

# Passenger Flow Direction Analysis of Urban City Using Bus Card Data\*

Man Li<sup>1</sup>, Zhihong Chen<sup>2</sup>, Liying Sui<sup>2</sup>

<sup>1</sup>Department of Computer Science, Beihang University, Beijing

<sup>2</sup>Beijing Traffic Development Research Center, Beijing

Email: liman@nlsde.buaa.edu.cn

Received: Aug. 22<sup>nd</sup>, 2012; revised: Sep. 18<sup>th</sup>, 2012; accepted: Sep. 20<sup>th</sup>, 2012

**Abstract:** While Automated Fare Collection System is widely used in many cities all over the world, large amount of swiping card data has been recorded. By using the data, this paper proposed a novel method to recognize passenger flow direction in urban cities. At the beginning of this paper, subway stations are divided into three kinds, including stations in residences, stations in work area and stations in mixed area. Then a model is built to recognize the passenger flow directions in peak hour. Finally, passenger flow patterns in Beijing are analyzed.

**Keywords:** Bus Card; Subway Transportation; Passenger Flow Direction

## 基于一卡通数据的城市客流流向分析\*

李 曼<sup>1</sup>, 陈志宏<sup>2</sup>, 隋莉颖<sup>2</sup>

<sup>1</sup>北京航空航天大学, 北京

<sup>2</sup>北京交通发展研究中心, 北京

Email: liman@nlsde.buaa.edu.cn

收稿日期: 2012年8月22日; 修回日期: 2012年9月18日; 录用日期: 2012年9月20日

**摘 要:** 基于城市公交一卡通刷卡数据, 本文提出一种新的方法来识别城市客流流向, 该方法首先根据轨道站点的特征把站点分为三类: 住宅区站点、工作区站点和混合区站点, 然后通过进一步分析得到城市客流流向规律; 最后, 以北京市轨道交通为例, 分析了客流流向规律。

**关键词:** 公交一卡通; 轨道交通; 客流流向

### 1. 引言

随着城市交通规划建设的发展, 城市公共交通的交通分担率也有所增长, 目前主要有地面公交和轨道交通。近年来, 城市交通一卡通的使用也日益普遍, 通过一卡通, 城市居民可以方便地乘坐公交和轨道进行出行。同时, 一卡通刷卡会产生大量的交易数据, 包括乘坐线路和时间等信息, 利用这些数据, 有助于了解居民出行特征, 从而为城市交通规划和运行提供

实践支持。与移动手机数据不同, 刷卡数据提供更多出行的信息, 包括出行时间、出行方式、起止位置以及消费金额等<sup>[1]</sup>。一卡通数据中的时空信息可以提高城市交通车辆调度效率、实现运行监测等。

Bachi<sup>[2,3]</sup>分析了城市一卡通数据对出行特征的潜在支持, 并提出公交一卡通数据按照一定规则处理后可以得到居民的出行换乘率、连续出行比例及出行者特征。

目前, 城市居民的出行特征比较复杂, 包括出行时间和距离, 出行目的(工作、上学及其他)以及出行

\*基金项目: 特大城市先进交通运行协调指挥平台(课题编号: 2010 ZX01045-001-009-1)。

模式(私家车、地面公交、轨道交通及步行)。因此,有必要对城市居民的出行模式进行分析,以便对城市的出行规律有一定认识,指出当前城市交通中的不足并为城市规划和建设提供建议。对于一卡通数据的使用和处理在国内外已有很多人在研究。通过伦敦的一卡通(Oyster Card)数据被用来讨论当前城市客流需求与实际运能的对比以及在交通服务的稳定性,包括在在站点、线路路段及线路级别<sup>[4]</sup>。刘<sup>[5]</sup>研究了个体和整体上的城市居民出行模式,并通过深圳的一卡通数据给出了实例验证。Kitamura et al.<sup>[6]</sup>基于一卡通数据得到了基于时间变化的时空出行规律。Garling and Axhausen<sup>[7]</sup>也讨论了基于一卡通数据的居民出行规律。

本文主要基于城市轨道刷卡数据研究不同区域的居民出行规律。首先,通过客流模式识别站点性质;然后识别城市主要的客流流向模式;最后,利用北京一卡通刷卡数据进行客流分析。本文研究思路见图 1。

## 2. 一卡通数据处理

### 2.1. 一卡通数据

本文中用到的一卡通数据格式如下:

- 1) 卡号,一卡通编号,一卡通的唯一标识;
- 2) 进站时间,进站时刷卡的时间;
- 3) 进站站点,通过进站时的线路和站序标识;
- 4) 出站时间,出站时刷卡的时间;
- 5) 出站站点,通过出站时的线路和站序标识。

通过一卡通刷卡数据,可以得到乘客精确的 OD 出行信息,包括时间和位置。

### 2.2. 站点客流模式

一般情况下,城市的客流量在一天之内有两个高

峰时段,早高峰时段和晚高峰时段。客流早高峰是由于大量居民上班或上学引起的;同样地,下班高峰期即是晚高峰。

通过地铁站点在时间维度上的客流变化趋势,可以得到站点附近区域的三种不同的用地性质:

Morning-only 模式,如图 2 所示,只在上午有个高峰时段;

Afternoon-only 模式,如图 3 所示,只在下午有个高峰时段;

Morning-afternoon 模式,如图 4 所示,在上午和下午均有高峰时段。

## 3. 站点模式识别模型

地铁站点的位置及周边用地性质,是影响站点客流的重要因素。因此,通过研究站点的客流模式可以有助于分析站点周边区域的居民的出行规律。这部分我们主要讨论站点的客流模式识别方法。

### 3.1. 数据预处理

本文中,考虑一天从上午 5 点到晚上 22 点的客流,半个小时为一个时段,共有 34 个时段。

站点客流量记作  $PT = (PT_1, PT_2, \dots, PT_N)$ , 其中,  $PT_i$  表示站点在第  $i$  个时段的客流量,  $N$  表示时段总个数, 这里  $N = 34$ 。

通过如下表达式,可以得到归一化的客流量表达式。

$$PT'_i = \frac{PT_i}{\sum_k PT_k}$$

### 3.2. 早高峰时段和晚高峰时段

对于不同站点,居民的出行时间规律不同,早高

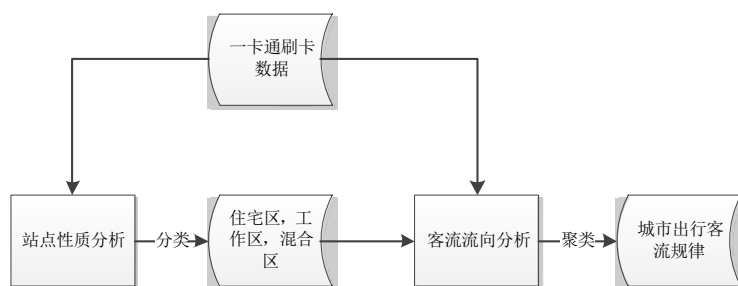


Figure 1. Ideas of this paper  
图 1. 本文研究原理图

基于一卡通数据的城市客流流向分析

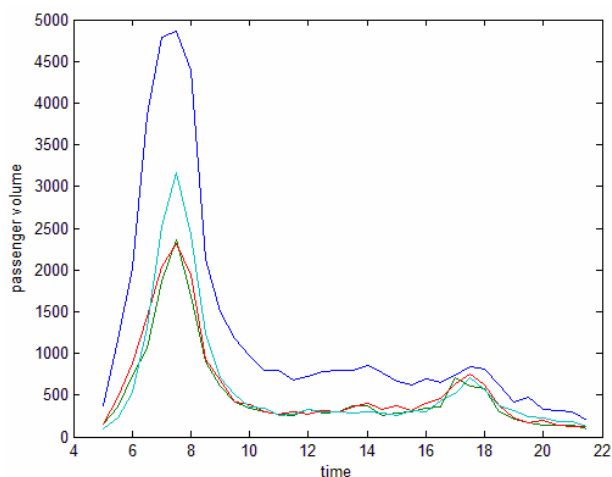


Figure 2. Passenger pattern of morning-only station  
图 2. Morning-only 模式站点

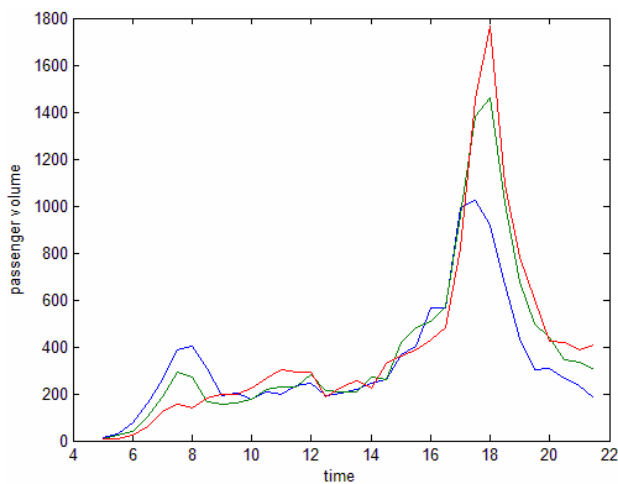


Figure 3. Passenger pattern of afternoon-only station  
图 3. Afternoon-only 模式站点

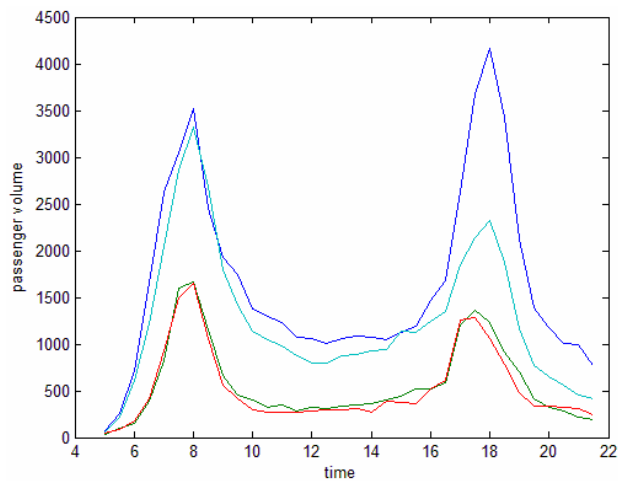


Figure 4. Passenger pattern of morning-afternoon station  
图 4. Morning-afternoon 模式站点

峰时间和晚高峰时间也不同。因此需要针对不同的站点动态地确定早晚高峰时间。以中午 12 点作为中间界限,上午有最多客流量的连续的2个小时(四个时段)称为早高峰时段;同理,下午有最多客流量的连续的2个小时(四个时段)称为晚高峰时段。

高峰时段表示为  $[t, t+4]$ , 则高峰时间的客流量为  $f(t) = \sum_{k=t}^{t+4} PT_k$ , 早高峰客流量和晚高峰客流量应该满足如下表达式:

$$\text{type} = \begin{cases} \text{morning-only,} & \text{if } f(t^{(\text{morning})}) > \alpha, f(t^{(\text{afternoon})}) < \max\left\{\alpha, \frac{f(t^{(\text{morning})})}{\beta}\right\}; \\ \text{afternoon-only,} & \text{if } f(t^{(\text{afternoon})}) > \alpha, f(t^{(\text{morning})}) < \max\left\{\alpha, \frac{f(t^{(\text{afternoon})})}{\beta}\right\}; \\ \text{both,} & \text{if } f(t^{(\text{afternoon})}) > \alpha, f(t^{(\text{morning})}) > \alpha, \frac{1}{\beta} \leq \frac{f(t^{(\text{afternoon})})}{f(t^{(\text{morning})})} \leq \beta. \end{cases}$$

其中,  $\alpha, \beta$  均为参数。 $\alpha$  为高峰时段内客流量的最小值, 也就是说, 如果一个站点的在早高峰(晚高峰)时间内的客流量小于这个阈值  $\alpha$ , 则认为这个站点没有早高峰(晚高峰)。 $\beta$  表示站点的早高峰客流量与晚高峰客流量之间的差异限制, 如果早晚高峰客流量差异过大, 则认为该站点只有一个高峰时段(早高峰时段或晚高峰时段)。

#### 4. 客流流向识别模型

已有许多研究者针对单个站点的客流流向进行研究。本文提出了一种客流流向识别模型, 通过这个模型, 可以得到城市级别的客流流向规律。

##### 4.1. OD 矩阵

OD 矩阵表示城市中每两个站点之间的客流流向, 有如下表示:

$$P = (p_{ij})_{n \times n}$$

其中  $p_{ij}$  表示从站点  $i$  到站点  $j$  之间的客流量。 $N$  表示城市站点总数。

对站点间的客流量进行归一化,

$$p'_{ij} = \frac{p_{ij}}{\sum_k p_{ik} * \sum_k p_{kj}}$$

$$f(t^{(\text{morning})}) = \max_{1 \leq i \leq 10} f(t); \quad f(t^{(\text{afternoon})}) = \max_{15 \leq i \leq 30} f(t)$$

其中,  $[t^{(\text{morning})}, t^{(\text{morning})} + 4]$  是早高峰时段,

$[t^{(\text{afternoon})}, t^{(\text{afternoon})} + 4]$  为晚高峰时段。

##### 3.3. 模式识别模型

结合站点高峰时段的定义, 通过如下的公式把站点分为三种模式。

从站点  $sta$  出发的客流量记作  $\text{PassIn}_{sta}$ ,

$$\text{PassIn}_{sta} = \{p_{sta,1}, p_{sta,2}, \dots, p_{sta,n}\} = P_{sta}$$

同理, 到达站点  $sta$  的客流量记作  $\text{PassOut}_{sta}$ ,

$$\text{PassOut}_{sta} = \{p_{1,sta}, p_{2,sta}, \dots, p_{n,sta}\} = P'_{sta}$$

##### 4.2. 两站点相似度

对于工作区的两个站点, 相似度定义为从该站点出发的客流量的相似度。

$$\text{Similar}_{i,j}^{(\text{Work})} = \text{PassIn}_i * \text{PassIn}'_j$$

对于住宅区的两个站点, 相似度定义为到达该站点客流量的相似度。

$$\text{Similar}_{i,j}^{(\text{Resi})} = \text{PassOut}_i * \text{PassOut}'_j$$

对于混合区的两个站点, 相似度定义为两种客流量的相似度的最小值。

$$\text{Similar}_{i,j}^{(\text{Mix})} = \min \left\{ \text{PassIn}_i * \text{PassIn}'_j + \text{PassOut}_i * \text{PassOut}'_j \right\}$$

其中  $i, j$  是两个站点。

两站点之间的客流流向特征的距离定义如下:

$$d_{i,j} = 1 - \text{Similar}_{i,j}$$

## 5. 结果分析

本文的测试是基于北京市轨道交通一卡通 2011 年 7 月 4 日(周一)的刷卡数据及城市轨道线路和站点地图数据。

### 5.1. 站点客流识别

通过站点模式识别模型, 我们可以得到站点周围区域的用地性质(结果见表 1)。

### 5.2. 客流流向识别

通过客流流向识别模型, 对住宅区站点的客流流向进行聚类, 聚类采用最小距离法, 类间距离阈值设置为 0.1, 也就是说如果两站点客流流向特征相似度大于 0.9 就认为是同一类。

聚类结果如图 5 和图 6, 其中横轴表示每个站点的标号, 纵轴表示到达该站点的客流百分比, 每个类别中的客流流向特征都很明显的相似。

以类别 1 为例, 计算累加客流量, 得到客流累加曲线(如图 7), 该曲线比较光滑, 我们用 6 阶多项式进行拟合。可以看出, 大约 25%的目标站点(D 站点)包含了 60%多的出客流量。

增长模式的拟合结果如下(拟合系数见表 2):

$$f(x) = p_1 * x^6 + p_2 * x^5 + p_3 * x^4 + p_4 * x^3 + p_5 * x^2 + p_6 * x + p_7$$

### 5.3. 在地图中显示客流流向

为了能够定位客流 OD 流向, 我们利用 Google 地图上显示方向特征。在以下的地图中, 红色表示站点出发的客流量, 绿色表示到达该站点的客流量。每个站点位置圆形的面积代表客流量的大小, 面积越大, 该站点的客流量越大。

对于在 5.2 中提到的类别 1, 如图 8, 出发站点位于 1 号线的最西侧的几个站点及八通线所有的站点(连接了 1 号线的东侧并通向郊区); 目标站点大部分是 1 号线中间的站点以及其他线路的少量站点。

在分类 2 中, 出发站点只有 3 个: 安河桥北、北宫门站和西苑站。如图 9 所示, 除了中关村, 海淀黄庄和中国人民大学三个站点外, 其他的目的站点分布比较均匀, 并且客流量较少。从地图上的位置看, 我们还可以推断出这种模式的居民出行距离较短。

根据类别 6 的客流模式, 在地图上显示如图 10。许多的郊区居民从北边方向到达市区工作。很明显, 大多数的目的站点在市中心位置偏北, 主要有西直门、五道口、上地等站点。

## 6. 结论

本文基于城市一卡通刷卡数据, 首次根据站点客流分析站点性质, 把站点分为三类: 住宅区站点、工作区站点和混合区站点。并在此基础上提出了一种客流流向识别模型, 对住宅区站点的客流流向进行聚类分析, 从而得出在城市级别上的客流流向, 而不仅仅是单个站点级别的客流流向; 并以北京为例, 讨论了北京的客流情况, 着重分析了北京市客流流向模式。

Table 1. Station pattern recognition  
表 1. 站点模式识别

模式名称	数量	区域性质
Morning-only	53	住宅区
Afternoon-only	32	工作区
Morning-afternoon	81	混合区

Table 2. Coefficients (with 95% confidence bounds)  
表 2. 系数(95%置信区间)

$p_1 = -5.848e-013$ (-6.241e-013, -5.454e-013)
$p_2 = 3.481e-010$ (3.283e-010, 3.679e-010)
$p_3 = -8.318e-008$ (-8.699e-008, -7.937e-008)
$p_4 = 1.03e-005$ (9.95e-006, 1.065e-005)
$p_5 = -0.0007333$ (-0.000749, -0.0007176)
$p_6 = 0.03355$ (0.03325, 0.03385)
$p_7 = 0.03522$ (0.03336, 0.03708)

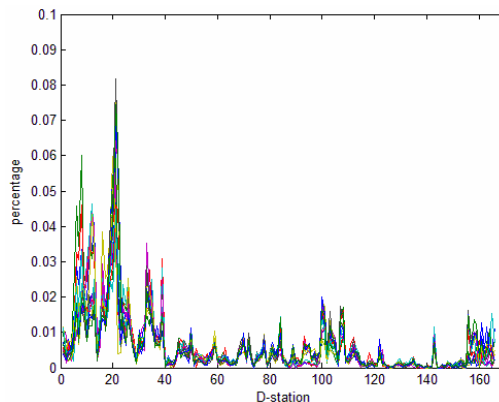


Figure 5. Cluster pattern 1  
图 5. 聚类模式 1

基于一卡通数据的城市客流流向分析

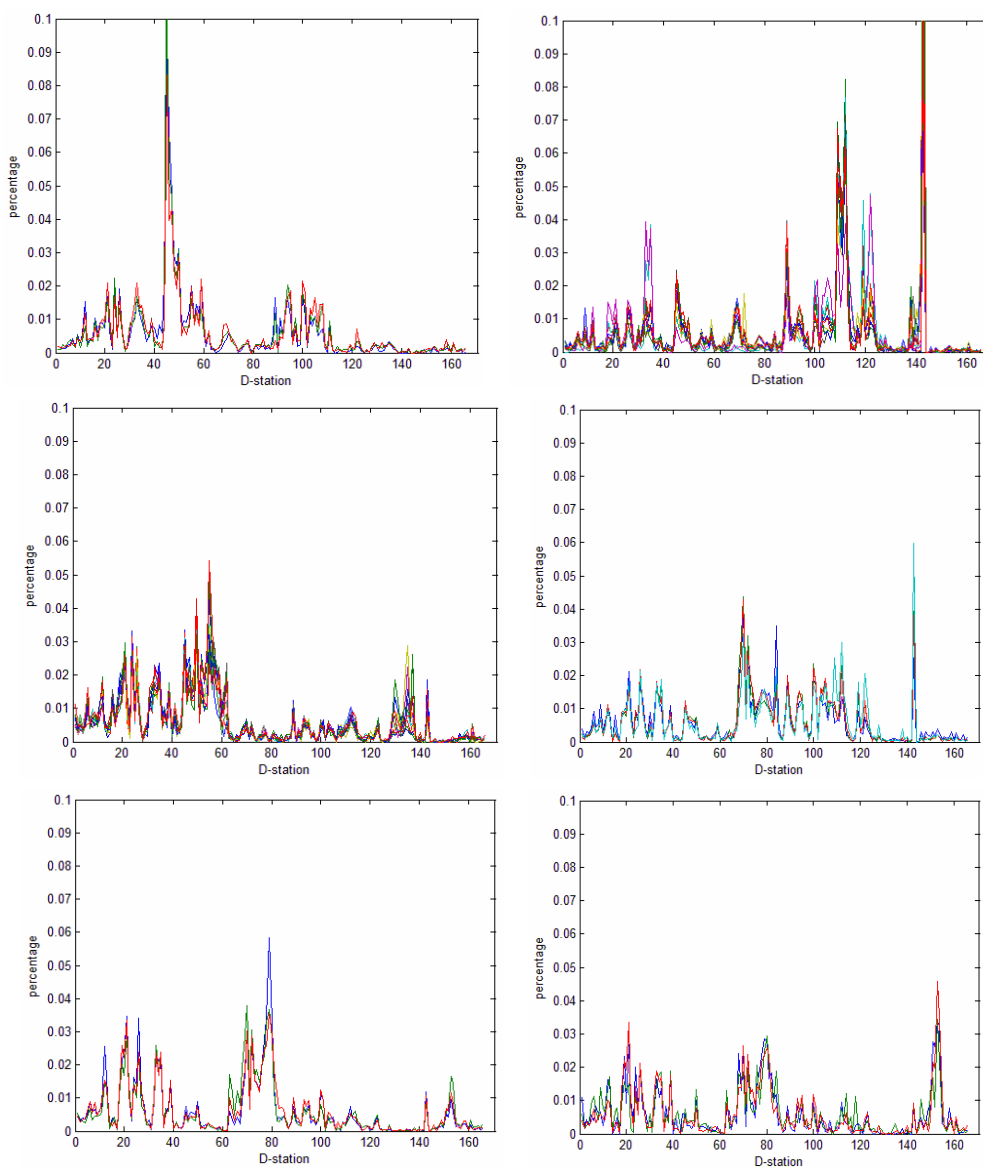


Figure 6. Cluster pattern 2-7  
图 6. 聚类模式 2-7

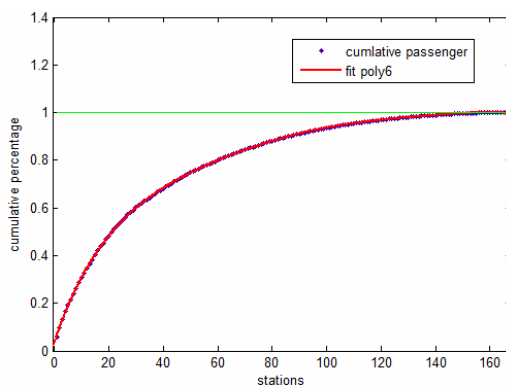


Figure 7. The increment pattern on number of stations  
图 7. 站点客流增长模式

基于一卡通数据的城市客流流向分析

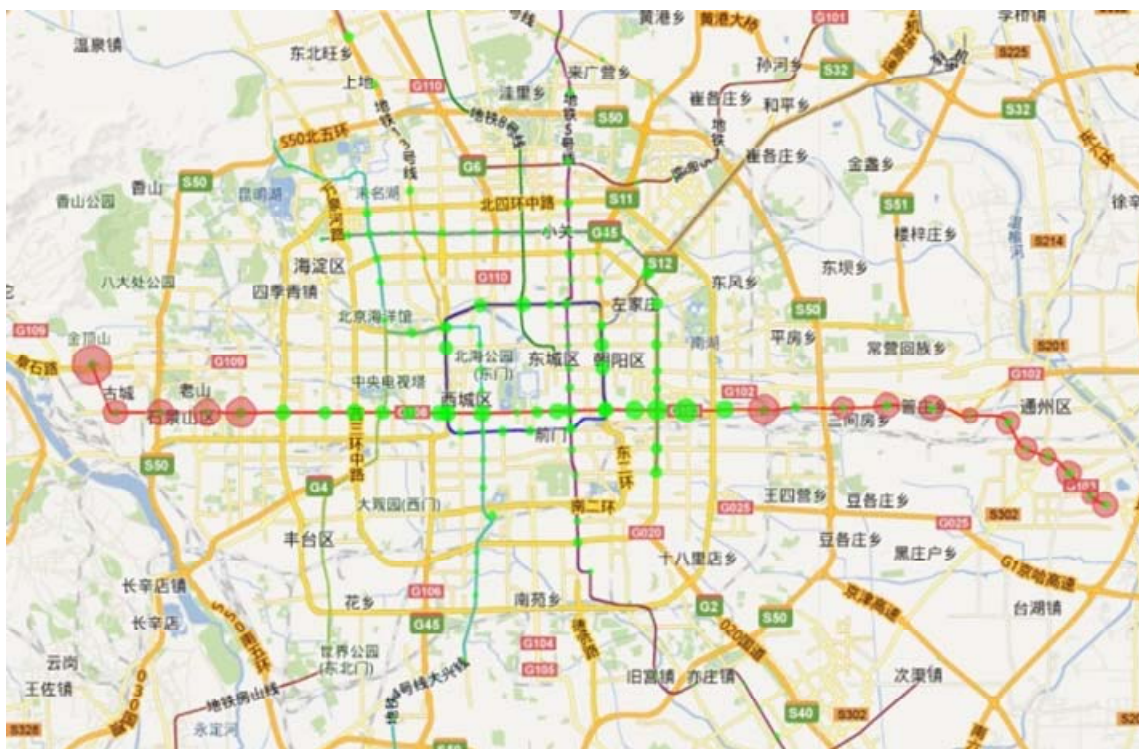


Figure 8. Travel pattern: obvious direction of large amount  
图 8. 区域出行模式：大量客流



Figure 9. Travel pattern: main destination with short distance  
图 9. 区域出行模式：短距出行



Figure 10. Travel pattern: passenger from suburbs to urban are  
图 10. 区域出行模式：城郊出行

但是，地铁客流不能准确地反映城市所有居民的出行，城市居民的出行方式是多种多样的，还包括地面公交、出租车、私家车、自行车甚至步行。因此需要综合考虑多种出行方式，得到城市的出行规律。另一方面，出行规律也受到时间的影响，周末节假日的出行规律与平时是不同的。这将成为下一步的研究内容。

### 参考文献 (References)

[1] N. Lathia, L. Capra. How smart is your smartcard? Measuring travel behaviours, perceptions, and incentives. 13th ACM International Conference on Ubiquitous Computing (UbiComp), Beijing, September 2011.

[2] M. Bagchi. Use of smartcard data from bus systems for travel behavior analysis, and implications for marketing. Ph.D. Thesis, University of Westminster, 2003, unpublished.

[3] M. Bagchi, P. R. White. Use of public transports smart card data for understanding travel behavior. Proceedings of the European Transport Conference, Strasbourg, 8-10 October 2003.

[4] J. Chan. Rail transit OD matrix estimation and journey time reliability metrics using automated fare data. Master's Thesis, Massachusetts Institute of Technology (MIT), 2007.

[5] C. Ratti, L. Liu, A. Hou, A. Biderman, J. Chen, et al. Understanding individual and collective mobility patterns from smart card records: A case study in Shenzhen. 12th International IEEE Conference on Intelligent Transportation Systems, St. Louis, 3-7 October 2009: 1-6.

[6] T. Garling, K. W. Axhausen. Introduction: Habitual travel choice. Transportation, 2003, 30(1): 1-11.

[7] R. Kitamura, T. Yamamoto, Y. O. Susilo and K. W. Axhausen. How routine is a routine? An analysis of the day-to-day variability in prismvertex location. Transportation Research Part A, 2006, 40: 259-279.