

大数据与城市交通战略决策

同济大学 杨东援

2015年9月

代引言——学习“促进大数据发展行动纲要”

“大数据已成为国家重要的**基础性战略资源**，……，充分利用我国的数据规模优势，实现数据规模、质量和应用水平同步提升，发掘和释放数据资源的潜在价值，有利于更好发挥数据资源的战略作用，……，有效提升国家竞争力。”

“大数据成为**提升政府治理能力的新途径**。……，建立‘用数据说话、用数据决策、用数据管理、用数据创新’的管理机制，实现基于数据的科学决策，将推动政府管理理念和社会治理模式进步，……，逐步实现政府治理能力现代化。”

不同的
领域，
不同的
动机！

智能商务——通过大数据“沙里淘金”；
信息服务商——探寻新的服务内容与商业模式；
自动化领域——新的控制理论与方法；
信息领域——数据挖掘与信息融合；
交通运输领域——……？？？

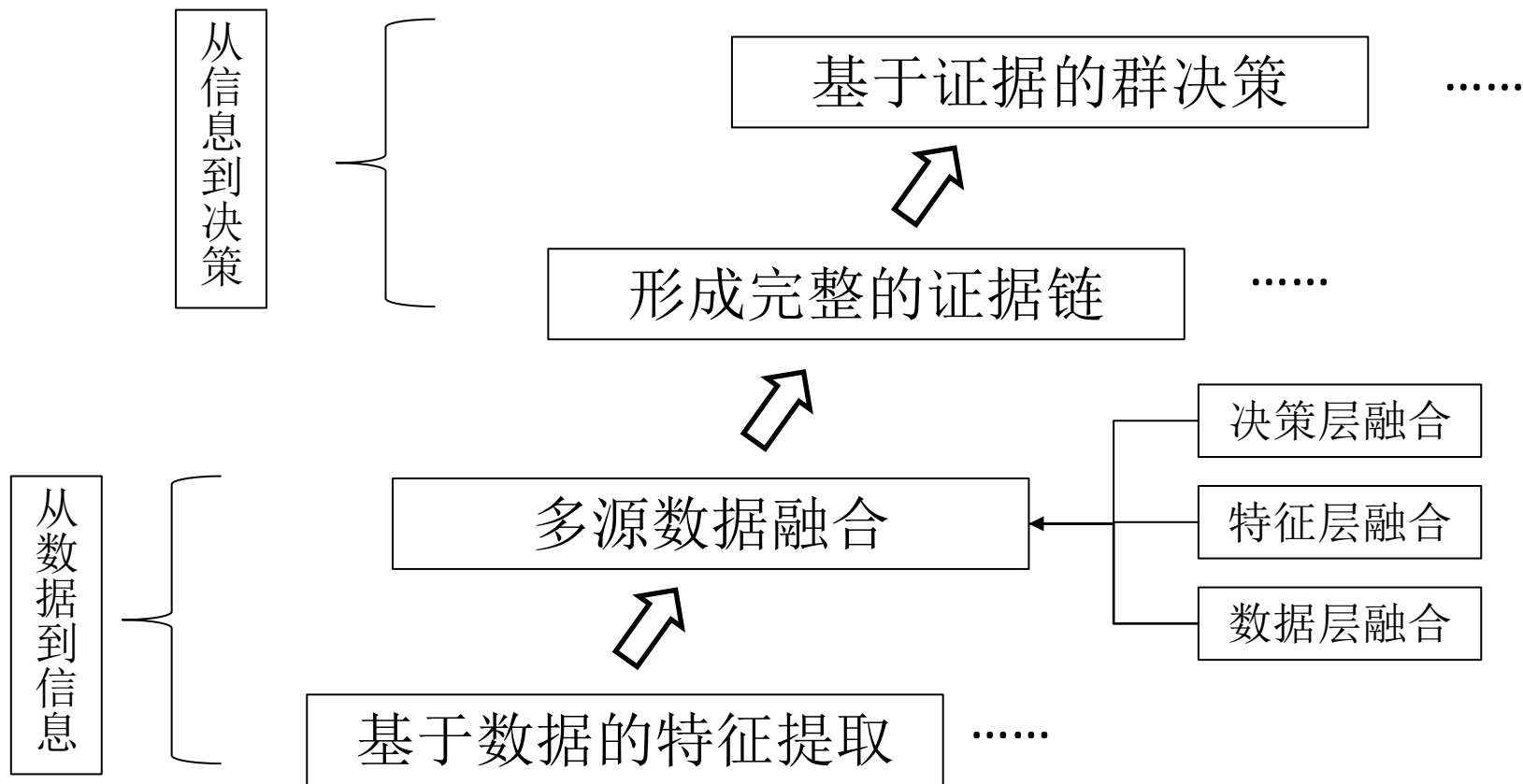
智慧交通需要一个高效思考的大脑

- 交通的复杂性首先表现为“综合”：病因是多种因素相互影响；药方需要“复方合剂”；治理需要多方协同和社会的支持。
- 交通不能如同西医简单地“治病”，而必须如同中医“医人”。因此需要一个精明的政府，实施有效的过程管理！
- “智慧交通”与“智能交通”相比，其最主要的任务区别在于要建设一个充满智慧的政府管理！
- 比采集数据更加重要的是使用信息的能力，大数据强调的是基于数据的决策判断，智慧交通同样强调的是基于信息“搭脉”基础上的系统治理！

相比大数据技术，我们需要更加重视大数据思维

- 大数据研究，方法隐藏在目的之后；
- 交通领域的大数据研究，远比“啤酒和尿布”问题复杂的多；
- 交通问题之所以复杂，就在于其综合性，利用大数据增强把脉能力，是城市的希望；
- 交通领域的大数据技术应用，既非将大数据技术塞进传统理论的框架，也非对数据挖掘、数据关联技术的简单套用；
- 交通大数据：并非简单的“量”大，而是更加全面地应用多源数据，多角度、连续、多粒度、多层次地观察研究对象，以应对具有关联性，复杂且具有自主性的演变挑战。

任重道远的“长征”——交通大数据发展的阶段性



内容提要

- 理解：大数据的“长”与“短”
- 理念：城市交通战略管理
- 技术：从“预测”走向“基于证据的决策判断”
- 行动：技术与管理的协同变革
- 没有结束的结束语

相对于传统技术，大数据提供了一种全面连续观察研究对象的手段，为城市交通分析技术带来变革的机遇，同时也不能将大数据视为万能药。

理解：大数据的“长”与“短”

城市交通:新的技术环境带来分析资源的拓展

单次计量 观测数据



轨交AFC数据



出租车营运数据



定点检测器数据



停车管理数据



个体连续 追踪数据



移动通信数据



IC卡数据



GPS数据



车牌照数据



其它基础数据



住宅价格数据



人口普查数据



城市规划管理数据



POI数据



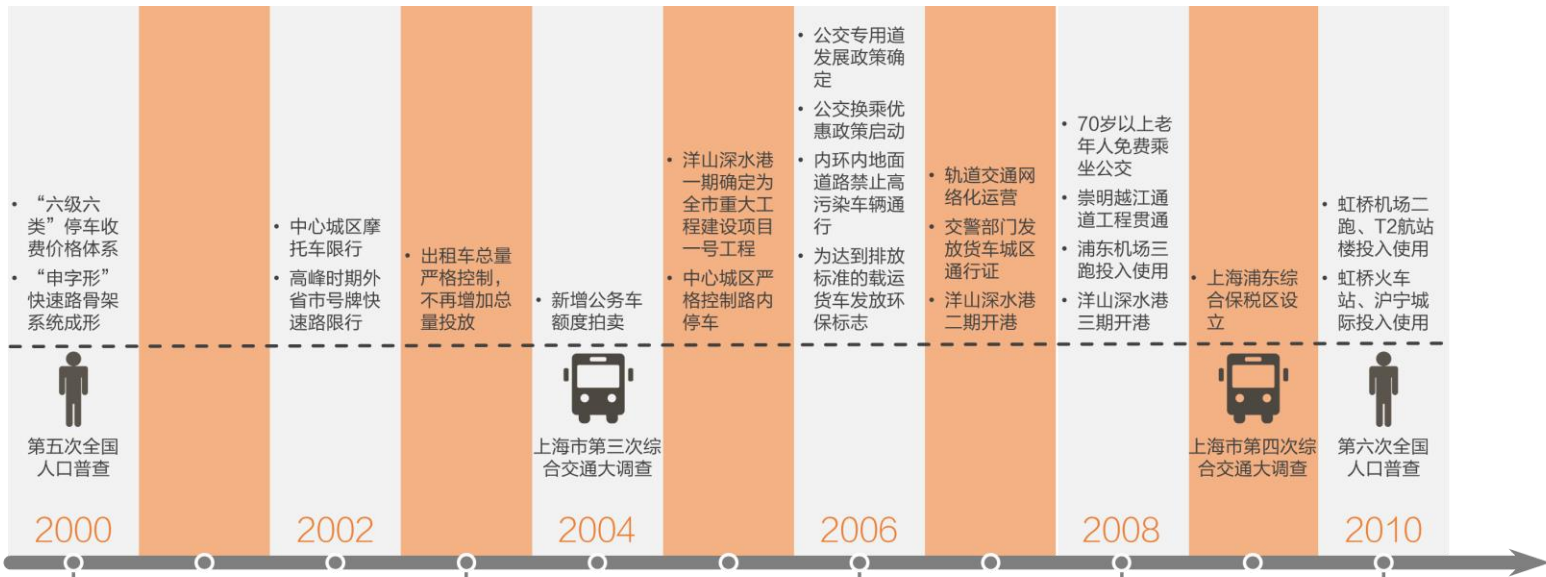
全国机动车缉查布控系统



全国联网接入高速公路、国省道以及省际、市际、县际等重要点位公路卡口**22220**个，累计汇聚机动车通行信息**310**多亿条，日均新增**9000**多万条。

技术手段的瓶颈：连续的数据采集与追踪

重大交通基础设施建
设、交通政策发展



传统数据的采集与更新



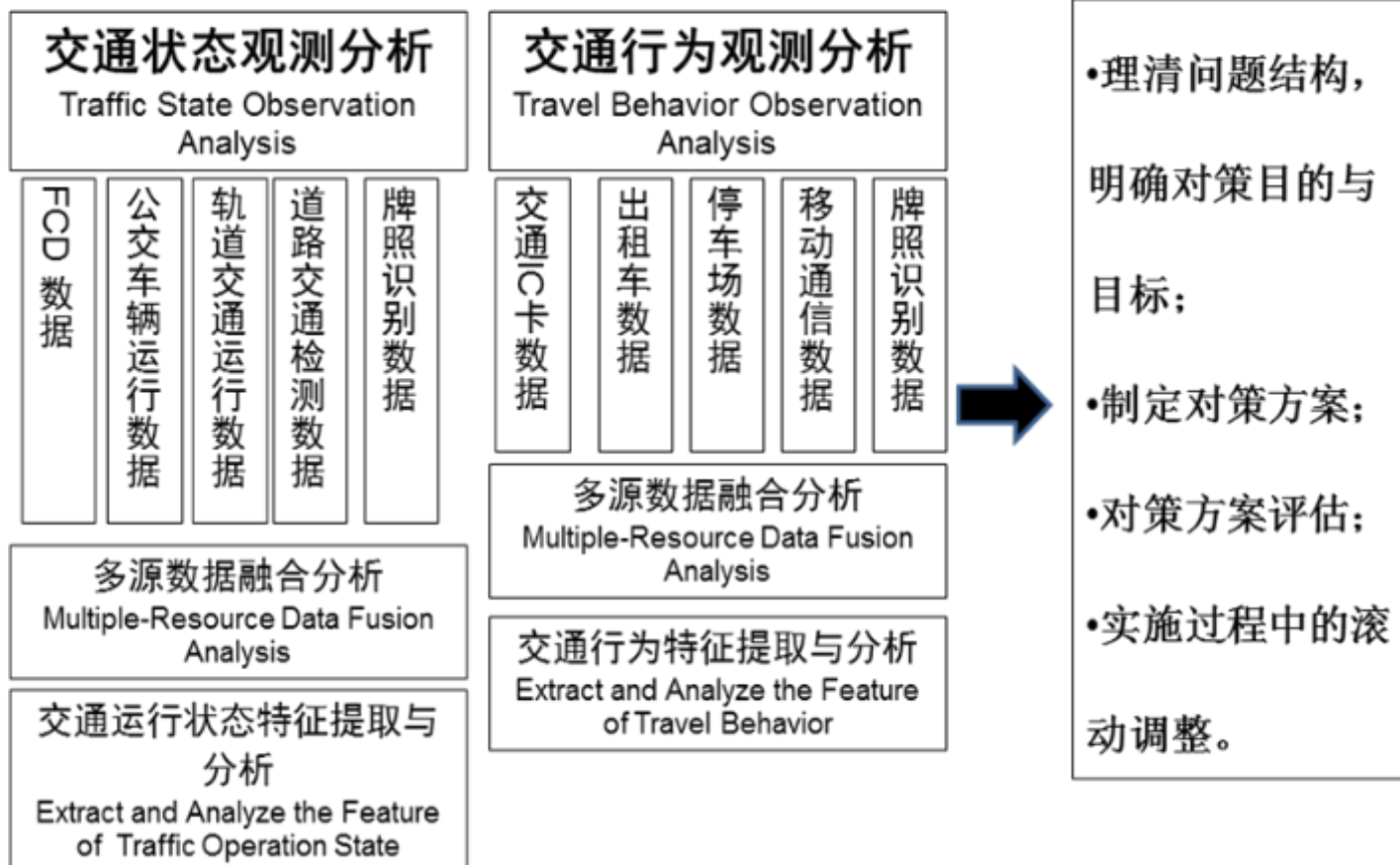
城市的发展与扩张



传统5-10年一次的交通调查已经难以满足城市交通战略过程管理的需求，近年来交通信息系统建设形成的数据资源亟待开发利用。对于城市交通来说，大数据不是一种时髦，而是技术推动、需求拉动和任务理念变革所共同呼唤的创新！

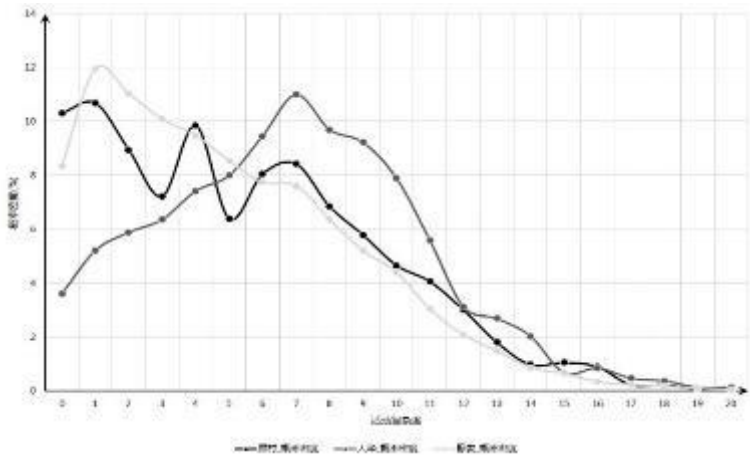
交通大数据环境与政府决策支持

识别现状、认识历史规律、把握系统特征



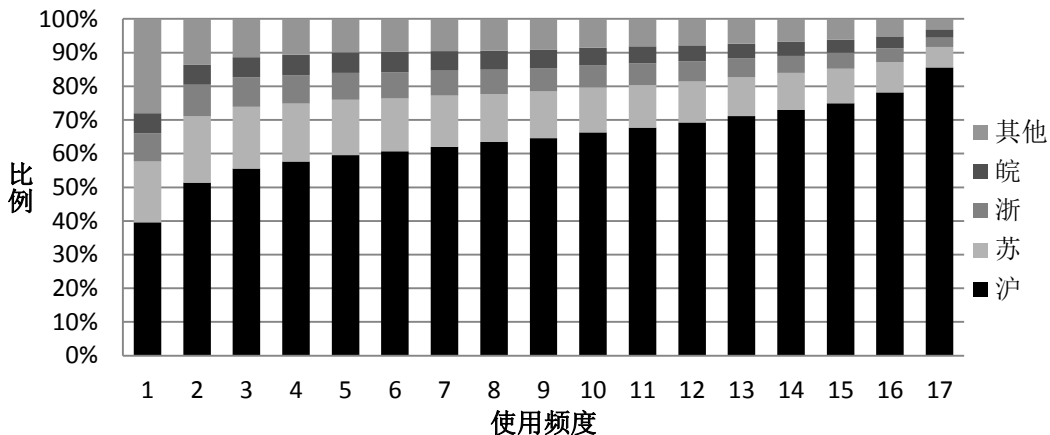
大数据对于城市交通的价值

大样本帮助我们详细了解分布特征



上海市轨道交通7号线沿线顾村、大华、静安社区居民空间活动复杂度对比

连续追踪数据通过频度细分研究对象

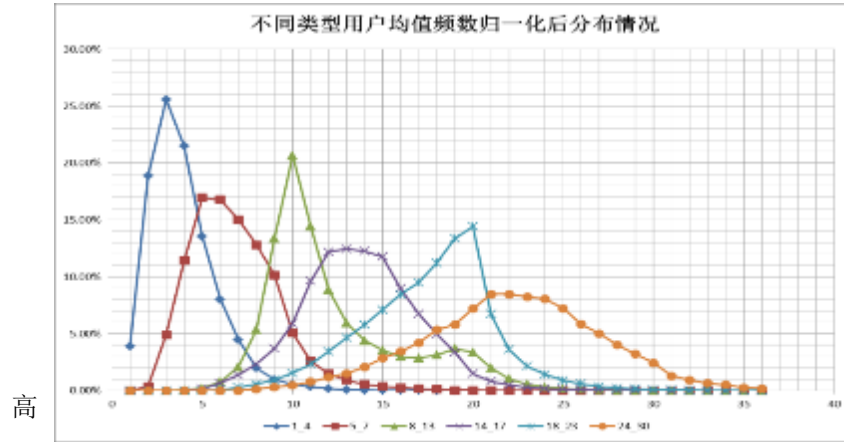
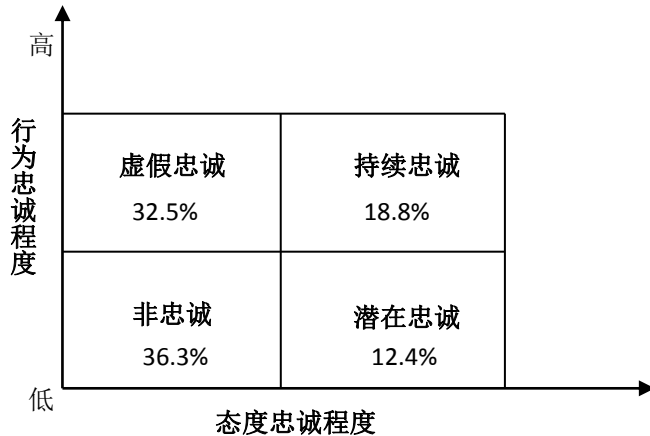


多角度多层次观察帮助全面把握情况

城市空间结构与交通网络关联问题的相关数据资源

数据类型	信息内涵	备注
移动通信数据	城市不同区域居民的活动空间	
公交IC卡数据	城市不同区域居民在公交网络上的活动空间	仅说明了IC卡用户情况。
车辆牌照检测数据	城市中各种类型车辆在道路网络上的活动空间	并非整个路网均被检测系统覆盖，一般主要集中于快速道路。
人口普查数据	城市不同区域居民社会结构	不同区域居民社会经济属性结构的主要数据源。
综合交通调查数据	城市不同区域居民交通行为特征	定制数据因此针对性强。
轨道交通运行数据	城市不同区域通过轨道交通网络形成的可达性	
公交GPS数据	城市不同区域通过常规公交网络形成的可达性	一般需要与调度数据配合使用。
FCD数据	城市不同区域通过道路交通网络形成的可达性	数据质量取决于采样车辆规模。
城市POI数据	城市公共服务设施分布	
大众点评数据	城市公共服务口碑	
经济普查数据	城市就业岗位分布	专项数据，质量容易得到保障。

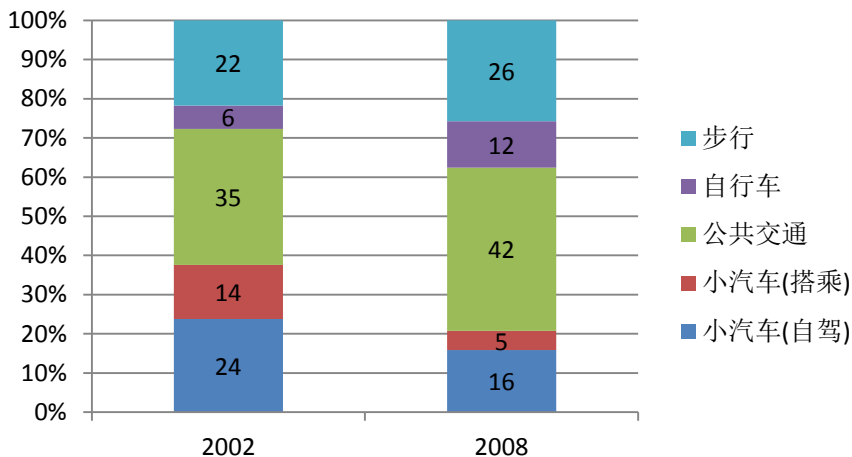
案例1：现有公交乘客是“不得不坐”还是“能够接受”？



利用IC卡数据对不同
类型乘客乘坐频次
的分析

看到这组调查数据对挑战的思考

柏林18-24岁居民出行方式结构

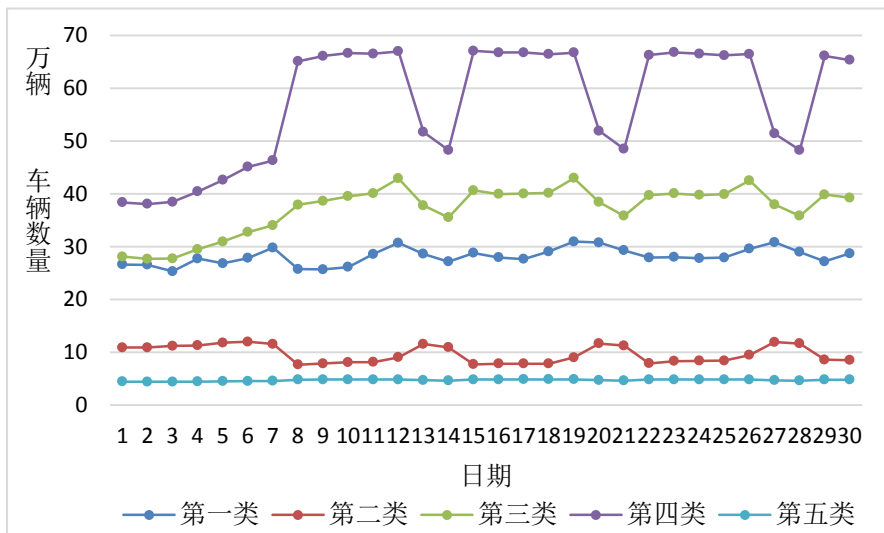


柏林交通模式变化的原因：

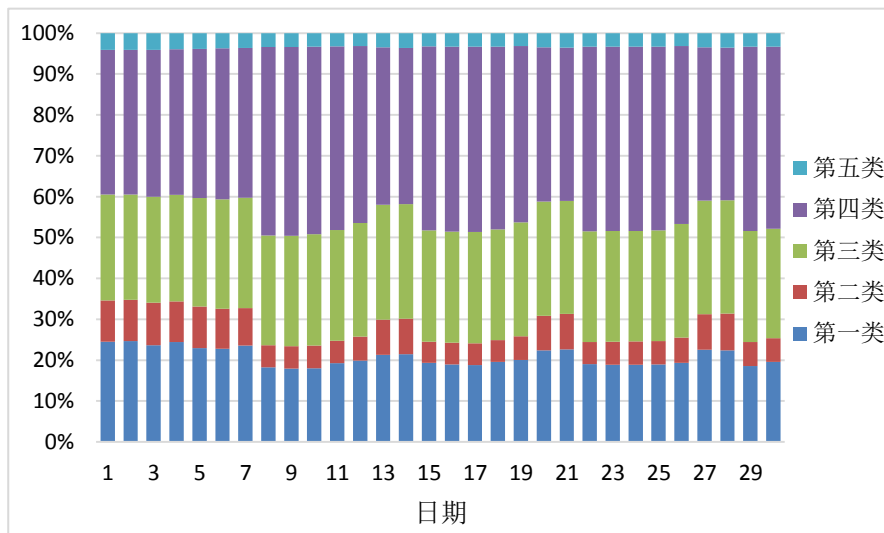
城市多样化功能和混合用地开发；
短距离出行，无汽车依赖；
便于使用的公共交通、自行车、步行；
观念的改变，小汽车不再是身份的象征。



案例2：交通需求管理抑制与服务提升



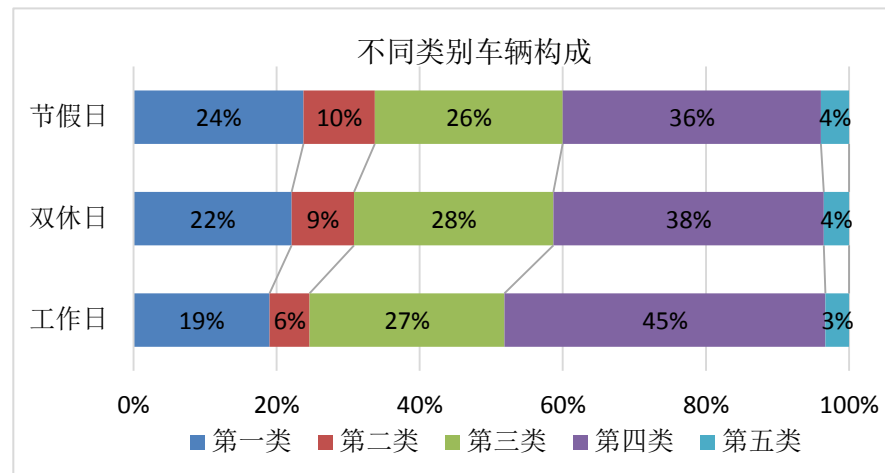
每天不同类别车辆数量



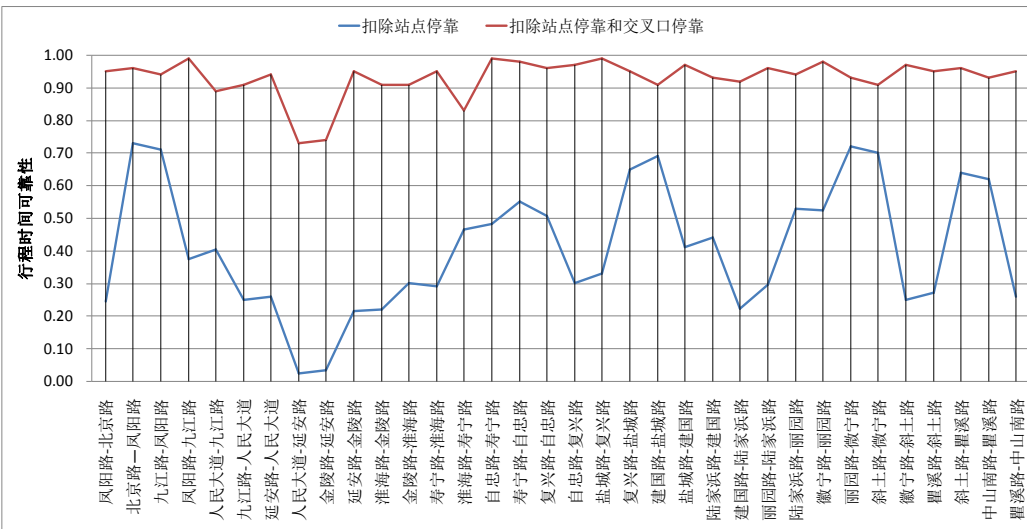
每天不同类别车辆构成比例

拥堵的延安路高架的车辆构成

对工作日而言，48%的第四、五类车辆产生了71%的记录量，而第一、二、三类车辆以51%的车辆仅产生了29%的数据记录量；对双休日而言，42%的四、五类车辆产生了62%的记录量，节假日40%的四、五类车辆产生了60%的记录量。不难发现，第五类车辆每天以3%-4%的车辆产生了17%-18%的记录量。

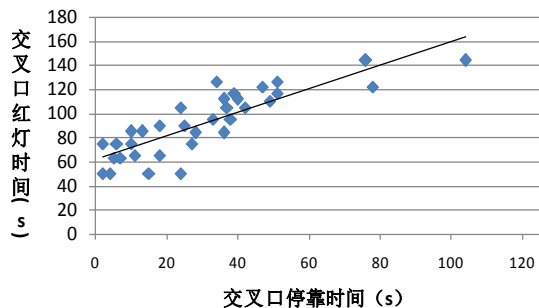


案例3：如何精细地提升公交服务水平

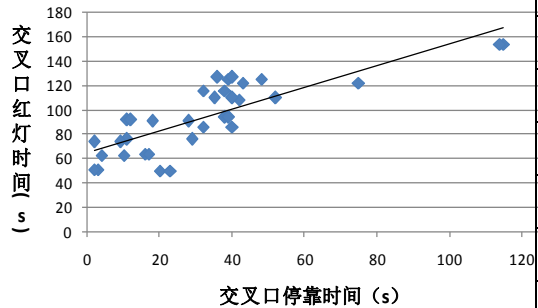


行駛型態	機車	市區公車	小汽車
時間占比			
怠速停等	37.16%	43.85%	40.13%
巡航	10.13%	10.43%	10.35%
加速	24.81%	22.03%	24.60%
減速	27.91%	25.63%	24.92%

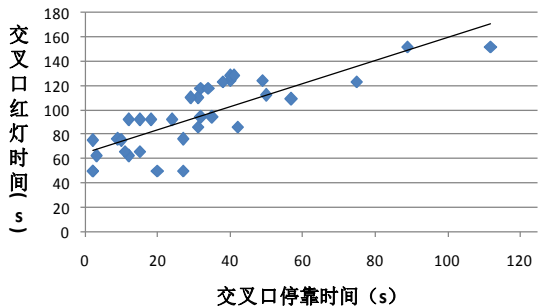
早高峰交叉口停靠时间与红灯时间关系图



晚高峰交叉口停靠时间与红灯时间关系图



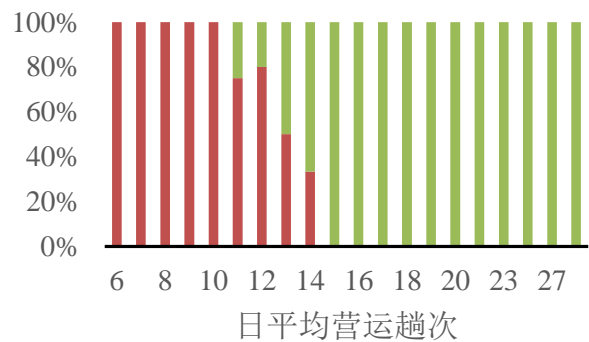
平峰交叉口停靠时间与红灯时间关系图



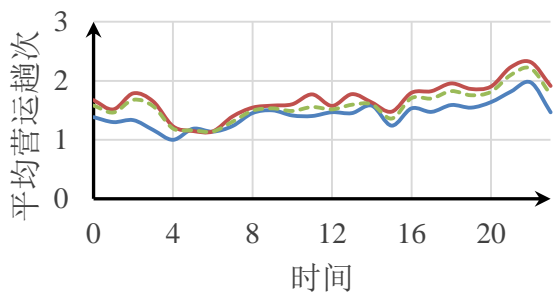
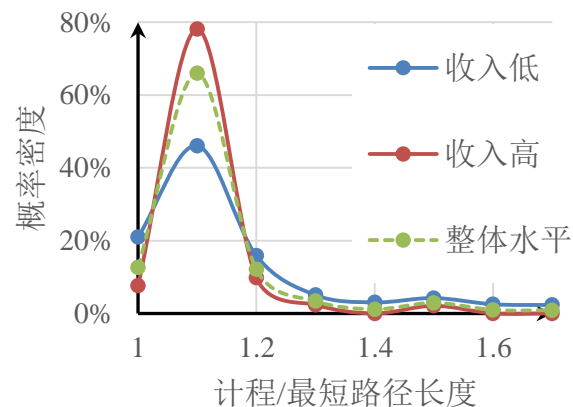
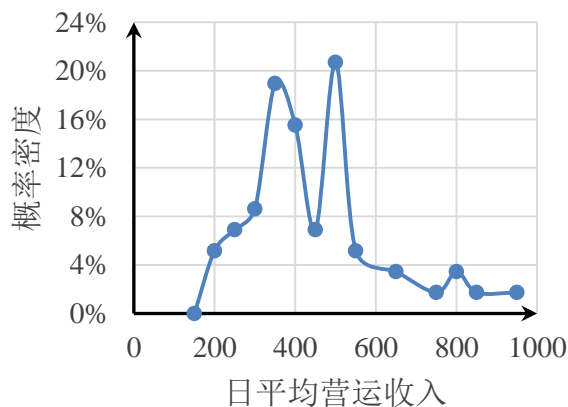
有针对性地改善信号配时，调整线路衔接，提高公交可靠性。

公交站点	位置	车站形式	泊位数	复线系数	A类停靠时刻表的可靠性	B类停靠时刻表的可靠性	C类停靠时刻表的可靠性
西藏中路 北京东路站 (802路北向南)	凤阳路 北进口 上游 49 米	路侧式	2	7	0.47	0.75	0.89
西藏中路 北京东路站 (802路南向北)	北京路 南进口 上游 81 米	港湾式	1	6	0.50	0.74	0.88
人民广场 (福州路)站 (802路北向南)	福州路 北进口 上游 148 米	路侧式	2	8	0.57	0.78	0.89
人民广场 (福州路)站 (802路南向北)	九江路 南进口 上游 145 米	港湾式	3	11	0.54	0.76	0.93
人民广场 (武胜路)站 (802路北向南)	延安路 北进口 上游 255 米	路侧式	2	8	0.54	0.81	0.87
人民广场 (门东路)站 (802路南向北)	福州路 南进口 上游 108 米	路侧式	2	5	0.48	0.62	0.77
西藏南路 淮海东路站 (802路北向南)	寿宁路 北进口 上游 112 米	路侧式	2	8	0.54	0.63	0.82
西藏南路 淮海东路站 (802路南向北)	淮海东路 南进口 上游 137 米	路侧式	2	9	0.50	0.74	0.84
老西门 (西藏南路)站 (802路北向南)	盐城路 南进口 上游 95 米	路侧式	1	6	0.78	0.83	0.99
老西门 (西藏南路)站 (802路南向北)	盐城路 南进口 上游 90 米	路侧式	1	7	0.82	0.88	0.95
老西门 (西藏南路)站 (969路南向北)	盐城路 南进口 上游 90 米	路侧式	1	7	0.81	0.96	0.99
西藏南路 陆家浜路站 (802路北向南)	陆家浜路 北进口 上游 190 米	路侧式	2	7	0.91	0.99	0.99
西藏南路 陆家浜路站 (969路北向南)	陆家浜路 北进口 上游 190 米	路侧式	2	7	0.86	0.97	0.99

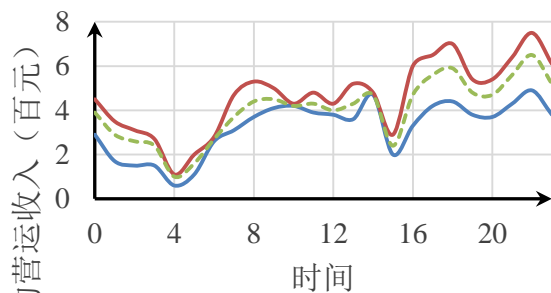
案例4：某市电召出租车营运收入出现分化征兆的原因分析



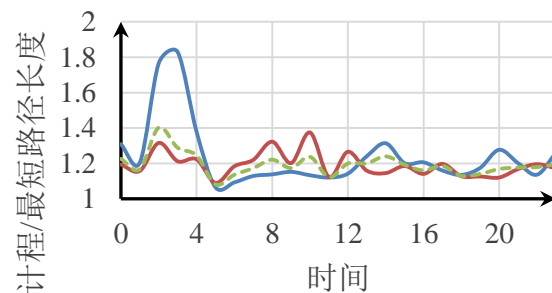
■ 收入低 ■ 收入高



— 收入低 — 收入高
- - - 整体水平

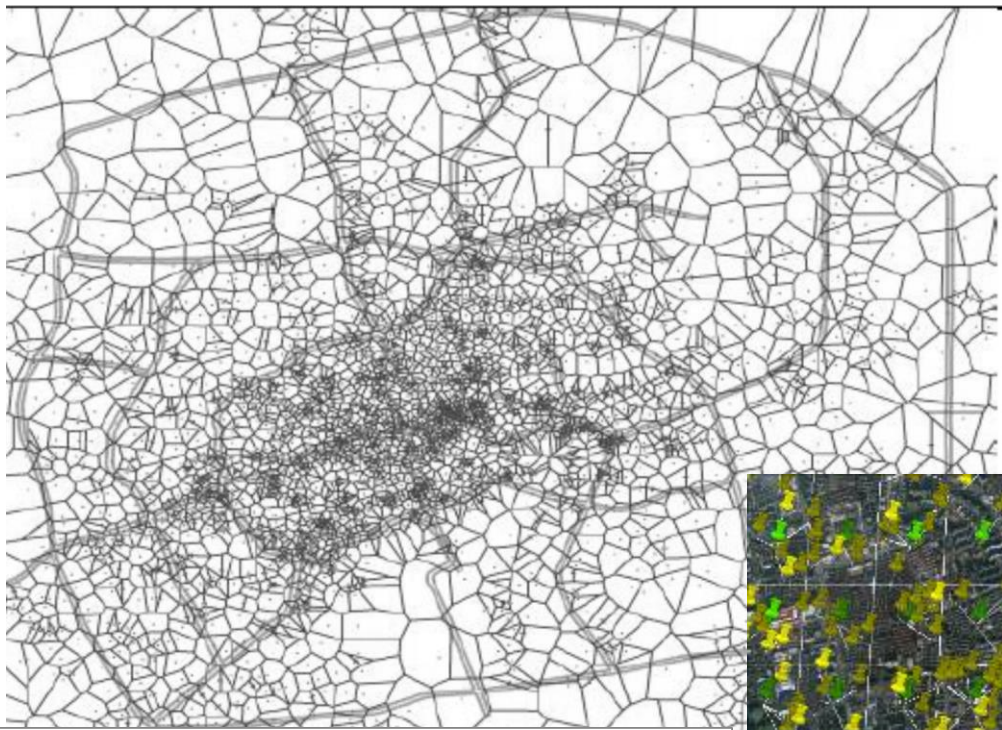


— 收入低 — 收入高
- - - 整体水平



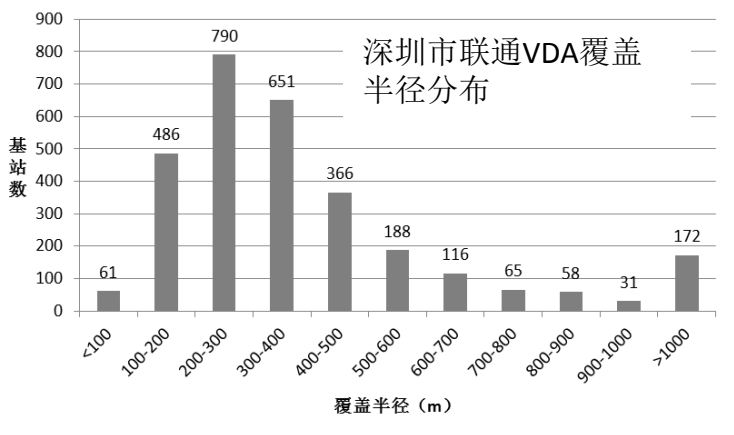
— 收入低 — 收入高
- - - 整体水平

大数据并非完美信息环境：信令数据的空间分辨率



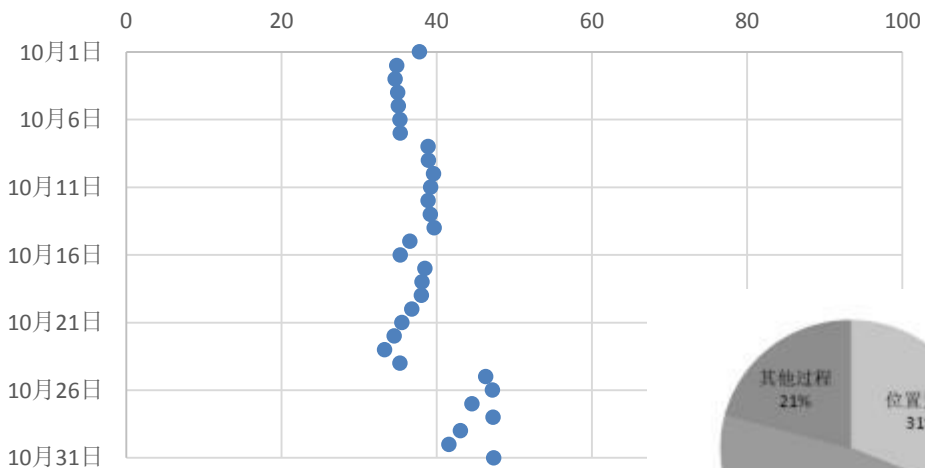
上海市VDA的几何特征

	统计指标	统计值
所有 VDA	平均面积 (平方千米)	0.95
	平均周长 (千米)	2.99
	平均半径 (千米)	0.48
外环以内的 VDA	平均面积 (平方千米)	0.24
	平均周长 (千米)	1.57
	平均半径 (千米)	0.25

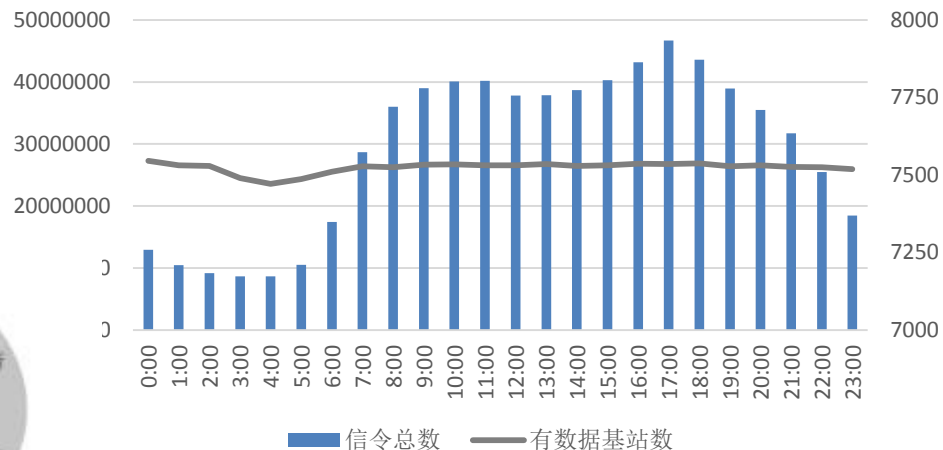


大数据并非完美信息环境：信令数据的时间分辨率

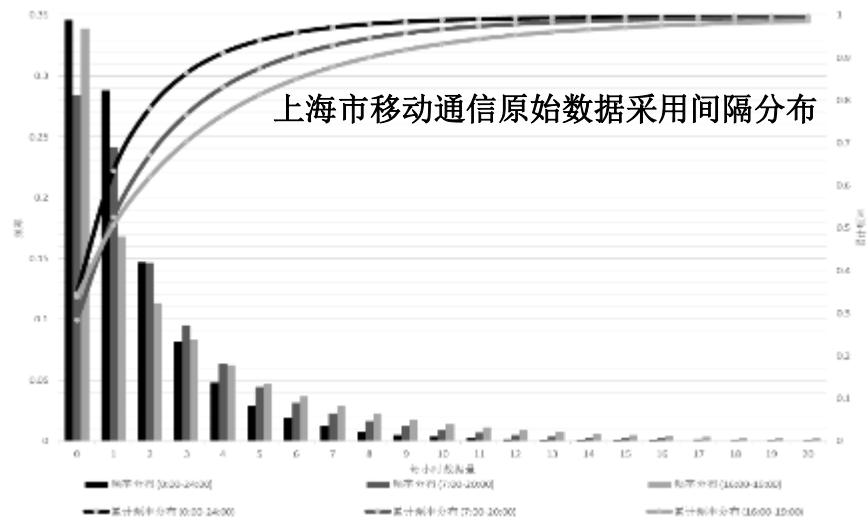
单用户日均信令条数



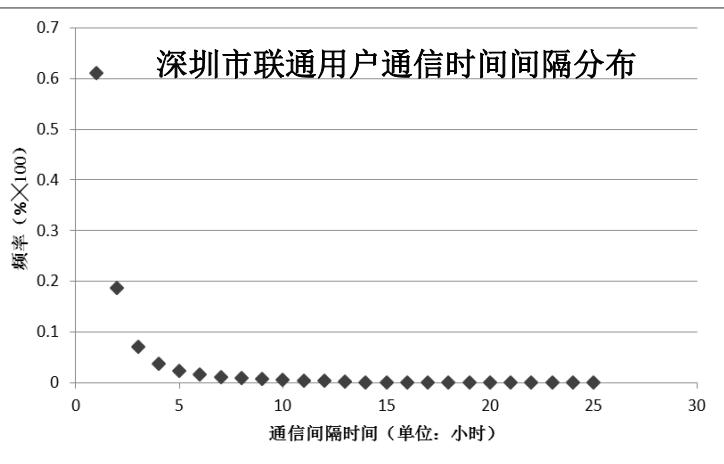
移动通信信令数量时间分布



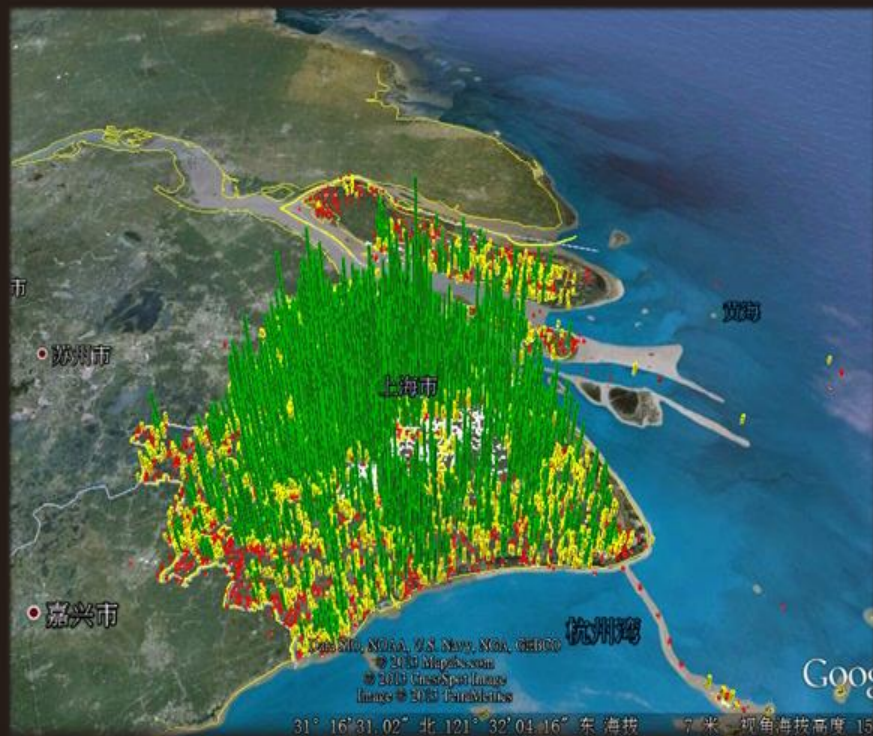
上海市移动通信原始数据采用间隔分布



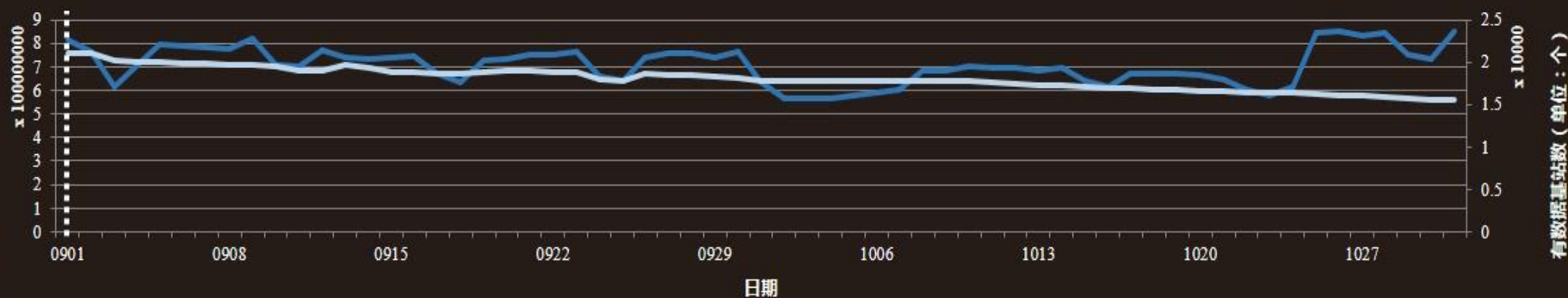
深圳市联通用户通信时间间隔分布



数据本身的不完备——缺失



数据质量 (指令数, 单位: 条)



— 数据质量 — 有数据基站

数据本身不完备：未识别用户

COLUMN	TYPE	DESCRIPTION
MSID	Varchar2(32)	移动用户唯一识别号，32位大写。 是IMSI或TMSI单向加密的结果，规避用户隐私
TIMESTAMP	NUMBER(14)	信令记录发生时间戳，单位：秒。 格式：YYYYMMDDHH24MISS。
LAC	Number(5, 0)	位置区编号，唯一标识一个LAC
CELLID	Number(5, 0)	LA内基站小区编号（只在LA内唯一）。 LAC&CELLID才能唯一标识一个基站小区
EVENTID	Number(3, 0)	信令事件类型
CAUSE	Number(3, 0)	信令时间的触发原因
SECID	Varchar2(13)	辅助标识MS（可忽略）
FLAG	Varchar2(3)	表征是否捕捉到IMSI
RES	Varchar2(8)	MSC加密后的标识（可忽略）
RES1	Number(8, 0)	BSC加密后的标识（可忽略）
CODE	NUMBER（3）	000、100在一定范围内一致唯一（约为92%）； 101、001 无法标识用户（约为8%）

交通领域大数据应用的难点

- 非定制数据造成的一定程度信息不确定性；
- 单一数据源如同“盲人摸象”，不能给出完整的对象描述图像；
- 不同的数据源具有不同的参照标准，并且服从不同的误差分布，缺乏统一的数据标准；
- 数据工程师独特的思维方式，造成拘泥传统模型思维的技术人员难以适应和沟通困难。

交通运输的复杂性在于“综合”，如同中医所谓的“气虚”，一大堆“小问题”综合成为“大问题”，其对策需要向中医学习，不能局限于“治病”，而是需要“医人”。

理念： 城市交通的战略管理

城市交通中的重大战略问题

- 交通是城市的骨架和血脉，交通问题并非“保障措施”，而是城市空间架构、城市发展模式中的核心问题！
- 城市骨架与城市空间结构的协同
- 城市血脉与城市空间功能布局

什么是城市的空间结构？

城市是一个生命体，需要从城市的活动和成长角度，看待城市空间结构，需要认识城市的社会因素影响！

与规划师想象的差异

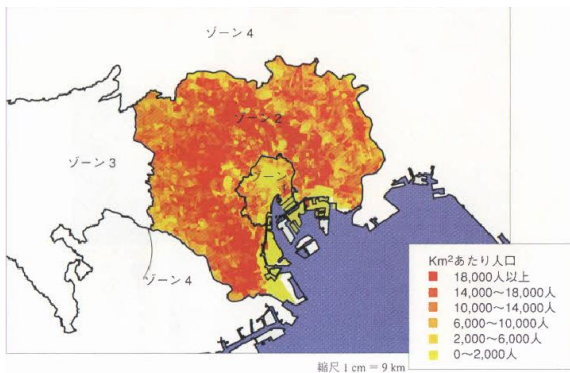


单纯理念的
城市空间结构

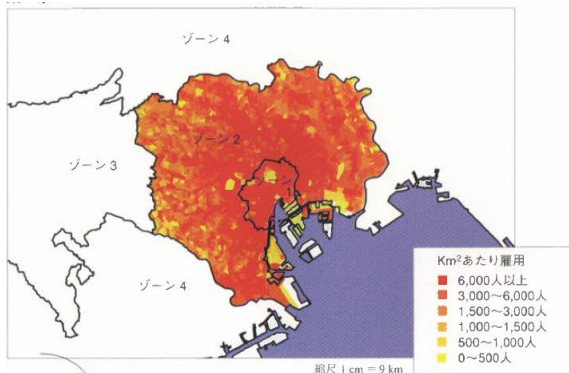


静态分布的观点

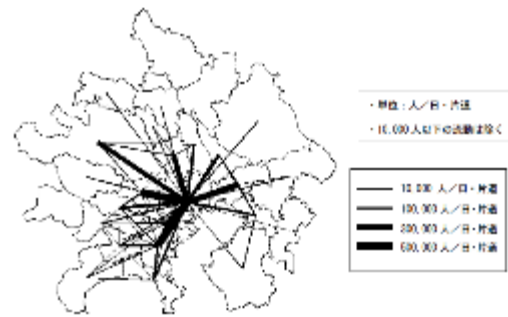
东京都区部人口密度分布图



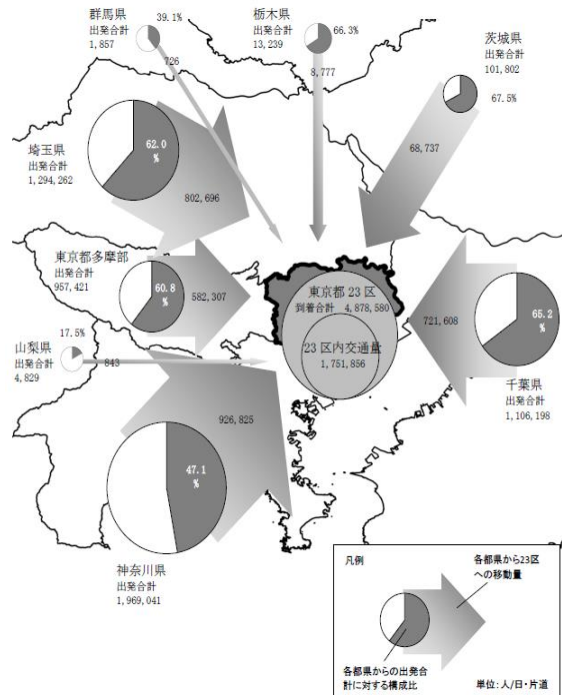
东京都区部就业岗位密度分布图



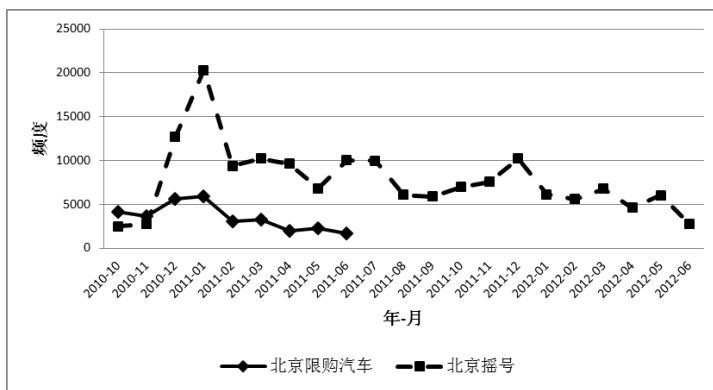
单纯交通流量观点



空间联系的观点



交通面临的挑战比单纯的供需矛盾复杂得多



一方面是车满为患， 另一方面是针对需求管理激烈的争论
问题的本质是人居环境与机动化需求的矛盾

首尔市清溪川改造工程



改造前



改造后(2005年9月)

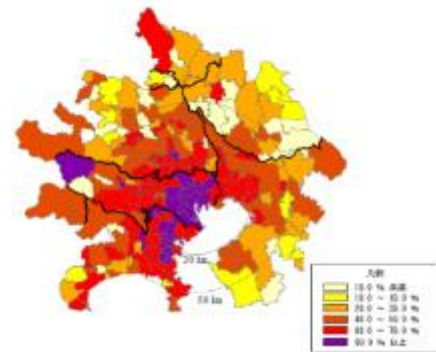
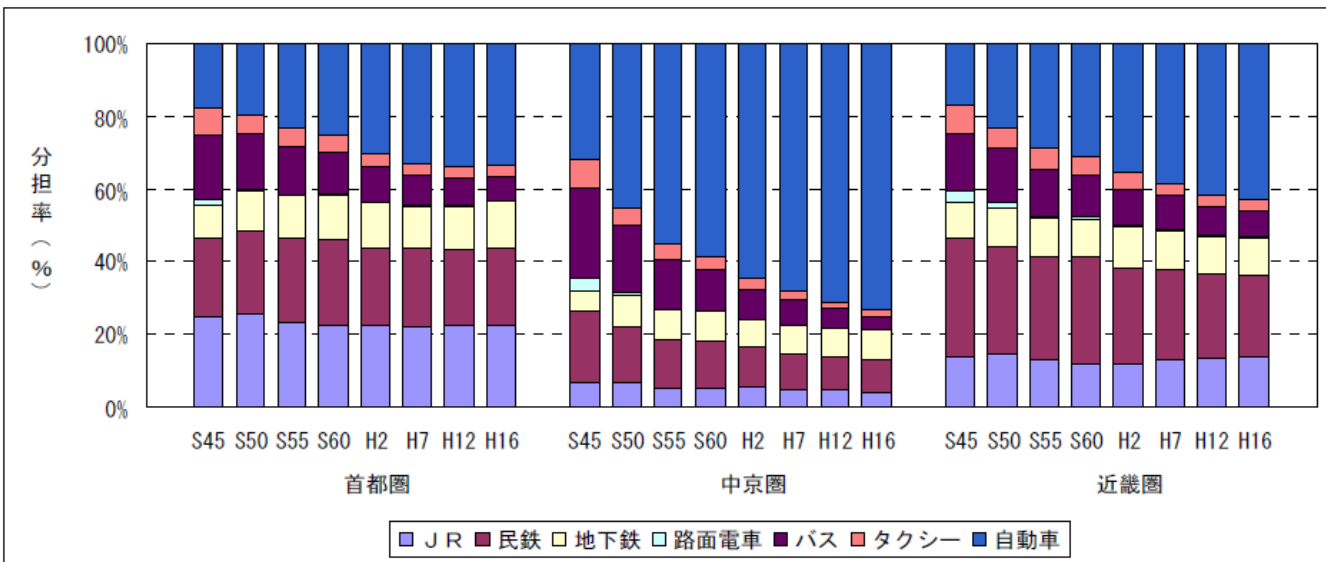
拆除了5.8公里的高架路，创造宜居的生态环境
 减少车流：125,000 辆/日



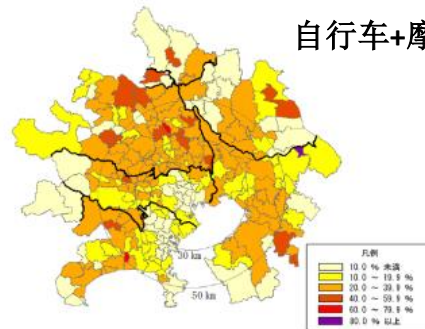
不时出现的瘫痪性拥堵影响了社会与经济的正常运行。

单纯交通设施建设并不一定能够改变城市交通困境

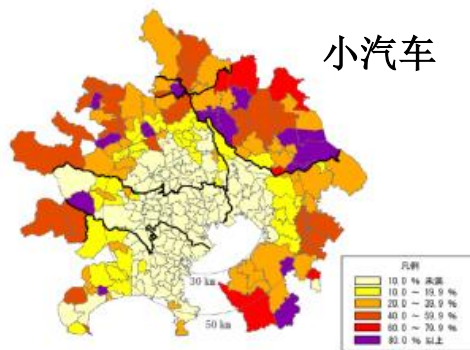
日本三大都市圈交通模式的演变



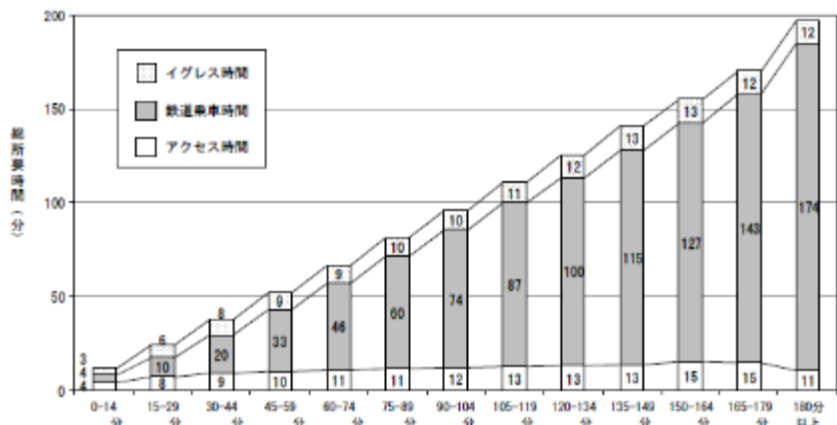
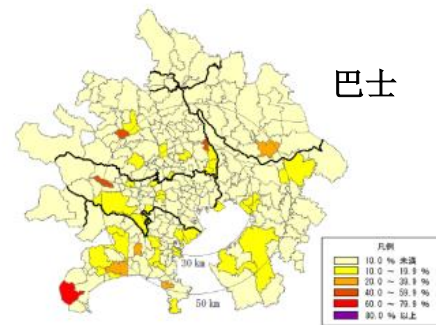
自行车+摩托



小汽车



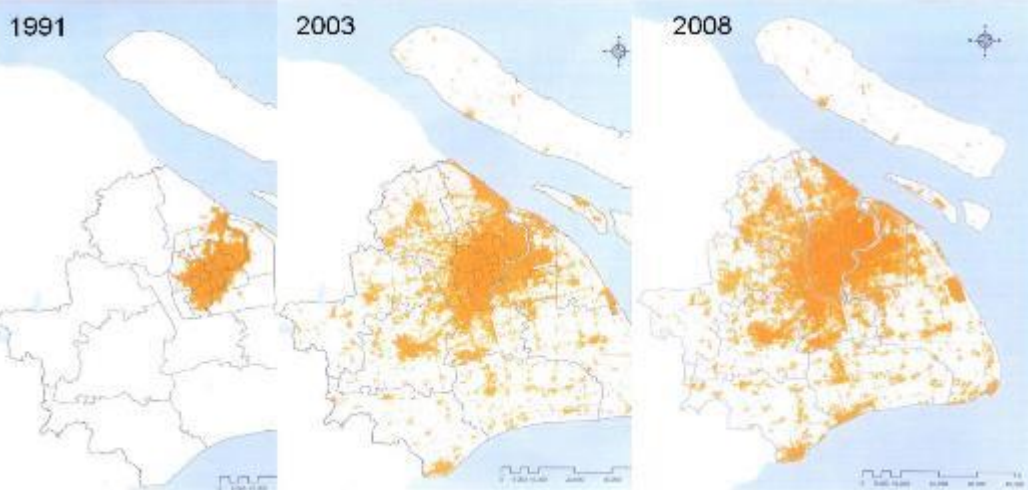
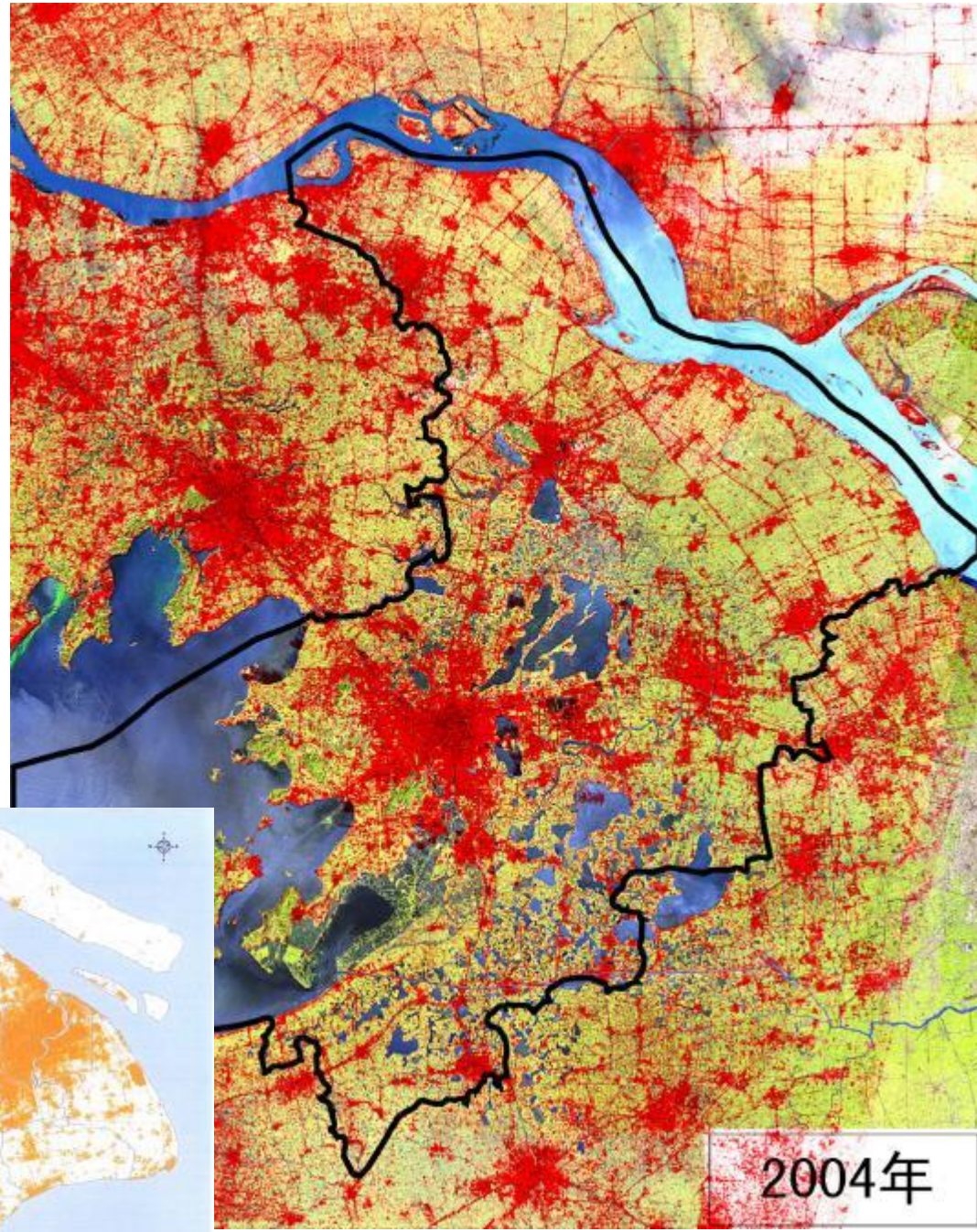
巴士



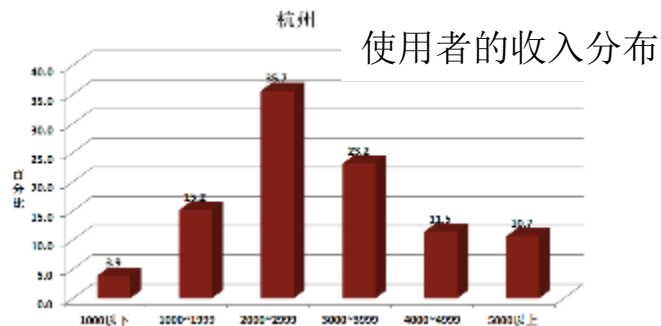
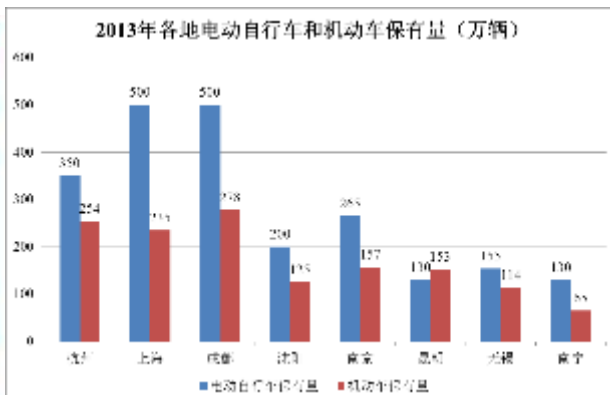
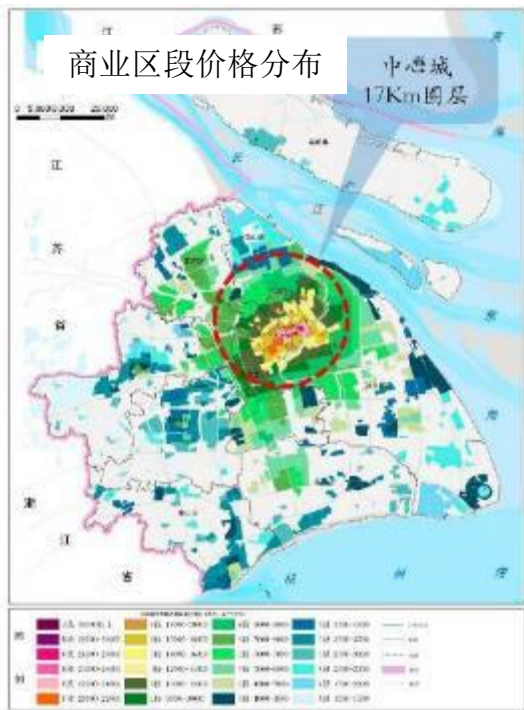
东京圈轨道与接驳时间的关系

通过轨道交通接驳方式理解东京轨道为什么有很高的分担比

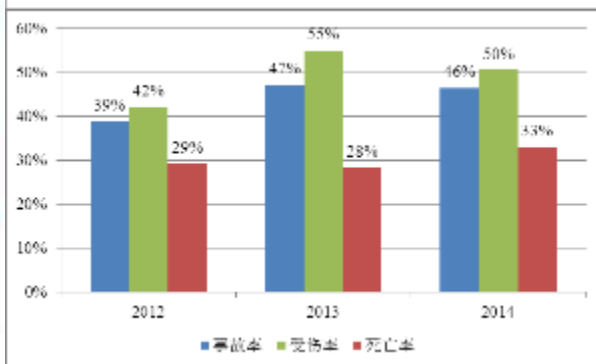
土地资源成为强约束



数据还在不断告诫我们新的挑战



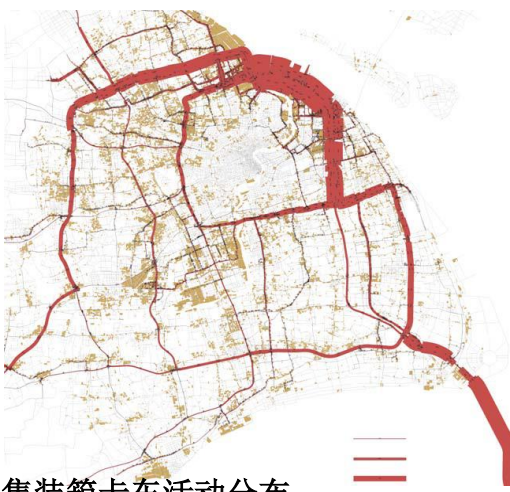
快速增长的电动助动车是中低收入阶层的主要私人交通工具



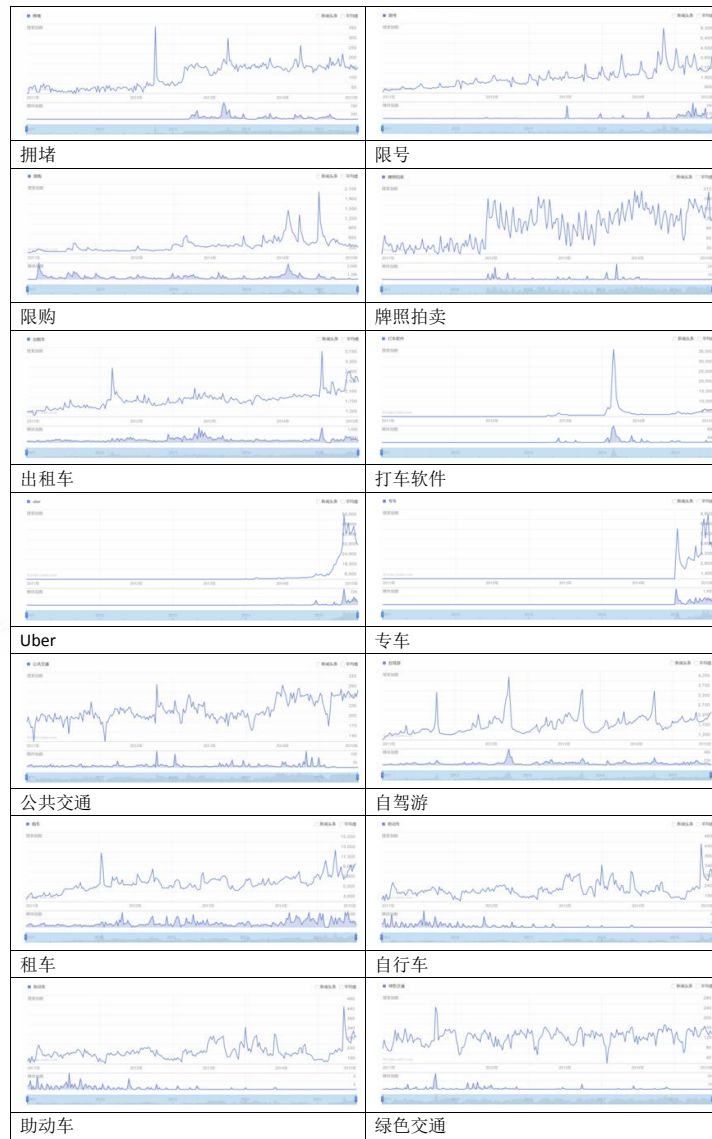
缺乏预见性的行业管理，难以适应市场参与形式的变化.....



根据上海市2011年商业区段价格分布图，在以人民广场17Km半径范围内，环境的非均质分布产生的价值级差覆盖3-4个量级（M级8000至S级3500），宝山新城明显受此影响（由集卡GPS数据绘制的上海市集卡运输路径及流量分布图可见，其处于集卡路线覆盖区范围）。



面对“互联网+”，我们不能做“鸵鸟”



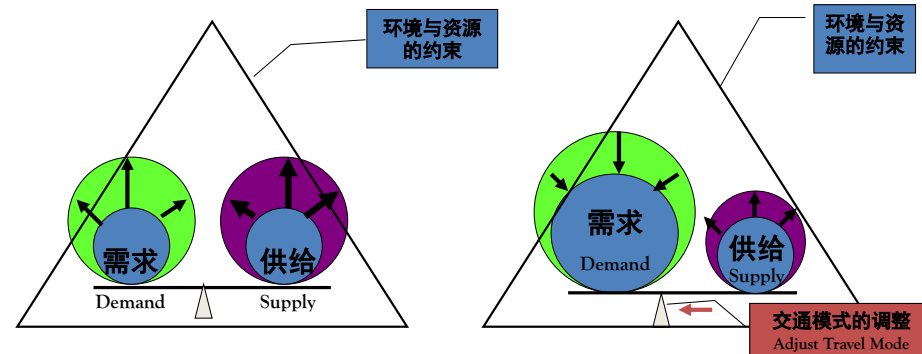
国家正在酝酿的“十三五”重大科技项目中，“互联网+”作为一个单独领域，至少有两个项目有可能涉及交通：“互联网+服务”，“互联网+交通”。我们已经熟知的“公共交通”、“个体（私人）交通”之外，正在形成新的交通模式！

“交通最重要的任务就是进行基础设施建设”正在成为过去时！

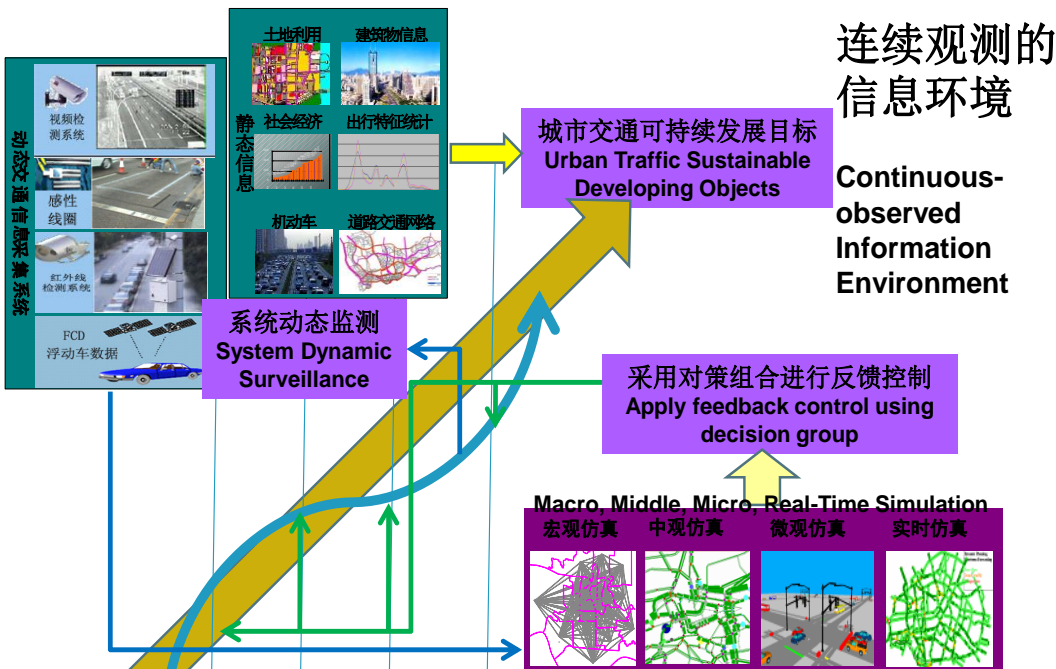
交通需要对演变过程进行战略调控



战略调控的核心是交通模式

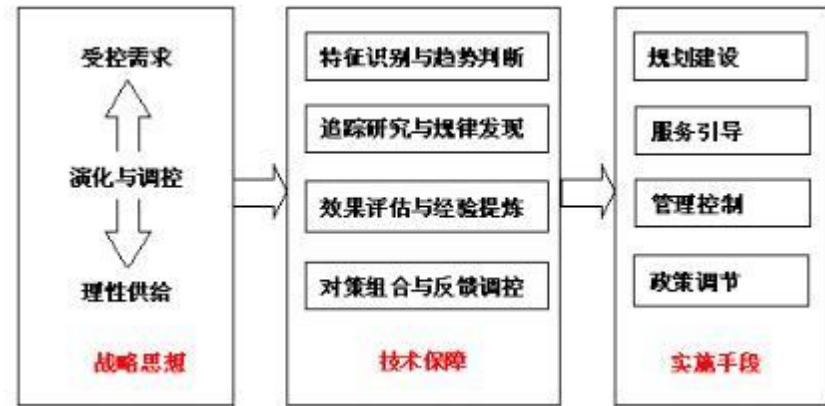


战略调控的手段是交通政策



系统演变规律的认识 Knowledge of System Evolvement Rules

战略调控的目标是遵循可持续发展的轨道



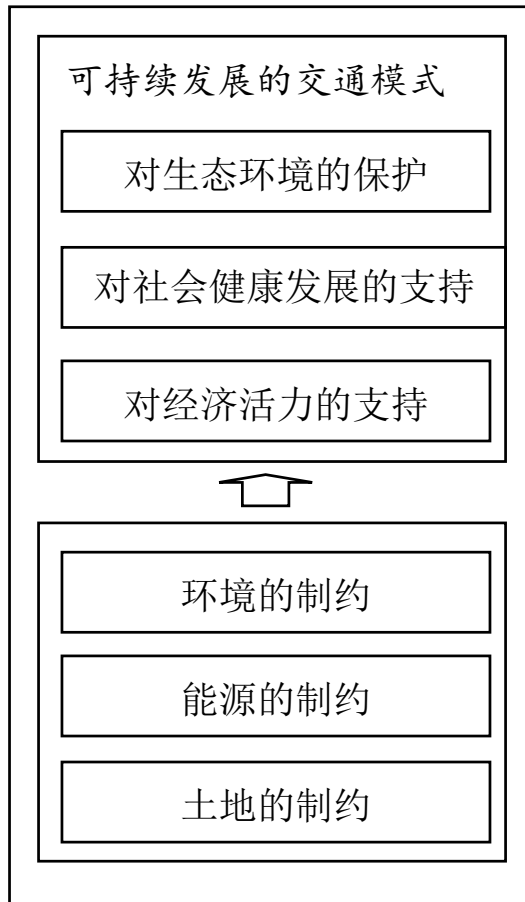
战略调控的关键是建设智慧政府!

从制定“规划方案”走向“战略过程管理”

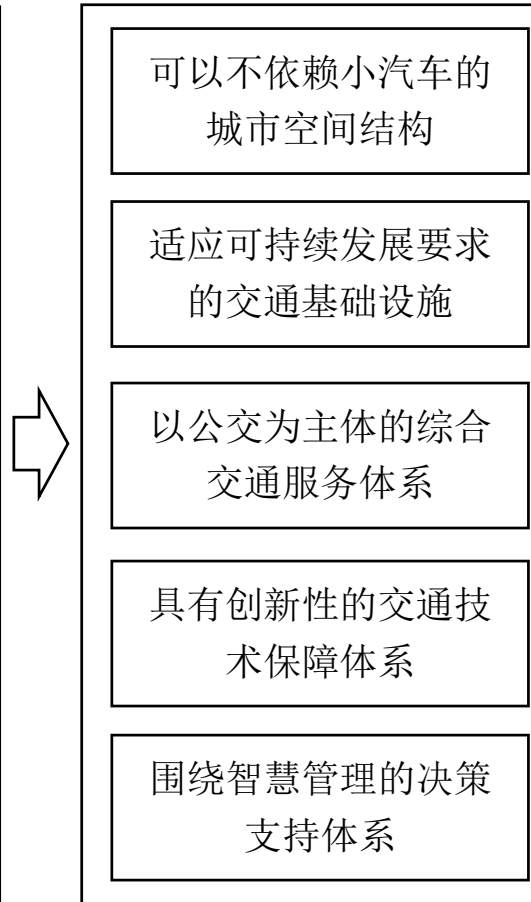
- 基础设施建设、运输组织、综合服务服务，我们正在经历一个提升的演化过程。
- 过程管理需要具有“洞察力”，这意味着发现重大变化，相应调整工作部署。需要承认面对未来，我们的经验是不够的，现有工作内容必须“与时俱进”进行调整。
- 现有理论面临从“西医模式”向“中医模式”，从“还原论”向“整体论”的深刻变革。
- 而大数据提供的观察、诊断能力，将成为促进管理和技术变革的“推手”！

城市交通战略调控的体系架构

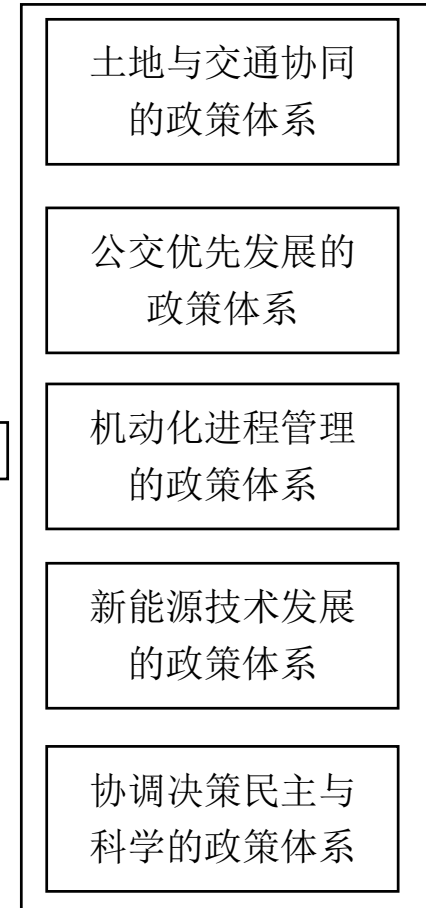
目标



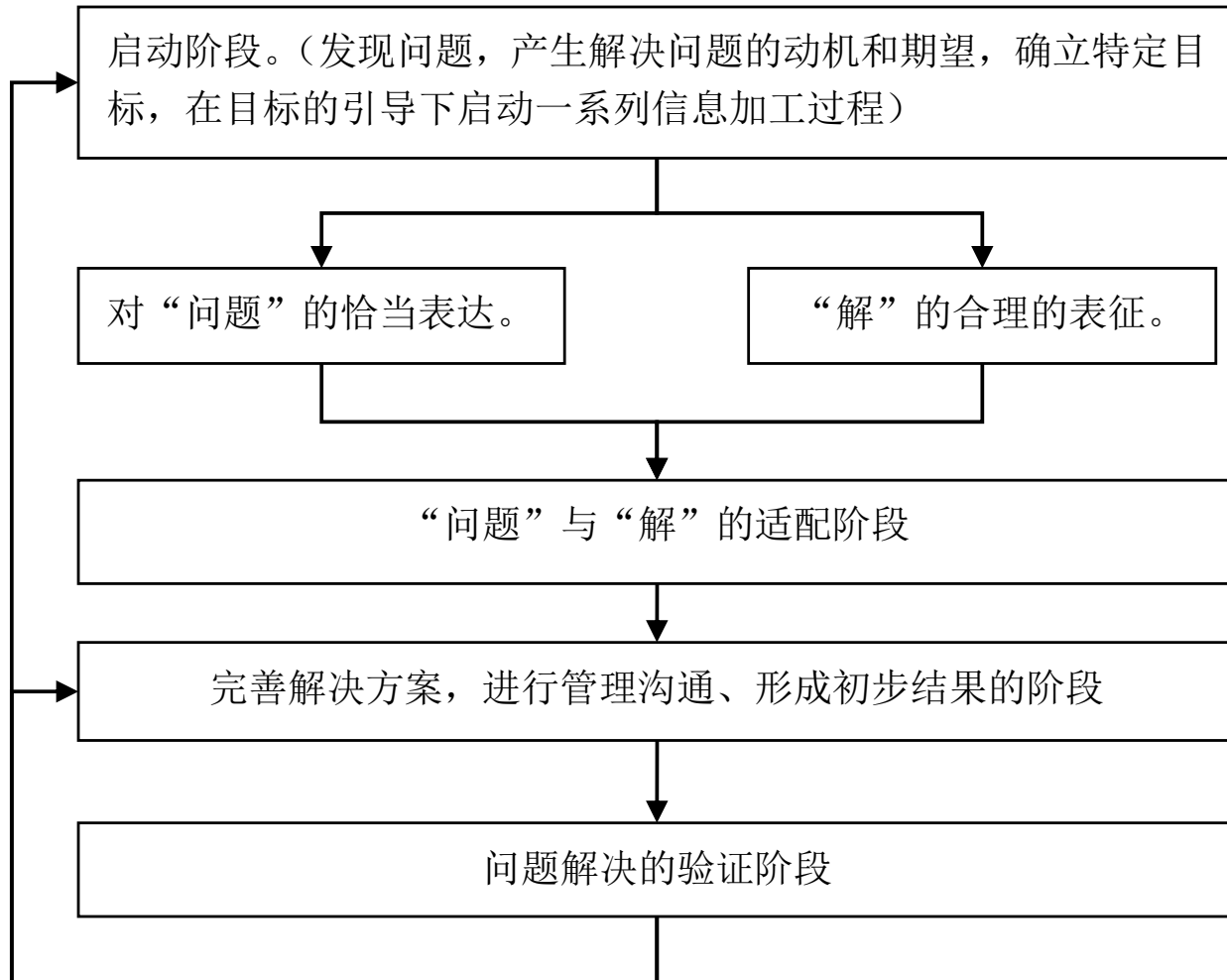
要素结构



杠杆



不是“以不变应万变”而是要“适时反应”

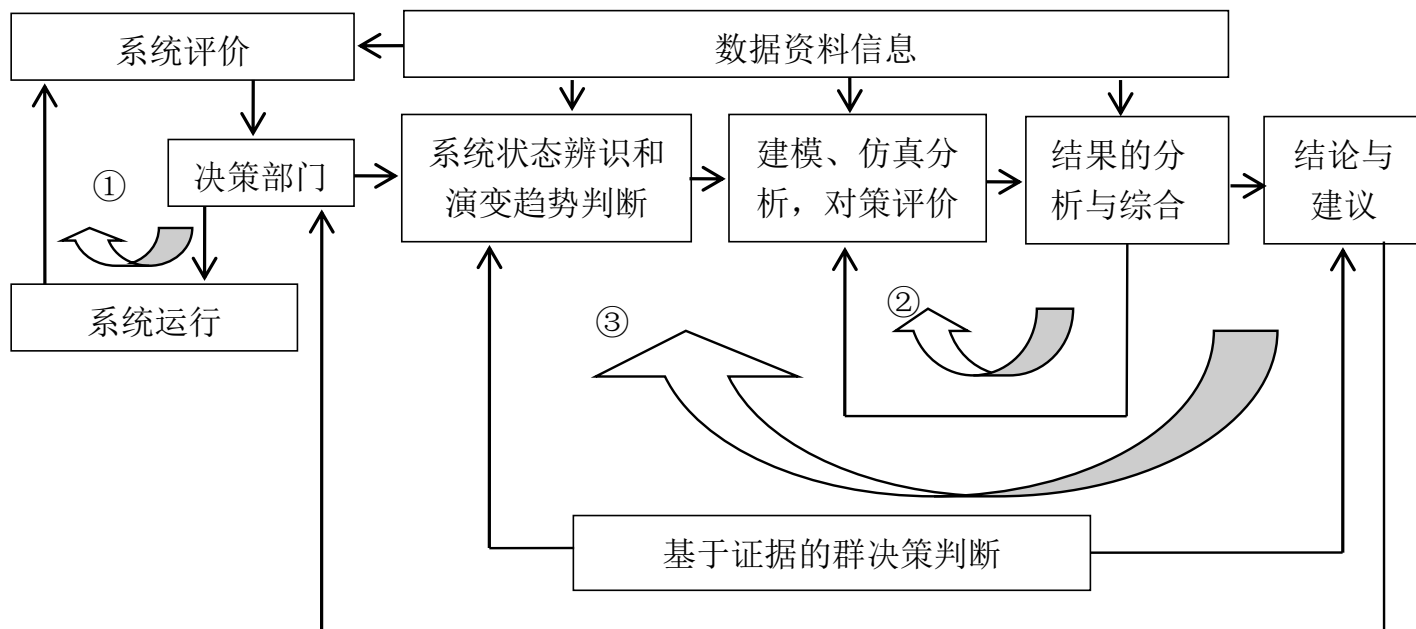


大数据环境下城市交通决策支持技术包括三个支柱：数据驱动分析、仿真实验和模型分析；将其分析结果整合在一起的是基于证据的决策框架。

技术：从“预测”走向“基于证据的决策判断”

政府管理框架下的技术分析

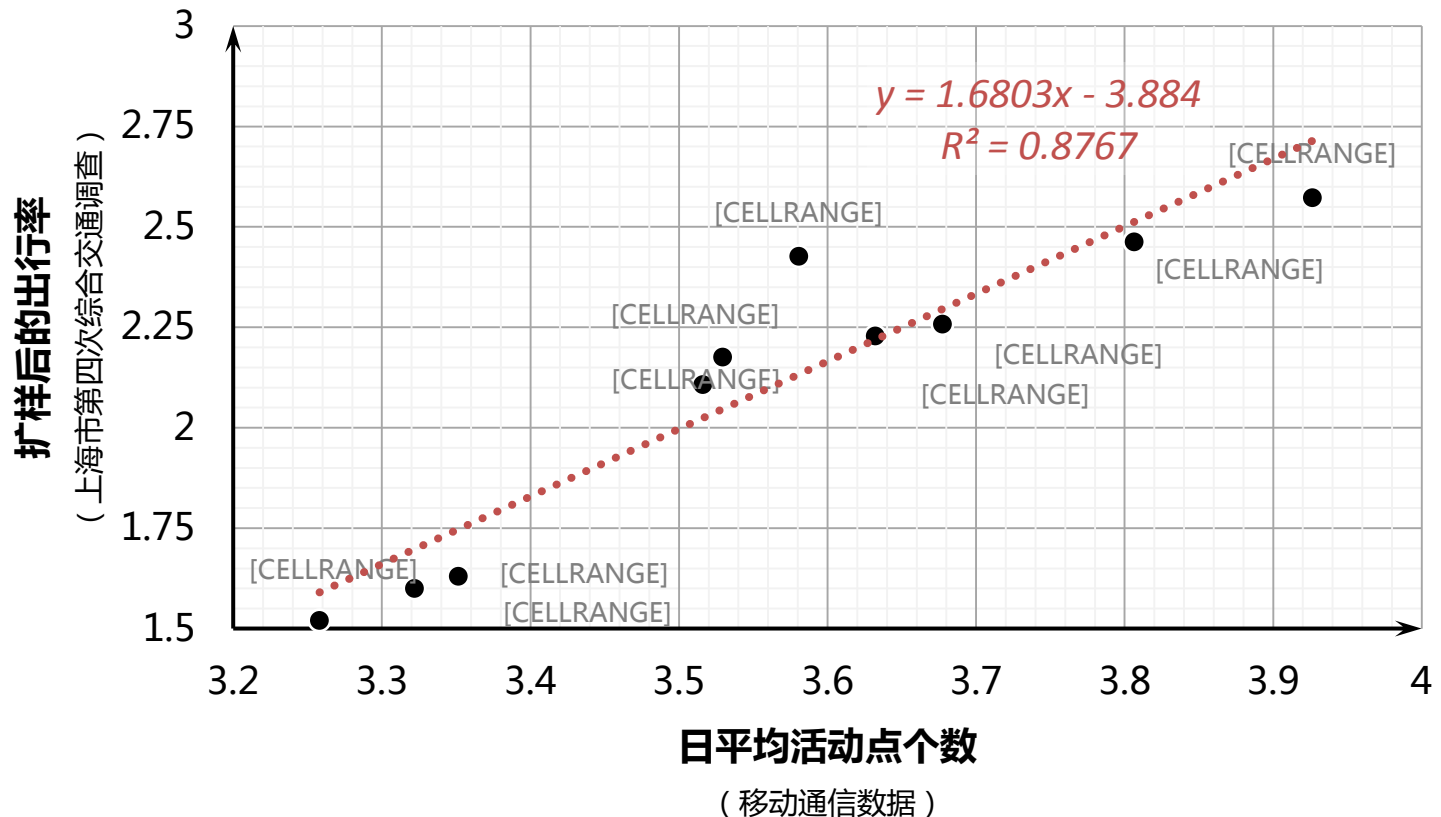
1. 日常监测：系统是否按照预想在运行？
2. 与传统分析技术结合：补充信息来源，增强分析效果。
3. 基于证据的决策判断：预见、预警、预判、评估.....



融入管理流程的技术分析——基于证据的决策

理解大数据中的技术概念：关联分析

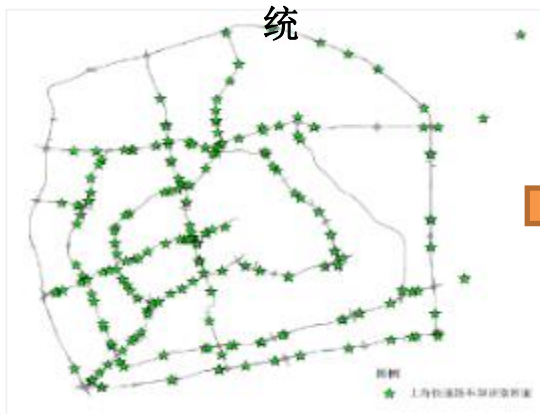
- 关联分析不同于因果研究，是发现同时出现的事物，警示背后可能存在因果。



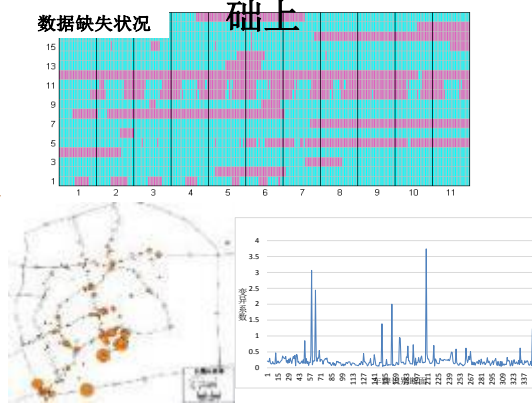
理解大数据中的技术概念：聚类分析

聚类实际上是在建立一种更加方便的“度量”方法。

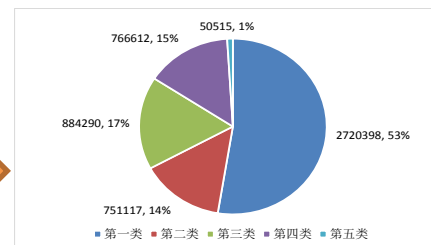
依托不同位置布设的牌照识别系统



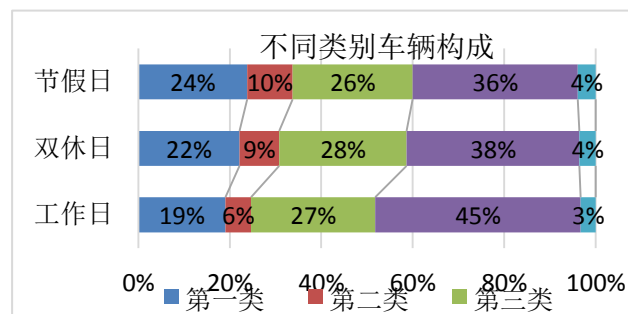
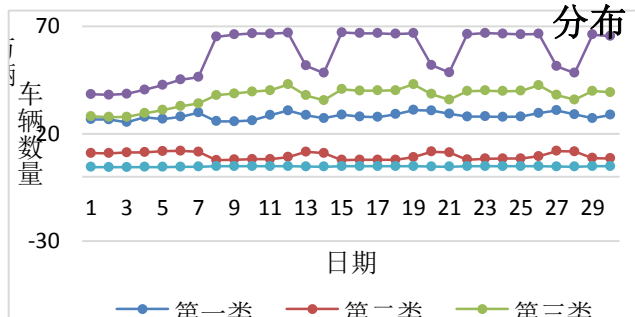
在对数据质量分析和可信度判断基础上



通过聚类进行对象分组

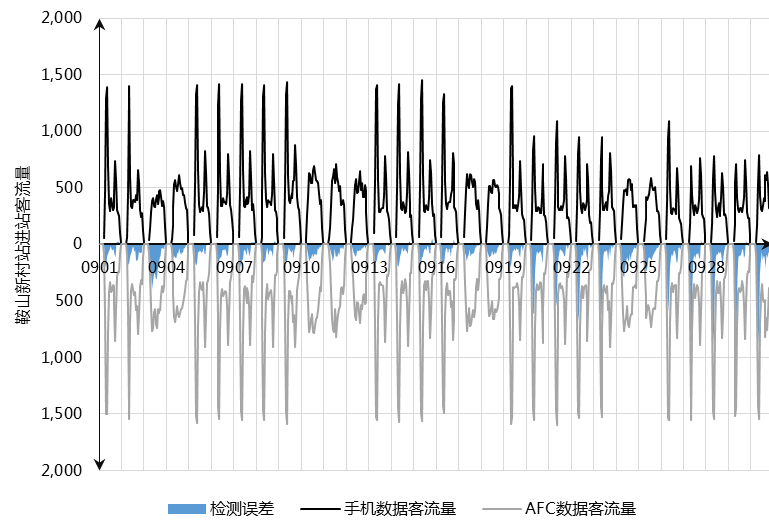
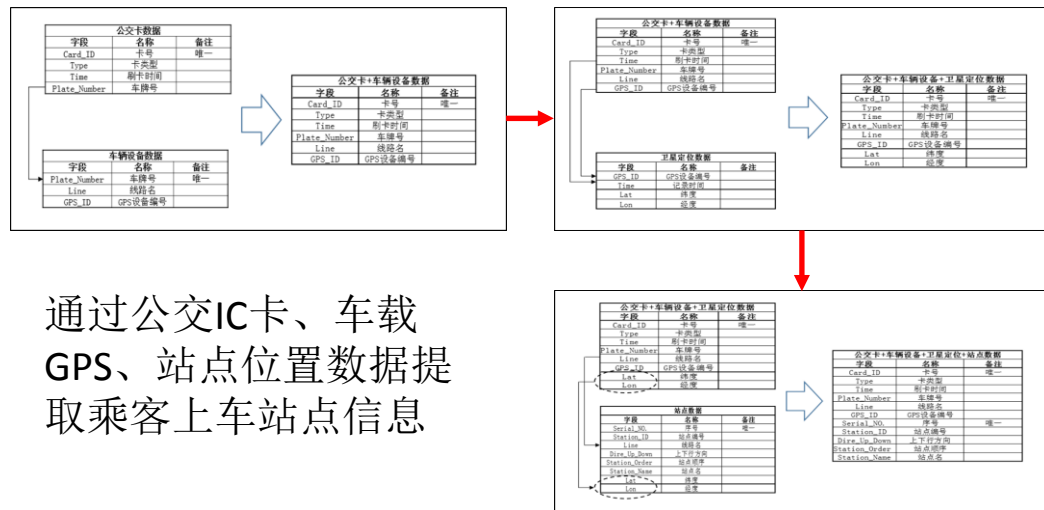


研究不同使用特征类型车辆的时空分布

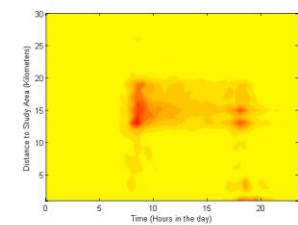


对交通需求管理涉及的对象进行精细化研究

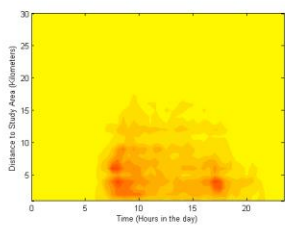
理解大数据中的技术概念：数据融合



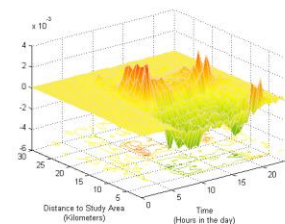
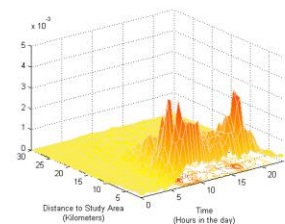
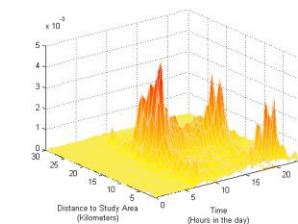
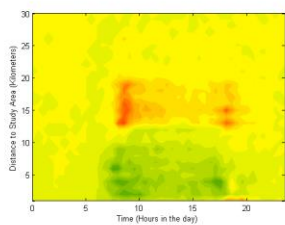
第一类居民
活动时空间分布



第二类居民
活动时空间分布



第一类 - 第二类
绿色表示负值



利用轨道AFC数据校核移动通信数据提取的特征精度

利用移动通信数据精细化分析居民在城市中的活动，协调交通与城市公共服务设施之间的关系。

能否在间接证据基础上做出决断： 构建尽可能完整的证据链

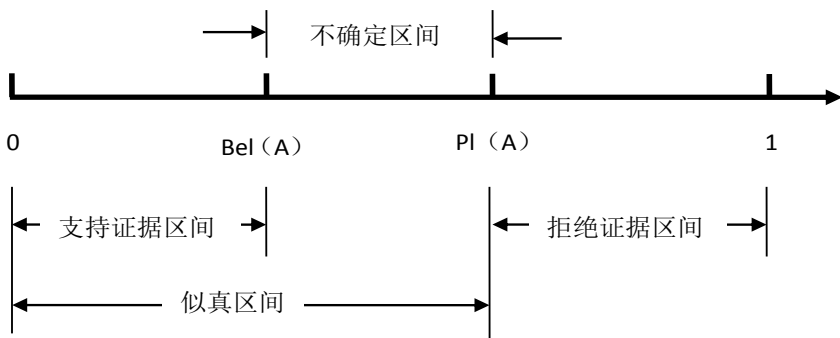
由于并非专门“定制数据”，往往所提供的只是间接证据，如何使用间接证据对重大问题形成决策判断，成为基本应用逻辑问题。

间接证据应用理论与信息融合技术的关联

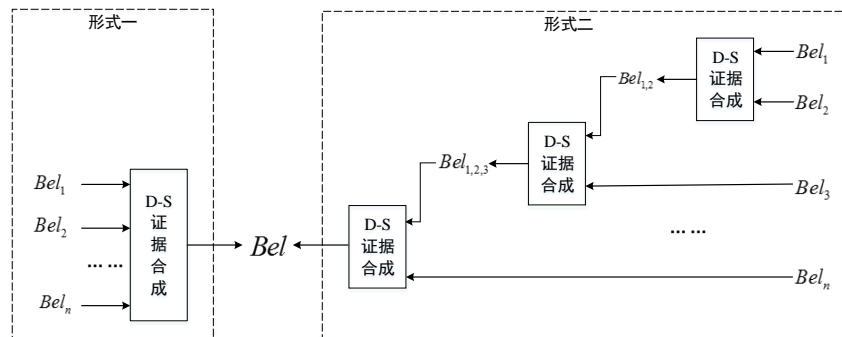
	数据层融合	特征层融合	决策层融合
证据可信度评估	**	***	
证据间关联辨识		***	***
多角度证据汇集		**	***
构建证据链			***

将数据组织成为信息，从信息中提取特征，依据特征形成判断证据。

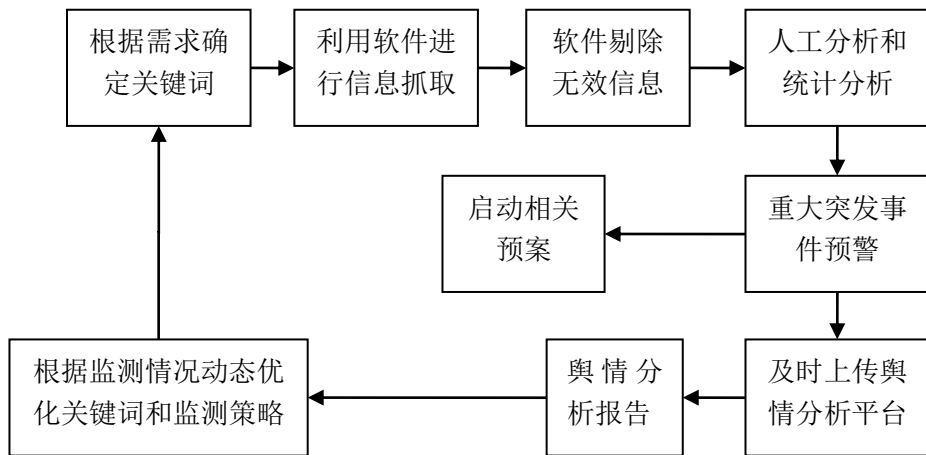
具有盖然性的证据区间



基于专家群体的多个证据合成



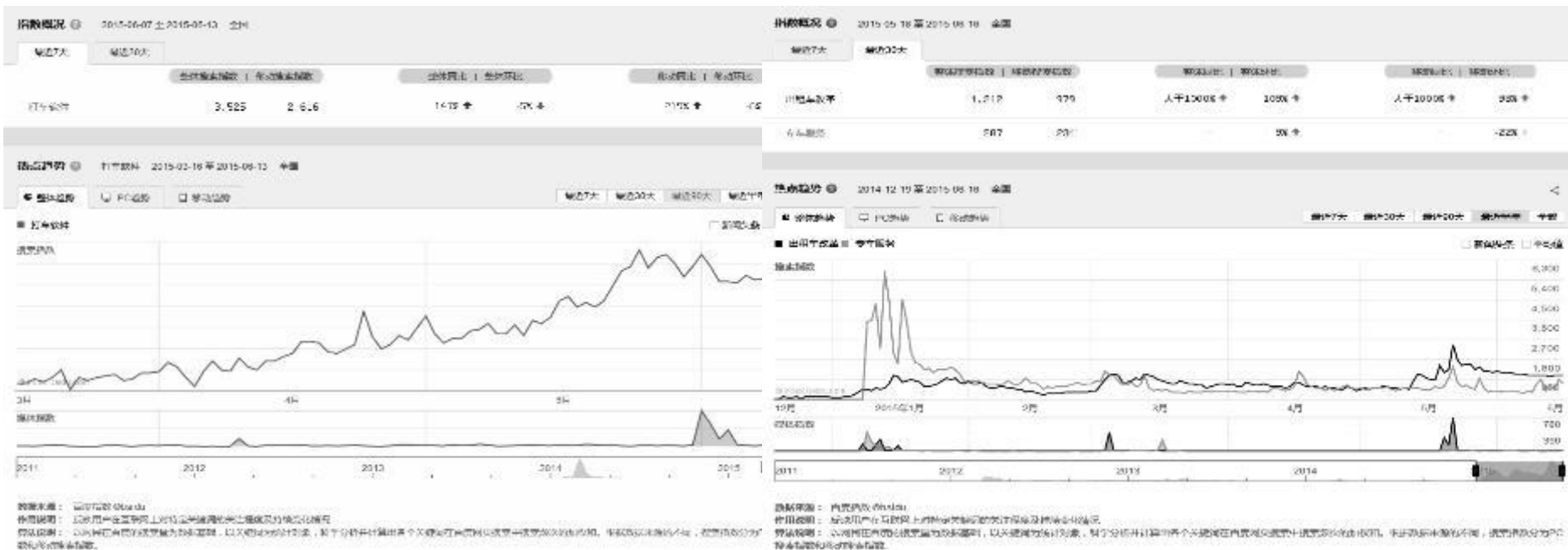
舆情分析：管理需要心中有数



舆情分析的流程



网络舆情的形成过程

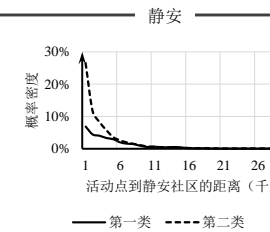
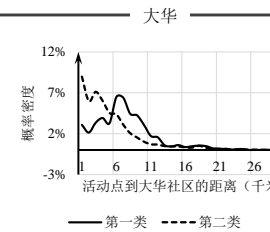
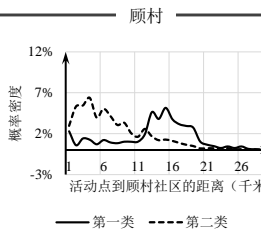
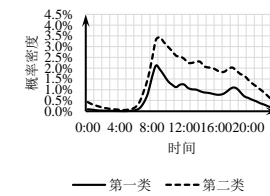
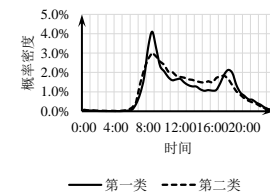
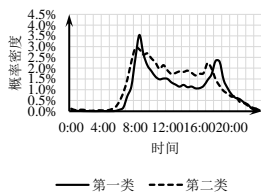
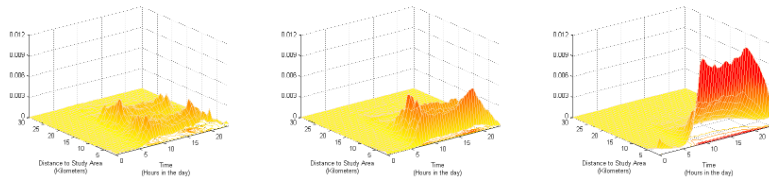
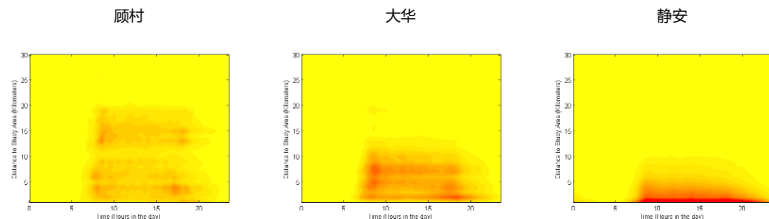


行动：技术与管理的协同变革

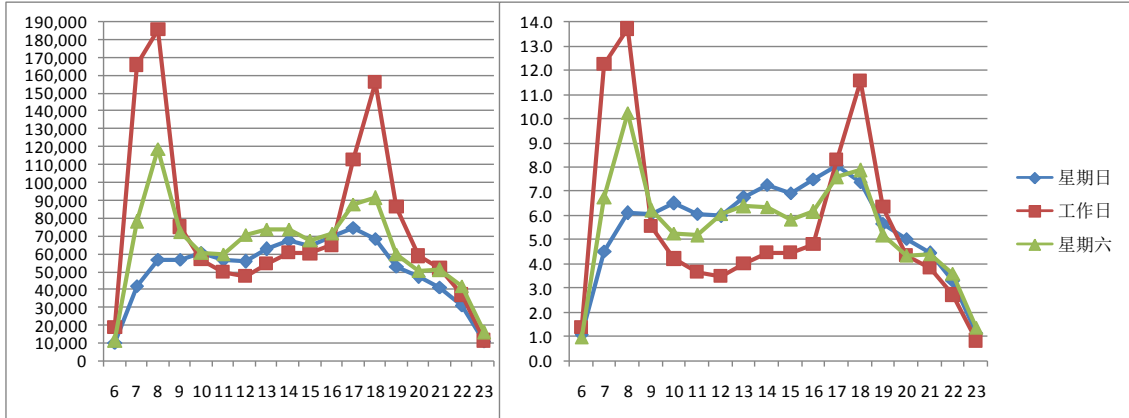
更加准确、完整、清晰地设定目标

- ◎ 减少交通工具产生的传统排放物，达到不会对任何地方的公共健康产生严重危害的程度；
- ◎ 把交通所产生的温室气体排放限制在可持续的程度；
- ◎ 大量减少全世界由于道路交通事故而死亡和受伤的人数，特别是需要关注交通能力迅速提高的发展中国家；
- ◎ 降低交通所产生的噪声；
- ◎ 缓解交通拥堵；
- ◎ 缩小在“发达”与“欠发达”、“贫困”与“非贫困”之间的机动化分化；
- ◎ 提高大众的机动化机会。

目标导向1：城市空间联系结构

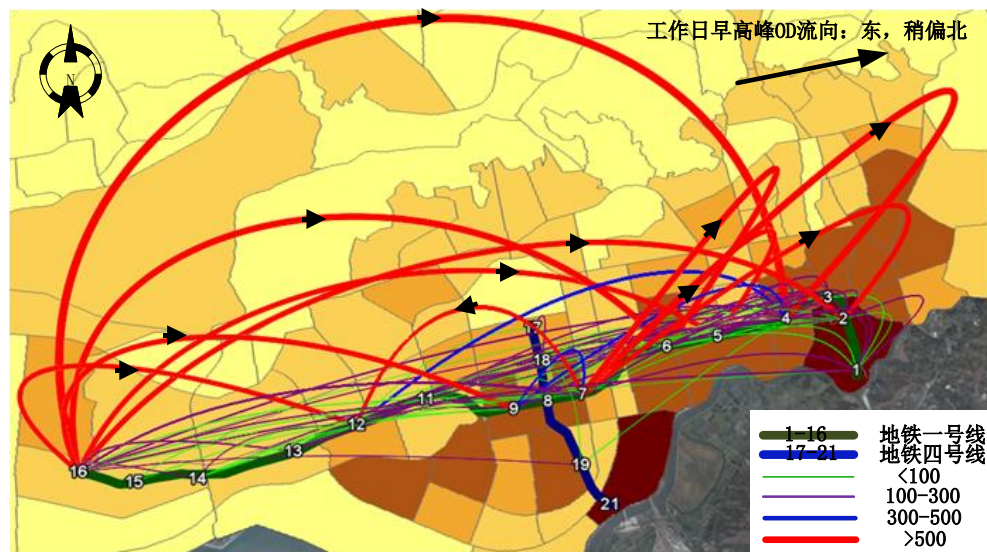
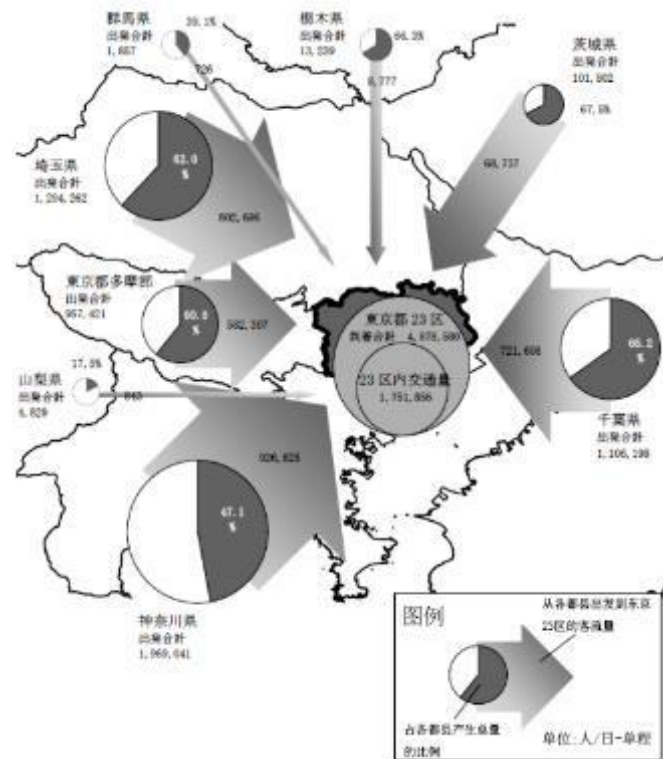


目标导向2：建设公交都市

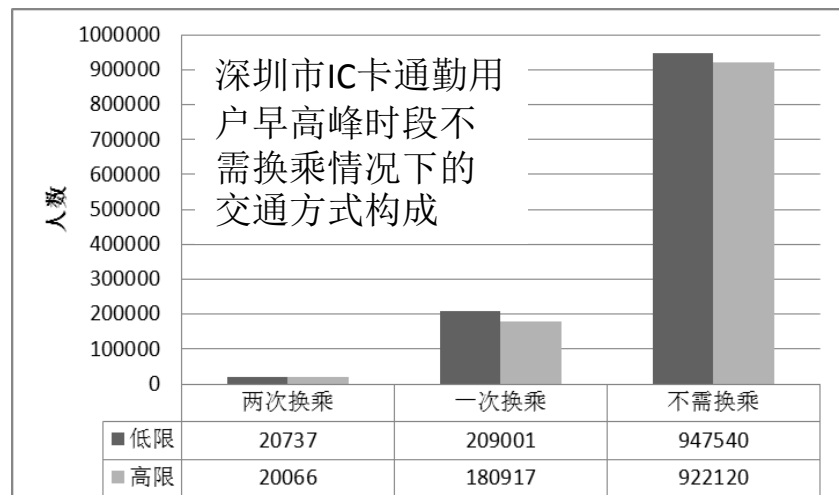


(a) 出行总量(人次) (b) 出行比例(%)

公交使用人数时变特征

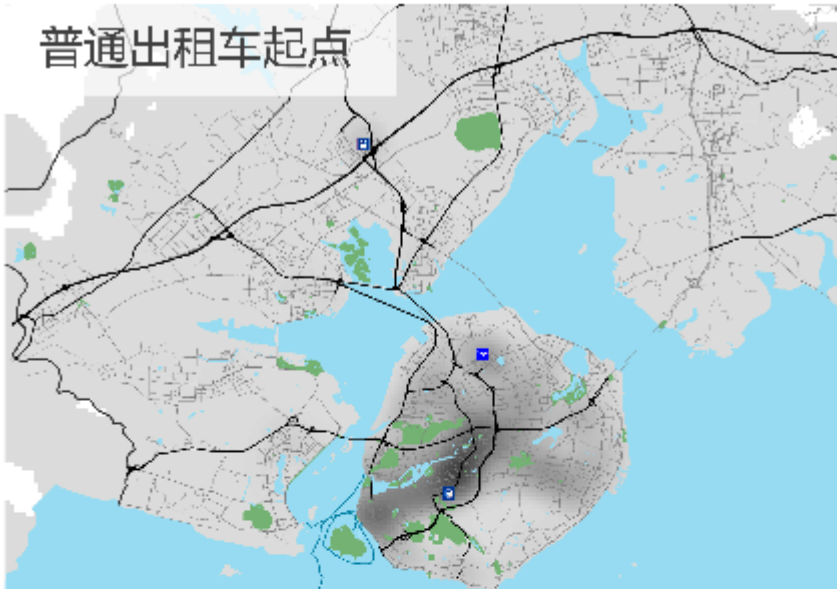


通勤交通公交OD Commute Transit OD

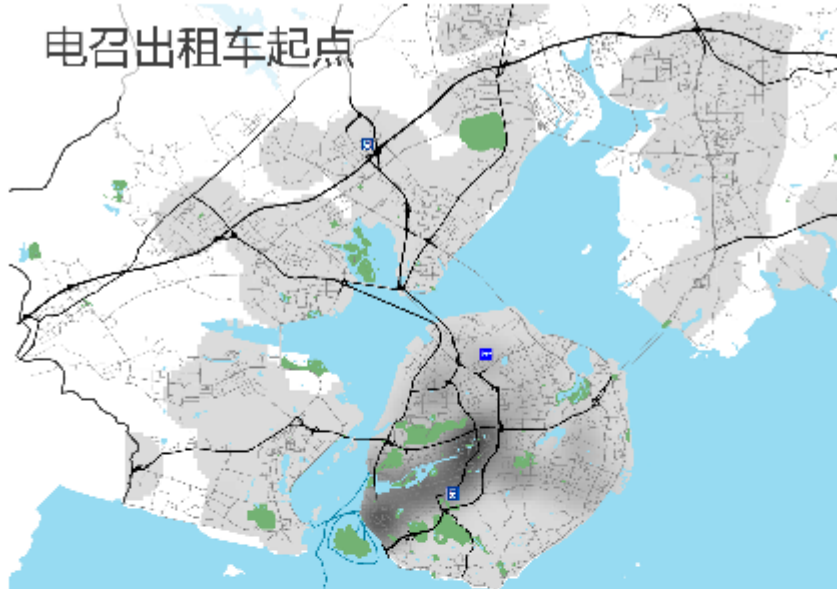


目标导向3：精细化行业管理

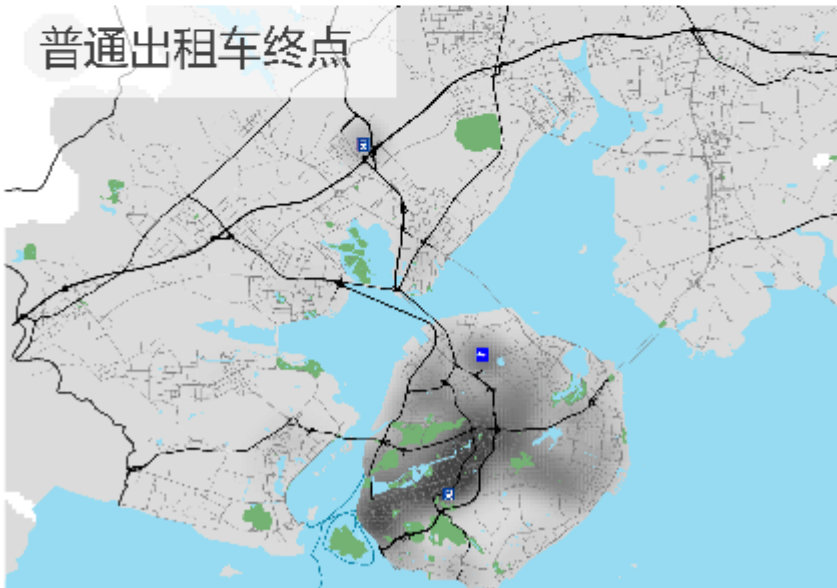
普通出租车起点



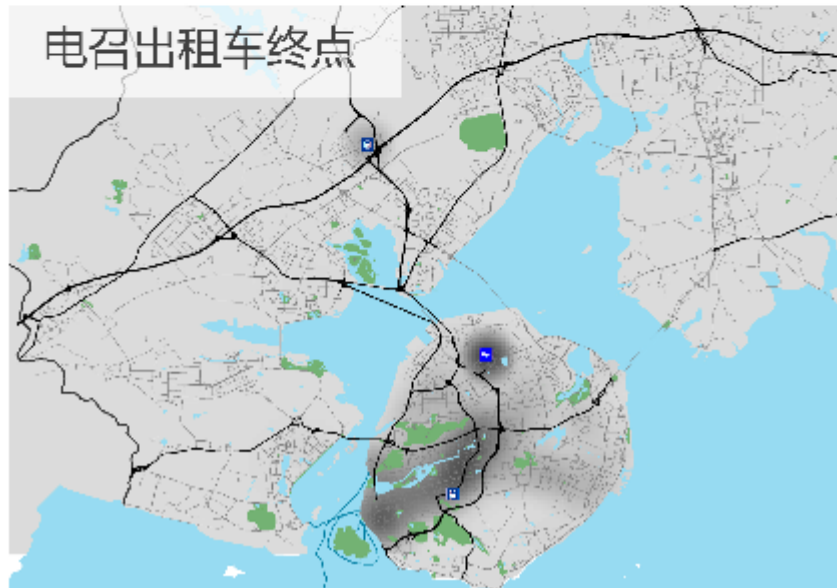
电召出租车起点



普通出租车终点



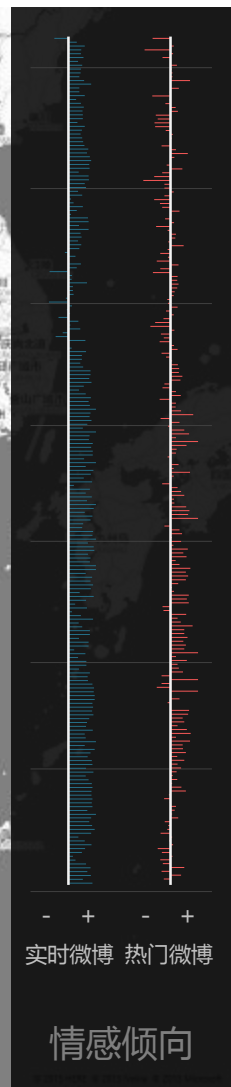
电召出租车终点



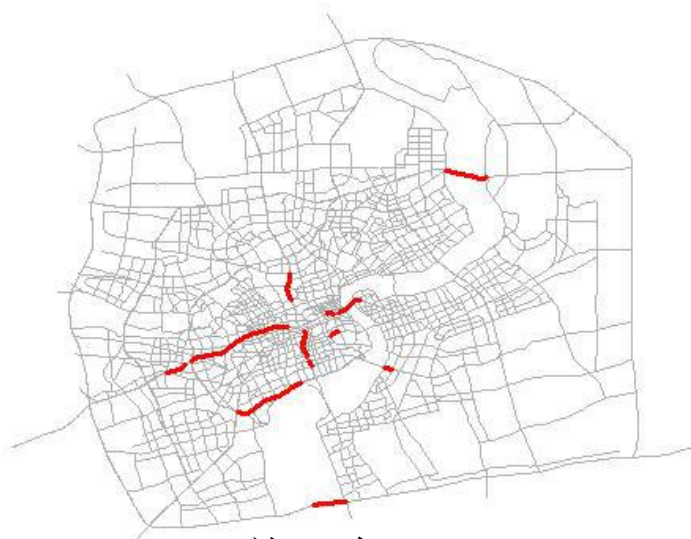
目标导向4：支持交通政策制定

“专车” 相关热点话题

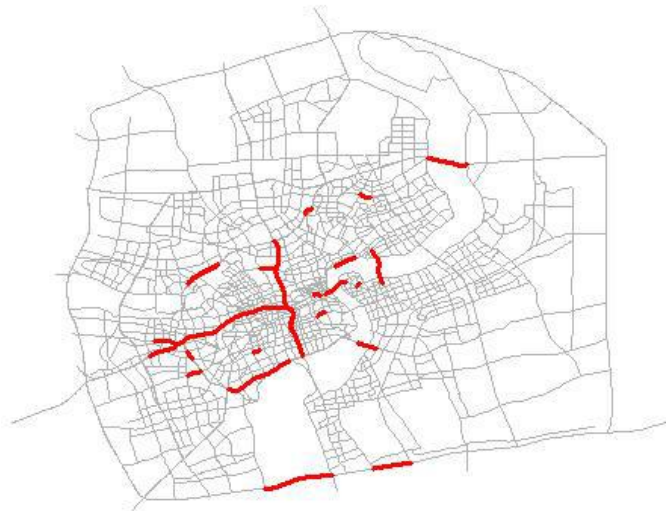
微博关注度·情感倾向



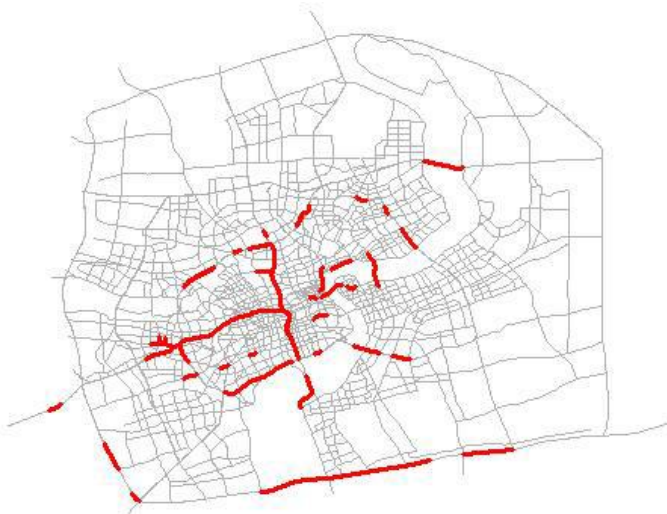
目标导向5：精细化的拥堵治理



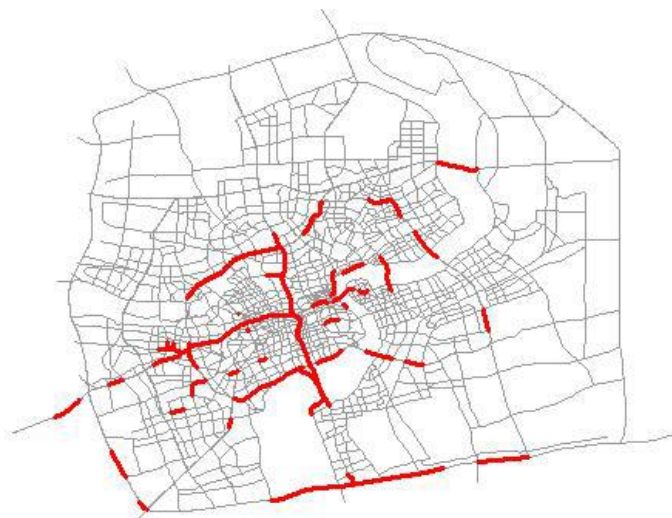
前50个



前100个



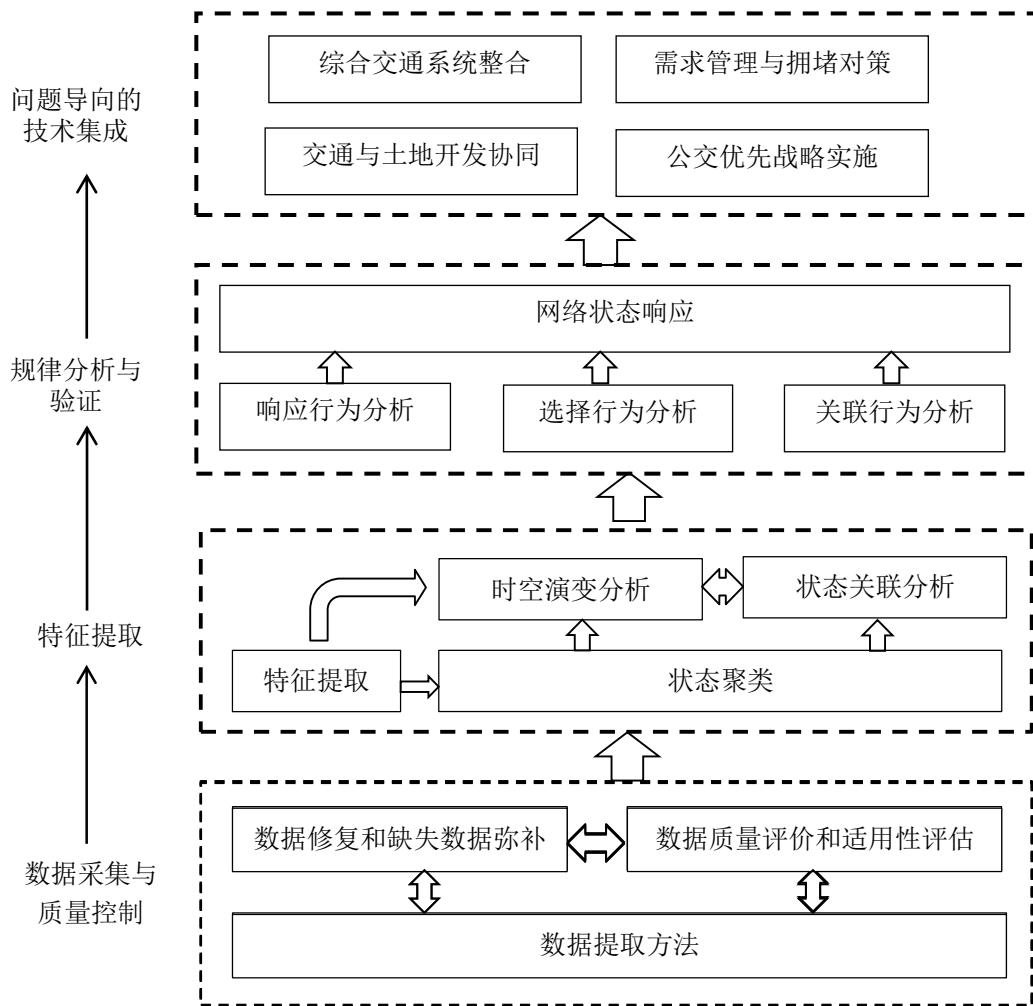
前150个



前200个

哪些路段拥堵
对全路网拥堵
影响大？

从数据加工流程看交通运输领域内的大数据技术应用



从本质上来说，大数据环境下城市交通分析技术所完成的是一种将数据组织成为信息，从信息提炼特征，从特征变化中发现规律，就对策进行追踪评估的信息处理过程。

从信息处理角度来看，可以划分为数据采集与质量控制层、特征提取层、规律辨识与分析层和问题导向功能综合层四个层次，以及通过功能衔接和信息传递实现的有机整合。

没有结束的结束语

- 大数据的价值，在于为我们提供一个更为深入了解城市交通的环境。
- 大数据带来的不仅是技术，而且将引发思维模式和决策模式的变革。
- 应用大数据是为了发现我们所不知道的，而不是验证或者重复我们已经知道的，不要受到传统思维的束缚。
- 在城市交通领域，大数据需要与传统技术一起，融合在一个新的决策分析框架之中。

希望了解具体技术，请参考《大数据环境下城市交通分析技术》，同济大学出版社，2015年1月



透过大数据
把脉城市交
通（写作
中，预计
2016年出
版）

大数据与居
民空间活动
分析（写作
中，预计
2016年出
版）

个人观点，“抛砖引玉”，谢谢！

YANGDYK@TONGJI.EDU.CN