

• 方法研究 •

移动用户画像构建研究

黄文彬 徐山川 吴家辉 王 军

(北京大学信息管理系, 北京 100871)

〔摘 要〕基站通信网络数据蕴含着丰富的移动用户行为, 从移动用户频繁活动、规律行为以及移动速度 3 方面建构移动用户行为画像, 可以为个性化服务提供更完整丰富的信息。在分析和挖掘某电信运营商 3 万位移动用户记录的基站数据的基础上, 本文采用频繁模式挖掘、构建概率矩阵、计算熵等方法, 从用户基站日志中所包含的地理位置信息中构建移动用户行为画像。研究结果表明, 该画像模型可显示移动用户的频繁活动规律、周期性行为及出行方式, 可作为分析移动用户群体行为及用户间交互行为的基础。

〔关键词〕移动数据; 移动行为; 移动用户; 用户画像; 周期性行为; 频繁序列

DOI: 10.3969/j.issn.1008-0821.2016.10.009

〔中图分类号〕G252.0 〔文献标识码〕A 〔文章编号〕1008-0821(2016)10-0054-08

The Profile Construction of the Mobile User

Huang Wenbin Xu Shanchuan Wu Jiahui Wang Jun

(Information Management, Peking University, Beijing 100871, China)

〔Abstract〕Cellular mobile data service logs contains massive mobile user behavior, based on which mobile user behavior profile, including frequent activities, periodic behaviors and user speed, can be built to provide rich information for personalized services. Analyzing 30,000 users' station data provided by a telecommunication operator, user profile is built from location sequence by means of frequent pattern mining, probability matrix constructing and entropy calculating. The user profile proposed shows that users' frequent behavior patterns, periodic behavior patterns and speed distribution could be an important reference for further research on population mobile behavior analysis and user interaction behavior.

〔Key words〕mobile data; user behavior; mobile user; user profile; periodic behavior; frequent sequences

由于智能手机的普及, 人们的生活行为与移动智能设备连接的更紧密, 截止到 2014 年 5 月, 苹果系统与安卓系统为主的智能手机操作系统占据了智能手机市场份额的 96.4%^[1]。相对于传统的桌面端, 人们可以随时随地通过移动终端设备取得服务, 各种与移动终端相结合的新兴商业模式与日俱增, 依据用户的地理位置为用户提供基于用户地理位置的服务 (LBS) 便是一个典型应用, 与此同时用户所携带的手持设备由于服务的需求被动的生成了一系列包含网络服务基站信息 (如基站 ID, 基站坐标)、时间信息等内容, 并且针对这些数据的分析与挖掘为理解用户多方面的行为模式提供了至关重要的途径。

移动数据与传统桌面端用户日志主要差异在于提供了用户的地理位置变化, 许多研究将用户的地理位置按时间排序, 采用序列挖掘的算法挖掘用户移动序列的频繁模式^[2-6], 并且 Yavas 使用所构建的频繁模式对用户将来的地理位置进行了预测, 应用在基站资源的预分配^[4]。同时, 探讨如何从比较长期的地理位置日志中挖掘周期性的行为也是研究热点之一^[7-10]。从移动用户数据中构建用户特征属性可作为识别用户重要依据^[11-12], 其中 Zhu 等人除了考虑地理位置频繁序列之外, 通过在手机设备中安装专门的位置记录软件收集用户地理位置数据并提出频繁停留地点、频繁移动路径、有意义的地理位置与交通工具 4 种用户特

收稿日期: 2016-07-05

基金项目: 国家自然科学基金项目“面向电子商务生态平衡的目录导购机制研究”(项目编号: 71373015)的研究成果。

作者简介: 黄文彬 (1977-), 男, 讲师, 研究方向: 数据科学。

征来描述用户, 构建了更有意义的用户属性^[11]。刘瑜等人提出了利用大数据的方法通过对社交网站签到数据以及其他信息进行了人类移动模式的研究的流程^[13]。除了地理位置相关的研究之外, Tseng 等人把用户的地理位置序列与对应所使用的服务进行了组合, 通过挖掘两者共同出现的频繁模式来预测用户将来的行为^[14-15]。此外, Zhu 等人把所有的手机所反映的行为考虑在内, 通过手机的原始传感器数据与自然语言的词相对应, 采用自然语言处理中的方法构建语言模型反映用户行为, 并应用于设备被盗检测、移动应用安全等方面^[16]。张慷等人则提出了一个整合客户资料、语言行为、移动互联网行为等分析用户画像的总体架构, 但在用户属性上缺少细化挖掘分析^[17]。由于用户数据多涉及隐私问题, 现有的研究多采用模拟数据或是通过在手机端安装软件收集少量用户 GPS 或基站数据改进用户频繁序列挖掘算法, 并且主要着重刻画某些方面的移动用户行为, 在国内外的研究中, 通过海量通信网络服务基站信息数据建构移动用户画像相对都比较少。

本文采用北京某电信企业记录 3 万移动用户 1 个月的移动数据, 其中包含用户的通话时间和时长、通话基站坐标、网络请求时间、网络请求基站坐标等记录。笔者利用该数据从频繁活动规律、周期性活动规律、移动速度 3 个方面建构移动用户画像的行为属性, 实证研究结果显示这 3 个方面能够反映真实用户整体的行为习惯, 这对移动用户行为相关研究及移动服务提供商的用户研究提供了重要

的参考。

1 数据集的处理

本研究的数据为北京某电信企业服务器 1 个月的 3 万移动用户数据, 时间为 2013 年 12 月。数据包含用户的基本信息记录、用户通话日志以及用户网络请求日志, 具体的数据项目如表 1 所示。

表 1 所获取的用户数据的基本描述

数据集名称	数据集描述
用户通话日志	拨打时间、基站坐标
用户网络请求日志	请求发起的时间、请求服务类型、基站坐标

该数据是典型的基站记录, 只在用户手机与基站通讯时产生的日志, 其中用户通话日志共有 6 百万条记录与网络请求日志共有 4 亿条记录。相比之下通话日志记录更稀疏, 本研究重点采用网络通讯数据分析用户移动行为。本研究的网络日志数据预处理的流程见图 1。首先将网络数据拆分成基站坐标数据及网络请求数据; 由于用户访问页面的时候通常伴随着图片、脚本、及其他等诸多请求, 笔者再利用域名、位置等信息对网络请求数据进行压缩, 以提炼出用户真实网络行为; 同时, 利用国内企业提供的地图 API 获取每一个基站坐标的具体地址、周围的建筑和场所的名称标签、以及商圈资讯等语义化信息。

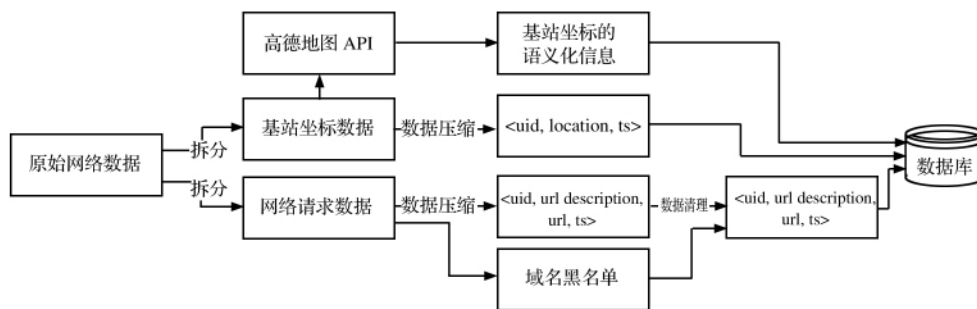


图 1 数据预处理流程

通过数据构建用户画像的前提是该用户有足够的通话及网络请求日志记录, 这批数据里有些用户 1 个月中仅仅发起了数十条网络请求, 因此必须筛选清除网络请求记录稀疏的用户。本研究将 1 天按小时分割成 24 个时段, 若用户在某时段中产生了数据请求, 则标记该用户在该时段为活跃时段, 通过筛选活跃时段至少为 8 的天数大于 20 天的用户共计 1.6 万用户作为主要的研究对象。

2 移动用户行为画像的构建

移动属性的画像构建流程见图 2。笔者从数据中所获取的用户行为坐标并根据用户停留时间的长短分为停留点

与移动点。通过停留点的坐标集合建构用户的周期性活动规律和频繁活动规律研究用户的生活习惯。其中周期性活动规律是指该用户在 1 天中各时段停留地理位置的概率分布, 结合基站坐标的语义化信息推断出用户的住家、工作场所、偏好地点以及作息规律。频繁活动规律是指该用户的停留地点之间存在明显的顺序关系, 结合基站坐标的语义化信息可推断出用户的活动目的与服务需求。通过移动点的坐标集合建构用户在停留点间的移动路径, 通过该两点时间区间的交通路径、时间和速度展现该用户的交通方式以及选择偏好。

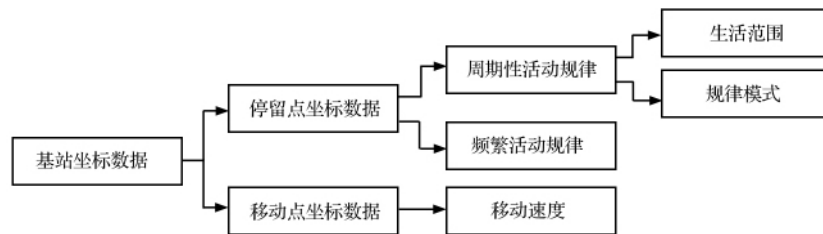


图2 地理位置属性构建流程图

2.1 停留点与移动点

本研究采用的数据是电信运营商服务器记录的用户基站服务日志，并非通过 GPS 记录用户精确地理坐标，由于基站通讯的数据限制，无法细粒度的描述用户的精确坐标与对应时间，该数据只有当用户与基站通讯时的信息记录，笔者利用时间区间区分用户停留坐标为停留点或移动点的定义。假设用户 u 共有 n 个基站请求记录，利用时间 ts_j 依序标记该用户的坐标为 loc_j ，则该用户位置与时间序列可表示成：

$$Traj^u = \{[ts_j, loc_j], \forall j \in [0, n-1]\} \quad (1)$$

其中，将第 $k \in (0, n-m)$ 段长度为 $m \in N$ 的用户子连续位置序列表示为：

$$Traj_{k,m}^u = \{[ts_j, loc_j], \forall j \in [k, k+m-1]\} \quad (2)$$

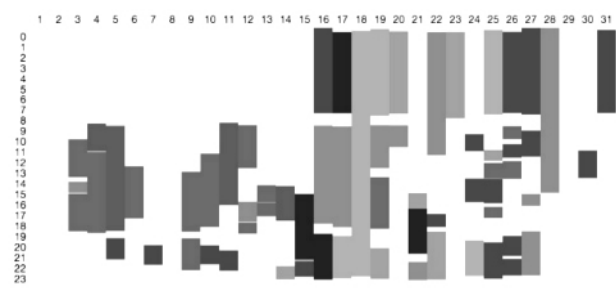
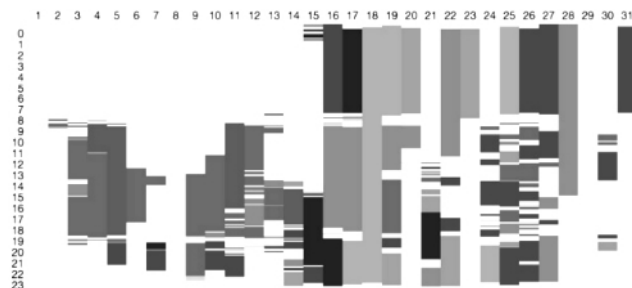


图3 左图为某用户31天的地理位置展示图，右图为记录该用户的暂留点位置展示图

笔者提取数据中某用户31天位置记录并采用上述的定义进行计算，图3左图显示了该用户整月的停留点和移动点，其中同颜色是指该时间段属于连续位置子序列，白色为信息缺失的部份。图3右图显示将移动点从图上移除后的结果，可见地理位置快速切换的移动数据已经被剔除，且综合观察仍可得到长时间驻留地点的信息。

2.2 周期性活动规律计算

为了构建用户在1天中各时段停留地理位置的概率分布，首先笔者将该用户位置时间序列 $Traj^u$ 转变成以时间区间表示用户活动序列 $BTraj^u$ ，通过用户的停留点数据构建用户在不同的时间区段停留的概率值，以进一步清理在某时间区段中的不可能长时间规律存在的停留点，分析用户的周期性活动规律。假设用户记录的数据总时长个数 L 下有 n 个停留区间，按照停留点坐标 loc_i 、停留区间的起始时间 ST_i 、停留区间的结束时间 ET_i 来表示用户活动序列 $BTraj^u$ 为：

— 56 —

如果用户子连续位置序列 $Traj_{k,m}^u$ 满足：

$$loc_i = loc_j, \forall j, i \in [k, k+m-1], \text{ when } ts_j - ts_{j+1} \leq \Delta t \text{ and } ts_{m-1} - ts_0 \geq \Delta T \quad (3)$$

则这该连续子序列 $Traj_{k,m}^u$ 所记录的移动坐标即为停留点，而 ΔT 即是这个暂留点当时连续的停留时间。如果用户子连续位置序列 $Traj_{k,m}^u$ 满足：

$$loc_i \neq loc_j, \forall j, i \in [k, k+m-1], \text{ when } ts_j - ts_{j+1} \geq \Delta t \quad (4)$$

则该 $Traj_{k,m}^u$ 是一个用户移动路径，所记录的移动坐标即为移动点，而 ΔT 是移动路径的时间区间。本文针对该批数据分析处理设置 $\Delta t = 5$ 分钟和 $\Delta T = 30$ 分钟，并按照以上所定义的概念构建用户的停留坐标、停留时间区间、用户的移动路径以及移动时间区间。

$$BTraj^u = \{[loc_i, ST_i, ET_i], \forall i \in [0, n-1]\} \quad (5)$$

则给定任意时段 t 中用户是否停留在任意 loc 可表示为：

$$g_{loc,t} = \begin{cases} 1, loc = loc_i \wedge ST_i \leq t \leq ET_i \\ 0, loc \neq loc_i \wedge ST_i \leq t \leq ET_i \end{cases} \quad (6)$$

若以周期 T 的时间观察用户活动规律，且周期 T 包含 d 个离散的时间区段 $t_j \in d$ ，将则认为用户在时段 t_j 处于位置 loc_i 的概率为：

$$p(loc_i, t_j) = \frac{\sum_{k=0}^{L-1} g_{loc_i, t_j + dk}}{L} \quad (7)$$

其后将获取一个 $n \times d$ 的概率矩阵，其中 n 为该用户的暂存点坐标个数， d 为周期 T 内的时段个数。本文设置 $T = 24$ 小时、 $d = 48$ ，即以1天为周期、30分钟为时间区段构建概率矩阵，并针对每个时间区段建构最有可能长时间规律驻留的停留点，其方法如下：

输入：用户的地理位置概率矩阵。

输出：该用户各个时段最有可能长时间规律的停留点。

(1) 构建用户 $t_{j \in d}$ 个时段处于各个停留点 $loc_{i \in n}$ 的概率和 $\sum p(loc_i, t_j)$ 。

(2) 计算各个时段可能长时间停留坐标的阈值。

a. 预先设置用户在某时段处于该位置的概率值 τ 。

b. 针对每个停留点计算，若 $\sum p(loc_i, t_j) < \tau$ ，则不认为该时段存在长时间规律的暂留点，此时设置 $\Gamma_{ij} = -1$ ；否则设置该时段的长时间规律停留点的阈值 $\Gamma_{ij} = \frac{1}{2} \sum p(loc_i, t_j)$ 。

(3) 对各个时段采用阈值列表 Γ_{ij} 获取长时间规律的停留点坐标。

最后通过设置概率阈值获取用户各个时段长时间活动规律停留点的序列，笔者将在本文的 3.1 节说明本方法的实证结果与示例讨论。

2.3 频繁活动规律计算

为了分析用户的频繁活动规律，笔者首先将用户停留点序列 S_Traj^u 转变成用户连续的停留点序列：

$$CS_Traj^u = \{[loc_i], \forall i \in [0, S-1]\} \quad (8)$$

S 是用户所有停留点的个数，由于笔者是以天为单为进行用户序列分析，因此 CS_Traj^u 等于每天的子序列 $D_CS_Traj_j^u$ 的组合。

$$CS_Traj^u = \{[D_CS_Traj_j^u], \forall j \in [1, d]\} \quad (9)$$

那么，我们可以将用户频繁活动规律的挖掘转化为频繁序列挖掘的问题。我们采用 GSP (Generalized Sequential Patterns)^[18] 算法对每一个用户的停留坐标序列集挖掘其中的频繁停留坐标序列，算法如下。

输入：用户每天的子序列 $D_CS_Traj_j^u, j \in d$

输出：频繁停留坐标序列

(1) 设置 $k=1$ 。

(2) 获取长度为 k 的子序列以及子序列的支持度，并删除支持度小于 \minSup 的子序列。

(3) 利用长度为 k 的频繁序列生成长度为 $k+1$ 的序列。

(4) $k=k+1$ ，跳转到 2，直到找不到频繁序列或者不再有新的子序列。

其中，支持度是指在序列集 $D_CS_Traj_j^u, j \in d$ 中包含长度为 k 的子序列的频次，并且最小支持度是 \minSup 实验参数。在获取频繁序列之后，在从中选择最长频繁序列作为序列挖掘的结果，笔者将在本文的 3.2 节说明该方法的实证结果与示例讨论。

2.4 移动速度计算

笔者引入了熵^[11]的概念建构用户当时的移动熵用以刻画用户在各个时刻的移动速度，利用用户移动点序列 M_Traj^u

进行用户移动熵的计算，并根据时间点 t 在时间范围 ΔT 内用户出现在不同地点的概率刻画用户的移动速度，其方法如下：

$$H(t, \Delta T) = - \sum_{i=1}^N P_{t, \Delta T}(C_i) \log_2 P_{t, \Delta T}(C_i) \quad (10)$$

其中，

$$P_{t, \Delta T}(C_i) = \frac{1}{\Delta T} \int_{t-\frac{\Delta T}{2}}^{t+\frac{\Delta T}{2}} I(C_i, r_\tau) d\tau \quad (11)$$

是用户在时间 $\left[t-\frac{\Delta T}{2}, t+\frac{\Delta T}{2}\right]$ 范围内位于地点 $C_i \in M_Traj^u$ 的概率其中 r_τ 为用户在时间 τ 所处的基站，并

$$I(C_i, r_\tau) = \begin{cases} 1, & C_i = r_\tau \\ 0, & \text{others} \end{cases} \quad (12)$$

所计算的移动熵说明用户在时间 ΔT 所变化的基站坐标越多，相应的其移动速度就越大。如果移动熵为 0 则说明用户没有进行基站的切换。用户的停留时间区间移动熵为 0，用户的移动时间区间通过取等时间跨度的时间节点可以获取多个时间点的移动坐标，这些坐标可以构建用户的移动速度图。

3 移动用户画像示例

本章节笔者从周期性活动规律、频繁活动规律、用户移动速度 3 个方面通过 4 个用户的真实数据的实证结果与示例说明用户的移动行为属性。

3.1 周期性活动规律示例

利用 2.2 节提出的算法建构每个用户的坐标概率矩阵，用户 1 的坐标概率矩阵见图 4，横坐标表示 1 天 48 个时间段，纵坐标表示该用户的所有基站坐标。为了包含数据隐私，基站坐标末两码已模糊化处理，图中方块的颜色由浅到深表示该时段在对应坐标的概率逐渐增大。由于基站数据仅能记录用户使用基站的记录，而用户在有 Wi-Fi 无线网络的环境使用的网络请求是没有记录的，因此每个用户都会有数据缺失，即用户在各个时段的概率之和小于 1。图 4 显示该用户在时段 1~17 (0 点至 8 点 30 分) 使用坐标 $\langle 116.154xx, 40.052xx \rangle$ 基站的概率大，在时段 19~44 (9 点钟至 22 点钟) 使用坐标 $\langle 116.348xx, 39.921xx \rangle$ 基站的概率大，由此推断该用户住家是在基站 $\langle 116.154xx, 40.052xx \rangle$ 的覆盖范围内，利用语义化信息查知该基站位于北京市海淀区温泉镇，而用户办公场所是在基站 $\langle 116.348xx, 39.921xx \rangle$ 的覆盖范围内，并位于北京市西城区万通新世界大厦，同时也可以判断该用户的工作时长大约都在时段 19~44 之间，且可能因公务需求偶尔住宿在其他地方。利用阈值 Γ_{ij} 删除概率获取该用户的时间规律，通过与基站的语义化信息进行整合，可以清晰的展现用户的活动规律表，其结果如表 2 所示。

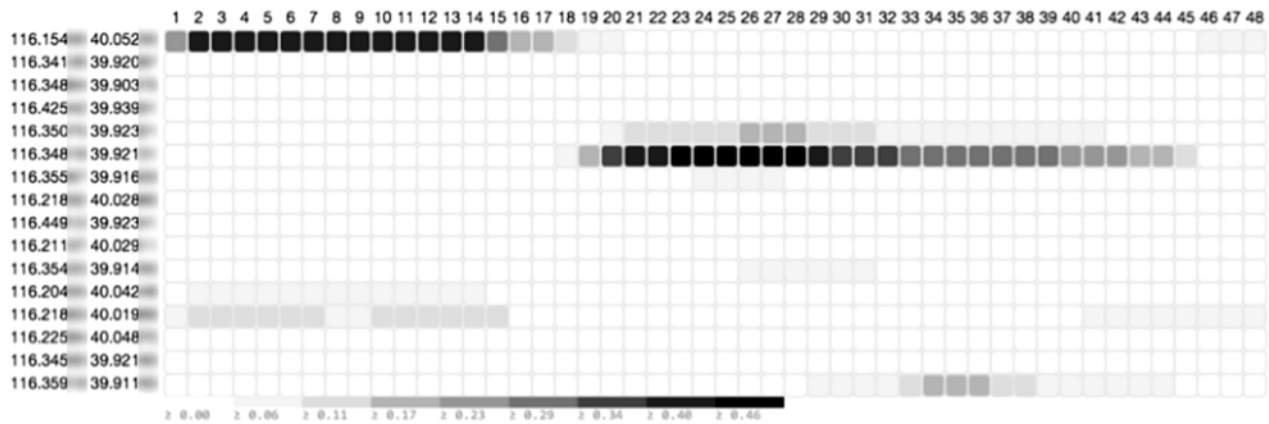


图 4 用户 1 的概率矩阵

表 2 用户 1 的周期性活动规律表

时 间	基站坐标	语 义
0 点到 8 点 30 分	116.154xx 40.052xx	住家（海淀区）
8 点半到 9 点 30 分	数个坐标值	移动中
9 点半到 21 点 30 分	116.348xx 39.921xx	办公场所（西城区）
22 点到 23 点	数个坐标值	移动中
23 点以后	116.154xx 40.052xx	住家（海淀区）

图 4 表明用户 1 生活规律性很强，往返于家与办公地，而停留在其他地理区域的概率非常小，那么虽然用户的办公地与住家的地理位置相差较远，我们依然认为该用户的生活范围是比较小的。图 5 则展示用户 2 的基站坐标概率矩阵，在时段 1~14（0 点到 7 点）位于基站〈116.424xx，39.892xx〉的概率比较大，而在剩余时段该用户没有明显的停留坐标点，可推测该用户没有固定的工作地点，并且该用户停留的地理位置要比用户 1 多且物理距离较远，说明用户 2 的生活范围较大。

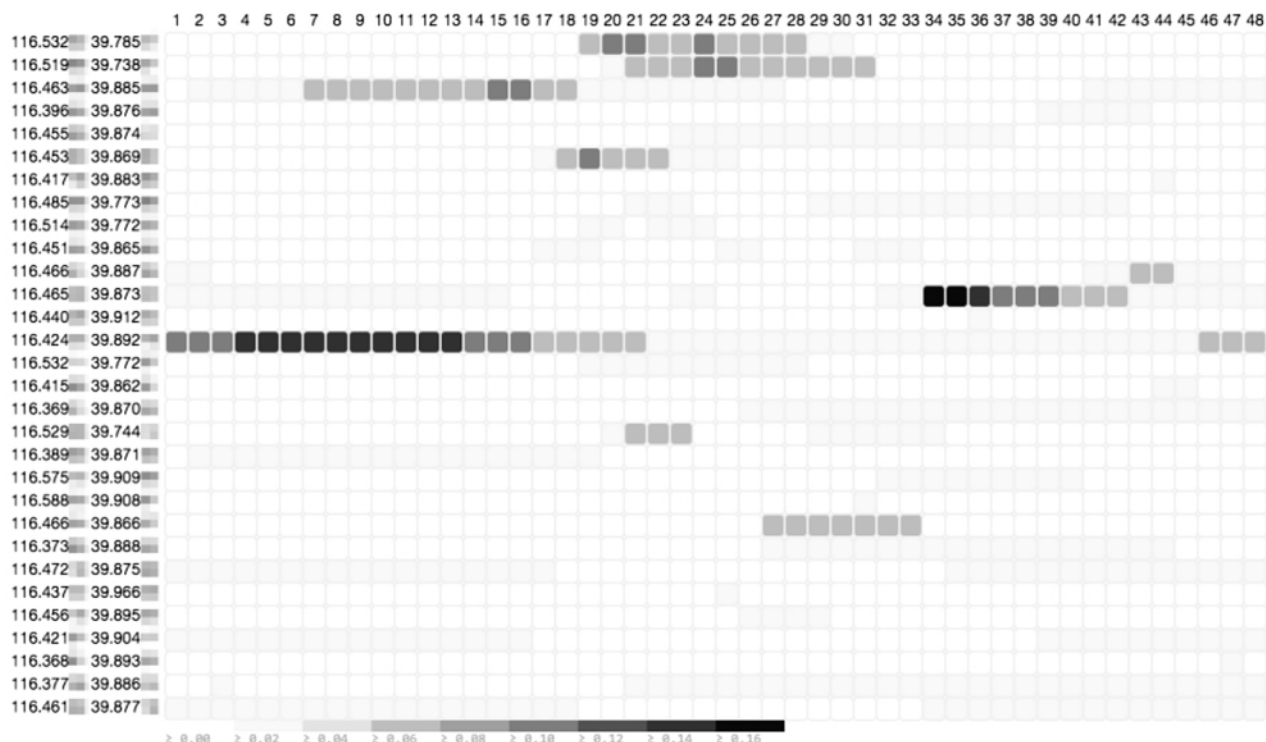


图 5 用户 2 的概率矩阵

3.2 频繁活动规律示例

用户 3 的频繁停留坐标的序列见表 3，该用户频繁从基站〈116.381xx，39.973xx〉到基站〈116.301xx，

39.969xx〉，或者经由〈116.410xx，39.906xx〉再到〈116.412xx，39.968xx〉。结合表 4 基站的语义化信息可以看出用户频繁从裕民路出发，经过东单再到北京化工大学

或直接到中国人民大学活动。

表 3 用户 3 的频繁活动序列

标号	频 繁 序 列
1	116. 381xx 39. 973xx → 116. 410xx 39. 906xx → 116. 412xx 39. 968xx
2	116. 381xx 39. 973xx → 116. 301xx 39. 969xx

表 4 用户 3 频繁停留坐标序列中坐标的语义化信息

坐标	北京市地址	标 签
116. 381xx 39. 973xx	朝阳区安贞街道裕民路 12—7 号	学校, 图书馆, 培训机 构, 餐饮
116. 410xx 39. 906xx	东城区东华门街道东长 安街甲 2 号	地铁站, 银行
116. 412xx 39. 968xx	朝阳区和平街街道北京 化工大学	科教文化, 宾馆酒店, 住宅
116. 301xx 39. 969xx	海淀区海淀街道中国人 民大学	学校, 住宅, 中餐厅, 公园广场

用户 3 的周期性活动规律 (见图 6) 表明, 该用户晚

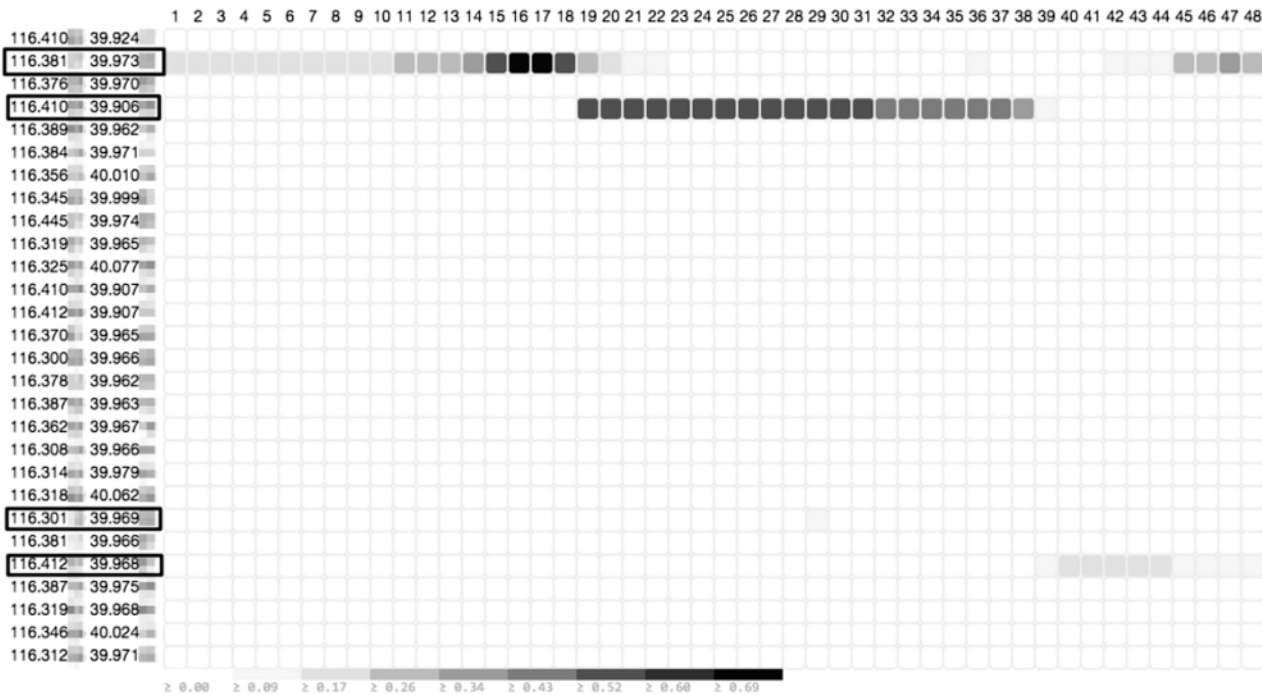


图 6 用户 3 停留基站的概率矩阵

表 5 用户 4 移动路径中各个移动点的移动熵

时 间	移动熵	时 间	移动熵
10 : 23	0. 995	14 : 30	0. 891
11 : 45	0. 921	14 : 50	1. 466
13 : 02	1. 202	15 : 25	1. 000
13 : 18	2. 139	16 : 03	0. 999
13 : 23	2. 164	16 : 22	0. 000
13 : 26	2. 487		

• 10 点 23 分该用户的移动熵开始发生变化, 并在 30 分钟内移动路径上有两个邻近坐标的变换且移动熵小于

间和晨间停留在基站 <116. 381xx 39. 973xx> 的概率非常大, 而在上午 9 点半至 19 点间停留在基站 <116. 410xx 39. 906xx> 的概率大。结合地理语义信息, 推断用户 3 住在裕民东里附近并在上午 9 点半后东单附近工作, 再加上该用户标号 1 的频繁活动序列说明该用户经常从居住地去工作地后在晚上 8 点去基站 <116. 412xx 39. 968xx> (北京化工大学) 附近, 推断用户 3 习惯在下班后在该区域进行夜间上课学习到深夜才回家。标号 2 的活动序列说明用户 3 从居住地到人民大学区域活动, 该序列属于假日, 推断该用户在周末会频繁去人民大学上课或进行学术活动。

3. 3 移动速度计算示例

图 7 是用户 4 某一天的坐标记录, 横纵坐标分别为时间、基站坐标。提取该用户的移动路径, 使用 3. 4 节的算法计算移动熵, 如表 5 所示。考虑到城市交通的拥堵以及在各个时段交通的速度是有明显的区别的, 笔者在这只区分用户步行还是使用其他交通工具, 综合观察图 7 和表 5, 做如下描述:

1. 0, 推断该用户在基站范围内或附近慢速度的行走移动或进行购物等活动;
 - 13 点 02 分该用户开始以较高的移动熵进行运动, 推断是该用户使用交通工具进行活动, 结合坐标语义可以看出用户从北京市朝阳区安贞街道经过北太平庄到达海淀黄庄附近, 并随后 1 个小时内用户正在中关村商圈范围内活动;
 - 随后 14 点 50 到 15 点 25 之间, 用户再次使用交通工具低速移动, 从中关村商圈回到北京市朝阳区安贞街道附近区域并之后停止活动。

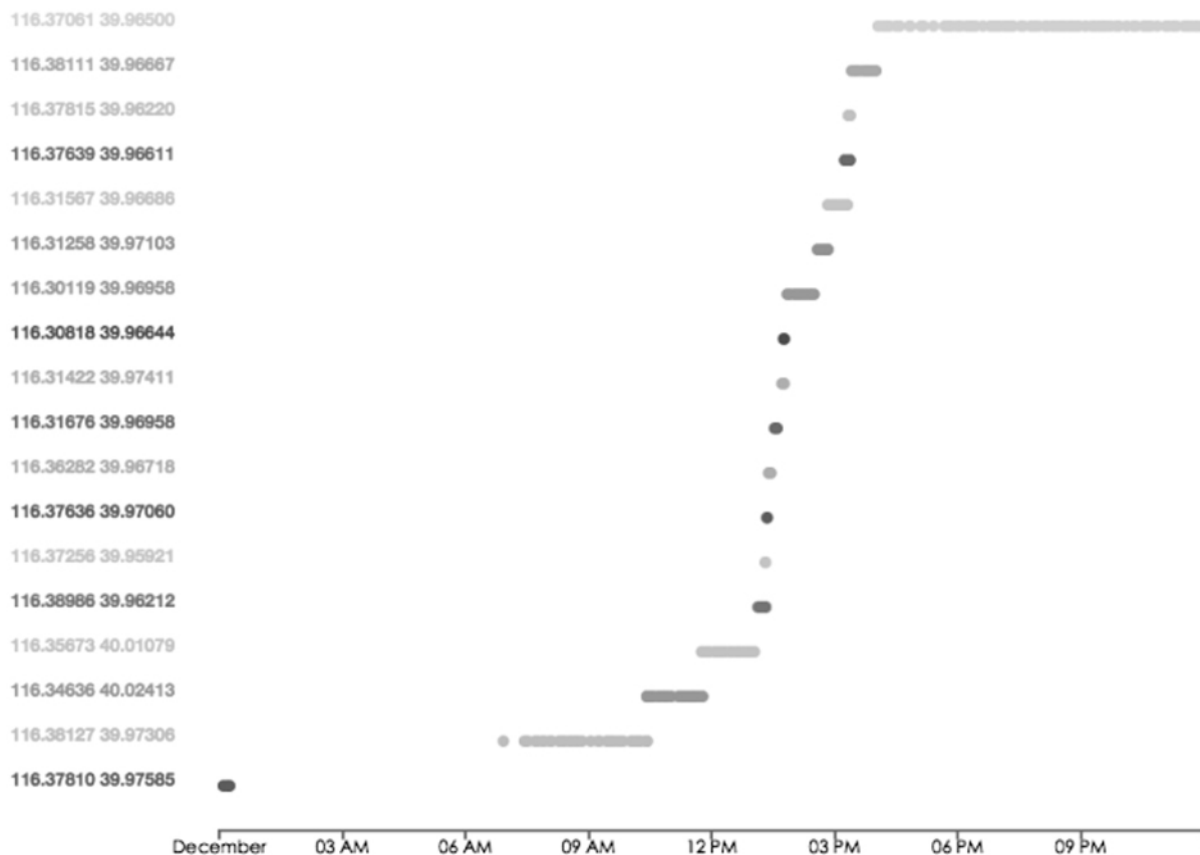


图 7 左图是用户 4 按照时间的移动点图

综合上述的周期性活动规律、频繁活动规律、用户移动速度的分析方法并添加坐标语义信息后所获得的用户行为生活习惯和活动规律容易暴露个人隐私或识别出用户个人身份，同时也说明对通信数据进行挖掘分析对用户具有信息安全的疑虑。

4 结 论

本文利用移动网络基站数据从周期性活动规律、频繁活动规律以及移动速度三方面构建了移动用户的用户画像，结合所提供的可视化视图可以清晰地展现出用户地理位置相关的行为，所获取的属性相互印证和补充，展现了 1 个用户每天的生活规律和移动情况。由于本文所获取的用户数据时间跨度仅 1 个月，笔者只以 1 天为周期的活动规律进行挖掘，但这分析结果已经可以推测出用户的工作方式、住家地点、工作地点、生活规律、出行方式与模式。由于该用户数据存在部分空缺，对用户的微观行为刻画比较有限，未来笔者会将这移动属性与网络访问信息相结合获取更准确更细粒度的用户移动行为和偏好。若能将所有用户的移动用户画像并且支持用户属性检索，便可以快速检索出具有相似特定属性的用户群，便可以针对需求提供用户相应的个性化服务。

参 考 文 献

[1] Idc. Worldwide Quarterly Mobile Phone Tracker. 2014.

[2] Lee S C, Paik J, Ok J, et al. Efficient mining of user behaviors by temporal mobile access patterns [J]. Int'l J. Computer Science Security, 2007, 7 (2): 285—291.

[3] Chen T S, Chou Y S, Chen T C. Mining user movement behavior patterns in a mobile service environment [J]. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 2012, 42 (1): 87—101.

[4] Yavaş G, Katsaros D, Ulusoy Ö, et al. A data mining approach for location prediction in mobile environments [J]. Data & Knowledge Engineering, 2005, 54 (2): 121—146.

[5] 曾灿灿. 基于数据挖掘的移动行为预测 [D]. 武汉: 华中科技大学, 2011.

[6] 王晓明. 基于移动数据的轨迹模式挖掘算法的研究与应用 [D]. 沈阳: 东北大学, 2013.

[7] Baratchi M, Meratnia N, Havinga P J M. Recognition of periodic behavioral patterns from streaming mobility data [M]. Mobile and Ubiquitous Systems: Computing, Networking, and Services. Springer International Publishing, 2014: 102—115.

[8] Li Z, Ding B, Han J, et al. Mining periodic behaviors for moving objects [C] // Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2010: 1099—1108.

[9] Zhang M, Kao B, Cheung D W, et al. Mining periodic patterns with gap requirement from sequences [J]. ACM Trans-

- actions on Knowledge Discovery from Data (TKDD), 2007, 1 (2): 7.
- [10] Ji Y, Zhang C, Zuo Z, et al. Mining user daily behavior based on location history [C] // Communication Technology (ICCT), 2012 IEEE 14th International Conference on. IEEE, 2012: 881-886.
- [11] Zhu Y, Zhang Y, Shang W, et al. Trajectory enabled service support platform for mobile users' behavior pattern mining [C]. Mobile and Ubiquitous Systems: Networking & Services, MobiQuitous, 2009. MobiQuitous' 09. 6th Annual International. IEEE, 2009: 1-10.
- [12] 薛冉. 基于地理位置服务的用户移动属性的构建和比较 [D]. 济南: 山东大学, 2013.
- [13] 刘瑜, 康朝贵, 王法辉. 大数据驱动的人类移动模式和模型研究 [J]. 武汉大学学报: 信息科学版, 2014, (6): 8.
- [14] Lu E H C, Tseng V S, Yu P S. Mining cluster-based temporal mobile sequential patterns in location-based service environments [J]. IEEE Transactions on knowledge and data engi-

neering, 2011, 23 (6): 914-927.

- [15] Tseng V S M, Lin K W C. Mining sequential mobile access patterns efficiently in mobile web systems [C] // Advanced Information Networking and Applications, 2005. AINA 2005. 19th International Conference on. IEEE, 2005, (2): 762-767.
- [16] Zhu J, Hu H, Hu S, et al. Mobile behaviorometrics: Models and applications [C] // Communications in China (ICCC), 2013 IEEE/CIC International Conference on. IEEE, 2013: 117-123.
- [17] 张慷. 手机用户画像在大数据平台的实现方案 [J]. 信息通信, 2014, (2): 266-267.
- [18] Srikant R, Agrawal R. Mining sequential patterns: Generalizations and performance improvements [M]. Springer Berlin Heidelberg, 1996.

(本文责任编辑: 马 卓)

(上接第 53 页)

用。因此, 本文在参考国内外相关文献的基础上, 构建了一套可用于了解用户对分类机制设置偏好的指标体系, 并通过发放调查问卷的方式, 了解当前用户使用微信订阅服务主要情况, 以及用户对信息分类机制的设置偏好, 继而根据对调研结果的分析, 提出了建立微信订阅服务中信息过载防控机制的建议, 以期对解决微信订阅服务中信息过载问题能够起到一定的借鉴意义。但是, 本次调查对象的年龄主要集中在 20~29 岁、学历主要在本科以上, 对于其他年龄段、受教育阶段的用户所知甚少, 还有待于后续的进一步研究, 此外, 本文只是从理论方面提出了建立微信订阅服务中信息过载防控机制的建议, 并未进行相关的实证研究, 这也需要在以后进一步深化研究。

参 考 文 献

- [1] Toffler A. Future shock [M]. New York: Bantam Books, 1970.
- [2] Wurman, Richard Saul: Information Anxiety [M]. New York, Doubleday, 1989.
- [3] Klapp, Orrin E: Overload and Boredom: Essays on the Quality of Life in the Information Society [M]. New York: Greenwood, 1986.
- [4] Casey CJ. Coping with information overload: the need for empirical research [J]. Cost and management, 1992, (4): 31-37.
- [5] 于文莲. 网络环境下的信息过载研究 [J]. 农业图书情报学刊, 2008, 20 (11): 51-54.
- [6] 蔺丰奇, 刘益. 网络化信息环境中信息过载问题研究综述 [J]. 情报资料工作, 2007, (3): 36-48.
- [7] 王娜, 任婷. 移动社交网站中的信息过载与个性化推荐机制

研究 [J]. 情报杂志, 2015, (8): 190-194, 176.

- [8] 马捷, 闫明, 靖继鹏. 信息生态视角下电子商务网站信息分类优化研究 [J]. 情报科学, 2012, (3): 360-363, 407.
- [9] 高翔. 基于混合特征的微博信息分类方法研究 [D]. 北京: 北京工业大学, 2013.
- [10] 吴砥, 赵姝, 杨晓露, 等. 学习资源的分类机制与分类标准的研究 [J]. 中国远程教育, 2009, (1): 56-60, 79.
- [11] Thomas Vander Wal. Folksonomy Explanations [EB/OL]. <http://www.vanderwal.net/random/entrysel.php?blog=1622>, 2016-03-08.
- [12] A. Mathes. Folksonomies: Cooperative Classification and Communication through Shared Metadata [EB/OL]. <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>, 2016-03-08.
- [13] Shirky, C (2005) Ontology is overrated: Categories, links and tags [EB/OL]. <http://www.shirky.com/writings/ontology-overrated.html>, 2016-03-08.
- [14] 陈树年. 网络信息分类法研究 [J]. 现代图书情报技术, 2002, (3): 54-57, 80.
- [15] 吴胜, 张智光, 张长江. 网站信息分类体系质量评价指标体系的研究 [J]. 图书情报工作, 2009, (6): 58-61, 82.
- [16] Eppler M J, Mengis J. A frame work for information overload research in organization [EB/OL]. http://www.bu.lunich.it/cerca/bu1/publicazioni/com/pdf/w_pcao30.pdf, 2016-03-08.
- [17] Wurman, Richard Saul Bradford, Petered. Information Architects [J]. Library Journal, 1996, 121 (16): 120.
- [18] 邵国川, 李晓霞. 网络不良信息监督管理对策研究 [J]. 科技情报开发与经济, 2011, (4): 142-145.

(本文责任编辑: 马 卓)