

# 一种个体地铁出行预测的马尔可夫模型

于晨<sup>1</sup>, 王斌<sup>1</sup>, 翟希<sup>2</sup>, 张倩<sup>1</sup>, 汪洋<sup>1</sup>

(1. 上海师范大学信息与机电工程学院, 上海 201418; 2. 上海市城乡建设和交通发展研究院, 上海 200032)

**摘要:**为了解和挖掘城市人类移动行为的规律性和差异性,提出了一种马尔可夫模型来预测个体地铁出行行为。首先通过数据分析挖掘出城市出行个体的类别差异,根据出行数据将用户分为不同属性;然后标记出行记录标签,统计分析不同属性用户的出行时间分布差异;最后分时间段建立了用户个体出行的马尔可夫模型,预测其出行状态。对10万名用户的1000余万条地铁出行记录进行交叉验证,结果显示个体马尔可夫模型的平均预测精度达到72.73%,优于传统的时序预测模型。本文模型有效地预测了地铁出行中的个体出行行为,有利于解决城市发展面临的交通拥堵问题,进一步合理分配交通资源。

**关键词:**人工智能;个体出行预测;马尔可夫模型;人类移动行为;城市计算

中图分类号:TP312

文献标志码:A

文章编号:2095-2783(2019)07-0705-06

## A Markov model for individual travel prediction by subway

YU Chen<sup>1</sup>, WANG Bin<sup>1</sup>, ZHAI Xi<sup>2</sup>, ZHANG Qian<sup>1</sup>, WANG Yang<sup>1</sup>

(1. College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 201418, China;

2. Shanghai Urban-Rural Construction and Transportation Development Research Institute, Shanghai 200032, China)

**Abstract:** In order to understand and explore the regularity and difference of human mobility in cities, a Markov model was proposed to predict individual travel behavior by subway. Firstly, through the analysis of data, the category differences of urban travel individuals were mined, and users were divided into different attributes based on travel data. Then, the travel record label was labeled to analyze the difference of travel time distribution among users with different attributes. Finally, a Markov model was established to predict the travel status of individual users in different time periods. After cross-validating more than 10 million travel records of 100 000 users by subway, the results show that the average prediction accuracy of the individual Markov model is 72.73%, which is better than the traditional time series prediction model. This model effectively predicts individual travel behavior in subway travel, which is conducive to solving the traffic congestion problem in urban development and further rational allocation of traffic resources.

**Keywords:** artificial intelligence; individual travel prediction; Markov model; human mobility; urban computing

人类在社会中的移动现象以多种方式存在,国家和城市间的人类移动表现为移民或者旅游,城市内的移动则主要是通勤行为。与前者相比,城市内的移动行为更加频繁,结构更加复杂。城市规模的快速发展使得交通拥堵和公共交通资源配置不合理的问题越来越突出,越来越多的学者开始关注并研究人类移动行为。

Gonzalez等<sup>[1]</sup>通过研究波士顿10万名匿名手机用户的轨迹后发现个体的轨迹具有高度的时空规律性,这打破了以往对人类移动具有随机性的认识。在一个城市里,大多数人都有很高的可能性往返于几个固定的地方,偶尔也会探索新的地方,并且个人出行模式往往具有单一的空间概率分布。Yong等<sup>[2]</sup>从最大熵原理的角度分析了北京地铁交通出行数据分布,发现人类进入地铁的时间间隔服从指数截断的幂律分布,这说明人类出行时间的分布符合一定规律。Jiang等<sup>[3]</sup>研究发出现出租车和公交巴士

出行位移分布符合指数分布,地铁出行位移分布更适合用伽马分布来拟合,这说明人类出行的空间分布也符合一定的规律。Yong等<sup>[4]</sup>将人类的工作日和周末出行分开研究,发现它们可以分解为不同的模式,并且关注到出行的一些兴趣点,例如住宅、娱乐场所和交通枢纽,这给本文的研究带来启发。Yang等<sup>[5]</sup>从网络的角度出发,证明人类移动网络可以分为两类:协议人类移动网络和社会人类移动网络,并讨论了链接预测社会人类移动网络的可行性。Alexander等<sup>[6]</sup>分析了美国波士顿地区的手机通话细节记录(call details record, CDR),利用CDR所含的时间和坐标信息将用户的地点转换为集群位置,进而判断为家庭、工作或其他地点,并且和人口普查结合起来,汇总出行信息后建立出行矩阵来分析其活动,从群体层面对人类出行特征进行建模。Jiang等<sup>[7]</sup>提出了一种时间分辨率为10 min、空间分辨率为数百米的移动TimeGeo模型。在获得城市结

收稿日期:2019-06-23

基金项目:上海市科学技术委员会科技创新行动计划项目(18511104202)

第一作者:于晨(1995—),男,硕士研究生,主要研究方向为智能交通

通信作者:王斌,副教授,主要研究方向为计算机视觉、机器学习、城市计算等,binwang@shnu.edu.cn

构与出行距离的关系后,利用通信技术生成个人信息,无需依赖繁琐而昂贵的出行调查来构建城市交通模型。随着近几年深度学习方法的发展,也有很多学者致力于使用深度学习的方法解决人类出行预测问题。Song 等<sup>[8]</sup>收集了大量的异构数据,例如全球定位系统(global positioning system, GPS)记录和交通运输网络的数据,基于深度长短期记忆(long short-term memory, LSTM)架构建立了一个智能系统,可以在全市范围内预测和模拟人类的出行和交通模式,这也是第一个将深度学习应用于预测人类出行和交通运输模式的系统,但是此模型并没有建立稀疏轨迹用户类别的模型。针对地铁出行数据,Yang 等<sup>[9]</sup>根据出行数据的熵值将用户分为3类,并对用户的出行数据进行了相关的统计分析,根据相关的数据分布结果分别分析不同类别用户的出行行为。Gambs 等<sup>[10]</sup>使用聚类算法发现个体经常访问的兴趣点(points of interest, POI),然后根据出行数据生成  $n$ -MMC 预测模型预测下一个位置,但是获取前  $n$  个历史地点带来了很大的额外开销,并且不能预测个体的全天出行。人类移动行为的研究可以带来诸多的社会效益,Xu 等<sup>[11]</sup>通过整合手机活动数据、人口普查数据和电动汽车司机调查数据,提出了一种在精细的时间和空间分辨率上估计电动汽车司机移动的方法,通过解决电动汽车移动和充电的需求节省了电力开支。

在以往的人类移动行为研究中,由于用户覆盖面广,使用移动电话通信数据或 GPS 数据的居多,但是从移动设备端获取到的信息也有诸多不足,特别是冗余信息过多,比如用户仅在某地接通或拨打一次电话,就更新一次用户位置,或者用户仅在本文可以忽略的有限范围内活动,例如在公司处理日常事务的走动。这些位置的小变动对于本文所研究的出行预测问题来说都是噪声信息,更严重的是这些信息在数据中往往占据了大部分,因为大多数用户 1 d 中的出行时间相比于非出行时间来说少得多。在最近的研究中,越来越多的学者关注公共交通智能卡数据(smart card data, SCD),因为这种数据在保证信息准确的情况下,噪声信息非常少,易于处理和分析。在城市中,特别是上下班高峰期,由于地铁的准时,越来越多的人更偏向于乘坐地铁上下班,SCD 可以反映大量城市居民出行的时空特征。除此之外,大多数地铁站点建设在城市重要交通枢纽、商业中心、购物商场附近,利用城市土地建设特征可以提高模型的预测准确率。由于地铁数据的优越特点,本文利用上海市 2017 年 5 月—8 月的地铁出行数据按用户类别分别建立了个体出行模型。本文所研究的 SCD 来源于上海市 2017 年 Soda 大赛(<http://shanghai.sodachallenges.com>),其中包含 10 万名用户经过地铁闸机所产生的匿名刷卡数据,包括卡片 ID、刷卡日期、刷卡时间、进站站点、出站站点等。

## 1 SCD 处理与分析

上海市轨道交通系统比较成熟,并且还在不断地完善当中。目前开通的有 16 条线 400 余个站点,日均客运量为 1 083 万人次,年客运量约为 39.5 亿人次,运营里程居世界第一。本文所研究的 SCD 采集自 2017 年 5 月—8 月,共 4 个月超过 10 万名用户的刷卡信息,总数据量达 1 000 余万条。刷卡记录仅在出站时上传到服务器的数据库中,这些数据包括不含姓名的用户 ID、出站刷卡日期、出站刷卡时间、进站站点 ID 和出站站点 ID 等。从庞大的数据量中挖掘用户的出行规律对进一步规划城市轨道交通系统具有重要意义。

### 1.1 数据预处理与用户属性判定

通常来讲,用户在工作日偏向于通勤出行,周末的出行轨迹则比较杂乱。某些用户周末可能前往商场等休闲娱乐场所,但这往往也不限于一个固定的场所。由于工作日出行的规律性,在数据预处理阶段去掉了 2017 年 5 月—8 月间的周末,以便下一步的分析。从工作日数据中分析得到大部分用户往返于 2 个固定站点的频率较高,这符合通勤用户的出行习惯,往返于住宅和公司的次数占比最大。由于通勤用户与非通勤用户出行的差异性,需要根据出行记录将用户分为通勤与非通勤用户分别建模。根据出行规律,制定了一个用户属性判决规则,如图 1 所示。首先根据出行记录判断用户的住宅点和工作点<sup>[12]</sup>,取出每名用户 10:00 前的第一条和 17:00 后的最后一条出行记录,统计第一条记录的出发站点和最后一条记录的到达站点作为住宅地候选站点。住宅地候选站点中出行频率最高且大于设置的阈值的站点定义为住宅地附近站点,若出行频率均小于阈值,则判断为无住宅站点。工作地附近站点则统计第一条记录到达站点和最后一条记录出发站点作为候选站点,按此规则判断工作地附近站点。通过早晚出行记录确定住宅地附近的站点后,通勤用户与非通勤用户的最大区别在于是否存在一个明确的工作地附近站点,存在则为前者,否则为后者。由于数据缺失或者用户每天出行习惯不固定,可能存在少数用户无法确定住宅站点,这部分用户称为无住宅(站点)用户。分析全部用户出行信息后,结果显示通勤用户占到 80% 左右,非通勤用户和无住宅用户分别占到 15% 和 5%。

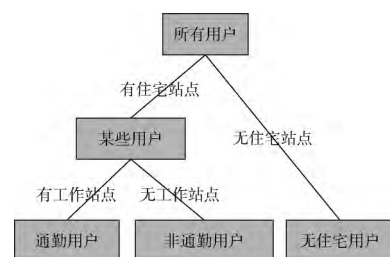


图 1 用户属性判决规则

在将用户合理分为3类后,得到各类用户4个月内出行总次数(不含周末)时间分布如图2所示。

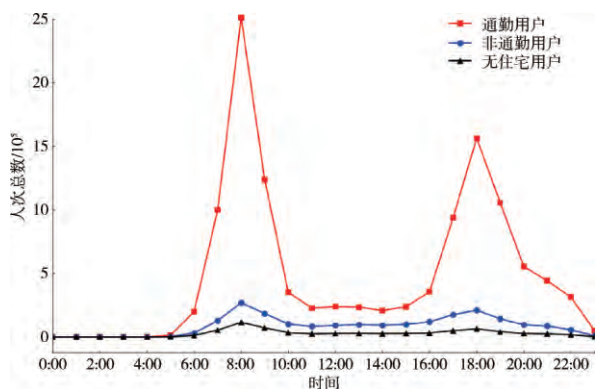


图2 各类用户出行总次数时间分布

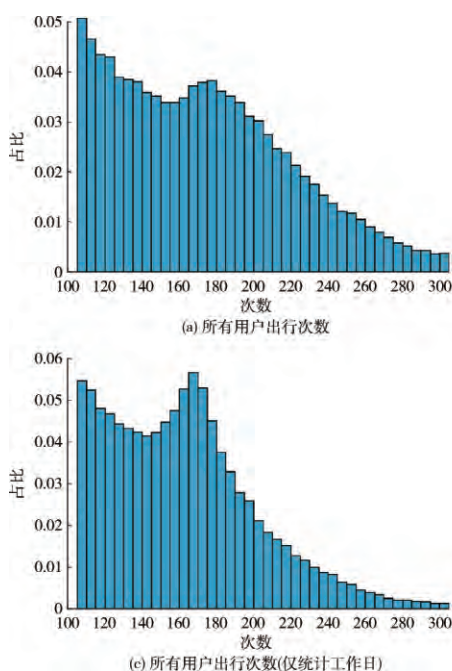


图3 所有用户出行次数、出行天数分布直方图

在对用户出行有了整体了解后,为了分析用户的每条出行记录,在1.1节判断完用户的住宅(附近)站点和工作(附近)站点后,对每个用户的每条出行记录都添加标签。除了住宅和工作站点外,一律称为其他站点,图4和图5分别给出了通勤用户与非通勤用户的几种主要活动的出行时间分布直方图。图4(a)中可以明显地看出通勤用户在住宅和公司之间的出行集中在9:00和19:00左右,这也符合对早晚高峰的预判。由于非通勤用户没有一个合适的工作站点,所以没有住宅和公司之间的出行。图4(b)和图5(a)分别为通勤用户和非通勤用户从住宅往返其他站点的出行分布,图5(a)下午出行比图4(b)略高,而在晚上恰恰相反。通勤用户由于有较为规律的下班时间,所以在上下班后由住宅往返其他站点的出行会相对较多,下午则相对较少。在不基于

从图2中可以明显地看到,无论何种用户出行高峰都集中在7:00—9:00和17:00—19:00时间段,其中通勤用户占绝大多数,这符合早晚高峰的出行情况和用户的分布情况。

## 1.2 用户出行标签与统计分布

在1.1节中对数据进行预处理后,统计了所有用户4个月内的出行次数和出行天数(图3),以对用户出行有进一步的了解。从图3(a)中可以看出用户出行次数主要集中在110次和170次左右,除去周末后的统计结果如图3(c)所示,此时170次更加突出。在图3(b)和图3(d)中,用户的出行天数更多地集中在85d左右。统计结果的原因可能是在4个月的出行记录中,多数出行记录为通勤用户在工作日的2次通勤出行。

住宅的出行中,如图4(c)和图5(b)所示,非通勤用户在下午(非上下班时间)明显比通勤用户占比要多,说明非通勤用户在非上下班时间活动更频繁一些,而在上下班高峰期,两者没有很大区别。所有的出行分别展示在图4(d)和图5(c)中,非通勤用户的全天出行分布相比于通勤用户明显更加均匀。

## 1.3 用户出行距离分析

目的地距出发地位移通常也是决定用户是否出行的重要因素。在地铁站点经纬度坐标已知的情况下,可以根据始发站和到达站的ID计算用户每次出行的位移。图6为根据每次出行的位移绘制出的分布图。从图6中可以看出,出行距离在2~3 km处有一明显的高峰,这个距离在城区约为2~4站,说明乘坐地铁短程出行的人占比较大。另外在6~7 km处达到另一个高峰,之后出行呈下降趋势。与文献[2]中

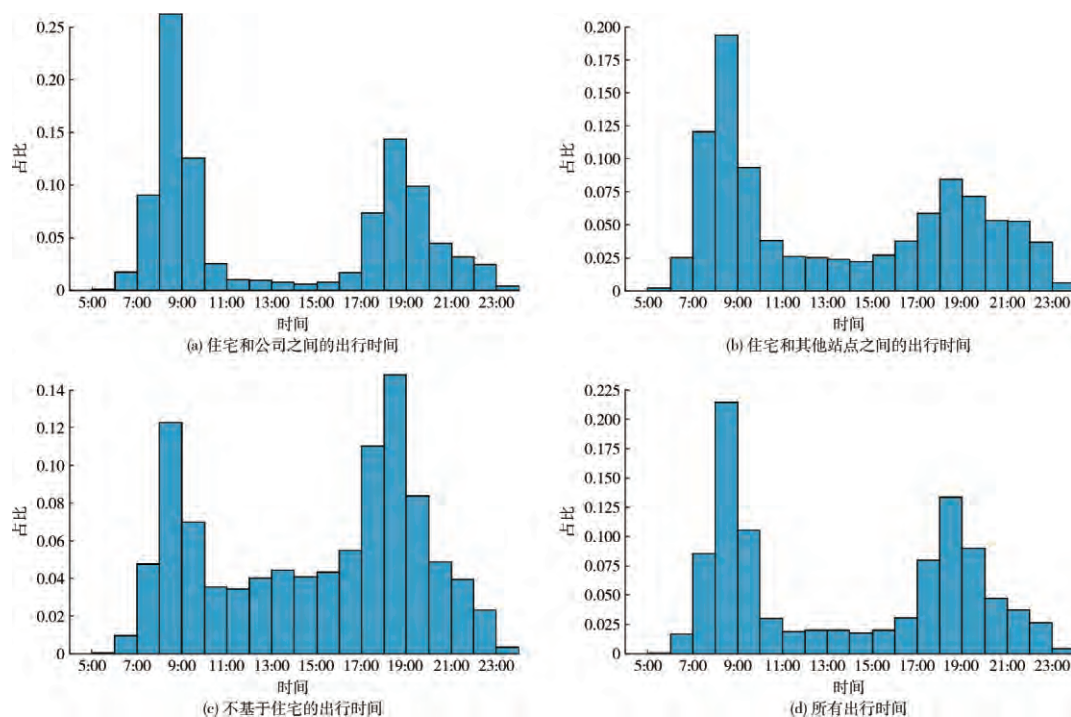


图4 通勤用户按标签统计出行时间分布直方图

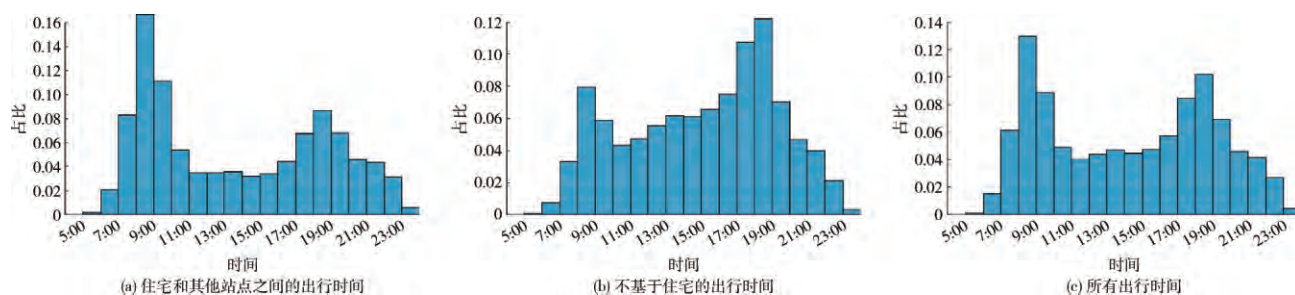


图5 非通勤用户按标签统计出行时间分布直方图

所提出的地铁分布符合伽马分布不同,上海地铁出行位移分布更符合卡方( $\chi^2$ )分布。

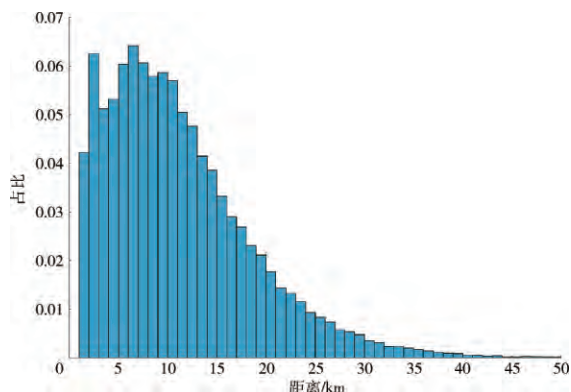


图6 出行位移分布直方图

## 2 基于马尔可夫的出行预测模型

### 2.1 马尔可夫模型

马尔可夫(Markov)模型是一种广泛应用在语音

识别、自然语言处理等领域的统计模型。在马尔可夫模型中,若给定当前时刻信息,则过去的状态(指当前时刻以前的状态)对于预测将来的状态(指当前时刻后的状态)是无关的。另外一种比较简明的阐述是,过程中某一时刻的状态只依赖于其前  $n$  个状态,  $n$  取不同的值代表不同阶数的马尔可夫过程。 $n=1$  时的马尔可夫过程是一阶马尔可夫模型,即某一时刻的状态只依赖于其前一个状态。

设  $\{X(t), t \in T\}$  为马尔可夫过程状态向量,其中  $T = \{t_1, t_2, \dots, t_n, t_{n+1}, \dots\}$  为马尔可夫过程的时刻集合。由定义知一阶马尔可夫模型  $X(t_{n+1})$  的取值只取决于  $X(t_n)$ , 而与  $X(t_1), X(t_2), \dots, X(t_{n-1})$  的取值无关。若过程状态向量空间为  $S$ , 状态向量表示为  $X(t_i) = x_i, x_i \in S, i = 1, 2, \dots, n$ 。则  $X(t_{n+1})$  的条件概率分布函数  $P(X(t_n))$  可以表示为

$$P(X(t_n)) = P(X(t_n) = x_n | X(t_{n-1}) = x_{n-1}). \quad (1)$$

当马尔可夫过程中的时间和状态变量均为离散值时称为马尔可夫链。状态转移概率指从某一状态

转移到其他状态的概率,通常可以表示为状态转移概率矩阵  $A$ , 其表达式为

$$A = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nm} \end{bmatrix} \quad (2)$$

式中:  $p_{ij}$  ( $i, j=1, 2, \dots, n$ ) 表示从状态  $S_i$  转换到状态  $S_j$  的概率,  $p_{ij} \in [0, 1]$  并且满足  $\sum_{j=1}^n p_{ij} = 1$ 。

## 2.2 建立出行预测的马尔可夫模型

根据 1.1 节预处理地铁出行数据,主要选定了 3 个出行目的地,即住宅、公司和其他。一个含有 3 个状态的转移概率示意图如图 7 所示。

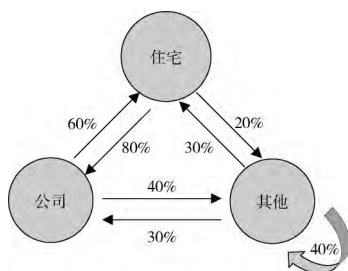


图 7 状态转移概率示意图

由图 2、图 4 及图 5 对用户出行次数的时间分布图来看,无论何类用户的出行在 1 d 中的不同时间段都呈现较大的差异,例如在早晚高峰和下午时间段的出行次数相差甚远,因此出行的状态转移概率也应该分别计算。为了更好地计算转移概率,将每天的时刻量划分为 4 个时间段:11:00 之前为第 1 个时间段,称为上午(M);11:00—16:00 为第 2 个时间段,称为下午(A);16:00—20:00 为第 3 个时间段,称为傍晚(E);20:00 之后为第 4 个时间段,称为晚上(N)。分为 4 个时间段后,相应地计算  $A(M)$ 、 $A(A)$ 、 $A(E)$ 、 $A(N)$  这 4 个转移矩阵。

对于每一个用户来讲,每次出行都表示为含有 3 个元素的元组,即  $(P, S_1, S_2)$ , 其中  $P \in \{M, A, E, N\}$  代表行程发生时间段,即上午(M)、下午(A)、傍晚(E)和晚上(N);  $S_1 \in \{1, 2, 3\}$  代表始发站点,  $S_2 \in \{1, 2, 3\}$  代表到达站点,其中 1 代表住宅站点,2 代表工作站点,3 代表其他站点。每个时间段有 7 种转移状态,行程元组可以取值的集合如图 8 所示,其中  $U$  表示所有元组组成的集合,该集合为  $\{(M, 1, 2), (M, 1, 3), (M, 2, 1), (M, 2, 3), (M, 3, 1), (M, 3, 2), (M, 3, 3), (A, 1, 2), (A, 1, 3), (A, 2, 1), (A, 2, 3), (A, 3, 1), (A, 3, 2), (A, 3, 3), (E, 1, 2), (E, 1, 3), (E, 2, 1), (E, 2, 3), (E, 3, 1), (E, 3, 2), (E, 3, 3), (N, 1, 2), (N, 1, 3), (N, 2, 1), (N, 2, 3), (N, 3, 1), (N, 3, 2), (N, 3, 3)\}$ 。以  $(A, 2, 1)$  为例,表示该用户在下午时间段从公司到住宅的行程。

$U$							
	(M,1,2)	(M,1,3)	(M,2,1)	(M,2,3)	(M,3,1)	(M,3,2)	(M,3,3)
	(A,1,2)	(A,1,3)	(A,2,1)	(A,2,3)	(A,3,1)	(A,3,2)	(A,3,3)
	(E,1,2)	(E,1,3)	(E,2,1)	(E,2,3)	(E,3,1)	(E,3,2)	(E,3,3)
	(N,1,2)	(N,1,3)	(N,2,1)	(N,2,3)	(N,3,1)	(N,3,2)	(N,3,3)

图 8 行程元组集合

## 3 仿真结果与性能分析

### 3.1 训练马尔可夫模型

在数据预处理阶段,以 10 min 为间隔将用户出行时间量化为时隙便于建模分析。采用交叉验证方法训练和测试模型,将 4 个月的原始出行数据按月份分为 4 组,每次使用 3 个月的数据作为训练集,剩余 1 个月的数据作为测试集,交叉预测用户出行记录 4 次以检验模型性能。由于选取的状态空间有 3 个状态变量,所以每一个状态转移矩阵的维度为  $3 \times 3$ 。虽然每一类用户有相似的出行行为,但是个体间出行差异很大,因此为每个用户单独训练出行模型,同时根据不同的时间段分别训练不同的状态转移矩阵。某通勤用户 M 时间段的状态转移概率矩阵见表 1。

表 1 某通勤用户 M 时间段的状态转移概率矩阵

出发站点	住宅	公司	其他
住宅	0	0.917	0.083
公司	0	0	1
其他	0	0.667	0.333

从表 1 可以看出,每一行的概率和等于 1,同时认为住宅到住宅和公司到公司的行程无效,它们的转移概率总是为 0。其余 7 种状态转移情况根据每个用户 3 个月内的行程数据进行训练,分别得到每个用户 4 个时间段的状态转移矩阵  $A(M)$ 、 $A(A)$ 、 $A(E)$ 、 $A(N)$ 。

### 3.2 预测用户下一时隙目的地

每个用户的出行模型训练完毕后,首先在测试集中随机挑选 2 条连续记录,第一条作为已知的  $n$  时隙记录,后一条作为下一时隙出行时需要预测的真实记录,其到达站点为所需要预测的目的地。然后根据随机挑选的  $n$  时隙,将训练集中  $n+1$  时隙的所有记录取出,若在历史记录中  $n+1$  时隙的出行概率足够大,认为下一时隙将会出行,否则认为该用户不会出行,仍然停留在  $n$  时隙的到达站点。若认为下一时隙会出行,首先判断下一时隙位于哪个时间段,将  $n$  时隙的到达站作为下一时隙的始发站,根据



此时间段的状态转移矩阵,将转移概率最大的状态判定为下一时隙的目的地。此外,如果在挑选测试记录时跨越了2 d,即后一条的时隙要比前一条小,则认为需要预测的出行为第2天的首次出行。最后将预测的目的地与真实记录进行比对,若一致,则判定此用户预测成功。采用差分整合移动平均自回归(ARIMA( $p, d, q$ ))模型作为对比模型,该模型是时间序列预测分析方法之一,将自回归模型、移动平均模型和差分法结合,其中参数 $p$ 为自回归项数, $d$ 为对数据进行差分的阶数, $q$ 为移动平均项数。经过对ARIMA模型调参优化后,选定 $p=4, d=0, q=1$ 为模型参数。2种模型测试方法均采用4折交叉验证,在测试完全部10万个用户后,2种模型的预测准确率见表2。结果显示个体马尔可夫出行模型的平均准确率可以达到72.73%,优于传统时序预测模型。本文的实验代码采用Python语言编写,实验所用CPU为Intel Core i5-6500, RAM为8 GB。

表2 2种模型的预测准确率 %

测试集	差分整合移动 平均自回归模型	马尔可夫模型
5月	34.26	67.63
6月	36.93	72.29
7月	37.92	73.40
8月	37.70	77.61
平均值	36.70	72.73

#### 4 结 语

准确地预测用户出行对于理解人类移动行为有很大帮助,同时也有利于部署智能交通分布,缓解交通拥堵,合理配置交通资源。在本项研究中,使用公共交通智能卡数据来预测出行,数据集为上海轨道交通系统在2017年5月—8月的刷卡信息。首先判断用户属性、住宅和工作站点,以便进一步分析。之后从多方面统计分析了出行信息,这有利于理解用户出行方式的共性,提高建立模型的合理性。最后使用3个月数据作为训练集对每个用户建立了出行预测模型,剩余1个月数据测试模型的准确性。测试结果显示本文的模型可以达到一个理想的准确率。本文的创新点在于将用户和站点合理分类,同时建立了每个用户的个体出行模型并研究了稀疏乘客的出行行为。本文的不足是对于其他类型的站点预测没有达到一个理想的结果,这可能需要进一步理解用户的其他出行习惯。未来的工作除了要建立高阶次预测模型外,还要着眼于人与人之间的社会关系带来的出行影响,这可能对提高预测准确性有极大帮助。

(由于印刷关系,查阅本文电子版请登录:<http://www.paper.edu.cn/journal/zgkjlw.shtml>)

#### [参考文献] (References)

- [1] GONZALEZ M C, HIDALGO C A, BARABASI A L. Understanding individual human mobility patterns [J]. *Nature*, 2008, 453(7196): 779-782.
- [2] YONG Nuo, NI Shunjiang, SHEN Shifei, et al. An understanding of human dynamics in urban subway traffic from the Maximum Entropy Principle [J]. *Physica A: Statistical Mechanics and its Applications*, 2016, 456: 222-227.
- [3] JIANG Shixiong, GUAN Wei, ZHANG Wenyi, et al. Human mobility in space from three modes of public transportation [J]. *Physica A: Statistical Mechanics and its Applications*, 2017, 483: 227-238.
- [4] YONG Nuo, NI Shunjiang, SHEN Shifei, et al. Uncovering stable and occasional human mobility patterns: a case study of the Beijing subway [J]. *Physica A: Statistical Mechanics and its Applications*, 2018, 492: 28-38.
- [5] YANG Yang, CHAWLA N V, BASU P, et al. Link prediction in human mobility networks [C]// *IEEE/ACM International Conference on Advances in Social Networks Analysis & Mining (ASONAM 2013)*. Piscataway, NY: IEEE Press, 2013: 380-387.
- [6] ALEXANDER L, JIANG Shan, MURGA M, et al. Origin-destination trips by purpose and time of day inferred from mobile phone data [J]. *Transportation Research Part C: Emerging Technologies*, 2015, 58: 240-250.
- [7] JIANG Shan, YANG Yingxiang, GUPTA S, et al. The TimeGeo modeling framework for urban mobility without travel surveys [J]. *Proceedings of the National Academy of Sciences*, 2016, 113(37): E5370-E5378.
- [8] SONG Xuan, KANASUGI H, SHIBASAKI R. DeepTransport: prediction and simulation of human mobility and transportation mode at a citywide level [C]// *International Joint Conference on Artificial Intelligence*. Palo Alto, CA: AAAI Press, 2016: 2618-2624.
- [9] YANG Chao, YAN Fenfan, UKKUSURI S V. Unraveling traveler mobility patterns and predicting user behavior in Shenzhen metro system [J]. *Transportmetrica A: Transport Science*, 2018, 14(7): 576-597.
- [10] GAMBS S, KILLIJIAN M O, del PRADO CORTEZ M N D P. Next place prediction using mobility Markov chains [C]// *Proceedings of the First Workshop on Measurement, Privacy, and Mobility*. New York: ACM, 2012: 3.
- [11] XU Yanyan, COLAK S, KARA E C, et al. Planning for electric vehicle needs by coupling charging profiles with urban mobility [J]. *Nature Energy*, 2018, 3(6): 484-493.
- [12] TOOLE J L, COLAK S, STURT B, et al. The path most traveled: travel demand estimation using big data resources [J]. *Transportation Research Part C: Emerging Technologies*, 2015, 58: 162-177.