

引用格式:方志祥,倪雅倩,张韬,等.利用终端位置时空转移概率预测通讯基站服务用户规模[J].地球信息科学学报,2017,19(6):772-781. [Fang Z X, Ni Y Q, Zhang T, et al. 2017. Using terminal location spatio-temporal transfer probability to predict subscriber base size of communication base station. Journal of Geo-information Science, 19(6):772-781.] DOI:10.3724/SP.J.1047.2017.00772

利用终端位置时空转移概率预测通讯基站服务用户规模

方志祥^{1,2},倪雅倩¹,张 韬³,冯明翔¹,于 冲¹

1. 武汉大学 测绘遥感信息工程国家重点实验室,武汉 430079; 2. 地球空间信息技术协同创新中心,武汉 430079;
3. 中国移动通信集团湖北有限公司业务支撑中心,武汉 518055

Using Terminal Location Spatio-temporal Transfer Probability to Predict Subscriber Base Size of Communication Base Station

FANG Zhixiang^{1,2*}, NI Yaqian¹, ZHANG Tao³, FENG Mingxiang¹ and YU Chong¹

1. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China;
2. Collaborative Innovation Center of Geospatial Technology, Wuhan 430079, China; 3. Business Support Center, Hubei Mobile, Wuhan 430040, China

Abstract: The prediction of the service population of cell phone tower plays an important role in the optimization of the spatial location of the cell phone towers and the configuration of the communication service bandwidth in mobile communications, and also provides the decision-making basis for early warning of human convergence and prevention of group incidents in urban management. This study proposed a prediction method of population in a region on a citywide scale, considering the human movement between cell phone towers. Based on the Markov chain and Bayesian probability theorem, we calculated the transition probability between different cell phone towers using massive mobile phone trajectories and we acquired the transition probability matrix which was distinct in different periods of one day. We made full use of the spatiotemporal transition probability to quantify the spatial and temporal characteristics of the mobile phone user's intra-urban movement. This study applies massive historical mobile phone location data to model training and proposes a citywide prediction model of the service population of the cell phone tower based on the spatiotemporal transition probability model. The algorithm proposed is verified by a mobile phone location dataset within thirty days collected by Hubei Mobile. This study shows that the prediction accuracy rate can reach about 94.8% and the proposed algorithm performed a good prediction with a temporal granularity of 60 minutes. Moreover, this study analyzed the prediction performance of the spatiotemporal transition probability model in different time granularities, and made comparison with other methods, such as the Castro model and moving average method. The results indicated that the proposed algorithm outperforms the Castro's model and moving average method when the temporal granularity is larger than 20 minutes. The proposed prediction method takes account of the spatial and temporal characteristics of human mobility and provides more accurate prediction results.

Key words: mobile phone location data; human mobility; spatiotemporal transition probability; population pre-

收稿日期 2017-01-03;修回日期:2017-05-10.

基金项目 国家自然科学基金项目(41231171、41371420);湖北省青年英才开发计划项目;武汉大学自主科研项目拔尖创新人才类资助项目(2042015KF0167)。

作者简介 方志祥(1977-),男,教授,主要从事时空行为建模、导航与位置服务研究。E-mail: zxfang@whu.edu.cn

diction; Markov chain

*Corresponding author: FANG Zhixiang, E-mail: zxfang@whu.edu.cn

摘要 基站服务用户数的预测对通信基站的空间位置选取、通讯服务带宽的配置优化等有重要作用,并为城市公共安全管理方面中的人群聚集预警与群体事件预防提供辅助决策支持。本文利用海量手机轨迹数据,运用时空转移概率量化描述城市内不同区域间人群流动的时空特性,结合马尔可夫链和贝叶斯定理,构建手机用户群体在基站间的时空转移概率模型,并以此提出一种城市区域尺度上的基站服务用户数预测方法。利用湖北省某城市长达30 d的手机轨迹数据进行模型训练与预测方法验证,实验结果表明:在时间粒度为60 min时,本文提出的方法对8-22时各时段基站服务用户数预测准确率都大于94.8%;在不同时间粒度下对比本文方法、Castro模型、移动平均法,发现在时间粒度大于20 min时,本文方法预测准确率高于另外两种方法。

关键词 手机数据;人群移动;时空转移概率;人数预测;马尔可夫链

1 引言

随着手机用户群体的不断增大,基站的空间位置设定和通讯服务带宽的动态配置成为保障高质量通讯的关键。通讯基站的容量管理需要考虑当前用户数及未来的发展趋势,有些学者采用遗传算法^[1]、启发式算法^[2]等对基站容量规划进行研究。Hakim等^[3]考虑交通出行引起的人口密度变化因素,对基站容量规划进行优化,但此类研究还缺乏对手机用户在基站间移动规律的理解^[4],对基站服务用户数的预测缺乏用户的动态溯源分析及其影响评估。从人群移动角度对基站服务用户数的预测,为人群聚集预警、城市安全管理等提供辅助决策支持。

近年来,手机位置数据在人类移动模式^[5-7]、移动行为预测^[8-10]、城市计算^[12-14]等方面的研究中被广泛使用。与本文研究有关的个体移动行为预测方面:Song等^[9]发现人类移动在时空上表现出高度的可预测性,93%的人类移动可以被预测;Monreale等^[10]根据人的移动模式构建决策树预测个体移动。在利用手机位置数据进行区域人数预测方面,Dash等^[15]利用动态贝叶斯对个体进行预测,基于当前位置和时刻聚类个体移动来实现群体移动的预测。此外,时间序列预测^[16-17]、马尔可夫模型^[18]、神经网络^[19]等方法也常被用于区域人数预测,如戴蓉蓉等^[16]使用ARIMA模型对武汉市的人群变化进行季节性模拟和预测;刘红梅等^[17]利用灰色模型预测景区旅游人数;Fan等^[18]构建基于相似度聚类的马尔可夫模型实现人数预测。

综上所述,现有区域人数预测研究一般由个体预测扩展到群体^[15],然而由于城市区域人数人口较多,导致计算量过大,并且个体预测的不可确定性

对城市区域尺度上人口预测影响程度也有待深入研究,需要建立群体层面高效的预测方法。时间序列预测方面的研究^[16-17]主要关注区域人数的时序变化,较少考虑现实中人群移动带来的动态累积影响。基于Markov模型的预测方法^[18]考虑当前时段区域人数和人群转移特性,并作为下一时段的预测规律,但由于人类移动存在时空差异,需要集成相邻时段人群转移规律的差异。

因此,本文从群体角度出发,采用手机位置数据,考虑人群移动的时空差异,结合马尔可夫链的无后效性^[21]和贝叶斯定理^[20],提出一种城市区域尺度上的基站服务用户数预测方法。该方法计算手机位置的时空转移概率,构建基站服务用户数预测模型,实现城市区域内的基站服务用户数预测。

2 手机位置时空转移概率

2.1 时空转移概率

转移概率是马尔可夫相关理论^[21]中的重要概念,用来描述状态空间中从一个状态到下一状态的转换过程。根据马尔可夫理论的定义,设定状态 n 马尔可夫链 $\{X_n, n \in T\}$ 的转移概率为条件概率 $p_{x_n, x_{n+1}}^n$,计算公式如式(1)所示。

$$p_{x_n, x_{n+1}}^n = P\{X_{n+1} = x_{n+1} | X_n = x_n\} \quad (1)$$

式中: $p_{x_n, x_{n+1}}^n$ 表示从当前状态 n 到下一状态 $n+1$,状态由 x_n 转化为 x_{n+1} 的概率。

手机用户在时刻 t 的空间位置表示为 $l_n = (\text{longitude}_n, \text{latitude}_n, t)$,其中 t 表示时刻, $(\text{longitude}_n, \text{latitude}_n)$ 表示用户所在的手机基站的经纬度坐标,简记为 (lon_n, lat_n) 。不同时刻的用户空间位置构成移动轨迹 $L_n = (l_0, l_1, l_2, \dots, l_n)$,其中 n 为轨

迹记录条数。本文将空间位置的变化视为状态变化,即时段 t 用户所在的基站位置 $l_n=(lon_n, lat_n, t)$ 视作状态 x_n ,相应地下一状态 x_{n+1} 表示为 $l_{n+1}=(lon_{n+1}, lat_{n+1}, t+1)$,则用户群体的位置时空转移概率为 $p'_{x_n, x_{n+1}}=P(l_{n+1}|l_n)$ 。

手机位置时空转移概率用于定量描述手机用户群体在不同基站间移动的可能性。从手机定位数据中可以统计到各基站的服务用户数 N'_i 、基站 i 和基站 j 间的手机用户人流数量 $flow'_{i,j}$ 。 $flow'_{i,j}$ 表示时段 t 手机用户轨迹中出现当前位置为基站 i ,下一位置为基站 j 的次数。根据贝叶斯定理^[20,23],手机用户群体从基站 i 流向基站 j 的时空转移概率 $p'_{i,j}$ 的计算公式为:

$$p'_{i,j}=p(l_j|l_i)=\frac{p(l_j)p(l_i|l_j)}{\sum_{k=1}^m p(l_k)p(l_i|l_k)}=\frac{\frac{N_j^{t+1}}{N_j^{t+1}} \frac{flow'_{i,j}}{N_j^{t+1}}}{\sum_{k=1}^m \frac{N_k^{t+1}}{N_j^{t+1}} \frac{flow'_{i,k}}{N_k^{t+1}}} \quad (2)$$

$$=\frac{flow'_{i,j}}{\sum_{k=1}^m flow'_{i,k}}$$

式中: i, j 表示时段 t 和时段 $t+1$ 用户群体所在的基站; m 表示城市区域的手机基站个数; $flow'_{i,j}$ 表示时段 t 从基站 i 移动到基站 j 的手机用户数; N_j^{t+1} 表示时段 $t+1$ 基站 j 的服务用户数; N^{t+1} 表示 $t+1$ 所有基站服务的总用户数。

2.2 时空转移概率矩阵

手机用户群体在基站间移动的可能性可以通过时空转移概率矩阵表示,如图1给出了基站间人群流动的示意图。当只考虑手机用户从单个基站 i 移动到 m 个基站时(图1(a)), t 时段手机用户群体从基站 i 移动到基站 $1, 2, \dots, m$,时空转移概率构成一个行向量 P'_i ,可表达为 $P'_i=[p'_{i,1}, p'_{i,2}, \dots, p'_{i,m}]$ 。

图1(b)表示手机用户在多基站间移动的示意图。多基站间的人群移动时空转移概率可表达为一个 $m \times m$ 的二维矩阵 \mathbf{P}' ,如式(3)所示。其中,转移概率矩阵 \mathbf{P}' 满足矩阵元素 $0 \leq p'_{i,j} < 1$,并且每行所有元素相加之和为1。

$$\mathbf{P}'=[p'_{i,j}]=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1j} & p_{1m} \\ p_{21} & p_{22} & \cdots & \vdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ p_{i1} & \vdots & \cdots & p_{ij} & \vdots \\ p_{m1} & \cdots & \cdots & \cdots & p_{mm} \end{bmatrix} \quad (3)$$

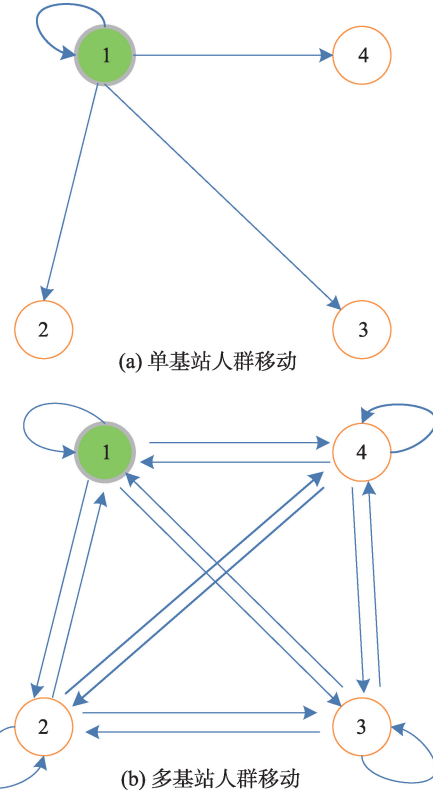


图1 基站间人群移动示意图

Fig. 1 The crowd movement between cell phone towers

时空转移概率矩阵是对基站区域间人口流动的量化表达,时空转移概率矩阵 \mathbf{P}' 的不对称性能体现基站间用户流动的不平衡性,可以反映城市基站间人口流动的空间特征。不同时段人群移动时空转移概率矩阵存在的差异,则体现出手机用户出行的时间特征。

3 基于时空转移概率的基站服务用户数预测

3.1 Castro的预测模型

Castro的模型^[22]利用马尔可夫链的无后效性,对不同时段的交通流量进行预测。该模型假设城市区域车辆数量保持不变,在不同时间粒度下统计各时段的交通流量,计算车辆的时空转移概率矩阵 \mathbf{P}' ,如式(4)构建交通流量预测模型。该模型被广泛用于城市交通流量的短时预测。利用马尔可夫链中状态转移的思想,在研究基站服务用户数预测时,可用当前时段各基站服务用户数 N' 和转移概率矩阵 \mathbf{P}' 对下一时段各基站服务用户数 $N_predict^{t+1}$ 进行预测,如式(4)所示:

$$N_predict^{t+1} = \mathbf{P}^t \times \mathbf{N}^t \quad (4)$$

式中: \mathbf{P}^t 为时段 t 用户群体的转移概率矩阵, $\mathbf{P}^t = [p_{ij}^t]$, $0 \leq p_{ij}^t \leq 1$, 并且满足 $\forall i, j \leq m, \sum_{j=1}^m p_{ij}^t = 1$; \mathbf{N}^t 为 t 时段各基站的服务用户数向量。

3.2 模型改进

本文的基站服务用户数预测方法将贝叶斯理论^[20,23]和马尔可夫链的无后效性^[21]结合起来, 计算手机用户群体在基站间的时空转移概率, 在 Castro 模型的基础上进行改进, 构建基站服务用户数预测方法。对 Castro 模型的改进具体为如下 2 个方面:

(1) 现实中由于存在城市间的人口流动、手机开关机等现象, 城市内基站服务的手机用户总数不断变化, 使模型中基站服务的手机用户总数不变的假设不成立。本文针对手机用户总数波动的现象, 利用历史轨迹数据确定相邻时段基站服务的手机用户总数的改正值, 根据相邻时段基站服务用户数变化的历史规律, 加入相邻时段基站服务用户数的改正项 $\Delta N^{t \rightarrow t+1}$ 、转移概率矩阵改正项 $\Delta \mathbf{P}^{t \rightarrow t+1}$, 对 Castro 预测模型中各基站的服务用户数进行改进。

首先, 计算相邻时段基站服务用户数的改正项 $\Delta N^{t \rightarrow t+1}$ 。利用训练数据统计得到各时段基站服务用户数, 分别计算每天相邻时段的基站服务用户数变化量 $\Delta N^{day, t \rightarrow t+1}$, 上标 day 用来标识不同日期。根据训练数据中基站服务用户数变化量的最大值 $\Delta N_{\max}^{t \rightarrow t+1}$ 和最小值 $\Delta N_{\min}^{t \rightarrow t+1}$, 平均分为 n 个状态区间 $[a_k, b_k]$, 统计不同日期基站服务用户数变化量 $\Delta N_{t \rightarrow t+1}^{day}$ 在各个区间 $[a_k, b_k]$ 的概率 p_k 。按照式(5)对区间的均值进行加权平均得到基站服务用户数的改正项 $\Delta N^{t \rightarrow t+1}$ 。取区间均值进行计算, 尽量减少突发事件引起的人数变化极端值对基站服务用户数改正项 $\Delta N^{t \rightarrow t+1}$ 的影响。

$$\Delta N^{t \rightarrow t+1} = \sum_{k=1}^n p_k \times \frac{1}{2}(a_k + b_k) \quad (5)$$

式中: $a_k = \Delta N_{\min}^{t \rightarrow t+1} + \frac{k}{n}(\Delta N_{\max}^{t \rightarrow t+1} - \Delta N_{\min}^{t \rightarrow t+1})$, $b_k = a_k + \frac{1}{n}(\Delta N_{\max}^{t \rightarrow t+1} - \Delta N_{\min}^{t \rightarrow t+1})$, $k=0, 1, \dots, n-1$ 。 $\Delta N_{\max}^{t \rightarrow t+1}$ 、 $\Delta N_{\min}^{t \rightarrow t+1}$ 分别表示训练数据中相邻时段 $t, t+1$ 基站服务用户数变化量的最大值、最小值。

同样, 计算相邻时段人群时空转移概率矩阵的改正项 $\Delta \mathbf{P}^{t \rightarrow t+1}$ 。不同时段人群移动规律的不同会引起基站间人群移动时空转移概率的改变, 时空转

移概率矩阵的改正项 $\Delta \mathbf{P}^{t \rightarrow t+1}$ 体现了人群在不同时段移动特性的差异, 其计算是以矩阵中基站 i, j 间的时空转移概率差异 $\Delta p_{ij}^{t \rightarrow t+1}$ 为基础的。由每天相邻时段 $t, t+1$ 的转移概率 p_{ij}^t 、 p_{ij}^{t+1} 计算基站 i, j 间转移概率的差异值 $\Delta p_{ij}^{day, t \rightarrow t+1}$ 。将每天相邻时段基站间的转移概率差异值 $\Delta p_{ij}^{day, t \rightarrow t+1}$ 划分 n 个区间, 按式(6)计算基站 i, j 间时空转移概率的改正项 $\Delta p_{ij}^{t \rightarrow t+1}$ 。由此可以得到, 相邻时段人群移动时空转移概率矩阵的改正项 $\Delta \mathbf{P}^{t \rightarrow t+1} = [\Delta p_{ij}^{t \rightarrow t+1}]$ ($i, j=1, 2, \dots, m; m$ 表示基站个数)。

$$\Delta p_{ij}^{t \rightarrow t+1} = \sum_{k=1}^n p_k \times \frac{1}{2}(a_k + b_k) \quad (6)$$

式中: $a_k = \Delta p_{ij, \min}^{t \rightarrow t+1} + \frac{k}{n}(\Delta p_{ij, \max}^{t \rightarrow t+1} - \Delta p_{ij, \min}^{t \rightarrow t+1})$, $b_k = a_k + \frac{1}{n}(\Delta p_{ij, \max}^{t \rightarrow t+1} - \Delta p_{ij, \min}^{t \rightarrow t+1})$, $k=0, 1, \dots, n$ 。 $\Delta p_{ij, \max}^{t \rightarrow t+1}$ 、 $\Delta p_{ij, \min}^{t \rightarrow t+1}$ 分别表示训练数据中, 相邻时段基站 i 与基站 j 间时空转移概率差异值的最大值、最小值。

(2) 具体到各基站服务用户数变化的计算, 此前 Castro 预测模型的流量分配没有考虑相邻时段移动规律的改变。本文的模型加入时空转移概率矩阵改正项 $\Delta \mathbf{P}^{t \rightarrow t+1}$, 利用改进后的转移概率矩阵 \mathbf{P}' 对基站服务用户数 $\Delta N^{t \rightarrow t+1}$ 进行基站服务用户数的分配。基站间时空转移概率改正项 $\Delta \mathbf{P}^{t \rightarrow t+1}$ 的实际意义在于定量描述不同时段城市区域间的人群移动规律差异。按照式(7)可得到更接近时段 $t+1$ 实际用户移动规律的转移概率矩阵 \mathbf{P}' , 对 \mathbf{P}' 进行按比例计算使之依然满足各行元素之和为 1 的条件。最后, 按照改进后的转移概率矩阵 \mathbf{P}' 将时段 $t+1$ 的基站服务用户数变化量 $\Delta N^{t \rightarrow t+1}$ 分配到各个基站, 按式(8)得到 $t+1$ 时段各基站服务用户数的预测值。

$$\mathbf{P}' = \mathbf{P}^t + \Delta \mathbf{P}^{t \rightarrow t+1} \quad (7)$$

$$N_predict^{t+1} = \mathbf{N}^t + \Delta N^{t \rightarrow t+1} \times \mathbf{P}' \quad (8)$$

3.3 方法主要流程

图 2 为本文方法的基本流程, 主要分为 3 个阶段: 数据准备、模型训练、模型预测与评估。首先, 统计不同时段手机用户在基站间的流量 $flow_{ij}^t$ 和基站的服务用户数 N^t , 按照式(2)计算用户群体在基站间的时空转移概率, 构建时空转移概率矩阵 \mathbf{P}^t 。然后, 利用历史轨迹数据进行预测模型的训

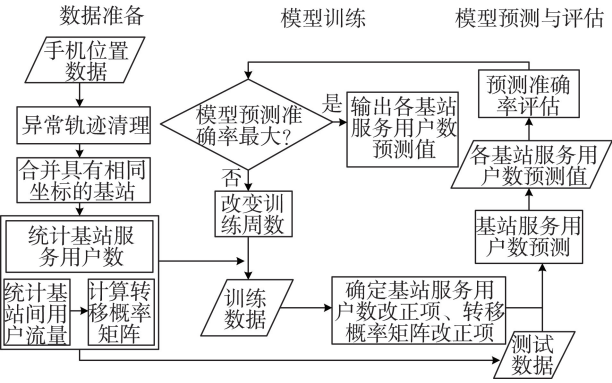


图2 方法基本流程

Fig. 2 Basic procedures of the prediction method

练,按式(5)、(6)计算相邻时段基站服务用户数改正项 $\Delta N^{t \rightarrow t+1}$ 和转移概率矩阵改正项 $\Delta P^{t \rightarrow t+1}$ 。最后,运用式(8)进行基站服务用户数预测,利用预测准确率对本文提出的方法的预测性能进行分析和评估。通过实验确定最佳的训练周数、计算改正项。

4 实验结果与分析

4.1 实验数据

本研究采用的实验数据为某城市2015年8月10日至2015年9月27日长达7周的手机位置数据。该城市位于湖北省东南部,总人口80.52万人,面积1200.35 km²。手机位置数据以手机基站的经纬度坐标记录手机用户的位置,如表1所示,包括经过脱敏处理后的手机用户ID、记录产生的时间、基站编号、基站经纬度坐标。手机位置数据的匿名化处理机制,有效地保护了用户隐私,同时为群体角度的城市区域人群移动研究提供了数据基础。图3为手机基站分布密度图。

本实验区域的手机位置数据包括43.2万手机用户的主动和被动定位信息,即当用户位置发生基站间变更、接打电话、收发信息时用户的位置信息将被记录(被动记录),当用户长时间(超过1 h小

表1 手机用户轨迹数据实例

Tab. 1 The examples of one user's records

用户ID	时间	基站编号	基站经度	基站纬度
58****	07:32	11**	115.****	29.****
58****	08:32	11**	115.****	29.****
58****	09:27	12**	115.****	29.****
...
58****	21:06	10**	115.****	29.****
58****	21:34	11**	115.****	29.****

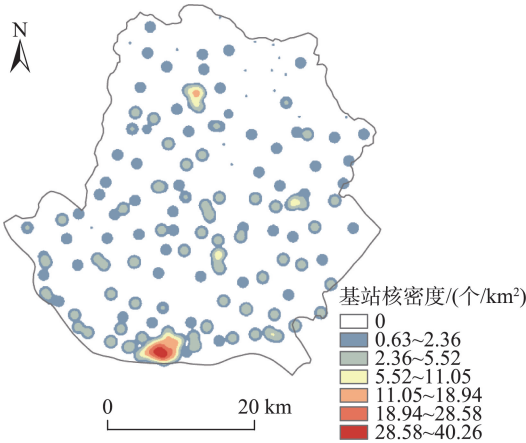


图3 基站核密度图

Fig. 3 Spatial kernel density of cell phone towers

时)未发生上述行为时,手机将会主动捕捉其所在的位置(主动记录)。因此每个手机用户一天中至少会产生24条定位数据。考虑实际生活中人们的日常出行,本次实验采用该城市每天活动记录在30~1500条之间的24万用户的手机轨迹数据,一天内手机用户记录数目多达940万条。实验对原始轨迹数据进行相同坐标的基站合并和针对基站信号跳变的轨迹清洗等预处理,并统计了不同时段各基站的服务用户数和基站之间的用户的流动信息,用于对工作日基站服务用户数的预测。由于部分日期的手机轨迹数据残缺不全,实际可用于训练的数据为26天的工作日数据。

4.2 实验结果分析

为了方便评价不同基站服务用户数预测的准确率,实验按照式(9)定义预测准确度 $R_{i,t}$,评价不同时段各基站服务用户数预测的准确性。考虑到人群移动的时空特性,不同时段基站服务用户数的预测准确度会发生改变。为了更好地分析不同时段基站服务用户数预测的准确率,实验统计该城市所有基站的预测准确度 R 的平均值、中位数,用来评价方法的准确性,用四分位数之差(3Q-1Q)评价模型的稳定性。

$$R_{i,t} = \left(1 - \frac{|N_{i,t}^{pr} - N_{i,t}|}{N_{i,t}} \right) \times 100\% \quad (9)$$

式中: $N_{i,t}$ 和 $N_{i,t}^{pr}$ 依次表示时段 t 时基站 i 的服务用户数的统计值和预测值。

实验前,本研究对训练数据的天数进行了分析。实验以周为单位依次使用长达1周、2周、3周、4周、5周、6周的手机定位数据分别进行模型训练,

采用第7周的基站服务用户数进行预测,并评估模型的准确性和稳定性。时间粒度为60 min时,不同训练周数对应的模型预测准确率变化如图4所示,随着训练周数的增加,模型预测准确率稳步增加,当训练周数增长到3周以上时模型预测准确率趋于稳定,在94.7%上下波动。从实验结果来看,在对工作日基站服务用户数进行预测时,选取长达3周的工作日数据进行训练,可有效掌握相邻时段历史人群移动规律,也能保证对现实变化过程的捕捉。实验采用3周的工作日数据作为模型的训练数据,进行后续的对比分析。

4.2.1 不同方法的结果对比

本实验选取3.1节中Castro的预测模型、基于

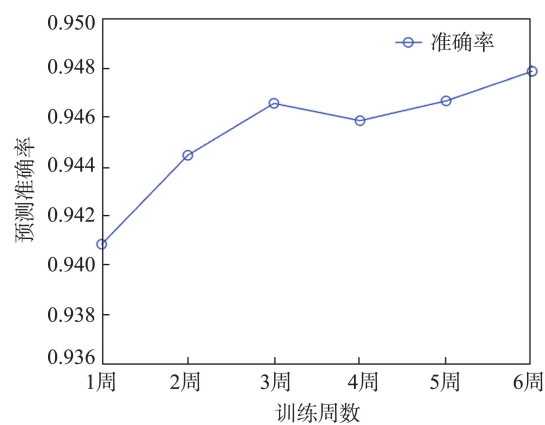


图4 不同训练周数下的准确率比较
Fig. 4 Accuracy comparison with different training days

时间序列的移动平均法^[24]进行对比。移动平均法是一种经典高效的时间序列预测方法,仅使用各个时段的基站服务用户数即可进行预测。实验以60 min为时间粒度,根据式(9)定义的预测准确度,对本文提出的时空转移概率模型与Castro的预测模型、基于时间序列的移动平均法进行基站服务用户数预测的对比分析,以证明本文考虑人群移动时空特性和使用历史数据训练之后方法的有效性与准确性。

表2列出了3种模型预测准确率的统计值。对比60 min为时间粒度时,3种方法预测准确率的平均数、中位数,结果表明本文提出的时空转移概率方法在一天的多个时段预测准确率都保持较高水平,其中对基站服务用户数的平均预测准确率不低于94.6%,比Castro的模型预测准确率高5.1%,特别在人数波动较大、预测难度较大的8-9时和19-20时,运用前3周工作日手机数据进行训练获取时空改正项之后,模型准确率至少提高了4.2%(见表2的相应时段)。从时间序列角度出发的移动平均法在各个时段的预测准确率波动不大,平均预测准确率为79.79%。移动平均法在不同时段预测准确率的四分位数之差Q3-Q1较大,这主要是由于移动平均法是一种常用的时间序列预测方法,该方法受历史数据的影响很大,没有考虑基站之间的人群流动的时空特性,当基站服务用户数出现突增突减时,预测准确率明显低于另外两种方

表2 3种模型的准确率对比
Tab. 2 Accuracy results of three prediction models

时刻	时空转移概率模型			Castro的模型			移动平均法		
	均值	中值	Q3-Q1	均值	中值	Q3-Q1	均值	中值	Q3-Q1
8时	0.950	0.956	0.064	0.896	0.907	0.105	0.819	0.821	0.171
9时	0.963	0.962	0.051	0.939	0.942	0.079	0.821	0.821	0.165
10时	0.966	0.965	0.051	0.943	0.944	0.079	0.800	0.805	0.179
11时	0.963	0.962	0.050	0.943	0.944	0.078	0.802	0.811	0.175
12时	0.955	0.955	0.058	0.937	0.940	0.100	0.778	0.791	0.197
13时	0.966	0.966	0.048	0.942	0.943	0.094	0.774	0.788	0.196
14时	0.964	0.964	0.054	0.941	0.941	0.084	0.777	0.788	0.192
15时	0.966	0.967	0.051	0.947	0.950	0.075	0.782	0.792	0.186
16时	0.961	0.964	0.054	0.941	0.943	0.078	0.791	0.800	0.182
17时	0.957	0.960	0.057	0.918	0.925	0.092	0.786	0.803	0.183
18时	0.946	0.948	0.079	0.915	0.920	0.131	0.802	0.799	0.180
19时	0.948	0.951	0.080	0.895	0.906	0.192	0.798	0.799	0.177
20时	0.965	0.963	0.050	0.922	0.916	0.128	0.785	0.790	0.181
21时	0.968	0.967	0.050	0.903	0.906	0.142	0.770	0.782	0.203

法。图5(a)给出了不同模型预测准确率分布情况,本文方法的预测准确率优于Castro的模型、基于时间序列的移动平均法。

图5(b)、(c)、(d)表示时间粒度为60 min时,3种方法所对应的基站服务用户数预测准确率的箱图,超出区间 $[Q1-1.5IQR, Q3+1.5IQR]$ 的值为异常值,图5中红色的小圆表示异常值。其中, $Q1$ 、 $Q3$ 分别表示下四分位数、上四分位数, $IQR=Q3-Q1$ 表示四分位距。四分位数之差($Q3-Q1$)主要用于测度各种方法预测准确率的离散程度。由表2中可知,本文提出的时空转移概率预测模型对不同基站的预测准确率的四分位数之差处于0.048–0.080之间,Castro的模型预测准确率四分位数之差处于0.075–0.192,基于时间序列的移动平均法则处于0.165–0.203。本文方法的预测准确率四分位数之

差小于其他2种方法,说明本文的预测方法对不同基站服务用户数的预测具有很好的稳定性和适用性。相对移动平均法来说,时空转移概率模型异常值较多,但从异常值的分布来看,时空转移概率模型的异常值也保持在较高水平,在0.7–0.8之间,Castro的模型则集中在0.6–0.75之间,移动平均法的异常值准确率则低于0.5。因此,60 min为时间粒度时,本文提出的时空转移概率预测模型,能更加准确和稳定地预测基站服务用户数。

4.2.2 不同时间粒度下的模型结果分析

对20、40、60、80、100和120 min 6个时间粒度下本文模型的预测准确率进行对比分析。图6为不同时间粒度下预测准确率对比图,除了20 min这一时间粒度,基站服务用户数的预测准确率随着时间呈现相似的变化趋势。在人群流动频繁的17–19时

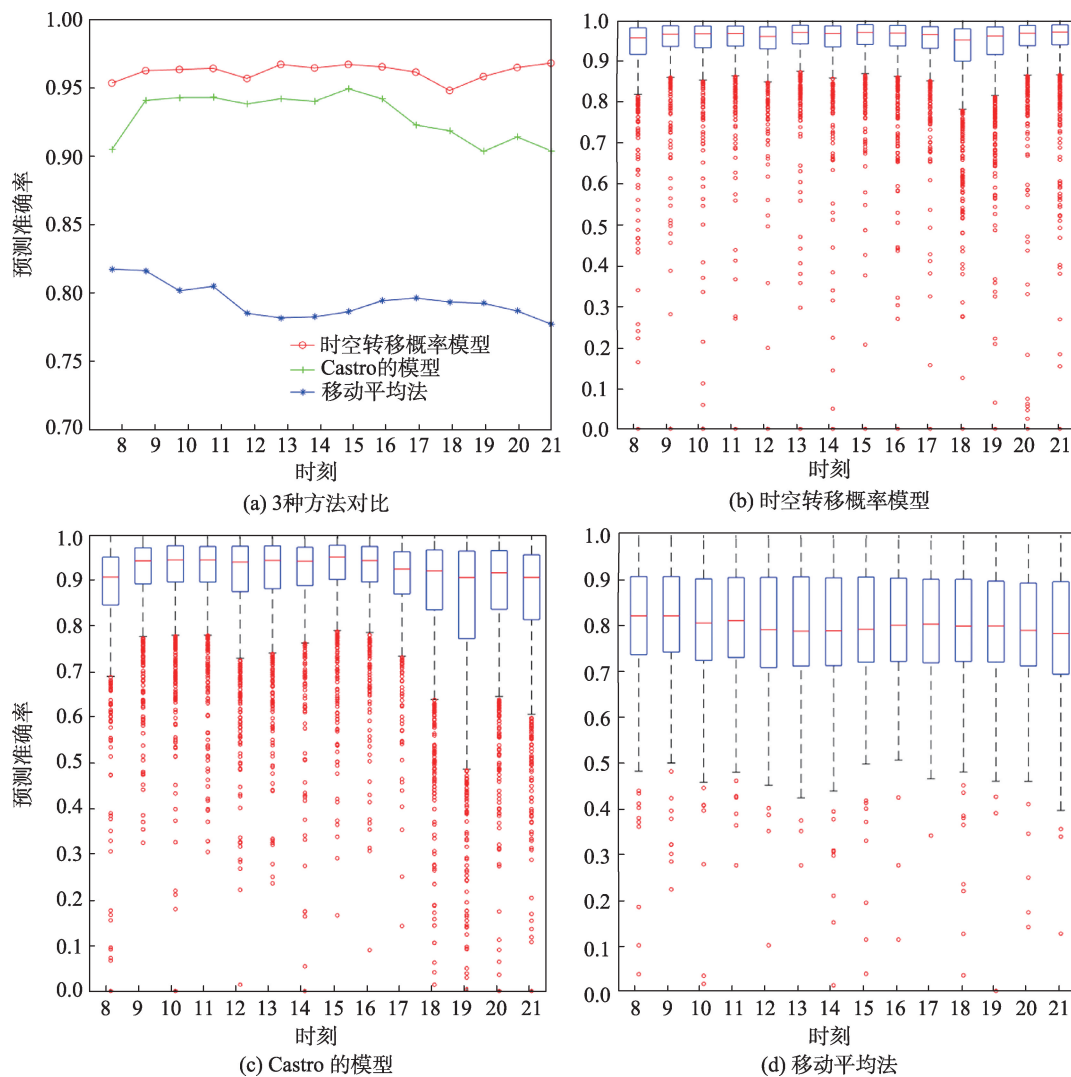


图5 模型对比和预测准确率分布

Fig. 5 The accuracy distributions of three prediction models

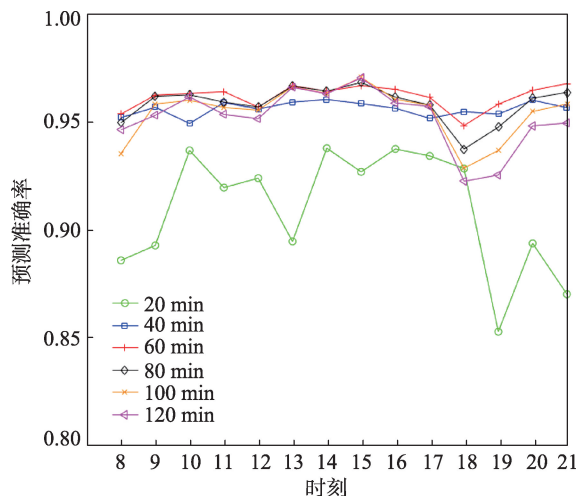


图6 不同时间粒度下时空转移概率模型的准确率对比
Fig. 6 Comparisons of prediction accuracy in different temporal granularities

晚高峰时段,模型预测准确率降低了2%。14–16时为一 天中基站服务用户数预测最为准确的时段,偏差小于4%。

以 20 min 为时间粒度时,本文模型的预测准确率较低,且各时段波动较大,在人群流动频繁的早晚高峰和午休时段,体现得尤为明显。例如,在 13–14 时和 19–20 时,以人群流动频繁的时段的基站服务用户数和基站间用户流动为基础进行下一时段的预测,时空转移概率模型预测准确率大大降低。

在 40 min、60 min 为时间粒度时,基站服务用户数的预测准确率随着时间粒度的增大逐渐提高,并且各个时段预测效果稳定,具体的准确率统计见表 3。以 60 min 为时间粒度的基站服务用户数预测

表3 不同时间粒度下模型平均预测准确率

Tab. 3 Average prediction accuracy in different temporal granularities

时间粒度/min	时空转移概率模型	Castro的模型	移动平均法
20	0.910	0.924	0.819
40	0.957	0.939	0.807
60	0.963	0.931	0.799
80	0.959	0.910	0.804
100	0.955	0.890	0.809
120	0.953	0.871	0.812

平均预测准确率比 40 min 粒度时高出 3%。以 80、100 和 120 min 为时间粒度时,预测准确率并没有因为时间粒度增大、基站间的手机用户流动数据更加充足而有所提升,这是由于时间粒度越大,受用户移动规律的影响,前一时段的基站服务用户数与所预测的时段用户数相差较大。

表 3 列出了不同时间粒度下 3 种模型的平均预测准确率。在 20 min 的短时预测中,Castro 的模型略优于时空转移概率方法,当时间粒度变为 40 min 及以上时,本文方法的预测准确率在各个时段都高于 Castro 的模型。移动平均法对时间粒度不敏感,在 6 个时间粒度下,预测准确率几乎不变,都处在较低水平。从现势性上来说,20 min 相对而言时间粒度小,需要更好地反映当前的基站真实服务用户数,但受临时性事件等不确定因素的影响较大。图 7 给出了 20 min 和 40 min 时间粒度下 3 种方法在 8–21 时的预测准确率,20 min 时间粒度下在 12–13 时、18–19 时等人口流动频繁的时段,Castro 的模型优

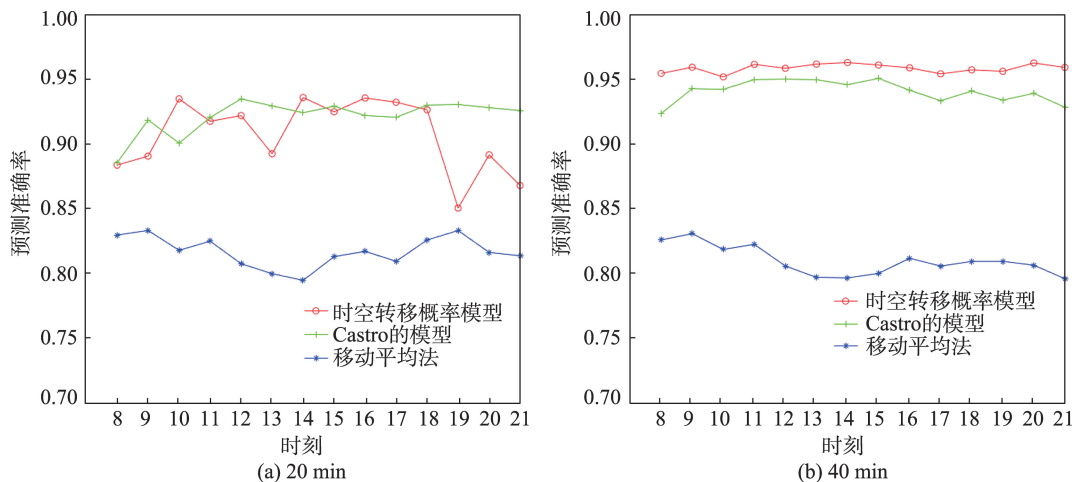


图7 3种方法在不同时间粒度的准确率对比

Fig. 7 Accuracy comparison between three methods in different temporal granularities

于本文模型。在人口流动相对较少的时段,如10-11时、14-16时,本文模型优于Castro的模型,历史轨迹数据在基站服务用户数的短时预测上的作用并不明显。参照图5(a),当时间粒度变为40 min和60 min时,历史轨迹数据对基站服务用户数预测准确率提高了5.1%。因此,在小于20 min的短时预测上,Castro的模型具备一定的优越性;在大于20 min的时间粒度下,本文所提出的模型预测性能相对较好。

5 结语

本文结合马尔可夫链和贝叶斯定理,从群体角度出发提出了一种基于手机位置时空转移概率的基站服务用户数预测方法。该方法综合考虑城市人群移动的时空特性,适用于城市区域尺度的区域人数预测。实验表明:该方法的平均预测准确率达到94.8%,比Castro的模型高5.1%,比基于时间序列的移动平均法高17%。本文提出的方法定量化表达人类移动时空特性,具有较好的手机基站服务用户数预测准确率。该方法除了为基站设置和带宽配置提供智能决策依据,还可的人群聚集预警、公众出行服务等提供城市人口分布预测信息参考,辅助预防开放区域人群踩踏等安全事故。本文研究目前只关注工作日的规律,对周末出行模式以及周五的模式由于数据受限,尚未做深入对比,将在后续研究中陆续展开;此外,该预测算法应用于不同规模城市区域人数的预测效果还有待验证。

参考文献(References):

- [1] Meng H, Long F, Guo L, et al. Cooperating base station location optimization using genetic algorithm[C]// Chinese Control and Decision Conference, 2016:4820-4824.
- [2] Lin C H R, Liao H J, Lin Y C, et al. An efficient tabu search for cell planning problem in mobile communication[J]. Wireless Communications & Mobile Computing, 2016,16(4):486-496.
- [3] Ghazzai H, Yaacoub E, Alouini M, et al. Optimized LTE cell planning with varying spatial and temporal user densities[J]. IEEE Transactions on Vehicular Technology, 2016,65(3):1575-1589.
- [4] 萧世伦,方志祥.从时空GIS视野来定量分析人类行为的思考[J].武汉大学学报·信息科学版,2014,39(6):667-670. [Shaw S L, Fang Z X. Rethinking human behavior research from the perspective of space-time GIS[J]. Geomatics and Information Science of Wuhan University, 2014,39(6):667-670.]
- [5] 周涛,韩筱璞,闫小勇,等.人类行为时空特性的统计力学[J].电子科技大学学报,2013,4(4):481-540. [Zhou T, Han X P, Yan X Y, et al. Statistical mechanics on temporal and spatial activities of human[J]. Journal of University of Electronic Science and Technology of China, 2013,4(4):481-540.]
- [6] Hasan S, Schneider C M, Ukkusuri S V, et al. Spatiotemporal patterns of urban human mobility[J]. Journal of Statistical Physics, 2013,151(1):304-318.
- [7] 杨喜平,方志祥,赵志远,等.城市人群聚集消散时空模式探索分析——以深圳市为例[J].地球信息科学学报,2016,18(4):486-492. [Yang X P, Fang Z X, Zhao Z Y, et al. Exploring urban human spatio-temporal convergence-dispersal patterns: A case study of Shenzhen city, Journal of Geo-information Science, 2016,18(4):486-492.]
- [8] Do T M T, Gatica-Perez D. Where and what: Using smartphones to predict next locations and applications in daily life[J]. Pervasive & Mobile Computing, 2014,12(10):79-91.
- [9] Song C, Qu Z, Blumm N, et al. Limits of predictability in human mobility[J]. Science, 2010,327(5968):1018-1021.
- [10] Monreale A, Pinelli F, Trasarti R, et al. WhereNext: a location predictor on trajectory pattern mining[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28 - July. 2009:637-646.
- [11] Xu Y, Shaw S L, Zhao Z, et al. Another tale of two cities: understanding human activity space using actively tracked cellphone location data[J]. Annals of the American Association of Geographers, 2016,106(2):489-502.
- [12] 许宁,尹凌,胡金星.从大规模短期规则采样的手机定位数据中识别居民职住地[J].武汉大学学报·信息科学版,2014,39(6):750-756. [Xu N, Yin L, Hu J X. Identifying home-work locations from short-term, large-scale, and regularly sampled mobile phone tracking data[J]. Geomatics and Information Science of Wuhan University, 2014, 39(6):750-756.]
- [13] Deville P, Linard C, Martin S, et al. Dynamic population mapping using mobile phone data[J]. Proceedings of the National Academy of Sciences of the United States of America, 2014,111(45):15888-93.
- [14] 吴健生,黄力,刘瑜,等.基于手机基站数据的城市交通流量模拟[J].地理学报,2012,67(12):1657-1665. [Wu J S, Huang L, Liu Y, et al. Traffic flow simulation based on call detail records[J]. Acta Geographica Sinica, 2012,67

- (12):1657-1665.]
- [15] Dash M, Koo K K, Krishnaswamy S P, et al. Visualize people's mobility-both individually and collectively-using mobile phone cellular data[C]//Mobile Data Management (MDM), 2016 17th IEEE International Conference on. IEEE, 2016,1:341-344.
- [16] 戴蓉蓉,朱海红,李霖.基于ARIMA模型的市内人群移动预测[J].测绘工程,2016,25(2):38-41. [Dai R R, Zhu H H, Li L. Intra-urban human mobility prediction based on ARIMA model[J]. Engineering of Surveying and Mapping, 2016,25(2):38-41.]
- [17] 刘红梅,刘建平.基于灰色模型的韶山红色旅游景区客源预测[J].经济地理,2010,30(6):1047-1051. [Liu H M, Liu J P. Predicting tourists of Shaoshan red tourism scenic spots on gray model[J]. Economic Geography, 2010,30(6):1047-1051.]
- [18] Fan Z, Song X, Shibasaki R, et al. CityMomentum: an on-line approach for crowd behavior prediction at a citywide level[C]//Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2015:559-569.
- [19] Liang V C, Ma R T B, Ng W S, et al. Mercury: Metro density prediction with recurrent neural network on streaming CDR data[C]//Data Engineering (ICDE), 2016 IEEE 32nd International Conference on. IEEE, 2016: 1374-1377.
- [20] Spiegelhalter D J. Bayesian Analysis in Expert Systems [J]. Statistical Science, 1993,8(3):219-247.
- [21] Meyn S P, Dsc R L T. Markov Chains and Stochastic Stability[M]. Singapore: World Scientific, 1999.
- [22] Castro P S, Zhang D, Li S. Urban traffic modelling and prediction using large scale taxi GPS traces[C]//International Conference on Pervasive Computing. Springer Berlin Heidelberg, 2012:57-72.
- [23] Sun S, Zhang C, Yu G. A Bayesian network approach to traffic flow forecasting[J]. IEEE Transactions on Intelligent Transportation Systems, 2006,7(1):124-132.
- [24] Box G E P, Jenkins G M, Reinsel G C, et al. Time series analysis: forecasting and control[M]. New Jersey: John Wiley & Sons, 2015.