学 号\_\_\_\_\_\_\_\_\_\_\_\_

密 级\_\_\_\_\_\_\_\_\_\_\_\_

哈尔滨工程大学本科生毕业论文

面向群智感知任务推荐算法的设计与实现

院（系）名 称：

专 业 名 称：

学 生 姓 名：

指 导 教 师：

哈尔滨工程大学

2019年6月

**面向群智感知任务推荐算法的设计与实现 哈尔滨工程大学**

学 号\_\_\_\_\_\_\_\_\_\_\_\_

密 级\_\_\_\_\_\_\_\_\_\_\_\_

面向群智感知任务推荐算法的设计与实现

Design and Implementation of Crowd-sensing Oriented Task Recommendation Algorithm

**学 生 姓 名：**

**所 在 学 院：**

**所 在 专 业：**

**指 导 教 师：**

**职 称：**

**所 在 单 位：**哈尔滨工程大学

**论文提交日期：**

**论文答辩日期：**

**学位授予单位：**哈尔滨工程大学

摘　　要

移动群智感知作为一种高效的数据获取方式，以数量庞大、分布广泛且背景不同的用户作为感知源实体，可实现社会化推荐以及城市服务的应用。在现有的移动群智感知平台中随着任务数量的不断增长，用户将面对严峻的信息过载问题。如果用户有明确的目标，用户可以主动通过搜索引擎来过滤信息并找到合适的任务。但是如果用户没有明确的目标时，则需要推荐系统为用户自动生成推荐列表。

通过研究国内外的相关推荐算法，发现能够适用群智感知平台的推荐算法存在计算量大、维护成本高的弊端，因此本文针对移动群智感知平台引入一种考虑兴趣衰减并融合用户能力属性的任务推荐算法。在设计面向群智感知的任务推荐算法框架时，发现本文设计的算法框架存在着两个冷启动问题，于是对冷启动问题进行了分析并提出了解决方法。本文以移动群智感知平台作为推荐算法的运用场景，在用户都具有大量有效历史记录的条件下，算法从用户的投标记录入手分析用户的兴趣和能力。

本文的推荐算法综合考虑召回率和准确性，使用F1值定量描述系统推荐效果。通过实验结果分析，该算法可以较好地考虑到用户兴趣的衰减和用户能力的水平，并且该任务推荐方法在准确度上的比传统的仅考虑用户兴趣的任务推荐方法高大约百分之八。

关键字：任务推荐；冷启动；群智感知；协同过滤

**ABSTRACT**

As a kind of efficient data acquisition method, mobile crowd-sensing can realize social recommendation and even serve the urban by leveraging a large number of normal users, who are widely distributed and have different backgrounds, as the source of perception. While the number of tasks increases continuously, in the existing mobile crowd-sensing platforms, users will face severe information overload problems. If the user has a clear target, he can use the search engine to filter information on his own, which can help him find the right task. However, if the user does not have a clear target, the recommendation system is required to automatically generate a recommendation list for the user.

By studying the related recommendation algorithms at home and abroad, it is found that the recommendation algorithm that can be applied to the crowd-sensing platform has the drawbacks of large computational complexity and relatively high maintenance cost. Therefore, this paper brings in a certain recommendation algorithm for the mobile crowd-sensing platform which comprehensively considers the user’s interest and capability attributes. What’s more, the interest model takes user’s interest attenuation into consideration. In the process of designing the task recommendation algorithm framework for the mobile crowd-sensing platform, it is found that there are two cold start problems in the algorithm framework of this paper. Therefore, through analyzing the cold start problem, a certain solution is proposed in this paper. For the task recommendation algorithm will be used on the mobile crowd-sensing platform, users have a large number of effective historical records. So that the algorithm can analyze the user's interest and ability from the user's bid record.

The recommended algorithm in this paper uses the F1 value to describe the effect of the task recommendation, taking the recall rate and the precision rate comprehensively into consideration. Through the analysis of experimental results, the algorithm can give the attenuation of user interest and the level of user ability better consideration, and the effect of the task recommendation method has an advantage about eight percent over the traditional task recommendation.

**Keywords:** task recommendation; cold start; crowd-sensing; collaborative filtering

目　　录

[第1章 绪论 1](#_Toc9856415)

[1.1 研究背景、目的及意义 1](#_Toc9856416)

[1.2 国内外研究状况 2](#_Toc9856417)

[1.2.1 任务推荐算法研究现状 2](#_Toc9856418)

[1.2.2 推荐系统冷启动研究 3](#_Toc9856419)

[1.3 论文研究内容及预期结果 3](#_Toc9856420)

[1.4 论文后续组织结构 4](#_Toc9856421)

[第2章 考虑兴趣衰减的任务推荐算法设计 6](#_Toc9856422)

[2.1 推荐算法框架 6](#_Toc9856423)

[2.2 任务模型 9](#_Toc9856424)

[2.3 用户模型建立 10](#_Toc9856425)

[2.3.1 基于兴趣衰减的用户模型 10](#_Toc9856426)

[2.3.2 融入胜任力的用户模型 12](#_Toc9856427)

[2.4 相似度计算 14](#_Toc9856428)

[2.5 推荐生成 15](#_Toc9856429)

[2.7本章小结 16](#_Toc9856431)

[第3章 任务推荐算法的冷启动问题研究 17](#_Toc9856432)

[3.1 用户冷启动 17](#_Toc9856433)

[3.1.1 用户冷启动问题分析 17](#_Toc9856434)

[3.1.2 用户冷启动问题的解决方法 19](#_Toc9856435)

[3.2 任务冷启动 20](#_Toc9856436)

[3.2.1 任务冷启动问题分析 20](#_Toc9856437)

[3.2.2 任务冷启动问题的解决方 20](#_Toc9856438)

[3.3 本章小结 21](#_Toc9856439)

[第4章 任务推荐算法的实验分析 23](#_Toc9856440)

[4.1 实验目的及参数选取 23](#_Toc9856441)

[4.2 实验环境 24](#_Toc9856442)

[4.3 实验结果分析 24](#_Toc9856443)

[4.3.1 实例化实验 25](#_Toc9856444)

[4.3.2 算法实验 27](#_Toc9856445)

[4.4 推荐效果优化实验 28](#_Toc9856446)

[4.5 网站实验 30](#_Toc9856447)

[4.6 本章小结 33](#_Toc9856448)

[结 论 36](#_Toc9856449)

[参考文献 37](#_Toc9856450)

[攻读学士学位期间取得的学术成果 39](#_Toc9856451)

[致 谢 40](#_Toc9856452)

# 第1章 绪论

## 1.1 研究背景、目的及意义

群智感知是一种群体配合下的数据获取方式，通过人们已有的移动设备形成交互式的、参与式的感知网络，并将感知任务发布给网络中的个体或群体来完成。2009年2月，以AlexPentland为首的多名教授在美国《科学》(Science)杂志上发表文章并提出了“计算社会学”概念[1]，认为大量感知数据可以用来理解个体、组织甚至社会。2012年清华大学的教授刘云浩在之前的概念上发展成为“群智感知计算”，即利用大量用户的移动设备作为基本感知单位，通过互联网实现感知任务的分配和感知数据的收集感知数据，最终实现大规模、复杂的城市与社会感知任务。移动群智感知平台的用户在移动端提交各种任务的解决方案，被采纳的方案提供者可以获得一定的奖励，这种奖励有任务发布者决定，可以是金钱上的，也可以是知名度或声誉上的。目前，移动群智感知的应用主要定位在城市服务这个方面，如创始于以色列的Waze地图。它的中文翻译是位智，是一款基于社区的导航软件。由其他大量用户产生的实时交通位置信息，为使用者提供导航最佳的服务。同时，使用者也为平台的其他用户贡献路况和道路信息。这种模式就是体现了群智的理念，人人为我，我为人人。移动群智感知还可以为社会的管理提供智能支持，如公共安全、社会化推荐、环境监测等[2]。

随着当今网络技术的不断发展，快速增长的信息总量超过了人和系统所能接受、处理或有效处理的范围。为了解决信息过载的问题，科学家和工程师提出了很多解决方案，其中最具代表的是分类目录和搜索引擎。推荐系统是一种不需要用户主动提供关键词的情形下帮助用户快速发现有用信息的工具，通过分析用户的历史行为来形成用户画像（profile），以此为基础向用户推荐满足他们特征的信息。如今推荐算法已经广泛地运用于互联网电商和社交平台，创造了巨大的收益和流量。其中，抖音APP就是一个靠优秀推荐系统获取短视频软件，能为用户推荐感兴趣的内容，做到所谓“千人千面”。它创造了月活跃用户三亿的流量奇迹，2018年底达到市值750亿美元，超越Uber成为最大的“独角兽”企业。国外的Netflix视频推荐网站和Amazon在线购物商城也都集成了推荐系统。推荐系统在电商领域的价值还在于挖掘长尾效应(Long Tail Effect)，所谓长尾就是那些种类多样但销量很低的产品，这些产品加起来所占的市场份额可以和少数热销产品的市场份额接近，甚至超越了热销产品。成立于1994年的亚马逊以线上书城起家，在1998年的时候开始使用协同过滤推荐算法。之后的二十年，推荐算法每年贡献百分之三十左右的创收[3]。但是由于移动群智感知平台的形成时间较短，没有如电商平台那般高度依赖推荐算法。用户还是依靠搜索引擎来避免信息过载问题，因此如何设计一个面向移动群智感知平台的有效的推荐系统成为群智感知领域的热点问题。具体针对移动群智感知平台的推荐系统应当能够发现感知用户的“千人千面”，也就是推荐那些小众的任务给对应的用户，让拥有独特技术或独特兴趣的用户能够充分发挥作用。这个过程同时也节约了用户挑选任务的时间，达到降低信息过载的目的。若是把所有任务随机放在首页，那么用户将面对大部分对他而言是无用的信息，此时则需要推荐系统为用户过滤掉无关的信息。

## 1.2 国内外研究状况

### 1.2.1 任务推荐算法研究现状

关于任务推荐算法的研究，国外和国内有不同的研究方式。国外主要面向粒度较细的微观任务，这类任务的目的是收集大量的答案，然后进行整合得到一个结果。所以这些任务不需要工人使用如何高深的知识去解决一个问题，对用户来说每一个任务都是平等的。

在Ayswarya·R·Kurup的论文[4]中提出基于工人报酬的推荐算法，利用历史投标信息建立工人-任务的成功概率矩阵，通过贝叶斯个性化排序，从大量任务中选取数量相对小的匹配任务推荐给工人。由于该算法精确地计算了用户对于平台上每个任务的投标成功概率，它的前提条件是精确的任务识别和任务匹配，通过评估信息之间的相关性来减少任务增加带来的信息过载问题，具有较好的推荐准确性。但同时不可避免地其维护的工人-任务矩阵的工作量过于庞大，并且后期随着任务数量的不断增长，成为用户数量的十倍甚至更多倍，矩阵将会表现出严重的稀疏性。

Vamshi Ambati等人提出使用用户的历史信息来实现用户的个性化任务推荐，并根据技能和兴趣隐式建模[5]，不仅使输出质量最大化了还使用户完成任务后得到利益得到了最大化。Vishal和Rajan在研究中群智感知平台时都得到一个规律，即用户花在寻找一个符合个人偏好的任务的时间要比完成这个任务所需的时间更多[6]，所以如果政府想通过移动群智感知的方式掌握社会的运行状况或检测突发事件，由较低级的控制转变为较高级的管理，那么信息的汇报系统也就是群智感知中信息汇报的算法需要进一步优化。

国内对于群智的运用更多地在于众包平台的建设，通过群智实现社会化的任务推荐。这些众包平台主要关注那些能够由工人独立完成的宏观任务，这需要工人具有相关的知识技能。相比于国外的细粒度的任务，这种众包平台的用户不仅仅出于兴趣选择一个任务，还需要考量知识和技能水平。这种模式下，任务发送方也就是雇主，会从众多提交的任务结果中选取一个或多个满意的方案，所以这个过程涉及用户兴趣和能力。

仲秋雁提出考虑工人兴趣和能力的推荐算法，建立融合了KSAO胜任力的工人兴趣模型，通过计算用户之间的相似度选取近邻集，从而产生top-N推荐列表[7]。她将兴趣和能力组合起来描绘用户，具有现实意义并且能够提高推荐任务的完成概率。该方法能够较好地解决信息过载并且突出个性化推荐。

邱丹逸提出在难度（技能）、报酬、时间、竞争水平基础上建立任务模型和用户模型，根据任务-用户的匹配算法，从任务池中选取top-N匹配对最高的任务推荐给目标用户[8]。她在用户建模的时候考虑了更多的信息，使用户画像更加地完整并且接近现实。同时，她还提出利用TF-IDF算法运用到任务难度的计算上，增加了任务描述的文本处理步骤，使用户和任务之间的关系多了一种量化方法。

据此，本文确立将以缓解移动群智感知平台的用户所面临的信息过载问题为目的，并在用户建模的阶段根据现实情况把对推荐有较大意义的用户的兴趣信息和用户能力考虑在内，以提高任务推荐的准确性。

### 1.2.2 推荐系统冷启动研究

推荐系统中应用最广泛的非协同过滤算法莫属，但随着运用场景的不断扩大，协同过滤推荐算法面临着一些挑战。其中尚未得到有效解决的一个关键问题就是冷启动问题。冷启动问题的研究在推荐系统的研发中处于不可或缺的地位，对于任务推荐系统的冷启动问题还缺乏成熟的解决方法。

现有协同过滤算法主要基于分析与计算评分矩阵，找到与目标用户兴趣相似的近邻用户，根据这些最近邻的建议得到推荐。所以面对没有评分信息的新用户、新项目，系统无法对其进行有效的推荐。李改和李磊在研究中提出基于K近邻属性的特征映射算法获得新用户和新项目的特征向量[9]，解决了基于矩阵分解的协同过滤算法所面临的冷启动问题。高亨德等人在文献[10]中提出从文本入手为用户建立主题、挖掘兴趣并进行个性化推荐，在此基础上融入信任关系机制实现对新用户的特征建立。他的方法适合用来解决新浪微博、推特或其他社交媒体的用户冷启动问题。罗喜军等人在研究中指出基于类别的推荐[11]，核心思想是构造用户群体和资源类别之间的对应关系。从而对于新用户，根据他的属性将用户分到某一群体，对于新项目利用贝叶斯分类算法，把它归为某一分类。进而根据用户群体和项目类别的对应关系，进行冷启动推荐。孙冬婷针对协同过滤算法建立了一个解决冷启动问题的算法框架[12]，利用用户-项目的评分矩阵对项目和用户分别进行聚类，然后建立分类模型对新用户或新项目进行分类，最后在协同过滤的基础上进行个性化推荐。Sedhain Suvash等人在ACM上发表的论文[13]提出通过人口统计学数据和社交网络信息建立广义矩阵代数框架解决冷启动。

通过研究国内外对冷启动问题的解决方法，本文在学习他们的思想方法后提出一种较为简便的方法针对任务推荐算法冷启动的情况，即将用户和任务分别按职业标签和任务属性进行分类，然后通过类似于协同过滤的思想为新用户推荐任务或者为新任务匹配用户。具体的解决方案将在本文第三章进行介绍。

## 1.3 论文研究内容及预期结果

根据上一节对国内外推荐算法和冷启动研究现状的分析，可以知道国内外学者在任务推荐和推荐系统方面进行过大量的研究分析，但是传统电商网站的推荐算法并不能适应群智感知平台的特征。本文为了解决移动群智感知平台信息过载的问题，在协同过滤算法思想的基础上提出一种修改过的协同过滤算法以适应移动群智感知平台。算法在考虑用户兴趣衰减情况下建立用户兴趣模型，并结合用户胜任力模型组成综合的用户模型。然后依据相似度计算公式建立目标用户和平台用户之间的相似关系，获得与目标用户相似度最高的近邻集，然后预测目标用户从近邻用户的任务池中选择任务的概率，最后给将任务按概率降序推荐给用户。此外，针对任务推荐框架所面临的任务冷启动和用户冷启动问题，本文设计利用用户职业标签为新用户进行冷启动推荐，利用用户-任务匹配方法解决任务冷启动。因此总结起来本文重点针对移动群智感知平台的任务推荐算法，从以下两个方面进行研究：

一，提出针对群智感知任务推荐平台特征的任务推荐算法。算法综合考虑用户的兴趣和工作胜任力，为用户推荐符合其兴趣且在能力范围内的任务。其中用户兴趣模型引入了衰减因子，可以对用户兴趣的转变做出反馈。

二，针对推荐算法的用户冷启动问题，为用户引入职业标签并快速地为新用户产生推荐。针对任务冷启动问题，提出基于分类的匹配算法让任务的分类和用户的兴趣进行较为直接的匹配。对于匹配成功的新任务将其推荐给用户。

算法在建立模型时用到两个参数，参数在0到1之间的调整可以使推荐考虑到不同程度的兴趣衰减和能力融合。为了使推荐效果达到框架下较高的水准，也为了遵循了实验的严谨性，这两个参数将在第四章通过实验来确定。实验通过计算召回率、准确率的综合指标F1值，使用计算结果出于0和1之间的F1值来代表任务推荐的效果。最终结果显示该算法效果比传统匹配用户和任务的方法高出百分之八左右。

## 1.4 论文后续组织结构

论文总共分为四个章节，第一章为绪论，文章后续的组织结构如下：

第二章，考虑兴趣衰减的任务推荐算法的设计。利用用户的投标行为建立用户的兴趣模型和能力模型，其中兴趣模型着重考虑了兴趣衰减的情况，并且在兴趣和能力融合的时候引入融合系数。在用户模型基础上，通过基于用户的协同过滤算法计算用户之间相似度，生成近邻集K。然后介绍推荐生成的过程，即预测指定用户选择投标他近邻用户投标过的任务的概率。

第三章，任务推荐冷启动研究。介绍冷启动问题产生的原因，阐述解决两种冷启动问题的方法，即引入用户标签和基于内容的推荐算法。然后在现有算法框架的基础上实现冷启动解决模块。

第四章，任务分配算法的实验分析。根据国内一家众包网站的数据创建任务和用户，通过准确率和召回率的综合指标来评测推荐算法的精度，量化该推荐系统的实际效果。通过参数实验获得框架下最佳推荐效果。而后在web端进行各模块的测试，以展示成果。

最后，是本文的结论。总结了本文的工作内容和算法的创新点，并对该领域的未来进行了展望。

# 第2章 考虑兴趣衰减的任务推荐算法设计

由于移动群智感知平台面临信息严重过载的问题，为了让推荐系统准确推荐符合用户条件的任务，我们需要在算法设计阶段深入了解用户，然后结合平台的特征为平台设计一个推荐算法自动为用户推荐任务。目前，个性化推荐算法主要分为基于规则的推荐算法、基于内容的推荐算法、协同过滤算法以及混合式推荐算法。

## 2.1 推荐算法框架

移动群智感知模式就是汇集群体的智慧进行任务的解决或者发现创新，这也是开源精神的体现。近年来任务推荐平台发展迅速，这需要考虑这种模式兴起的内在驱动力，才能更好地利用这种模式来提高生产力。

本文将任务推荐算法的运用环境设定为移动群智感知平台，该平台的前身是众包平台。众包模式可分为四类：集体智能、大众创造、大众投票和大众集资[8]。移动群智感知模式就是来源于集体智能该分类。该平台的主要功能如图2.1。平台雇主可以发布任务，从工人提交的任务结果中选择满意的结果并决定给用户的悬赏金额。工人用户在平台上根据自己情况选择任务，提交任务的最终结果。移动群智感知的商业模式体现的是集体智能，本文研究对于系统来说怎么推荐才能让用户最大发挥他们的兴趣和才能。



图2.1 移动群智感知平台

本文的推荐算法是以协同过滤算法为基础。协同过滤算法的原理是根据用户对项目的评分矩阵来获得用户的兴趣分布，然后利用相似度计算找到相似的用户，通过研究相似用户的行为为目标用户生成推荐。该算法的思想可以在生活中找到依据，即兴趣爱好相似的人会成为朋友，朋友之间会相互分享自己接触过的有兴趣的事物或话题。举个例子，和你很像的一个朋友喜欢电影《复仇者联盟》，那么系统就会把这个电影推荐给你，这是最简单的基于用户的协同过滤算法,如图2.2。

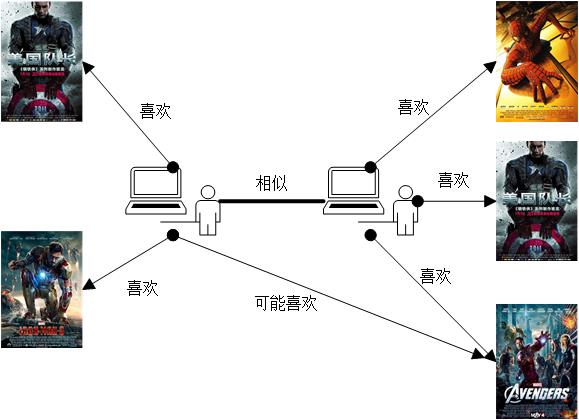


图2.2 基于用户的协同过滤算法示意图

还有一种是基于项目的协同过滤算法，它是亚马逊书城的经典推荐算法。原理是，电影《美国队长》和用户看过并喜欢的《钢铁侠》有相似的关系，那么系统就会给用户推荐他没看过的《美国队长》，如图2.3所示。该算法使用用户对项目的偏好来挖掘项目之间的相似度 [14]。

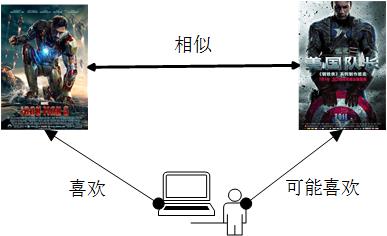


图2.3 基于内容的协同过滤算法示意图

这两种方法都是将用户的所有数据读入到内存中然后进行运算的，因此又被称为Memory-based Collaborative Filtering。基于记忆的协同过滤非常依赖相似性度量来匹配相似的用户或项目，如果有一矩阵其中每一行代表一个用户，每一列代表一个物品，则基于记忆的方法对这一矩阵的行或列使用相似性度量来获得一个相似度值[15]，但是当数据量很大的时候这个矩阵就显现出稀疏性。为了应对矩阵稀疏性问题，产生了基于模型的协同过滤算法，它利用用户对项目的评分学习出一种有效的模型，尝试量化用户没有遇到过的项目，也就是通过学习填充原本稀疏的矩阵，有很多机器学习方法可以用于建立模型，比如，贝叶斯网、聚类、分类、回归、矩阵分解、受限玻尔兹曼机等。

总之，协同过滤算法的优势在于只需要用户数据以及很小的工程量就能得到不错的结果，当然也有限制。比如，CF趋向于推荐比较流行的商品，很少会考虑到冷门的商品，这个问题可以通过基于内容的推荐来解决；另一个重要的问题就是“冷启动问题”，比如新注册用户问题(关于用户的数据较少甚至没有，可以通过基于流行度的推荐解决)，新商品问题(关于商品的数据较少甚至没有，可以通过基于内容的推荐解决)。本文推荐算法基于协同过滤，结合了移动群智感知运用环境的特征设计了推荐框架，如图2.4，



图2.4 任务推荐算法框架

根据上图的框架可知，该算法大致分为三个基本步骤：工人用户模型建立、工人间相似度计算和生成推荐结果。

第一步，建立工人模型。工人用户选择任务时会受自身兴趣爱好的影响。传统方法采用用户-任务意向矩阵来描述用户兴趣，当任务数据规模不断增大时，矩阵呈现出严重的稀疏性，这将占用大量内存导致算法效率底下。本文借用TF-IDF的思想，将用户兴趣点类比文章中的词语，将其投入到整体用户的兴趣中考量。对于文章的词语来说，TF-IDF可以过滤高频的词汇，给低频的词语赋予较高的权重，于是类比到本文的兴趣刻画上就将热度高的兴趣点给予较低的权重，而让算法更注重用户独特的、小众的兴趣。本文考虑用户兴趣衰减，在计算用户在某兴趣分量上的偏爱程度时，与相对陈旧的投标记录相比，让近期的投标记录产生更大的影响。类似地在众包商业模式下设计推荐算法也常常考虑建立用户模型，邱丹逸提出在难度（技能）、报酬、时间、竞争水平这些维度的基础上分别建立任务模型和用户模型，根据任务-用户匹配算法从任务池中选取top-N匹配对最高的任务推荐给目标用户[8]。本文参考邱丹逸的建模思想，将难度、报酬竞争水平作为用户胜任力模型的属性，为用户建立基于胜任力和兴趣的综合模型。

第二步，计算工人间相似度。推荐系统领域内相似度的计算有很多种方法，如使用欧式距离、皮尔森相关系数或杰卡德系数。使用余弦相似度计算用户的兴趣相似度是比较方便的一种选择，同时减少雇主的悬赏差异导致的误差。该相似度由投标任务的分类属性决定，然后在同样的方法下计算用户的能力相似度，通过实验选取线性融合的系数，最终融合得到工人综合相似度。

第三步，生成推荐。通过排序将相似度最高的k位用户组合成为近邻集K，然后获取近邻用户的历史任务池，根据公式计算用户选择任务池中每个任务的概率。最后根据概率高低为用户以top-N的形式依次推荐。

## 2.2 任务模型

群智感知平台的后端会有成千上万个结构类似的任务，所以通过将所有任务都按一个统一的模型来描述，在实现的时候再将具体的信息赋予这个模型。在抽象描述一类物体时，我们需要思考这个物体有哪些属性会被我们使用到。比如商城把水瓶归为一类处理时，买家通过水瓶的容积、保温性以及外观来衡量每个不同水瓶。鉴于任务的固有属性包含四类，如图2.5分别为任务id、任务名、任务分类和任务描述。任务id用于遍历和查找任务时的关键字，任务名和任务描述属于用户在浏览器上与平台互动时所必需的文字描述信息，任务分类是本文设计的算法需要经常使用到的一个关键信息。任务模型文件中提供返回一级分类、二级分类、任务名和任务描述的接口。

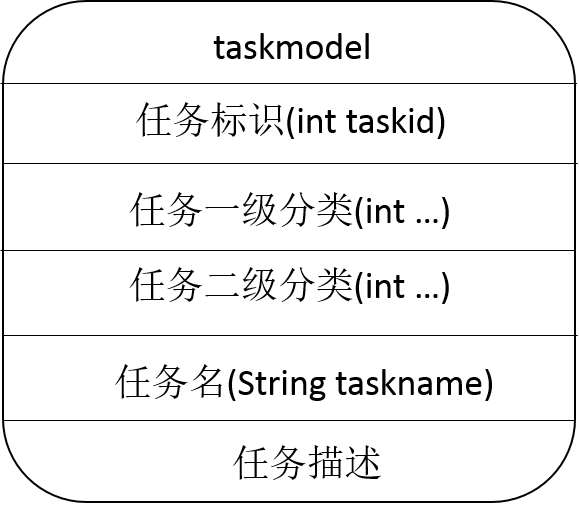


图2.5 任务模型

在设计之初，将平台的所有任务划分为五个大类，虽不能涵盖现实生活中的所有工作类别，但是选择的五个类别都可以通过群体智慧的方式使任务更好地完成。而在现实生活中，有些工作就不适合通过群智来完成，比如其中任务的一级分类情况如下表2.1：

表2.1 一级分类情况

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 一级分类 | | | | |
| 品牌设计 | 营销推广 | IT | 装修 | 工程设计 |

每个一级分类都有若干个属性值，属性值将分类进行细化。因为现实生活中，一个行业内有多张具体的工作，有些工作之间联系比较少。本文设计的二级分类如下5个表：

表2.2 品牌设计的二级分类情况

|  |  |  |  |
| --- | --- | --- | --- |
| 品牌设计 | | | |
| 品牌建设咨询 | 海报设计 | 包装设计 | 品牌取名 |

表2.3 营销推广的二级分类情况

|  |  |  |  |
| --- | --- | --- | --- |
| 营销推广 | | | |
| 微信代运营 | 微博推广 | 抖音营销 | 小红书测评 |

表2.4 IT的二级分类情况

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| IT | | | | | |
| 网站开发 | 微信小程序 | 软件开发 | 数据处理 | 软件测试 | UI设计 |

表2.5 装修的二级分类情况

|  |  |  |  |
| --- | --- | --- | --- |
| 装修 | | | |
| 店铺装修 | 办公室装修 | 酒店空间装修 | 家庭装修 |

表2.6 工程设计的二级分类

|  |  |  |
| --- | --- | --- |
| 工程设计 | | |
| 工程效果图 | 旅游景区设计 | 商业景观设计 |

## 2.3 用户模型建立

### 2.3.1 基于兴趣衰减的用户模型

在传统电商推荐系统中，将用户对商品的点击或购买视为对商品的兴趣。不同的动作可能代表不同的权重。而在移动群智感知平台上，用户对任务的操作主要是投标，因此统计用户的投标情况才能获得用户兴趣分布。本文将用户对任务的投标行为作为对该任务的兴趣量化，一次投标使兴趣数值增加1。同时，将用户对不同任务的兴趣度视为对任务不同属性值的兴趣度。设计时将平台上所有任务分为m+1个分类，按0到m将每个分类进行编号。计算用户兴趣模型时，根据投标列表统计出工人在任务分类的属性值上的兴趣次数，兴趣向量具有m+1个维度，且每个分类有若干个属性值，也就是二级分类：

|  |  |
| --- | --- |
|  | (2-1) |
|  | (2-2) |

根据用户投标列表的内容，可以得到一个代表用户投标分布的用户模型文件中设置了用户类包含的基本数据成员和成员函数，包括用户id和自定义的用户投标记录类。该文件主要维护用户id与用户投标记录相联系，并且由投标记录引申出投标的具体结果，例如报酬和获赞数。文件的结构如下图2.6：

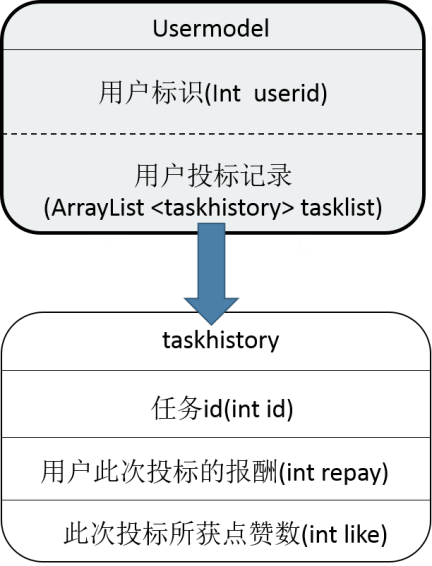


图2.6 用户模型文件的结构

建立用户兴趣模型时通过对历史记录列表进行一轮扫描，得到用户对各任务分类下的兴趣值，根据兴趣的热门程度利用TF-IDF技术强化兴趣偏好受到平台整体投标情况的影响。当然，本文假设平台用户不存在作弊行为，不会恶意刷任务热度。

在视图层，分类和属性值都是具有现实意义的字符串；在数据层，分类和属性值是自然数，这是根据枚举思想而设定的。从用户历史任务中，具有属性值Savij的任务数量越多代表了用户对该属性值的兴趣越浓厚，Sai描绘了用户在i分类上的兴趣分布，m+1个向量组合在一起就形成了初步的兴趣模型，即向量Sa描绘了用户在各个分类上的兴趣分布。具体计算时会利用关键词提取技术TF-IDF来处理属性值的重要程度，某一任务分类中的高频属性值同时在平台用户历史记录中呈现低频，则会产生高权重的TF-IDF，其公式如下：

(2-3)

(2-4)

其中，TFij表示某用户兴趣j在属性i上出现的频率，值越接近1说明任务属性分类为<i,j>的任务越符合用户兴趣；在文本关键词权重计算中，IDF指逆向文件频率，对于某个词(term)来说，其IDF值由总文件数除以包含该term的文件数目，再将得到的商取对数，本文中的IDFj表示兴趣独特性，与平台的用户数量和有兴趣j的用户的比值有关，User表示用户总数，Userj是具有j偏好的用户数量，结果越大代表独特性越高。根据TF-IDF方法，用户u在任务分类i的j属性值上的兴趣程度表示为：

|  |  |
| --- | --- |
|  | (2-5) |

所以用户兴趣偏好在整体任务上的分布可以用向量表示为：

|  |  |
| --- | --- |
|  | (2-6) |
|  | (2-7) |

因此使用TF-IDF处理过后的基于用户兴趣的用户模型可表示为：

|  |  |
| --- | --- |
|  | (2-8) |

式子中的uid是用户唯一标识。至此，考虑用户兴趣的用户模型就建立完成。接下来将介绍，在用户兴趣模型中引入衰减因子，使模型能够体现用户兴趣的衰减及转变。

由于用户兴趣随时间变化，越早的任务对用户当前的兴趣影响越弱[15]，王林等人以遗忘曲线为基础建立了兴趣的指数衰减曲线，并运用到传统的推荐算法中[16]。本文将新旧历史记录合并时对久远的历史记录进行线性衰减考虑。在算法设计中用户历史记录是根据时间先后不断增长的，所以历史记录是有时间顺序的，将后半段历史记录视为最近的投标记录，用其计算得到的兴趣偏好模型Interestur来描述用户最近的兴趣。将先前计算得到的完整兴趣模型与最近兴趣模型融合成为综合兴趣模型，基于线性合并公式(2-9)：

|  |  |
| --- | --- |
|  | (2-9) |

其中，λ的数值将通过实验来确定。由此我们获得了考虑兴趣衰减的用户兴趣模型:

|  |  |
| --- | --- |
|  | (2-10) |

### 2.3.2 融入胜任力的用户模型

通过第二章的分析，用户选择任务时会重点考虑自己的能力是否能够胜任这个任务。对不同的用户来说，他们自身具备不同程度的知识、技能、业务能力，并且深度也是不同。仲秋雁曾在研究众包任务推荐方式时提出构建工人KSAO能力集并融入到工人模型中，使胜任力理论的运用更为丰富[7]。所以本文融入KSP胜任力模型全面地刻画用户知识(knowledge)、技能(skill)、业务能力(profession)的特征，并遵循算法所必备的科学性、完整性和可操纵性的原则[17]，选择三种类型的数据来量化描述胜任力模型的三个分量。其具体内容如下：

(1)知识特征(K)：移动群智感知商业平台上发布有各类任务，面向拥有不同知识结构和水平的工人用户。例如有的工人专注于前段开发领域，但也有工人同时掌握数据挖掘和算法这样多领域的知识，本文将用户u投标过某分类下任务的总数代表用户在该分类下的知识储备程度Ki，具体表达式如(2-11)和(2-12)

|  |  |
| --- | --- |
|  | (2-11) |
|  | (2-12) |

其中的Sa是用户u的兴趣分布向量。

(2)技能特征(S):类比知识特征，技能特征反映用户在某分类下的技能水平。用户所获得的报酬能直接体现了技能水平，因为对于单个任务来说难度越高报酬越高，只有熟练掌握了技能才能完成难度高的任务单次获得较高赏金或者多次被雇主打赏累积较高赏金。假设本文的算法是运用于一个任务粒度很细的平台，因为任务步骤具体而又简单，不需要用户特别掌握某项技能，则算法可以不用考虑用户的技能水平。因此在移动群智感知商业模式下用第i个分类下用户获得的报酬总和表示技能水平，即用户在技能维度上的向量表示为：

|  |  |
| --- | --- |
|  | (2-13) |

(3)业务能力特征(P):鉴于移动群智感知商业模式的开放性，工人用户需要有较强的业务能力，能在同他人竞争脱颖而出较好地完成任务才能获得更多的报酬，业务能力是由客户反馈，通常人们用金钱换取等价的服务，只有当服务方具备超乎寻常的真诚、高效、优质的服务能力时，人们内心受到触动才会额外给服务方一个好评，所以本文设定在移动群智感知平台上雇主用户可以查看工人用户提交的方案并通过点赞对方案表达满意，以工人好评数描述其业务能力P，用到的数据就是历史任务类中的成员变量点赞数，向量的形式如下：

|  |  |
| --- | --- |
|  | (2-14) |

本节中K、S、P三个维度的计算的数据来源依赖于用户历史记录这个数据结构。其中知识维度的计算只需扫描历史记录，根据任务id获取任务分类，这个过程和兴趣建模时计算用户投标分布情况很相似，只是精确到一级分类；另外两个维度的计算各自的数据落在一级分类内的总和，以下伪代码表2.7解释了技能维度和业务水平维度的计算过程：

表2.7 技能、业务能力维度计算过程伪代码

|  |
| --- |
| 伪代码名称：技能、业务能力维度计算过程 |
| 输入参数：用户历史记录数组int history[]  输出内容：ArrayList<Integer>形式的S、P向量  计算过程：  for ( int i = 0 ；i < 投标次数 ；i++) {  通过getCate函数获得任务的分类<first，second>；  在S的对应位置中的Integer元素增加本次投标的获得的报酬；  在P的对应位置中的Integer元素增加本次投标的获得的点赞数；}} |

综上，用户胜任力集合包含知识特征、技能特征和业务能力特征三个元素，表示为：

|  |  |
| --- | --- |
|  | (2-15) |

根据3.3.1建立的考虑兴趣衰减的用户兴趣模型UIMnow，再融入用户胜任力集合Competencyu之后的用户模型表示为：

|  |  |
| --- | --- |
|  | (2-16) |

## 2.4 相似度计算

常见的协同过滤思想是通过发掘相似的集合来产生推荐，当用户A需要个性化推荐时，系统首先通过计算找到和他有相似兴趣的其他用户。本文使用基于用户的协同过滤方法，本节介绍在综合用户模型的基础上使用容易实现的余弦相似度公式计算出用户画像之间的相似度。由于余弦相似度公式对向量长度不敏感，可以在一定程度上减弱由雇主主观决定的报酬大小对用户能力产生的影响。

在协同过滤推荐算法中，生成推荐这一步骤之前都是需要一个特定的相似度计算公式，来定量表示项目与用户画像之间的相关程度。常用的方法有欧几里得度量、余弦相似度、皮尔逊相关系数和广义杰卡德相似系数。本文选用易于理解和现实概念相近的余弦相似度公式。

余弦相似度是将个体用向量表示，并在向量空间中计算两向量的夹角余弦值，作为衡量两个个体之间相似程度的数值。余弦相似度的值范围在-1到1之间，越接近1说明两个体越相似。

|  |  |
| --- | --- |
|  | (2-17) |

为了适应本文设计的任务推荐算法，将修改后的相似度计算细分为三个模块，首先是兴趣相似度计算，然后是能力相似度计算，最后将两者线性融合。计算用户P，Q在第i个分类下的兴趣相似度时，具体的余弦相似度公式如下：

|  |  |
| --- | --- |
|  | (2-18) |

所以由m+1个分类共同作用下的兴趣模型，需要循环使用m+1次公式分别计算各分量余弦相似度，最终兴趣相似度的计算公式如下：

|  |  |
| --- | --- |
|  | (2-19) |

通过公式可以发现所有分量的地位是平等的，这样的设置是从任务分类的角度出发，考虑到实际上任务的分类是互相独立的，分类i不会对分类j 产生影响。

类似地，计算能力相似度时，采用余弦相似性分别获得用户P，Q在K，S，P维度上的相似度，然后求平均。以K维度上的相似度计算为例：

|  |  |
| --- | --- |
|  | (2-20) |

在KSP三个维度共同作用下，用户的胜任力相似度表示为：

|  |  |
| --- | --- |
|  | (2-21) |

最终将兴趣相似度和能力相似度以线性组合的方式获得一个综合相似度的值，该值的取值在0到1之间，计算方法如下：

|  |  |
| --- | --- |
|  | (2-22) |

其中θ是一个可调节的融合系数，范围在0和1之间。θ取1时该用户模型就是单纯考虑用户兴趣的模型

## 2.5 推荐生成

根据上一节里的方法，计算登陆用户p与平台其他用户之间的相似度，然后选取相似度最高的前k位用户组成近邻集K，设置一个任务池,提取K集中每个用户的历史记录放入任务池，随后分别计算用户选择任务池中每个任务的概率，用户p选择任务i的概率计算公式如下：

|  |  |
| --- | --- |
|  | (2-23) |

式(2-23)中是近邻用户u对i的选择情况，其值只能为0或1，如果用户u没有选择过任务i，则值为0，否则为1。

整个推荐生成的过程需要使用sort对用户相似度的表进行排序，并找出前k位，然后对这些用户的历史投标记录计算选择概率，过程的伪代码如下：

表2.8 推荐生成的方法

| 伪代码名称：推荐生成方法 |
| --- |
| 输入参数：近邻集K  输出内容：元素为任务id的推荐列表int rec[ ]  计算过程：  1.将近邻集中每个用户的历史记录加到taskPool中；  2.对于taskPool中每个任务，计算目标用户u选择该任务的概率，即计算近邻集中做过该任务的用户p与目标用户u之间的相似度之和simChosen，计算近邻集所有用户和目标用户的相似度之和sim，用simChonse和sim的比值作为选择该任务的概率；  3.按概率大小将任务id加入到推荐列表rec[ ]； |

融合系数的值会对相似度结果产生影响，进而影响推荐生成，所以该值将在第四章通过控制变量实验来确定。

## 2.6 本章小结

本章由移动群智感知平台引出算法的框架结构，并详细介绍了各个模块的设计。突出强调了本文在用户模型的建立上时综合考虑用户的兴趣和能力，同时还考虑到用户兴趣的衰减情况。环环相扣将基于用户的协同过滤算法的流程以文字的形式结合了公式、图片、伪代码等翔实地展示了一遍。但是系统使用单一协同过滤算法的话，会遇到冷启动问题。下一章，将阐述针对本次毕业设计项目如何解决冷启动问题。

# 第3章 任务推荐算法的冷启动问题研究

推荐系统需要根据大量用户行为记录，分析、预测用户的行为和兴趣，所以推荐系统是依赖于用户行为的。实际平台运作时会遇到用户冷启动和任务冷启动问题。用户冷启动是指新注册用户没有行为数据，无法为新用户找到相似邻居，无法预测其兴趣或技能；任务冷启动是指新任务到达平台时，根据基于用户的协同过滤算法的要求，没有用户操作过的任务是不会被推荐给任何用户[19][20]。对于冷启动的相关问题，许多学者提供了多种解决途径。本章将在现有理论的基础上，寻求一个可靠而又简便的方法解决该任务推荐算法中的冷启动问题，使系统变得完整可行。本章将分别对这两个冷启动问题提供可操作的解决方案，以及合理的解释。

## 3.1 用户冷启动

### 3.1.1 用户冷启动问题分析

本文在群智感知的背景中研究推荐算法。传统推荐算法主要分为两种，分别为基于内容的推荐算法和协同过滤推荐算法。基于内容的推荐是根据用户过去喜欢的东西，给用户推荐与其相似的物品。以今日头条为例，该算法为每篇文章设置几个主题词，用主题词组成的特征向量来描述文章，同时利用用户的历史喜好记录描绘出用户的画像即兴趣特征，利用余弦相似度计算公式得出item 和profile之间的相似度，而后产生个性化推荐。协同过滤推荐算法主要通过寻找行为相似的其他用户，假设行为相似的用户爱好也会相近，所以通过计算用户之间的相似度，让最相似的用户为目标用户产生推荐。在不同运用环境中，推荐算法会有不同的推荐目标，会具有环境的特色。所以根据第二章介绍的基于用户的协同过滤算法，新用户没有产生行为就无法和其他用户产生联系，算法也就不会给用户推荐任何任务。

面对用户冷启动问题，现有以下几种解决途径。在微信、QQ和微博主导社交网络的今天，很多小网站都可以使用社交网站账号快速登陆，这个方式一方面免去了用户注册的过程，节约了用户的时间，另一方面网站可以获得用户在社交平台的信息和行为记录，方便建立初步的用户画像，从而减少冷启动带来的影响。刘华锋等人对基于矩阵分解的协同过滤算法冷启动问题进行了总结，介绍浏览融入社交信息的矩阵分解推荐方法[18]。[张栩晨](http://xueshu.baidu.com/s?wd=author%3A%28%E5%BC%A0%E6%A0%A9%E6%99%A8%29%20%E5%A4%8D%E6%97%A6%E5%A4%A7%E5%AD%A6%E8%AE%A1%E7%AE%97%E6%9C%BA%E7%A7%91%E5%AD%A6%E6%8A%80%E6%9C%AF%E5%AD%A6%E9%99%A2%EF%BC%9B&tn=SE_baiduxueshu_c1gjeupa&ie=utf-8&sc_f_para=sc_hilight%3Dperson" \t "_blank)在研究推荐系统冷启动问题时，针对矩阵分解模型增加上下文信息，通过机器学习方法扩展矩阵分解模型[19]。 Lin,J等人研究使用Twitter中的社交信息预测用户的兴趣爱好，为用户推荐APP[20]。但是，利用社交信息这种行为通常会让用户感到隐私泄露的担忧，所以很多时候用户会把不常使用的社交账号来快速登陆各种网站，这就使得系统获得的社交信息并不完整或者说不能真实反映用户的画像。另一种常见的方法是利用人口统计学信息，包括用户的性别、年龄、职业、籍贯和学校等。许多网站在用户注册时需要用户填写这些人口统计学信息，从而给用户大致分类，并根据平台上已有分类的用户记录给新用户进行推荐。这些特征信息可能对预测起到不小的作用，以用户性别为例，用户在选择电视剧时会呈现比较明显的“男女有别”。根据Netflix统计，《衰姐们》和《破产姐妹》是比较受女性欢迎的美剧，而男性观众则更偏爱《黑道家族》和《神盾局特工》。王春才在研究物品冷启动中利用物品的内容信息，通过计算物品相似度，为用户推荐感兴趣的物品[21]。传统用户冷启动解决方法有以下几种：

1、平均值法

平均值法是认为新用户对项目的评分是接近于平台所有用户对该项目评分的平均水平，通过平台均值评分来预测生成新用户的兴趣。这种方法类比到任务推荐算法中就是将所有用户的兴趣值和能力值做个平均，作为新用户的模型。该平均值法在用户兴趣、能力处于整体的平均位置时推荐效果较好，这样的概率很小。平均值的做法无法考虑用户的个人爱好。

2、众数法

众数法将所有用户对某项目的评分中频率最高的一个数来预测新用户对该项目的评分。首先，这种方法无法使用在本文的任务推荐算法上，因为本文的算法没有用户对任务的评分，根据现实用户对任务的评分很难去衡量用户在选择其他任务时的概率。其次，这种方法也存在忽视个性化的问题。

3、信息熵法

系统中评估的元素包含不同的信息，其中一些包含更多信息，另一些包含更少信息。 信息熵类似于学习决策树的选择分类，并且基于信息熵大的项目值作为未评估的新项目的期望值。信息熵方法基本上是一种均值预测方法，但它不是所有项目的平均值。当然这个方法也会有缺陷，首先，含有的信息熵越大并不一定能代表用户对没有评估过的项目的偏好。其次，它需要花费大量的时间来计算用户已评估的项目评分，但最终的结果可能仍然无法准确地预测。

4、利用用户注册信息

对于新注册的用户，我们通常无法了解他们的偏好，只能为他推荐一些受欢迎的产品。此类推荐不是个性化推荐。但是，在许多网站的注册过程中，系统可以充分利用用户注册信息。系统首先确定年龄，性别，民族等信息分类，为用户提供多个标签供用户选择最初定义用户偏好，然后根据这些标签向用户提出建议；最后，系统可以通过登录其他合作伙伴站点的帐号以检索其他网站上的用户行为和社交数据，以便可以将数据用于个性化推荐。在现实生活中，系统可以完全整合上述不同用户的信息，并为用户提供一致的建议。

本文所实现的移动群智感知平台是纯粹的社会化任务推荐网站，在新用户开始阶段没有大量的用户数据，所以在本章下一节将介绍一种方法以解决平台用户冷启动问题。

### 3.1.2 用户冷启动问题的解决方法

本文的解决方案是利用用户注册信息中的职业选择，为用户提供粗粒度的热门推荐。正如大家所知，隔行如隔山，这句话应证了从事不同行业的人之间的差别是相当大的，就以律师和程序员为例，律师可以在自由时间兼职做企业知识产权的顾问，程序员可以利用空闲时间开发小程序，他们很难交换任务因为知识和技能的限制。具体来说，在用户注册页面让用户选择目前的职业pro，然后利用设定的映射函数使该pro标签匹配一个任务一级分类，从属于该一级分类的所有任务中随机挑选出5个任务，作为用户的职业相关推荐。同时考虑到用户兴趣和能力的广泛性，冷启动算法还会从平台任务池中随机选取10个任务作为推荐，方便系统发掘用户潜在兴趣。如果一开始提供给用户的所有推荐都集中在某个分类范围内的话，可能会造成推荐范围逐渐集中，使用户视野变得狭窄。范围内的随机推荐可以提高推荐的准确率，同时随机推荐的计算量小方便实现。用户冷启动算法的伪代码如下表3.1：

表3.1 用户冷启动解决方法伪代码

| 伪代码名称：用户冷启动解决方法 |
| --- |
| 输入参数：用户职业标签int proLabel  输出结果：元素为任务id的推荐列表int rec[ ]  计算过程：  1.匹配用户职业，获得任务分类fcate；  2.从一级分类为fcate的一组任务中随机选择5个任务，加入推荐列表rec[ ]；  3.从不含fcate的任务池中随机选择15个任务，加入推荐列表rec[ ]； |

算法的核心问题在于用户职业到任务分类的映射关系，本文在第二章任务分类的基础上，只是根据作者的常识将几个常见的职业和任务分类联系起来。举个例子，将设计师职位映射到室内设计这个任务分类。本文推荐框架在设计之初制定了任务的分类为五类，所以为了方便解释，冷启动模块的用户职业分为五种。职业标签和任务分类的匹配原则如下表3.2：

表3.2 职业-任务分类匹配

|  |  |  |
| --- | --- | --- |
| 编号 | 职业 | 推荐的任务分类 |
| 0 | 平面设计师 | 品牌设计 |
| 1 | 销售 | 营销推广 |
| 2 | 程序员 | IT |
| 3 | 室内装修 | 装修 |
| 4 | 工业设计师 | 工程设计 |

## 3.2 任务冷启动

任务冷启动跟用户冷启动比较类似，其产生的主要原因是推荐系统采用基于用户的协同过滤算法。该算法要求根据相似用户的投标集合产生推荐，所以新到达平台的任务由于系统找不到用户对该任务的历史行为，将一直无法出现在推荐列表中。

### 3.2.1 任务冷启动问题分析

本文在研究基于内容(content-based)的推荐算法时发现这种算法根据用户过去感兴趣的项目，为用户推荐相似的项目，其中用户的兴趣是由项目的属性特征描述的，利用用户曾经喜欢（或不喜欢）的项目特征向量来学习出用户的画像，推荐时计算项目特征和用户画像直接的相似度，从而将最相似的若干项目推荐给用户[19]。这种算法具有用户独立性的优点，每个用户的画像都是依据个人历史记录建立的，推荐时与其他用户的画像无关，所以面对新项目冷启动问题时，使用基于内容的推荐算法可以绕过项目缺少评分这个不足之处。李龙生等人研究物品冷启动时提出结合用户行为和物品标签的协同过滤算法[22]。这个算法解除了对用户-物品评分矩阵的依赖，只需要根据物品的标签便可进行推荐。

### 3.2.2 任务冷启动问题的解决方法

通常纯粹的基于项目的推荐算法该算法主要分为三步，首先是项目描述，为每个项目抽取一些特征作为区别于别的项目的属性；其次，通过朴素贝叶斯算法学习用户的历史记录，从而建立用户模型；最后，生成推荐，只要把算法预测的用户最可能感兴趣的n个项目经过排序，作为推荐列表返回给用户即可。本文根据移动群智感知平台这个应用限定，提出一个较为方便实现的基于分类的任务匹配算法(Classification-Based Matching algorithm,CBM)，该方法的思想是基于CB(content-based)算法的，其步骤伪代码大致如下：

表格3.3 任务冷启动方法

| 伪代码名称：物品冷启动解决方法 |
| --- |
| 输入参数：用户id，待推荐的任务列表int newlist  输出结果：元素为任务id的推荐列表int rec[ ]  计算过程：  1.根据登陆用户u的兴趣模型，获得用户u的兴趣模型；  2.扫描兴趣模型得出用户u最感兴趣分类，记为favourite< r , s >；  3.用favourite匹配newlist中每一个新项目的cate，匹配成功的加入推荐列表rec[ ]； |

该CBM算法巧妙地利用用户登陆时检测这一个用户的最感兴趣分类，避免了遍历计算平台所有用户的兴趣分类这种耗时的计算。同时也得益于先前用户兴趣模型计算时基于任务分类，使得任务冷启动时使用任务的分类可以容易匹配用户兴趣。以下用流程图更直观地反应CBM算法的具体操作：

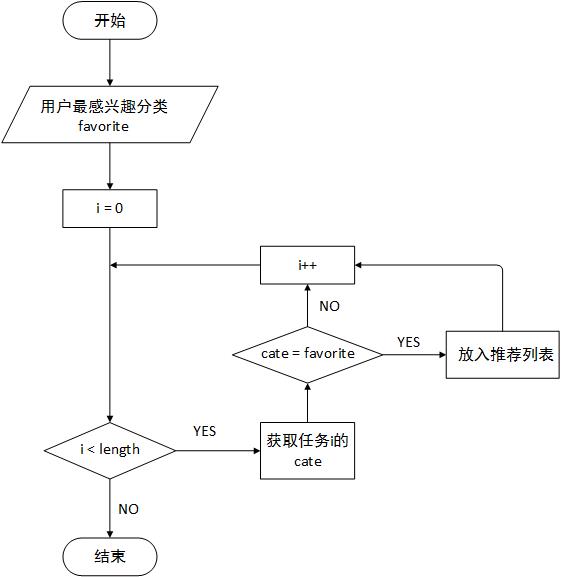


图3.1 任务冷启动流程

## 3.3 本章小结

本章在分析了两种冷启动问题产生原因之后，介绍了现有的一些解决方案，发现它们对于任务推荐系统并不完全适用。因此，引出本章对该推荐框架下冷启动情况的解决方案。应对两种不同的冷启动情况，提出了不同的方案。面对用户冷启动时，根据用户注册时的职业标签，系统在职业对应的任务分类内和所有任务范围内分别进行随机选择。这种方法优于完全随机推荐且算法简单。面对任务冷启动问题，算法从已知的用户模型中提取用户最擅长的分类，与待启动的任务逐个匹配。将匹配成功的任务推荐给登陆用户，解决了任务的冷启动。至此，一个考虑了冷启动问题的任务推荐框架就介绍完毕，下一章将基于实验中具体分析算法数据层的运作方式，并介绍参数选定和推荐效果等细节。

# 第4章 任务推荐算法的实验分析

通过第二章对移动群众感知平台的特征分析和流程介绍，可以知道这个毕设项目的最大特点是考虑了用户兴趣和能力建立用户模型。本章将主要介绍对设计的算法进行的三部分测试实验：第一部分实验为了检测代码各部分可行性，将进行模块测试，并检测算题整体运行情况，进行整体的系统测试；第二部分实验为了优化用户模型，对两处待定参数（兴趣衰减参数和兴趣能力融合参数）的设置进行参数测试，根据推荐系统测评指标分析获得最佳参数取值；第三部分实验为了对比本文设计的算法和其他算法的推荐效果，证明该算法的优势。

## 4.1 实验目的及参数选取

移动群智感知模式曾在中国盛行过，但它对大众的冲击只是处于思维层面，换句话说只是在媒体的主导下冲击了大众的思维。从宏观角度来看，群体的智慧是惊人的。中国政府网发布的文章指出，中国拥有世界第一大市场的广阔空间和巨大的人口人才红利，拥有13多亿人口、大约9亿劳动力、1.7亿多受过高等教育或具有专业技能的人才。如果正确利用了移动群智感知，将激发创新推动社会发展和民族振兴。所以应该深入分析面向移动群智感知的任务推荐算法，研究任务推荐算法的效果。具体来说就是分析兴趣模型、考虑兴趣衰减的兴趣模型和基于知识、技能和业务能力的胜任力模型，通过实验数据分析用户相似度的计算并衡量推荐效果。实验通过编程将算法思路落实成代码，实现构建一个可以登陆后个性化推荐的框架。实验将对网站的功能进行测试和分析。

任务推荐效果的评价指标根据平台的设计特点决定。由于算法不需要用户-任务评分矩阵来进行协同过滤，而是通过用户执行任务的结果来描绘用户画像。因此，本章采用推荐系统评价标准中的准确率（precision）和召回率（recall）来量化本任务推荐项目的推荐效果[23]。其公式分别如下所示：

|  |  |
| --- | --- |
|  | （4-1） |
|  | （4-2） |

其中，n为平台用户总数量，rec为系统为登陆用户自动生成推荐列表，history为用户的实际投标列表。通过这两个公式可以发现这两个指标与推荐列表的长度有关，所以根据Pazzani等人提出的将准确率和召回率综合起来的指标F作为最终评价标准[24]。

综合评价标准公式如下：

|  |  |
| --- | --- |
|  | （4-3） |

## 4.2 实验环境

实验所用硬件为联想Thinkpad E450c笔记本，处理器型号为英特尔酷睿i5-4210U 处理器主频为1.70GHz，RAM内存为8.00 GB，硬盘容量为500GB 7200转。

实验的软件环境：

1. 操作系统：微软window8.1；
2. 编程软件：Eclipse4.7；
3. 编程语言：JAVA语言；
4. 服务器：Apache-Tomcat-8.5.6；
5. JRE版本：1.8.0\_111。

## 4.3 实验结果分析

本次毕业设计项目主要由两部分组成，分别为算法部分和前端网站部分。本节的模块测试是对算法各部分进行测试，以控制台的输出为模块测试结果。算法部分源代码组织结构如图4.1：

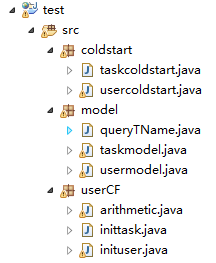


图4.1源代码文件组织结构

由于算法涉及的事物主要为用户和任务，所以在model文件夹下设计了用户模型类和任务模型类。userCF文件夹是算法的主体部分，对用户模型和任务模型进行实例化，然后再arithmetic文件中进行计算。coldstart文件夹包含了任务冷启动和用户冷启动的解决算法。

### 4.3.1 实例化实验

本项目的数据来源于国内猪八戒众包平台的部分用户投标记录和部分的任务，两个模型实例化文件的作用是实例化本项目的任务。在逻辑上，实例化属于算法中模型建立的部分，所以项目将实例化这部分的功能连同相似度计算和推荐产生一起放在同一个package里，具体可以回顾图4.1。

首先是任务实例化。其实际过程是向inittask类维护的一个HashMap数据结构中增加任务数据。由于HashMap可以维护key和value的对应关系，在项目中它维护着任务id和任务分类、任务名和任务描述之间的映射关系。实例化代码如图4.2所示：

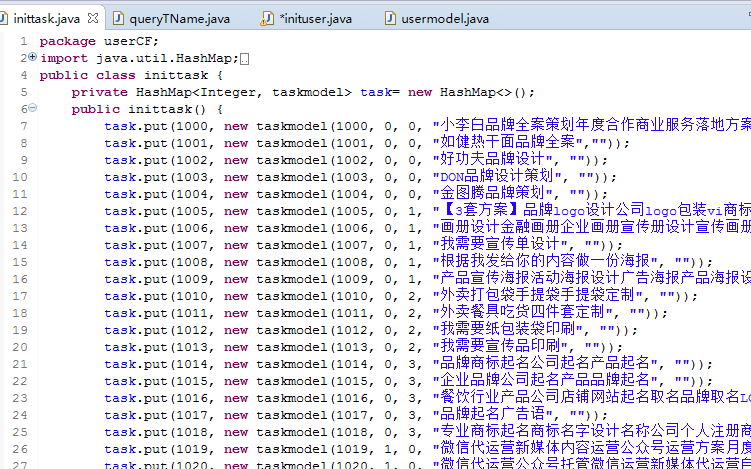


图4.2 任务实例化过程

项目中用户的实例化是在inituser文件中完成，该类中一个核心的成员类型是HashMap结构的chart，它存储了平台所有实例化后的用户。根据第二章的介绍，用户实体是由用户id和用户的历史投标记录组成的。所以在该类中存在一个HashMap结构的表map，用于存储并管理用户的投标记录，同时使用表map的数据循环创建usermodel并添加到chart中，代码如图4.3和图4.4所示。



图4.3 用户实例化函数

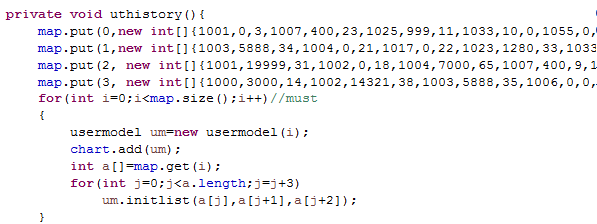


图4.4 用户实例化过程

对于实例化后的用户，就可以根据算法计算用户的兴趣模型、能力模型以及融合之后的综合用户模型。根据第二章介绍，两种模型的计算均以投标的分类< i ，j >有关，所以提取重复的过程设计为一个课重复使用的函数Sa。由于每个一级分类下的二级分类数目不尽相同，因此无法使用矩阵这种数据结构来存储，改用两层嵌套的ArrayList作为数据结构。同时，分类的获取需要调用inittask类中的getName函数。

任务分类统计部分的代码如下图4.5表示：

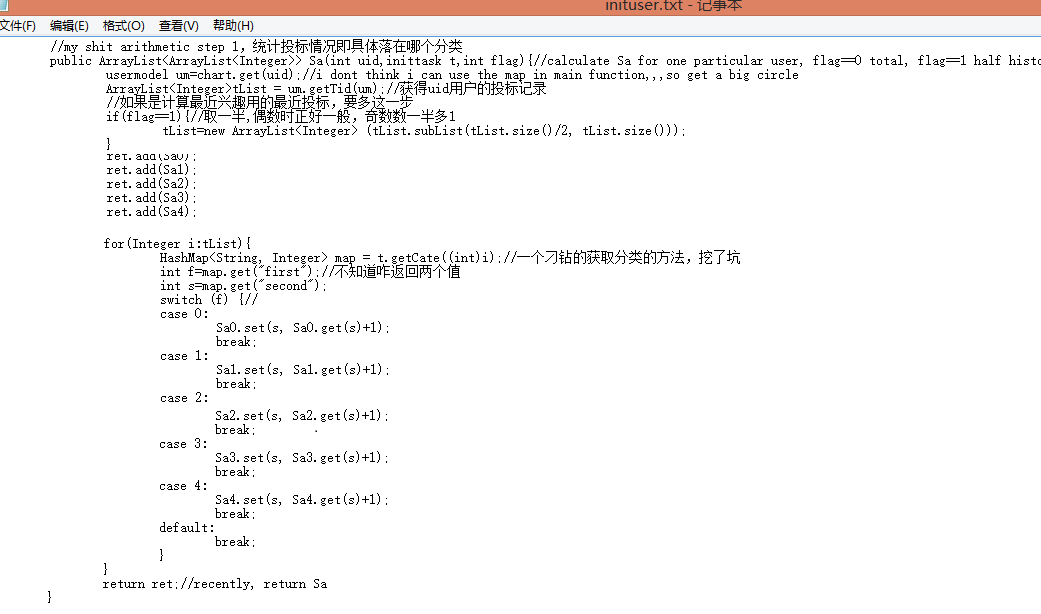


图4.5 任务分类统计

在获得用户投标的分类之后就可以根据公式计算用户的兴趣模型和任务模型。在兴趣衰减融合的函数blendInterest中先设定衰减参数α为0.6，在之后的参数实验中会对参数取不同值计算系统的推荐效果。这里的模块测试就先按照指定参数进行。测试中对20位用户进行了实例化计算，以用户0的计算结果为例如图4.6：

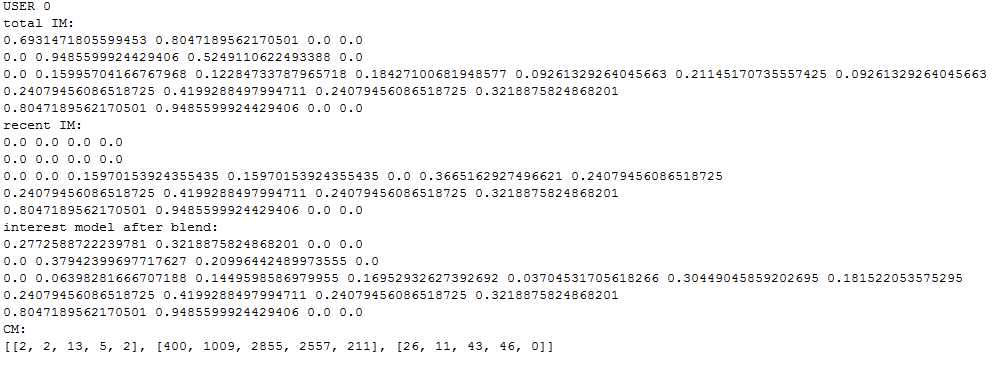


图4.6 用户兴趣和能力模型计算结果

对比total IM和recent IM数值上的变化，可以明显看出兴趣的衰减痕迹。根据用户0的数据结果，其在分类0上的总体兴趣向量为{0.6931，0.8047，0，0}，但由于最近没有对这一分类进行投标，所以该分类上的兴趣向量被衰减至{0.2773，0.32189，0，0}。

### 4.3.2 算法实验

根据第二章介绍的理论，在arithmetic文件中将算法的过程完全实现。在算法文件可以调用用户和任务的建模和实例化，然后根据数据进行计算用户之间的相似度，选取前3的用户作为近邻集。测试对用户0进行，得到平台用户与该用户0之间的相似度，并预测了该用户选择紧邻用户任务的概率，最后按序为该用户推荐任务，算法测试的中间相似度结果和预测结果如图4.7。

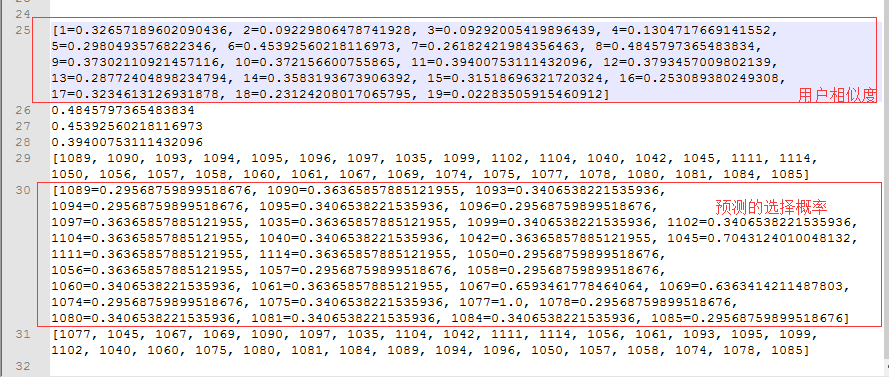


图4.7 算法中间结果

首先算法计算了登陆用户和其他用户之间的综合相似度。可以看出与登陆用户0最相似的用户为8号用户、6号用户和11号用户，目标用户与这三者的相似度分别为0.4845797365483834、0.4539256021811697和0.39400753111432096，和其他用户的相似度都在0.2~0.37甚至低于0.1。所以这三个用户的相似度接近或大于0.4属于这三个用户最为相似。然后，算法将这三个相似用户的任务列表以HashSet形式组织到一起，成为待推荐的任务池。任务池在图4.12中第29行列出。接着，对任务池的每个任务根据2.6节的公式预测目标用户选择它的概率，计算的结果如图4.12第二个红框中所示。最终经过排序，将概率从大到小的顺序把任务的id组织形成推荐列表，实验的推荐列表如图4.12中第31行所示。

## 4.4 推荐效果优化实验

由于项目推荐结果采用top-N列表的形式进行展示，就可以获得推荐列表与用户实际投标列表中重合的数量，从而根据本文使用的准确率和召回率公式，计算综合指标F1来衡量推荐效果。计算公式如(4-1)，(4-2)，(4-3)。

在4.3.2的算法实验中将衰减参数α指定为0.6，融合参数θ指定为0.7的情况下，保留四位小数后系统的推荐效果指标为F=0.3271。在本节，通过控制变量实验重复一百次，每次在α和θ取不同值的情况下计算系统的推荐效果F1值，实验结果保存在excel中，如图4.8所示。

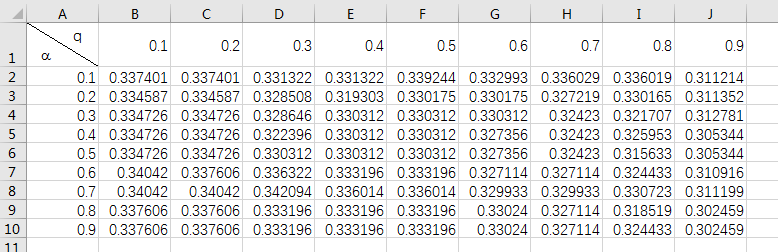


图4.8 α、β不同取值下的系统评价指标值

观察上图中的数据，可得出兴趣衰减参数α在0.6和0.8之间时，系统推荐效果的综合评价指标F11值大于α取其他值的时候产生的结果。因此认为，在该任务推荐框架下，α取值在0.6至0.8之间能使推荐效果达到更优。

为了使数据更加的清晰明了，方便不同参数之间的比较以及对实验结果进行分析，本文使用Matlab软件，将不同α取值下的综合评价值F1绘制成折线图，α的取值分别为0.6、0.7、0.8。这三组数据绘制成的折线图，如图4.9所示：

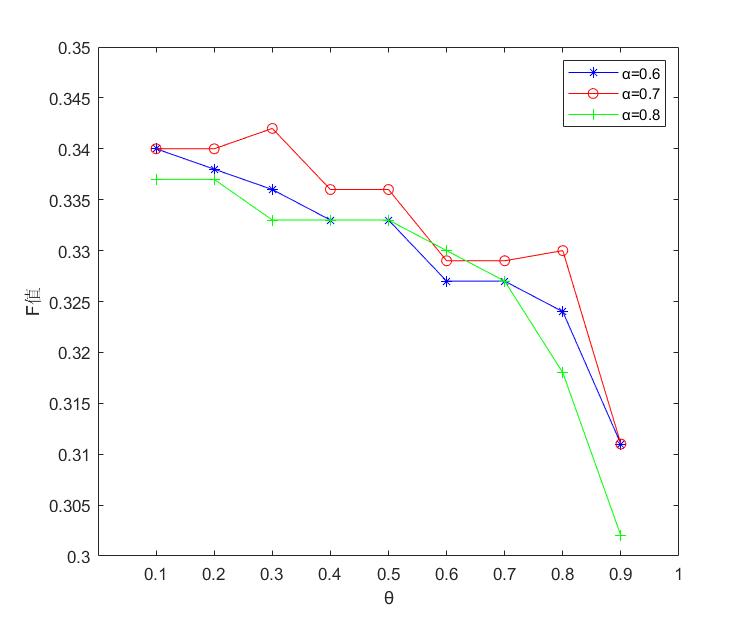


图4.9不同衰减参数下的F1值

观察上图中的三条折线，不难发现，它们具有一个相同的现象：当θ为0.9的时候，F值会有一个明显的断层式的下降。从该现象本文可以分析出，仅考虑用户兴趣的模型而不考虑能力模型的时候，推荐算法会具有较大的不准确性，所以融入用户胜任力的模型对任务推荐算法有比较显著的效果提升作用。

同时本文也注意到一个不足之处，即测试平台的用户较少，系统的推荐效果指标主要分布在0.3025至0.3421之间。推荐效果指标在0.34以上的只有4组数据，所以判定兴趣衰减参数取值为0.7且融合参数取值为0.3时系统达到最佳的推荐效果，其指标为0.3421。

根据准确率的公式，在推荐列表极端长的情况下，准确率将为投标列表长度与推荐列表长度的比值。在极端短的情况下，准确率将为1。同样地在极端情况下，对于召回率公式来说，过长的推荐列表会导致召回率为1，过短的推荐列表会使召回率接近0。因此本文使用综合评价指标F1。在上文两个参数设定的情况下，对不同推荐列表长度，综合指标F的数值如下表：

表4.1 不同推荐列表长度下的F1值

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 推荐列表长度 | 15 | 20 | 25 | 30 |
| F1值 | 0.3421 | 0.3674 | 0.3785 | 0.3731 |

从上面的实验结果可以看出，个性化推荐列表长为25的时候，综合评价指标达到0.3785较推荐列表长度为15或20的时候有明显优势。同时，作为一个显示推荐列表的页面，25个任务足够用户从中做出选择，再多可能就会让用户产生疲累的感觉。所以，综合考虑网页显示情况后，本文将推荐列表中用户选择概率最大的前25个任务作为用户登录网站后显示的任务。

## 4.5 网站实验

本项目使用Tomcat-8.5.6服务器。Apache是一种只能支持静态页面的普通服务器，Tomcat是扩展Apache之后的java服务器。目前，Tomcat服务器是世界上最流行的服务器之一，它具有简单、速度快、性能稳定的特点，而且可以方便地在Eclipse上进行调试开发[25]。所以本项目选用轻量级的Tomcat服务器。

项目的服务器端是用JSP开发的。JSP(Java Server Pages)是一种服务器端的脚本语言，它在HTTP页面中引入了Java的程序段从而成为动态页面技术。用JSP开发的web应用具有较好的移植性，可以在多种操作系统运行。JSP执行的过程如图4.10所示，服务器在页面被客户端请求之后对页面中的Java代码进行处理，再将生成的HTML页面返回给客户端。它的底层功能是由Java Servlet实现的。

图4.10 JSP执行过程

本项目以Eclipse开发软件为依托，使用Tomcat服务器搭建了一个可以实现个性化推荐的简单的动态web网站项目，使用户可以通过较好的UI界面直观感受个性化任务推荐的算法。该项目的文件组织结构如下图4.11。Web后端通过检测用户登录的ID，调用本文设计的任务推荐算法，并为登陆用户返回推荐结果的页面。其中页面引入了Bootstrap工具包，著名的推特社交网站的前端就是使用了Bootstrap。它的引入使前端页面更加优雅简洁大方。

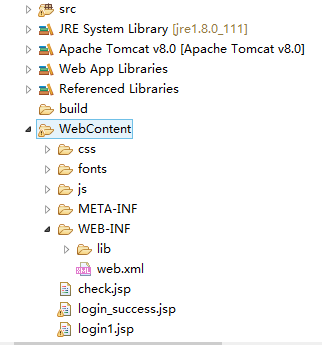


图4.11 web项目文件组织结构

实验的预备工作是配置服务器，通过浏览器访问localhost:8080检验服务器配置的成功与否。接着为项目添加服务器并部署，然后将项目在服务器上运行。访问该应用时通过谷歌浏览器，在地址栏输入localhost:8080/test/login1.jsp就可以访问项目的登陆页面。网页的登陆界面如图4.12所示：

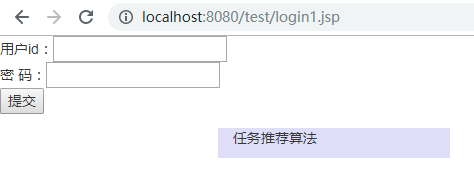


图4.12 用户登录界面

登陆成功后，页面根据登陆的JSP文件中重定向代码的设置自动跳转至login\_success页面。在登陆成功的JSP文件中导入src文件中的arithmatic文件，使得服务器在登陆成功页面执行协同过滤算法，并展示算法自动推荐的任务列表。不同ID的用户有不同的投标历史记录，算法根据用户的模型自动为用户生成任务推荐列表，使用户能在范围广泛且数量成百上千的任务中发现自己爱好并擅长的任务。至此，任务推荐系统的缓解信息过载的问题得到了解决。

接下来，本文作者通过实验，展示算法考虑兴趣衰减的特性。我们将ID为20的用户设置成一个兴趣由“品牌设计”方向转变为“装修”的特殊用户，为他创建相关的投标记录。然后，登陆该ID检验系统是否会对他兴趣的转变有所察觉。根据第二章的设计，算法应当重点考虑用户最近产生的投标记录，突出强调最近的兴趣，在寻找近邻集的时候会选出与用户最近投标相似的用户。为了方便实验观察，让控制台打印ID为20的用户的投标记录列表。该投标列表是按时间顺序进行安排，先做的任务序号较小，所以会先打印出来。该用户的投标记录如下图4.13所示：

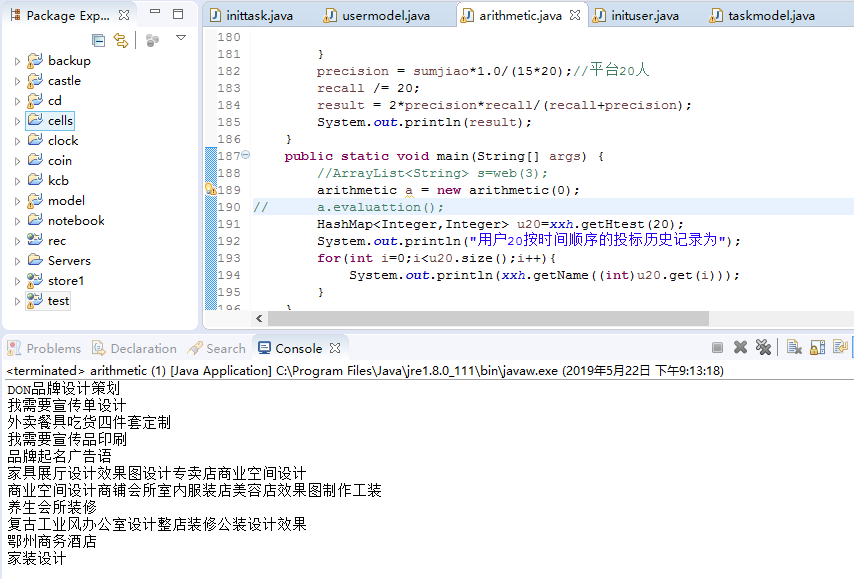


图4.13 用户20的投标记录

我们发现他的历史记录中前一部分主要是品牌设计的任务，后一部分主要是装修方面的任务，也就是说这是一个具有明显兴趣衰减特征的用户。接下来登陆该账户，得到的推荐列表如下图4.14所示：



图4.14 用户20的推荐列表

由上图系统为用户20自动推荐的任务来看，系统主要为该用户推荐了装修类的任务组成，可以判断是算法的兴趣衰减模型产生了效果。因此该任务推荐算法具有识别用户兴趣衰减的特性，能够很好地产生符合用户的心意推荐。

## 4.6 本章小结

本章基于第二章设计的算法，对算法的各个步骤以Java代码的方式实现，对算法各部分的可行性和整体推荐算法的可行性进行了实验，并证实了算法可以实现任务的个性化推荐。进而，对算法的推荐效果加以量化考虑，使用了召回率、准确率以及能够综合评价的F1值，使得该算法的推荐效果以数字的方式直观、清晰地展现。接着，对第二章提出的两个参数，分别为衰减参数α和融合参数θ，进行了不同取值。目的是找出能够使推荐算法呈现最佳效果的参数组合。通过实验，得到α、θ分别取0.7和0.3时达到最优推荐效果。同样地，通过观察不同推荐列表长度下的F1值，确定了web前端显示的任务数量为25。最后，我们将项目以动态网站的形式在服务器上运行，展示了算法在网站上的运用。

# 结 论

随着海量信息压倒性地进入当今时代，信息过载的问题在互联网领域显得日益严峻了起来。如果有效解决信息过载问题，将给人们的工作、学习、生活带来效率上的提高，从而激发灵感的产生以及技术问题的解决。移动群智感知作为一种新兴的合作型商业模式，在产生之际就吸引了研发、设计人员的参与。这种开源的理念促进了合作。在现有的任务推荐平台上每天都有新的任务发布出来，这就导致了用户要花更多的时间在找到一个合适的任务上。为了使平台更好地起到服务作用，本文提出在平台上使用推荐算法，根据用户的投标记录发掘用户的兴趣和能力水平，为用户自动地提供符合其属性的任务，以此提高用户的工作效率和工作积极性。

本文的工作主要在以下两个方面：

（1）根据移动群智感知平台的特征，提出使用考虑用户兴趣衰减的兴趣模型和用户能力模型，从两个方面描绘用户的特征，并通过基于用户的协同过滤算法为用户寻找相似用户的集合。进而算法从相似用户的历史投标记录中选择用户最可能产生投标行为的一组任务作为推荐列表。

（2）针对该推荐算法的用户冷启动问题和任务冷启动问题，分别提出了解决的途径。面对用户冷启动问题时，算法给新用户职业选项，然后根据用户提供的职业给用户推荐相应分类下的任务，体现术业有专攻的思想。面对任务冷启动问题时，算法将可以和用户兴趣想匹配的任务推荐给用户。

由于群智感知的概念提出不久，成熟的商业化应用还是比较少，并且主要是社会化推荐这个方面的应用。针对群智感知的任务推荐算法和常见的电商或音乐类的推荐算法有所不同前者类似于对企业的员工推荐任务未来对面向群智感知的任务推荐可以从以下角度再深入研究：

（1）对用户胜任力建模时虽然在一定程度上能够发挥用户的专长，但是任务报酬这个数值是由雇主确定的，完成类似的任务有的雇主提供的报酬会偏高，所以会存在一定的偏差。在今后的研究中应当考虑到这方面的误差，寻找一种更精确描述用户技能水平的方法。

（2）在应对任务冷启动问题时，冷启动推荐算法在用户登录之后才进行。这是一种在线计算的机制，今后应当对离线计算模式深入研究，对推荐算法框架进行优化，寻求算法的更高效率。

# 参考文献

1. Lazer D , Pentland A , Adamic L , et al. Social science. Computational social science.[J]. Science, 2009, 323(5915):721-3.
2. 郭斌, 翟书颖, 於志文, et al. 群智大数据:感知、优选与理解[J]. 大数据, 2017(05):60-72.
3. Smith B , Linden G . Two Decades of Recommender Systems at Amazon.com[M]. IEEE Educational Activities Department, 2017.
4. Kurup A R, Sajeev G P. Task recommendation in reward-based crowdsourcing systems[C]// International Conference on Advances in Computing. 2017.
5. Ambati V , Vogel S , Carbonell J G . Towards Task Recommendation in Micro-Task Markets[C]// Human Computation, Papers from the 2011 AAAI Workshop, San Francisco, California, USA, August 8, 2011. AAAI Press, 2011.
6. Vishal C , Shivnesh V. R , Romil Kumar V , et al. A crowdsourcing-based platform for better governance[M]// Proceedings of the International Conference on Soft Computing Systems. Springer India, 2016.
7. 仲秋雁, 张媛, 李晨, et al. 考虑用户兴趣和能力的众包任务推荐方法[J]. 系统工程理论与实践, 2017, 37(12):3270-3280.
8. 邱丹逸. 众包模式下产品设计任务推荐[J]. 机械设计, 2017(12):48-52.
9. 李改, 李磊. 一种解决协同过滤系统冷启动问题的新算法[J]. 山东大学学报(工学版), 2012, 42(2):11-17.
10. 高亨德, 王智强, 李茹. 基于信任关系和词相关关系的冷启动用户词特征重建[J]. 中文信息学报, 2018, v.32(05):85-93+101.
11. 罗喜军, 王韬丞, 杜小勇, et al. 基于类别的推荐——一种解决协同推荐中冷启动问题的方法[C]// 中国数据库学术会议. 2007.
12. 孙冬婷. 协同过滤推荐系统中的冷启动问题研究[D]. 国防科学技术大学.
13. Sedhain S , Sanner S , Braziunas D , et al. [ACM Press the 8th ACM Conference - Foster City, Silicon Valley, California, USA (2014.10.06-2014.10.10)] Proceedings of the 8th ACM Conference on Recommender systems - RecSys \"14 - Social collaborative filtering for cold-start recommendations[J]. 2014:345-348.
14. 牛亚男. 教学资源个性化推荐系统的设计和实现[D]. 大连理工大学, 2013.
15. 常晓雨, 余正生. 引入时间衰减项的兴趣点推荐算法[J]. 杭州电子科技大学学报, 2016, 36(3):42-46.
16. 王林, 刘继源, 马安进. 基于兴趣衰减的个性化排序算法[J]. 计算机工程, 2017(09):220-225+233.
17. 吕英杰, 张朋柱, 刘景方. 众包模式中面向创新任务的知识型人才选择[J]. 系统管理学报, 2013, 22(1):60-66.
18. 刘华锋, 景丽萍, 于剑. 融合社交信息的矩阵分解推荐方法研究综述[J]. 软件学报, 2018.
19. 刘畅, 王玉龙. 推荐系统冷启动问题分析[J]. 电信网技术, 2017(1):65-68.
20. 杨博, 赵鹏飞. 推荐算法综述[J]. 山西大学学报(自然科学版), 2011, 34(3):337-350.
21. 张栩晨. 利用Tri-training算法解决推荐系统冷启动问题[J]. 计算机科学, 2016(12):115-121.
22. Lin J, Sugiyama K, Kan M Y, et al. Addressing cold-start in app recommendation:latent user models constructed from twitter followers[C]// International Acm Sigir Conference on Research & Development in Information Retrieval. 2013.
23. 王春才, 邢晖, 李英韬. 个性化推荐系统冷启动问题研究[J]. 现代计算机（专业版）, 2015(19).
24. 李龙生, 艾均, 苏湛等. 结合用户行为和物品标签的协同过滤推荐算法[J]. 计算机应用与软件, 2018, v.35(06):254-259.
25. 刘建国, 周涛, 郭强, 等. 个性化推荐系统评价方法综述[J]. 复杂系统与复杂性科学, 2009, 6(3).
26. Pazzani M , Billsus D . Learning and Revising User Profiles: The Identification of Interesting Web Sites[J]. Machine Learning, 1997, 27(3):313-331.
27. 胡莉萍. Tomcat+JSP+MySQL整合配置初探[J]. 中国科技信息, 2010(5):102-103.

# 攻读学士学位期间取得的学术成果

# 致 谢

感谢XXX对我毕设的指导和帮助，感谢团队里的每个导师，是他们的负责才让我从毕业设计中学到很多。XXX和蔼可亲，平易近人，严谨治学。他在学术上已将取得了巨大的成就，依然保持着谦逊的态度。毕设开始的阶段我怀着忐忑不安的心情学习基础的知识，后来经过XXX及XXX多次的指导我才逐渐明晰了思路。当我在写论文的阶段遇到前所未有的低谷时，XXX利用个人时间对我论文的行文思路提供了详细的指导。

感谢父母从小给我提供良好的家庭环境，从小教育我做人要诚信。这是我大学四年坚持诚信考试的原因，这也使我有足够的信心说出这份大学最后的考试毕业设计不含虚假部分。感谢父母当年没有过分阻止我选择哈工程，在我选择留学的时候也表示支持，感谢他们在我在小学以后就给予我足够的平等、尊重和关爱。

感谢哈尔滨工程大学让我在四年之间学到无法量化的知识和技能，老军工的自律精神在我人格上留下了深深的雕琢痕迹。自律的老军工营造了浓厚的学术氛围，感谢图书馆陪伴我走过整个大学四年，在那些寂寞而寒冷的日子给我光明给我温暖。相信今后老军工会成为我国乃至世界范围内一流的学校。感谢港中文对我的认可，让我有机会接触更高深的计算机科学。

感谢我的那些可爱的同学们，参与我人生美好的四年大学生活。感谢Lisa作为我大学最好的朋友，永远会在我需要帮助的时候站出来给我人生建议。感谢Christy同学在整个毕设阶段都和我一起泡图书馆，教会我用Matlab作折线图，教我论文写作的术语，督促我抓紧写代码写论文。感谢室友MissChe在论文排版方面为我提供耐心的帮助。感谢帮助过我的学姐们学长们，感谢XXX对我毕设的悉心指导，感谢XXX为我论文写了详细的批注。

感谢拜仁慕尼黑球队，勇得德甲七连冠，足球让我永远保持年轻，永远热泪盈眶。

感谢祖国的稳定，让我们可以安心的上学。