



基于位置信息的用户行为轨迹分析与应用综述^{*}

陈 康¹, 黄晓宇^{2,3}, 王爱宝⁴, 陶彩霞¹, 关迎晖¹, 李 磊³

(1. 中国电信股份有限公司广东研究院 广州 510630;

2. 华南理工大学经济与贸易学院 广州 510006;

3. 中山大学软件研究所 广州 510275; 4. 中国电信集团公司 北京 100032)

摘 要:近年来,随着空间数据采集技术的发展,基于位置信息的用户行为轨迹分析及其应用的研究引起了广泛关注,并已展现了良好的商业前景。根据应用的领域,对这一问题的研究主要可以分为智能交通应用和用户行为分析应用两种类型。本文分别对这两类应用的研究现状进行了较为全面的总结,对每类应用,都概括了在该领域中研究的典型问题和代表性结果。最后,讨论了在用户行为轨迹分析研究中的主要技术特点,并对未来的研究工作进行了展望。

关键词:位置服务;行为轨迹分析;智能交通;社交网络;大数据

doi: 10.3969/j.issn.1000-0801.2013.04.023

Analysis and Application Review of User Behavior Trajectory Based on the Location Information

Chen Kang¹, Huang Xiaoyu^{2,3}, Wang Aibao⁴, Tao Caixia¹, Guan Yinghui¹, Li Lei³

(1. Guangdong Research Institute of China Telecom Co., Ltd., Guangzhou 510630, China;

2. School of Economics and Commerce, South China University of Technology, Guangzhou 510006, China;

3. Software Institute, SUN Yat-Sen University, Guangzhou 510275, China;

4. China Telecom Corporation, Beijing 100032, China)

Abstract: With the rapid development of spatial data collection technology in recent years, many efforts have been devoted to the research on trajectory based user behavior analysis and its applications, due to their academic interests; they also show promising business opportunities. From the application aspect, in principle, current issues in trajectory based user behavior analysis can be classified into two major categories: the intelligent transportation applications and the user behavior based applications. The state-of-the-art progress of these categories were reviewed. For each category, its main focus and related achievements were discussed. What's more, the main technical characteristics of the means used were also summarized in current studies.

Key words: location based service, behavior trajectory analysis, intelligent transportation, social network, big data

^{*} 国家高技术研究发展计划(“863”计划)“业务受控的精密定位信息移动广播系统技术”基金资助项目(No.2012AA12A203)

1 前言

近年来,随着以 GPS 导航仪和智能手机为代表的智能终端的普及与应用,人们已经能够以相对低廉的代价获得大量的用户实时位置数据,如在 GPS 导航系统的支持下,可以实时获得汽车驾驶员当前所在的经、纬度位置信息和行驶方向信息;对于随身携带移动电话的用户,能以基站定位的方式,估计出该用户所在的大概区域。特别地,对于给定的用户,将其在一组连续时间点上的位置“串联”起来后,就形成了他在这个时间段内的行为轨迹数据。

在大量用户位置和行为轨迹数据的背后,隐含了丰富的空间结构信息和用户行为规律信息,通过对这些信息进行深入的挖掘和利用,不仅有可能发现个体用户的日常行为规律和群体用户的共性行为特征,甚至还有可能掌握其社交关系信息,这对智能交通、广告推荐等应用具有非常重要的意义。早在 2003 年,Rao 与 Minakakis 预测,如下 4 类基于用户位置信息的服务应用蕴藏着巨大的商机^[1]:

- 用户空间定位及驾驶的路径诱导服务;
- 基于用户位置分析的精准广告投放服务;
- 基于用户行为的市场细分及应用服务;
- 面向企业的商业合作应用服务。

目前,这些预测大多已成为现实,如美国的 Inrix 公司采集道路上的汽车行为轨迹信息,通过分析、挖掘为在途驾驶员提供交通信息发布和路径诱导服务,其产品用户已经覆盖了北美和欧洲的 20 多个国家;腾讯公司开发了带位置服务功能的即时通信软件——微信,在 14 个月内即

积累了超过 2 亿名注册用户。这些应用的成功吸引了广泛的关注,越来越多的研究人员投入用户行为轨迹分析的研究领域中。

对于用户行为轨迹的分析,一般可以总结为“数据采集—位置匹配—分析应用”3 个步骤,具体技术架构如图 1 所示。

在图 1 中,用于轨迹分析的数据主要来源于车载的 GPS 定位数据和以智能手机为代表的基站定位数据(部分智能手机也支持 GPS 定位,也有可能提供 GPS 定位数据),一般这些数据都可以使用四元组<数据源 ID,时戳,经度,纬度>表示;在使用这些数据之前,需要把它们匹配到地图上^[2]以关联某些兴趣点(point of interest, POI),但由于地图匹配所需的电子地图通常难以获得,因此研究人员对这一步骤做了简化处理,只是简单地栅格化,将每个原始的数据点映射到栅格中,对用户行为轨迹的分析和应用在完成上述位置匹配过程之后的数据上进行^[3-5]。

虽然对用户日常行为轨迹的分析吸引了众多的研究兴趣,然而总结起来,相关的工作主要分为两种类型:其一是传统的智能交通服务应用,研究所需的数据主要来源于由 GPS 终端和智能手机采集获得的用户空间位置信息;其二是对用户的行为模式识别与社交关系的发现,这一工作的开展除了需要用户的空间位置信息之外,还需要其他的数据信息,如用户的网络浏览信息、用户在社交网站上的活动记录信息等。

2 智能交通应用

随着位置数据采集设备的普及,近年来,在智能交通

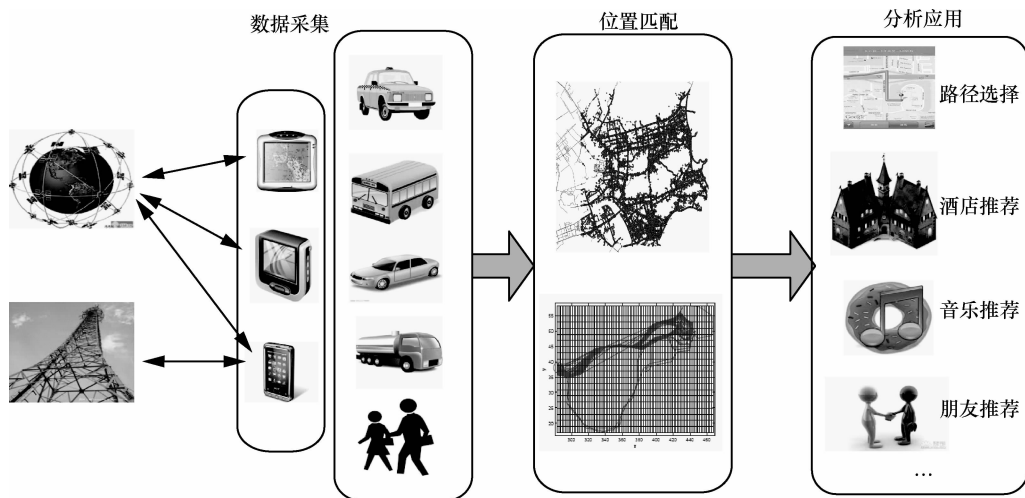


图 1 用户行为轨迹分析技术架构



领域,研究手段也出现了革命性的变化:在早期的研究中,由于数据采集与处理的成本过高,数据的主要作用在于对已有模型进行检验和校正;在最近10年中,随着采集成本的降低,已经可以用非常低廉的价格获得大量的动态交通数据,以此为基础,相关研究者提出了以数据为驱动的智能交通系统^[6](data-driven intelligent transportation system, DITS),其核心在于通过数据挖掘、机器学习等手段,从数据中提取交通系统的运行模式,从而研究新的交通应用。

2.1 电子地图

在智能交通的各种研究中,电子地图是最为底层的数据资源,为地图匹配、用户兴趣点分析以及路径诱导等各种应用提供基础性的支撑。然而,传统电子地图的编制代价非常高昂,需要投入大量的人力和物力进行精确的测量与校正。

为解决这一问题,Cao与Krumm^[7]利用了如下事实:用户驾车经过的地方必然有道路的存在。由此提出了一种基于用户出行轨迹的地图编制算法,认为车辆在一条道路上行驶的GPS轨迹由该道路上各点的空间位置与一组相互独立的以0为均值的误差叠加而成,因此,当多条这样的轨迹合并时,得到的结果即该道路的实际位置。另一方面,在电子地图中,车道标识是一个很重要的组成部分,参考文献[8]中研究了一种完全基于用户假设轨迹数据的算法,他们注意到道路上不同车道的通行速度是有区别的,而道路上的通行速度可以视作所有车道速度的综合,因而使用混合高斯分布对同一道路上的通行速度进行描述,进而利用采集到的速度数据对模型进行拟合分析,最后获得的高斯成分个数即该道路上的车道数目。

2.2 路径诱导

对行驶中的车辆进行路径诱导是智能交通研究中的核心问题之一,其目标是驾驶员寻找“最合适”的道路。众所周知,Dijkstra算法能在 $O(n^2)$ 的时间内找到路网中两点间的最短路径,然而对驾驶员而言,除路径的长短外,交通状况、驾驶成本等也是影响其对路径进行选择的重要因素,所以“最短”的路径并不等价于“最好”的路径。

为解决上述问题,需要对所谓的“最合适”的路径给出可操作的定义。当前一个被广泛接受的定义是:选择走一条路的车辆越多,这条路的“合适程度”就应该越高。以这一假设为基础,越来越多的工作转向基于驾驶轨迹的路径

诱导研究。

参考文献[9]中提出了一种有别于传统的最短路径算法的路径诱导策略,驾驶员普遍倾向于选择高等级的道路(等级越高,速度越快),而不同等级的道路又把路网划分成大小不一的封闭区域(一般路网的等级越高,区域内的面积越大,反之亦然),由此设计了一种基于A*搜索的最快路径启发式策略。在出发端,该策略引导用户由低等级路网围成的区域尽快进入通往目标方向的高等级区域;在目标端,则引导用户尽快由高等级区域进入包含目标地址的低等级区域。在该算法的设计中,Gonzalez等人还注意到,路网中存在部分路径,虽然级别较低,但通行的车辆却非常多,因此对道路等级的评定方式也做了修改,即道路的等级不仅与其静态特性相关,还与其上通行车辆的数目和通行速度这些动态特征相关,这些特征的取值可以通过对用户出行轨迹的分析获得。

参考文献[9]中的算法能很好地解决中长途旅行(如在两个城市之间)的路径诱导问题,然而对于城市内的路径选择,其效果则差强人意。主要原因在于:在同一城市内,各道路的通行与拥堵状况在整体上已经渐趋一致,难以给出明晰的等级划分,因而该算法的应用也受到了限制。对此,T-drive^[10]提供了另一种解决方案:以城市内出租车的驾驶轨迹数据为基础进行路径诱导,由于出租车的数量众多,活动范围广泛,而且司机的经验也非常丰富,因此可以认为,对于城市内的任意两点,连接这两点的出租车行驶轨迹接近最佳路线。

除了指导人们在驾车出行过程中选择合适的路线外,对用户出行轨迹的分析还有助于“绿色出行”的实现。这里的“绿色”指鼓励人们尽可能使用公共交通工具,从而减少碳排放量,减轻道路的交通拥堵状况。为减少用户在使用公共交通工具出行过程中的不便,参考文献[11]提出基于出租车的历史行驶轨迹数据的公交线路规划算法,该算法使用出租车的起止位置表示它的每一条行驶轨迹,并对这些轨迹进行层次聚类以生成若干轨迹簇,这些簇的中心轨迹即公交线路的规划结果。与此相类似,参考文献[5]中也研究了根据大规模的出租车驾驶轨迹进行夜间公交线路规划的应用。

2.3 行为模式识别

在城市居民的日常出行中,一条出行的行为轨迹可能包含多种不同的行为模式,如步行、公共汽车接驳以及地铁换乘等,参考文献[12]中提出了对用户出行行为模式识

别的研究,对比了条件随机场、支持向量机、贝叶斯网络与决策树在这一问题上的表现,令人吃惊的是,在这些工具中,决策树取得了最好的结果。然而,对于这一结果,还需结合数据的实际分布因素来考虑:对于大多数用户,在他/她的一次出行行为中,绝大多数时间都是在交通工具上度过的,而步行、骑自行车接驳公共交通工具等行为在总体数据中仅占了非常小的比例,对分类算法而言,这是一个类别不平衡的问题^[13]。在这种情况下,分类器仅需简单地把所有结果都判别为优势类别,便能获得很高的准确率,所以简单地以分类结果的准确率来度量分类器的好坏是不合适的。

Stenneth 等人对参考文献[12]的结果做了改进,当用户使用不同的公交工具时,公交服务设施可以作为识别其行为模式的重要启发信息,因此公交站、地铁线路等的位置分布也可以用于辅助计算^[14],结果与参考文献[12]类似,以决策树为基础的随机森林(random forest)算法获得了最好的表现。

此外,对用户交通行为模式的研究还衍生了非常丰富的其他应用,如 Kjargaard 等人^[15]研究了群体用户的聚集行为,他们的工作以智能手机采集的数据为基础,汇集从智能手机获得的用户行为加速度数据、移动方向数据以及信号强度数据,通过层次聚类生成用户的群体聚集与移动信息;Zhang 和 Li^[3]、Ge 和 Xiong^[4]等人则研究了基于出租车历史行为轨迹的驾驶员异常行为探测算法,对于给定的两点(起点和终点),多数人选择的路径就是“正常”的路径,因此若有某些驾驶轨迹明显背离了这些正常路径,则这些行为属于异常行为。

3 用户行为分析

除了智能交通应用,基于位置信息的行为分析也已渗入用户的日常生活中,从个人的活动模式到群体的社交关系,都有可能从用户的历史活动轨迹记录中提取获得,以此为依据,位置服务的提供者又能进一步改进他们的产品,为用户提供更为个性化的服务。

3.1 用户行为理解

对用户行为的意图分析是用户轨迹分析的一个重要研究内容,其目的在于从用户的历史行为轨迹中挖掘和解释用户的日常行为规律。在对这一问题的研究中,聚类分析是最为常用的技术手段,然而在另一方面,由于缺乏必要的验证信息,对聚类结果的解读通常需要结合特定的时

空上下文进行。如 Kirmse 等人对用户日常行为特点的研究^[16],使用 meanshift 聚类根据用户的历史行为轨迹数据生成了其日常驻留区域,并结合时间特点给出了结果的语义解释(如用户在白天驻留时间最长的地点是工作场所,晚上驻留时间最长的地点是家里);而在 Ying 等人^[17]的研究中,则充分利用空间信息辅助对结果的解读:首先把聚类获得的用户日常驻留点与其周边的兴趣点(如公园、学校、银行、酒店等)相关联,进而根据用户的出行轨迹把这些驻留点“串联”起来,如“出门—学校—单位—下班”等,以此实现对用户出行轨迹的语义解读。

也有部分研究者尝试直接根据用户的行为轨迹数据建立其统计生成模型,主要做法是引入文本处理的相关技术。首先在轨迹数据与文本数据之间建立如下映射关系:

- 轨迹数据中的一个区域(如一个 POI 或一条街道)对应文章组成中的一个单词,用户经过一个区域的次数相当于单词在文章中出现的次数;
- 一条行为轨迹对应一篇文章;
- 一组轨迹构成的集合对应一个文本集合。

在以上对应关系下,Zheng 与 Ni 把对用户出行行为的理解映射为文本处理中的主题抽取^[18],进而构造了 LDA 模型^[19]的一个变体,以解释用户出行轨迹的生成过程;与此相类似的还有 Yuan 等人的工作^[20],把 LDA 聚类的结果与城市中的服务设施位置相结合,从而实现基于用户行为轨迹的城市实际功能的分区别。

3.2 社交关系理解

众所周知,在社交网络中,存在著名的“六度分隔”理论,即任意两个用户之间的通信最多只需由 6 个顺次认识的用户进行信息交换即可实现,而这一理论的正确性也已被 Jure 与 Horvitz 使用微软公司的 3 000 多万名 MSN 用户的历史通信记录所验证^[21]。此外,参考文献[22]中对多个社交网络数据集的分析结果还指出,在社交网络中,随着网络半径的扩大,人与人之间的交互距离反而缩小;另一方面,对于社交关系时空分布的特点,当前还没有定论。近年来,随着社交网络与位置服务的盛行,对用户的物理位置与其社交联系两者之间的关联性分析也激发起了广泛的兴趣。

在基于位置信息的用户关系分析中,Crandalla^[23]等人做了开创性的工作。以照片分享网站 Flickr 上的数据为研究对象,由于用户在向 Flickr 上传照片的同时也留下了时



间信息,此外,在照片上载过程中有部分用户提交了他所在的位置信息,根据这些记录可以获得用户的时空位置信息;另一方面,Flickr 还提供了社交网络服务,由此,又可以确定用户在网络上的社交关系信息。以这些数据为依据,Crandalla 等人研究了用户在物理空间中的行为轨迹与其社交关系之间的联系,在一定的简化假设下,证明了如下结论:若两位用户在连续 k 天内访问了相同的区域,则他们在社交网络中存在朋友关系的概率可以写为 $c_1 e^{c_2 k}$, c_1 和 c_2 是常数因子。

由这一结果可知,当两个人连续出现在相同的场合时,他们间彼此认识的概率将以共现的次数呈指数增长。这一发现为探索人群中的社会关系提供了新的手段:对所关心的目标对象,只需简单地比较他们的活动轨迹在一定时间周期内的重合次数,即可断言其是否存在朋友关系。

与参考文献[23]相反,Cho 及其合作者则研究了用户间的社交关系对他们出行活动的影响^[24]。分析了位置服务社交网站 Gowalla 与 Brightkite 的用户数据以及由欧洲某电信运营商提供的用户定位和联系数据。结果显示,一方面,两个用户成为朋友的概率随物理距离的增加而降低;另一方面,当用户做旅行规划时,他的社交关系对目标地点的选择存在一定的影响:对于短途旅游,用户对目标地点的选择一般独立于其社交关系;对于长途旅游,目标地点一般会被优先考虑在有朋友居住的地方。

与参考文献[24]相类似的工作还有 Ye、Yin 和 Lee 的研究^[25],在这一工作中,他们分析了位置服务社交网站 Foursquare 的用户数据,结果显示,若两个用户间存在社交联系,则在相同的地理位置共同出现的概率要远远高于他们间没有社交关系时的概率,即对于用户 A 和 B,记他们去过的地方集合分别为 L_A 和 L_B ,令 $r = \frac{|L_A \cap L_B|}{|L_A \cup L_B|}$,则在 A、B 两者是朋友时, r 的取值远大于两者不是朋友时的取值。

容易看出,参考文献[24,25]的发现对旅游行业的广告推荐具有非常重要的意义,由于用户在规划长途旅游时会倾向于选择有朋友或朋友去过的地方,因此基于用户社交网络联系的目标地点推荐有可能会获得良好的效果。

3.3 服务推荐

推荐服务是在线社交系统与内容服务系统的核心功

能之一,常见的推荐内容包括广告、电影、音乐、朋友关系以及旅游线路等,为能向用户提供个性化的推荐服务,近年来,大量的工作投入对用户行为特点与待推荐目标两者之间联系的分析之中。

在社交网络服务中,一个核心的应用是朋友关系的发现和推荐,参考文献[23]的结论为此提供了支持,如 Ling 等人的工作^[26]可以视为对参考文献[23]的直接应用,在这一工作中,Ling 等人设计了向量化的数据结构来表示用户在空间上的历史访问记录和共同出现记录,进而使用欧氏距离度量用户间的熟悉程度,并在社交网络中“相熟”用户间进行朋友推荐。

另一方面,在参考文献[23,26]等的工作中,用户间“相似性”的计算依据是用户与物理空间中特定参照物间的相对位置关系,这一计算方式虽然直观而简单,却不能直接应用于远距离用户间的相似度计算。事实上,当两个用户间的物理距离充分大时,容易猜测,他们多次在同一区域中共同出现的概率将趋向于 0,但在现实中,即使在这种极端的情况下,有很多用户仍然在社交网络中结为好友。为解决这一问题,Xiao 等人^[27]的做法是把原始的空间位置抽象为更高一级的概念,使用位置空间的语义代替具体的物理地址进行相似度计算,在这一计算模式下,可以认为,当两个用户的行为轨迹语义相近时,则他们间有可能存在朋友关系。例如,若有两位用户的出行轨迹都是“购物中心—餐馆—电影院”,则认为他们是相似的,因而可以在他们之间互相做朋友关系并推荐位置服务。

除了发现新的社交关系外,协助用户强化现有的联系也是社交网络平台所关注的焦点之一。Braga 等研究人员提出了一种通过分析用户的线下行为来促进其线上交互的策略^[28],以用户的行为轨迹为研究对象,把最小有界矩形(minimum bounding rectangle)应用于好友用户的行为轨迹,从中提取这些用户的日常公共活动区域,进而在社交网络上为他们共享这些结果,以此促进彼此的线上互动。

旅游推荐是用户行为轨迹分析的另一个重要应用。由于社交关系对用户旅游地点的选择起着重要的影响,因此可以认为,这一结论从侧面证实了病毒营销策略在旅游广告推荐中的有效性,此外预期它还将深刻地影响未来旅游广告的推荐行为,社交关系将成为广告推荐的主要依据。事实上,在 Ye 等人^[29]的研究中,已经对此做了初步的尝试。

4 结束语

对于基于用户行为轨迹的分析应用,一方面,在内容上,智能交通领域的主要焦点仍然聚焦在对驾驶路径诱导、交通状态判断等传统问题的研究上;相对而言,基于互联网的应用研究则更为活跃,特别地,随着社交网络因素的引入,相关工作已经展现了乐观的前景。然而,需要指出的是,作为一个新兴的研究方向,有众多的问题有待进一步探讨,其中最为核心的一个问题是如何针对特定的商业需求,从用户的历史轨迹中提取深层的关联行为,这也是支持和推动这一方向深入发展的动力。另一方面,在手段上,“大数据”已经成为对用户行为轨迹分析的主流方向。

对于智能交通系统的研究,近年来,随着 DITS 的兴起,“大数据”已经成为这一领域研究中必不可少的组成部分。这是由于在动态交通数据采集系统中,随着采集手段的完善,需要处理的数据规模随之高速增长。这里仅以国内某省级公路交通数据中心为例进行说明。

浮动车数据是该数据中心的主要数据资源之一,目前接入中心内的浮动车数目已经接近 100 000 辆,平均每车以 15 s/次的频率向数据中心发送实时交通状态信息(每条信息包括记录编号、车辆识别号、车辆位置的经纬度、车行方向、瞬时速度以及时戳),据此估算,每天该数据中心将新增约 17 亿条记录,以每条记录占 30 byte 计算,则每天该数据中心新增约 48 GB 的浮动车数据。

在其他各项应用中,“大数据”依然占有十分显著的位置。如 Crandalla 等人^[23]为分析用户的物理距离与社交关系的联系,使用了 Flickr 上的 3 800 万张照片记录;Cho 及其合作者^[24]使用了位置服务社交网站 Gowalla 的 640 万条用户登录数据和发生在 196 591 个用户间的 950 327 对用户关系数据、网站 Brightkite 的 450 万条用户登录数据和发生在 58 228 个用户间的 950 327 对用户关系数据以及由欧洲电信运营商提供的发生在 200 万用户间的 4.5 亿次通话联系数据。

之所以需要如此规模庞大的数据,原因之一在于数据获取的便利性,对于车辆的 GPS 轨迹数据,相关的运营商处有完整的记录;对于用户的社交网络数据和其他定位数据,由于当前主流的位置服务和社交应用平台(如 Facebook、Twitter 以及 Flickr 等)已经积累了大量的用户数据,通过使用网络爬虫等工具,研究人员可以较为方便地

抓取获得这些数据。但根本的原因还是数据的稀疏性。以 Cho^[24]的工作为例进行说明:容易理解,虽然参考文献[24]中所使用的 3 个数据集都包含了大规模的用户活动记录,然而在任一数据集中,相对于用户总量的规模,发生在用户间的关系(无论是社交关系,还是在物理空间上的接近关系)仍然是极为稀疏的,试图在这两种关系之间建立概率联系,显然会导致一个更为稀疏的结果,为了克服这一问题,研究者只能通过提升原始数据的规模以获得更多的支持。

在“大数据”的背景下,对高效快速的数据分析处理算法的研究,已成为用户行为轨迹的分析与应用研究的核心内容之一。参考文献[29]中提出了一个著名的观点:“更多的数据胜过更聪明的算法”,这一论断在现有的工作中已经得到了充分的证明。事实上,在现有的对用户行为轨迹的研究工作中,大多强有力的结果都建立在高度简化的假设以及对现有数据的统计、分析和拟合之上^[23,24];也有其他工作是对某些成熟高效的机器学习算法的应用,如决策树^[12,14]、mean shift 聚类^[16]等。由此,在这一研究领域中,对现有算法和模型的合理选择与高效实现,也许比新型算法的设计更为重要。

参考文献

- 1 Minakakis R. Evolution of mobile location-based services. *Communication of the ACM*, 2003, 46(12)
- 2 Qudus M A, Ochieng W Y, Noland R B. Current map-matching algorithms for transport applications: state-of-the art and future research directions. *Transportation Research Part C*, 2007(15):312~328
- 3 Ge Y, Xiong H, Liu C, *et al.* A taxi driving fraud detection system. *Proceedings of the 11th IEEE International Conference on Data Mining (ICDM'11)*, Vancouver, Canada, 2011: 181~190
- 4 Zhang D Q, Li N, Zhou Z H, *et al.* iBAT: detecting anomalous taxi trajectories from GPS traces. *Proceedings of the 13th ACM International Conference on Ubiquitous Computing (UbiComp'11)*, Beijing, China, 2011: 99~108
- 5 Chen C, Zhang D Q, Castro P S, *et al.* Real-time detection of anomalous taxi trajectories from GPS traces. *Proceedings of the 8th Annual International ICST Conference on Mobile and Ubiquitous System(MobiQuitous'11)*, Copenhagen, Denmark, 2011: 63~74
- 6 Zhang J P, Wang F Y, Wang K F, *et al.* Data-driven intelligent transportation systems: a survey. *IEEE Transactions on Intelligent Transportation Systems*, 2011, 12(4)



- 7 Cao L, Krumm J. From GPS traces to a routable road map. 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS 2009), Seattle, WA, 2009:3~12
- 8 Chen Y H, Krumm J. Probabilistic modeling of traffic lanes from GPS traces. 18th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS 2010), San Jose, CA, USA, 2010
- 9 Gonzalez H, Han J W, Li X L, *et al.* Adaptive fastest path computation on a road network: a traffic mining approach. VLDB 2007, Vienna, Austria, 2007
- 10 Yuan J, Zheng Y, Zhang C Y, *et al.* T-drive: driving directions based on taxi trajectories. Proceedings of ACM SIGSPATIAL 2010, New York, USA, 2010
- 11 Bastani F. A greener transportation mode: flexible routes discovery from GPS trajectory data. GIS 2011, San Jose, California, USA, 2011
- 12 Zheng Y, Liu L, Wang L H, *et al.* Learning transportation mode from raw GPS data for geographic applications on the web. Proceedings of International Conference on World Wide Web (WWW 2008), Beijing, China, 2008
- 13 He H B, Garcia E A. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(9)
- 14 Leon Stenneth, Ouri Wolfson, Philip S Yu, *et al.* Transportation mode detection using mobile phones and GIS information. GIS 2010, San Jose, CA, USA, 2010
- 15 Mikkel Baun Kjargaard, Martin Wirz, Daniel Roggen, *et al.* Detecting pedestrian flocks by fusion of multi-modal sensors in mobile phones. UbiComp 2012, Pittsburgh, Pennsylvania, United States, 2012
- 16 PabloBellver A K. Extracting patterns from location history. GIS 2011, San Jose, California, USA, 2011
- 17 Ying J C, Lee W C, Weng T C, *et al.* Semantic trajectory mining for location prediction. GIS 2011, San Jose, California, USA, 2011
- 18 Zheng J C, Lionel M Ni. An unsupervised framework for sensing individual and cluster behavior patterns from human mobile data. UbiComp 2012, Pittsburgh, Pennsylvania, United States, 2012
- 19 David M, Andrew Y, Michael I. Latent dirichlet allocation. Journal of Machine Learning Research 3 (4-5): 993~1022
- 20 Yuan J, Zheng Y, Xie X. Discovering regions of different functions in a city using human mobility and POIs. KDD 2012, Beijing, China, 2012
- 21 Leskovec J, Horvitz E. Planetary-scale views on a large instant-messaging network. Proceedings of the 17th International Conference on World Wide Web, New York, USA, 2008
- 22 Leskovec J, Kleinberg J, Faloutsos C. Graphs over time: densification laws, shrinking diameters and possible explanations. KDD 2005, Chicago, IL, USA, 2005
- 23 David J Crandalla, Backstrom L, Cosley D, *et al.* Inferring social ties from geographic coincidences. Proceedings of the National Academy of Sciences of the United States of America, 2010
- 24 Cho E, Myers S A, Leskovec J. Friendship and mobility: user movement in location-based social networks. KDD 2011, San Diego, CA, USA, 2011
- 25 Xiao X Y, Zheng Y, Luo Q, *et al.* Finding similar users using category-based location history. GIS 2010, San Jose, CA, 2010
- 26 Huy Pham, Ling Hu, Cyrus Shahabi. Towards integrating real-world spatiotemporal data with social networks. GIS 2011, San Jose, California, USA, 2011
- 27 Ye M, Yin P F, Wang-ChienLee. Location recommendation for location-based social networks. GIS 2010, San Jose, CA, USA, 2010
- 28 Braga R B. A trajectory correlation algorithm based on users' daily routines. GIS 2011, San Jose, California, USA, 2011
- 29 A few useful things to know about machine learning. Communications of the ACM, 2012, 55 (10):78~87

[作者简介]



陈康,男,硕士,中国电信股份有限公司广东研究院工程师,主要研究方向为基于位置的服务技术。

黄晓宇,男,华南理工大学讲师,主要研究方向为机器学习与数据挖掘。

王爱宝,男,博士,中国电信集团公司技术部高级业务督导,目前主要从事移动互联网、位置服务、社会化网络等领域的工作。

陶彩霞,女,硕士,中国电信股份有限公司广东研究院高级工程师,主要研究方向为移动互联网、大数据和个性化推荐技术。

关迎晖,女,硕士,现就职于中国电信股份有限公司广东研究院,主要研究方向为移动互联网应用。

李磊,男,中山大学教授、博士生导师,主要研究方向为人工智能、智能规划与机器学习理论。

(收稿日期:2013-03-05)