**班级 011214**

**学号 01121390**

**西安电子科技大学**

本科毕业设计论文



**题 目**  **基于位置社交网络中的好友**

**推荐系统开发**

**学 院**  通信工程学院

**专 业**  通信工程

**学生姓名** 王长通

**导师姓名** 朱晓妍

毕业设计(论文)诚信声明书

本人声明：本人所提交的毕业论文《先进无线（光）通信试验室主页设计与实现》是本人在指导教师指导下独立研究、写作的成果，论文中所引用他人的无论何种方式发布的文字、研究成果，均在论文中加以说明；有关教师、同学和其他工作人员对本文的写作、修订提出过并为我在论文中加以采纳的意见、建议，均已在我的致谢词中加以说明并深致谢意。

本论文和资料若有不实之处，本人承担一切相关责任。

论 文 作 者： （签字） 时间：2016年5月23日

指导教师已阅： （签字） 时间：2016年5月23日

**毕业设计（论文）任务书**

学生姓名 王长通 学号 01121390 指导教师 朱晓妍 职称 副教授

学院 通信工程学院 专业 通信工程

题目名称 基于位置社交网络中的好友推荐系统开发

任务和要求

1. 了解国内外相关推荐算法，并完成推荐系统的编写；
2. 阅读和翻译不少于10000单词的相关外文文献；
3. 独立完成毕业论文工作，论文字数应不少于15000；
4. 要求论文结构合理，概念清楚，逻辑清晰，语言通顺以及文笔流畅；
5. 工作作风严谨，刻苦钻研，每周与指导老师至少一次沟通联系一次；
6. 按时参加论文答辩工作。

开始日期 2016年3月6日 完成日期 2016年5月23日

院长（签字） 2012年1月10日

毕业设计（论文）工作计划

学生姓名 王长通 学 号 01121390

指导教师 朱晓妍 职 称 副教授

学 院 通信工程学院 专 业 通信工程

题目名称 基于位置社交网络中的好友推荐系统开发

1. 毕业设计（论文）进度

|  |  |
| --- | --- |
| 起止时间 | 工作内容 |
| 2016.3.1-2016.4.1  2016.4.2-2016.4.9  2016.4.10-2016.5.15  2016.5.16-2016.5.23 | 准备相关参考资料，了解相关推荐算法，选出适合该系统的算法。学习android开发，为编写系统做准备。  找寻地理位置相关数据集，并对数据进行预处理。  具体代码编写，将推荐算法在系统中实现。研究不同算法效率，提出改进算法。  完善系统，撰写毕业论文，准备答辩。 |

1. 主要参考书目（资料）

[1] 詹尼士，《推荐系统》 人民邮电出版社，2013.

[2] 查九, 李振博, 徐桂琼. 基于组合相似度的优化协同过滤算法[J]. 计算机应用

与软件, 2014, 323-328.

[3] 李刚，《疯狂Android讲义》 中国工业出版社，2015.

[4] Liao, Ll, Patterson, D.J., H. Building Personal Maps from GPS Data. In proc. of *IJCAI MOO05*, Springer Press(2005), 249-265

1. 主要仪器设备及材料

硬件：计算机、Andriod手机、Internet等

软件：Eclipse、SQLite等

四、教师的指导安排情况（场地安排、指导方式等）

1、每周集中汇报、指导一次；

2、采用电话、邮件随时联系指导。

西安电子科技大学毕业设计（论文）中期检查表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **学 院** | 通信工程学院 | | **专 业** | 通信工程 | |
| **学生姓名** | 王长通 | **学 号** | 01121390 | **班 级** | 011214 |
| **导师姓名** | 朱晓妍 | **职 称** | 副教授 | **单 位** |  |
| **题目名称** | 基于位置社交网络中的好友推荐系统开发 | | | | |
| **检 查 内 容** | | **检 查 结 果** | | | |
| **题目是否更换及更换原因** | |  | | | |
| **学生出勤情况** | |  | | | |
| **进度评价 （完成总工作量的百分比）** | |  | | | |
| **质量评价、进度描述** | |  | | | |
| **总体评价 （按优、良、中、及格、不及格五挡评价）** | |  | | | |
| **存在的问题与建议** | |  | | | |
| **学院审核（盖章）** | |  | | | |

西 安 电 子 科 技 大 学

毕业设计（论文）成绩登记表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 学 院 | 通信工程学院 | | | 专 业 | 通信工程 | |
| 姓 名 | 王长通 | 学 号 | 01121390 | 成 绩 |  | |
| 题目名称 | 基于位置社交网络中的好友推荐系统开发 | | | | | |
| 指导教师 | 朱晓妍 |  | | 职 称 | 副教授 |  |
| 指导教师评语及对成绩的评定意见 | 签名 2012年6月16日 | | | | | |

|  |  |
| --- | --- |
| 评阅人评语及成绩评定意见 | 签名 2012年6月16日 |
| 答辩小组意见 | 签名 2012年6月16日 |
| 学院答辩委员会意见 | 签名 2012年6月16日 |

# **摘 要**

随着物质生活的不断丰富，人们在精神上的需求日益增加。能结识到情投意合的朋友也是精神世界的重要组成部分，但这在线下生活中往往是可欲而不可求得，那么帮助用户发现潜在好友，并向其进行推荐就成为设计本系统的主要目的。 我们知道兴趣类似的人们一般会出现在相似的场所，如音乐爱好者一般会经常出没KTV，篮球爱好者必然会经常出没于球场等等。本系统通过获取用户地理位置信息来推测用户兴趣进而挖掘用户相似度，并进行排序和推荐。智能手机的出现极大的方便了获取用户地理位置的环节，提高了该系统的可行性，该系统使用Android手机作为开发平台。为了数据的真实性和研究该系统所使用推荐算法的有效性，我们取得了主打签到功能的移动社交网络Brightkite的真实用户数据，将其导入到SQLite数据库中提高运算效率，并用其中真实的好友关系对不同算法的推荐结果进行评估。

本文首先介绍了本课题的研究背景和意义，然后介绍了协同过滤的概念和其中具体推荐算法的种类，如：Jaccard相似度算法，ConsineSimilar余弦相似度算法等。随后介绍了以上算法在本课题系统中的具体实现方法，并通过计算得出各种算法在该数据集中的表现，通过分析提出优化后的算法。最后是对本课题的展望与总结、本文的不足之处和对本课题可能存在的创新点。

**关键词： 相似度算法 位置信息 Android 数据库**

# ABSTRACT

With the wealth of material life, people's spiritual needs are increasing. To get to know an important part of the spiritual world is like-minded friends, but offline life is often desirable and can not be obtained, then the potential to help users find friends, and to make recommendations which would be the main purpose of the design of the system. We know that people with similar interests generally occur in similar places, such as music lovers usually frequented KTV, basketball fans will inevitably haunt the court and so on. The system by obtaining user location information to speculate further mining user interest similarity, sorted and recommendation. Smart phones which greatly facilitates the user to obtain location links and improve the feasibility of the system, which uses your Android phone as a development platform. In order to study the data the authenticity and validity of the system using the recommended algorithm, we have made mobile social network Brightkite main sign of real user data capabilities, import it into SQLite database to improve operational efficiency, with which the real relationship between friends recommended result of different algorithms for evaluation.

This paper introduces the background and significance of this issue , and then introduced the recommendation algorithm related species , such as : Jaccard algorithm , ConsineSimilar algorithm. Then the algorithm described above in the subject system implementation methods, and by calculation algorithms performance in the data set , the analysis put forward by the optimized algorithm . Finally, on the subject of the outlook and summary , deficiencies and innovation on this topic may be present in this paper .

**Keywords : similarity algorithm location information Android**

**Database**

# 目 录

**第一章 绪论**............................................................................................................................ 1

1.1

1.2

# 第一章 绪 论

1.1 推荐系统研究的背景和意义

随着互联网的不断发展和生活节奏的加快，人们产生了越来越多的信息，这些信息往往是破碎的，混乱的。从这些零散的信息中挖掘出有价值的信息，会极大方便人们的日常生活，并获得巨大的商业价值，所以数据挖掘作为一门挖掘互联网信息的学科变得越来越引人关注。推荐系统便是在数据挖掘的基础上运行的，它成为一个相对独立的研究方向一般被认为始自1994年明尼苏达大学GroupLens 研究组推出的GroupLens系统。该系统有两大重要贡献：一是首次提出了基于协同过滤来完成推荐任务的思想，二是为推荐问题建立了一个形式化的模型。基于该模型的协同过滤推荐引领了之后推荐系统在今后十几年的发展方向。GroupLens所出的推荐算法实际上就是目前广泛使用的基于用户的协同过滤推荐算法。在之后的十几年中，其它一些著名的协同过滤算法被陆续提出，主要的有基于物品的协同过滤算法，基于矩阵分解的协同过滤算法等等。当然，基于其它方法而非协同过滤的推荐算法也在不断地发展，这些不同的推荐算法之间的取长补短、相互融合也成为一个重要的研究方向。

通过推荐算法进行好友推荐的系统极大的提高了用户认识潜在好友的效率，提供了难得的兴趣交友渠道。研究并实现推荐系统，使科技进步和算法发展直接可以提高人们的生活质量。

1.2 推荐系统的发展现状

推荐算法作为一门技术广泛应用在电子商户、大规模零售业和各种知识管理应用中。与此同时，随着各个学科研究人员的参与及方法的引入，推荐系统的研究得到迅猛发展。来自人工智能、信息检索、数据挖掘、安全与隐患以及商业与营销等各个领域的研究，都在为推荐系统提供了新的分析和方法。由于可以获得海量的数据，算法研究方面取得了巨大进步。

同时推荐系统的商业化也在不断的推进，在巨大的商业价值的推动下，各种有效的推荐算法层出不穷，大多数的电子商务平台尤其是网络购物平台，都不同程度地集成了推荐算法。比较著名的有Facebook根据好友关系向用户推荐可能认识的好友、Amazon网络购物商城向用户推荐已购买货物周边商品、Google新闻个性化推荐等。国内如淘宝网会基于近期用户搜索向用户推荐商品、豆瓣网会根据用户近期的浏览条目推荐类似的内容、滴滴打车会根据用户和司机的地理位置推荐合适的司机等等。基于的信息Amazon 发布的数据显示，亚马逊网络书城的推荐算法为亚马逊每年贡献近三十个百分点的创收。

1.3 主要研究内容

本文介绍的系统基于地理位置信息作为数据进行挖掘，并根据用户间地理位置的相似度大小，向用户推荐潜在的好友。系统采用真实的数据集和好友关系，最终用三种方法实现好友推荐，在实现推荐后对每种算法得出的结果进行评价，并分析出每种算法在该数据集中表现出的特点，并提出作者自己的见解。

# 第二章 协同过滤和算法介绍

2.1 协同过滤

2.1.1 协同过滤的概念及一般步骤

在用户群过去的行为或意见的基础上，通过算法找出指定用户兴趣相投、拥有共同经验的用户群体，之后可以用该用户群体的喜好来推荐指定用户感兴趣的信息。

（1）收集可以代表用户兴趣的信息。包括影评网站使用评分的方式让用户对不同电影打分、本系统中用户在不同的地点签到，本这种方式被称为“主动评分”。另外一种是“被动评分”，是指系统根据用户的行为来代替用户完成评价，不需要用户自己录入或输入数据。比如电子商务网站根据购买记录向用户推荐可能需要的商品，搜索引擎根据用户的历史搜索记录向用户推荐可能感兴趣的新闻等。

（2）通过算法计算该用户和其他用户的相似度，找到其中相似度较高的其他用户作为最近邻集合。这一步骤有不同的算法实现，根据数据的不同、场景的不同它们的准确度各不相同，目前较多使用的相似度算法有Jaccard相似度、余弦相似度、Pearson相似度等。

（3）有了最近邻集合，就可以根据该集合中用户的历史数据对目标用户的兴趣进行预测，产生推荐结果。

2.1.2 优点

（1）能够过滤机器难以自动内容分析的信息，如艺术品，音乐等。

（2）共用其他人的经验，避免了内容分析的不完全或不精确，并且能够基于一些复杂的，难以表述的概念（如信息质量、个人品味）进行过滤。

（3）有推荐新信息的能力。可以发现内容上完全不相似的信息，用户对推荐信息的内容事先是预料不到的。可以发现用户潜在的但自己尚未发现的兴趣偏好。

（4）推荐个性化、自动化程度高。能够有效的利用其他相似用户的回馈信息。加快个性化学习的速度。

2.1.3 缺点

（1）新用户问题系统开始时推荐质量较差，即冷启动问题。

（2）新项目问题质量取决于历史数据集，对最新数据的增添时间不敏感。

（3）稀疏性问题，数据间联系不密切。

（4）系统延伸性问题，扩展性不好。

2.2 Jaccard相关系数

Jaccard相关系数强调两个用户数据并集的损益，用来比较样本集中的相似性和分散性的一个概率，其值等于样本交集与样本集并集的比值。给定用户u和用户v，其中N(u)表示用户u的评分项集合，N(v)表示用户v的评分项集合[1]。可通过公式（2-1）计算用户u和v的兴趣相似度。

（2-1）

Jaccard相关系数越接近1，两个集合重合度越高，两个数据集越相似。越接近0，两个集合重合度越低，两个数据集越不同。

2.3 ConsineSimilar余弦相关系数

ConsineSimilar余弦相关系数强调两个用户数据的方向一致性。余弦相关系数用向量空间中两个向量夹角的余弦值作为衡量两个个体间差异的大小。用户对 n个数据的评分可视为 n维向量，用户u和用户v的相似性即为相应两个n维 向量的夹角余弦，可通过公式（2-2）计算用户u和v的兴趣相似度。

（2-2）

ConsineSimilar余弦相关系数越接近1，两个向量方向越一致，两个数据集越相似。越接近0，两个向量方向越分歧，两个数据集越不同。

2.4 Pearson相关系数

Pearson相关系数强调两个用户数据一同变化移动的相同趋势。表示用户u的平均评分，表示用户v的平均评分，可通过公式（2-3）计算用户u和v的兴趣相似度。

（2-3）

Pearson相关系数越接近1，两种变化趋势越相关，两个数据集越相似。越接近

-1，两种变化趋势越矛盾，两个数据集越不同。

# 第三章 好友推荐系统实现

3.1 实现平台

3.2 数据集预处理

3.3 SQLite数据库

3.3.1 SQLite数据库介绍

3.3.2 表的设计

3.4 算法实现

3.4.1 计算Jaccard相关系数

3.4.1 计算Cosine余弦相关系数

3.5 系统界面

# 第四章 算法分析和改进

4.1 从公式上分析

4.2 系统数据集验证

4.3 提出改进算法

**致 谢**

**参考文献**

[1] 查九, 李振博, 徐桂琼. 基于组合相似度的优化协同过滤算法[J]. 计算机应用与软件,

2014, 323-328.

[2]

[3]

[4]

[5]