

人迹地图:数据增强设计的支持平台

Human Activity Map: The Platform for Data Augmented Design

茅明睿 储妍 张鹏英 沈忱

文章编号1673-8985 (2016) 03-0022-08 中图分类号TU981 文献标识码A

摘要 手机信令、公共交通刷卡记录等大数据,以及来自商业网站和政府网站的开放数据共同促进了“新数据环境”的形成;时空行为数据是新数据环境下最有价值的数据类型之一,但规划设计行业的新数据应用问题仍然没能得到有效解决。分析了数据来源、分析工具和应用结合3方面的难点,在获取了全国范围的互联网LBS (Location Based Service) 数据的基础上,对其从时间、空间和组群3个维度进行了解构和重组,率项目团队开发了人迹地图时空行为规划分析平台,提供了人流分析、锚点分析、人群分析和城市画像等功能,编写了基于人迹地图平台的《数据增强设计手册》框架,初步支撑了若干个规划项目,并探讨了人迹地图在未来规划设计中的价值。

Abstract ‘Big data’ such as mobile phone signaling, public transportation smart card records and ‘open data’ from commercial websites and government websites jointly promote the formation of the ‘new data environment’. Spatial-temporal behavior data is one of the most valuable types in the new data environment, but the problem of new data application in urban planning and design still can’t be effectively resolved. This paper analyzes three difficulties: the data source, analysis tools and applications combination. Through cooperating with internet DMP (data management platform), the author obtains internet LBS (Location Based Service) data of 3 billion mobile devices and deconstructs and reorganizes the data from three dimensions: the time, the space and the group and develops a planning analysis platform: human activity map. The platform provides analyses about stream of people, anchor, groups and city profile. The author writes the frame of Data Augmented Design Manual based on the platform of human activity map, assists of several planning projects, and discusses the value of human activity map in the future planning.

关键词 时空行为 | 数据增强设计 | 人迹地图 | 新数据环境

Keywords Spatial-temporal behavior | Data Augmented Design | Human activity map | New data environment

作者简介

茅明睿

北京市城市规划设计研究院北京市城规技术服务中心

副总经理,高级工程师,硕士

储妍

北京市城市规划设计研究院 规划信息中心

助理工程师,硕士

张鹏英

北京城垣数字科技有限责任公司

工程师

沈忱

北京交通大学

博士研究生

0 引言

随着ICT技术 (Information and Communication Technology,信息与通信技术) 的发展——尤其是互联网的迅猛发展、开放数据运动的推动和志愿者数据的出现,数据的可获得性大为增强。手机信令、公共交通刷卡记录等大数据,以及来自商业网站和政府网站的开放数据共同促进了“新数据环境”的形成^[1]。有别于传统的来自于官方渠道的地形图、统计资料,这类新数据不仅能够对传统数据形成一定程度的替代,并且,来自于手机信令、互联网用户行为数

据等的个体数据还能够在微观尺度以及更丰富的维度上描述人的行为活动和个体特征,所以新数据环境的出现对于规划设计工作而言,不仅意味着对传统数据的依赖性降低,同时也意味着新的分析视角出现,以及由此产生的潜在的规划设计方法变革。为此龙瀛等提出了数据增强设计 (Data Augmented Design, DAD) 这一规划设计新方法论^[2]。

在新型城镇化的背景下,原有的宏大愿景式规划正在悄然改变,大量扩张型的增量规划在向存量规划转变。在这一过程中,规划的内

涵、价值、类型、形式、路径、对象、市场和技术手段都在发生深刻的改变:从单纯关注物质空间,变成全面关注市民、经济、社会、生态与空间的可持续发展。时空行为研究不仅为理解人类活动和城市环境在时空上的复杂关系提供了独特的视角,同时已经发展成为城市规划学一种很具影响力的研究方法^[3-5],时空行为研究使城市研究者能够从个体日常生活经历的视角理解中国城市转型的过程和结果^[6],成为理解中国城市社会转型的行为范式^[7],弥补传统城市规划对居民时空行为的规律与决策机制考虑的不足,传统的基于土地的、静态的、蓝图式的城市规划亟待与时空行为研究相结合,从而转向基于人的、动态的、精细化的规划^[8],因此时空行为研究和应用实践也成为数据增强设计方法变革中备受关注的组成部分。

1 难点

尽管近年来在规划设计领域也出现了不少基于时空行为数据(手机信令、出租车轨迹、IC卡刷卡记录等)的应用案例,例如在上海^[9]和北京^[10]等城市进行的出行调查使学者对于个体的活动类型、出行目的地、出行方式、出行路径有了更为深入和完整的分析,但分析结果发现,在数据精度和有效性上不可避免地存在误差,同时规划设计行业的几个应用问题仍然没能得到有效解决。

1.1 数据来源

开放的数据不大,大数据都不开放,真正有价值的时空行为数据大多都处于非开放状态,比如手机信令数据都由运营商所掌握,虽然可以开展商业合作,但价格不菲,且基本无法让规划设计人员接触到原始数据;IC卡刷卡记录则在政府或者一卡通公司手里,数据协调成本高昂。所以当前的大量城市研究和行业应用的数据基础并不稳固,已有的若干应用案例(比如上海SODA开放数据大赛所公开的某周的出租车轨迹数据、北京某几周的IC卡刷卡记录等)都基于非常偶然的数据样本,或者基于爬取的位置微博、百度热力图等互联网数据来

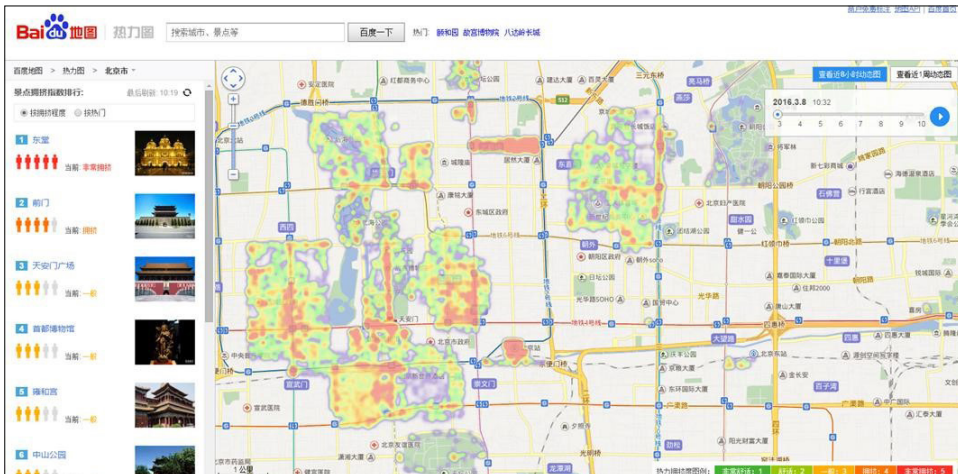


图1 百度热力图
资料来源:百度地图。

开展。

片段性的数据样本对于时空行为研究方法的探索有价值,也能够一定程度上反映某个时期的某个特定城市的空间结构以及规律性的任职关系等问题,但并不具有普遍意义,缺乏全局和组合度量数据的考虑^[11]不能解决数据的覆盖广度和时间跨度等问题。

而位置微博和百度热力图等互联网数据则具有很大的局限性:前者定位过于稀疏,人群样本有偏性严重,而如果数据样本没有代表性,直接利用有偏的样本分析,纵然样本量非常大,不仅无法得出正确的分析,而且无法反映出行链^[12];通常科学的数据具有复杂性、综合性,及高度可利用性的集成性^[13],而后者则只能表现某些城市空间的密度变化,而无法分析个体、群体的空间行为。更重要的是,基于爬取所获得的数据其本身也仅能限于学术研究领域,而无法成为规划设计行业的商业应用基础。

1.2 分析工具

作为大数据的一种,时空行为数据同样具有难以处理、清洗、建模等技术难题,普通的规划设计机构并不具备相应的软硬件架构基础,而普通规划设计人员也基本都不具备相应的IT技能。相比传统的行业软件基础——CAD制图软件和GIS空间分析软件,时空行为数据并没有成型、通用的应用软件平台来供城市规划

设计人员开展工作,已有的应用大多是规划设计机构通过与高校或者数据企业开展案例合作来进行,这种案例级别的专题合作形式也无法让时空行为数据真正成为行业转型的数据基础。

1.3 应用结合

要让数据真正起到增强设计的作用,使这一方法论能够成立,必须将基于物质空间的蓝图式规划设计方法向基于市民和社会的动态规划设计方法转变,而这个转变不仅需要相关理论方法的革新,还需要对规划设计行业的工作流程进行再造。

对于时空行为数据(及其他新数据)的统计、分析和可视化反映的都是现状或某个历史片段里个体、群体的活动情况,若要基于这些数据来辅助,甚至驱动规划设计,仅仅做现状统计、描述是不够的,所以需要行为论、时间地理学、活动分析法、复杂系统等传统城市规划、建筑设计领域相对陌生的理论来发现个体、群体的空间偏好,分析人群集聚、扩散规律,挖掘个体行为模式,探寻行为与空间互动关系,从而在传统规划设计工作基础上形成一套依托传统数据和新数据、针对不同规划设计工作类型的完整的工作流程和解决方案,而这个过程不可能一蹴而就,需要时间去实现数据积累、样本收集、理论引入、实践尝试、流程再造和形成解决方案。



图2 LBS数据的维度结构
资料来源：作者自绘。

1.4 小结

上述3大问题成为了规划设计行业在新数据环境下产业升级的3大痛点,不能解决这3个层面的问题,“新数据环境”仅能具有研究意义,数据增强设计也仅在个案情景下有效,而无法成为普遍的生产力。所以笔者认为要从数据基础、面向规划设计行业的分析工具平台建设,以及数据增强设计解决方案3个角度入手实现规划设计行业的技术升级,让新数据成为行业的生产资料,让数据增强设计的方法成为行业的生产力。

2 数据基础

因为上述种种原因,无论在城市规划领域还是交通规划领域,规划设计人员都将过多的精力放在协调资源上,而无暇顾及深层次的问题;已有应用局限在现状描述上,而没有触及规划的真问题^[14]。针对这个问题,笔者认为亟需建立一个面向规划设计行业的时空行为数据平台,降低行业的数据获取、应用成本和使用门槛,让新数据成为规划设计行业的生产资料和生产力,解放规划设计人员和城市研究者的精力,让他们专注到规划真问题的解决上来。

笔者对国内主要的的数据服务商进行了调查和沟通,并获得了TalkingData、中国联通等企业的合作,加上部分城市的IC卡刷卡记录以及位置微博等时空行为数据,为规划设计行业建立一个覆盖全国的时空行为分析平台奠定了数据

基础。

在这些数据中,当前最重要的组成部分是来自TalkingData的LBS数据。TalkingData是一家第三方移动DMP(Data Management Platform)平台,在其平台上拥有十余万款APP的应用数据,是目前互联网上唯一商业开放的能覆盖全国的时空行为数据源。在TalkingData平台上有超过30亿台移动智能终端的数据,其日活跃用户数超过1亿,月活跃用户数超过6亿,日定位请求超过30亿次。

与运营商的数据不同,互联网的LBS数据不仅包括来自于基站的定位,同时还包括GPS和Wi-Fi等其他定位方式的数据,由于DMP平台不仅采集定位数据,还会采集用户的其他行为数据,所以DMP平台能够对用户进行维度更加丰富、立体的画像,从而对用户赋予多种多样的标签。TalkingData用户标签包括人口属性、设备属性、地理位置、应用兴趣、游戏偏好、消费偏好、游戏深度、行业标签、定制标签等9大一级类500余二级类标签,其数据规模仍在迅速扩张中。

3 建设思路

在互联网上已有若干基于时空行为数据的应用平台,比如百度热力图(图1)、腾讯的宜出行平台等,这些平台都有着高质量的时空行为数据,但是这些平台大多仅能反映短期内的分时人口密度情况,而百度等与交通、旅游部门合作的应用平台也大多停留在瞬时流量的预警和

短期预测上,所以这些平台仅仅是利用单一维度的人数统计来进行“现象描述”。

对于规划设计而言,将时空行为数据仅用来进行现象描述是远远不够的。无论是手机信令数据,LBS数据还是IC卡刷卡记录,它们都是由“ID、时间和位置”构成的文本记录,在空间上其最直接的表现为一组时间序列的点的集合。笔者认为对这个集合可以有3个不同的维度来进行解构和重构:时间、空间和组群。

3.1 解构

(1) 空间维度

空间并不是没有意义的坐标,它是有异质性的,这种异质性以边界的形式呈现,每个边界都有其自然属性和社会属性。所以在空间维度上,时空行为数据的组织单元可以从微观到宏观,按照网格、地块、街区、乡镇、区县或者城市及以上更宏观的单元来组织(图2)。

(2) 时间维度

在时间维度上,时空行为数据可以分时、分日、分特定时间段来组织,此外通过个体行为的时空分布规律发现其锚点。通过不同的时间段来推测个体、群体的行为类型,定义他们的行为空间,发现个体的锚点,并进一步描绘个体、群体的生活圈。在通勤行为空间里,锚点为居住地和工作地;在日常生活行为空间里,锚点为居住地和常去的购物场所以及接送孩子的学校等;在假日的休闲行为空间里,锚点为居住地以及常去的休闲场所。

(3) 组群维度

在组群维度上,通过画像技术,时空行为数据又可以根据画像标签分为若干类组群,例如人口特征:年龄、性别等;设备特征:机型,语言,操作系统,价位等;行为特征:上班族,居家族,加班族,经常出差者等;以及职业特征和消费特征(表1)。这些特征标签里,有的来自于个体所持设备所直接贡献的信息,有的来自于设备安装的APP以及APP的应用情况,但是与规划设计关系最密切的标签——行为特征,则来自于个体的时空行为建模。比如笔者曾利用北京地铁IC卡刷卡记录识别了地铁里行为异常的481

个灰色人群^[15]（乞丐、发小广告、快递等），龙瀛等曾利用北京公交IC卡刷卡记录识别了18.9万个低收入持卡人^[16]。

3.2 重构

通过3个维度的解构，一组看似意义单一的时空日志文本变成了具有多种组合条件的数据集。它能够帮助规划设计人员在不同的时间范围上去观察城市时空异质性和人群异质性，发现不同组群的职住分布特征、联系特征和行为模式，了解他们空间选择偏好以及变化趋势。比如不同国籍的人都在城市里哪些小区居住，从事何种工作，喜好城市的哪些场所（图3）等；或者软件开发者群体都住哪些小区（图4），他们的生活圈又是如何构成的，他们与商务人士的居住、就业和日常活动场所的差异如何等。

另一方面，人的行为也可以反过来定义城市的空间，过往的城市规划、土地管理领域，城市的空间被自然属性、功能属性和权属性质所定义，规划设计人员对于空间最熟知的定义方式是用地性质。而时空行为数据赋予了从“行为空间”视角——也就是从人的活动来定义城市的全新条件。比如根据空间单元里不同国籍、不同地区的人群的活动强度来计算城市单元对不同国籍、不同省份人群的吸引力；根据单元居民的加班频次，计算其加班率；根据单元居民的出差频次，计算其差旅度；根据单元居民的出行目的地的不规律性，计算其活跃度；根据夜间单元里非本单元居民的密度来计算该区域的夜生活服务度；根据单元居民的通勤距离、通勤结构和内外通勤比，计算其职住、通勤特征；根据单元居民在春节、长假的去向，计算单元与其他区域的联系度……这些指标可以有助于规划设计人员来用于精细化的评估城市中微观空间的活力、品质，以及其他无法用测绘、遥感等传统手段观察的动态特性。基于这些行为活动特性，结合城市的区位、用地、景观、建筑、路网、步行可达性、公交可达性、公共服务设施分布、产业分布、房价等其他特性，使规划设计人员和城市研究者有可能



图3 2015年8月韩国人在北京活动位置热力图
资料来源：人迹地图平台。

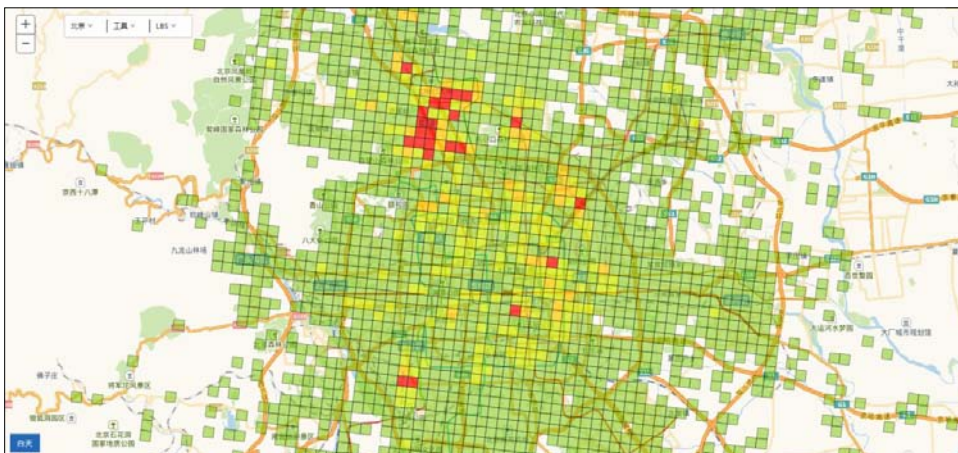


图4 研发设计人群在北京的居住地分布图
资料来源：人迹地图平台。

表1 人群标签体系

标签类型	标签	选项
人口属性	性别	男，女，未知
	年龄段	老，中，青，少，未知
	婚否	是，否，未知
	育否	是，否，未知
	有车族	是，否，未知
设备属性	价位	1 000元以下，1 000—1 999元，1 999—2 999元，3 000元以上
	系统	安卓，苹果，其他
行为属性	涉外人群	英系，日系，韩系，法系，德系，西系，中繁，其他，否
	居住地、就业地	坐标
	居住城市、就业城市、老家	城市名或者行政区划代码
	常出行方式	公共交通，出租车，私家车，非机动，未知
职业属性	身份标签	工作人员，大学生，中小学生，上班族，低收入人群，灰色人群，夜行者，居家族，加班族，二居所人士，早起者，晚归者，长距离通勤者……
	职业	公务员，科研，体育，医生，文化，宗教，商业，艺术传媒，商务，社会服务，教育，军人，外事，安保，乡镇企业，农业，矿业，工业，研发设计，交通，市政，仓储物流……



图5 人迹地图平台结构
资料来源:作者自绘。

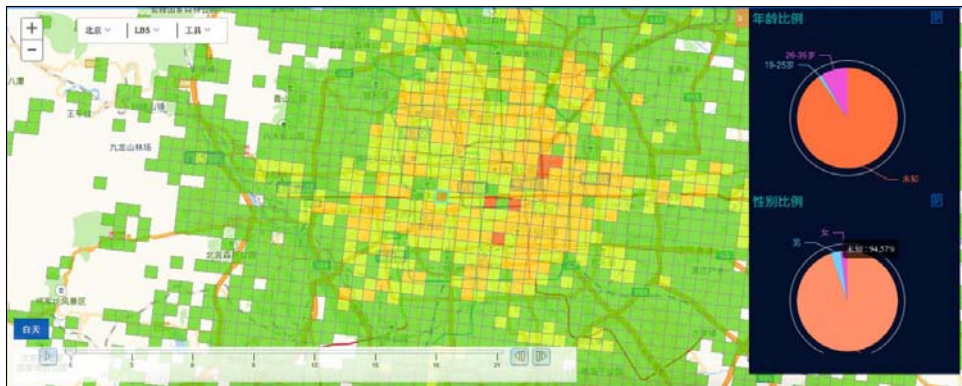


图6 北京2015年某日瞬时人流密度分布图
资料来源:人迹地图平台。

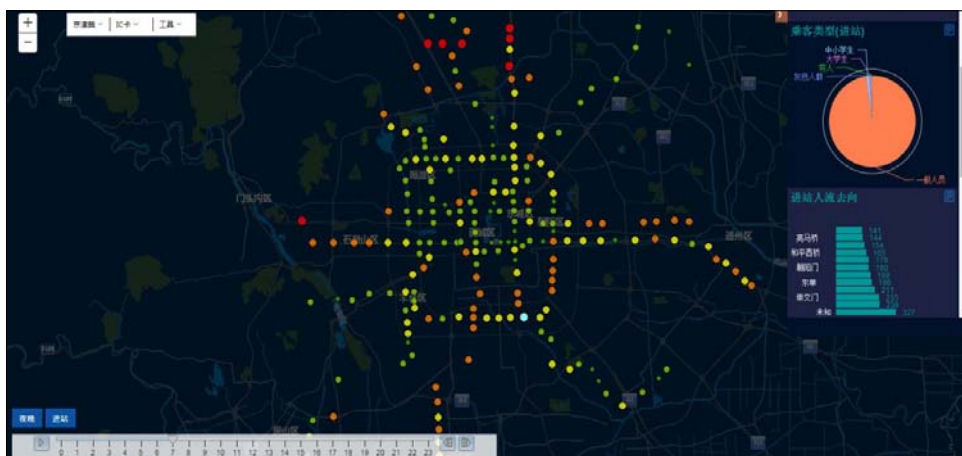


图7 2015年某日地铁站进站人流分析图
资料来源:人迹地图。

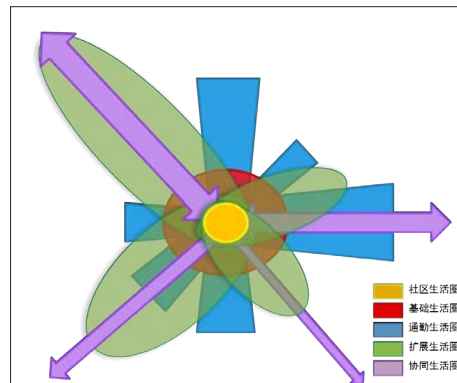


图8 基于锚点的生活圈示意图
资料来源:作者自绘。

建立起空间、人群、行为和活力之间的关系。这个指标体系的建立和计算过程，笔者将其称为“空间画像”。

4 平台功能

根据上述思路，项目团队进行了平台的研发，基于大数据的处理技术，对数据进行抽取，空间统计和建模画像，并对空间单元计算指标，通过人迹地图时空行为规划分析平台发布。人迹地图平台目前具有人流分析、锚点分析、人群分析、城市画像等模块（图5）。

4.1 人流分析

人流分析功能提供了人流密度分时变化特征。通过人迹地图的人流分析功能，可以查看全国的任何一个地方随着时间变化的人口密度分布，以及该单元的人口瞬时结构特征（图6），例如年龄、性别、来源等属性；此外，平台还可以集成多源数据：北京的IC卡刷卡数据，可以按时间查询各站进、出站人流数据：某时刻进站人数，进站客流的去向，出站人数，出站客流的来源（图7），以及人群的类型（例如普通乘客、大学生、中小學生、低收入人群、灰色人群等）。

4.2 锚点分析

北京大学柴彦威教授总结了生活圈规划思想，通过社区居民的出行行为日志，观测居民的出行规律和锚点，界定出不同的空间范围，并赋予其特殊的意义^[17]。依托该思想，通过锚点绘制

空间单元的居民群体的不同生活圈（图8）。

(1) 社区生活圈,即市民的居住社区范围,由居民的居住锚点结合空间特征聚类形成;

(2) 基础生活圈,即市民日常生活的基础生活范围,由居民居住锚点和购物等锚点构成;

(3) 通勤生活圈,即市民居住社区与工作地之间构成的通勤范围,由居民居住锚点和就业锚点构成;

(4) 扩展生活圈,即市民居住社区与他们中长距离出行探亲访友、休闲旅游的活动范围,由居民居住锚点和长距离休闲锚点构成;

(5) 协同生活圈,即市民居住社区与商务联系经常出差、或者回老家等多次反复长距离出行构成生活圈范围,由居民居住锚点与商务出差或回老家等长距离出行锚点构成。

按照不同的空间单元对单元内的人群锚点进行统计,从而形成不同单元间的住一职、住一闲(职一闲)、住一假(职一假)等联系特征。用户可以查看居住在不同单元的人群的就业地、休闲地、度假地的分布,或者在不同单元就业的人群的居住地、休闲地、度假地分布,以及查询单元间的联系强度(人数)和单元的各类指标。联系特征通过色差图、连线形式等不同可视化方式来展现。

单元的指标则包括人口结构,比如性别结构、年龄结构等;职住通勤结构,比如:平均通勤距离,职住比,内外通勤比(图9);职业结构,比如单元内不同职业的人数和比例等(图10)。

4.3 人群分析

人群分析可查询不同标签人群的居住地、就业地分布以及典型工作日、休息日的空间轨迹,通过空间可视化展现人群的异质性(图11)。

4.4 城市画像

城市画像通过空间专题图和仪表盘对城市单元的各类指标进行可视化展现,描绘城市不同区域的空间异质性。指标从行为空间、景观、用地等视角来组织(图12)。

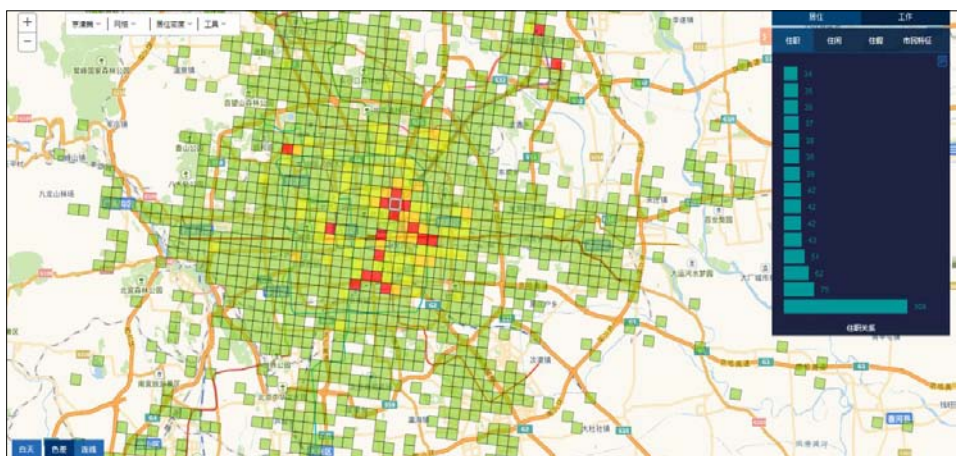


图9 京津冀地区某网格单元的职住色差图
资料来源:人迹地图。

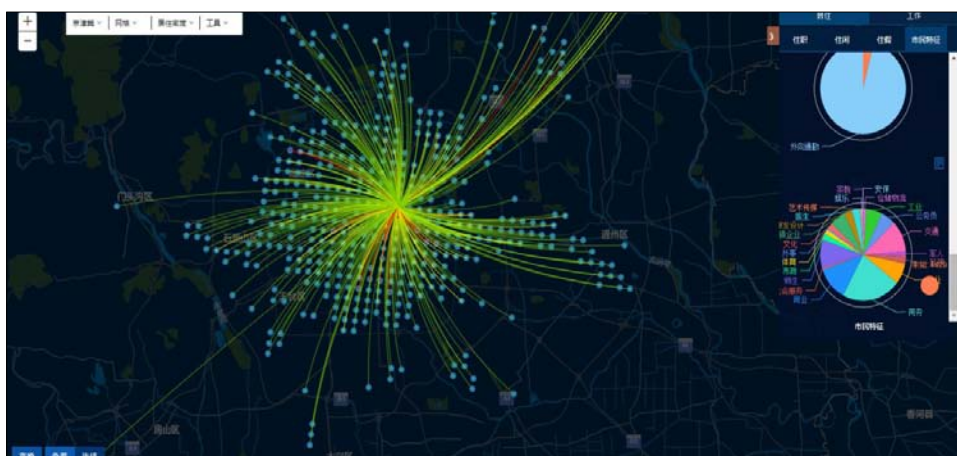


图10 京津冀地区某网格单元的职住特征连线图
资料来源:人迹地图。



图11 人迹地图的人群分析功能展示
资料来源:人迹地图。

5 平台应用

人迹地图平台是互联网上第一个全国范围的时空行为规划分析平台,由于其覆盖全国所有城市的数据,属性维度也比较丰富,所以具备

了按照任意单元、任意属性维度开展数据分析,以及多个地区间进行横向对比的条件。目前平台已在“通州行政副中心规划研究”、“顺义城市品质提升研究”等项目中得到应用(图13)。

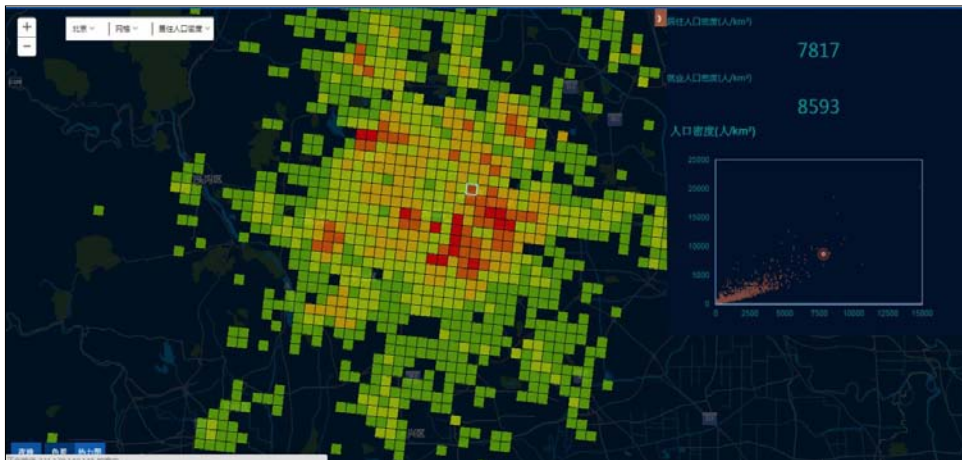


图12 城市画像:人口密度指标
资料来源:人迹地图平台。



图13 早晚高峰顺义地铁进、出站客流情况
资料来源:人迹地图平台。

在上述项目中,规划师利用人迹地图平台的锚点分析、人流分析和语义分析等功能,基于LBS数据对通州、北三县和顺义的职住关系、基于IC卡刷卡数据对轨道交通的通勤情况和基于微博语义对城市的消极空间等内容进行了分析。

为了实现对规划设计的有效支撑,项目团队正在结合具体项目,以人迹地图平台为应用载体,编写《数据增强设计手册》(以下简称“手册”)^[18],针对不同规划层次,从现状分析、方案设计到评估运营等阶段分别提出开展规划设计的工具、方法、指标等内容,使之成为新数据环境下规划设计的完整解决方案。手册选取用地、区位、空间结构、人口、就业、居住、功能品质、产业、交通等要素类别,通过前期分析、方案设计、评估运营3个流程来指导用户基于数据进行规

划设计(图14)。

6 结论

目前人迹地图平台还处于研发阶段,无论是功能层面、数据层面还是数据增强设计的解决方案层面都还有非常多的工作要做,比如数据源的有偏性问题亟待解决:互联网LBS数据与其来源——互联网DMP平台所管理的APP种类、APP的覆盖客户群体以及APP的活跃程度高度相关,目前来看其覆盖的人群偏一线城市的年轻人,不能代表全体市民,在不同区域其覆盖人群的比例也不一样;而IC卡刷卡记录也仅能代表乘坐公共交通的人群。但是这是规划设计行业走出的第一步,利用本专业的领域知识,规划设计人员有能力对海量的时空行为数据进行处理、组织、加工、统计,并通过建模计算

出对城市研究有价值的标签和指标,随着数据源的丰富和完善,理论上可以实现对任意时空范围的分布、联系、人口和空间特征的可交互式的提炼、统计、输出和可视化,这使行业的应用平台比现有的互联网时空平台从简单的“现象描述”更进一步,达到较复杂的建模和“特征提取”层面。

随着时空行为数据的沉淀积累,多源时空数据的融合,以及相应时期的其他数据——POI数据、政策数据、现状用地数据、规划许可数据、企业数据等,动态跟踪规划项目、公共政策的变化和行为空间的指标以及人群结构的变化,平台就有了发现时空行为与城市、规划、政策、经济等相关规律的可能性,从而向“规律发现”发展。

大数据——包括时空行为数据及其有关技术,在现阶段并不具备对城市这样的复杂系统进行远期预测的能力,目前已有的预测算法都是基于历史统计规律对短期行为(比如人流量与流向)、个体的出行方式和位置等进行预测,这类预测与规划项目、公共政策实施后的结果预测或者预评估是不同的,所以平台对于规划设计的“增强”意义并不在于为规划设计人员提供了预测未来的手段,而是通过动态的指标对规划项目的前期研究、目标选择、实施过程和运营管理进行跟踪、评估和反馈,协助规划设计从蓝图式规划到动态的、实施性、政策性规划转变。

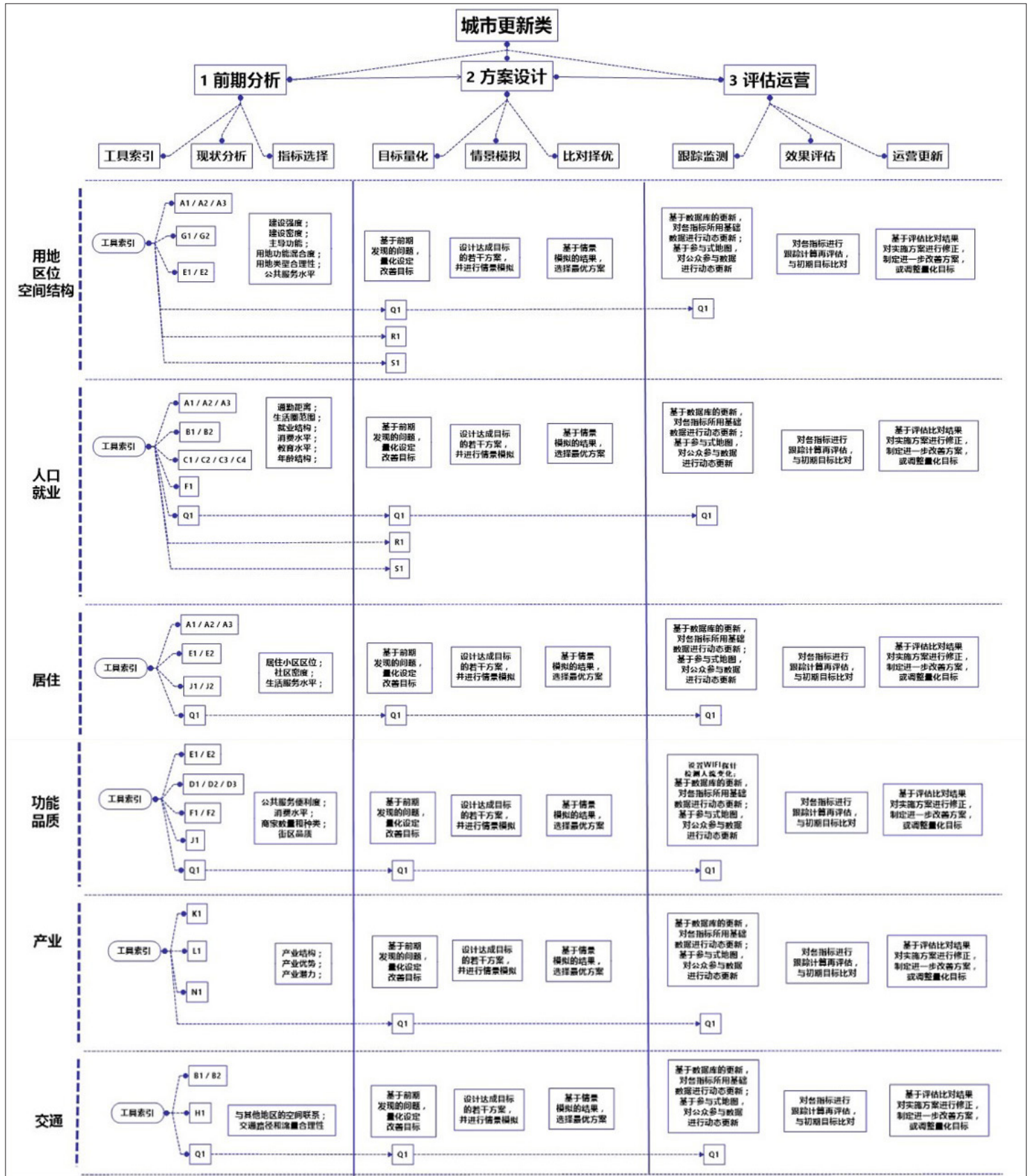


图14 《数据增强设计手册》
资料来源:参考文献[9]。