基于随机矩阵理论的城市人群移动行为分析*

徐赞新 王 钺† 司洪波 冯振明

(清华大学电子工程系,北京 100084)

(2010 年 5 月 20 日收到;2010 年 7 月 13 日收到修改稿)

移动通信应用为人类移动规律的研究提供了独特的数据来源. 本文通过城市手机用户的分布数据,研究城市移动人群的整体动力学行为. 借助随机矩阵理论的方法,通过比较移动人群数据与随机数据在互相关矩阵谱分布上的差异,发现移动人群数据互相关矩阵的相关系数均值、最大本征值及其对应的本征向量明显偏离于随机互相关矩阵的分布,指出这种差异体现了城市移动人群的整体行为特性,且这种差异在不同时间段也会有所不同. 研究结果体现出相关系数的均值和最大本征值的波动趋势,并指出本征向量成员权重的时空模式与城市移动人群整体行为特征的波动过程非常符合.

关键词: 随机矩阵理论, 移动人群, 宏观行为

PACS: 05.40.-a

1. 引 言

认识和理解城市人群的移动规律对于提高城市人口、资源和环境的综合管理能力具有重大的意义. 而越来越丰富的手机信息则为研究城市人群的移动行为提供了独特的数据来源[1]. 通过上百万手机用户通话记录的分析,得以发现人们社会网络结构与通信网络结构的耦合关系[2]. 带有手机用户位置信息和移动轨迹的数据则揭示了个人移动的规律性,以及个人移动模式之间的相似性[3]. 而对短消息记录的研究则发现了人们在短信活动中体现出来的重尾现象[4]. 此外,利用手机用户在每天不同时间段的位置信息,研究人员得出米兰城市中多个区域手机用户的活动强度分布[5]. 而 MIT 媒体实验室则将手机话务量信息与人群的移动模式联系起来,发现了手机用户移动行为的周期性[6].

相对于个体移动规律的研究,目前对于群体移动规律的关注还较少. 与个人移动规律的研究不同,群体移动规律的研究则更关注宏观层次的行为特征. 针对群体移动规律的现有研究,主要是基于随机过程理论的方法,将个人移动模型加以合成得到群体移动的随机过程模型[3,7,8];也有从社交网络、城市规划、流行病学等视角出发的研究[9—11]. 本文则借助随机矩阵理论的方法,研究城市移动人群的宏观行为.

随机矩阵理论(random matrix theory, RMT)通过比较随机的多维时间序列统计特性,可以体现实际数据中对随机的偏离程度,并揭示实际数据中整体关联的行为特性. 正是这种特定的视角,使得 RMT 被广泛应用于物理、金融数学、生物统计、网络科学等广阔的应用领域[12—19]. 基于 RMT,研究人员分析了美国股票收益率互相关矩阵最大本征值在 1962—1997 年的变化趋势,突出了 1987 年 10 月美国股市的崩盘行为特征[18]. 在大规模互联网流量的动力学研究中,也发现最大本征值对应的本征向量各成员权重的时空模式,能够很好的刻画各网络节点流量的时空关联性,以及互联网流量的整体动力学行为[19].

本文借助随机矩阵理论的方法研究了城市移动人群的整体动力学行为. 基于网格化的城市手机用户的分布数据,分析比较了移动人群数据与随机数据在互相关矩阵谱分布上的差异,发现移动人群数据互相关矩阵的相关系数均值、最大本征值及其对应的本征向量明显偏离于随机互相关矩阵的分布,指出这种差异体现了城市移动人群的整体行为

* 国家自然科学基金(批准号:60674048, 60603068, 60772053, 60672142, 60932005),国家重点基础研究发展计划(批准号:2007CB307100-2007CB307105)资助的课题.

† 通讯联系人. E-mail:wangyue@tsinghua.edu.cn

©2011 中国物理学会 Chinese Physical Society

特性,且这种差异在不同时间段也会有所不同. 研究结果体现出相关系数的均值和最大本征值在的波动趋势,并指出本征向量成员权重的时空模式与城市移动人群整体行为的波动过程非常符合.

2. 移动人群数据互相关矩阵的谱分析

2.1. 数据来源与构成

如图 1 所示为某城市的某一移动运营商通过“移动基站手机用户信息监测系统”得到的城市移动人群密度分布情况. 在空间尺度上,此系统使用网格化的建模方式将城市划分成 2136 个 1 km × 1 km 的有效单元网格,然后通过蜂窝通信系统采集到的每个基站数据,估算出每个单元网格的移动人群分布. 而在时间尺度上,系统每隔一个小时采集一次数据,即每当整点时刻完成一次数据的收集、汇总、计算和存储. 该系统估算的手机用户数与实际的移动人群数相比,90% 以上的网格估算误差可以控制在 5% 以内,99% 以上的网格估算误差可以控制在 10% 以内.

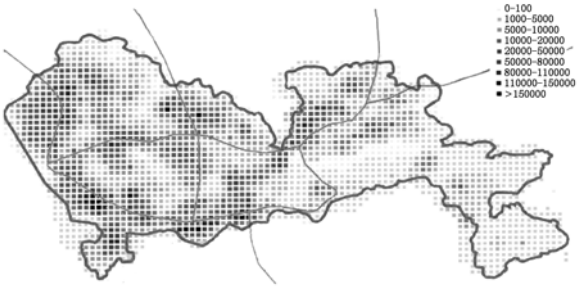

图 1 城市移动人群密度分布

原始数据包含从 2008 年 12 月—2009 年 10 月这 11 个月的移动人群分布,我们选取 1419 个网格构成三组数据矩阵. 如表 1 所示,三组数据分别记为 $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ 和 $\boldsymbol{P}_3$,对应的互相关矩阵分别记为 $\boldsymbol{C}_1$,$\boldsymbol{C}_2$ 和 $\boldsymbol{C}_3$. 每组数据中所有网格数据的时间跨度都为 90 天,由时间分辨率为 1 h 可知网格数据长度为 90 × 24 = 2160 h,得到矩阵的行为 $N = 1419$,列为 $L = 2160$.

表 1 三组数据构成

起止时间	数据矩阵	互相关矩阵
2008 年 12 月—2009 年 02 月	$\boldsymbol{P}_1$	$\boldsymbol{C}_1$
2009 年 03 月—2009 年 05 月	$\boldsymbol{P}_2$	$\boldsymbol{C}_2$
2009 年 06 月—2009 年 08 月	$\boldsymbol{P}_3$	$\boldsymbol{C}_3$

2.2. 相关系数分布

为了计算各网格移动人群数据的相关系数,定义网格 i 的归一化人数为 $m_i(t)$,以及 i,j 两网格归一化人数的相关系数为 C_{ij}

$$m_i(t) = \frac{M_i(t) - \langle M_i(t)\rangle}{\sigma_i}, \tag{1}$$

$$C_{ij} = \langle m_i(t) m_j(t)\rangle, \tag{2}$$

其中 $M_i(t)$ 代表第 i 个网格在第 t 时刻的人数,$\sigma_i = \sqrt{\langle M_i^2(t)\rangle - \langle M_i(t)\rangle^2}$ 是第 i 个网格人数的标准差,$\langle\cdots\rangle$ 表示时间平均.

图 2 给出 $\boldsymbol{C}_1$,$\boldsymbol{C}_2$ 和 $\boldsymbol{C}_3$ 与随机互相关矩阵的相关系数分布差异,其中 $\boldsymbol{R}_{\rm rm}$ 是仿真生成的由 $N = 1419$ 行,$L = 2160$ 列的高斯序列组成的随机矩阵的互相关矩阵. 可见,$\boldsymbol{R}_{\rm rm}$ 的相关系数是中心为零的对称分布,而 $\boldsymbol{C}_1$,$\boldsymbol{C}_2$ 和 $\boldsymbol{C}_3$ 的相关系数是中心大于零的非对称分布,即三者的相关系数均值都大于零,且以 $\boldsymbol{C}_1$ 的均值最大.

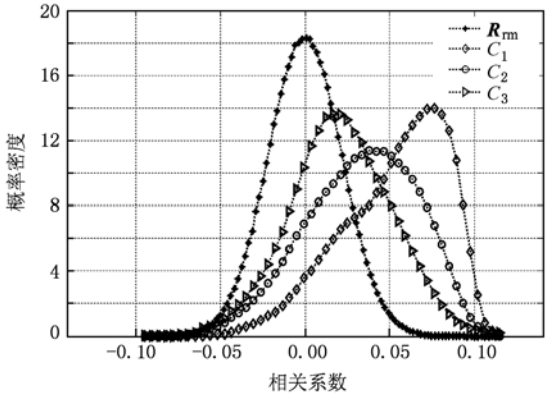

图 2 相关系数的概率密度

与随机互相关矩阵在相关系数分布上的差异说明各网格间移动人群的行为并不是随机独立的,它们之间存在内在的关联性. 此外,相关系数均值大于零意味着不同网格移动人群的行为特征更趋于一致. 且在不同时间段,相关系数均值也有所不同,说明在不同时间段,不同网格间移动人群在行为特征上的关联强度也会所有不同.

2.3. 本征值分布

比较互相关矩阵 $\boldsymbol{C}_1$,$\boldsymbol{C}_2$ 和 $\boldsymbol{C}_3$ 与随机互相关矩阵的本征值分布差异之前,首先分析随机互相关矩阵的本征值分布. 如(3)式所示,$\boldsymbol{R}$ 为一随机互相关矩阵.

$$\boldsymbol{R} = \frac{1}{L}\boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}, \tag{3}$$

其中 $\boldsymbol{E}$ 是一个 $N \times L$ 的随机矩阵,包含 N 行长度为 L 的时间序列,这 N 行时间序列独立同分布,其均值和方差分别为 0 和 1. 定义 $Q = L/N(>1)$,当 Q 值固定,且 $N \to \infty$ 和 $L \to \infty$ 时,$\boldsymbol{R}$ 的本征值概率密度 $\rho(\lambda)$ 为[20]

$$\rho(\lambda) = \begin{cases} \dfrac{Q}{2\pi}\dfrac{\sqrt{(\lambda_{\max}-\lambda)(\lambda-\lambda_{\min})}}{\lambda}, & \lambda_{\min} \leqslant \lambda \leqslant \lambda_{\max}, \\ 0, & \text{其他} \end{cases} \tag{4}$$

$$\lambda_{\min} = (1 - 1/\sqrt{Q})^2, \tag{5}$$

$$\lambda_{\max} = (1 + 1/\sqrt{Q})^2, \tag{6}$$

其中 $\lambda_{\min}$ 和 $\lambda_{\max}$ 分别是 $\boldsymbol{R}$ 的最小和最大本征值. 由 $N = 1419$ 和 $L = 2160$ 可知 $Q = L/N = 1.5222$,代入(5)和(6)式可求得 $\lambda_{\min} = 0.0383$ 和 $\lambda_{\max} = 3.2355$.

图 3(a)给出 $\boldsymbol{R}_{\mathrm{rm}}$ 的本征值概率密度和 $\rho(\lambda)$ 的分布. 可见,$\boldsymbol{R}_{\mathrm{rm}}$ 的本征值概率密度分布与 $\rho(\lambda)$ 非常符合,而且几乎所有 $\boldsymbol{R}_{\mathrm{rm}}$ 的本征值都位于 $[\lambda_{\min}, \lambda_{\max}]$,说明随机互相关矩阵 $\boldsymbol{R}_{\mathrm{rm}}$ 在 N 和 L 不满足趋于无穷大的条件时,它的本征值概率密度分布与 $\rho(\lambda)$ 的分布差异不大.

图 3(b)给出 $\boldsymbol{C}_1, \boldsymbol{C}_2, \boldsymbol{C}_3$ 与 $\boldsymbol{R}_{\mathrm{rm}}$ 的本征值概率密度分布的比较,由图可知 $\boldsymbol{C}_1, \boldsymbol{C}_2$ 和 $\boldsymbol{C}_3$ 的大部分本征值都位于 $[\lambda_{\min}, \lambda_{\max}]$,但约有 20 个以上的大本征值以及几个小本征值偏离 $[\lambda_{\min}, \lambda_{\max}]$,其中以最大本征值的偏离程度最大. 此外,$\boldsymbol{C}_1, \boldsymbol{C}_2$ 和 $\boldsymbol{C}_3$ 的最大本征值分别为 724.27, 483.14 和 316.16,分别是 $\lambda_{\max}$ 的 221, 147 和 96 倍,远偏离于 $\lambda_{\max}$,而且以 $\boldsymbol{C}_1$ 的偏离最大. 上述的分析已知这种偏离不是由 N 和 L 不满足趋于无穷大的条件导致,而是由各网格移动人群在行为特征上的关联性导致,可以说这种偏离体现了城市移动人群的整体行为特性.

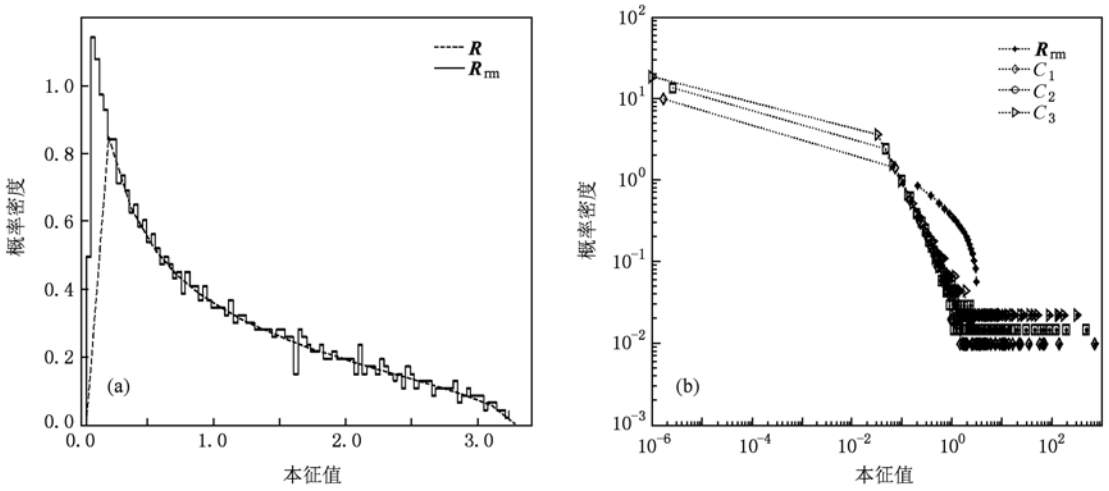

图 3 本征值概率密度 (a) $\boldsymbol{R}_{\mathrm{rm}}$ 与 $\boldsymbol{R}$ 的本征值概率密度;(b) $\boldsymbol{C}_1, \boldsymbol{C}_2, \boldsymbol{C}_3$ 和 $\boldsymbol{R}_{\mathrm{rm}}$ 的本征值概率密度

2.4. 本征向量分布

本节将比较互相关矩阵 $\boldsymbol{C}_1$ 与随机互相关矩阵 $\boldsymbol{R}_{\mathrm{rm}}$ 的本征向量成员分布的差异,并进一步分析本征向量与相关系数的关系. 以本征值的大小标记本征值与本征向量的下标:$\lambda_N \leqslant \cdots \leqslant \lambda_2 \leqslant \lambda_1$,对应的本征向量分别为 $\{w_N, \cdots, w_2, w_1\}$.

图 4(a),(b),(c)和(d)分别给出 $\boldsymbol{C}_1$ 与 $\boldsymbol{R}_{\mathrm{rm}}$ 的本征向量 $w_i(\lambda_{\min} < \lambda_i < \lambda_{\max})$,$w_N, w_2$ 以及 w_1 的成员概率密度分布. 由图 4(a)可知,当 $\lambda_{\min} < \lambda_i < \lambda_{\max}$ 时,w_i 成员的概率密度分布与 $\boldsymbol{R}_{\mathrm{rm}}$ 的分布非常符合. 在图 4(b)和(c)中,w_N 和 w_2 成员的概率密度分布与 $\boldsymbol{R}_{\mathrm{rm}}$ 的分布虽然存在一些差异,但都近似满足对称分布.

观察图 4(d)发现 w_1 成员的概率密度分布与 $\boldsymbol{R}_{\mathrm{rm}}$ 的分布差异最为显著,这说明最大本征值对应的本征向量 w_1 最能体现城市移动人群数据互相关矩阵和随机互相关矩阵在本征向量成员分布上的差异.

进一步观察图 4(d)发现 w_1 的各成员基本同

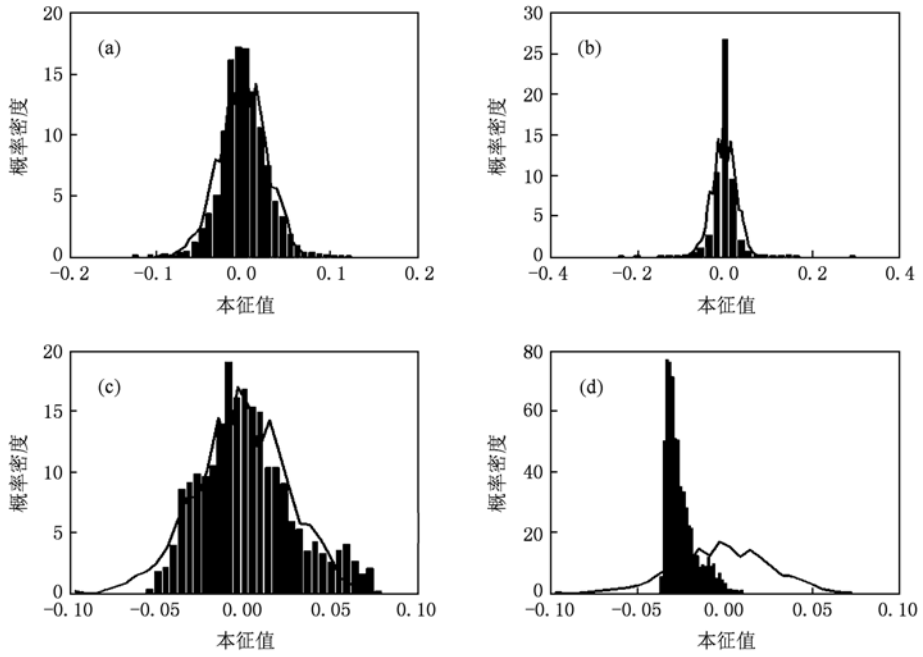

图4 $\boldsymbol{C}_1$ 和 $\boldsymbol{R}_{\rm rm}$ 的本征向量成员概率密度分布 (a) $w_i(\lambda_{\min}<\lambda_i<\lambda_{\max})$; (b) w_N; (c) w_2; (d) w_1

向,下面分析本征向量各成员具有同向特性的含义. 由本征值分解公式可知,相关系数可以写成本征值与本征向量的叠加

$$C_{ij} = \sum_{k=1}^{N} \lambda_k w_{ki} w_{kj}, \tag{7}$$

其中 w_{ki} 是第 k 个本征向量的第 i 个成员. 由本征向量的正交归一化条件可知

$$\sum_{l=1}^{N} w_{il} w_{jl} = \sigma_{ij}. \tag{8}$$

由(7)式可知,本征向量的各成员越趋于同向,对相关系数均值偏离零的贡献就越大. 由(8)式的正交归一化条件,以及 w_1 各成员基本同向的特点可知,w_1 的各成员基本满足 $w_{1k} \sim 1/\sqrt{N}, k = 1, 2, \ldots, N$,因此 w_1 对相关系数 C_{ij} 的贡献满足 $\lambda_1 w_{1i} w_{1j} \sim \dfrac{\lambda_1}{N}$.

实际上 $\boldsymbol{C}_1, \boldsymbol{C}_2$ 和 $\boldsymbol{C}_3$ 的 λ_1/N 值分别为 0.51, 0.34 和 0.22,相应的相关系数均值分别为 0.46, 0.26 和 0.11. 可以看出除 $\boldsymbol{C}_3$ 外,$\boldsymbol{C}_1$ 和 $\boldsymbol{C}_2$ 的 λ_1/N 值与相关系数均值非常接近,尤其以 $\boldsymbol{C}_1$ 的两者最为接近. 这说明最大本征值对应的本征向量成员对相关系数均值的贡献最大,且这种贡献在不同时间段也会有所不同.

3. 城市移动人群的整体行为分析

3.1. 移动人群总数的波动

图5给出了城市移动人群总数从2008年12月1日—2009年10月31日的逐日变化. 可以看出城市移动人群在一定时期内的变化不大,然而在节假日期间人数会发生突变现象. 大约从第50天开始到第60天结束的这段时间,移动人群开始大幅度减少,直到第60天达到谷底. 实际上这段时间正处于春节期间,春节期间由于大量人群返乡或外出旅游,城市移动人群减少一半左右,最低值出现在大

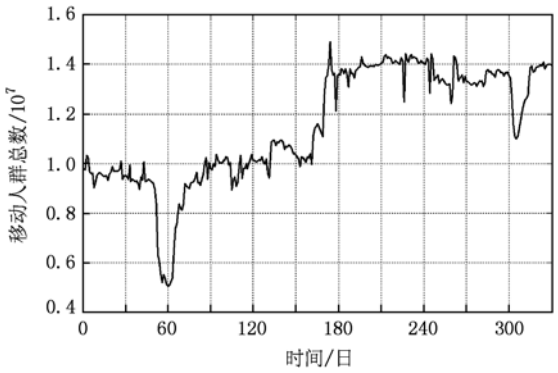

图5 城市移动人群总数的逐日变化

年初三(2009 年 1 月 28 日). 春节 7 天长假结束时约 50% 的人员返回该城市,当元宵节结束后,90% 以上的返乡人员均已返回该城市,之后移动人群数基本恢复到正常.

3.2. 相关系数均值和最大本征值的波动

为了分析相关系数均值和最大本征值的波动趋势,选取 70 个网格的移动人群数据,每个网格移动人群数据的长度为 7985 h,组成一个 70×7985 的数据矩阵. 以滑动时间窗口 720 h,离散步长 24 h 的方式对该数据矩阵进行互相关运算得到相关系数均值,以及通过本征值分解得到最大本征值.

图 6(a)和(b)分别给出了相关系数均值和最大本征值的波动趋势. 由图可知相关系数均值和最大本征值的波动趋势大致相同,都在第 32 天左右达到峰值,在第 58 天左右下降,峰值持续时间约为 27 天. 对比城市移动人群总数的波动趋势,可以发现相关系数均值和最大本征值的峰值起始时间不是在城市移动人群总数最少的第 59 天(2009 年 1 月 28 日),而是在移动人群总数开始逐渐减少的第 32 天(2009 年 1 月 1 日)左右,但两者的峰值持续时间与移动人群总数从大幅度减少到大幅度回升过程的持续时间一样,约为 27 天左右.

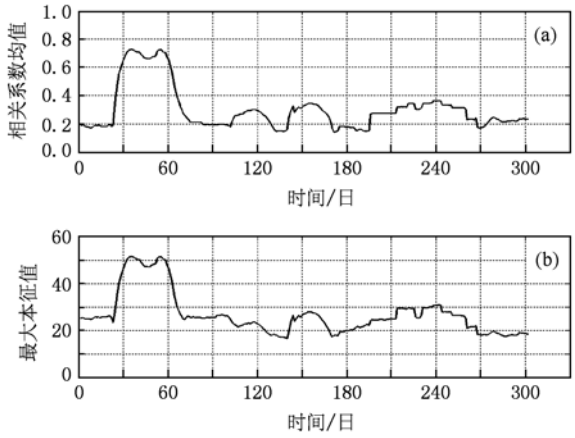

图 6 (a)相关系数均值分布;(b)最大本征值分布

值得注意的是,图 6(a)中相关系数均值分布的变化和图 6(b)中最大本征值分布的变化均超前于图 5 中城市移动人群总数的逐日变化,尤其以包含元旦和春节的这段时间最为明显. 首先,我们观察不同互相关运算时间窗口值对互相关系数均值和最大本征值分布的影响. 图 7(a)和(b)分别给出了互相关运算时间窗口值为 720,480 和 360 h 下的互相关系数均值和最大本征值的时间变化趋势. 由图 7 可知,随着互相关运算时间窗口值的变小,相关系数均值分布和最大本征值分布的变化对城市移动人群总数变化的超前时间随之变小,但是仍然存在时间上的超前. 其次,相关研究也曾发现类似的现象. 文献[19]通过观察最大本征值对应的本征向量各成员权重的时空变化模式,发现权重模式的变化在网络节点流量从一个稳定状态变化到另一个稳定状态的过渡阶段最为明显. 文献[18]使用最大本征值的变化趋势分析美国股票收益率的变化趋势,也发现最大本征值的变化行为在股市崩盘之前的过渡时间里表现的最为显著,一旦股市崩盘,最大本征值的变化反而变得缓和. 因此,针对本文的这种时间上的超前现象,我们也尝试给出一种可能的理解和解释,我们认为这种时间上的超前现象与城市移动人群总数变化的整体特性有着内在关联性. 观察图 5 可知,在元旦之前,城市移动人群总数相对稳定;在春节期间,城市移动人群总数达到谷底,但其变化幅度较小,并处于相对稳定状态;然而,从元旦到春节开始的这段时间,城市移动人群总数开始大幅度变化,处于一个波动的过渡过程. 此时,图 6 中相关系数均值和最大本征值的幅度也开始迅速增大,两者的时间变化趋势超前捕捉到城市移动人群总数的整体变化趋势.

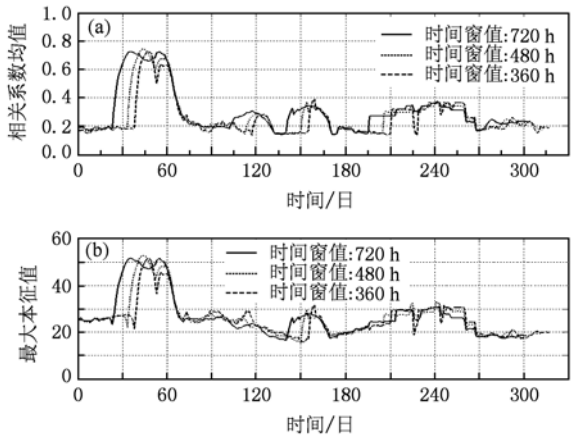

图 7 不同互相关运算时间窗口值的比较 (a)相关系数均值分布;(b)最大本征值分布

此外,在约第 300 天开始的一个星期时间,城市移动人群总数也出现类似的较大波动过程,但由于受互相关运算时间窗的约束,无法观察到相关系数均值和最大本征值在第 300 天后的波动趋势.

3.3. 本征向量成员权重的时空模式

由第二部分的分析比较结果可知,最大本征值对应的本征向量对相关系数均值的贡献最大,且相关系数均值越大,城市移动人群的整体行为特征越明显. 因此,最大本征值对应的本征向量与城市移动人群的整体行为特征应存在内在的关联性. 为了观察这种关联性,定义本征值对应的本征向量 w_k 成员的权重为

$$S_i^k = w_{ki}^2, \tag{9}$$

上式中,S_i^k 表示本征向量 w_k 的第 i 个成员的权重.

图 8 给出最大本征值对应的本征向量 w_1 的成员权重分布,即 S^1 的分布. 其中横轴为时间,纵轴为网格编号,高度为权重值. 由图可知在横向上可观察同一网格权重随时间的波动趋势;在纵向上可观察不同网格权重的大小. 结合横纵方向则可观察权重的时空联合模式,体现的是城市移动人群的宏观行为特征. 在图 8 中的第 32 天左右到第 58 天左右的时间段,所有网格的权重出现突变行为,形成明显的波谷,与城市移动人群整体行为的波动过程非常符合,且在时间上也有明显的提前.

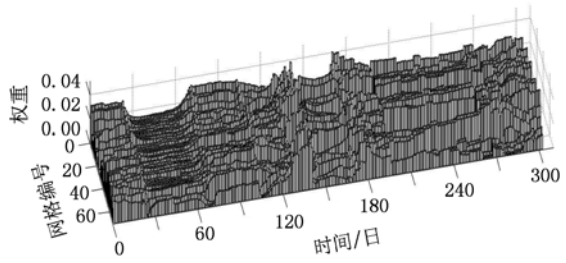

图 8 S^1 的分布

4. 结 论

本文借助随机矩阵理论的方法研究了城市移动人群的整体动力学行为. 基于网格化的城市移动人群的分布数据,比较了城市移动人群数据的互相关矩阵与随机互相关矩阵在谱分布上的偏离,具体表现在相关系数,本征值以及本征向量分布上的差异,发现正是这种差异体现了城市移动人群的整体行为特性.

在不同的时间段,移动人群数据互相关矩阵的相关系数均值和最大本征值的偏离程度,以及相应的本征向量对相关系数的贡献程度也存在着差异. 这种差异体现出不同时间段内城市移动人群的整体行为特性的不同,这促发本文进一步对相关系数均值和最大本征值的波动趋势进行分析,发现在城市移动人群总数发生大幅度波动的过程中,相关系数均值和最大本征值同时出现大幅度的波动,且在时间上比前者要提前,对于这种时间上的超前现象,我们也尝试给出了一种可能的理解和解释.

此外,为了观察最大本征值对应的本征向量与城市移动人群整体行为特征的关联性,定义了本征向量成员的权重. 通过观察最大本征值对应的本征向量成员权重的时空模式,发现城市移动人群总数发生大幅度波动的过程中,权重的时空模式会出现突变行为,在时间上也比前者要提前. 一旦城市移动人群总数的波动趋于缓和,权重的时空波动也趋于缓和.

[1] Bohannon J 2006 *Science* **314** 914

[2] Onnela J O, Saramaki J, Hyvonen J, Szabo G, Lazer D, Kaski K, Kertesz J, Barabasi A L 2007 *PNAS* **104** 18

[3] Marta C G, Cesar A H, Barabasi A L 2008 *Nature* **453** 7196

[4] Hong W, Han X P, Zhou T, Wang B H 2009 *Chinese Phys. Lett.* **26** 2

[5] Ratti C, Williams S, Frenchman D, Pulselli R M 2006 *EPB* **33** 5

[6] Reades J, Calabrese F, Sevtsuk A, Ratti C 2007 *IEEE Pervas Comput* **6** 3

[7] Barabasi A L 2005 *Nature* **435** 7039

[8] Vazquez A, Oliveira J, Dezso Z 2006 *Phys. Rev.* E **73** 3

[9] Watts D, Strogatz S 1998 *Nature* **393** 6684

[10] Musolesi M, Mascolo C 2007 *Mobile Computing and Communications Review* **11** 3

[11] O'Neill E, Kostakos V, Kindberg T 2006 *Ubiquitous Computing* **4206** 1

[12] Wigner E P 1967 *SIAM Review* **9** 1

[13] Chen Z Q, Zheng R R, Chen H, Yao C Q 2000 *Acta Phys. Sin.* **49** 5 (in Chinese) [陈志谦、郑仁蓉、陈 洪、姚纯青 2000 物理学报 **49** 5]

[14] Chen Z Q, Zheng R R 2001 *Chin. Phys.* **10** 12

[15] Li R, Yan P L, Chen J, Li J, Li J, Zhang K W, Zhong J X 2009 *Acta Phys. Sin.* **58** 10 (in Chinese) [李 蓉、颜平兰、陈 健、李 俊、李 金、张凯旺、钟健新 2009 物理学报 **58** 10]

[16] Zhang F Z, Wang J, Gu Y 1999 *Acta Phys. Sin.* **48** 12 (in Chinese) [张飞舟、王 娇、顾 雁 1999 物理学报 **48** 12]
[17] Xing Y Z, Xu G O 1999 *Acta Phys. Sin.* **48** 5 (in Chinese) [邢永忠、徐躬耦 1999 物理学报 **48** 5]
[18] Plerou V, Gopikrishnan P, Rosenow B, Amaral L A N, Guhr T 2002 *Phys. Rev.* E **65** 6
[19] Yuan J, Mills K 2005 *IEEE T DEPEND SECURE* **2** 4
[20] Sengupta A M, Mitra P P 1999 *Phys. Rev.* E **60** 3

Analysis of urban human mobility behavior based on random matrix theory*

Xu Zan-Xin Wang Yue† Si Hong-Bo Feng Zhen-Ming
(*Department of Electronic Engineering, Tsinghua University, Beijing* 100084, *China*)
(Received 20 May 2010; revised manuscript received 13 July 2010)

Abstract

Mobile communication applications provide a unique data source for the research of human mobility pattern. Based on the distribution data of urban mobile phone users, in this paper is explored the macroscopic dynamical behavior of urban mobility human by using the method of random matrix theory. The largest eigenvalue and the corresponding eigenvector of mobile phone user data deviate far from the distribution of random matrix. The deviations from random matrix vary with time. We find that the largest eigenvalue corresponds to a whole behavior common to all urban human mobility. The results indicate the temporal trends of the mean of correlation coefficient and the largest eigenvalue. We also find that the spatio temporal evolution of the weight of eigenvector components for the eigenvector corresponding to the largest eigenvalue is very consistent with the fluctuation pattern of the macroscopic behavior of urban human mobility.

Keywords: random matrix theory, mobility human, macroscopic behavior
PACS: 05.40.-a

* Project supported by the National Natural Science Foundation of China (Grant Nos. 60674048, 60603068, 60772053, 60672142, 60932005), the State Key Development Program for Basic Research of China (Grant No. 2007CB307100-2007CB307105).

† Corresponding author. E-mail: wangyue@tsinghua.edu.cn