

# 20亿“电子脚印”绘出市民出行图

## 济南首次利用手机大数据规划公交线路

大数据已成当下最时髦的IT术语之一,如果你觉得“大数据”离我们的生活还很远就大错特错了。近日,济南市城市交通研究中心利用大数据做了一份出行调查分析,通过“省城市民2个月20亿条手机“电子脚印”历史数据,分析出行特征与规律,进而为公交线路设计和优化做科学准备。

本报记者 孟燕

### 20亿“电子脚印” 来自手机大数据

“其实,大数据分析早就有了,不过这是第一次和咱济南的公共交通结合。”济南市城市交通研究中心主任刘彤说,“以往的出行调查多是抽样调查,费时费力效果还不一定好。为什么要用大数据?一是这是一种新兴的潮流,二是做出来结果更实用。”

大数据就是海量数据,把过去人工无法分析的海量数据进行分析处理。在大数据时代,毫不夸张地说,每一天每个人的“电子行为”都已经被模糊“记录”了下来,成为大数据里的一部分。早晨起来,你在山东路的家里打开IPAD登录了微博,并转发了一条搞笑的信息;然后开着私家车去上班,和平路上又堵成一团了,你拿出手机拍了照发到了微信朋友圈里;快到高新区的时候,部门主任打来电话给你安排了当天的工作……

这次调查分析主要采用的是手机大数据。“早上我在凤凰路打了个电话,会和附近的基站产生联系,下午我又去西客站发了短信,会和西客站附近的基站产生交互。”这样就形成了一个“电子脚印”,要是数以千计的这样的电子轨迹叠加起来就有意义了。

这意味着,当有人使用手机通话或发短信时,移动数据就会被创建。用户的行为将会通过通讯基站记录,并被确认在基站的范围内的某个位置。当呼叫转移到一个新的基站或者一个新的呼叫连接到不同的基站时,用户的移动轨迹将会被确定。

为了避免隐私泄露,所有数据已经过处理,防止任何人从中识别出用户。但有用的信息就是用户的移动轨迹,成为这次大数据分析的基础信息。

“我们获得了近2个月20亿条剔除了个人信息的电子脚印。”刘彤强调,通过对海量的手机数据的时间、位置的规律分析,可以获取城市交通相关数据包括人口分布、客流分布、客流走廊分布以支撑城市交通的发展规划、公共交通的发展规划、公交线路开辟与优化,以及公交运营计划的改善。

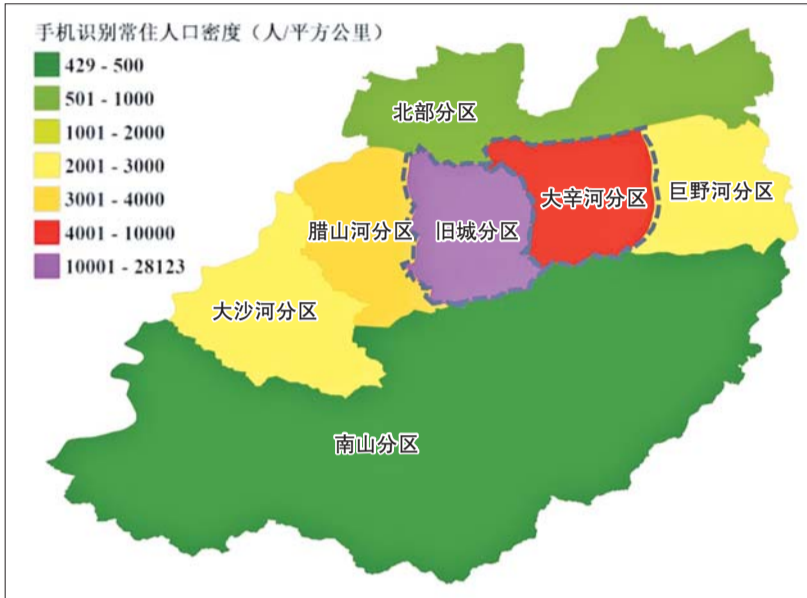
### 线网规划将更合理 等公交时间将变短

省城每隔几年都要进行一次出行调查,为城市公共交通的规划和优化做前提准备。以往的调查都是“抽样调查”,需要耗费大量的人力和物力不说,样本数往往有限。“手机大数据优势在于‘全样本’,我们关注的不是精确度,而是一种趋势。”刘彤举例,以前像公交线路的开辟和优化,更多凭借经验和市民意见。“以前从热线、媒体或者进社区调查发现某个区域缺少公交线路,市民反映比较集中,我们去现场调研。现在基于大数据能进行量化,更科学有效。”

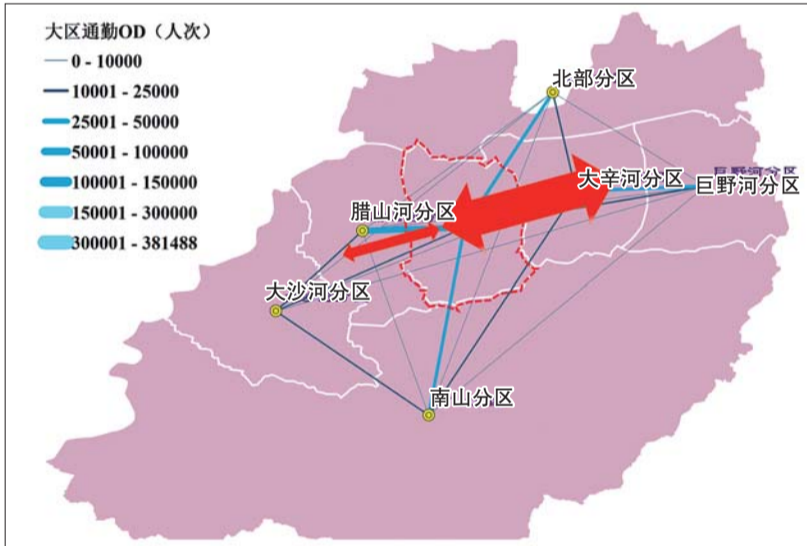
手机大数据提供了动态掌握城市居民出行特征与规律的途径,更清晰了解居民的交通需求,从而可以提供针对性的交通供给,解决城市居民的出行需求供给矛盾,带给市民最优的出行体验。

目前,济南有220余条公交线路,4600余辆公交车。结合手机大数据、IC卡数据和GPS数据,能够利用大数据来减少交通拥堵时间,提高车辆运转效率。“将天气、是否工作日等代入这个科学模型,可以准确知道哪个时段需要增加车辆加密班次,或者在哪个路段发送区间车,保证市民及时乘车。”刘彤表示,目前该模型在K56路测试,通过模型预测的周转时间正确率达到95%,以后即使堵车严重,省城市民等车的时间也会缩短。

### 手机识别常住人口



### 客流总体格局:常规客流走向(通勤OD)



区域	平均通勤出行距离(km)
旧城分区	5.19
大辛河分区	6.36
腊山河分区	7.34
大沙河分区	9.04
巨野河分区	11.07
北部分区	11.93
南山分区	14.39

### 图示解读

## 全市通勤客流呈东西向带状

本报记者 孟燕 整理

### ◎各片区人口分布、职住比

从整个城市的职住布局看,旧城分区(规划范围东至二环东路,南至中心城区南边界,西至二环西路,北至黄河)和大辛河分区(规划范围东至绕城高速东环线,南至中心城南边界,西至二环东路,北至小清河和济青高速公路)交通强度较大,两个区域的居住人口占全市72.3%,工作人口占全市75.5%。南山分区以居住为主,职住比最低,大辛河分区含济南奥体中心、济南高新技术开发区,职住比最高。职住比越大,交通吸引力越大。

作用:把握各个交通小区的出行需求,直观了解各区域的交通需求和供给的矛盾。

### ◎通勤客流分布

从通勤客流分布看,全市通勤客流总

体呈现以旧城分区为中心的的东西向(偏东)的带状客流。东向“旧城分区与大辛河分区”交通通过量最大,占50%;西向“旧城分区与腊山河分区”,占18%。

各分区平均通勤出行距离(直线距离),发展较为成熟的区域出行距离相对较短,旧城分区平均通勤出行距离约为5公里,大辛河分区为6公里,而南山分区以居住为主,往其他分区出行的需求较大且出行距离长,平均通勤出行距离达到14公里。

### ◎典型区域客流特征

通过高新区居住人口与工作人口的出行分布特征分析,可了解市民从哪里来,到哪里去,这些数据为居民交通出行方式选择、出行特征分析提供支持,并为区域内公交线路的布设的走向提供参考,为线路长度的确定提供依据。同时可分析通勤客流与公交运力的匹配度,区域内首末站匹配度分析,是否存在公交站点服务盲区。以找准公交线网、运力投放的调整方向。

### 多知道些

## 大数据那些事儿

本报记者 孟燕

商业营销、政府管理、公共交通、社交网络、医疗健康……都打上了“大数据”的深深烙印。大数据时代已然到来,并以迅雷不及掩耳之势颠覆着我们的传统思维。什么是大数据?大数据能为我们带来什么?本报为您科普一下“大数据”那些事儿。

### 什么是大数据?

最早提出“大数据”的是全球知名咨询公司麦肯锡。麦肯锡在其研究报告中指出,“数据已经渗透到每一个行业和业务职能领域,逐渐成为重要的生产因素;而人们对于海量数据的运用将预示着新一波生产率增长和消费者盈余浪潮的到来。”

大数据(Big data),又称巨量资料,是指数据规模极其巨大,以致很难通过一般软件工具加以抽取、管理、处理并整理成为有用资讯的海量数据。目前业界公认大数据的特点有以下4个V:即Volume(大量)、Velocity(高速)、Variety(多样)、Value(价值)。

第一,数据体量巨大。大数据的起始计量单位至少是P(1000个T,1T=1000G)、E(100万个T)或Z(10亿个T);第二,数据类型繁多。网络日志、视频、图片、地理位置信息等。第三,价值密度低,商业价值高。在数以万计的信息中,能用的可能仅仅有几个。第四,处理速度快,要在秒级时间范围内出分析结果,时间太长就失去价值了。速度要求是大数据处理技术和传统的数据挖掘技术最大的区别。

### 大数据从何而来?

仅以互联网为例,一天之中,互联网产生的全部内容可以刻满1.68亿张DVD;发出的邮件有2940亿封之多;发出的社区帖子达200万个,相当于《时代》杂志770年的文字量……

根据相关统计,互联网上的数据每年将增长50%,每两年便将翻一番,目前世界上90%以上的数据是最近几年才产生的。全世界的工业设备、汽车、电表上有着无数的数码传感器,随时测量和传递着有关位置、运动、震动、温度、湿度乃至空气中化学物质的变化,也产生了海量的数据信息。

物联网、云计算、移动互联网、车联网、手机、平板电脑、PC以及各种各样的传感器,无一不是大数据来源或者承载的方式。

### 大数据如何使用?

但是,巨量资料本身其实并不多大用处。我们一般讲“大数据”的时候,指的是大数据的处理和分析,大数据经过的处理分析过速提取以及可视化后才变得“有用”。

大数据如何使用并没有“模式”可言。现在大数据的影响并不仅仅限于技术层面,它颠覆了传统上我们看待世界的思维方法。海量的数据资源让各领域开始了量化,决策行为由过去更多凭借经验和直觉做出,变成了基于数据分析做出。这就是“大数据思维”,一方面样本来源从抽样变成了“全部数据”样本,另一方面是关注效率和趋势而不是精确度,关注相关性而不是因果关系。

拥有大数据思维的人,可以将大数据的潜在价值转化为实际利益。大数据的真正价值在于创造,在于填补无数个还未实现过的空白。