

移动轨迹数据分析与智慧城市

潘 纲 李石坚 齐观德 张王晟
浙江大学

关键词：智慧城市 移动轨迹

引言

近年来，以智能手机等移动设备为载体的移动社交网络的兴起，极大地方便了人与人之间的社交活动，并正深入影响人们的社会生活方式。作为移动社交网络的主体，人的移动性带来的位置轨迹不仅记录了人的行为历史，也记录了人与社会的交互活动信息。移动社交网络中位置轨迹数据的分析与利用，为解决城市问题提供了一种新的思路。

我国各大中城市普遍面临着交通拥堵、环境污染、社会老龄化、能源紧缺等问题，严重影响了城市的宜居程度和持续发展。这是我国城市化进程中面临的挑战，也是必须要解决的问题。信息技术的发展和广泛应用为城市问题的缓解与改善提供了机会，市政府、学术界和工业界均期望利用信息技术让城市更加智能，通过打造覆盖广、交互深、协同好的“智慧城市”来惠及市民生活。

动态性是城市的重要特性。城市里各种移动对象的位置变化，如人的移动、车辆的运动等，反映了城市环境的变动，是城市动态性的直接体现。不仅如此，移动对象的位置轨迹还体现了人们在城市中各种社会活动，也是人们移动社交网络的一个重要体现。分析大量移动对象的历史轨迹数据，可以定量描述和估计人们的社会活动特征，进而可发现人们的行为（包括社交行为）在不同时空粒度下的统计规律，从微观到宏观的不同尺度上认知和把握纷繁多变的城市动态性。

移动轨迹数据分析不仅移动社交网络的重要

研究内容，也是“智慧城市”的研究对象和研究热点，它综合利用了传感、识别、预测等手段，使人们能以前所未有的深度认知自己的生活轨迹、社交行为、环境的变动和城市的演进，为政府、商家和市民提供各类丰富、准确、动态的信息服务。

移动轨迹数据

如何获得城市中众多移动对象的轨迹数据？目前手机与公共交通设施是两类最常见的移动轨迹数据采集途径。从已有的研究工作看，移动轨迹数据的来源主要分为四种：

手机 已成为现代城市人群的标准通信设备。通过手机GPS或通信基站可获取机主的位置信息。利用通信基站信息测算的位置数据精度较低，通常误差大于200米，而利用手机自带的GPS接收机获取的位置数据精度则较高。该类数据覆盖范围很广，蕴含丰富的个性化内容，但由于用户GPS数据涉及个人隐私，此类数据目前尚不能广泛利用。

公共自行车 近年来部分城市为满足市民短途出行需求，提供公共自行车租赁，成为一项绿色环保的公益服务，巴黎、巴塞罗那、杭州、苏州等城市已开始运营。系统可记录所有自行车的出借/归还时间和位置。但考虑到设备成本，自行车上并未装载定位设备，因此无法准确记录每辆自行车的行驶轨迹。此类系统可以粗略反映一部分城市人群的出行信息，但信息量较少且覆盖人群较为有限。

表1 目前研究中使用的部分移动轨迹数据集

数据类型	简介	采集地点	时间(天)	数据量
手机基站数据	用户通信的时刻与估计位置	美国马萨诸塞州	44	1百万用户 ^[1]
	基站在每个小时的通信强度	意大利罗马	49	398个基站 ^[2]
	用户通信时对应的基站	葡萄牙	365	1百万用户 ^[3]
公共自行车租赁点数据	租赁点空车与借出车辆数	英国伦敦	100	358个租赁点 ^[4]
		西班牙巴塞罗那	98	390个租赁点 ^[5]
公共交通卡消费数据	用户公共交通的时间、消费、起始、终止位置等信息	英国伦敦	166	约30万人 ^[6]
出租车轨迹数据	包括出租车的GPS位置、行驶速度、载客状态	中国北京	205	约3万辆出租车 ^[7]
		中国杭州	385	约5000辆出租车 ^[8]
		中国深圳	25	5631辆出租车 ^[9]
	除一般数据外,还包括出租车收费情况	新加坡	630	15000辆出租车 ^[10]

公共交通卡 目前已被广泛用于公共交通费用的支付,如地铁、公交车等。公共交通卡消费记录可反映海量城市用户的活动情况,能较全面覆盖城市人群。但此类数据只能粗粒度描述移动对象在不同刷卡位置间的移动,而无法获取移动对象在其他时间或地方的位置。

出租车 目前国内大部分城市的出租车都已装备GPS,可记录各出租车位置轨迹。出租车也是一类常用的交通工具,在城市内的覆盖范围广,位置数据定位精度高、连续性强,且没有手机数据涉及的隐私问题,因此不仅在学术界还是在工业界,正成为一种相当重要的移动轨迹数据源。

表1列出目前部分研究工作用到的真实移动轨迹数据集,供读者参考。

主要研究内容

目前移动轨迹数据的研究主要集中在三个关键问题上:

移动轨迹特征分析 对移动轨迹数据的特征分析是后续分析的基础。轨迹特征分析主要研究如何从含噪音的海量原始数据中提取出可信的移动对象运动轨迹,进而分析车辆或人群运动轨迹的特征,在此基础上考察其与城市道路拓扑、司机驾驶习惯或个人行为习惯等因素之间的关系,以及轨迹特征在交通和规划等领域的应用。

人群移动规律发现 人群移动规律在城市规划、卫生防疫、公共安全等领域有极高的理论和应用价值。车辆作为人群移动的载体,其轨迹包含着丰富的人群移动信息。主要研究包括如何从移动轨迹等基础数据中挖掘人群移动信息,如流向、密度等特征量,并进而从挖掘得到的人群移动信息中分析时空环境下人群移动变化规律,如周期性、无标度性等普适规律,同时分析区域社会功能、土地用途等宏观因素对人群移动规律的影响。

社会活动模式挖掘 社会活动或社会事件是人群移动数据中蕴含的高级语义信息,也是人群移动社会性的体现。基于人群移动的社会活动模式发现主要研究如何从人群移动信息中发现特定社会事件及其涌现模式,如人群的快速聚集行为往往由突



图1 移动轨迹分析的科学问题与典型应用

发性群体事件诱发, 据此可通过分析人群的快速流动、停留时长和活动区域对群体事件进行判定和预警。进一步可分析社会活动模式和当地社会经济文化等方面的关系以及形成此种社会活动模式的内在原因, 并探索社会活动模式研究成果在城市管理中的现实应用。

移动轨迹分析的“智慧应用”

移动轨迹数据可看作一种在城市中人们社会活动的独特采样。从应用的角度看, 研究移动轨迹数据到底会给智慧城市带来哪些帮助? 我们认为有以下几方面。

监控道路环境 道路环境包括道路基础设施与道路交通情况。监控道路环境是智慧交通的重要应用需求, 也是移动轨迹分析(特别是车辆数据分析)最早的应用领域。微软研究院的约翰·克鲁姆(John Krumm)等人用车辆轨迹数据还原车辆行驶的道路信息, 分析并计算道路宽度、道路通行方向与道路交叉情况, 在城市道路地图重建方面进行了一系列工作^[11,12]。中科院深圳先进技术研究院的樊建平等人使用出租车行驶中的速度信息, 挖掘道路上的交通热点^[9]。

改进交通服务 交通相关的各类主体(包括市民、驾驶者与政府)均可得益于车辆轨迹数据的研究, 使总体交通情况得以改善。新加坡管理大学的拉杰什·克里希纳·巴兰(Rajesh Krishna Balan)等人通过分析出租车轨迹数据, 帮助市民在乘车前就清楚地了解两地的预期通行时间与花费^[10]。由于交通情况的复杂性, 通行速度并不仅仅依赖于两地距离, 为帮助驾驶者寻找最优路径, 微软亚洲研究院的谢幸与郑宇等人研究出租车轨迹数据, 利用出租车司机丰富的驾驶经验为

人们提供导航建议^[13,14]。

监测城市动态 城市动态性通常体现为城市人群活动的动态变化, 车辆轨迹反映的人群移动信息是城市动态的重要指示器。美国麻省理工学院的Senseable City Lab对城市动态可视化进行了大量研究, 通过手机基站数据的可视化, 将难以把握的城市动态变为直观的图像^[1,16]。事实上, 城市动态可视化也可以通过分析车辆轨迹数据来实现。不仅如此, 研究者甚至可以通过分析历史数据, 对未来的城市动态进行预测, 例如对个体移动轨迹^[17]以及公共交通使用情况的预测^[5,18]。

评估城市规划 我国的城市规划一般是由政府部门根据城市发展中长期目标制定的, 难以对其效用进行细粒度的定量考察和评估。通过分析车辆轨迹数据, 可以对城市规划情况做更为精确地了解 and 评估, 从而发现城市规划中的问题。例如, 浙江大学从出租车轨迹数据中提取人群移动数量的时序变化, 发现人群移动的时序数据可以反映城市对应区域的社会功能^[8]。微软亚洲研究院的郑宇与谢幸等人则从交通规划的角度, 分析出租车轨迹数据中由于交通规划造成的绕路、低速等情况, 检测目前交通规划中存在的问题^[7]。

我们的工作

近两年来, 我们对出租车轨迹数据进行了较深



图2 杭州市主城区乘客密集的区域(不同颜色用于区分相邻的区域)

人的研究，分别从出租车司机、乘客以及政府的角度，研究出租车数据可能为智慧城市带来的各类智能服务。通过分析出租车轨迹数据，我们希望帮助出租车司机更好地盈利；帮助乘客规避黑车、服务不太规范的出租车司机；帮助政府了解城市的“移动”情况以便更好地对城市进行规划。

快速寻找乘客^[19]

对出租车司机来说，最重要的问题当然是盈利，而最关键的则是如何

最快找到乘客。一个熟练的出租车司机可以几乎不间断地搭载客人，而缺乏经验的出租车司机却往往找不到最快搭载乘客的路线。幸运的是，移动轨迹分析可以帮助出租车司机们解决这个问题。我们根据历史数据，分析并预测一个区域未来的乘客数量，可为出租车司机提供最快找到乘客的路线。

“依据历史数据预测乘客数量”是一个时间序列预测问题，我们需要根据这个序列中的若干既往乘客数量值，预测乘客数量在未来的变化趋势。解决这个问题的困难主要集中于两方面：第一，大部分区域都只有很少的乘客数量，这些区域的乘客数量具有很大的随机性和不可预测性；第二，预测模型往往需要根据数据的特征进行设计，不同区域的乘客数量变化具有不同的时序特征。

针对这两个困难，我们首先通过聚类算法找到那些乘客密集的区域（图2显示了主城区的密集区域），并且将这些区域作为推荐给出租车司机的候选区域。其次，注意到所有区域的乘客数量都会以一天为变化周期，我们对不同日但相同时间段的数据使用改进的差分自回归移动平均模型（ARIMA）进行预测，将乘客数据分成等长的时间段，然后对

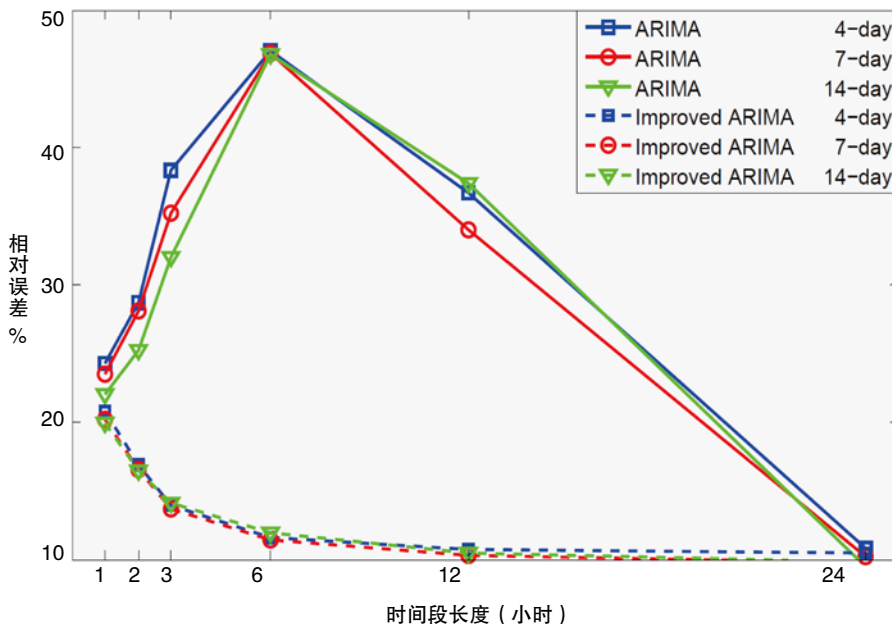


图3 预测效果比较：ARIMA模型 Vs. 改进模型

每个时间段的乘客数量进行预测。如图3所示，我们分别使用4天、7天、14天数据进行训练，测试了不同的时间段长度（从一小时到一天不等）预测的相对误差，发现针对时间段进行预测的改进ARIMA模型要比一般ARIMA好，预测的相对误差在10~20%之间。

在掌握了区域未来的乘客数量之后，又该如何推荐最快的载客路线呢？我们考虑了载客过程中两段最主要的时间开销，即从当前地点到载客点的行驶时间与在载客点等待乘客上车的时间。我们将历史数据中对应路段的平均行驶时间作为到载客点行驶时间的估计，并假设乘客到来是随机而均匀分布的，建立简单的排队模型计算在载客点等待乘客上车的时间。我们将这两个时间之和最小的路线作为最快的载客路线。从实验结果可发现，我们推荐的载客路线可以帮助出租车司机平均节省37%的时间。

异常轨迹检测^[15]

在乘坐出租车时，不熟悉路途的乘客经常会产生这样的疑问：这么长时间了怎么还没到目的地，

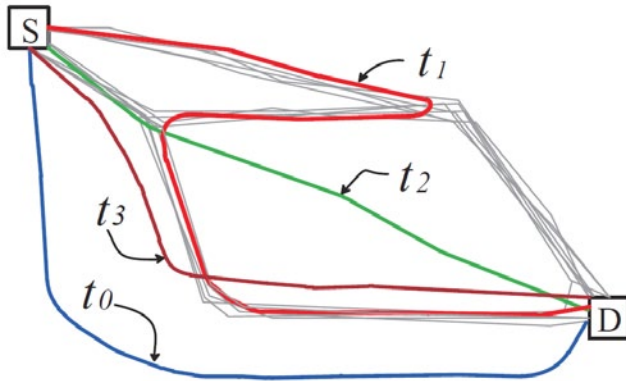


图4 灰色轨迹为频繁轨迹， t_0 、 t_1 、 t_2 、 t_3 均为异常轨迹

出租车司机是不是绕路了？之所以会有这一问题，是因为一些出租车司机违反服务规范，绕路、违规通行等；即使出租车司机选择的路线距离是最短的，但由于途经红绿灯多，反而行驶时间较长；此外，有的道路可能是限制通行的，但有的出租车司机违反了规定。定义“正常”的出租车行驶行为需要考虑诸多因素，十分困难。但是，如果假定大多数司机的行为都是正常的，据此检测不频繁的出租车轨迹就可以很容易地发现那些出租车中的“少数派”。我们与法国国立电信学院张大庆教授的课题组合作对此进行研究。

我们将那些不频繁的轨迹定义为异常轨迹。这些异常轨迹反映了出租车司机不合常态的驾驶行为，这些行为很有可能也是不合理或不恰当的。我们可以将两地之间的所有轨迹简单地堆积，如图4所示，S和D之间存在三条比较频繁且不长行驶路线，而 t_0 、 t_1 、 t_2 、 t_3 是四条不频繁的轨迹，被定义为异常轨迹。

应如何从众多轨迹中检出异常轨迹？我们将轨迹映射成字符串然后使用字符串匹配的算法来寻找异常轨

迹。从数据结构来说，由于GPS采样的不连续性，每条轨迹实际上由一组点构成。为了去除同一条道路上不同位置的GPS差异，我们将每个GPS点映射为一个 $250\text{m} \times 250\text{m}$ 的方块，并用唯一的ID表示。一条轨迹就可以表示为由这些ID构成的字符串。对这些字符串匹配可以比较轨迹的相似程度，使用Isolation Forest算法就可以将所有字符串表达为树的一个结点，而结点的深度对应于它与其它结点的匹配程度，从而快速找到那些异常轨迹。

区域功能识别^[8]

自动地进行城市土地用途分类是城市规划领域的一项重要工作。政府部门做城市规划时希望某个区域具备哪些主要社会功能，但是规划实施后的区域不一定与预期完全相符。传统的土地用途分类算法往往使用卫星遥感技术，所得到的分类也多为物理形态上的类别。我们从城市人群流动数据出发，揭示区域功能和区域人流的关系，对城市区域的社会学功能进行分类。

我们经常可以看到区域功能影响人流的例子：

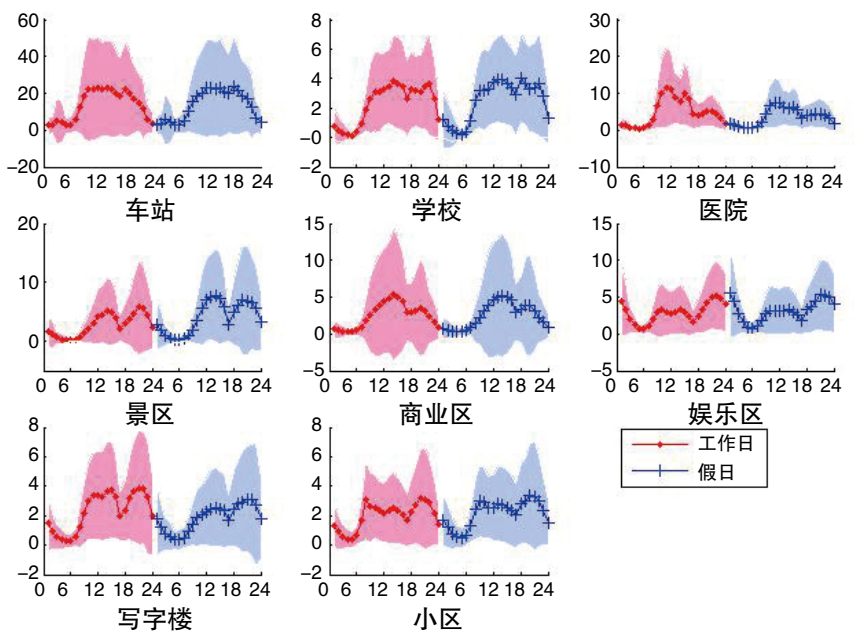


图5 不同类别区域的平均人群流动模式与标准差（阴影部分）

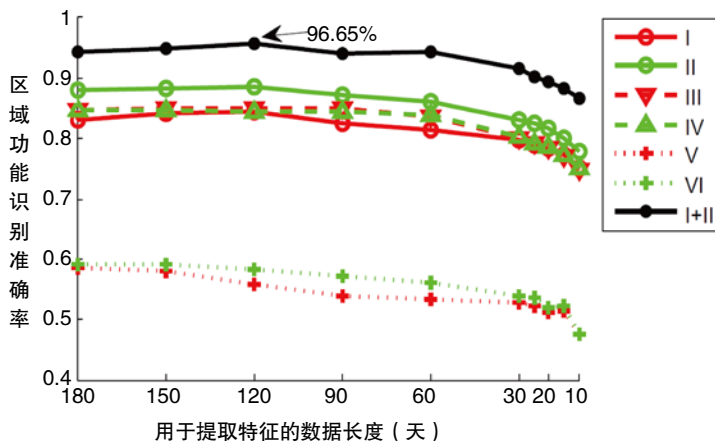


图6 六种特征与最优组合 (I+II) 在不同数据长度下的分类结果

假日时景区的人比较多，晚上娱乐场所的人会显著多于白天。反过来，区域人流的时序变化是不是也可用于识别区域的社会功能呢？的确，我们发现特定功能区的人流时序变化确实具有某种模式（图5），并且探讨了将其用于区域功能分类的可能性。

识别区域功能是一个分类问题。首先，可以从出租车轨迹数据中提取乘客数量的时序数据；其次，根据乘客数量的空间分布，对城市区域进行聚类，以找出城市的热点区域；再次，设计六个特征用于描述这些时序数据中反映的每个热点区域的六种对应的乘客变化模式。最后，使用SVM、KNN、LDA与反向传播算法对这些特征进行分类，从而识别原区域的功能。在具体实验中，我们测试不同数据长度用于提取这些特征，并且在不同算法下选择了最优特征组合。图6所示为最佳的分类算法（SVM）在不同时序长度和不同特征（以及最优特征组合I+II）的算法性能，其中最好的特征可以做到约95%的识别率。

结语

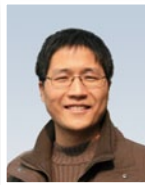
移动轨迹数据分析之所以如此重要，是因为单一对象的活动反映了其自身的行为特征，群体的活动反映了该群体共同的行为特征，而同一城市大量

移动对象的活动则反映了该城市的社会活动特征。分析移动对象轨迹中蕴含的知识是认识城市行为不可或缺的步骤。我们以出租车轨迹数据为研究对象，围绕出租车司机、乘客以及政府的不同需求开展分析，并尝试相关人员提供更智能、更有效的信息服务。我们相信，随着更多更丰富的数据积累和应用，基于这些异构数据集的跨界挖掘和知识的融合将带来大量有趣的理论问题和更具潜力的新型应用。■



潘纲

CCF普适计算专委会秘书长。浙江大学计算机学院教授。主要研究方向为普适计算、计算机视觉、智能系统等。
gpan@zju.edu.cn



李石坚

CCF普适计算专委会委员。浙江大学计算机学院副教授。主要研究方向为普适计算、移动互联网、社会计算等。
shijianli@zju.edu.cn



齐观德

浙江大学计算机学院博士生。主要研究方向为普适计算、移动数据挖掘等。
qiguande@zju.edu.cn



张王晨

浙江大学计算机学院博士生。主要研究方向为社会计算和复杂网络等。
zws10@zju.edu.cn

参考文献

[1] Calabrese, F. et.al., The geography of taste: analyzing cell-phone mobility and social events, The 8th International Conference on Pervasive Computing, 2010, 22~37

- [2] Reades, J. et.al., Cellular census: Explorations in urban data collection, *IEEE Pervasive Computing*, 6(3): 2007,30~38
- [3] Calabrese, F. et. al., Interplay between telecommunications and face-to-face interactions: A study using mobile phone data, *PLoS ONE*, 6(7):e20814, 2011
- [4] Lathia, N. and Ahmed, S. and Capra, L., Measuring the impact of opening the London shared bicycle scheme to casual users, *Transportation Research Part C: Emerging Technologies*, 22: 2012, 88~102
- [5] Froehlich, J. and Neumann, J. and Oliver, N., Sensing and predicting the pulse of the city through shared bicycling, *International Joint Conference on Artificial Intelligence (IJCAI 2009)*, 1420~1426
- [6] Lathia, N. and Capra, L., How smart is your smartcard? Measuring travel behaviours, perceptions, and incentives, *Proceedings of the 13th International Conference on Ubiquitous computing*, 2011, 291~300
- [7] Zheng, Y. et. al., Urban computing with taxicabs, *Proceedings of the 13th International Conference on Ubiquitous Computing*, 2011, 89~98
- [8] Pan, G. and Qi, G. et. al., Land-use Classification Using Taxi GPS Traces, *IEEE Transactions on Intelligent Transportation Systems*, 2012. Accepted
- [9] Liu, S. et. al., Towards mobility-based clustering, *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2010, 919~928
- [10] Balan, R.K. and Nguyen, K.X. and Jiang, L., Real-time trip information service for a large taxi fleet, *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services*, 2011, 99~112
- [11] Cao, L. and Krumm, J., From GPS traces to a routable road map, *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2009, 3~12
- [12] Chen, Y. and Krumm, J., Probabilistic modeling of traffic lanes from GPS traces, *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2010, 81~88
- [13] Yuan, J. and Zheng, Y. and Xie, X. and Sun, G., Driving with knowledge from the physical world, *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2011, 316~324
- [14] Yuan, J. et. al., T-drive: driving directions based on taxi trajectories, *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2010, 99~108
- [15] Zhang, D. et. al., iBAT: detecting anomalous taxi trajectories from GPS traces, *Proceedings of the 13th international conference on Ubiquitous computing*, 2011, 99~108
- [16] Calabrese, F. et. al., Real-time urban monitoring using cell phones: A case study in Rome, *IEEE Transactions on Intelligent Transportation Systems*, 99, 2010, 1~11
- [17] Song, C. et. al., Limits of predictability in human mobility, *Science*, 327(5968) , 2010, 1018~1021
- [18] Kaltenbrunner, A. et. al., Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system, *Pervasive and Mobile Computing*, 6(4): 2010,455~466
- [19] Li, X. et. al., Prediction of urban human mobility using large-scale taxi traces and its applications, *Frontiers of Computer Science*, 6(1): 2012, 111~121