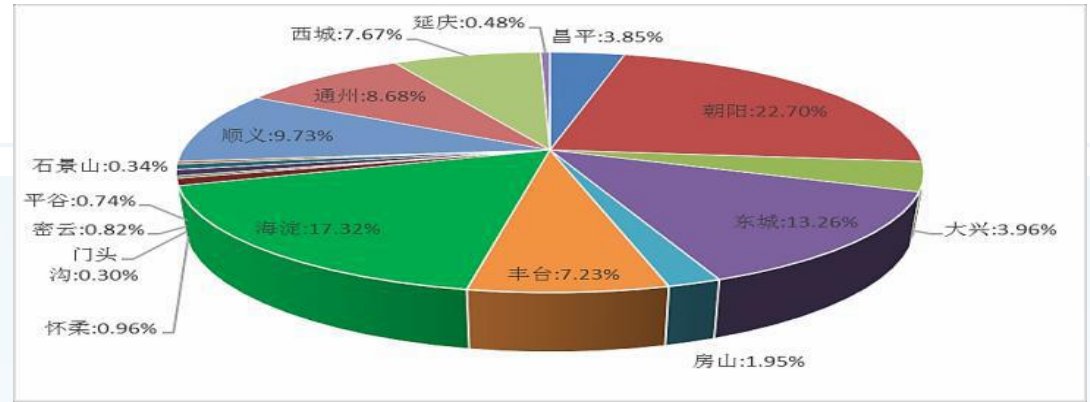
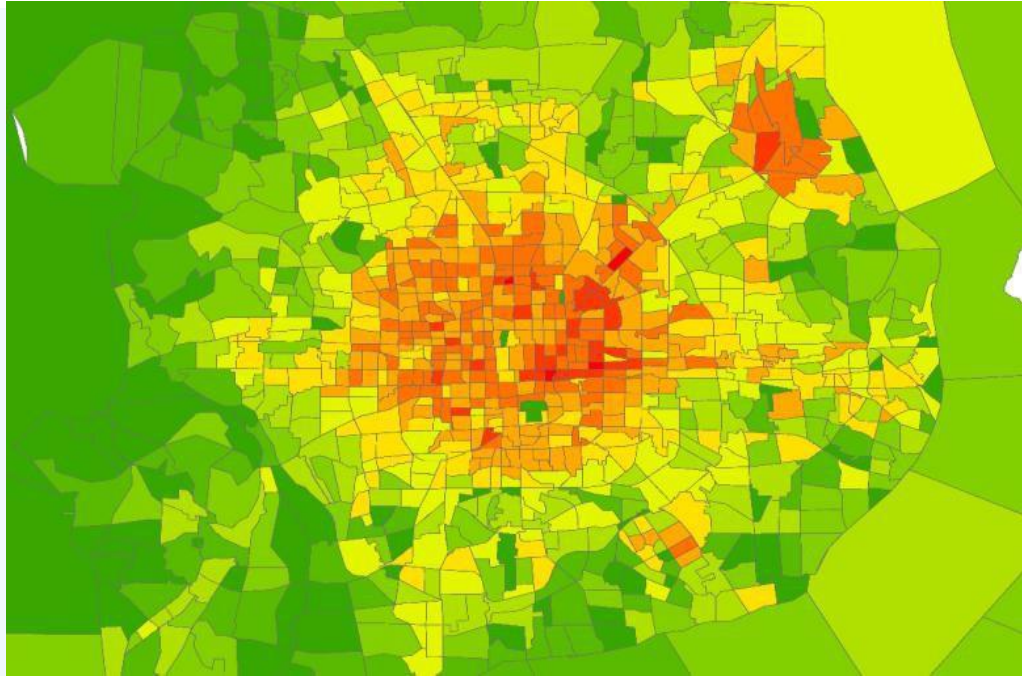
- 早高峰小汽车出行总量51.8万辆次/小时

应用案例：通州副中心对交通的影响

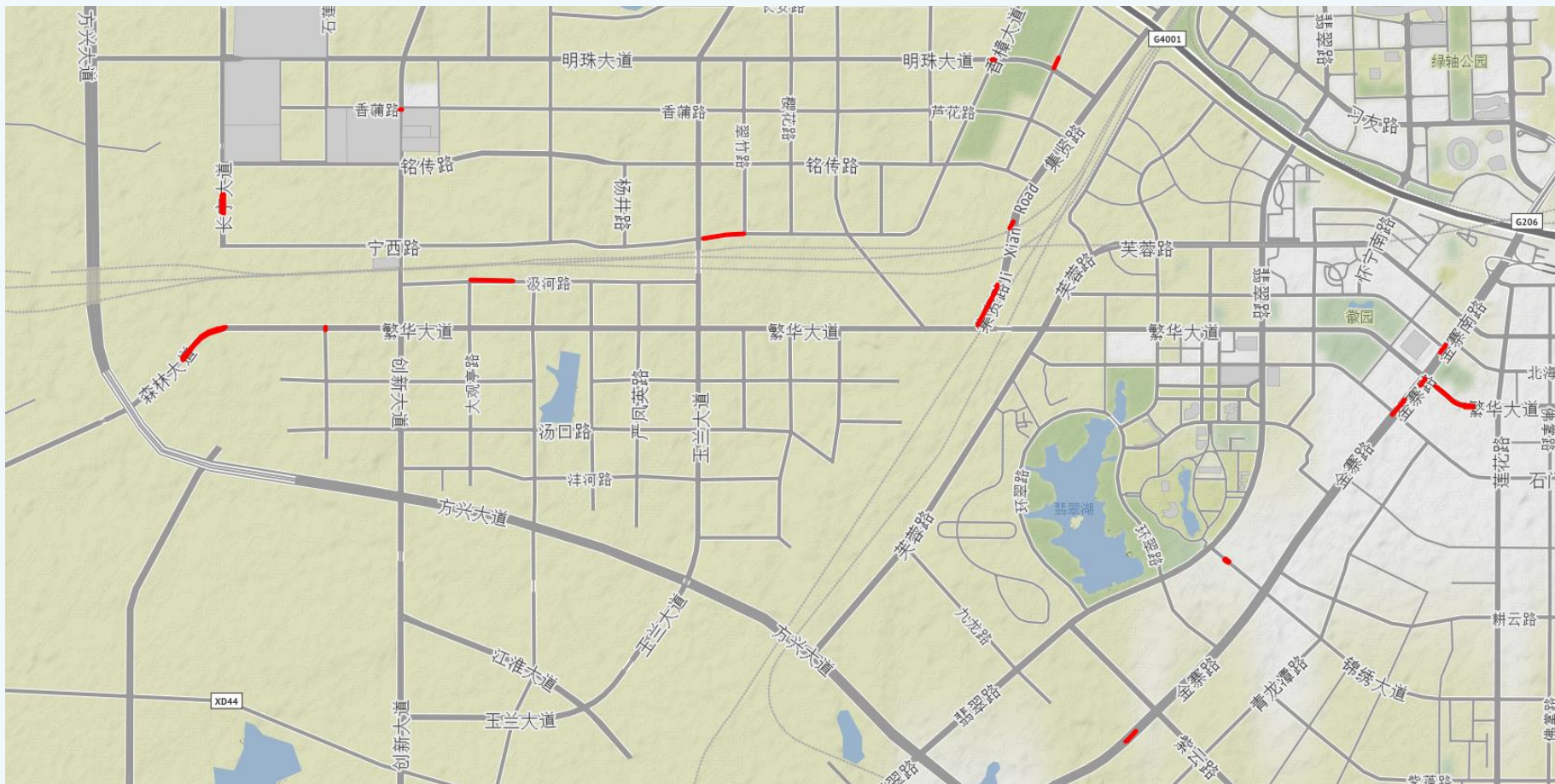
数据科学「算法」(Python)



应用：第二机场选址论证



核心拥堵点分析



城市运行监测

北京市

预警专区

运行总览

城市生命线

城市生活线

城市事件线

城市社情民意线

热点专题

城市运行监测

昨日市情

今日舆情

12345

人口动态监测

重点工作专题

70周年专题

世园会

金融风险防控

污染攻坚战

企业法人

综合指数计算结果及排名

排名	区县	生命线指数
1	朝阳区	23.70
2	顺义区	23.54
3	海淀区	23.46
4	房山区	23.19
5	东城区	22.55
6	西城区	22.36
7	大兴区	22.31
8	昌平区	22.21
9	延庆区	22.13
10	丰台区	21.79
11	石景山区	21.22
12	密云区	21.22
13	通州区	20.96
14	怀柔区	20.96
15	平谷区	20.26
16	门头沟区	19.04

数据来源: 手机信令 2019/03/21 13:10



- 工作批示
- 电话连线
- 视频连线
- 待办事项

城市生命线



暂未监测到预警信息

上次监测时间: 2019-07-30 20:07:76

城市生活线



北京石景山一居民楼发生...

气象预警

2019-3-19
北京市

发现 3 预警

* 数据来源: 北京市突发事件预警信息发布...

城市事件线



北京西城一小发生伤害...

突发事件

2019-1-8
宣武附一小

发现 3 预警

* 数据来源: 互联网

城市社情民意线



居民区幼儿园小子万...

投诉举报

2019-1-20
北京市朝阳区东坝乡和敬路4号院

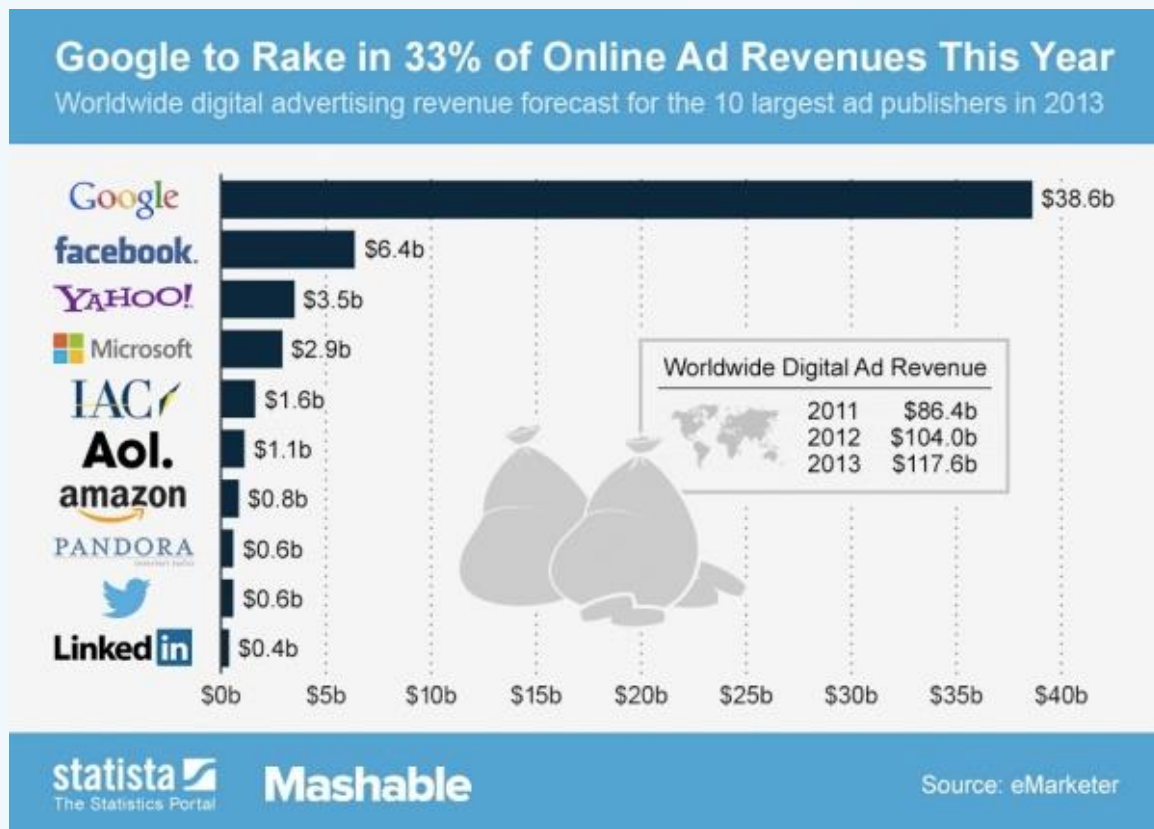
发现 3 预警

* 数据来源: 12345

北京什刹海游船公司

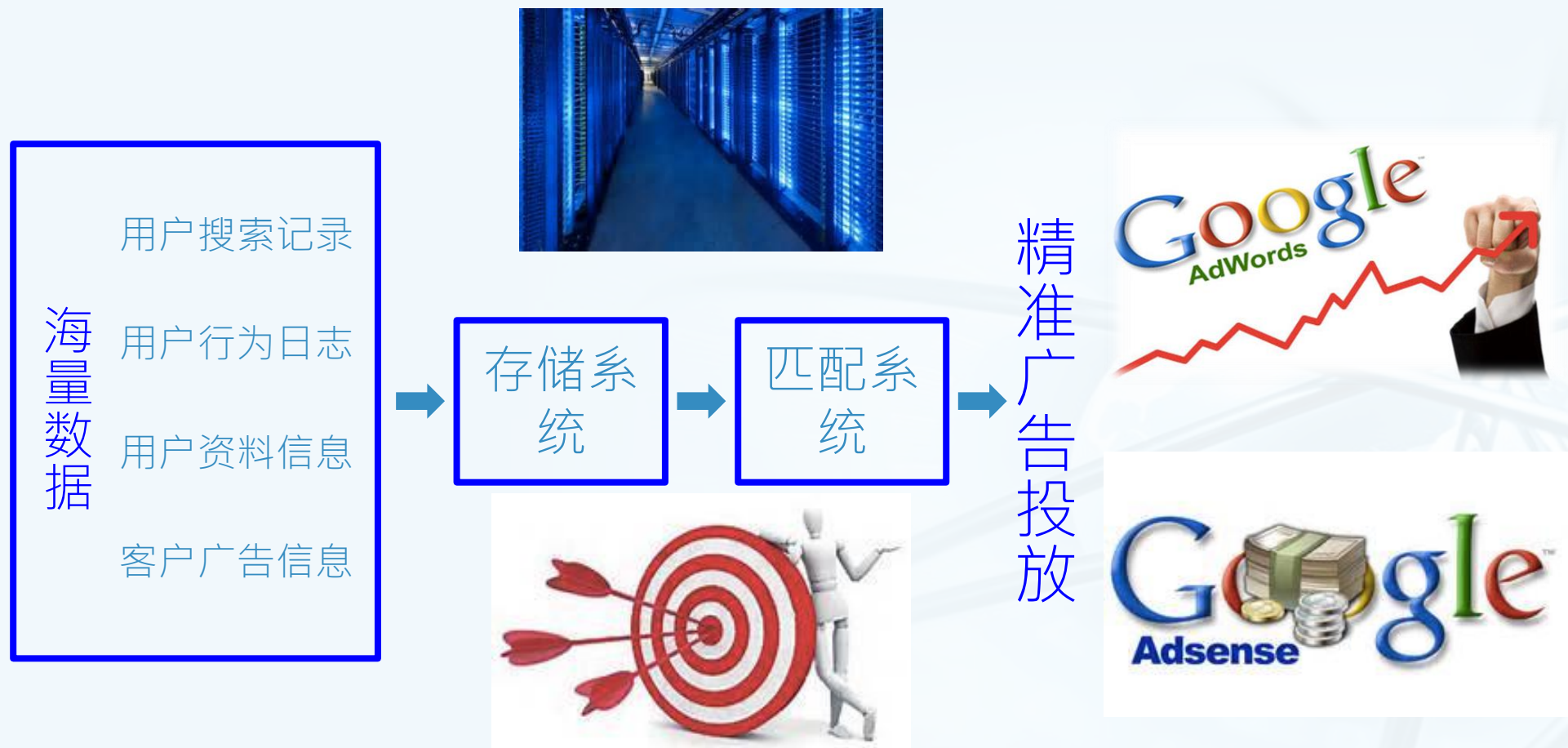
大数据的商业价值

谷歌的金库：精准广告投放



2013 年，谷歌从在线广告获得 386 亿美元的收益，基本相当于全球在线广告收入的三分之一。这 386 亿美元的收益占到谷歌 2013 年总营收的 65%

谷歌的金库：精准广告投放



大数据下淘宝 “千人千面”

2017年11月12日凌晨，阿里公布了淘宝天猫“双十一”购物狂欢节全天的销售额：

支付宝全天成交金额为：

2135亿

比2015年的1682亿增长：

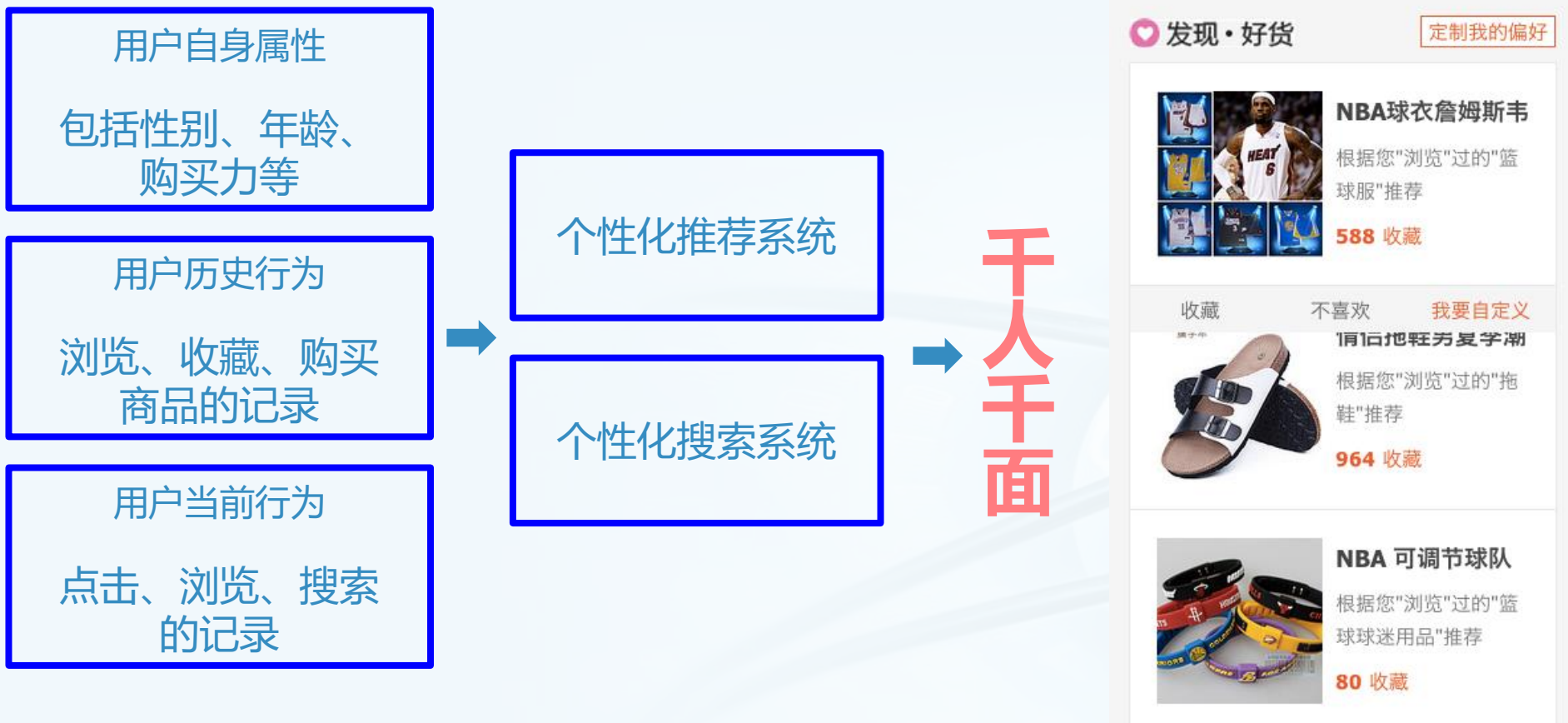
27%

订单：

12.35亿



大数据下淘宝“千人千面”



爱奇艺的大数据-用户规模

- › 超过**10亿**的客户端下载量, **3.4亿**全平台月度活跃用户
- › 每个活跃用户平均每天使用**2小时45分钟**
- › 所有活跃用户一天使用总时长约为**10980年**
- › 重大直播事件**1000万**用户同时在线



月度用户数(UV) 每天覆盖人数(UV)

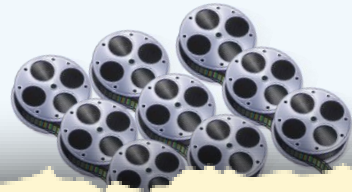
3.4亿

5000万

直播在线峰值

1000万

爱奇艺的大数据



《赘婿》 上线第一天，就带来了100万新增注册会员，超过3000万的收入，远超视频版权支出

服务器日志 (10+T/日)
Web server, LB, CDN server, 广告投放, 搜索, P2P server, application trace...

客户端日志 (10+T/日)
PC客户端, Flash/网页客户端, 移动App, P2P引擎...

大数据的价值 – 商业运营

- › **古装偶像剧**在上个月产生了多少次播放？上周呢？昨天呢？前一个小时呢？
- › 不同渠道带来的独立用户有多少？他们的停留时间和留存率如何？
- › 每一部视频的投入（版权、带宽）和产出（广告收入，付费点播）比如何？分地区分析？分终端分析？
- › 过去6个月**宋轶**、**郭麒麟**在视频观众里受关注程度变化趋势如何？



一个馒头引发的...

- › 淘宝：我吃完一个馒头，问我，你要不要来一个馒头？
- › 豆瓣：我买了个馒头，他问我，你要不要来碗米饭？
- › 百度：“老板，给我俩馒头。湖南株洲馒头机制造厂供应优质馒头机”
- › 谷歌：我想买个馒头-Z馒头无敌好吃！（Ad）为你找到X家馒头店，（正文）。
提示：是否使用Google shopping？
- › 腾讯：正当我要买馒头时，在后面拍了拍我，“同学，来我这买，一模样，还有豆沙馅，充值还有馒头皮换装！”
- › 360：让我摸一下，免费送馒头。
- › 小米：馒头便宜啦，就是卖完了，预订吧。

蚂蚁金服 “让天下没有难借的钱”

- › 传统银行借贷，即使不计其实体营业点运营成本和各种人力成本，借贷运营成本很高
- › 倘若借贷一元，其涉及借贷数据的硬件及软件花费将远超其借贷利息所得收益
- › 阿里小微信贷，利用大数据技术，使得信贷门槛降低，真正一元起贷



蚂蚁金服（网商贷）

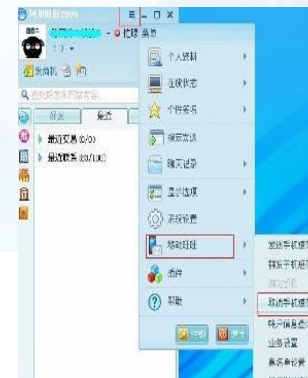
- › 基于大数据的放贷
- › 2013年阿里小贷全年新增投放贷款1000亿元
- › 截至2014年2月中旬，阿里小贷累计投放贷款超过1700亿元
- › 服务小微企业逾70万家，户均贷款余额不超过4万元，不良率小于1%
- › 远低于传统银行的10%-15%的不良率
- › 秒级放贷

阿里小贷发展路径图



蚂蚁金服的水文模型

- 某河道水位达到某个值，但人们无法依据这个数值采取应对，是准备防汛还是不做任何动作？下月河道水位走高还是走低，会否影响防汛等河道管理的措施？
- 但如果将这个值放到历史数据及周边河道数据中，就可以做出一定判断：如比过往同期，这个数据是否变高了，高了多少；以往这个时期后，河道水位又是怎么变化的。
- 通过该店铺自身数据的变化，以及同类目类似店铺数据的变化，判断客户未来店铺的变化。
- 如过往时点，该店铺销售会进入旺季，销售额就会增长，对外投放的额度就会上升，结合这些水文数据，系统可以判断出该店铺的融资需求；
- 结合该店铺以往资金支用数据及同类店铺资金支用数据，可以判断出该店铺的资金需求额度。
- 判断的数据甚至包括购物评价，阿里旺旺聊天内容



即时配送大数据

即时配送在中国

• 即时配送在中国快速增长

- 2021年全年超过287亿订单，年增长率24.3%
- 2026年，即时零售市场规模将达到10,000亿
- 占全部物流订单的25%以上



• 新型的O2O的产业模式

- 全国超过3亿的用户
- 250万的商户
- 超过85万的骑手

2021年平均每人21.4个即时配送订单

行业订单规模

• 多种服务模式

- 外卖、京东到家、盒马生鲜、同城快递、闪送



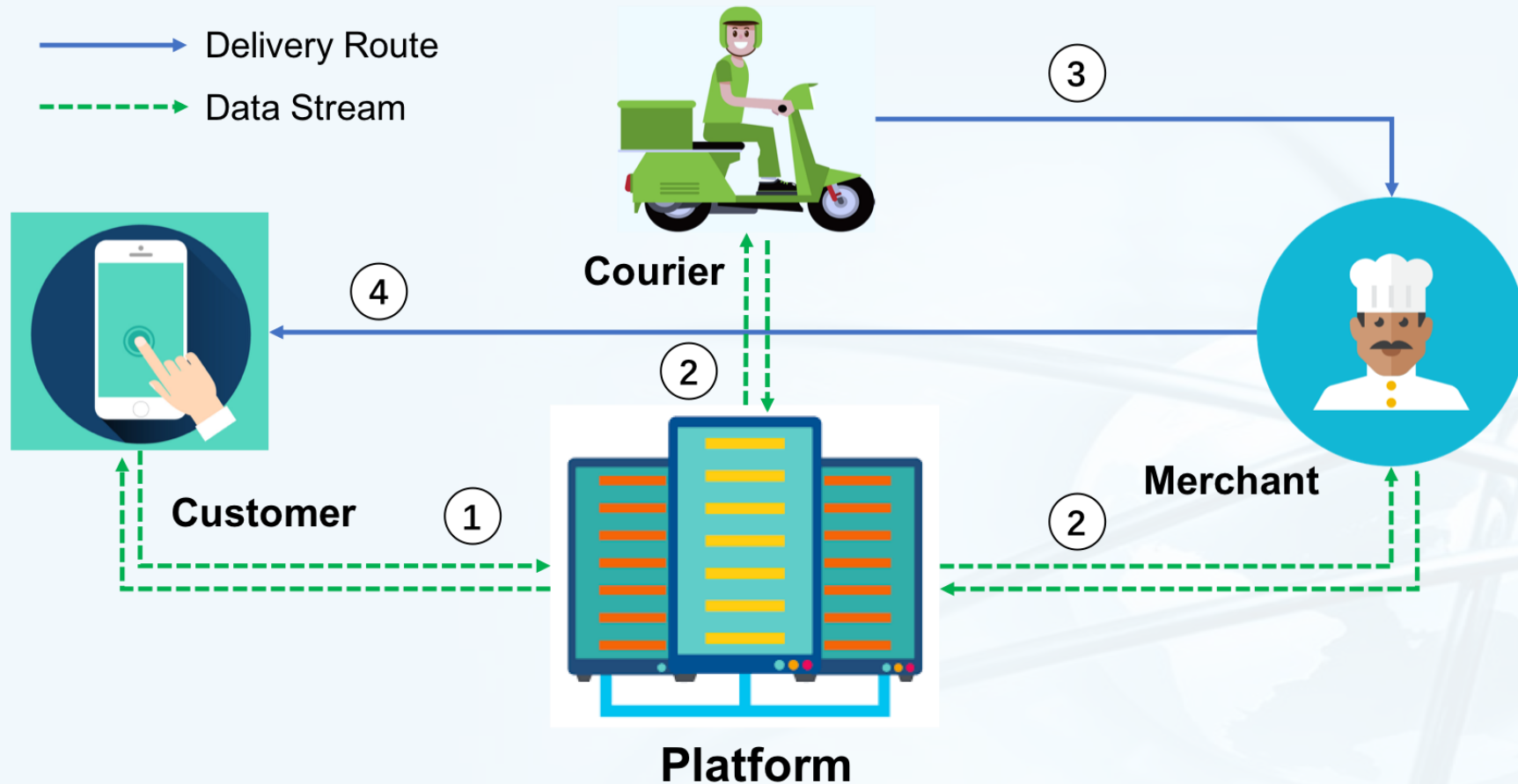
来源：综合企业及专家访谈，由艾瑞咨询研究院自主研究及绘制。

全世界范围的即时配送

2019年全球超过1019亿美元的市场 - *ResearchAndMarkets.com*



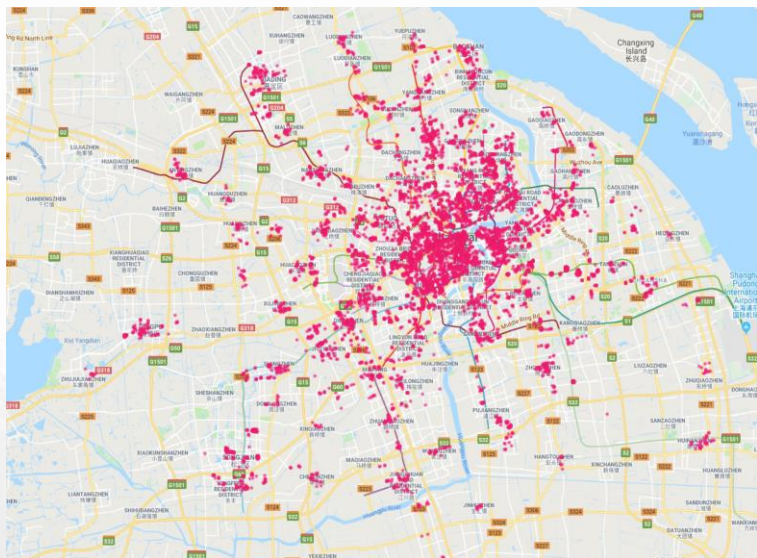
即时配送的典型应用场景



一个典型的即时配送业务场景: (1) 用户在即时配送业务平台下单; (2) 平台通知商家准备包裹并将订单分配给某个配送员; (3) 配送员到商家取包裹; (4) 配送员将包裹送到用户手中

即时配送的业务特点

- **平台的终极目标：保留客户的同时，尽可能降低超时率**
 - **配送时限**，能大幅度提高用户体验，保留客户粘度
 - 超时订单会给客户很大折扣，甚至**免单**
- **订单分配：平衡成本和超时罚款**
 - 平台唯一能控制的事情，每个配送员会同时配送多个订单
 - 传统物流没有时限和超时的要求



Merchants of Instant Delivery in Shanghai



A typical courier

相关工作：拼车场景

• 拼车场景

- 一名司机在一次行程中需要搭载多名乘客，然后分别将他们送达目的地
- 平台目标希望可以缩短司机的**行驶距离或服务时长**

• 拼车订单分配算法：优化建模

订单分配与司机路线规划被统一建模为优化问题，并可通过多种优化方法求得最优解（最短路径方案）

- 在Alonso-Mora等人的文章[1]中，司机与乘客被建模为RTV图，订单匹配问题被进一步抽象为**整数线性规划问题**，以贪心策略为始进行搜索，算法最终可以收敛到最优解。
- 在Zhao等人的文章[2]中，作者提出了支持灵活设定乘客接送地点的最佳司乘匹配策略和车辆路线计划。作者引入时空窗的概念，将该订单分配问题建模为时空网络中**带有时空窗口的收货和交付问题**（PDPSW）并设计了基于拉格朗日松弛的求解算法。

[1] Alonso-Mora J, Samaranayake S, Wallar A, et al. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment[J]. Proceedings of the National Academy of Sciences, 2017, 114(3): 462-467.

[2] Zhao M, Yin J, An S, et al. Ridesharing Problem with Flexible Pickup and Delivery Locations for App-Based Transportation Service: Mathematical Modeling and Decomposition Methods[J]. Journal of Advanced Transportation, 2018, 2018.

相关工作：拼车场景

• 场景相同之处

- 平台将订单分配给司机或骑手，由司机或骑手完成订单的取送工作
- 司机或骑手需要在**一次行程**中完成**多个订单**，包括乘客/物品的取、送

• 场景差异之处

- 优化目标不同
 - 拼车场景: 最小化行驶距离或服务时长
 - 即时配送: 最小化超时率
- 约束条件不同
 - 拼车场景: 乘客多在上车点等候上车
 - 即时配送: 订单包准备时间差异较大，部分餐品准备时间达数十分钟

• 拼车算法在即时配送问题中的表现

	拼车算法路径	骑手自行选择路径
超时率	2.32%	1.31%
平均配送时长 (分钟)	46.7	36.6

表1 实验结果对比

北京计算机学院/刘云淮 实验中，我们对拼车算法给出的最优路径与骑手自行选择的路径进行了对比。结果表明，拼车算法最优路径的超时率要远高于骑手的真实路径，骑手的单均配送时长也大大增加

- **难于求得最优解**

- 缺少实时信息，如交通状况、天气状况等
- 缺少未来的订单分布信息（极大地影响路径的最优性）
- 算法复杂度过高，为NP-hard

- **最优路径不被骑手所采纳**

- 骑手对于路况与未来的订单分布有自己的经验判断，对于推荐路径的采纳率极低

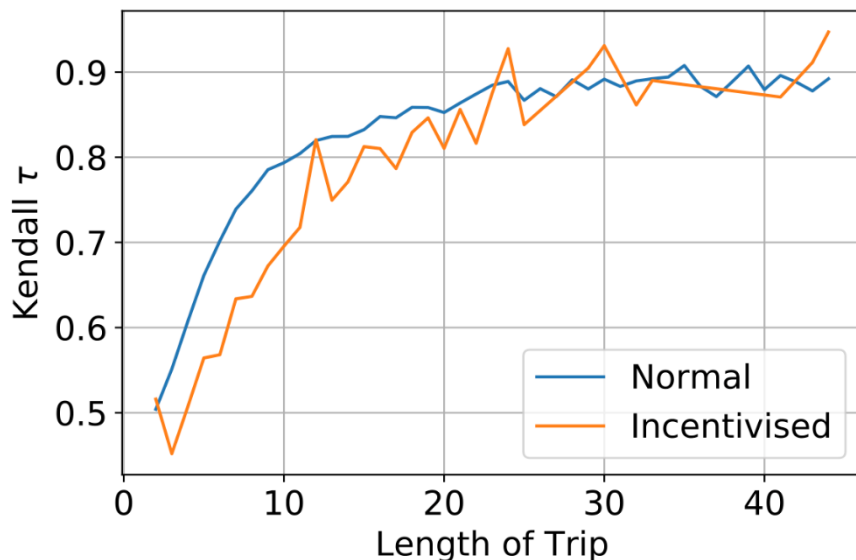


图3 次优路径与骑手自选路径的差异图

- **路径预测**

- 假定某位骑手被分配了 K 个订单，路径预测需给出骑手对**这 K 个订单的商家和顾客位置的访问顺序**

- **以路径预测为核心的订单分配策略**

- 订单分配以街区划分。通常有十几位或者几十位骑手负责附近几公里之内的所有订单的配送工作
- 对于某时刻给定区域内的 M 位骑手和 N 个待分配订单，订单分配算法需要给出 **M 位骑手和 N 个订单的匹配策略**

• 单点预测

- 在骑手的某个决策时刻，拥有多个待选位置，模型预测给出他下一个最可能去的位置。
- 建模为二分类问题：对于每个待选位置，如果它在此次决策中被骑手选中，则标签置为1，否则，标签置为0。

• 感知距离模型(PDM: Perceived Distance Model)

- **感知距离**：人对两个地点之间的距离的估计依赖于他对于实际距离的理解，受到个人的记忆、知识、经验等因素的影响。

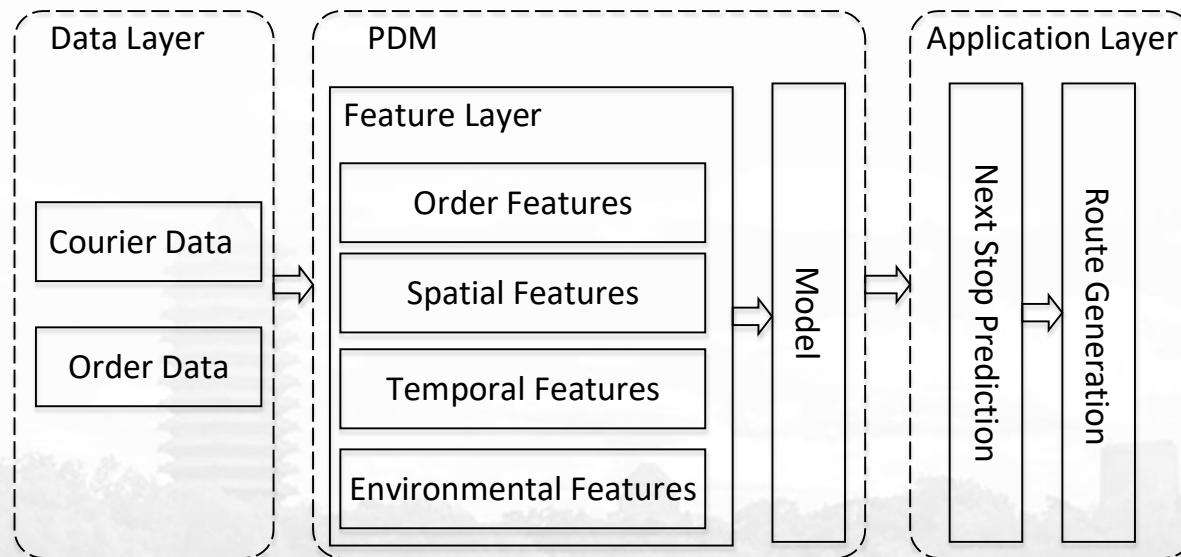


图4 路径预测模型

• 特征构建

特征类型	特征名称	特征含义
空间特征	D_sd	从骑手当前位置到该点的距离
	B_nearest	该点是否为距离骑手最近的候选点
	D_mc	该订单中的商家到用户之间的距离
	D_other	从该点到其他所有候选点的距离之和
时间特征	T_pick	距离取货时间的时间间隔
	T_ToD	距离订单超时的时间间隔
	B_urgent	该点是否为候选点中最紧迫的订单
订单特征	O_money	订单金额
	O_num	订单物品总数
	O_benefit	订单优惠金额
	O_m	商家编号

表2 特征含义表

• 数据集

- 数据采集于实际生产环境
- 2019.03.15 – 2019.04.15期间，上海市和成都市的6000名骑手的232万条订单信息

• 评价指标

- 准确率
- 编辑距离
- 超时率
- 平均超时时间

• 对比算法

- 距离模型：骑手被假设每次只选距离当前位置最近的候选点
- 拼车优化算法
- 最优路径

• 综合评价

Metric	Distance-based Method	Ride-sharing	OSquare Route	Optimal Route
Accuracy	0.6634	0.6857	0.7222	0.6977
Kendall	0.7060	0.7262	0.7486	0.7337
Edit Distance	3.3986	3.1920	2.9069	3.0904

表3 实验结果汇总

• 以时间维度观察

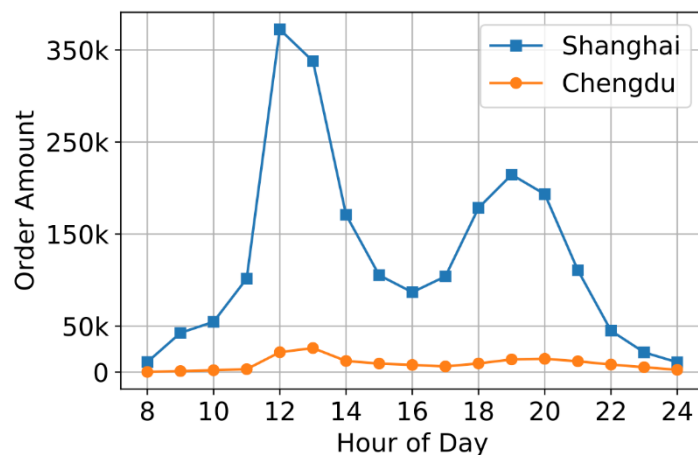
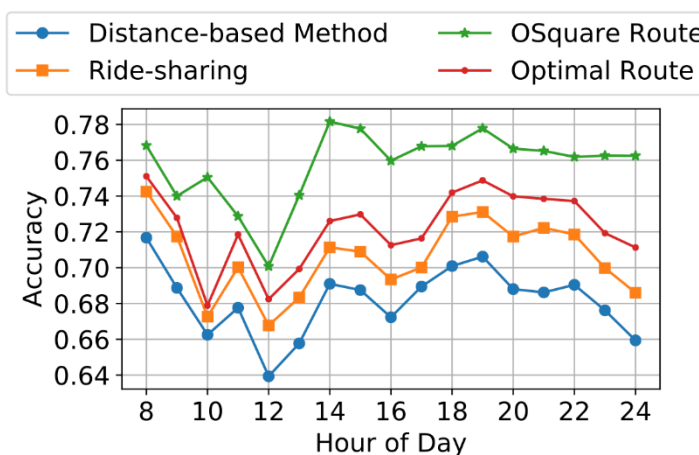
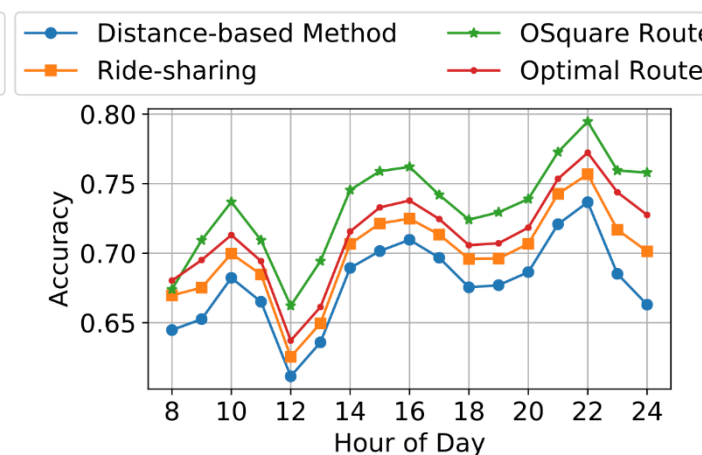


图6 (a) 订单量变化图



(b) 上海市预测准确率变化图



(c) 成都市预测准确率变化图

有趣的发现

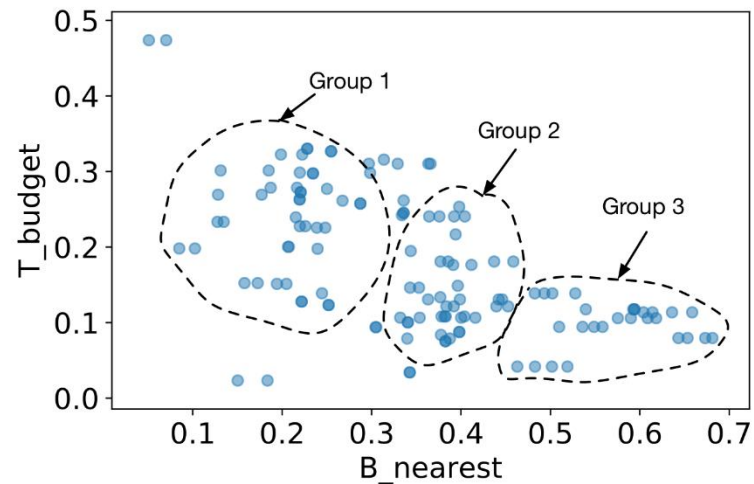
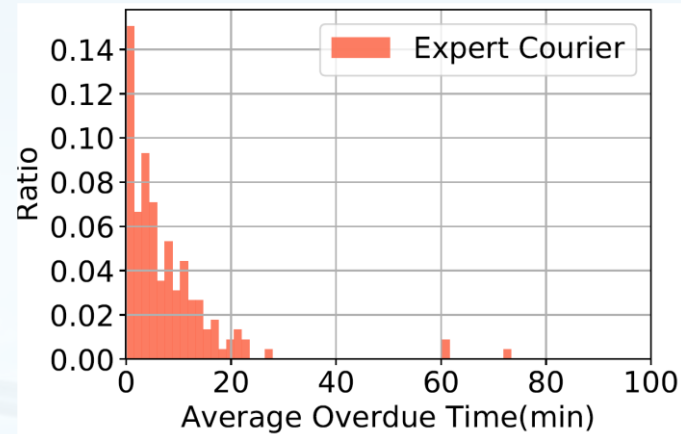
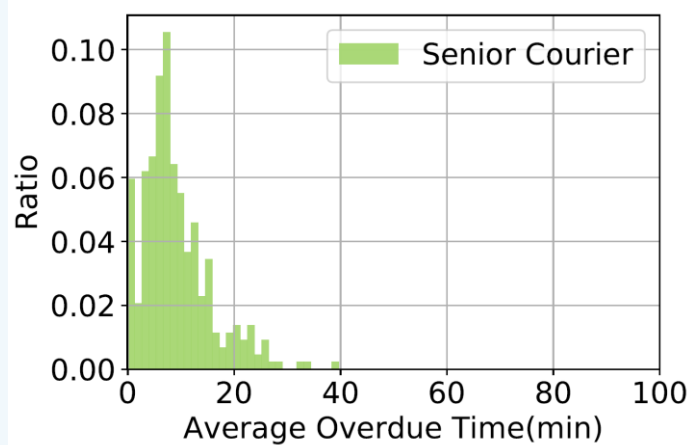
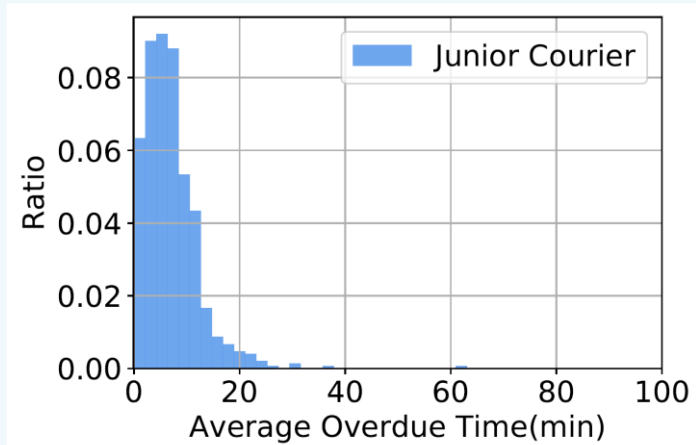


Table 5. Sample statistics

Group	Working Time	Order Amount	Overdue Rate
Junior Courier	<16 days (5% quantile)	499.0	2.6%
Senior Courier	235 days (50% quantile)	1311.9	0.5%
Expert Courier	> 827 days (95% quantile)	1872.9	0.4%

- 新手配送员倾向于优先配送临近超时的，专家配送员优先选近的
- 新手送的更少，超时更多；专家送的多(3x)，超时的少(1/5)
- 欲速则不达

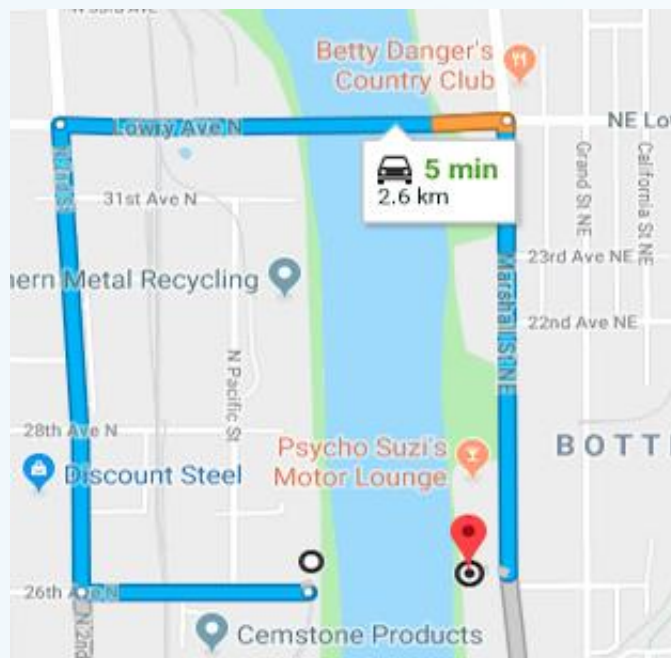
有趣的发现 (2)



- 新手配送员大多是5-10分钟超时
- 专家配送员有60-80分钟的超时订单

□ 商户位置欺诈问题

- 商户位置由商家自行注册，因此并非都是准确位置
- 商户谎报位置以获取更多用户或平台资源



100米的定位误差，造成2.6公里的绕行距离

□ 影响

- 骑手绕路，配送时间估计错误
- 图2：在位置欺诈商户中，超过36%的商户漂移距离超过500米
- 图3：100米直线误差造成765米绕行距离

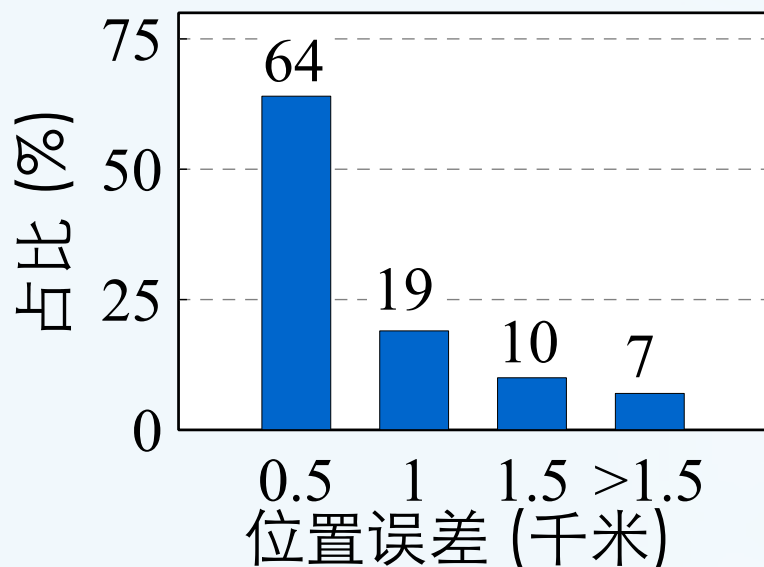


图2 位置欺诈商户的位置误差分布

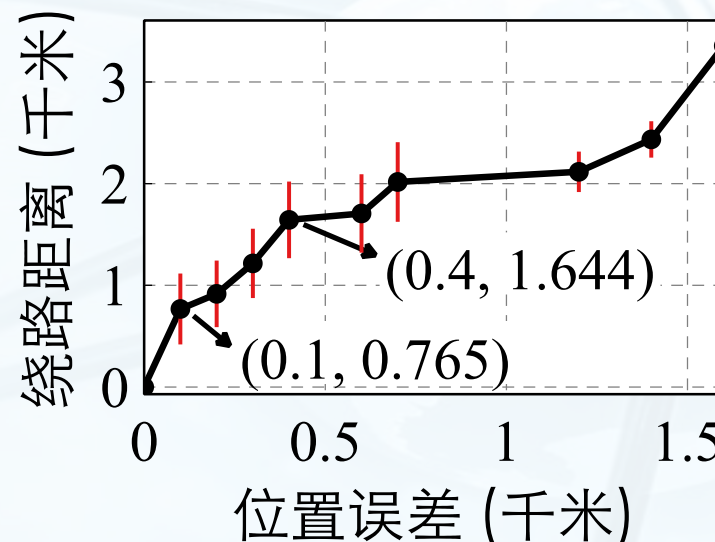


图3 位置误差与绕行距离关系



国内外研究现状

- 雇佣专业人士位置纠错
 - 成本高、无法大规模应用
- 众包解决
 - 雇佣零工或根据用户反馈纠正位置
 - 谷歌地图的反馈页面、高德淘金App等
 - 需要用户的积极参与或经济激励
 - 无法避免虚报瞒报的问题

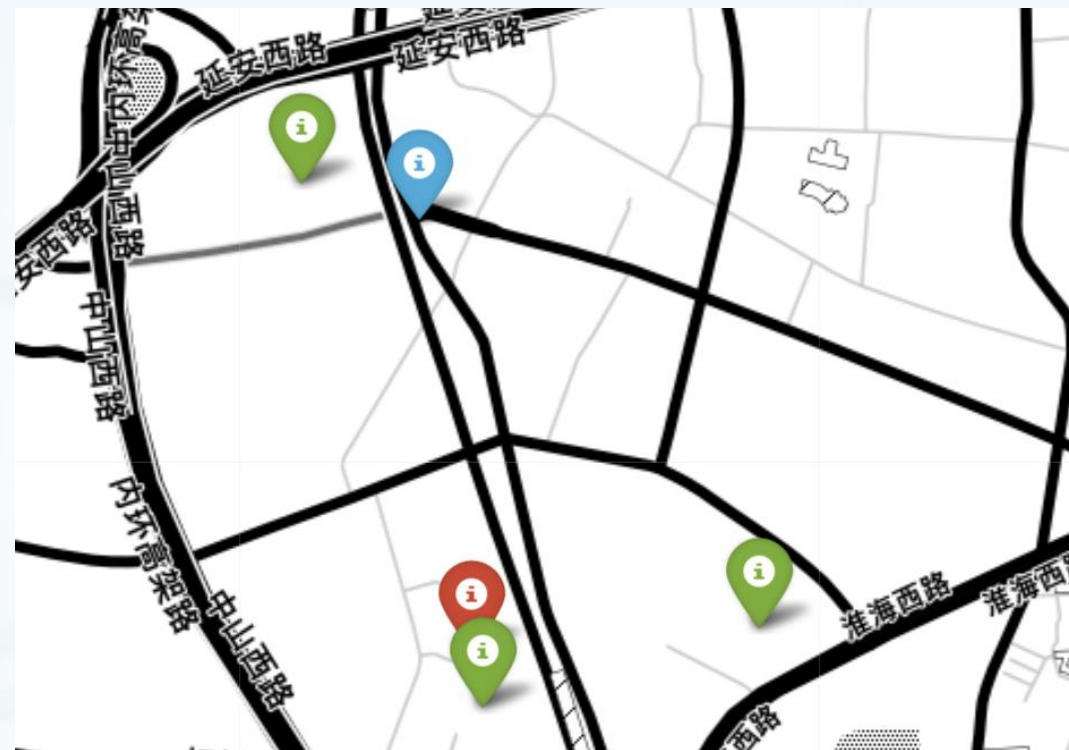
“Eager Beaver” 的问题，
难以回避

- 问题：自动识别和纠正商户位置欺诈，低成本大规模部署
- 在线即时配送平台：订单更新数据
 - 得到商户间行程时间

表1 订单更新数据示例

字段	数值
订单/骑手/商户 ID	O001/C001/M001
商户注册位置	116.418065, 39.916998
接单时间 & 位置	18-01-01 12:00:00 & GPS
到店时间 & 位置	18-01-01 12:10:00 & GPS
离店时间 & 位置	18-01-01 12:10:10 & GPS
送达时间 & 位置	18-01-01 12:25:00 & GPS

- 关键思想
 - 行程时间和行程距离的相关性
 - 行程时间估计的行程距离 vs. 商户注册位置得到的行程距离
- 检测的基础：数据量和商户密度分布
 - 58%的商户每月有300多个订单
 - 70%的商户周围有超过10个邻居



位置欺诈商户示例

ALWAES系统设计



输入：订单更新数据，商户注册位置

输出：商户估计位置；对比商户估计位置与注册位置，识别欺诈

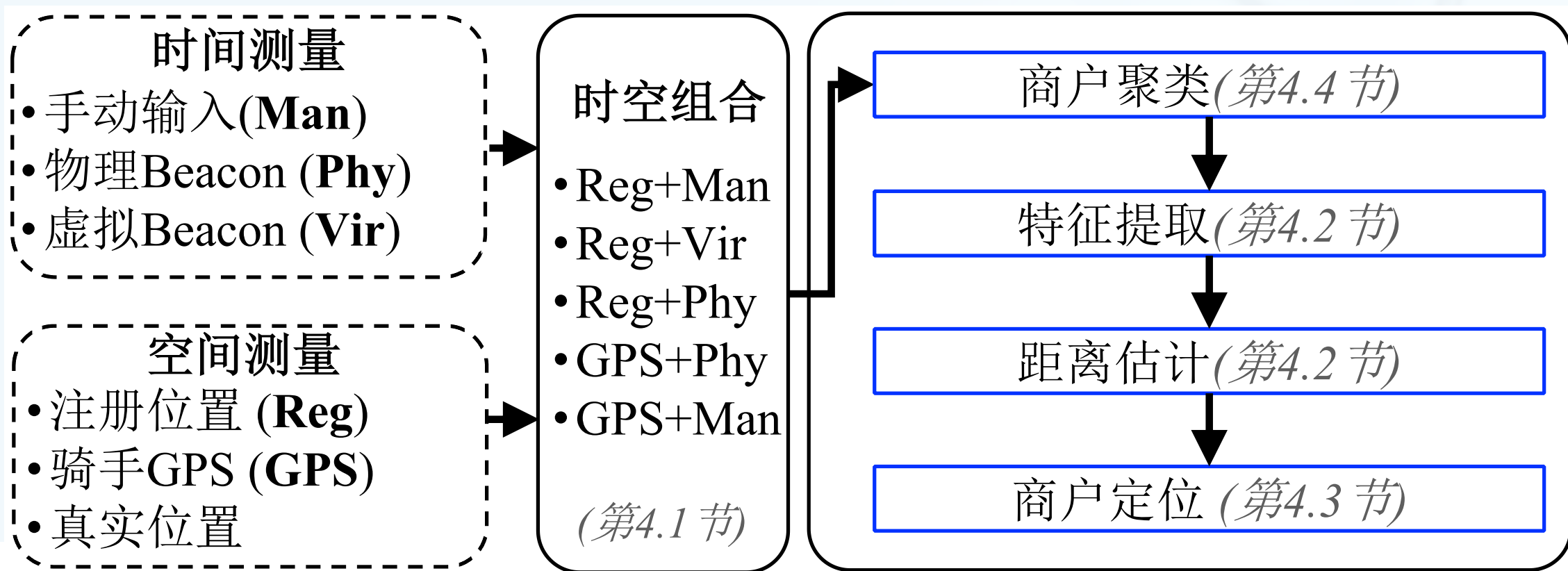


图7 ALWAES系统框架

- 骑手移动行为的不确定性
 - 骑手的移动行为受到许多因素的影响，需要考虑更复杂的模型来刻画关系

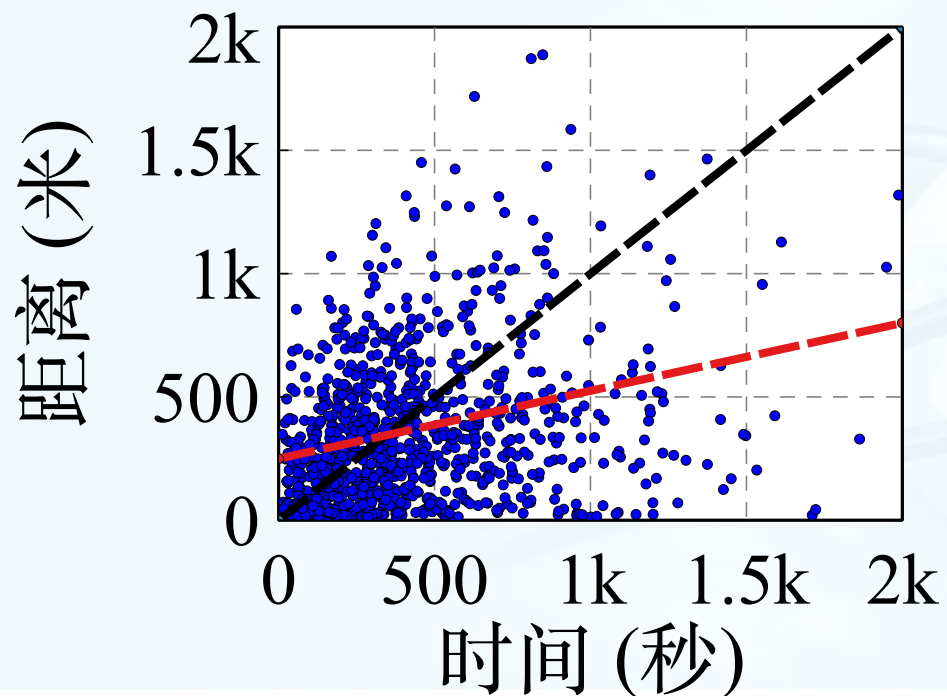
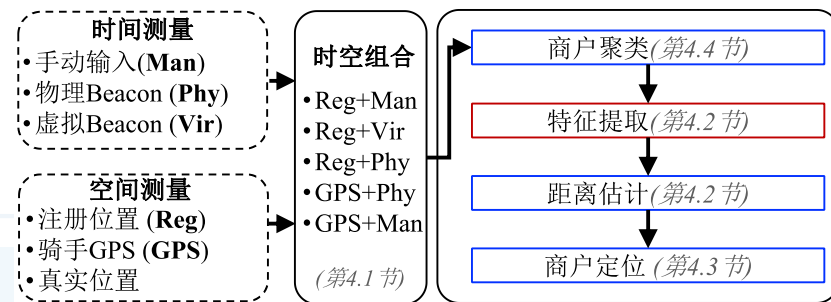


图6 行程时间与行程距离散点图

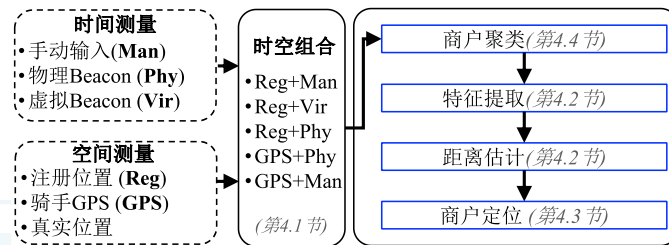
距离估算

- 输入：订单相关特征+行程时间
- 输出：商户间的行程距离
- 特征提取
 - 32个相关特征



特征类型	特征名称
时间特征	行程时间, 高峰时间段, 最长配送时间, 最短估计行程时间, 历史平均配送时长, 团队历史配送时长, 承诺送达时间, 时间片, 商户历史出餐时间, 商户历史取餐时间, 商户历史送餐时间, 网格平均出餐时间, 网格平均取餐时间, 网格平均送餐时间
空间特征	配送难度, 可能延迟比例, 商户热度, 最近 30 分钟的完成订单数量, 商场 ID, 网格单量, 网格配送难度, 网格历史速度, 网格骑手密度
个性化特征	背单量, 骑手等级, 超时单量, 骑手接单时长, 平均历史行程距离, 骑手热度, 商户历史延迟出餐率

系统评估与分析



□ 数据集

- 饿了么平台上海市2.4万名骑手在85天内的1082万份订单数据

□ 基线算法

基线算法	商户聚类	距离估计	商户定位
基于线性的算法	√	基于线性回归	√
基于轨迹的算法	√	基于GPS轨迹 (20s/打点)	√
基于MDS的算法	√	√	基于多维标度分析
无聚类模块的算法	无	√	√
ALWAES	√	√	√

系统评估与分析

□ 验证阶段总体结果

- 距离模块影响最大，其次是定位和聚类模块
- 在物理Beacon设备的帮助下，能够有比较好的表现
- 考虑到昂贵的部署成本，虚拟Beacon系统(Reg+Vir)能够达到和物理Beacon系统相似的性能，但依赖于高渗透率
- 在没有Beacon设备的条件下，手动打点+商户注册位置能够帮助商户位置纠错

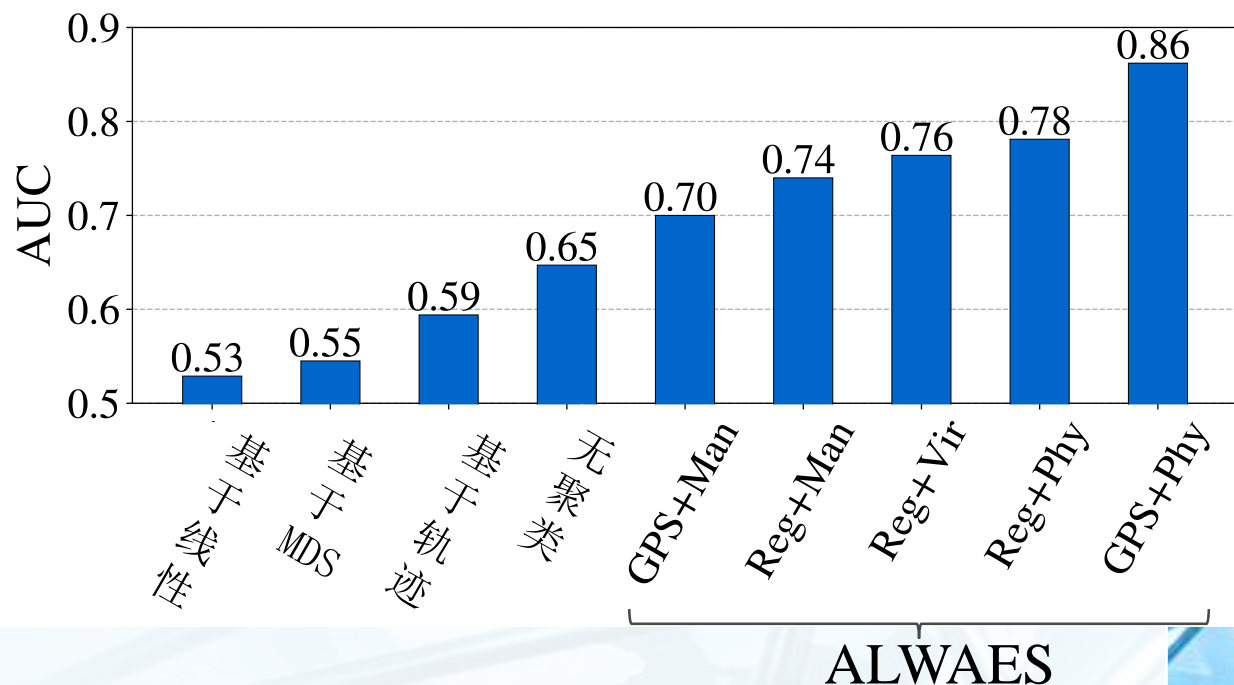


图9 总体实验结果比较

- 真实世界测试结果（一周的测试）
 - 发现了8家平台没有发现的欺诈商户
 - 骑手配送绕行距离缩减：节省3,846小时配送时间

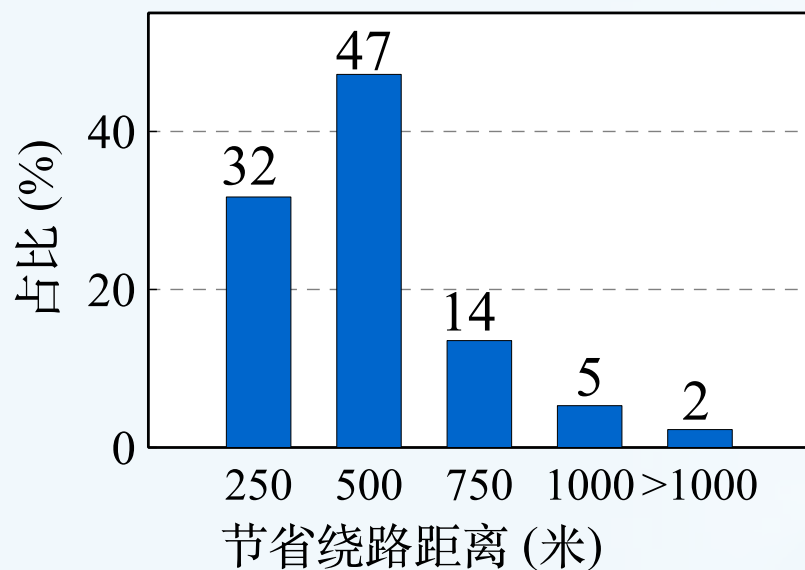
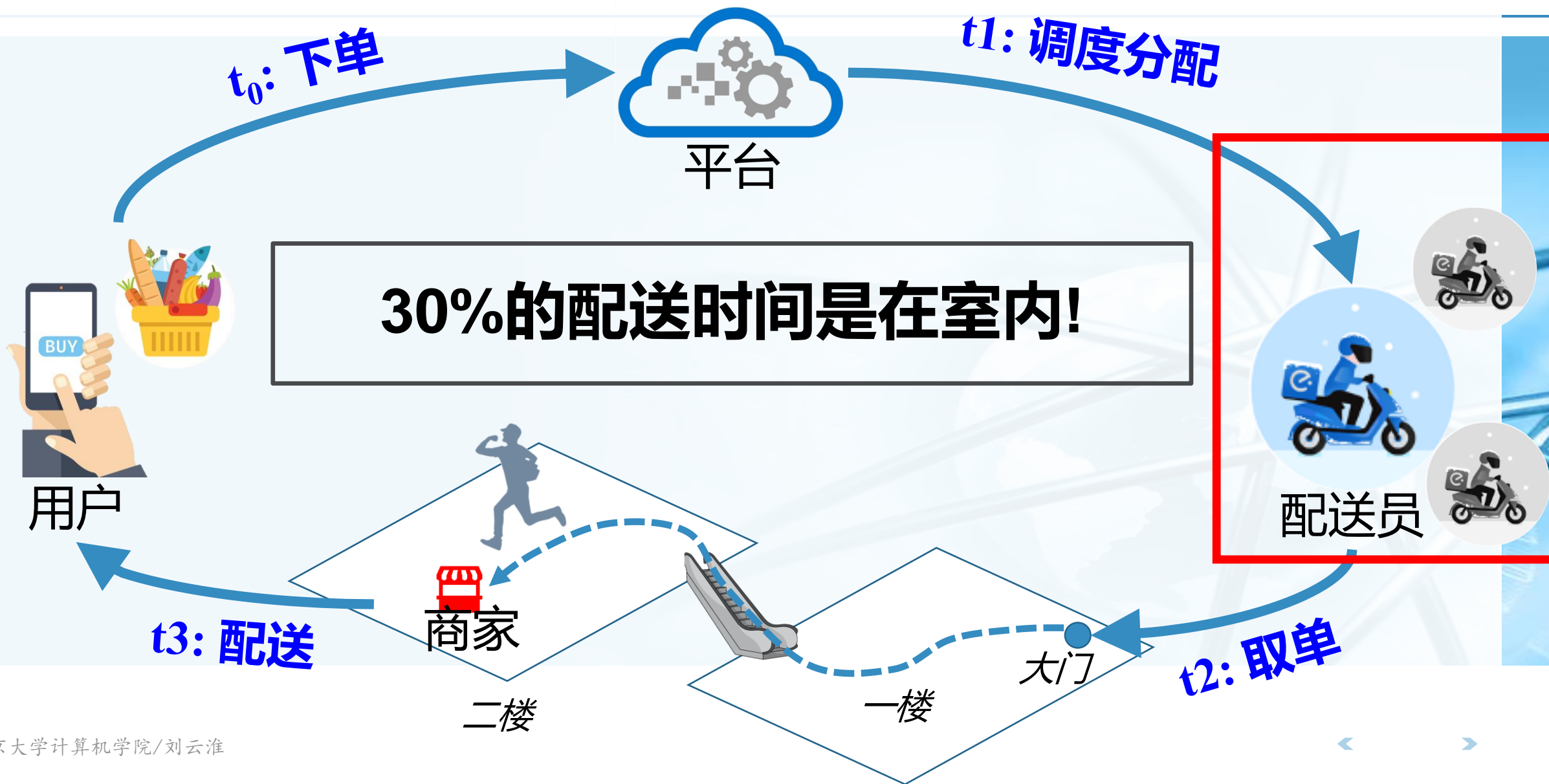


图10 节省绕路距离分布

室内定位非常重要



基于基础设施的

- Wi-Fi: [MobiCom'14,'17,'20]
- Camera: [INFOCOM'14]
- Light: [Mob]
- etc.

优势: 高精度
不足: 高成本

无基础设施的(Smartphones)

- Landmarks [MobiSys'12]
- Indoor map [UbiComp'12, MobiCom'14]

目标: 推测配送员的室内位置

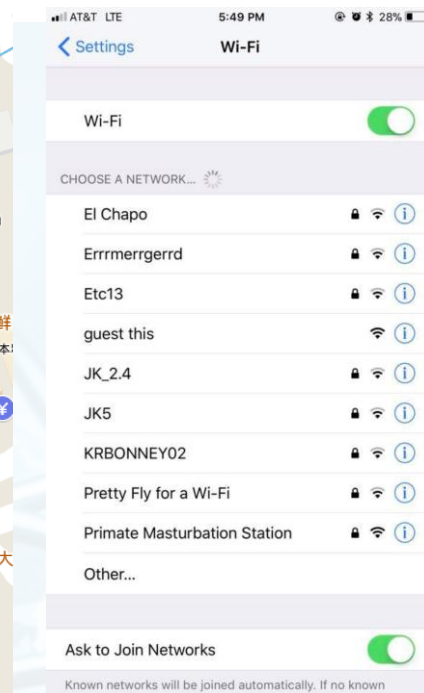
- 应用场景特定的解决方案
- 不需要额外的基础设施
- 无额外的人工

在我们的平台中很难实施

- 250万室内商家
- 85 万配送员
- 221个城市

技术挑战

- 室内环境，GPS漂移
- 商场通常多层楼
- 高动态环境
- 多种品牌的手机
(67个品牌, 700+ 型号)
- 各种WiFi信号，多达几十个



Bluetooth 作为技术路线

- BLE, 超低功耗蓝牙设备
 - 短距离通讯(~10m)
 - 超低功耗 (3% 只占手机设备3%的能耗)
- BLE beacon 协议
 - 无连接广播协议
 - 发送id信息
- BLE 设备
 - 电池供电
 - 持续广播
 - 基于未知的服务



我们的设计: aBeacon 系统

• 系统架构

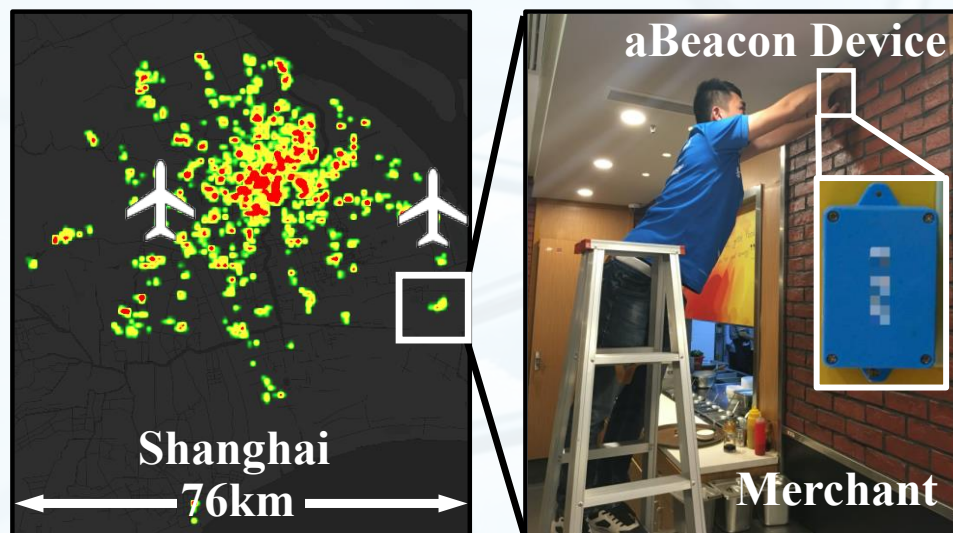
- 商家布设蓝牙设备
- 配送员手机蓝牙

• 目标

- 低成本
- 长时间运转
- 高可靠性
- 高产出

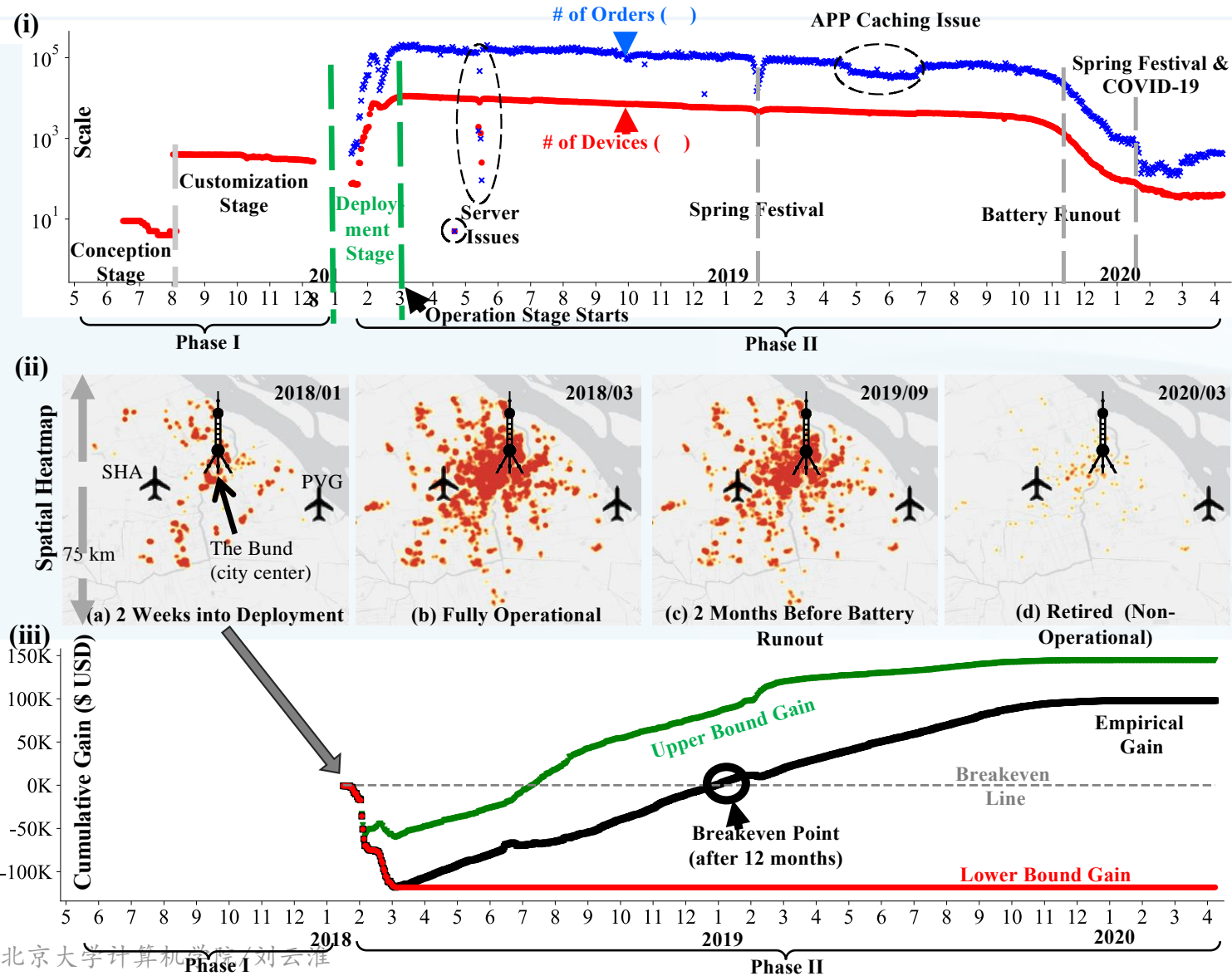
• 实施

- 在上海布设12,109商家
- 服务109,378配送员
- 共服务6400万订单
- 服务超过730万用户



在上海的12,109商家

系统概况



aBeacon 时间线

3个阶段

- (1) 构思 (可行性验证) ;
- (2) 定制 (A/B test) ;
- (3) 部署与运行;

时间维度:

- 部署后持续缓慢减少;
- 节日/技术故障/新冠 影响;

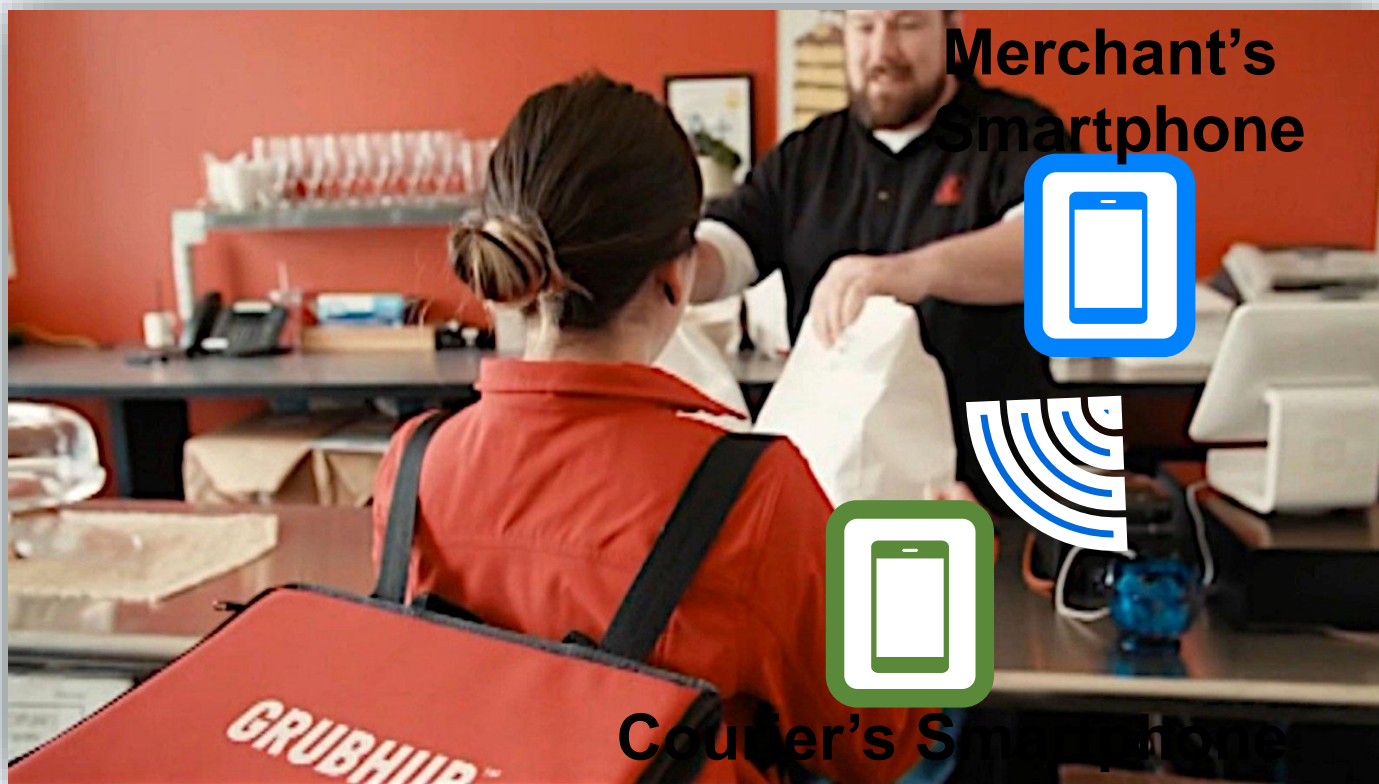
空间维度:

- 覆盖几乎所有商场和半数平台商户;
- 电池耗尽前, 保持一定范围的空间覆盖;

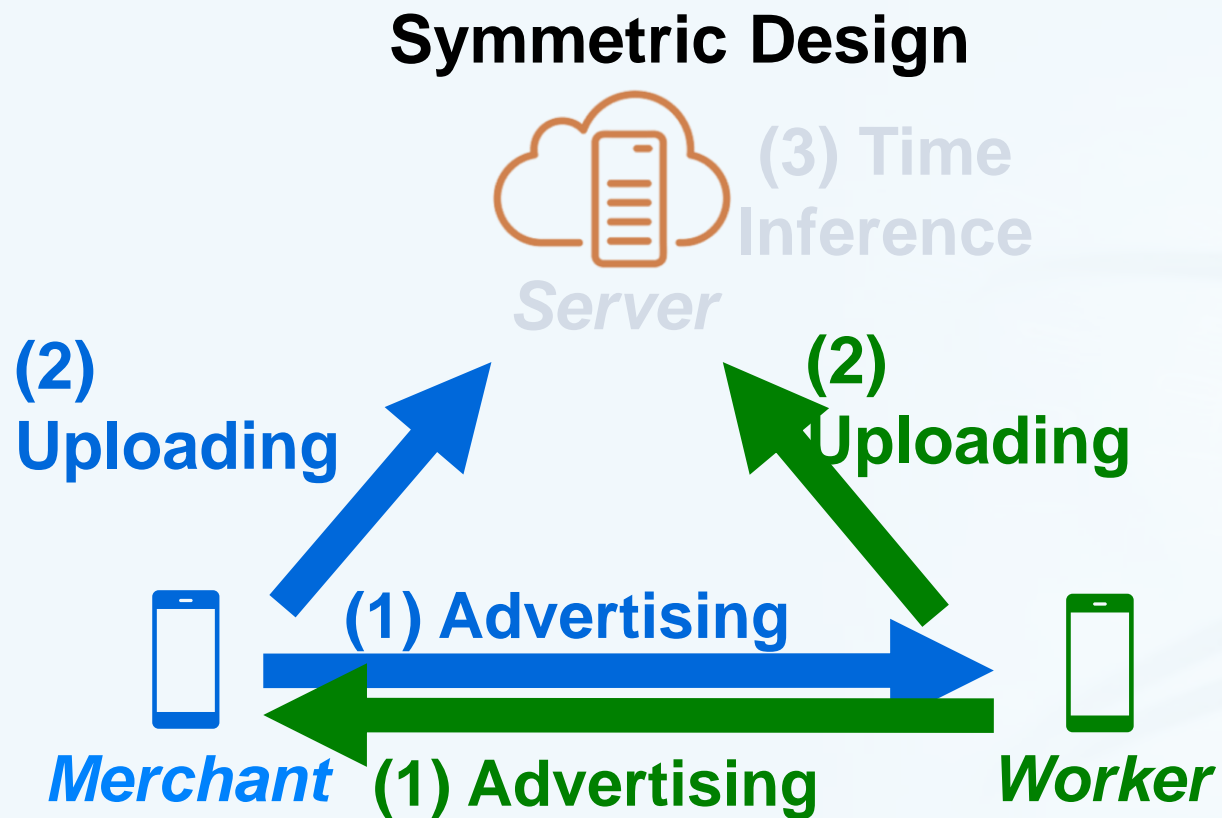
12月的时候收支平衡
(*基于理论计算)

全国范围内的室内抵达检测系统

观察：绝大多数的商户，都会使用手机作为应用终端



Virtual Beacon System 设计



商户手机广播信息

- 尽可能减少商户的负担
- 与实体Beacon系统向下兼容

能耗

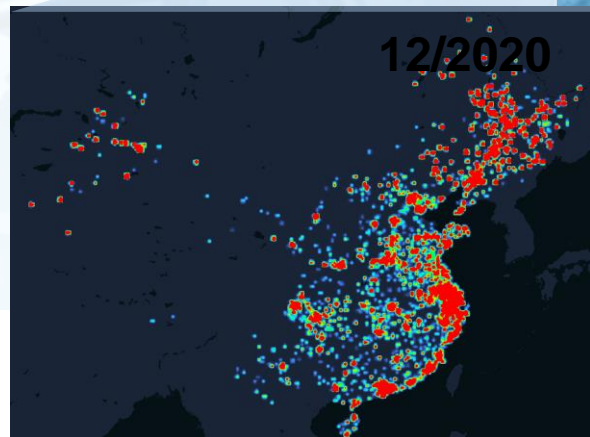
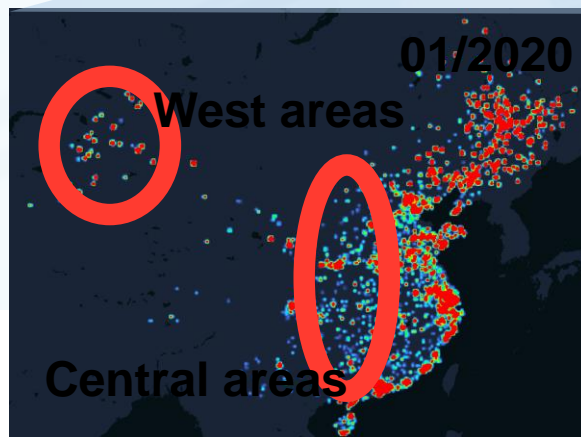
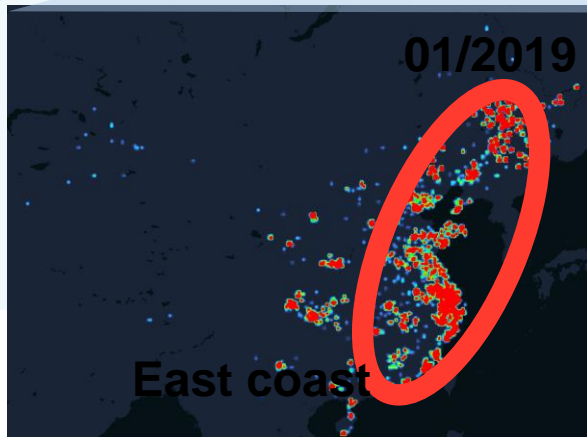
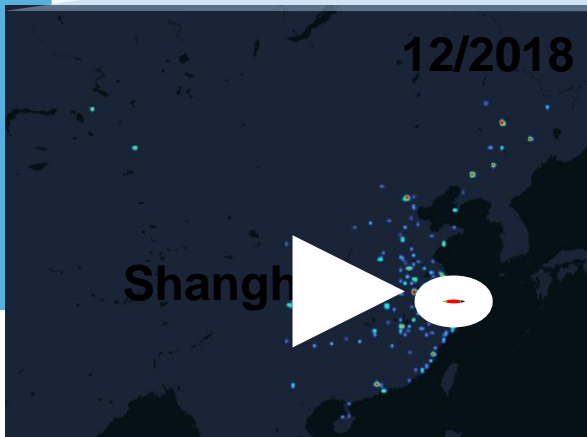
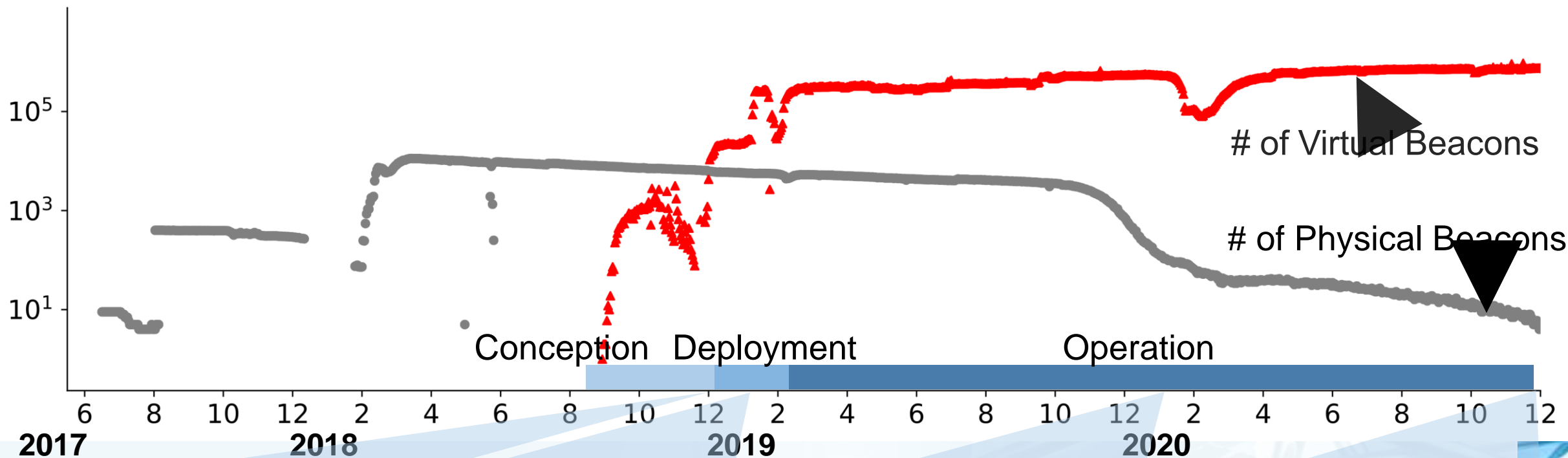
- 低频率

隐私保护

- 动态ID
- 可随时关闭

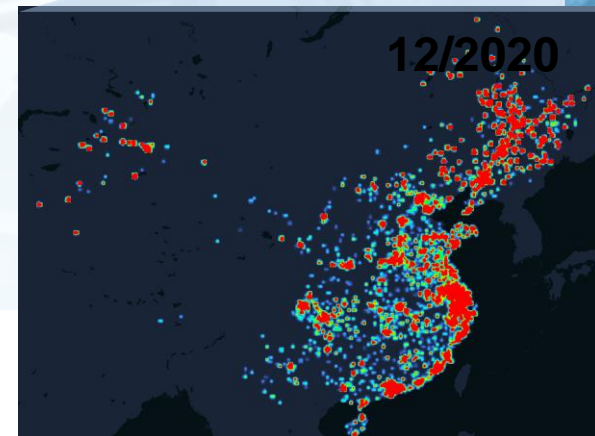
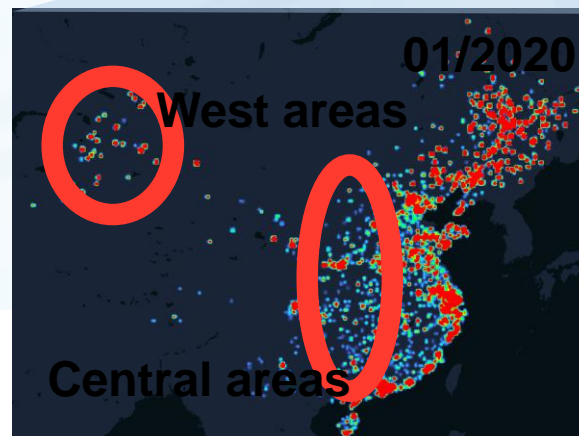
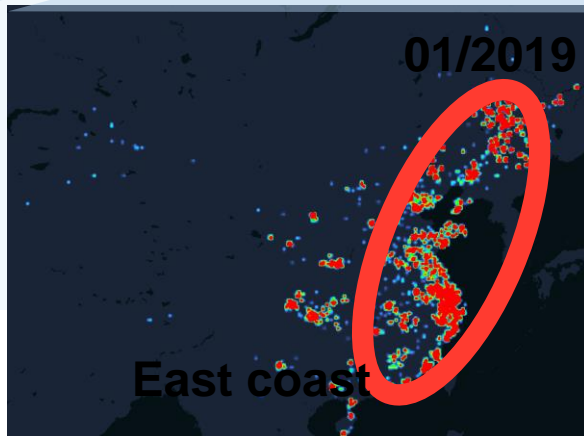
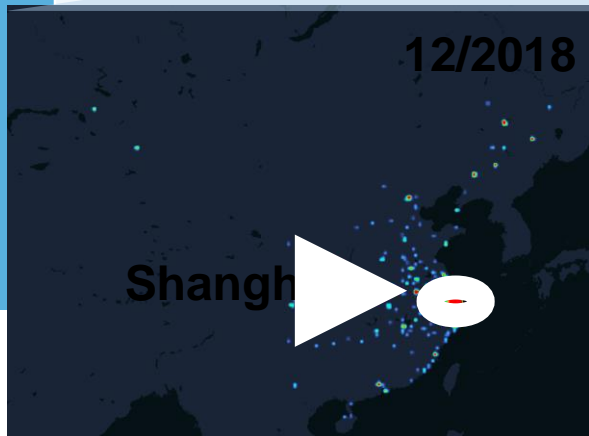
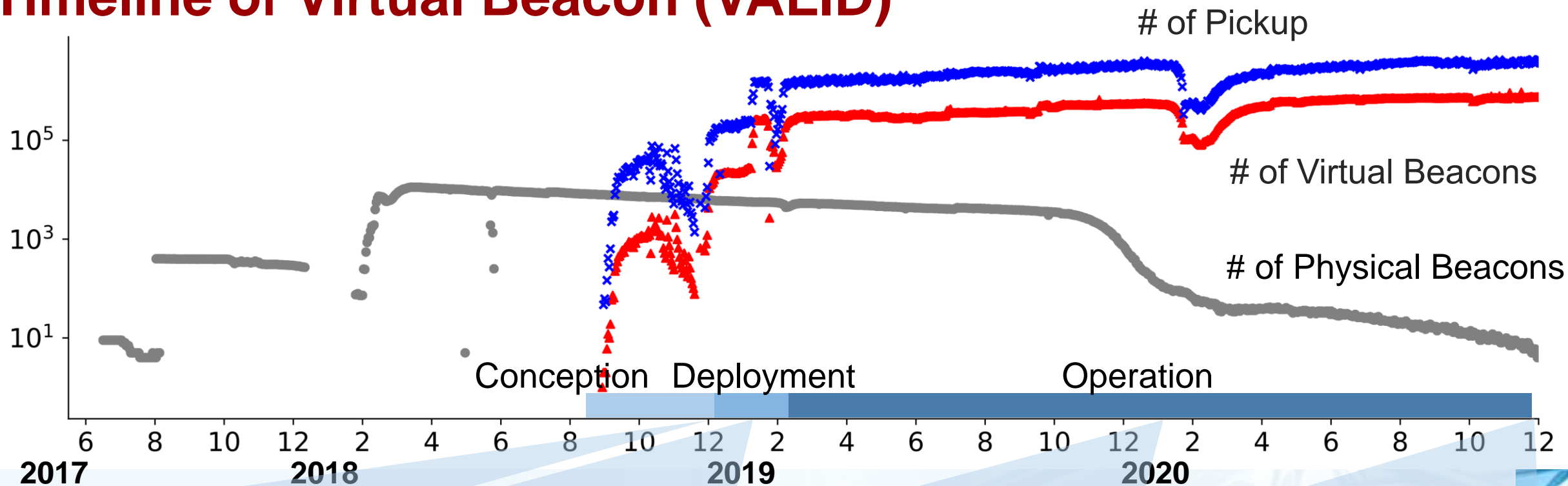
Timeline of Virtual Beacon (VALID)

数据科学「Python」



Timeline of Virtual Beacon (VALID)

数据科学「Python」进阶

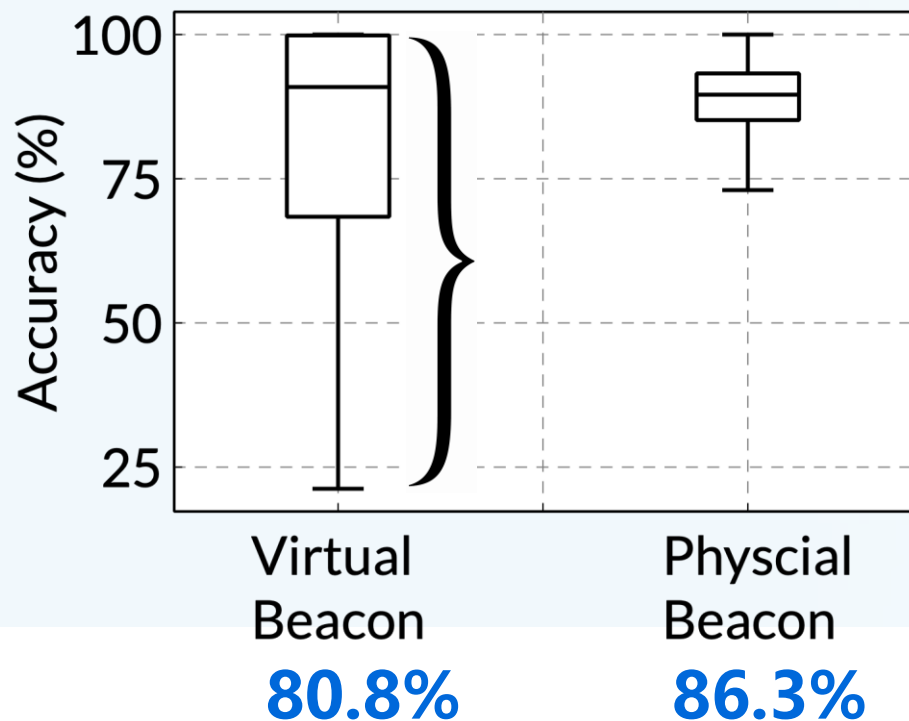


Evaluation (1/5): Reliability

Is VALID as accurate as Physical Beacon?

Setting:

- Ground Truth: Accounting data
- Baseline: Physical Beacon



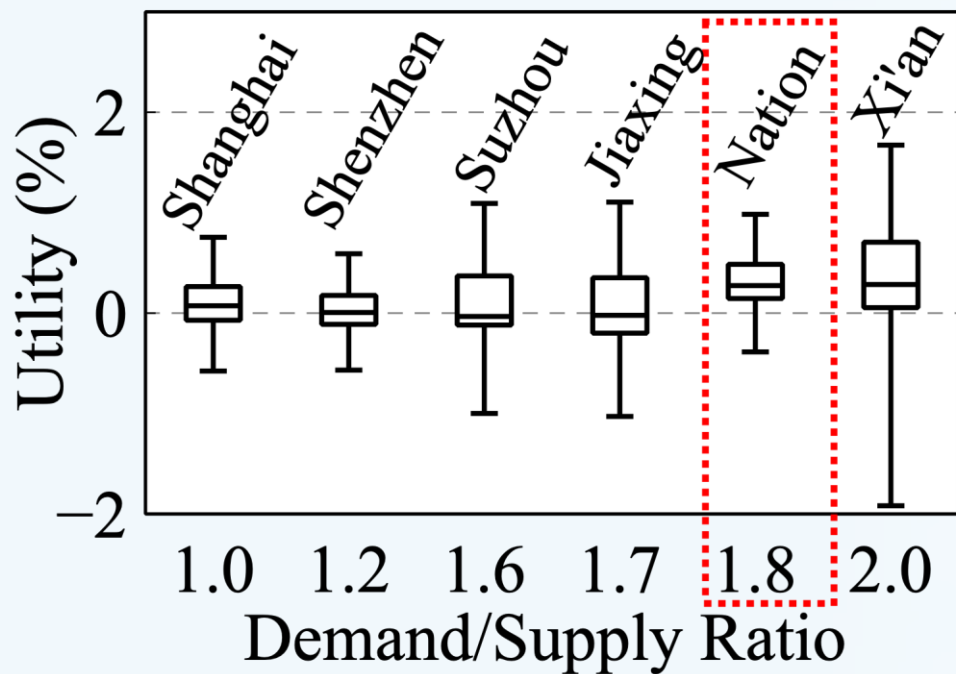
Smartphone Diversity

Merchant (Market Share) \ Worker (Market Share)	Apple(14%)	HUAWEI(15%)	Xiaomi(6%)	vivo(15%)	Samsung(1%)	Others(49%)
Apple(29%)	37%	80%	89%	77%	79%	86%
HUAWEI(20%)	38%	78%	87%	76%	79%	85%
Xiaomi(2%)	38%	72%	77%	75%	70%	81%
vivo(33%)	37%	77%	79%	76%	77%	84%
Samsung(1%)	39%	81%	83%	78%	82%	87%
Others(15%)	37%	78%	80%	77%	79%	85%

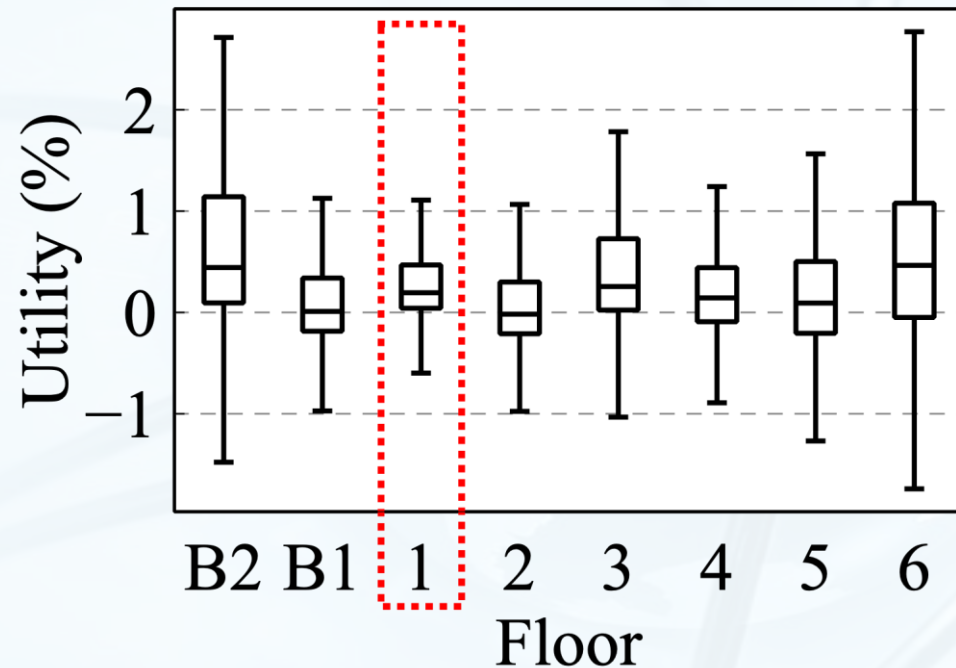
Evaluation (2/5): Utility

The empirical cumulative benefit 2021/01: \$7.9 million
— *Results from nationwide A/B test*

Impact of Demand/Supply Ratio



Impact of of Building Floors



Higher ratio leads to a higher utility **Higher floors and basements lead to higher utility**

可行性验证核心成果：两个结论

› 超大规模群智感知系统中，激励机制是**不存在的**

不同个体对于激励机制的承受力不同

由于“eager beaver”（热情海狸现象），不存在可达到纳什均衡状态的激励机制

只有不依赖于激励机制的群智感知系统，才能有效运行

› 多方受益的群智感知系统，**存在并能长期运行**

平台减少超时损益

骑手减少超时罚款

商户减少超时罚款

用户减少订单超时

- Yi Ding, Ling Liu, Yu Yang, **Yunhuai Liu**, Desheng Zhang, Tian He: From Conception to Retirement: a Lifetime Story of a 3-Year-Old Wireless Beacon System in the Wild. **NSDI 2021**: 859-872
- Yi Ding, Yu Yang, Wenchao Jiang, **Yunhuai Liu**, Tian He, Desheng Zhang: Nationwide deployment and operation of a virtual arrival detection system in the wild. **SIGCOMM 2021**: 705-717

大数据的思维变革

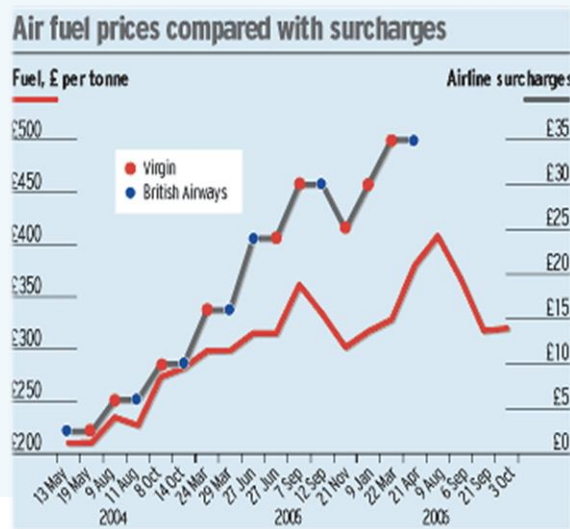
大数据之思维的变革

一旦完成了数据本身的目的之后，数据就没有用处了吗？

› 在飞机降落之后，票价数据就没用了吗？

› 一个检索命令完成之后，检索数据就没用了吗？

› 用户完成购物后，订单数据就没用了吗？



不会做饭的裁缝不是个好司机

发现·好货 定制我的偏好



NBA球衣詹姆斯韦

根据您"浏览"过的"篮球服"推荐

588 收藏



收藏 不喜欢 我要自定义

100%地轻万复学潮

根据您"浏览"过的"拖鞋"推荐

964 收藏



NBA 可调节球队

根据您"浏览"过的"篮球球迷用品"推荐

80 收藏

数据资源与传统资源（石油）的区别

› 越用越多

你有、我有、大家有

› 价值的增加通过**横向扩张**

单一数据资源用处有限

不同来源、不同类型的数据增加价值

› 随着时间流逝，价值**降低**

› 应用**松耦合**

› 越用越少

› 价值的增加通过纵向**拉长**产业链

石油烧掉最浪费

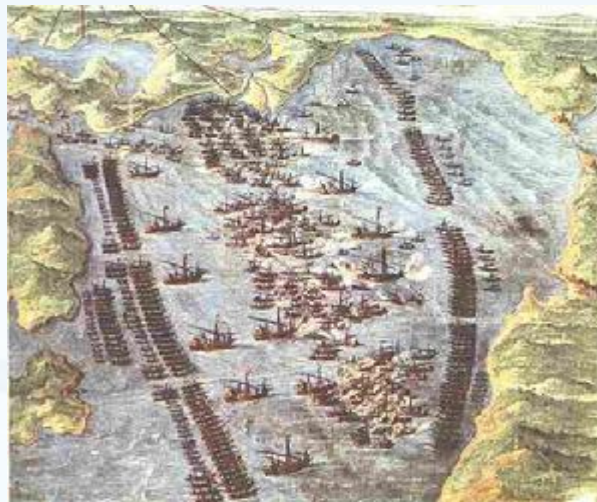
做成塑料、纤维、合成革、纺织服装

› 随着时间流逝，价值**越来越高**

› 应用**紧耦合**

勒班陀海战

- › 发生在1571年10月7日
- › 欧洲基督教联军：13000水手，28000步兵，127条战船
- › 奥斯曼土耳其军队：13000水手，34000步兵，278条战船
- › 战争结果：基督教联合舰队大胜，共击毁土耳其舰队舰船113艘，俘获117艘，缴获火炮274门以及无数金银财富和细软，击毙土军将士3万人，俘虏8000人，使土耳其舰队几乎全军覆没。而联合舰队只损失舰船12艘，被俘1艘，死伤1.5万人。
- › 作战部署比较严密；规模较大，场面惊人；火力战、冷兵器战、接舷战并用，战斗异常激烈、残酷，



勒班陀海战结果的不同解读

西班牙人的解读

- 要加强接舷战的能力
- 增加水手数, 增加接舷训练
- 强化装甲
- 提高冲撞能力



英国人的解读

- 接舷很重要, 但火炮是未来
- 提高射程,
- 提高射击精度
- 提高战舰速度



英西班牙大海战

- 1588年7月-10月
- 西班牙130艘各类战舰，30000水手和士兵，1124门火炮
- 英国197艘战舰，10000人，200余门火炮
- 战争持续了3个多月，集中战斗1周多
- 西班牙惨败，沉40多艘战舰，阵亡5000余人
- 英国无一沉船
- 英国从此走上了“日不落帝国”道路
- 西班牙从此开始衰落



三块钢板的故事

- › 二战后期，美军对德国和日本法西斯大规模战略轰炸，每天都有成千架轰炸机呼啸而去，返回时往往损失惨重。美国空军对此十分头疼：如果要降低损失，就要往飞机上焊防弹钢板；但如果整个飞机都焊上钢板，速度航程载弹量什么都要受影响。
- › 怎么办？空军请来数学家亚伯拉罕·沃尔德。沃尔德的方法十分简单。他把统计表发给地勤技师，让他们把飞机上弹洞的位置报上来，然后自己铺开一张大白纸，画出飞机的轮廓，再把那些小窟窿一个个添上去。画完之后大家一看，飞机浑身上下都是窟窿，特别是机翼，发动机。只有飞行员座舱、尾翼、机箱几个地方几乎是空白。



› 于是解决方法看起来很简单了

科学发展的四个范式

- 实验范式
 - 神农尝百草、
- 理论范式
 - 牛顿、爱因斯坦
- 仿真范式
 - 蒙特卡洛仿真法、三体
- 数据范式
 - 数据直接说话



神农尝百草，中医在这里



$$F_1 = F_2 = G \frac{m_1 \times m_2}{r^2}$$



为什么选Python?



› 代码短小精悍，干净整洁

没有变量声明，不需要花括号begin/end，也没有分号，比java短80%，比C短98%

› 解释执行，上手就玩，编程小白福音

不用焚香沐浴安装GB级别的开发环境compile/build，可以随问秒答，边玩边改

› “包装内附带电池”

自带大量运行库，网络、数据库、图形图像、GUI、压缩加密一应俱全，几行代码建网站

› 功能无比强大，开发左右逢源，最酷的网络应用都是用它

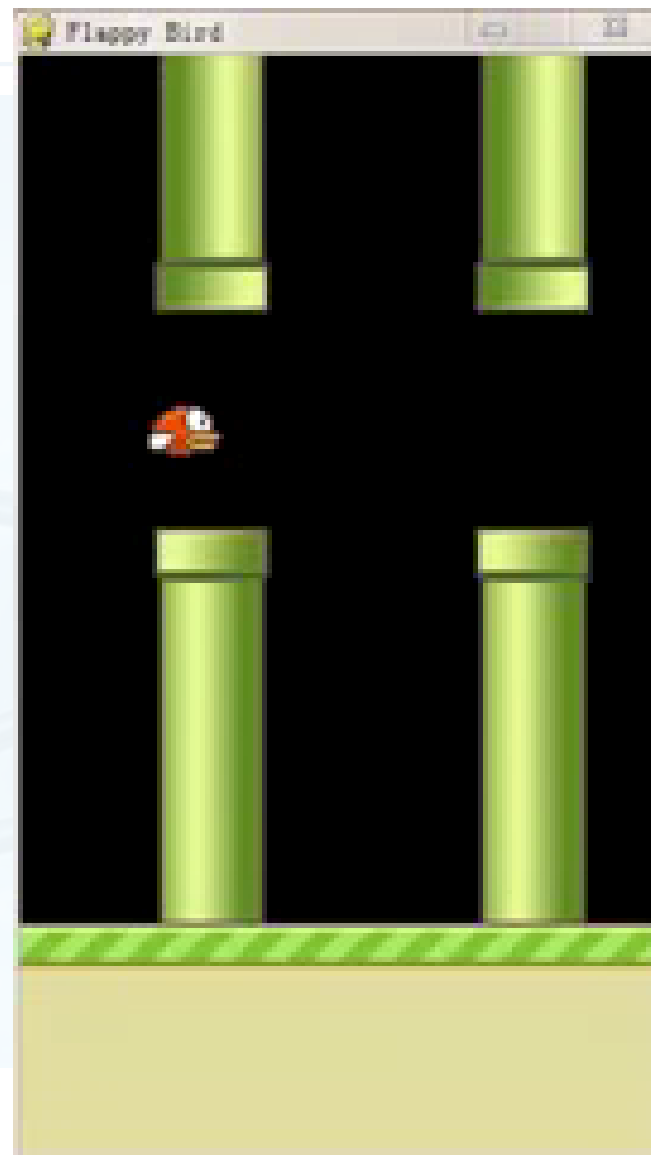
Google/Youtube/Instagram/豆瓣……，NASA也用它哦

› 搞大数据和AI的人们也爱它

有各种面向大数据处理的数据模型、数值分析、机器学习、空间分析等Python工具随时恭候

Python坐稳人工智能时代的头牌语言

- › Google开源的AI系统Tensorflow
- › 支持Python和C++开发
- › 160行Python代码可以让AI从游戏视频中学习玩Flappybird



课程目标和考评

› 课程目标

掌握数据结构和算法的基本概念，特别是数据结构的逻辑结构和物理结构
理解每种数据结构及其相应的运算
能采用“抽象”和“自顶向下”方法分析问题，使之简化，设计算法
能分析给定数据结构和算法的效率，特别是时间效率和空间效率

› 教学方式

课堂授课、课后作业、团队大作业、网络交流

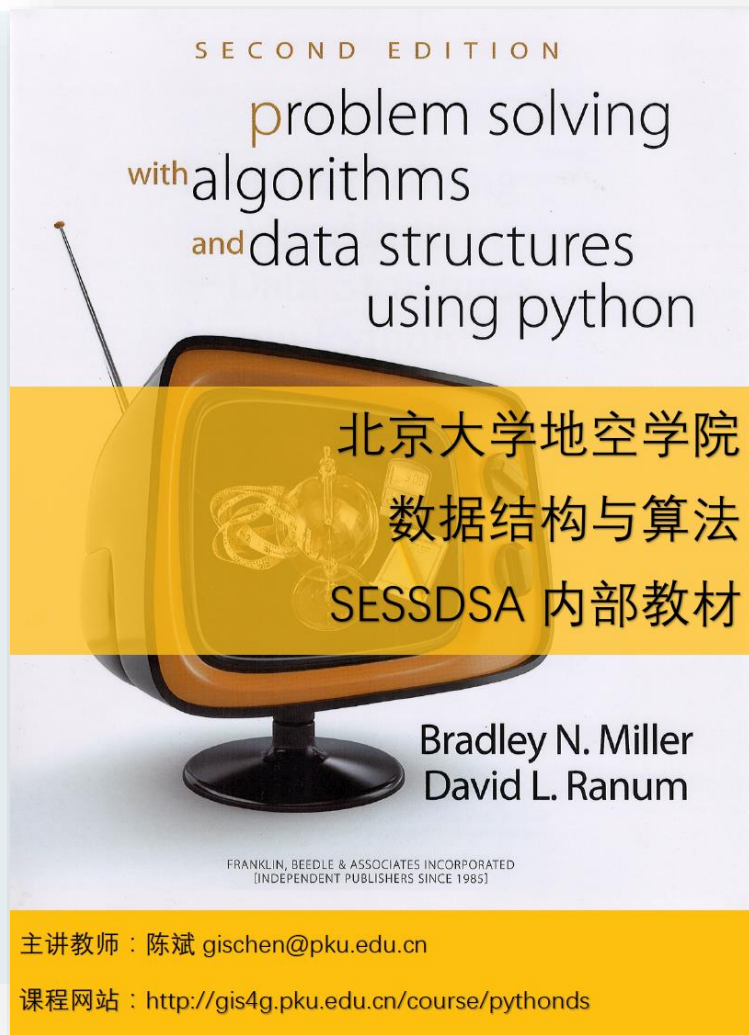
› 评分方式 (可能会微调)

作业（报告/上机之类）30%，一次大作业（团队竞赛）20%，
期末考试（机考）20%，期考考试（闭卷）30%

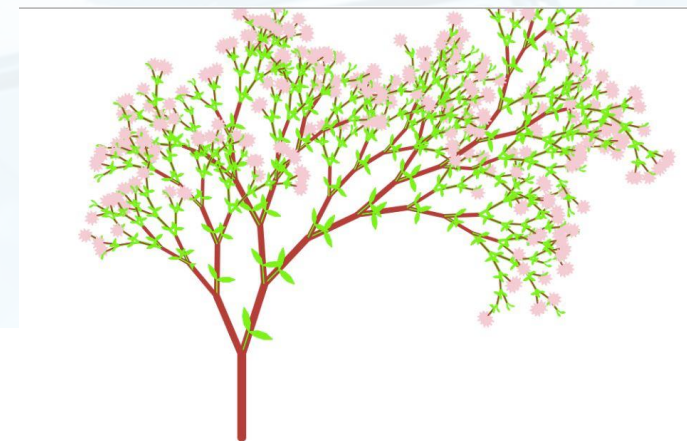
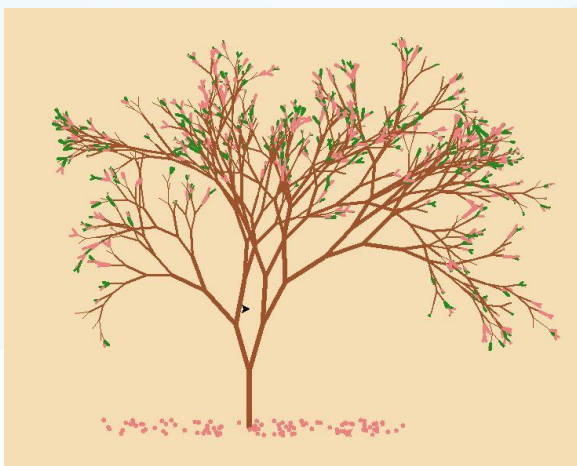
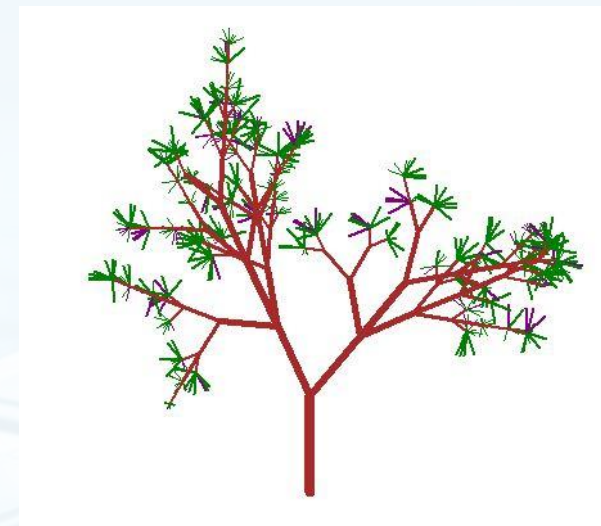
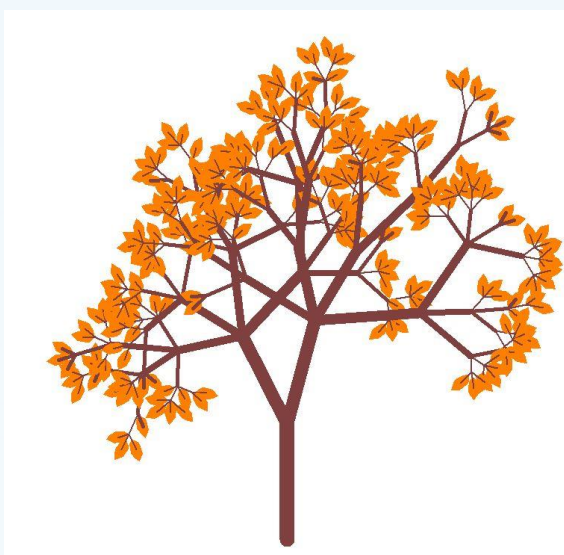
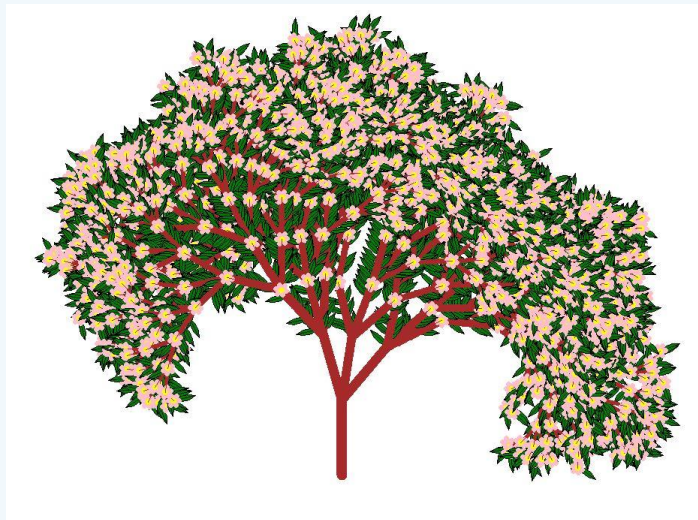
DSA'15和'16都做了什么?

	人数	平时表现	上机作业 / 报告	递归作业	额外加分	分组大作业	闭卷考试
2015	108 (3)	上机签到 随堂作业 10分	1次报告 9次作业 40分	(画二叉树)	教材翻译 竞赛场务 作业小组评议 加分 Scratch学习	黑白棋 15分	35分
2016	112	上机签到 随堂作业 10分	1次报告 6次作业 25分	递归视觉 艺术 10分	教材编辑完善 作业小组评议 加分	坦克大战 25分	30分

DSA'15: 教材众包翻译 / 助教之选优秀作业



DSA'18 : 二叉树的艺术



DSA'15 : 二叉树的艺术



DSA'15 : 黑白棋大战



```
SOUTH: GOLF >- BRAVO 39,25 (END, time: 21.925933:10.703209)
SOUTH: ALPHA <- GOLF 20,44 (END, time: 12.404031:56.17389)
SOUTH: GOLF >- ALPHA 36,28 (END, time: 29.095909:17.664121)
SOUTH: GOLF >- BRAVO 39,25 (END, time: 29.095909:17.664121)
SOUTH: ROMEO <- ALPHA 12,52 (END, time: 0.0061449999999995:41.687809)
WEST: INDIA <- KILO 27,37 (END, time: 73.501144:53.690597)
WEST: KILO <- INDIA 28,36 (END, time: 66.200688:62.920071)
SOUTH: ALPHA >- ROMEO 43,21 (END, time: 51.511866:0.00524000000001)
SOUTH: BRAVO >- ROMEO 36,28 (END, time: 23.439517:0.005974000000021)
SOUTH: ROMEO <- BRAVO 14,50 (END, time: 0.007491000000002:39.960571)
SOUTH: GOLF <- LIMA 9,55 (END, time: 18.68421:110.74275)
SOUTH: BRAVO <- GOLF 27,37 (END, time: 26.581381:84.9053)
```

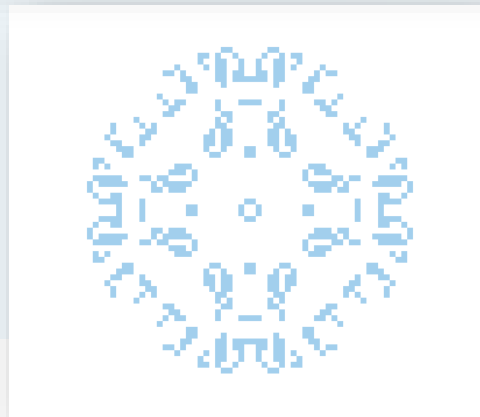
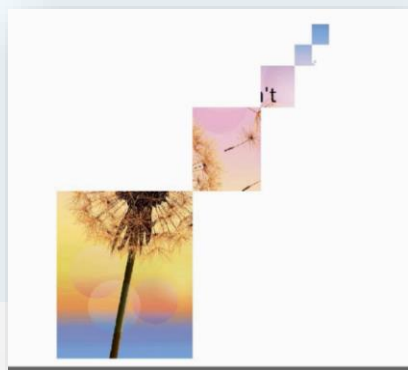
DSA'15 : 黑白棋大战



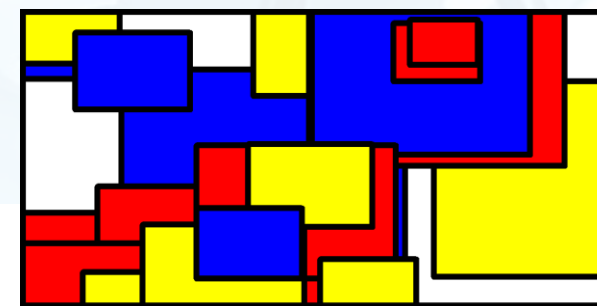
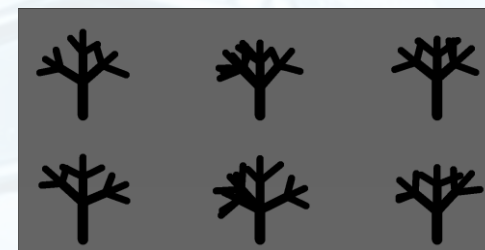
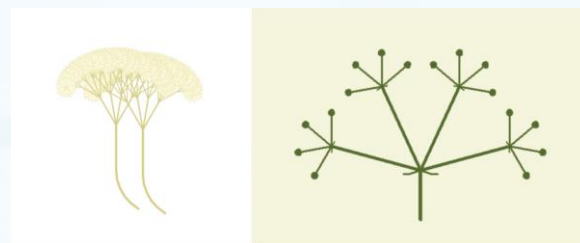
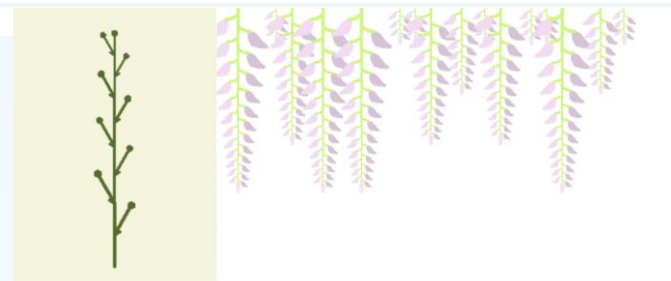
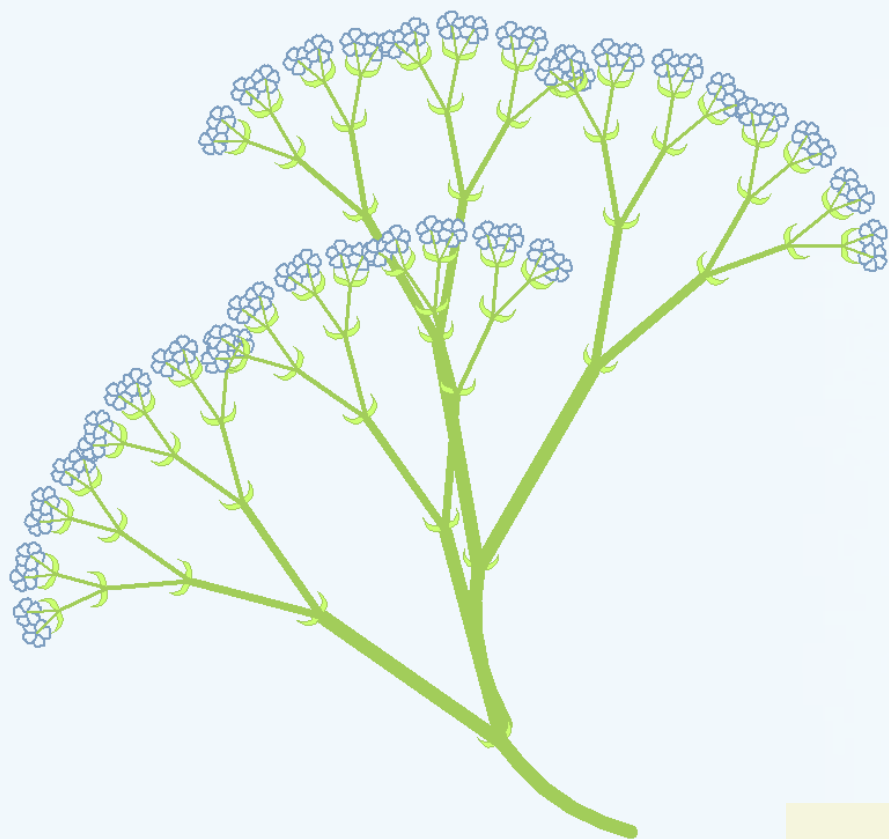
DSA'15 : 黑白棋大战



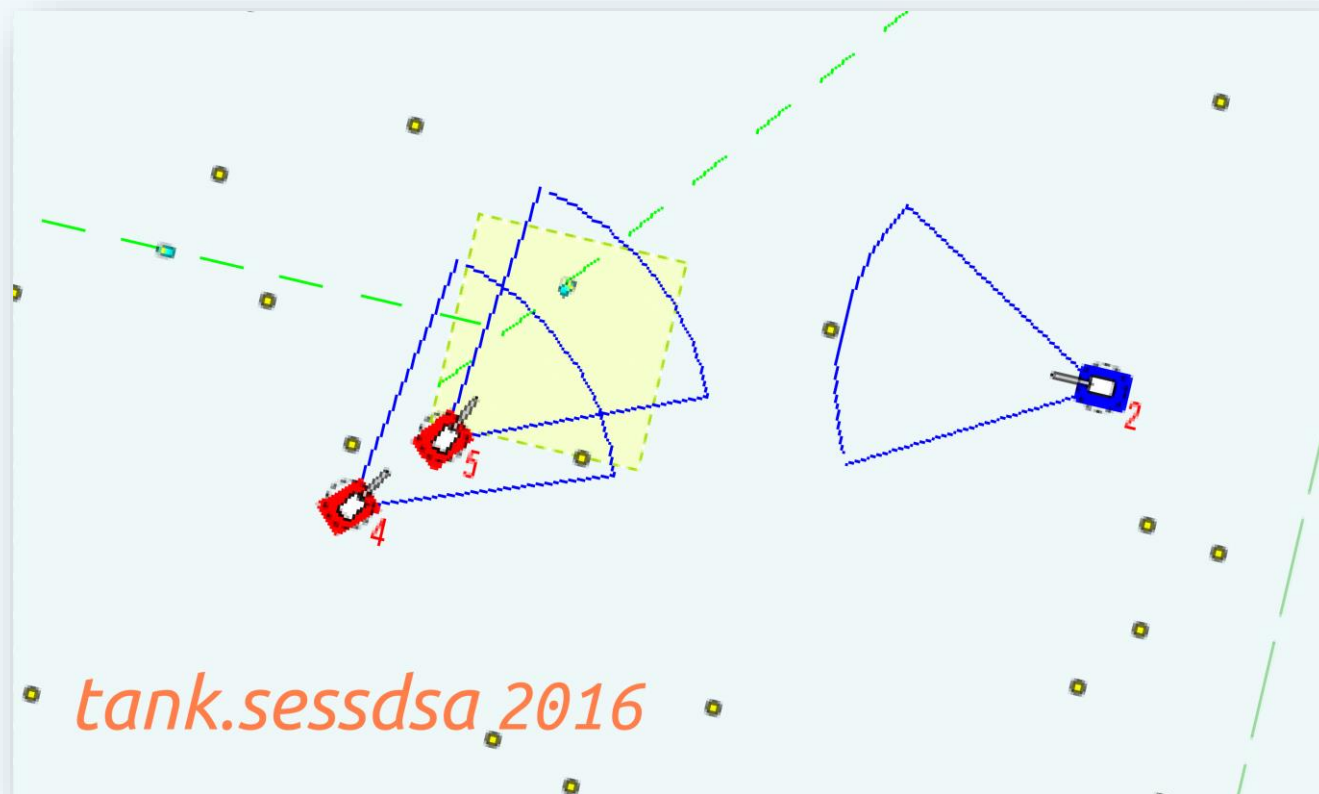
DSA' 16 : 递归视觉艺术



DSA' 16 : 递归视觉艺术



DSA' 16 : 坦克大战

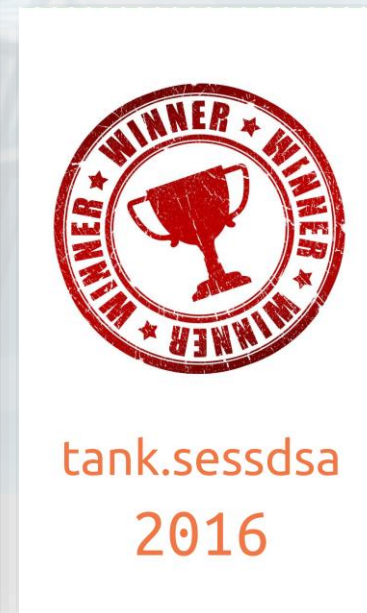
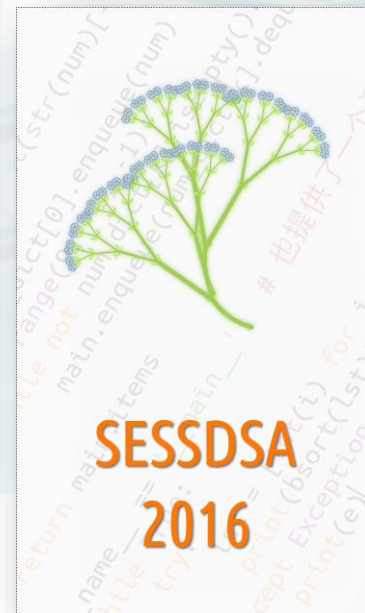


DSA' 16 : 坦克大战

数据结构与算法 (Python)



DSA' 16 : 坦克大战



联系方式与课程表

- › **微信群：2019城环数算课程**
教师：刘云淮（微信：yunhuailiu）
助教：张岩、江东哲
- › **上课地点：二教319**
- › **上课时间：**
每周二，7-8节；15:10-17:00
单周四，5-6节：13:10-14:50
- › **上机地点：计算中心7#机房**
上机时间：每周五(第三周)，11-12节



我们的教材

› Problem Solving with Algorithms and Data Structures

在线教材:

<http://interactivepython.org/runestone/static/pythonds/index.html>

› 好消息! 中文版教材已完成翻译! (by sessdsa'15&16)

› 参考资料

廖雪峰的Python教程

- <http://www.liaoxuefeng.com/wiki/0014316089557264a6b348958f449949df42a6d3a2e542c000>

数据结构与算法可视化

- <http://visualgo.net/>

› 课程网站

<http://www.yunhuai.net/DSA2021/CoursePage/DSA2021.html>

参考书

Python数据结构与算法分析

- [美] [布拉德利·米勒](#), [戴维·拉努姆](#)著, [吕能](#), [刁寿钧](#)译
数据结构与算法 (Python语言实现)

- 迈克尔·古德里奇等
数据结构与算法

- 张铭, 北京大学
算法与数据结构-C语言描述 (第3版)

- 张乃孝主编, 高等教育出版社, 2011, 6
数据结构-C语言版

- 严蔚敏等, 清华大学出版社

有用的软件和网站

python



powered

› 在浏览器里运行Python

<https://www.programiz.com/python-programming/online-compiler/>

<http://pythontutor.com/visualize.html>

<https://www.online-python.com/>



› 集成开发环境Geany

<https://www.geany.org/Download/Releases>

› 更高级的集成开发环境PyCharm

<https://www.jetbrains.com/pycharm/download/>



› 有用的地空学院陈斌老师的网站

<http://gis4g.pku.edu.cn/course/pythonds/>