

引用格式:王姣娥,杜方叶,靳海涛,等.基于交通出行链的就医活动识别理论框架与方法体系[J].地球信息科学学报,2020,22(4):805-815. [Wang J E, Du F Y, Jin H T, et al. Identifying hospital-seeking behavior based on trip chain data: Theoretical framework and methodological system[J]. Journal of Geo-information Science, 2020,22(4):805-815. ] DOI:10.12082/dqxxkx.2020.190566

# 基于交通出行链的就医活动识别理论框架与方法体系

王姣娥<sup>1</sup>,杜方叶<sup>1,2\*</sup>,靳海涛<sup>3</sup>,刘 瑜<sup>4</sup>

1. 中国科学院地理科学与资源研究所 中国科学院区域可持续发展分析与模拟重点实验室,北京 100101;

2. 中国科学院大学资源与环境学院,北京 100049;3. 北京信息科技大学计算机学院,北京 100085;

4. 北京大学遥感与地理信息系统研究所,北京 100871

## Identifying Hospital-seeking Behavior based on Trip Chain Data: Theoretical Framework and Methodological System

WANG Jiaoe<sup>1</sup>, DU Fangye<sup>1,2\*</sup>, JIN Haitao<sup>3</sup>, LIU Yu<sup>4</sup>

1. Key Laboratory of Regional Sustainable Development Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; 2. College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China; 3. School of Computer, Beijing Information Science and Technology University, Beijing 100085, China; 4. School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China

**Abstract:** Transportation is an important tool and carrier for people to realize their trip purposes. Thus, it's a vital measurement for studying spatio-temporal pattern of individuals. Trip chain refers that a series of displacements completed by an individual in order to do one or more activities using a transportation. The time period of trip chain is one day. There are lots of information on individuals' trip purpose contained in trip chains. Extract this information from trip chains help to explore individuals travel behavior, which help understand the urban space. In previous studies, researchers have been focused on inferring and exploring the dynamic characteristics of commuting behavior, go to school activities, go home activities, entertainment, and leisure activities in urban space with the help of smart card data and taxi trajectory data. But limited studies have been carried on detecting hospital-seeking behavior with the assistance of trip record. With this in mind, this paper attempted to extend the application of trip records on hospital-seeking behavior. Specifically, we proposed a theoretical framework used to detecting hospital-seeking behavior from trip record. It consists of six principle, such as proximity principle, ring-closure principle, single-purpose principle, infrequency principle, time-coherence principle, and accompany principle. Also, a methodology for detecting hospital-seeking behavior was put forward based on the theoretical framework. Taking Beijing as an example, we found the key parameters of detecting hospital-seeking behavior and calibrated their thresholds. Finally, spatial and temporal patterns of hospital-seeking behavior were explored to reveal the accuracy of our results. On the one hand, the spatial patterns of hospital-seeking behavior showed that patients were mainly concentrated in tertiary hospitals. Tertiary hospitals have better professional skills and a

收稿日期:2019-09-30;修回日期:2019-11-06.

基金项目:中国科学院战略性先导科技专项(A类)(XDA19040402);国家自然科学基金优秀青年基金项目(41722103)。

[ **Foundation items:** Strategic Priority Research Program of the Chinese Academy of Sciences, No.XDA19040402; National Natural Science Foundation of China, No.41722103. ]

作者简介:王姣娥(1981—),女,湖南涟源人,博士,研究员,主要从事交通地理与区域发展、城市交通大数据等研究。

E-mail: wangjiaoe@163.com

\*通讯作者:杜方叶(1994—),女,山东济南人,博士生,主要从事城市交通、健康地理等研究。E-mail: dufy.18b@igsnrr.ac.cn

larger service area than secondary and primary hospitals. Thus, they attracted and served more patients. On the other hand, patients' arrival time shows a high peak during 8:00 am and 10:00 am and a low peak during 13:00 pm and 15:00 pm, which closed to start time of registration and treatment. Two aspects above both supported the accuracy of results and rationality of the theoretical framework. The application of trip chains on detecting hospital-seeking behavior could make up for the shortages of traditional data, which is a small sample and difficult to access. This paper provided a new perspective, methodology, and data source for researching hospital-seeking behavior. Moreover, it could provide methodology references for other activities based on trip records.

**Key words:** big data; bus; taxi; hospital; trip purpose; trip chain; Beijing

**\*Corresponding author:** DU Fangye, E-mail: dufy.18b@igsrr.ac.cn

**摘要:**交通是人们实现出行目的的重要工具和载体,也是研究城市居民出行目的的重要手段。本文试图采用交通出行数据来识别就医活动目的的行程,以深化交通大数据研究的应用领域。在合并交通出行链的基础上,构建了就医活动识别的理论框架和方法体系,提出6大准则:邻近性准则、出行链闭合准则、单一出行目的准则、时间耦合性准则、路径偶发准则。以北京市为例,基于公交车刷卡和出租车GPS数据,明确就医出行的关键参数与阈值,最终甄别出以就医为目的的交通出行链,并对识别结果进行分析与验证。基于交通出行链的就医活动识别研究可以弥补传统研究中病例数据和问卷数据样本量小和难获取的不足,为就医活动研究提供了新的方法体系,也为基于其他交通出行目的识别研究提供理论和方法借鉴。

**关键词:**大数据;公交车;出租车;医院;出行目的;出行链;北京市

## 1 引言

交通出行链是指个体为完成一项或多项活动,而在交通系统中产生的一系列空间位移,且其时间周期为一天<sup>[1]</sup>。交通出行链中包含了大量的居民出行目的信息<sup>[2]</sup>,提取此信息不仅能够剖析居民出行行为的时空格局,还可以更全面且细粒度地揭示不同城市活动的动态特征,因而得到学术界和规划界的密切关注<sup>[3-4]</sup>。

目前交通出行链已被广泛用于识别就业<sup>[5-8]</sup>、回家<sup>[9]</sup>、上学、娱乐<sup>[10]</sup>和休闲<sup>[11]</sup>等活动中。例如,黄洁等<sup>[13]</sup>基于地铁刷卡数据构建的交通出行链,并结合个体上、下车点位置、停留时长等规则识别通勤者的工作地和居住地<sup>[12]</sup>;龙瀛等<sup>[13]</sup>集成公交刷卡、居民出行调查与土地利用类型等多种数据源,在构建交通出行链的基础上识别个体居住地、就业地并进一步揭示通勤出行的时空特征;Ma等<sup>[14]</sup>同样基于公交刷卡数据构建交通出行链,并结合就业活动的时间、空间以及频次等特征提取城市居民的通勤行为,揭示城市空间的职住关系。Algeter等<sup>[10]</sup>基于布里斯班的公交刷卡数据,并结合土地利用数据,识别就业、上学、回家和娱乐等活动。出租车的停靠点具有灵活性,乘客倾向于在尽可能接近目的地的地方下车,其下车点对活动具有较好的指示性,在活动识别和研究中得到广泛的应用<sup>[15-17]</sup>。例如,Axhausen等<sup>[18]</sup>使用出租车GPS数据基于土地利用数据、兴趣点数据、上下车位置、活动开始时间等指标

的特征,提取回家活动;Gong等<sup>[19]</sup>通过乘客下车点与POI的位置关系,判断回家、就业、换乘以及外出用餐等活动类型;Huang等<sup>[20]</sup>结合出租车GPS数据和各类POI在不同时间的吸引力识别购物、娱乐和外出就餐等活动类型。

就医活动是居民生活中重要组成部分,剖析居民就医活动时空特征对医疗资源优化布局具有重要指导意义。传统上,病例数据和问卷调查数据是研究居民就医活动的主要数据源,但因其具有涉及隐私、获取成本高、样本量小以及时效性差等缺点,导致就医活动的相关研究进展缓慢。海量、动态、多源的交通大数据将人、地点和活动等要素集成分析,为研究动态的城市活动提供了可能性,同时也为就医活动研究提供契机<sup>[21-24]</sup>。借助交通大数据进行就医活动的研究,能够更全面、高效地揭示城市中医疗资源布局的现状及问题,具有广阔的应用前景<sup>[25-28]</sup>。现阶段,交通大数据已在就医活动研究中得到了初步的应用,如基于出租车GPS数据提取就医活动、剖析就医活动时空特征<sup>[29-30]</sup>、划分医疗服务范围<sup>[31]</sup>、研究非就近就医行为<sup>[32]</sup>以及揭示交通拥堵对医疗设施可达性的影响<sup>[33]</sup>等。尽管如此,现有研究主要为通过上、下车点与医院的位置关系提取出行行为并直接判断为就医活动,缺少从交通出行链的视角对就医行为的精准识别及验证。

基于此,本文试图采用交通出行数据识别就医活动,以深化交通大数据研究的应用领域,构建基于交通出行链的就医活动识别理论框架和方法体

系。本文以北京市为例,进一步研究乘坐公交车和出租车就医者出行的关键参数与阈值,甄别以就医为目的的交通出行链,并最终通过就医出行统计特征、交通方式选择偏好以及抵院时刻分布规律等方面验证理论框架和方法体系的合理性。

## 2 基于交通出行链的就医活动识别理论框架

明确就医活动的时空特征是构建就医活动识别理论框架的基础。首先,就医活动的交通出行链由3部分组成:去程(始发地—上车点、上车点—下车点、下车点—医院)、就医(挂号、候诊、就诊以及取药等环节的统称)以及返程(医院—上车点、上车点—下车点、下车点—始发地)。其中,去程和返程均需通过交通系统完成。因此,一条完整的交通出行链包含交通出行时间(去程时间和返程时间)和交通系统外时间两部分(下车与再次进入公交系统之间的时间)。

基于以上要素,一条完整的就医活动交通出行链可以被表述为:就医者从始发地出发,然后进入交通系统,换乘或离开交通系统,随后进入医院。就诊完成后,再次进入交通系统,并返回始发地(图1)。就医者在交通系统外的时间大致等同于在医院的停留时长(简称就医时长)。由于就医活动属于偶发性活动,在个体交通出行链中出现的概率较小。因此,上、下车点与医院的位置关系、交通系统外时间以及出行链发生频次是交通出行链中的关键要素,也是识别就医活动的关键指标。

根据就医活动的交通出行链特征,本文构建了基于交通出行链识别就医活动的理论框架(图2),并提出了如下6个准则以尽可能精准识别就医活动。

(1)邻近性准则,即出行目的地越邻近医院,就医可能性越大。上、下车点和设施的位置关系常被用作判断活动目的地的重要指标<sup>[29-30,34]</sup>。因城市交通较为发达,居民对步行距离的接受范围有限,通常倾向于在活动目的地附近下车。

(2)出行链闭合准则,即往返行程形成闭环。在所有出行活动中,居民就医行为具有优先性,返程交通直接回到始发地的行为目的为就医的概率更大。

(3)单一出行目的准则,即交通出行链以医院为唯一目的地。就医属于特殊的出行目的,一般情况下不会与其他出行活动同时进行,出行目的具有单一性。

(4)时间耦合性准则,即交通系统外停留时间与平均就医时长越接近,就医可能性越大。充足的时长可以保证活动顺利进行,因此交通系统外停留时长是区分活动类型的关键指标<sup>[10,12]</sup>。若交通系统外停留时间与平均就医时长越接近,则就医出行概率越大。

(5)路径偶发性准则,即短期内就医活动的交通出行链不具有重复性。一般而言,就医活动属于偶发性行为<sup>[35]</sup>,在居民日常出行中占比较小<sup>[36]</sup>。若相邻时段内同一路径重复出行的概率越高,则就医可能性越低。

(6)陪同准则,即就医者陪同人员会出现平行的交通出行链。考虑到特殊群体的就医活动往往

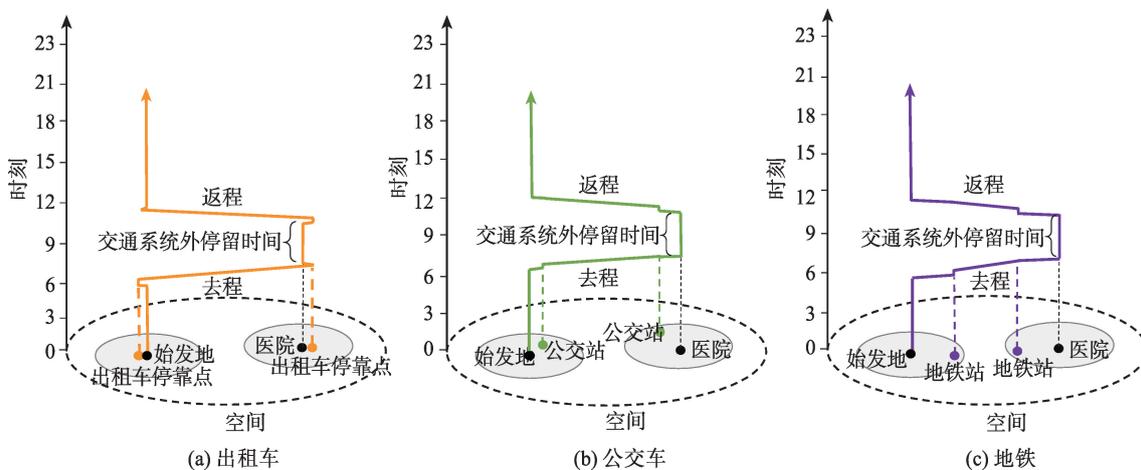


图1 基于交通出行的就医活动时空路径示意

Fig. 1 Sketch of space-time path of hospital-seeking behavior

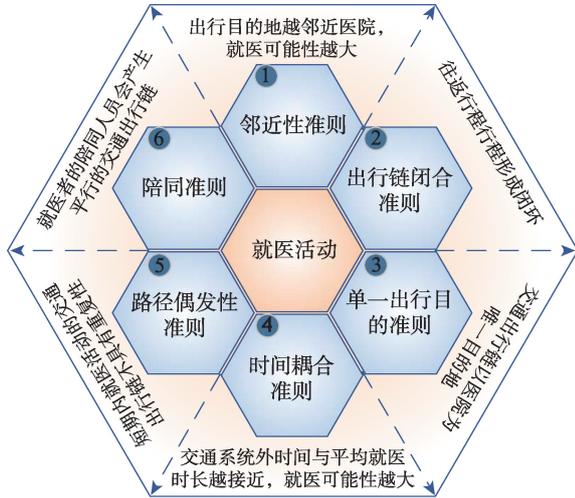


图2 基于交通出行链的就医活动识别理论框架

Fig. 2 Theoretical framework for detecting hospital-seeking behavior

需要陪同人员<sup>[37]</sup>,从而产生平行的交通出行链,但其出行目的为陪同而非就医。

### 3 基于交通出行链的就医活动识别方法体系

公交车和出租车是城市交通的重要组成部分,也是分析城市中居民出行活动的重要数据源。2017年北京市以上2种交通方式的出行比例高达37%<sup>[36]</sup>,对分析城市中居民的出行特征具有代表性。此外,基于交通出行链识别的就医活动中同时包含了本地就医和异地就医活动,这有助于更准确的理解城市中就医活动的动态特征,尤其是异地就医活动较为普遍的大城市。根据上述6个准则,构建基于交通出行的就医活动识别方法体系和数据处理流程。考虑到数据源的结构差异(表1),其中公交IC卡具有唯一编号,而出租车GPS数据无法关联乘客数据,因此基于不同数据源的就医活动识别方法体系构建不同。

表1 公交IC卡刷卡数据和出租车GPS数据结构

Tab. 1 Data structure of smart card data and taxi GPS data

	个体编号	上车点位置	上车时间	下车点位置	下车时间
公交IC卡刷卡数据	√	√	√	√	√
出租车GPS数据		√	√	√	√

#### 3.1 基于公交IC卡刷卡数据的就医活动识别方法体系

根据公交IC卡的唯一编号,可以构建基于个体

的公交出行链,并在此基础上,从上、下车站点的空间特征、交通系统外时间以及出行频次等方面来构建就医活动识别的方法体系(图3)。具体分为以下7个步骤:

(1)构建交通出行链。公交IC卡刷卡数据中记录了某一时间、某一地点发生的上车或下车刷卡行为,根据其唯一编号的追踪,可以生成持卡人上车站点(上车时间)一下车站点(下车时间)的单一行程信息;对一天内的行程信息进行汇总,则可以构成持卡人完整的交通出行链。这是识别和分析个体活动轨迹及其出行目的地的基础。

(2)合并换乘行为,构建无换乘的交通出行链。换乘行为指离开交通系统外、在短时间和距离内又重新进入该系统,换乘并未产生有效的活动目的,但会产生刷卡记录,从而对出行目的地和出行行为识别产生干扰。基于此,若某交通出行链的刷卡记录为站点A—B, B或C—D,若根据交通系统外的停留时间及出行距离,识别出持卡人在B或C仅为换乘行为,则将出行链合并为A—D,从而判断出持卡人在单一交通出行链中的最终目的地。

(3)识别单一出行目的且往返行程形成闭环的交通出行链。基于公交车运营路线的对称性,居民在单一出行活动中,通常会选择与去程相同的交通方式和路线返回始发地。基于出行链闭合和单一出行目的准则,提取仅有一个目的地且活动结束后返回始发地的交通出行链。

(4)提取以医院为目的地的交通出行链。基于邻近性准则,筛选去程下车点且返程上车点均与医院具有地理邻近性的交通出行链。即获得轨迹为始发地—医院—始发地的交通出行链。

(5)标定交通系统外时间的阈值范围。基于就医活动的平均时长及交通系统外时间与出行链数量变化关系<sup>[38]</sup>,标定交通系统外时间的阈值范围。

(6)根据偶发性准则,提高以就医为目的的交通出行链识别精度。《北京交通发展年报》<sup>[36]</sup>显示就医出行占居民所有出行的比例仅为2.1%。基于就医活动的偶发性准则,剔除短期内重复出现的交通出行链。

(7)筛减陪同者的出行链。陪同者指陪同活动者进行某项活动的人,其出行目的为陪同而非活动本身。本文定义若干条(大于1条)具有相同出发点和到达点,且彼此出发和到达时间差均小于1 min的交通出行链,其中一条为就医活动,其余为陪同

者的出行链。基于就医活动的陪同准则,识别并剔除陪同者的交通出行链。

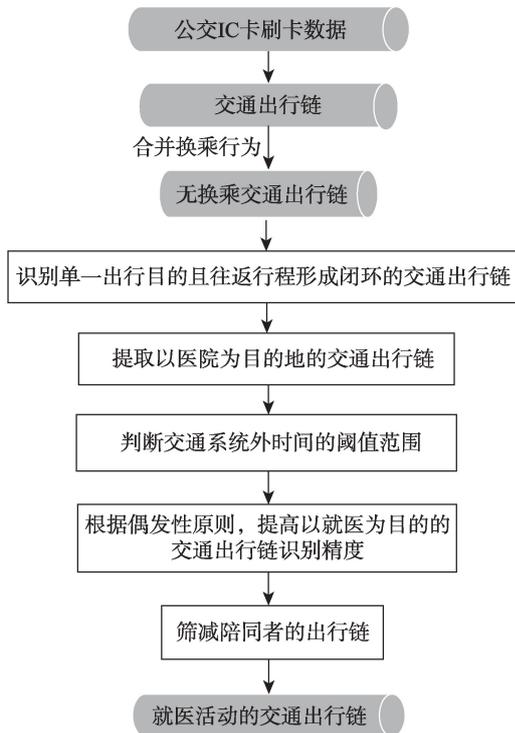


图3 基于公交IC卡刷卡数据的就医活动识别方法体系

Fig. 3 Methodology for detecting hospital-seeking behavior from smart card data

### 3.2 基于出租车GPS数据的就医活动识别方法体系

出租车的停靠位置较为灵活,乘客倾向于在目的地出入口的位置下车。但通常情况下,会因为交通拥堵等原因乘客的下车点会在目的地入口一定距离范围内。因此,乘客上、下车点的位置关系对活动目的地仍具有较强的指示性。通过判断乘客上、下车点与医院大门的空间位置关系,识别就医行为。依据就医活动识别理论框架中邻近性准则,设置上、下车点与医院的距离阈值,提取目的地为医院的行程(图4)。

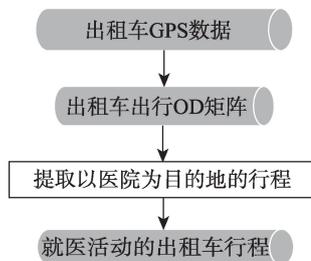


图4 基于出租车GPS数据就医活动识别的方法体系

Fig. 4 Methodology for detecting hospital-seeking behavior from taxi GPS data

## 4 案例分析与结果验证

### 4.1 研究对象和区域

北京市医疗机构由医院、疗养院、社区卫生服务中心等类型机构组成,其中医院和卫生服务中心提供门诊服务。由于社区卫生服务中心的服务范围为已划定,就医者基本为就近就医。从就医者具有自主选择医院的医院出发,本文以医院为研究对象,且研究对象为门诊就医。按照《医院分级管理办法》,北京市医院分为一级、二级和三级医院<sup>[40]</sup>,其中三级医院的规模和人才技术力量最强。北京市是全国医疗资源最集中、就医活动发生最密集的地区<sup>[41]</sup>。其六环内聚集了全市83.7%的一级以上医院和91.7%的三级医院,因此本文以六环内为研究范围。共涉及227家一级及以上的综合医院,其中包含50家三级医院和30家二级医院(图5)。

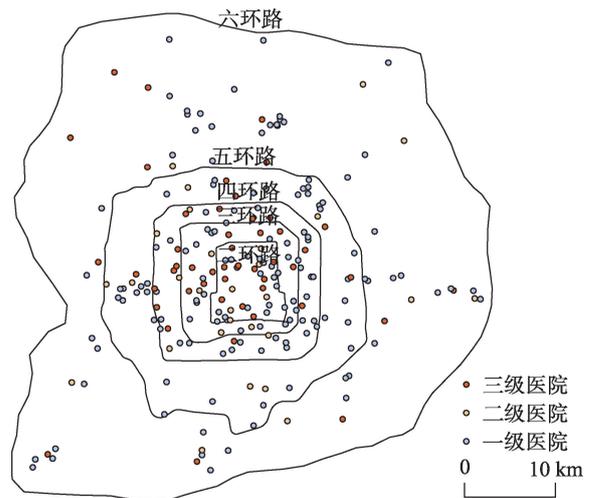


图5 2017年北京市六环内一级及以上医院的分布

Fig. 5 Spatial distribution of hospitals within Beijing's sixth ring road in 2017

### 4.2 数据处理

#### (1) 公交IC卡刷卡数据与交通出行链构建

本文公交IC卡刷卡数据来源于北京交通信息中心,涵盖了2017年6月5—7日(周一至周三)3 d内北京市所有公交IC卡刷卡数据,约3000万条。由于北京市公交车于2014年12月28日后全部采用分段计费的方式,且上、下车均需刷卡,因此可以通过公交IC卡刷卡信息追踪到持卡者的上车时间、上车站点、下车时间及下车站点(表2)。根据上文阐述的交通出行链构建方法,将持卡人一天中若干个单一行程信息(上车站点(上车时间)一下车站点

表2 公交IC卡刷卡数据示例  
Tab. 2 Sample of smart card data records

卡号	上车站点	站点经度/°E	站点纬度/°N	上车时间	下车站点	站点经度/°E	站点纬度/°N	下车时间
0002	酒仙桥	116.489	39.968	2017-06-05 10:44:00	首钢医院	116.204E	39.927N	2017-06-05 11:42:00

(下车时间))汇总,构建其完整的交通出行链。为减少数据运算量,本文首先筛选与医院周边站点相关的交通出行链,共筛选出约300万条数据。

#### (2)出租车GPS数据及OD矩阵构建

本文获取的出租车GPS数据来源于北京市某出租车公司,是2015年4月1—26日覆盖北京市六环内所有出租车的随机抽样数据,包含约18000辆出租车的出行轨迹,合计858.8万条数据记录。出租车GPS数据采样的时间间隔大约为30s,记录了出租车ID、经纬度坐标、时刻以及是否载客等信息。其中,是否载客字段显示1为载客,而0为未载客。其变化是判断乘客上车和下车的标志,是否载客字段由1变为0时为乘客的下车点;反之,为上车点。本文基于以上规则构建出租车出行OD矩阵(表3)。

表3 出租车OD矩阵示例  
Tab. 3 Table of taxi OD data

ID	O点经度	O点纬度	D点经度	D点纬度
	/°E	/°N	/°E	/°N
0000012	116.523	39.852	116.542	39.921

### 4.3 就医活动识别的关键参数与阈值标定

(1)基于公交出行链的就医活动识别参数与阈值甄别

首先,进行公交换乘时间和距离的阈值识别,以此合并换乘行为。其中有2个关键参数:持卡人在交通系统外的停留时间,以及离开站点与再次进入交通系统的站点之间的距离。通过交通系统外停留时间与持卡人数量的变化规律(图6)可看出,46.0%持卡人交通系统外停留时间小于20min且其中48.3%持卡人离开站点与再次进入交通系统站点之间的距离小于500m(北京市公交乘客可接受的步行距离<sup>[42]</sup>)。因此,本研究中换乘行为的时间和距离阈值为20min和500m,此阈值与靳海涛等<sup>[43]</sup>使用北京市公交IC卡刷卡数据识别的换乘行为阈值范围一致。共识别出约257.0万条无换乘交通出行链。

其次,判断首次乘车点和返程下车点的距离偏差参数。北京市公交乘客可接受的步行距离为500m<sup>[42]</sup>,设定持卡人首次乘车点和返程下车点的

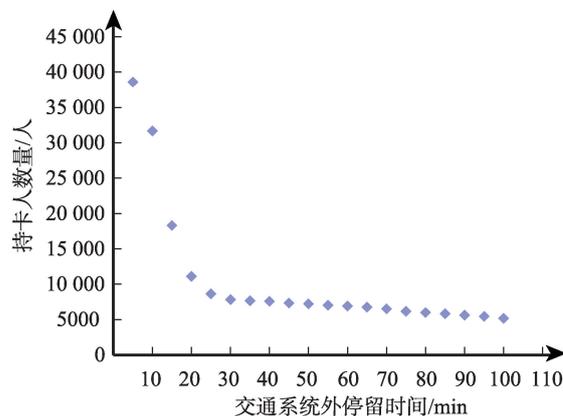


图6 换乘时间阈值的判断

Fig. 6 Threshold of transit time

距离偏差小于1000m,共提取出约384.0万条单一出行目的且往返行程构成闭环的交通出行链。

再次,持卡人通常使用步行的方式由下车点到达医院,因此结合北京市公交乘客可接受的步行距离,设定下车站点及再次进入交通系统站点与医院之间的距离均小于500m,共筛选出约22.2万条以医院为目的地的交通出行链。

进一步分析交通系统外停留时间与就医的可能性。通过交通系统外停留时间与持卡人数量的统计规律可以发现,随着交通系统外时间的增加,持卡人数量先减少再趋于稳定。北京市平均就医时长约为2~3h<sup>[38]</sup>,持卡人数量与此时间范围变化一致的阈值范围为就医的可能性大,因此设定交通系统外的时间阈值为60~280min(图7),共筛选出约

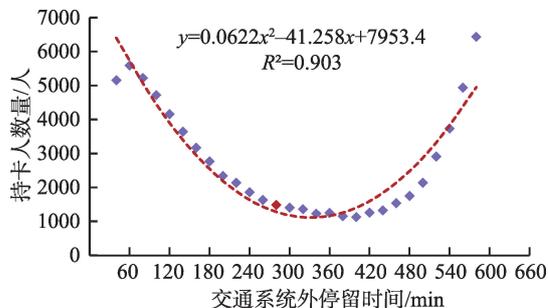


图7 交通系统外时间与持卡人数量的关系

Fig. 7 Relationship between interval time and the number of bus passengers

6.0万条数据。

根据上文中就医活动偶发性准则,设置同一个体以就医为目的的交通出行链在3 d内不重复出现。最后识别并剔除陪同者的就医出行链。经以上参数阈值的设定(图8),本文共提取出约5.0万条

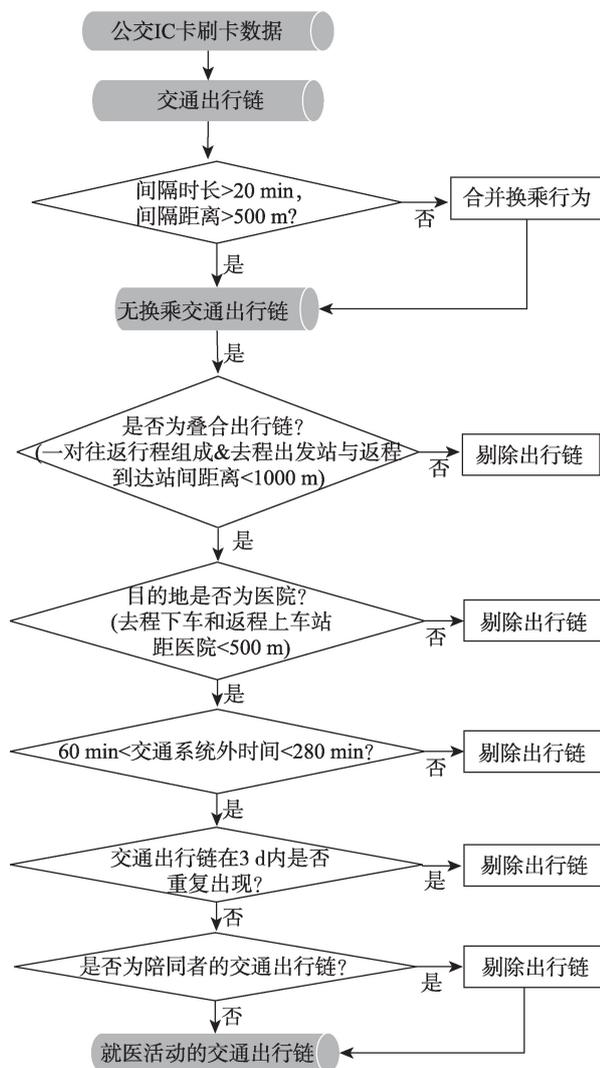


图8 基于公交IC卡刷卡数据的就医活动识别与参数阈值标定流程

Fig. 8 Flowchart of detecting hospital-seeking behavior with smart card data

以就医为目的的交通出行链。

(2) 基于出租车行程的就医活动识别参数与阈值甄别

通过统计医院(入口)不同距离范围内出租车上、下车点数量的变化规律,可发现当距离小于30 m时,出租车上、下车点数量与距离成正比,即随着距离的增加出租车上、下车点数量上升;当距离大于30 m时,出租车上、下车点数量呈逐步下降趋势;而

当距离超过50 m后,出租车数量变平稳并再次呈增加趋势(图9);这可能是由于医院附近存在其他的交通出行吸引目的地。因此,判断上、下车的距离医院50 m范围内的出租车行程为就医的可能性较大。基于以上参数的阈值标定,本文从约858.8万条出租车出行OD矩阵中识别出约10.0万条就医活

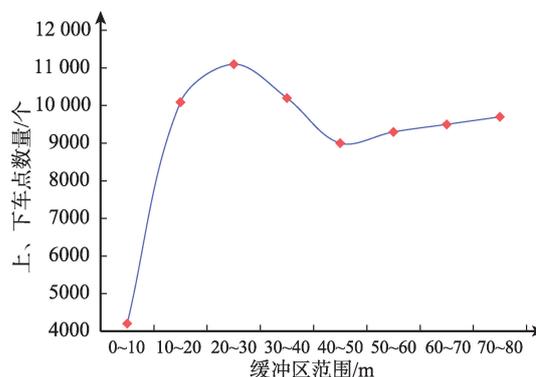


图9 医院(入口)不同距离范围内出租车上、下车点数量的变化规律

Fig. 9 Changes in the number of drop-off and pick-up points by the distance to hospital entrances

动的出租车行程。

#### 4.4 就医活动识别结果验证

基于上文中就医活动识别的理论框架和方法体系,共识别出约15.0万条以就医为目的的交通出行链。在我国医疗体系中,三级医院拥有最大的规模,最丰富的医师资源,最先进的技术水平及医疗硬件设备,因而具备最大的吸引力。为获取此类医疗资源,居民通常愿意克服大的距离成本,提前抵达医院获取号源,并花费更多的时间候诊<sup>[44]</sup>。且就医者的病症越严重越倾向于选择三级医院就医<sup>[45-46]</sup>。出租车的舒适程度和成本均高于公交车<sup>[30]</sup>,就医者病症严重且就医急迫时倾向于乘坐出租车就医,因此三级医院就医者的出租车选择偏好更明显。结合上述就医活动特征,下面将通过就医出行统计特征、交通方式选择偏好分析及抵院时刻分布规律方面评价识别结果的准确性,并以此验证理论框架和方法体系的合理性。

##### (1) 就医出行统计特征

统计各级医院的就医量可以发现,北京市居民偏好三级医院。在识别出的约15.0万次就医活动中,有75%以三级医院为出行目的地。且就医者集中的前十位医院均为三级医院,包括北京大学第三医院、中国人民解放军医院、中日友好医院西

区、北京协和医院、首都医科大学附属北京朝阳医院、首都医科大学附属北京安贞医院、首都医科大学附属北京同仁医院、首都医科大学宣武医院、北京市房山区良乡医院、火箭军总医院。三级医院不仅是各级医院中就医量最大的,也是平均就医时间最长的,比所有医院平均就医时间多4.2 min。此外,通过比较分析各级医院服务距离特征可以发现,三级医院就医者的平均交通出行时间也最长(表4),这表明所有医院中,三级医院的空服务范围最大。综上所述,三级医院的就医量最大、就医时间最长、空服务范围最大,符合医疗体系规划中三级医院的定位。

### (2) 交通方式选择偏好

通过对比各级医院乘坐公交车和出租车就医者数量可以发现,三级医院乘坐出租车的就医者数量明显大于公交车(表4),这表明三级医院就医者偏好出租车作为交通方式。而一级医院的就医者中,乘坐公交车就医者的比例大于出租车,表现出明显的对公交车的选择偏好。此外,对各级医院而言,乘坐出租车就医者的平均出行时间均短于公交车就医者,说明出租车的出行效率高于公交车。以上结果符合就医者病症严重程度、医院选择及交通方式选择偏好的关系。

### (3) 抵院时间分布规律

统计就医活动抵院时间分布规律,可以发现就医者的抵院时间呈现集中分布的态势,整体上呈现上、下午“高低双峰”的分布格局。抵院高峰和低谷分别位于上午8:00—10:00和下午13:00—15:00(图10(a))。此分布格局与《京医通大数据报告》中挂号时间的分布格局一致<sup>[47]</sup>。对比公交车和出租车抵院时刻分布特征可以发现,在夜间时段(晚上20:00—次日凌晨5:00)乘坐公交车抵达医院的就医者

较少(图10(b)),而乘坐出租车就医者在一天中任何时刻均可能抵达(图10(c)),说明出租车是夜间时段就医的主要交通方式之一。此分布格局与公交车和出租车的运营时间一致。比较分析不同等级医院抵院时刻分布特征,发现各级医院上午的抵院高峰时段存在差别,三级医院就医者集中在上午7:00—9:00抵达医院(图10(d)),而二级医院和一级医院分别位于上午8:00—10:00(图10(e))和上午9:00—10:00(图10(f)),晚于三级医院。这与各医院的号源紧缺程度存在很大关系,三级医院号源较为紧缺,并于早上7:00放号,因此就医者倾向于提前到达医院获取号源并提前候诊;而二级医院和一级医院的号源获取容易,无需提前抵达医院。此外,晚上20:00—次日凌晨6:00几乎没有就医者抵达二级医院和一级医院,因为二级医院和一级医院在此时段内不提供服务,证明了识别结果与医院服务时间的吻合性。

## 5 结论与讨论

大数据具有精粒度、细尺度的特征,在居民日常活动的识别和分析中的应用前景广阔。本研究深化了交通大数据在就医活动识别和分析中的应用,基于交通出行链识别就医活动,突破了传统病例数据和问卷数据样本量小、获取难及时效性差等不足对就医活动研究带来的局限。

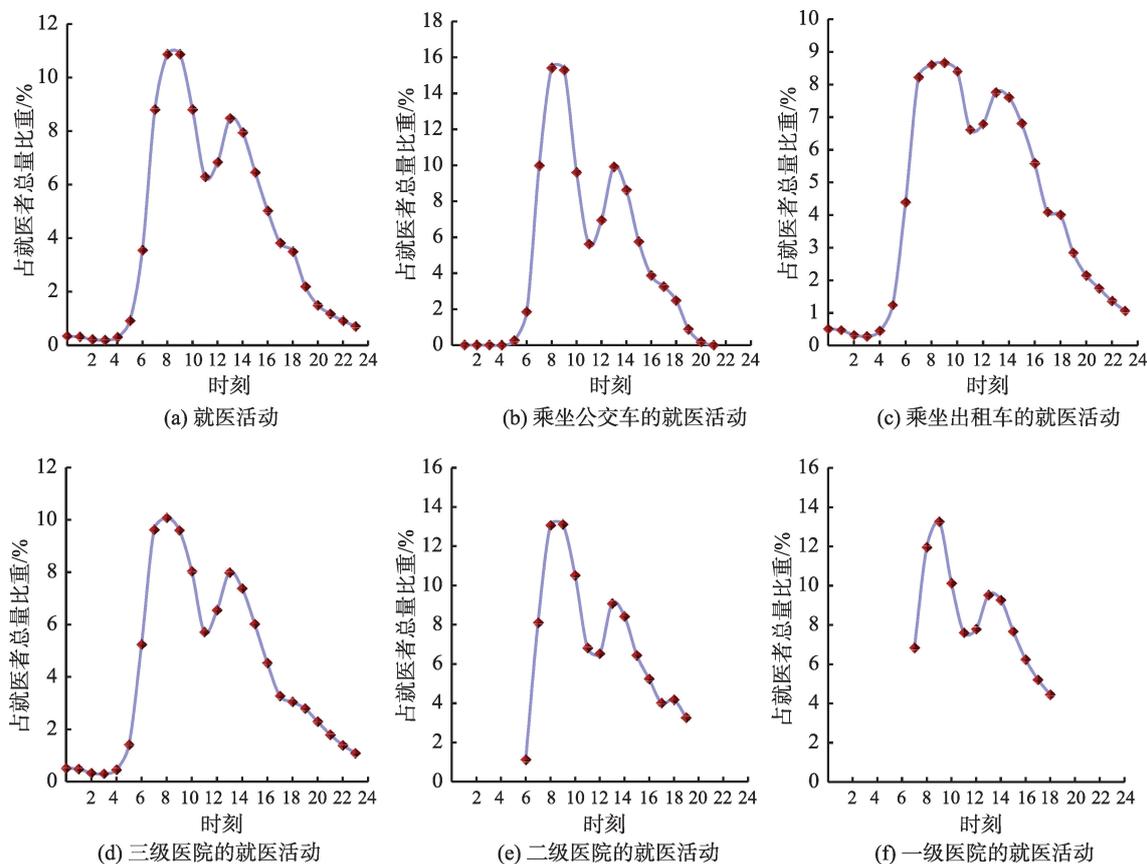
在构建公交车和出租车出行链的基础上,提出了一套基于交通出行链识别就医活动的理论框架和方法体系,并提出了就医活动识别的6个准则:邻近性准则、出行链闭合准则、单一出行目的准则、时间耦合性准则、路径偶发性准则及陪同就医准则。为验证上述理论框架和方法体系的合理性与可行

表4 就医活动识别结果的统计特征

Tab. 4 Statistics of hospital-seeking behavior

	就医者数量/人				交通出行时间/min				就医时长/min	
	公交车		出租车		公交车		出租车		公交车	
	均值	标准差	均值	标准差	均值	标准差	均值	标准差	均值	标准差
所有医院	220	231.1	441	959.6	21.1	16.7	18.7	13.4	138.2	56.7
三级医院	433	321.5	1520	1575.9	22.5	17.3	19.1	13.8	142.4	59.8
二级医院	271	217.7	287	245.8	20.9	17.0	18.0	12.8	139.3	58.2
一级医院	114	116.1	88	130.9	19.7	15.8	17.0	11.4	133.3	52.5

注:就医时长近似等同于交通系统外停留时间。因出租车行程中无法匹配乘客信息,故无法识别其交通系统外的时间,此处仅通过公交车的交通出行链分析就医时长的统计规律。



注:本文中就医者下车时间近似等同于抵院时间。

图10 就医者抵院的时刻分布规律

Fig. 10 Distribution of patients' arrival time

性,本文以北京市为研究对象,基于9000万条公交刷卡数据和858.8万条出租车GPS数据,识别出15.0万条以就医为目的的交通出行链,并从不同等级医院就医者数量、出行时间、就医时长等基本统计规律,交通方式选择偏好及就医者抵院时刻分布规律等方面验证了识别结果的可靠性及理论框架的合理性。

相比于就医问卷调查、统计等小样本数据,基于交通大数据的就医活动研究具有样本量大、覆盖面广的优势。随着理论框架和方法体系的不断完善,可以丰富就医活动研究的数据源,也为进一步开展基于交通大数据的其他出行活动等研究提供借鉴意义。当然,在后续研究中,仍需通过问卷调查来进一步验证基于交通出行链的就医活动识别理论框架和方法体系的合理性。此外,基于交通出行链的就医活动识别中在以下方面可能对识别精度造成影响,如未能将医生、探望者以及看护人员等的出行记录剔除,未能将两位老人同时就

诊的情况纳入到考虑范围内,干扰了就医行为的研究。

#### 参考文献(References):

- [1] Mcguckin N, Murakami E. Examining trip-chaining behavior: comparison of travel by men and women. *Transportation Research Record Journal of the Transportation Research Board*, 1999,1693(1):79-85.
- [2] Schwanen T. Geographies of transport II: Reconciling the general and the particular[J]. *Progress in Human Geography*, 2016,41(3):355-364.
- [3] Bagchi M, White P R. The potential of public transport smart card data[J]. *Transport Policy*, 2005,12(5):464-474.
- [4] Alsger A, Assemi B, Mesbah M, et al. Validating and improving public transport origin-destination estimation algorithm using smart card fare data[J]. *Transportation Research Part C: Emerging Technologies*, 2016,68:490-506.
- [5] Chen J, Yang D Y. Estimating smart card commuters origin-destination distribution based on APTS data [J]. *Journal of Transportation Systems Engineering and Informa-*

- tion Technology, 2013,13(4):47-53.
- [6] Zhou J, Murphy E, Long Y. Commuting efficiency in the Beijing metropolitan area: An exploration combining smartcard and travel survey data[J]. *Journal of Transport Geography*, 2014,41:175-183.
- [7] Devillaine F, Munizaga M, Trépanier M. Detection of activities of public transport users by analyzing smart card data[J]. *Transport Research Record: Journal of the Transportation Research Board*, 2012,2276:48-55.
- [8] Chu K, Chapleau R. Augmenting transit trip characterization and travel behavior comprehension: Multiday location-stamped smart card transactions[J]. *Transport Research Record*, 2010,2183:29-40.
- [9] Medina S, Erath A. Estimating dynamic workplace capacities by means of public transport smart card data and household travel survey in Singapore[J]. *Transportation Research Record*, 2013,2344:20-30.
- [10] Alsger A, Tavassoli A, Mesbah M, et al. Public transport trip purpose inference using smart card fare data[J]. *Transportation Research Part C: Emerging Technologies*, 2018,87:123-137.
- [11] 刘丽敏,虞虎,靳海涛.基于公交刷卡数据的北京城市居民周末户外休闲行为特征研究[J].*地域研究与开发*, 2018,37(6):54-59. [Liu L M, Yu H, Jin H T. Characteristics of outdoor recreation behaviors of beijing residents on weekends based on public transportation data[J]. *Areal Research and Development*, 2018,37(6):54-59. ]
- [12] Huang J, Levinson D, Wang J E, et al. Tracking job and housing dynamics with smartcard data [J]. *Proceedings of the National Academy of Sciences*, 2018,115(50):12710-12715.
- [13] 龙瀛,张宇,崔承印.利用公交刷卡数据分析北京职住关系和通勤出行[J].*地理学报*,2012,67(10):1339-1352. [ Long Y, Zhang Y, Cui C Y. Identifying commuting pattern of Beijing using bus smart card data[J]. *Acta Geographica Sinica*, 2012,67(10):1339-1352. ]
- [14] Ma X, Wu Y J, Wang Y, et al. Mining smart card data for transit riders' travel patterns[J]. *Transportation Research Part C: Emerging Technologies*, 2013,36:1-12.
- [15] Zhou J, Murphy E, Long Y. Commuting efficiency in the Beijing metropolitan area: An exploration combining smartcard and travel survey data[J]. *Journal of Transport Geography*, 2014,41:175-183.
- [16] 黄洁,王姣娥,靳海涛,等.北京市地铁客流的时空分布格局及特征——基于智能交通卡数据[J].*地理科学进展*, 2018,37(3):397-406. [ Huang J, Wang J E, Jin H T, et al. Investigating spatiotemporal patterns of passenger flows in the Beijing metro system from smart card data[J]. *Progress in Geography*, 2018,37(3):397-406. ]
- [17] Kuhlman W. The construction of purpose-specific OD matrices using public transport smart card data[D]. Delft University of Technology, TU Delft, 2015.
- [18] Axhausen K W, Wolf J, Schönfelder S, et al. 80 weeks of GPS-traces: Approaches to enriching the trip information [J]. *Transportation Research Record*, 2004,1870:46-54.
- [19] Gong L, Liu X, Wu L, et al. Inferring trip purposes and uncovering travel patterns from taxi trajectory data[J]. *Cartography and Geographic Information Science*, 2016, 43(2):1-12.
- [20] Huang L, Li Q, Yue Y. Activity identification from GPS trajectories using spatial temporal POIs' attractiveness [R]. 2nd ACM Sigspatial International Workshop on Location Based Social Networks, 2010.
- [21] Hu Y, Miller H J, Li X. Detecting and analyzing mobility hotspots using surface networks[J]. *Transactions in GIS*, 2014,18(6):911-935.
- [22] Zheng Z, Rasouli S, Timmermans H. Two-regime Pattern in Human Mobility: Evidence from GPS Taxi Trajectory Data[J]. *Geographical Analysis*, 2016,48(2):157-175.
- [23] Liu X, Gong L, Gong Y. Revealing travel patterns and city structure with taxi trip data[J]. *Journal of Transport Geography*, 2015,43:78-90.
- [24] Chen C, Bian B, Ma J. From traces to trajectories: How well can we guess activity locations from mobile phone traces?[J]. *Transportation Research Part C: Emerging Technologies*, 2014,46:326-337.
- [25] 杨格格,宋辞,裴韬,等.北京对外交通枢纽乘客OD时空分布特征[J].*地球信息科学学报*,2016,18(10):1374-1383. [ Yang G G, Song C, Pei T, et al. Passengers' OD temporal-spatial distribution characteristics of the external traffic hubs in Beijing. *Journal of Geo-information Science*, 2016,18(10):1374-1383. ]
- [26] Wu L, Zhi Y, Sui Z, et al. Intra-urban human mobility and activity transition: Evidence from social media check-in data[J]. *PLoS One*, 2014,9(5):e97010.
- [27] Yuan Y, Raubal M, Liu Y. Correlating mobile phone usage and travel behavior: A case study of Harbin, China[J]. *Computers, Environment and Urban Systems*, 2012,36(2): 118-130.
- [28] 陈卓然,黄翀,刘高焕,等.基于出租车GPS数据的居民就医时空特征分析[J].*地球信息科学学报*,2018,20(8): 1111-1122. [ Chen Z R, Huang C, Liu G H, et al. Analysis of spatial-temporal characteristics of resident travel for hospitals based on taxi GPS data[J]. *Journal of Geo-information Science*, 2018,20(8):1111-1122. ]
- [29] Kong X Q, Liu Y, Wang Y X, et al. Investigating public

- facility characteristics from a spatial interaction perspective: a case study of Beijing hospitals using taxi data[J]. *ISPRS International Journal of Geo-information*, 2017,6(2):38-51.
- [30] Pan X, Kwan M P, Yang L, et al. Evaluating the accessibility of healthcare facilities using an integrated catchment area approach[J]. *International Journal of Environmental Research and Public Health*, 2018,15(9):1-21.
- [31] Jia P, Wang F, Xierali I M. Using a Huff-based model to delineate hospital service areas[J]. *The Professional Geographer*, 2017,69(4):522-530.
- [32] Yang G G, Song C, Shu H, et al. Assessing patient bypass behavior using taxi trip origin-destination (OD) data[J]. *ISPRS International Journal of Geo-information*, 2016,5(9):157-177.
- [33] 陈建国,周素红,柳林,等.交通拥堵对急救医疗服务时空可达性的影响——以广州市为例[J]. *地理科学进展*, 2016,35(4):431-439. [Chen J G, Zhou S H, Liu L, et al. Estimating the effect of traffic congestion on accessibility to emergency medical care services: Take Guangzhou as an example[J]. *Progress in Geography*, 2016,35(4):431-439. ]
- [34] Boarnet M, Crane R. The influence of land use on travel behavior: specification and estimation strategies[J]. *Transportation Research Part A: Policy Practice*, 2001,35:823-845.
- [35] Ophem H V. The frequency of visiting a doctor: Is the decision to go independent of the frequency?[J]. *Journal of Applied Econometrics*, 2011,26(5):872-879.
- [36] 温慧敏,顾涛,王志丽,等.北京交通发展年报[R].北京:北京交通北京,2018. [Wen H M, Gu T, Wang Z L, et al. Annual report on Beijing's traffic development[R]. Beijing: Beijing Transport Institute, 2018. ]
- [37] 李佳贤,陈艳艳,韩志玲,等.北京市居民初诊就医出行行为特征分析[J]. *交通科技与经济*, 2019,21(5):34-38. [Li J X, Chen Y Y, Han Z L, et al. Analysis of first medical travel behavior characteristics of residents in Beijing[J]. *Technology & Economy in Areas of Communications*, 2019,21(5):34-38. ]
- [38] 许欣悦,曾多,田玮,等.北京某三甲医院预约挂号对患者候诊时间的影响分析[J]. *中国医院管理*, 2015,35(10):38-40. [Xu X Y, Zeng D, Tian W. Analysis of reservation registration effect on outpatient waiting time in one grade 3 first-class hospital[J]. *Chinese Hospital Management*, 2015,35(10):38-40. ]
- [39] 苑继承,高湘伟,石岳琴.门诊患者挂号与就诊时间调查分析[J]. *第四军医大学学报*, 2005,26(1):83-85. [Yuan J C, Gao X W, Shi Y Q. Survey and analysis of distribution of outpatients registration and visiting time[J]. *Journal of the Fourth Military Medical University*, 2005,26(1):83-85. ]
- [40] 中华人民共和国卫生部.医院分级管理办法[R].北京, 1989. [Chinese Ministry of Health. Measures for hierarchical management of hospitals[R]. Beijing, 1989. ]
- [41] Huang J, Levinson D, Wang J E, et al. Job-worker spatial dynamics in Beijing: Insights from Smart Card Data[J]. *Cities*, 2019,86:83-93.
- [42] Zhao P J, Bin Lü, Roo G D. Impact of the jobs-housing balance on urban commuting in Beijing in the transformation era[J]. *Journal of Transport Geography*, 2011,19(1):60-69.
- [43] 靳海涛,金凤君,陈卓,等.基于换乘链断裂点时空信息的公交换乘行为识别[J]. *交通运输工程学报*, 2018,18(5):176-184. [Jin H T, Jin F J, Chen Z, et al. Commute activity identification based on spatial and temporal information of transit chaining breaks[J]. *Journal of Traffic and Transportation Engineering*, 2018,18(5):176-184. ]
- [44] 史华英,李习平.患者就医时间影响就医决策的调查研究[J]. *社区医学杂志*, 2016,14(7):41-43. [Shi H Y, Li X P. Investigation of the influence of patient's medical treatment time on decision making[J]. *Journal of Community Medicine*, 2016,14(7):41-43. ]
- [45] 熊俊浩,夏曦,程小琴,等.基于结构方程模型的患者选择医院影响因素研究[J]. *中国卫生事业管理*, 2012,29(2):84-87. [Xiong J H, Xia X, Cheng X Q, et al. A SEM-based study on influence factors of hospital choice[J]. *The Chinese Health Service Management*, 2012,29(2):84-87. ]
- [46] 王森.我国居民的就医行为及其影响因素研究——基于CHNS调查面板数据的分析[J]. *西北人口*, 2015(3):32-36. [Wang S. Analysis on the Adjustment of Birth Control Policies and Development of China Population. Northwest Population, 2015(3):32-36. ]
- [47] 京医通大数据中心.2017京医通就医数据报告[R].北京, 2018. [Jingyitong big data center. 2017 Jingyitong medical report[R]. Beijing, 2018. ]