

学校编码: 10384

分类号_____密级_____

学号: 24320111152291

UDC _____

廈門大學

碩 士 學 位 論 文

基于位置社交网络的个性化地点
推荐算法研究

Research on Personalized Spot Recommendation Algorithm
Based on Location-Based Social Networks

王静金

指导教师: 林坤辉教授

专业名称: 计算机软件与理论

论文提交日期: 2014年4月

论文答辩日期: 2014年5月

学位授予日期: 2014年 月

指导教师: _____

答辩委员会主席: _____

2014年4月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1. 经厦门大学保密委员会审查核定的保密学位论文，于 年 月 日解密，解密后适用上述授权。

2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

摘要

随着社交网络和定位服务的发展,基于位置的社交网络逐渐兴起。在这个信息爆炸的时代,个性化推荐算法可以有效帮助社交网络的用户过滤掉自己不感兴趣的信息,更好的发掘用户的爱好,这样可以大大的增加用户在社交网络的活跃度。目前推荐算法已经在传统的在线社交网络平台上得到了广泛应用,对于基于位置的社交网络来说,推荐系统也是必不可少的。

相比传统在线社交网络推荐系统,位置社交推荐系统引入了地理位置信息,能更好地帮助推荐系统发现用户的喜好。目前,在位置社交推荐系统中,推荐类型主要包括好友推荐、活动推荐和地点推荐。其中地点推荐已经成为了最近研究的一个重点。然而由于基于位置的社交网络大部分的地点签到的人很少,甚至有些新开发的地点没有人签到过,使得地点推荐算法不可避免的面临数据稀疏和冷启动的问题。

本文对位置社交推荐系统进行了广泛的研究,介绍了基于位置的社交网络的结构、相关的推荐算法以及推荐系统的构成等。同时,针对个性化地点推荐这个类型,深入研究了影响地点推荐的三个因素:用户偏好、社交因素和地理位置,并针对这三个因素介绍了三种基本的个性化推荐算法:基于用户的协同过滤算法、基于社交影响因子的协同过滤算法和朴素贝叶斯分类。另外,本文深入分析了个性化地点推荐算法面临的数据稀疏、冷启动和签到地点距离常居地的远近这三个问题。并通过对这三个问题的分析,提出了一种改进的自适应地点推荐算法。该算法融合用户协同过滤、社交影响因子和朴素贝叶斯分类,并适应用户的当前位置,力求给用户推荐最适合用户当前位置的地点。

本文通过改进算法与其他推荐算法的实验对比分析,验证了改进算法的可行性和有效性。实验结果表明,改进算法的确可以解决个性化地点推荐面临的问题,可以更好地进行地点推荐。

关键词: 基于位置的社交网络; 协同过滤算法; 社交影响; 朴素贝叶斯分类

Abstract

With the development of the social networks and location-based services, location based social networks are increasingly popular. In this era of information explosion, personalized recommendation algorithms can effectively help social network users filter out information that they are not interested in. Using personalized recommendation algorithms also can explore users' interests and greatly increase users' activeness in social networks. At present, recommendation algorithms have been widely applied on traditional social network systems. And recommendation systems are essential to location-based social networks too.

Compared with traditional social network recommendation systems, location-based social recommendation systems can find user interests more effectively with the help of geographic location informations. At present, location-based social recommendation systems' contents mainly include friend recommendation, activity recommendation and spot recommendation. The spot recommendation has become one of the most hot research areas. However, most spots in the location-based social networks are rarely checked in. What's worse, some newly developed spots have none check in data. These make spot recommendation algorithms suffer from the sparsity of data and cold start problems.

This dissertation studies the location-based recommendation systems and introduces the structure of location-based social networks, related recommendation algorithms as well as the common composition of recommendation systems. As there are three main factors that influence the personalized spot recommendation, this dissertation studies the three factors: user preferences, social influence factors and geographic location, and introduces three basic personalized recommendation algorithms: the user-based collaborative filtering algorithm, the collaborative filtering algorithm based on social influence and the naïve bayesian classification. In addition, this dissertation deeply analyses the sparsity of data problem, cold start problem and check-in spots away from residence problem. And proposed an improved adaptive

recommendation algorithm based on the analysis of the three problems. The improved algorithm combines user collaborative filter, social influence factor and naïve bayesian classification, and it can adapt with users' current location for recommending the most suitable spot for the user.

In this dissertation, we compared the improved algorithm with other recommendation algorithms through comparison experiments and verify the feasibility and effectiveness of the improved algorithm. The experimental results show that, the improved algorithm can effectively solve the problems in personalized spot recommendation and can recommendation better spots for users.

Key Words: Location-Based Social Networks; Collaborative Filtering; Social Influence; Naïve Bayesian

目录

第一章 绪论	1
1.1 研究背景与意义	1
1.2 研究现状与面临的挑战	2
1.2.1 位置社交个性化推荐算法研究现状	3
1.2.2 面临的挑战	4
1.3 本文研究内容	6
1.4 本文组织结构	6
第二章 相关理论介绍	8
2.1 基于位置的社交网络	8
2.1.1 在线社交网络	8
2.1.2 基于位置的服务	10
2.1.3 基于位置的社交网络的层次结构	11
2.2 位置社交推荐系统	12
2.3 位置社交推荐算法概述	13
2.3.1 基于内存的算法	13
2.3.2 基于模型的算法	15
2.4 推荐效果的评测方法	16
2.4.1 度量方法	16
2.4.2 交叉验证	18
2.4.3 基于时间的分割方法	18
2.5 本章小结	18
第三章 个性化地点推荐算法研究	19
3.1 基于用户的协同过滤算法	19
3.1.1 用户间相似度的计算	20
3.1.2 用户对候选地点兴趣度的计算	22
3.1.3 Top-N 推荐	22

3.2 基于社交影响的协同过滤算法	24
3.2.1 社交影响因子	25
3.2.2 用户对候选地点兴趣度的计算	25
3.2.3 参数调节	26
3.3 朴素贝叶斯分类	26
3.3.1 朴素贝叶斯分类器简介	26
3.3.2 朴素贝叶斯在地点推荐中的应用	27
3.4 本章小结	30
第四章 一种改进的个性化地点推荐算法研究	31
4.1 个性化地点推荐算法面临的问题	31
4.1.1 数据稀疏问题	31
4.1.2 冷启动问题	32
4.1.3 签到地点离常居地的远近	33
4.2 问题解决方法的比较	34
4.3 改进的自适应地点推荐算法	35
4.3.1 UPS 算法	35
4.3.2 USG 算法	37
4.3.3 改进的算法 USPB	39
4.4 本章小结	41
第五章 实验设计与结果分析	43
5.1 实验数据集	43
5.2 实验环境及实验的算法	44
5.2.1 实验环境	44
5.2.2 实验验证的算法	45
5.3 标准情况算法实验	45
5.3.1 算法验证流程	45
5.3.2 参数调节	47
5.3.3 实验结果与分析	47
5.4 特殊情况下算法实验	49

5.4.1 数据稀疏	49
5.4.2 冷启动	51
5.5 本章小结	51
第六章 总结与展望	53
6.1 总结	53
6.2 展望	54
参考文献	55
攻读硕士学位期间发表的论文	60
致 谢	61

厦门大学博硕士论文摘要库

Contents

Chapter 1 Introduction	1
1.1 Background and Significance of the Research	1
1.2 Research Status and Challenges	2
1.2.1 Research Status of Location-Based Social Recommendation Algorithms....	3
1.2.2 The Challenges	4
1.3 Contents of this Dissertation	6
1.4 Structure of this Dissertation	6
Chapter 2 Related Theory Introduction.....	8
2.1 Location-Based Social Networks	8
2.1.1 Online Social Networks.....	8
2.1.2 Location-Based Services	10
2.1.3 Structure of Location-Based Social Networks	11
2.2 Location-Based Social Recommendation Systems	12
2.3 Outline of Location-Based Social Recommendation Algorithms	13
2.3.1 Memory-Based Algorithms	13
2.3.2 Model-Based Algorithms	15
2.4 Recommendation Effect Evaluation Methods	16
2.4.1 Measurement Methods	16
2.4.2 Cross-Validation.....	18
2.4.3 Time-Based Segmentation Method	18
2.5 Summary	18
Chapter 3 Personalized Spot Recommendation Algorithm Research	19
3.1 User-Based Collaborative Filtering Algorithm	19
3.1.1 User Similarity Calculation	20
3.1.2 User Interest Calculation	22
3.1.3 Top-N Recommendation	22

3.2 Social Influence Based Collaborative Filtering Algorithm	24
3.2.1 Social Influence Factor	25
3.2.2 User Interest Calculation	25
3.2.3 Parameter Adjustment	26
3.3 Naive Bayesian Classification.....	26
3.3.1 Naive Bayesian Classifier Introduction	26
3.3.2 Application of Naive Bayesian in Spot Recommendation	27
3.4 Summary	30
Chapter 4 Improved Personalized Spot Recommendation Algorithm	
Research.....	31
4.1 Problem of Personalized Spot Recommendation Algorithm.....	31
4.1.1 Data Sparsity Problem	31
4.1.2 Cold Start Problem	32
4.1.3 Distance Between Spot and Residence.....	33
4.2 Comparison of Solutions.....	34
4.3 Improved Adaptive Spot Recommendation Algorithm.....	35
4.3.1 UPS Algorithm	35
4.3.2 USG Algorithm	37
4.3.3 USPB Algorithm	39
4.4 Summary	41
Chapter 5 Experimental Design and Result Analysis	43
5.1 Experimental Dataset.....	43
5.2 Experimental Environment and Verification Algorithm	44
5.2.1 Experimental Environment.....	44
5.2.2 Experimental Verification Algorithm.....	45
5.3 Standard Case Algorithm Experiment.....	45
5.3.1 Algorithm Verification Process	45
5.3.2 Parameter Adjustment	47
5.3.3 Experimental Results and Analysis	47

5.4 Special Case Algorithm Experiment	49
5.4.1 Data Sparsity	49
5.4.2 Cold Start.....	51
5.5 Summary	51
Chapter 6 Conclusions and Prospect	53
6.1 Conclusions	53
6.2 Prospect	54
References	55
Publications	60
Acknowledgements	61

厦门大学博硕士论文摘要库

第一章 绪论

1.1 研究背景与意义

基于位置的社交网络 (Location-Based Social Networks)^[1]是在线社交网络的一种。随着获取地理位置定位技术的快速发展,传统的在线社交网络逐渐转向了基于地理位置的社交网络。随着智能手机的普及,以及 3G、4G、Wi-Fi 和 GPS 的发展,地理位置信息越来越容易获得。目前,几乎每个拥有智能手机的用户都能访问地理位置服务。随着能够获取地理位置数据的设备快速增长,传统的在线社交网络如国外的 Facebook 和 Twitter,国内的人人网、新浪微博等都添加了一个新的位置维度到现有的 OSN (Online Social Network)^[2]结构中。位置服务和在线社交网络正趋于融合,社交网络结合地理位置数据已经演变为基于位置的社交网络。

LBSNs 允许用户使用手机提供有价值的信息,这让系统用户和人气得到快速增长。用户可以使用智能手机对地点进行签到,公开他们的地理位置,记录他们在这个地方的体验。用户还可以对签到地评分并留下评论信息,这样他们在不同的地方签到,在网上发布他们的位置,看他们的朋友在哪里。此外,他们也可以对朋友的签到地进行评论。签到地点可以是饭店、咖啡馆、旅游景点、机场,甚至可以在用户家里等。对于一些商家,用户在该地点进行签到后,就有可能获得商家的奖励积分、赢取相应的徽章以及得到该门店特定的优惠折扣。目前国外比较主流的 LBSNs 有 Foursquare, Gowalla, Geolife, Brightkite 等,国内有新浪微博,嘀咕、街旁、玩转四方、切客、大众点评等。Foursquare 等 LBSNs 从运营开始就吸引了大量的用户,并且用户数量仍在迅速增长。Twitter, Facebook 等传统在线社交网络也意识到了位置服务的重要性,并增加了相应的位置服务。据世界著名市场调研公司 ABI Research 预测,基于位置的社交网络用户在 2013 年已经达到 8200 万,总收益已经达到 33 亿美元^[3]。

LBSNs 用户数量的迅速增长直接结果就是 LBSNs 中的信息过载。如此大规模的用户数量和信息为用户的生活带来了很大便利,方便了用户交友、学习等,但同时也给用户带来了困扰。用户很难去找到搜索的关键字,因而更期待智能终

端应用通过用户的位置信息感知用户周围的环境,挖掘地理位置背后的社会属性及其相互关联性,并对用户提供个性化的位置服务。

研究基于位置的社交网络,挖掘出网络中用户,位置,活动等节点之间隐藏的模式,同时为网络中的用户进行用户推荐,地点推荐和活动推荐等等,从而使得基于位置的社交网络的用户可以获得更加丰富的社交体验。因此可以看出对基于位置的社交网络的个性化推荐研究有一定的先进性和实用性。同时在推荐系统领域,基于位置的社交网络的个性化推荐算法研究是非常新颖,也是目前非常热门的一个课题。

基于位置的社交网络的推荐系统也叫位置社交推荐系统 (Geo-Social Recommender System)^[4]。基于用户提供的位置数据,位置社交推荐系统可以产生三种不同类型的推荐:好友推荐、地点推荐和活动推荐。在 LBSNs 中,两个人可以互相同意成为朋友。朋友间可以互相分享他们去过的地方,来自这些地方的图片、旅游路线等等。好友推荐是指推荐系统给目标用户提供一个他可能认识的用户的列表。当一个用户在一个不熟悉的地方,可能他想知道哪里有好吃的或者景点。地点推荐可以为用户提供有用的帮助,指导用户去有趣的地方。对于活动推荐来说,如果用户想去某个地方,位置社交推荐系统可以向用户推荐吃饭、跳舞等活动。

本文主要对位置社交推荐系统中现有的推荐算法进行研究,以地点推荐为代表深入研究基于 LBSNs 的个性化推荐问题。此外,针对位置社交推荐系统中存在的数据稀疏和冷启动问题,本文提出了一种融合用户偏好、社交影响以及地理位置影响的自适应地点推荐算法来提升推荐效果,并且在 Foursquare^[5]数据集上对算法进行了进一步验证。

1.2 研究现状与面临的挑战

LBSNs 除了传统社交网络中人与人的联系外,还可以跟踪和共享人的地理位置。在给传统在线社交网络中的用户或用户活动内容增加空间位置信息后,基于位置的社交网络呈现给用户好友的信息变得更加丰富,这也使得网络变得更加复杂。

1.2.1 位置社交个性化推荐算法研究现状

位置社交个性化推荐算法依赖于用户过去的签到历史,利用用户对去过的其他地方的访问时间、显示评分或者评论来将他们与用户偏好相似的其他人相关联,预测他对于没去过的那个地方是否有兴趣,然后推荐与目标用户有着相似偏好和品位的人喜欢的好友、活动和地点。因此位置社交个性化推荐算法有三种不同类型的推荐:好友推荐、地点推荐和活动推荐。

好友推荐在传统社交网络的中有了广泛的研究。传统的好友推荐方法是基于用户的个人信息,底层的社交结构和用户的交互模式。基于位置的社交网络提供了一个新的方法进行好友推荐,即推荐的时候同时考虑用户的历史位置信息。历史位置信息包含着丰富的上下文信息,同时它与用户真实的社交行为有着显著的相关性。一些研究也显示出地理位置信息对于确定社交网络中的用户关系起着重要的作用。基本的直觉是,用户的历史位置信息揭示了用户的偏好,因此有相似历史位置信息的用户有着相似的偏好,并更容易成为好友。一些论文研究了用户的地理位置对他们的社交关系的影响。Liben-Nowell 等在文献 [6] 中分析了在线社交网络中至少 2/3 的好友关系是由用户的位置决定的。Scellato 等在文献 [7] 中分析了从来自基于位置的社交网络 Foursquare 的数据,结果表明大约 40% 的好友关系是在 100 公里以内^[8]。DeScioli 等在文献 [9] 中揭示了用户的社交关系和他们的地理距离是高度相关的,居住在附近的用户更可能成为朋友。因此在 LBSNs 中,用户的历史位置信息可以提高好友推荐的有效性和效率。

基于 LBSNs 的活动推荐是指为一个特定的地点推荐一个或一个以上的活动。例如,可以为在北京奥林匹克公园附近的用户推荐观光、划船或慢跑等活动。对于一个地点,可以直接从用户标记的标签获取可能的活动列表,或者从用户的历史位置数据和位置的语义数据推断得到。在 LBSNs 中,一些用户的历史位置数据很少,而有些地点的访问人数也很少。为了解决数据稀疏问题,基于协作学习的方法利用来自所有用户的签到信息来发现活动。Sheng 等在文献 [10] 中为位置和活动历史数据使用用户-地点-活动张量进行建模。Bao 等在文献 [11] 中使用矩阵分解模型改进了之前的算法。

地点推荐已经成为最近研究的一个重点。地点推荐算法利用的数据主要有用户的个人信息数据,位置元数据以及用户的历史位置数据。一些地点推荐算法通

过将位置元数据(如位置描述和标签等)与用户的个人信息匹配来推荐地点。Park等在文献 [12] 中使用贝叶斯网络模型,对用户的个人信息(年龄,性别,饮食偏好)与餐馆的价格和类型做匹配。Lu等在文献 [13] 中探讨了各个点之间的内空间和时间关系轨迹以确定相关的用户偏好的活动的子序列,和分配他们一个语义。Ye等在文献 [14][15] 中提出一种方法基于用户签到的时间分布来提取位置的特点。用户的历史位置数据包括他们对历史位置的评分数据和签到数据。使用用户的历史位置数据进行地点推荐比单独的利用用户的个人信息数据推荐更有优势,它可以忽略跟用户很匹配但用户评分很差的地点。Chow等在文献 [16] 中,Horozov等在文献 [17]中, Ye等在文献 [18]中以及 Del Prete and Capra等在文献 [19]中都提出了使用协同过滤模型的个性化地点推荐算法。这些方法背后的直觉是用户与相似的用户有一样的对地点的偏好。

大多数的协同过滤算法有如下三个阶段:

- 1、基于历史评分数据计算用户之间的偏好相似度。
- 2、根据用户当前的位置为用户选择前 N 个用户最可能去的地点。
- 3、对推荐结果进行质量评估。

其中 Ye等在文献 [18]中使用用户的签到信息研究协同过滤模型,地理距离和社交结构对地点推荐的影响。作者发现地理距离在他们的模型中影响最大。Shi等在文献 [20]中根据用户历史位置数据建立了一个类别-正则矩阵,再基于这个矩阵进行地点推荐。这个地点推荐不仅考虑了用户的喜好也考虑了基于类别的地点的相似度。Bao等在文献 [11]中提出了地点推荐算法中三个关键的点。第一,用户的当前位置,这限制了地点推荐的候选。第二,用户的历史位置数据,这反映了用户的偏好。第三,其他用户(包括当地专家)的位置数据,这被认为是社会环境。

1.2.2 面临的挑战

虽然传统的 OSNs 很成功,但 LBSNs 向前跨出了一大步。向传统 OSNs 添加位置维度是一件具有挑战的事。传统基于 OSNs 的推荐算法与基于 LBSNs 算法的区别如表 1-1 所示。

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库