

# 基于位置的社交网络用户签到及相关行为研究

李 敏 王晓聪 张 军 刘正捷

(大连海事大学信息科学技术学院 大连 116026)

**摘 要** Web2.0 时代,空间定位技术不断成熟,使得基于位置的社交网络(LBSN)快速发展。LBSN 用户的典型行为是签到以及针对签到地进行评论等。探索用户签到及相关行为的规律及背后动机,可以更好地了解用户的需求,发现系统设计与用户需求的不匹配之处,这对 LBSN 类应用的设计和开发具有一定的指导意义。利用在线数据抓取工具 GooSeeker 抽样国内典型的 LBSN 嘀咕网的用户数据。通过对获取的数据进行处理、分析,获知用户签到行为特点。同时关注用户发布的签到地评论的内容,并且使用分类工具 SVMCLS 将用户对麦当劳的评论划分为不同的倾向级别,从而得到用户对麦当劳的主观情感倾向性。结果发现嘀咕网用户签到的时间和地点存在规律性特征。用户倾向于在签到地做出正面的评论,并且评论的内容比较简短。这些发现有助于 LBSN 类系统设计和开发人员更好地了解用户,获知用户的需求,最终完善自己的设计,为用户提供更好的应用服务。

**关键词** 基于位置的社交网络,签到行为,评论,文本分析

**中图分类号** TP393 **文献标识码** A

## Study on Check-in and Related Behaviors of Location-based Social Network Users

LI Min WANG Xiao-cong ZHANG Jun LIU Zheng-jie

(Information Science and Technology College, Dalian Maritime University, Dalian 116026, China)

**Abstract** In Web 2.0 era, location-based social networks(LBSNs) are increasingly developed along with the maturity of spatial positioning technology. The typical behaviors of LBSN users are checking-in and commenting at check-in places. Exploring the rules and motivations behind check-in and related behaviors can make better understand of users' needs and find out the mismatches between system design and user needs, and it is meaningful for the design and development of LBSN applications. GooSeeker, an online data capture tool, was used to crawl Digu which is one of the most typical Chinese LBSNs. After processing and analyzing the data, the features of check-in behaviors were known. At the same time, comments at check-in places were also analyzed. Taking comments at MacDonalad as example, classification tool SVMCLS was used to classify the comments into different sentiment inclination levels. Ultimately, the rules and features of check-in times and places were presented, and it was also found that users tend to leave brief and positive comments at check-in places. All of the findings can help designers and developers get better understand of users and what users really needs, and then they can refine the designs and provide more appropriate applications and services to users.

**Keywords** Location-based social network, Check-in behavior, Comment, Text analysis

## 1 引言

LBSN 作为一种在线社交网络,越来越受到人们的关注,因为其结合了用户当前地理位置这一要素,使得用户的体验更有时空感。用户可以对地点进行签到,公开他们的地理位置,在签到地留下评论信息。签到地可以是餐饮地点、旅游景点、名胜古迹,甚至还有企业的品牌等,例如麦当劳、肯德基、阿迪达斯及星巴克等。用户在特定地点进行签到后,有可能获得商家的奖励积分、赢取相应的徽章、成为这个地点的老大(该地点签到次数最多的人)以及得到实体店特定的折扣。目前国外比较主流的 LBSN 有 Foursquare、Loopt 和 Gowalla

等,国内有嘀咕、街旁、玩转四方、切客、大众点评等。嘀咕作为国内最为流行的 LBSN 之一,由火兔这一微博类应用发展成为基于位置的社交网络应用,据统计到 2012 年 1 月已有超过 160 万手机客户端用户,并且日签到数在 10 万次以上<sup>[2]</sup>。嘀咕的用户群主要是年轻人群体,年龄区间在 18 到 35 岁之间。用户使用嘀咕手机客户端进行签到就能够获得实体店的折扣或者优惠券。除了签到,用户还可以进行结交好友、玩游戏等活动,并且可以绑定微博或者其他 SNS 网站,把自己的最新活动发布到上面去,与其他社交平台上的好友进行分享。

当社交与位置结合在一起后,新的社交网络比传统的社交网络更具有空间感,用户活动的时间、地点、内容更能体现

到稿日期:2012-12-28 返修日期:2013-03-22 本文受国家自然科学基金(61173035)和中央高校基本科研业务费专项资金(3132013041)资助。  
李 敏(1982—),女,博士生,主要研究方向为可用性与人机交互设计、社交网络分析、基于位置的服务, E-mail: minleely@gmail.com; 王晓聪(1986—),男,硕士,主要研究方向为人机交互设计与 IT 产品可用性; 张 军(1977—),男,博士,副教授,主要研究方向为可用性与人机交互设计、社交网络分析; 刘正捷(1958—),男,教授,博士生导师,主要研究方向为可用性工程、用户体验、人机交互和信息无障碍化。

真实状态下用户的生活、学习、工作状态。以往的研究比较关注地理位置相关信息,如签到地点命名规律、用户结交好友的偏好与地点的关系、基于地点的推荐系统等,而对用户总体的签到相关行为的关注不足。所以,我们关注嘀咕这个中国典型的 LBSN 用户的行为特点,包括用户签到时间、签到地点以及对该地点做出的评价内容等。通过研究签到及相关行为,可以更好地了解用户的活动规律及对应用的需求。这对系统的交互设计、开发人员提供了很好的获知当前系统不足及用户潜在需求的机会,并可以在此基础之上,设计开发更好的应用服务。同时,用户可以获得更好的体验。

本文第 2 节介绍了相关工作;接下来的第 3 节概述了数据获取及预处理过程;然后在第 4 节分析签到及相关行为;最后对本文工作做了总结及对下一步工作进行展望。

## 2 相关工作

我们关注 LBSN 相关的研究,以及针对评论内容情感倾向性的研究。以下的部分会介绍与这两个内容相关的工作。

关于社交网络的研究有很多,如根据用户发布的信息内容进行灾难预测。Oklahoma 等研究员利用 Twitter 进行火灾和水灾的预测<sup>[1]</sup>。有的研究是侧重发现社交网络结构的<sup>[2-4]</sup>。而当社交网络与位置结合在一起时,就产生了新的可研究的问题,比如预测用户的地理位置。研究人员通过研究 Twitter 上用户产生的内容来预测用户的位置<sup>[6]</sup>。还有通过分析 Facebook 用户间的社会关系来预测用户的地理位置<sup>[5]</sup>。分析移动模式并将其可视化也是目前 LBSN 及数据挖掘的研究热点问题。例如,Gonzalez 等人通过分析 10 万多名用户的活动轨迹,发现用户的行为呈现“可重现(reproducible)”模式<sup>[10]</sup>。Zheng 设计了一个利用用户的 GPS 轨迹数据来挖掘用户的旅游线路的原型系统<sup>[11]</sup>。还有研究 LBSN 用户行为及背后动机的。如 Lindqvist 等人在研究中分析了人们是如何以及为什么使用位置分享服务<sup>[7]</sup>,还有研究用 Twitter 带有的位置信息发现影响用户进行签到的动机<sup>[9]</sup>。还有研究与位置分享服务相关的隐私问题<sup>[7,8]</sup>。还有与签到相关的涉及到位置推荐<sup>[12-14]</sup>、内容推送服务<sup>[15]</sup>以及好友推荐<sup>[16,17]</sup>的研究。以上的研究多是以西方及欧美比较流行的 LBSN 为对象,而我们的研究关注中国最为流行的 LBSN 应用嘀咕。我们也发现,社交网络的普及以及 LBSN 被广泛应用使得消费者的购买决策很大程度上受到了社交网络中其他人言行的影响。于是,我们特别关注用户在消费类地点的言行规律。

在线评论内容常常带有作者的主观倾向,通过观点挖掘可以发现评论者的主观倾向。观点挖掘任务可分为主观性分析、极性分析和观点总结。主观性分析是判断文本是客观的,还是带有主观极性的。极性分析是判断主观性是正面的还是负面的,以及积极性强度等。观点总结是把观点信息进行整理和展示。情感(sentiment)首次出现在 Das 等人的文章中,被用来对评价内容进行自动分析<sup>[19]</sup>。有的关于产品评论挖掘的研究工作采用关联规则算法<sup>[21]</sup>,根据产品特征完成了基于中文的语义倾向情感分析,并且研究结果可以接近基于英文的同类研究。Fei 等人分析文本句子的组成,提取有代表性的词形成模式对照表,然后计算不同词性的语义倾向度,去

匹配测试文本,最后可有效地判断出文本的情感倾向<sup>[20]</sup>。本文中收集到的签到地评论信息大部分是简短的语句,这也正是 LBSN 用户产生的典型的内容。因此,我们只需要考虑句子级别的情感主观性分析即可。情感主观性分析涉及到的方法有决策树分类法、神经网络分类法、最近邻分类法、贝叶斯分类法以及支持向量机分类法等。针对获得的数据特点及各种方法的优缺点,我们的研究采用支持向量机(SVM)方法对所获得的用户评论信息进行分类。

## 3 数据获取及预处理

嘀咕用户可以通过网页和客户端两种方式登录使用嘀咕应用,所有的操作数据都记录在 Web 端,所以,我们可以通过抓取 Web 端用户个人主页数据来获得用户完整的操作数据。我们使用 GooSeeker([www.gooseeker.com](http://www.gooseeker.com)) 免费工具抓取了嘀咕网([www.aidigu.com](http://www.aidigu.com)) 用户个人主页数据。GooSeeker 采用广度优先的方式来抓取数据,我们选择了嘀咕网的一个活跃用户(嘀咕数、签到数、好友数、粉丝数相对比较高)为起始点,抓取他的个人基本信息(用户名、性别、所在城市、生日、爱好)、嘀咕内容(updates/posts)、签到地详情(包括签到地点名称、嘀咕评价、用户本人来过的次数、来过的总人数、总签到数、本地老大——签到次数最多的人、签到地评论内容及评论时间)、好友(跟随的人)列表及粉丝(跟随自己的人)列表等 5 部分数据作为第 1 层数据。然后抓取该用户的好友及粉丝的以上 5 部分的信息作为第 2 层数据。如此循环,共得到 3 层数据。整个抓取过程从 2011 年 7 月持续到 2011 年 12 月。由于嘀咕网的限制,只能抓取到签到地详情中当前页面上显示的部分的签到地评论,另外,设置了隐私权限的用户的页面只能抓取到基本信息,嘀咕内容等是无法获得的。最终,共获得了 5.7 万多名用户的数据,其中签到地评论数据 100 多万条。

利用上述的工具会抓取到重复的数据,例如,一个用户是第 1 层节点的好友,同时也是第 2 层某个节点的好友,那他就会被抓取到两次。所以,我们进行了筛选剔除,把重复抓取的数据清理掉,得到了一个原始的数据集,详情见表 1。另外,得到的数据是非结构化的,我们用 JAVA 语言编程,把需要的内容提取出来存入到数据库中,待进一步分析时使用。数据库中划分为 6 个表,分别是个人基本信息表、嘀咕内容表、签到地信息表、签到地评论表、好友表以及粉丝表。为了更直观地观察签到地,要把它们描绘到地图上,以进行研究。我们只能抓取到签到地点的名称而没有附带经纬度信息,于是,需要找到地点对应的经纬度,通过第三方网站 <http://gpsppg.com/latitude-and-longitude.htm> 批量转换得到了经纬度信息。受转换工具限制部分地点的经纬度是有缺漏的。

表 1 数据集

原始数据集	对应目录	数量
基本信息	seuc_digu_1	57722
嘀咕内容	mydigu_content	7672649
签到地信息	seuc_digu_1_2_1	1048576
好友信息	mydigu_1_11_1	1247512
粉丝信息	mydigu_1_11_2	3516702

为了更直观地表示出收集到的 5.7 万用户之间的相互关

系,我们使用了马里兰大学联合其他机构研究人员开发的专门分析社交网络的工具 NodeXL(<http://nodexl.codeplex.com>)绘制了用户之间存在的社交关系图,如图 1 所示。在图 1 中,中心位置的绿色节点是数据抓取的起始节点,也就是第 1 层数据。中间红色节点是起始节点用户的好友及粉丝,这些用户作为数据的第 2 层。最外围黑色节点是第 2 层用户的好友和粉丝的好友及粉丝,这些用户作为数据的第 3 层。

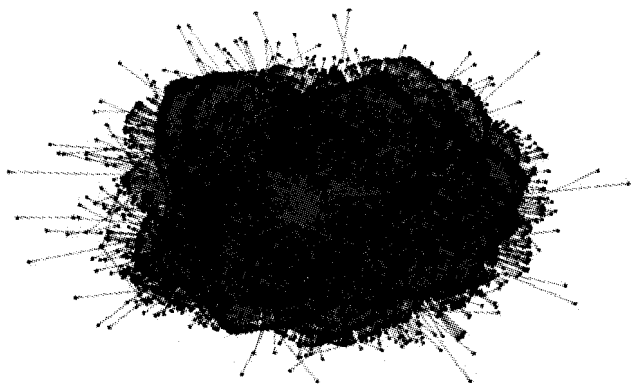


图 1 嘀咕用户社交关系图

## 4 签到及相关行为分析

根据收集到的嘀咕网的用户数据,我们重点关注中国 LBSN 用户签到及相关行为的特点,提出以下 3 个研究问题:

- (1) 嘀咕用户签到时间有什么规律?
- (2) 嘀咕用户签到地的地域分布有什么规律,哪些类型的地点更容易发生签到行为?
- (3) 嘀咕用户签到地评论的特点及情感倾向性是什么?

### 4.1 签到时间

为了获得签到时间上的规律,首先需要将包含签到地的信息取出来。由于包含签到时间的信息都包含在嘀咕内容中,我们从嘀咕内容表中提取签到地点不为空的签到时间列,筛选出小时和分钟,并计算出各个时段出现的频率,用户每天签到时间分布如图 2 所示。为了能够分析用户整段时间的签到规律,我们又计算了以周为单位的签到时间分布。通过计算得到日期和时间,再计算出周一到周日的签到时间分布,如图 3 所示。

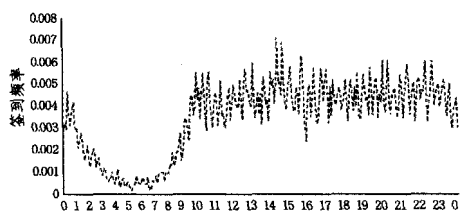


图 2 嘀咕用户一天签到时间分布

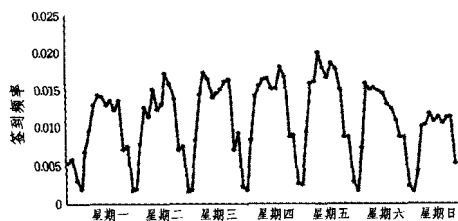


图 3 嘀咕用户一周签到时间分布

可以看出,一天当中用户在早上 8 点到 9 点之间签到数开始攀升,在中午 12 点到下午 1 点之间达到一天中的最高峰。可以推测早上 8 点到 9 点之间正好是员工去上班、学生去上课的高峰期,用户在这个时候喜欢用签到的方式记录自己一天的开始,或者是打发等车坐车的时间。而中午是就餐时间,最容易接触饭店、食堂这类地点,用户会利用这段休闲的时间进行签到。用户的评论内容也印证了我们的推测。从图中还可以看出用户在下午 6 点到晚上 12 点签到数平稳不下,原因可能是用户在这个时间属于下班后或者放学后的休闲时间,会在一些消费场所进行消费活动。而在早上 3 点到 5 点之间是睡眠时间,很少发生签到行为,这也符合客观规律。从签到时间的周分布图上看,周末的签到频率低于工作日,原因可能是周末的时候可以休息,不会像工作日那样有规律。另外,我们发现周五的签到频率最高,原因很可能是周五是工作日的最后一天了,下班之后就可以放松了,用户外出活动和消费次数明显增多,签到的机会也就增多。

### 4.2 签到地点

了解了用户在签到时间上的规律之后,我们想知道用户签到地的地域分布和签到地点选择上有哪些特点。由于有些地点是用户自己命名的,无法找到对应的经纬度信息,因此最终共得到 28.8 万个有完整经纬度信息的地点。然后,将同一经纬度的地点合并,共得到 14.5 万个不同经纬度地点(地点重复度达 50%)。利用这些签到地经纬度,通过 JavaScript 程序调用 Google Map API 的方式绘制了嘀咕用户在全国的分布图,如图 4 所示。从图中可以看出签到地热点主要集中在北京、上海、广州、深圳、重庆、成都等这些主干城市。我们根据签到地点,计算了地点所在省份的所占比例,发现广东省(包括广州和深圳两个城市)的签到最多,占到 50%。其次是上海和北京,各占 7%。而这一结果与分析街旁数据得到的结论相似<sup>[22]</sup>,这表明签到类应用主要在一线城市使用者较多。一线城市一直是众多服务提供商致力于抢占的高地,占据了这么大的战场,然后才推广到其他二三线城市。而且,很多服务提供商的本部都设立在这样的城市。所以,就近推广、发展也很合乎情理。另外,这类城市的用户群体更容易接受新事物,也促进了应用的推广。

了解了用户签到地的地域分布情况后,接下来要探究用户在这些地方的哪类场所进行签到。根据签到地关键字出现的频率筛选出频率较高的 36 大类,分别为:餐馆、茶馆、烧烤店、麦当劳、肯德基、咖啡店、酒店、旅店、洗手间、商场、商店、超市、机场、图书馆、理发店、公寓、会馆、公园、车站、地铁、广场、医院、银行、游乐园、体育馆、瑜伽店、健身馆、美容店、游泳馆、汽修厂、影城、KTV、俱乐部、旅行社、网吧及棋牌室。利用 Java 程序绘制了基于词频的标签云图,词频越高,字号越大,方便我们直观了解签到类型。使用 Java 的 awt 图形库制定字号,获取字符串将占用的矩形空间。文字的颜色、方向可直接随机生成。最后,生成了一幅嘀咕用户签到标签云图,如图 5 所示。从图中可以很清楚地看出最流行的签到地有餐馆、酒店、广场、咖啡店、公园以及其它能够反映日常活动的地点,如体育馆、麦当劳、银行、书店等。从这个图可以分析出,

用户签到最多的地点是餐饮类的场所,这可能与商家推出的鼓励签到的优惠活动有关,但也可以反映出嘀咕用户倾向于在餐饮消费场所进行活动。



图4 签到地域分布图

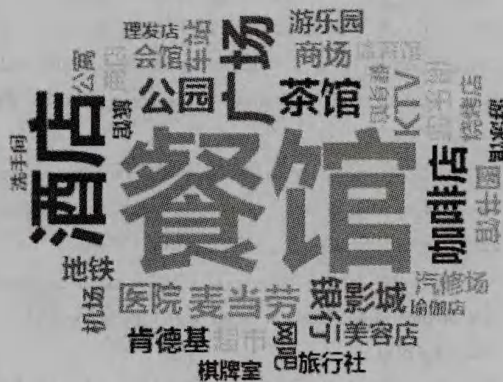


图5 签到地类型标签云图

### 4.3 签到地评论情感倾向性

许多用户都喜欢在签到的同时对地点发表评论,因此评论信息成为其他用户选择该地方消费的重要参考依据。如何挖掘签到地的在线评论信息,是我们研究的关注点。根据上面研究发现,用户在餐饮类地点签到较多,我们以快餐麦当劳这一典型签到地的评论信息为例进行分析。抽取了签到地评论中所有包含“麦当劳”这一关键字的信息,经过对外文信息、符号信息、无关的广告信息等的剔除处理后,共得到有关麦当劳的评论信息 50768 条。我们发现,这些评论信息都是很简短的,很少有长篇幅的内容,随机地抽取了其中的 5015 (约 10%) 条,利用该部分数据建立预测模型。首先,采用人工编码将评论信息分为 4 大类,分别是正面评论、中性评论、负面评论和无关评论。其中,正面评论指包含褒义词汇的评论,例如“美味”、“好吃”、“不错”。中性评论指包含一些中立词汇的评论,例如“一般”、“还行”、“还可以”。负面评论指包含贬义等负面情绪的词汇,例如“难吃”、“不好”、“太差了”等。无关评论指用户发布的与麦当劳没有明显关系的词汇以及我们无法判别的词汇。为使人工编码的结果更具有可靠性,我们采用 5 个人分别对 5000 多条评价进行编码,然后根据大于等于 3 的原则统一分类结果(大于等于 3 个人同意,才能确定该条

信息的类别)。我们一共得到 3585 条正面评论、93 条中性评论、359 条负面评论和 978 条无关评论,取其中的 80% 作为训练集,剩下的 20% 作为测试集,详见表 2。

表2 人工编码的实验语料

	正面	中性	负面	无关	总计
训练集	2868	73	287	782	4010
测试集	717	20	72	196	1005
所有语料	3585	93	359	978	5015

使用复旦大学的文本分类系统 SVMCLS 2.0<sup>[23]</sup>,辅助进行基于 SVM 的中文评论分类。本研究选择信息增益(IG)方法作为特征提取方法。其他参数的设置都采用工具提供的默认值。利用 SVMCLS 自带的结果分析功能建立的预测模型的总体准确率为 92.55%,说明我们的模型是可以预测剩余的大量数据的。各类查全率和查准率如表 3 所列。

表3 分类准确度

SVM	正面	中性	负面	无关	总体
查全率	98.89%	55.0%	72.22%	80.61%	92.55%
查准率	93.54%	91.67%	85.25%	90.81%	92.55%

得到了上述的预测模型,我们将对剩余的 4 万多条评论信息进行分类。最终得到 5130 条正面评论、397 条中性评论、301 条负面评论和 44939 条无关评论。由于要分析情感倾向性,我们剔除无关信息,对剩余的 5828 条正面、中性和负面评论进行分析,由此发现这些评价信息涉及的维度有口味、服务、环境、卫生和价格这几个方面。评价的维度中,口味是影响满意度的最重要的因素,其次是服务。正面评价中,对味道的评价占据 90%,服务次之,占据 6%。例如,有用户评价说:(2011-03-20 12:01 XX 在中山路好多麦当劳店)很好吃,服务也好,环境优雅。而负面评价中,对味道的评价占据了 78%,对服务的评价次之,占据了 18%。例如,有用户评价说:(2011-01-12 11:06 XX 在大连市天津街麦当劳)新推出的优惠套餐不好。环境、卫生和价格等因素也是用户评价的关注点。但是,与口味和服务相比,用户的顾虑不是太多。

我们又对带有倾向性的评论的时间分布情况进行了分析,当然,只抓取了一个时间段内的评论,所以时间只是抓取到该信息时显示的时间。我们从 2011 年 7 月开始抓取,到 2011 年 12 月截止,共抓取到了 2011 年 1 月到 2011 年 9 月这几个月的记录。结果发现,麦当劳在三月份的时候受到的关注最多,因为 2011 年 1 月 20 日嘀咕网与麦当劳合作推出特惠活动,网友可以在全国范围内的麦当劳签到获得勋章,从而获得优惠和免费的赠品,该活动持续到 3 月份。另外,我们还发现,用户倾向于在签到地留下正面的评论,而且没有随时间的变换而变化的趋势。这也是中国用户比较有特点的习惯,与我们之前进行用户研究的时候得到的结论一致。

**结束语** 通过对嘀咕网上抓取的信息进行分析,我们分别从用户的签到时间分布、签到地点分布和签到地评论情感倾向性这 3 个方面探索了 LBSN 用户的签到及相关行为主要有如下的发现:首先,用户签到时间存在周期性规律,中午达到峰值,下午签到要明显高于上午。周五是一周当中签到频率最高的一天。其次,一线城市签到频率明显高于其他城市,

而且主要是餐饮类地点。最后,用户趋向于在签到地留下好评,并且都很简短,不会随着时间而变化。

我们的研究发现对 LBSN 系统及其应用的设计和开发具有一定的指导意义,最终给用户带来更好的体验。例如,我们分析了用户的签到时间及地点的规律,让设计和开发人员了解用户的行为特点,从而提供更符合用户习惯的产品及功能。另外,利用分析结果,在设计 LBSN 类应用时可以提供更加个性化的功能,例如不同的时间、地点推荐不同的场所和服务,从而改善用户对产品和服务的使用体验。还有,对致力于研究 LBSN 的人员提供了一个新视角,使其可以从用户行为的角度来研究 LBSN。例如,根据用户倾向于在签到地评论及发布嘀咕内容的行为,可以做一些舆情分析,了解用户的动态,及时做出预测预警。还可以对用户的隐私关注问题进行研究,以及对数据做更深层次的挖掘。

本文的工作也存在着不足。我们会在后续研究中对比其他相似应用(如街旁等)来弥补只选取了嘀咕网作为我们研究平台的单一性。另外,获取数据的数量和质量还有待于补充和提高,主要是碍于抓取工具的局限性和嘀咕网站开放性等原因,我们后续会考虑使用网站的开放 API 接口结合 GooSeeker 工具获取数据。此外,根据地点经纬度信息绘制的地图的精度也存在不足,因为第三方经纬度转换网站得到的地点经纬度是存在偏差的。最后,对评论情感倾向性的分析,我们没有考虑到品牌或地点类型的差异,只以麦当劳为例,所得结论不一定都适用于其他类型品牌或地点。在未来的工作中,我们会改进和完善研究方法,更加准确地进行分析预测。基于研究发现,我们会进一步做基于个人偏好、位置或好友关系的推荐系统方面的研究。

### 参 考 文 献

- [1] Vieweg S, Hughes A L, Starbird K L. Microblogging during two natural hazards events; What twitter may contribute to situational awareness[C]//Proceeding of CHI. 2010; 1079-1088
- [2] Java A, Song X, Finin T. Why we twitter; Understanding microblogging usage and communities[C]//Proceeding of WebKDD/SNAKDD. 2007
- [3] Kwak H, Lee C, Park H, et al. What is Twitter, a social network or a news media[C]//Proceeding of WWW. 2010
- [4] Bernardo A H, Daniel M, Romero, et al. Social networks that matter; Twitter under the microscope[J]. First Monday, 2009, 14(1)
- [5] Backstrom L, Sun E, Marlow C. Find me if you can; improving geographical prediction with social and spatial proximity[C]//WWW. 2010; 61-70
- [6] Cheng Z, Caverlee J, Lee K. You are where you tweet; a content-based approach to geo-locating twitter users[M]. ACM CIKM, 2010; 759-768
- [7] Lindqvist J, Cranshaw J, Wiese J, et al. I'm the mayor of my house; examining why people use foursquare-a social-driven location sharing application[M]. ACM CHI, 2011; 2409-2418
- [8] Kostakos V, Venkatanathan J, Reynolds B. Who's Your Best Friend-Targeted Privacy Attacks In Location-sharing Social Networks[C]// Proceedings of UbiComp' 11. Beijing, China; ACM Press, September 2011; 177-186
- [9] Zhiyuan C, James C, Kyumin L, et al. Exploring Millions of footprints in location Sharing Services[C]//ICWSM, 2011
- [10] Gonzalez M C, Hidalgo C A, Barabasi A L. Understanding individual human mobility patterns[J]. Proceedings of ACL, 2008, 453(7196); 779-782
- [11] Zheng Y, et al. Mining interesting locations and travel sequences from GPS trajectories[C]//WWW'10. 2009
- [12] Zhuang J F, et al. When Recommendation Meets Mobile; Contextual and Personalized Recommendation on The Go[C]//Proceedings of UbiComp'11. ACM Press, 2011; 153-162
- [13] Zheng Y, Xie X. Learning travel recommendations from user-generated GPS traces[J]. ACM TIST, 2011, 2(1); 2
- [14] Zheng Y, Zhang L, Ma Z, et al. Recommending friends and locations based on individual location history[J]. ACM TWEB, 2011, 5(1); 1-5
- [15] Scellato S, Mascolo C, Musolesi M, et al. Track globally, deliver locally; improving content delivery networks by tracking geographic social cascades[M]. WWW, 2011; 457-466
- [16] Scellato S, Noulas A, Mascolo C. Exploiting place features in link prediction on location-based social networks[M]. ACM SIGKDD, 2011
- [17] Cranshaw J, Toch E, Hong J, et al. Bridging the gap between physical location and online social networks[C]// ACM UbiComp. 2010; 119-128
- [18] Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques[C]//The 2002 Conference on Empirical Methods in Natural Language. Philadelphia, Pennsylvania, USA, 2002
- [19] Das S, Chen M. Yahoo! for Amazon; Extracting market sentiment from stock message boards[A]//Asia Pacific Finance Association Annual Conference (APFA) [C]. Bangkok. Thailand, 2001
- [20] Fei Z C, Liu J, Wu G F. Sentiment classification using phrase patterns[C]// The Fourth International Conference on Computer and information technology. 2004; 1147-1157
- [21] 李实,叶强,李一军,等. 中文网络客户评论的产品特征挖掘方法研究[J]. 天津:管理科学学报, 2009(2); 146-156
- [22] 街旁网签到数据 用户主要集中在北上广四地[OL]. <http://www.yixieshi.com/it/8196.html>, 2011-07-13
- [23] SVM 分类器[OL]. [http://www.nlp.org.cn/docs/download.php?doc\\_id=283](http://www.nlp.org.cn/docs/download.php?doc_id=283). 2003-05-28