

基于社会化标签的个性化推荐系统研究

The Research on Personalized Recommendation System
Based on Social Tags

作者姓名 张 玉
学位类型 学 历 硕 士
学 科、专 业 信息管理与信息系统
研 究 方 向 信息化与管理创新
导 师 及 职 称 蒋翠清 教授

2011 年 3 月

合肥工业大学



本论文经答辩委员会全体委员审查，确认符合合肥工业大学硕士学位论文质量要求。

答辩委员会签名：（工作单位、职称）

主 席： { 刘 子 仙 子 子 子 子 }

委 员： (刘 子 仙 子 子 子 子 子 子)
安徽有轻件制地会子子
阳 子 子 合肥工业大学 副教授

导 师： 刘 子 仙

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标志和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得合肥工业大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签字：张玉 签字日期：2011 年 5 月 5 日

学位论文授权使用授权书

本学位论文作者完全了解合肥工业大学有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅或借阅。本人授权合肥工业大学可以将学位论文的全部或部分论文内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

(保密的学位论文在解密后适用本授权书)

学位论文者签名：张玉

导师签名：



签字日期：2011 年 5 月 5 日

签字日期：2011 年 5 月 2 日

学位论文作者毕业后去向：

工作单位：

电话：

通讯地址：

邮编：

基于社会化标签的个性化推荐系统研究

摘 要

随着互联网的发展, 标签技术的出现为个性化推荐带来了新的机遇。标签将用户和资源联系起来, 描述了资源的特征的同时体现了用户的兴趣偏好, 为个性化推荐服务带来了至关重要的数据源。然而, 现存的基于标签的个性化推荐系统还存在一些问题, 限制了标签在推荐领域的应用。主要包括以下三个问题: (1) 标签语义模糊问题, 模糊的语义难以准确描述用户兴趣, 降低了推荐精度; (2) 推荐效率问题, 随着用户标注急剧增加, 计算规模的增大, 推荐效率随之降低; (3) 推荐质量问题, 忽视了推荐项目本身的质量问题。

为了解决上述问题, 本文以智能推荐技术为背景, 以基于标签的个性化推荐系统为研究对象, 以消除标签语义模糊问题, 提高标签推荐系统的推荐精度、效率和质量为研究目的, 主要完成了以下工作:

(1) 基于大众标注的语义主题发现

分析标注中用户、资源和标签的潜在语义关系, 引入PLSA模型并对其进行扩展, 通过语义主题发现方法获得用户标注的潜在语义, 将用户标注映射到具有明确意义的语义主题上, 该方法较好的消除了标签语义模糊问题。

(2) 基于语义主题的用户兴趣模型

在语义主题发现研究的基础上, 利用语义主题所包含的语义信息对用户兴趣特征标签进行组织, 构建了三级的基于语义主题的用户多兴趣模型, 并提出了更新策略捕捉用户兴趣变化。该模型较好的描述了用户兴趣特征, 基于该模型的个性化推荐提高了推荐精度。

(3) 基于用户兴趣模型和评分的协同过滤推荐算法

研究了基于用户兴趣和评分的协同过滤算法, 借助用户多兴趣模型挖掘用户的潜在兴趣点, 然后在用户的兴趣范围之内推荐项目, 缩小了计算规模, 提高了推荐效率; 结合项目评分, 保证了推荐质量。

(4) 设计并开发了基于标签的图书个性化推荐系统, 介绍了推荐系统的体系结构和推荐流程, 并对其进行了比较完整的设计, 实现了系统总体框架中的大部分核心功能, 通过实验验证了系统的部分性能与实现效果。

关键字: 社会化标签, 推荐系统, 语义分析, 协同过滤

The Research on Personalized Recommendation System Based on Social Tags

ABSTRACT

With the development of the Internet, the emergence of tag technology has brought new opportunities to the personalized recommendation. The social tags connected users and items, described the characteristics of the items, embodied the user preference, and brought vital data sources for the personalized recommendation system. However, the existing personalized recommendation system based on the tags still has some problems, which limited the application of tags in the areas of recommendation. Mainly in the following three aspects: (1) Semantic fuzziness problem of tags. because of fuzzy semantics, the user preference was difficult to describe, which reduced the precision of recommendation system. (2) Efficiency problem of recommendation system. With increasing of user annotation and expanding of the scale of calculate, the efficiency of recommendation system was reduced. (3) Quality problem of recommendation system. The quality of the recommending items was neglected.

In order to solve the above problem, taking intelligent recommendation technology as background, taking tag-based personalized recommendation system as the research object, taking elimination for the tags semantic fuzziness problems and improvement of the accuracy, efficiency and quality of tag-based recommended recommendation system as research purposes, this paper mainly completed the following job:

(1) Discovery of semantic topic based on social annotation. This paper analysed the relation among users, items and tags, introduced latent semantic analysis probabilistic PLSA model and extended it, the latent semantic of user's annotation was gained by method of discovery of semantic topic, this method mapped the annotation to well-defined semantic topic, the semantic fuzziness problem of tags was eliminated.

(2) User interest model based on the semantic topic. On the basis of research on the discovery of the semantic topic, using semantic information which contained in semantic topic, this paper organized feature tags of user interest, then constructed the three-level user interest model based on semantic topic, and brought forward the refresh strategy to capture user interest changes. This model described user interest characteristics, and improved the precision of personalized recommendation system based on this model.

(3) Cooperative filter recommendation algorithm based on the user interest model and item rating. This paper researched cooperative filter recommendation algorithm based on the user interest model and item rating, mined user potential interest with the aid of the user interested model, then recommended the items to the user within the scope of the user interest, the computing scale was reduced, so the efficiency of recommendation systems was improved, the quality of the recommendation systems was guaranteed with item rating.

(4) This paper designed and developed a sample tag-based personalized recommendation system for book, introduced structure and recommend process of the recommendation system, realized most core functions of the system framework, verified part of performance and implementation effect of system by experiment.

Keywords: social tags, recommendation system, semantic analysis, collaborative filtering

致 谢

时光匆匆，如白驹过隙。两年半的硕士研究生的学习和生活就要随着这篇论文的答辩而结束了。籍论文完成之际，我要向我的师长、同学和家人表示由衷的感谢！

首先要感谢我的导师蒋翠清教授。两年多来，蒋老师在项目实践、论文的选题、构思及写作都给予了细心指引与教导，为我的研究工作指明了方向，使我顺利完成了硕士学业。蒋老师严谨的治学态度、丰富渊博的知识、敏锐的学术思维、精益求精的工作态度、积极进取的科研精神以及诲人不倦的师者风范是我终生学习的楷模。在此，谨向蒋老师表示我诚挚的感谢和崇高的敬意！

其次，衷心感谢梁昌勇院长、俞家文老师、丁勇老师、陆文星老师等所里的老师在我参与项目和学术研究期间给予的指导和教诲！同时感谢所有教育过我的老师！你们传授给我的专业知识是我不断成长的源泉，也是完成本论文的基础。

感谢管理学院 26 班以及信息技术与工程管理研究所的所有同学，与你们的交流给了我很多的启发，也使我在学习和生活中不断得到集体的温暖与关怀。

感谢多年来一直给予我鼎力支持的妹妹和无私奉献的父母。你们的支持一直是我前进的动力！

最后，谨向所有在攻读硕士学位期间曾经关心和帮助过我的老师、同学和家人表示最诚挚的谢意！

作者：张玉
2011年3月15日

目录

第一章 绪论	1
1.1 研究背景与意义	1
1.1.1 研究背景	1
1.1.2 研究意义	1
1.2 国内外研究现状	2
1.3 研究内容与结构安排	5
1.4 小结	6
第二章 基于标签的推荐系统及相关知识	7
2.1 社会化标签及标签系统	7
2.1.1 社会化标签	7
2.1.2 标签系统	7
2.2 个性化推荐技术和方法	9
2.2.1 协同过滤推荐	9
2.2.2 基于内容的推荐	11
2.2.3 基于混合的推荐	11
2.3 基于标签的推荐系统模型	12
2.4 本章小结	13
第三章 基于大众标注的语义主题发现	14
3.1 语义主题发现模型	14
3.1.1 语义主题分析	14
3.1.2 PLSA 扩展模型	16
3.2 基于大众标注的语义主题发现方法	17
3.2.1 语义主题发现基本流程	17
3.2.2 语义主题发现方法描述	18
3.2.3 基于语义主题标签聚类	20
3.3 实验验证	20
3.4 本章小结	23
第四章 基于标签的个性化推荐方法	24
4.1 基于语义主题的用户多兴趣建模	24
4.1.1 用户兴趣建模分析	24
4.1.2 基于语义主题的用户多兴趣建模	25
4.1.3 用户兴趣模型更新策略	26
4.2 基于用户多兴趣模型和评分的协同过滤推荐算法	27
4.2.1 算法描述	28
4.2.2 算法分析	30

4.3 实验验证 31

4.4 本章小结 33

第五章 原型系统实现 34

5.1 背景和目的 34

5.2 系统体系结构及工作流程 34

5.3 系统实现 36

5.4 实验与分析 37

5.5 本章小结 38

第六章 结论与展望 39

6.1 结论 39

6.2 展望 39

参考文献 41

插图清单

图 1-1 论文组织结构图	6
图 2-1 Delicious 标签系统	8
图 2-2 用户标签集合列表	8
图 2-3 基于标签的推荐系统模型	12
图 3-1 用户、标签、资源语义相关链	14
图 3-2 用户、资源和标签与潜在语义的关系	15
图 3-3 标注语义主题	15
图 3-4 标签语义对于用户群识别的影响	16
图 3-5 PLSA 算法示意图	16
图 3-6 三元 PLSA 模型	17
图 3-7 语义主题发现流程图	18
图 3-8 用户标签分布在各语义主题的概率	20
图 3-9 标签在潜在语义空间的概率分布	22
图 4-1 标签语义模糊	24
图 4-2 用户多兴趣	25
图 4-3 用户兴趣表示示意图	26
图 4-4 用户兴趣模型更新示意图	27
图 4-5 用户-语义主题相关性矩阵	28
图 4-6 基于用户多兴趣模型和评分的协同过滤算法流程	29
图 4-7 HR 评价指标对比图	32
图 4-8 ARHR 指标评价对比图	32
图 5-1 基于标签的图书推荐系统架构图	34
图 5-2 基于标签的图书推荐系统工作流程	35
图 5-3 系统提示用户定制标签界面	36
图 5-4 系统推荐页面	37
图 5-5 用户系统交互界面	37
图 5-6 兴趣变化统计直方图	38
图 5-7 用户评价系统推荐正确率对比图	38

表格清单

表 3-1 MovieLens10M100K 数据集具体数据格式.....21

表 3-2 UserTags 表.....21

表 3-3 潜在语义空间与标签.....21

表 3-4 标签与熵.....22

第一章 绪 论

1.1 研究背景与意义

1.1.1 研究背景

随着 Internet 的飞速发展,互联网实现了由 Web1.0 向 Web2.0 的升级换代,以博客、维基、RSS、SNS、标签等技术为代表的 Web2.0 体系由原来的自上而下的由少数资源控制者集中控制主导的 Web1.0 体系转变为自下而上的由广大用户集体智慧和力量主导。标签系统在这种环境下应运而生,在以用户为中心的 Web2.0 环境中,标签系统允许任意用户对感兴趣的网络资源进行无约束的标注,所有用户的标注都互为可见,这种开放的、共享的模式体现了以人为本的 Web2.0 理念,同时也为新环境中信息资源组织、推荐和共享带来了新的机遇。

虽然用户标注这个过程的本身非常简单,但是却有着很重要的意义。首先,用户标注反映了他们对网络资源的不同视角和理解,丰富了描述资源的多维角度;再次,标签系统能更好的为用户组织资源、浏览资源提供帮助。相比较之前的系统大多是依靠少数领域专家对资源进行科学分类,这种方法虽然比较科学和权威,但是不能很好的表达用户的个人想法和观点。而标签系统利用标签将网络资源按照用户的视角进行分类,能够更好的为用户组织资源、浏览资源提供帮助。最后,标签系统通过标签将用户和资源连接起来,通过三者的动态关系,可以分析出用户的兴趣偏好。例如,为同一项目资源标注相同标签的用户,很有可能具有共同的兴趣偏好,这对在标签系统中根据用户兴趣进行推荐提供了很好的基础。

总之,标签系统充分发掘了用户的积极性,使之参与到系统中来,发挥了广大用户所贡献的智慧和由用户联系形成的群体智慧的影响,解放了用户创作和贡献的潜能。

用户在标注网络资源时,不但更加准确客观的反应资源的相关特征,而且为个性化推荐服务带来了至关重要的数据源。结合标签系统的标签推荐系统,相比较传统的个性化推荐系统可以更加准确的获取用户的特征,为用户获取其“量身定做”的信息。从 GoogleNews、Amazon、豆瓣、MovieLens 等网站看,越来越多的网站已经慢慢开始在利用标签信息进行个性化推荐系统的摸索与应用,以此给用户带来更好的用户体验,从而提高网站流量及用户忠诚度和依赖度。

1.1.2 研究意义

20 世纪 90 年代中期,个性化推荐研究作为一个独立的概念被提出来,由于巨大的应用需求,推荐系统自提出以来得到了学术界和企业界的广泛关注。美国计算机协会多次把个性化推荐系统作为研讨主题,而国内外期刊也纷纷将推荐系统作为研究专题,明尼苏达大学教授 John Riedl 说:“推荐系统将成为未

来十年里最重要的变革，社会化网站将由推荐系统所驱动”^[1]。由于个性化推荐系统的良好发展和广阔的应用前景，目前，几乎所有大型电子商务系统和各种提供个性化服务的网站，如 Amazon、淘宝网和豆瓣网等，都在不同程度的使用着各种形式的个性化推荐系统，个性化推荐系统可以有效的保留客户提高客户忠诚度，为网站带来了巨大的效益。

自从 2003 年标签技术推出以来就得到了广泛的应用，结合标签的个性化推荐系统与以往推荐系统所能获得的信息有了本质区别，标签作为用户自由创造的关键词，体现了用户对资源的理解，也是用户之间联系和交流的纽带，因此使得个性化推荐技术的研究进入了一个新的阶段。标签从用户角度描述了信息资源的主要特征，涵盖了用户与资源之间，以及用户之间的关系，兼具内容与关联的特征，将标签作为推荐技术的基础数据来源，便能开发出同时具备基于内容和协同过滤优越性的推荐技术方法。因此研究基于标签的个性化推荐系统有着很重要的意义。

1.2 国内外研究现状

个性化推荐系统的发展源于二十世纪九十年代，它自产生以来引起了学术界和企业界的关注，对于它的研究一直持续至今。现在被广泛引用的个性化推荐系统的定义是 Resnick&Varian 在 1997 年给出的：“它是利用电子商务网站向客户提供商品信息和建议，帮助用户决定应该购买什么产品，模拟销售人员帮助客户完成购买过程”^[39]。实际中应用最多的是在网上购物环境下以商品为推荐对象的个性化推荐系统，它为用户推荐符合兴趣爱好的商品。

个性化推荐系统的最大的优点在于，它能主动的收集用户兴趣特征资料并根据用户兴趣特征，为用户作出有效的个性化推荐。同时，当系统中的商品库或用户兴趣特征发生改变时，系统给出的推荐实时更新，即给出的推荐序列会随着改变，总之，一个成功的个性化推荐系统的作用主要表现在以下三个方面：

(1)将浏览者转变为购买者。如果用户在电子商务系统的在浏览过程中没有购买的欲望，个性化推荐系统及时向用户推荐他们感兴趣的商品，就能促成购买过程^[40]。

(2)提高交叉销售能力。个性化推荐系统在用户购买过程中向用户提供其他有价值的商品推荐，用户能够从系统提供的推荐列表中购买自己确实需要但在购买过程中没有想到的商品，从而有效提高电子商务系统的交叉销售^[41]。

(3)提高客户忠诚度。个性化推荐系统分析用户的购买习惯，根据用户需求向用户提供有价值的商品推荐。如果推荐系统的推荐质量和精度很高，用户就会对推荐系统产生依赖，从而与用户建立长期稳定的关系，有效保留客户，提高客户的忠诚度^[42]。

作为推荐系统的核心，各种各样的推荐技术也被研究人员相继提出。目前，

比较成熟的推荐技术有基于内容的推荐、基于协同过滤技术的推荐、基于用户统计信息的推荐、基于效用的推荐、基于知识的推荐、基于规则的推荐等^[2]。其中协同过滤技术是推荐系统中最广泛使用的技术，协同过滤的概念是由 David Goldberg 等人在 1992 开发邮件系统 Tapestry 第一次提出，主要用于邮寄过滤，Tapestry 也成为第一个使用协同过滤推荐技术的系统^[3]。协同过滤技术提出以后，在不断地改进中成为当前应用最成功的推荐方法。随着互联网技术的发展，推荐技术也在向着更高效、更准确、更个性化发展。

随着 web2.0 的不断发展，社会化标签得到了广泛的应用。标签系统已经成为 Web2.0 网站的基本功能之一，为用户的个性化服务发挥了重要的作用，许多研究者对此进行的探讨，并取得了一定的成果^[3]。利用标签信息进行个性化推荐主要有以下几方面的研究：

(1)基于聚类分析的个性化推荐：聚类分析是用户模型建立的重要手段，有许多的研究者对此进行了探讨。对标签的聚类代表了不同语义主题标签间的分类，在这方面，Niwa 等人借助内容过滤分析，首先计算资源与标签的紧密度，得到用户-标签的紧密度；计算各个标签之间的相关性并将相关标签进行聚类，得到用户与已聚类标签的紧密度。再根据各个已聚类标签计算得出要推荐的网页，结合先前的紧密度得到最终的网页推荐顺序^[5]。对用户的聚类代表了不同兴趣的用户组，相似用户或邻居用户的识别对于推荐而言无疑具有重要的意义。Ae-Ttie 等人以用户-资源矩阵为切入点，通过计算目标用户和其他用户之间的相似性，得到目标用户的相似邻居，形成一个目标用户的候选标签集；再应用朴素贝叶斯法，结合标签-资源矩阵和用户-标签矩阵，计算出用户对特定标签的喜好程度；在综合各个标签的影响后，最终得到资源对用户的推荐度^[6]。Diederich 等人在数字图书馆领域根据相似用户的兴趣也相似的规律，通过了解相似用户对资源的标注为目标用户进行推荐^[7]。在对资源聚类方面，Sasaki 等人研究了基于标签对资源进行了内容聚类，通过假设检验计算了不同聚类间的相似性，最后给出了一个网页内容推荐系统^[8]。Yeung 等人运用贪婪算法，首先对单个用户所标注的资源进行内容聚类，再提取出已聚类资源的标签，将标注频率最高的标签引入用户模型进行推荐^[9]。

(2)基于矩阵处理的个性化推荐：标签系统可以分解为三类矩阵关系，即用户-资源矩阵、资源-标签矩阵和用户-标签矩阵。Tso-Sutter 等人通过矩阵扩展的方法，综合用户-资源矩阵与用户-标签矩阵，提出基于标签的协同过滤推荐算法，并综合多方的关系进行了分析^[10]。Yanfei Xu 等人对标签-资源矩阵进行 LSA 处理，得到 k 维的矩阵，将每个资源都表示成主流标签向量的形式。通过加和这些标签所标注的资源向量得到特定用户的标签，在此基础上计算用户标签和资源特征向量之间的相似性进行最终的推荐^[11]。在随后的研究中，Yanfei Xu 等人又进一步提出了用 HOSVD 算法将用户、标签和资源统一到一个框架中，

试图综合平衡各方面因素的相互影响。用户在选择一个特定标签时,该算法就可以为用户推荐相应的资源,该算法在一定程度上消除了标签的稀疏性和同义性问题^[12]。

(3)基于网络的个性化推荐:由于标签系统中的用户、资源和标签可以构建成网络的形式,Marlow 等人基于二分图构建了相应的网络后,利用社区内边比社区外边密集的原理,对网络进行社区划分,最终借鉴同一社区中其他用户的标注实现对目标用户的推荐^[13]。Michlmayr 等人探讨了三种用户模型构建的思想。其中两种思路是基于网络的方法:根据标签共现的思路,用点表示标签,边表示共现,画出相应的无向权图,取权重较大的 k 条边及其相应的点纳入用户模型。在该思路的基础上,应用蚁群算法中的蒸发技术将时间因素加入权重,实现新的用户模型构建^[14]。Zi-Ke Zhang 等人将经典物质扩散方法和标签结合起来,将用户-资源-标签三部图分解为用户-资源和资源-标签两个二部图,再分别利用物质扩散方法进行推荐后,使用线性组合的方法将两者有机的结合起来,从而充分的利用各自的优点,达到最优的推荐效果^[15]。

此外,还有一些文献从较为特殊的角度对基于标签的个性化推荐进行了探讨,如 Szomszor 等人对电影的推荐进行了分析^[16]。Plangprasopchok 等人研究了基于社会化标注的框架,通过条件概率建立了用户与资源的匹配模型^[17]。Eck 等人通过对 mp3 文件的分析,将音频特征自动转化为标签,实现了对音乐的推荐^[18]。

但是从目前实际应用中看,结合标签和推荐系统的标签推荐系统在推荐项目的效率和质量上还存在一些问题,主要问题有以下三点:

(1) 标签语义模糊问题。标签是由用户无约束的自由创造出来的,具有模糊性和社会性,存在着如同义词、多义词甚至是错误的拼写等情况,如果没有较好的语义描述,必然存在语义模糊等问题;如果不能准确理解用户标签语义,那么对于基于标签的用户兴趣获取也是模糊的,给出的推荐也是不准确的。因此,如何有效消除标签语义模糊问题,提高推荐精度是标签系统在推荐上的挑战之一。

(2) 推荐质量不高。基于社会化标签的推荐往往忽视了推荐项目本身的质量问题,网络信息资源浩如烟海,质量更是参差不齐,如何才能将高质量的信息资源推荐给用户,为用户提供更好的服务值得研究。

(3) 推荐效率问题。随着用户和网络资源的急剧膨胀,用户标注越来越多,计算规模越来越大,传统基于社会化标签的推荐的效率也就随着降低。因此,如何在利用标签优势的同时提高推荐效率也是值得考虑的问题。

本文主要从上述三个问题出发,针对传统基于标签的个性化推荐算法在标签系统中推荐精度、效率和质量上存在的瓶颈问题,结合标签语义主题发现和协同过滤推荐技术,通过引入 PLSA 模型并对其进行改进,获得用户标注的潜

在语义，解决标签语义模糊问题，进而发现出用户对各语义主题的关注度，建立基于语义主题的用户多兴趣模型，结合用户项目评分进行推荐，提高推荐的质量和效率。

1.3 研究内容与结构安排

针对标签推荐系统中存在的诸多问题，本文拟研究以下内容：

- (1)基于大众标注的语义主题发现研究；
- (2)基于语义主题的用户多兴趣模型研究
- (3)基于用户兴趣和评分的协同过滤推荐算法研究；
- (4)实现原型系统。

基于以上研究内容，本文的章节安排如下：

第一章，作为本文的绪论部分，主要介绍了基于标签系统的个性化推荐研究的背景和意义、国内外相关领域研究现状和本文的主要研究内容，并给出了论文的结构安排。

第二章，介绍了社会化标签及标签系统、主流的个性化推荐技术的概念、基础研究及关键技术等，对基于标签的推荐系统模型进行了分析研究，为后文的研究工作提供了理论基础。

第三章，基于大众标注的语义主题发现。主要分析了标注中用户、资源和标签的潜在语义关系，介绍了语义主题发现在标签系统中的作用，引入 PLSA 模型并对其进行扩展，详细阐述了基于大众标注的语义主题发现流程和方法描述并通过实验验证。

第四章，基于标签的个性化推荐方法。构建了基于语义主题的用户多兴趣模型，提出相关更新策略以捕捉用户兴趣变化，在此基础上，结合项目评分，详细描述了基于用户多兴趣模型和评分的协同推荐算法的流程，并对该算法进行了分析，最后通过实验验证了改方法的有效性。

第五章，原型系统实现。以图书推荐为切入点，设计并开发了一个简单的基于标签的个性化推荐原型系统，介绍了系统的体系结构和工作流程，并实现了系统的核心功能模块，验证了系统的部分性能与实现效果。

第六章，总结与展望。总结了研究过程中遇到的问题以及还没有解决的问题，并对下一步的研究提出进一步设想，为以后的工作指明方向。

本文的结构安排如图 1-1 所示。

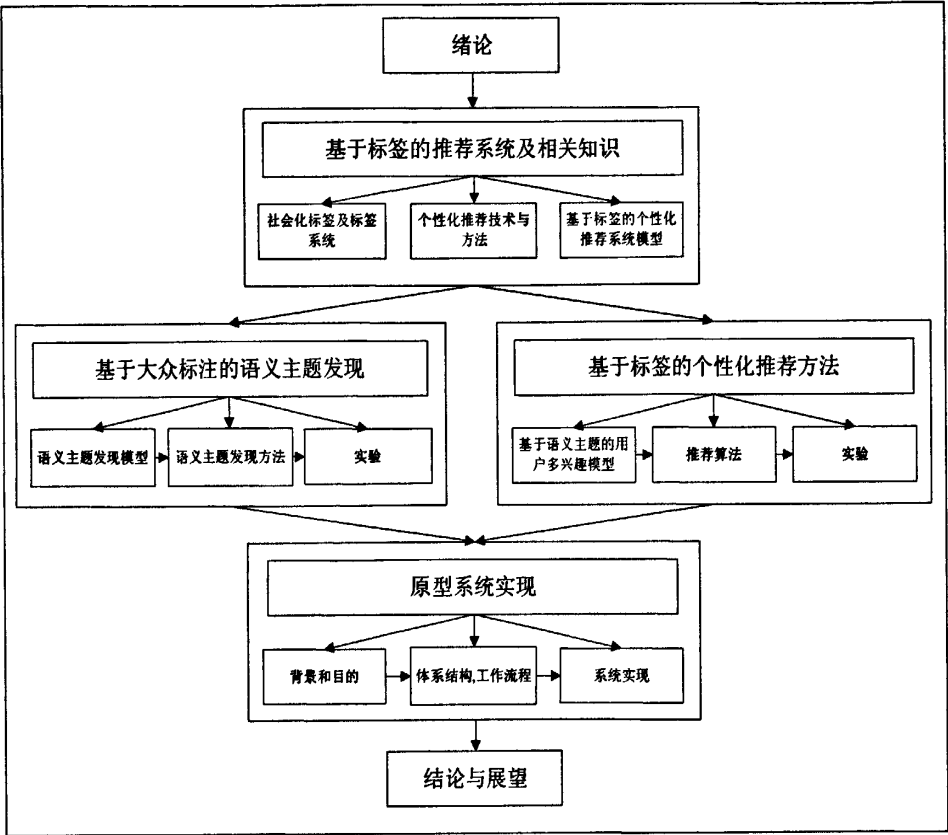


图 1-1 论文组织结构图

1.4 小结

本章介绍了论文的研究背景和研究意义，在对相关的国内外研究进行了综述的基础上，对研究的基于社会化标签的个性化推荐系统所存在的问题进行了总结，确定了本文的研究内容和论文结构，为下面章节作出了铺垫。

第二章 基于标签的推荐系统及相关知识

2.1 社会化标签及标签系统

2.1.1 社会化标签

针对越来越多的网络信息，标签以关键字的形式对资源进行分类，在个人理解角度对资源进行标注，不但丰富了描述资源的不同维度，而且展现了用户对资源的兴趣点。

社会化标签是一种灵活的、开放式的分类方式，又称为大众分类或分众分类(Folksonomy)，是用户无约束的运用自由定义的关键字的方式进行协作分类的一种工具^[19]。用户可以根据自己的需要自由的选择相应标签对网络资源进行标注，每个标签都是用户对网络资源的一个分类，这样，网络资源很自然地根据不同的标签被组织到不同的分类下，同时相同的标签也聚合了不同用户的相似资源^[51]。但是社会化标签不同于传统的目录结构的分类方法，标签之间是一种平行的关系，不考虑网络资源的层次关系。

社会化标签结合了“分类”与“关键字”的特点，对于从海量数据中发现用户潜在兴趣，为用户推荐个性化服务有着很大的价值。研究发现，社会化标签具有以下特点^[20]：

(1)自由性。社会化标签来源于网络用户，用户对网络上项目资源的标注具有完全的自由性和自主性。任何用户都可以对自己感兴趣的资源进行标注，提供一个或是多个标签，或不标注，仅仅浏览他人的标签。

(2)共享性。对于任何用户来说，所有的社会化标签都是共享的。每个用户都可以自由的查看或是利用别的用户所标注的标签，同时，自己标注的标签也可以被其他用户查看或使用。

(3)动态更新。随着用户的标注的不断增加，资源标签信息也越来越多，资源的标注信息得到不断的更新和丰富。

如今，标签技术不仅被广泛应用到网络产品上，同时也受到了国内外众多学者的重视。标签技术的应用与研究、基于社会化标签的检索、分类系统和标签推荐等都成为了近些年国内外研究讨论的热点问题。

2.1.2 标签系统

协同标签系统(Collaborative Tagging System)，简称标签系统，广泛的应用于 Web2.0 站点中，标签系统提供标签录入、标签存储、按标签组织及检索资源、标签分享和标签推荐等功能^[21]。标签系统是网络社会化发展的产物，它让用户之间建立起联系，充分满足了用户的个性化需求，使网站更具开放性。

Joshua Schachter 在 2003 年创建了网址书签收集分享网站 Delicious，成为最早的标签系统。用户在收藏网址的同时，使用标签对网址进行标记，进行分享^[22]。如图 2-1 所示，用户在收藏网址时，标注上基于自身理解的该站点的特征等。

Edit Bookmark

Title

合肥工业大学(HeFei University of Technology)

URL

http://www.hfut.edu.cn/ch/

Tags

myschool x

合肥 x

大学 x

Notes

1000

☐

Make private

图 2-1 Delicious 标签系统

在标签系统中，一个资源可以有多个标签，标签集合描述了资源各个特征，一个标签可以对应多个资源，这个标签描述资源的分类。在标签系统中，一个用户可以拥有多个标签，这个标签集合描述了用户的兴趣偏好，标签使用次数越多，用户对此类资源越感兴趣，如图 2-2 所示。

Tags	Options ▼
▼ Top 10 Tags	
job	3
movie	3
director	2
mtime	2
to	2
delicious	1
food	1
how	1
新浪微博	1
美食	1

图 2-2 用户标签集合列表

一个标签可以同时被多个用户共享，反映了大众的口味和偏好，标签被越多用户使用，说明大众越偏爱这类资源，从而可以预测流行趋势。国外流行的视频分享站点 YouTube、图片分享站点 Flickr、微博客站点 Twitter 等，国内流行站点豆瓣、优酷等也都使用了标签系统。标签系统具有以下特点^[20]：

- (1) 标签系统的使用门槛低，却对用户个体和群体的价值很大；
 - (2) 标签开放、灵活、且对用户兴趣的描述准确，在方便用户整理资源的同时有可为用户发现潜在兴趣，为用户提供个性化的服务；
 - (3) 使用群体智慧为资源描述特征，其描述准确、丰富且完整，能够较全面地概括资源。并且对于使用常规方法难以描述的资源(如音频、视频或图像等)，标签可以轻松地做到；
 - (4) 标签系统在用户方便自己的同时，也方便了别人，可谓“利己利人”。
- 标签系统目前比较成熟的应用有：

(1)标签系统应用于资源分类。标签是资源的特征的概括,可以作为资源的类别,这种分类方法较传统的分类方法更灵活,更快速。而且由于标签的数目和种类是无穷的,所以使用标签的分类也是无尽的^[23]。

(2)标签系统应用于资源检索。对于传统的搜索来说,搜索是基于资源中含有的关键字搜索,如文档中出现频度较高的词等。然而,关键字并不一定能全面、准确的描述文档;同时针对那些难以抽取特征的非文本资源来说,搜索旧更难以准确;标签是大众智慧的体现,使用标签来描述资源特征,是完全基于用户角度的理解,因此更加全面、准确,所以将标签作为关键字进行检索可以弥补传统检索方式的不足^[24]。

(3)标签系统应用于推荐。由于标签反映了用户的兴趣偏好,利用用户标签集合的相似性去寻找相似用户,为用户预测标签,进而推荐资源,帮助用户发现潜在兴趣,这是标签在推荐领域的非常有意义的新尝试^[52]。2006年,在推荐技术领域有着卓越贡献的 GroupLens 项目组^[25]将他们开发的 MovieLens 电影推荐系统引入了社会标签的特性,开始进行基于标签的推荐尝试。

2.2 个性化推荐技术和方法

2.2.1 协同过滤推荐

协同过滤推荐算法是应用最为广泛且最成功的个性化推荐方法,已被应用到很多领域中。与传统推荐方法不同,协同过滤推荐不比较资源与用户模型之间的相似性,而是通过用户对资源的评价来发现用户之间的相似性,把具有相似兴趣的用户分为一个用户群体,当用户对某资源感兴趣时,就把该资源推荐给同在一个群体的其他用户。

协同过滤推荐实质上是现实生活中经常采用的推荐方式,即两个兴趣相近的朋友相互推荐。协同过滤的实现主要分两步:第一步获得用户兴趣信息;第二步分析用户之间的相似性并预测目标用户对某一方面的偏好^[28]。协同过滤的算法可以分成基于用户的和基于项目的协同过滤。下面分别对这两种算法进行介绍:

(1)基于用户的协同过滤

基于用户的协同过滤方法用相似统计的方法得到具有相似爱好或者兴趣的相邻用户来进行推荐^[26],该方法的具体步骤如下:

①收集用户的兴趣信息,通常有“显式”和“隐式”两种方法,“显式评分”通常采用评分或评价的方式,在标签系统中,还可以使用标注的方式,但这种方式也有个明显的缺点,愿意主动评分的用户很少,收集数据比较困难^[46]。“隐式评分”是系统根据用户的行为特征如点击、浏览时间等信息代替用户完成评价,这种方式不需要用户直接输入评价数据,其中研究较多的方法是 Web 挖掘^[47]。

②最近邻搜索。基于用户的协同过滤技术的核心是寻找与目标用户有这相同兴趣爱好的邻居用户，一般通过相似度计算来求得^[48]。例如：寻找 N 个与目标用户有相似兴趣的用户，把他们对待预测的网络资源 R 的评分作为目标用户对 R 的评分预测。一般会根据不同的需要选择不同的相似度计算方法，目前使用相似度算法有很多，其中最常用的主要有三种：余弦相似性、皮尔森相关系数以及修正的余弦相似性^[50]。

③产生推荐结果。有了最近邻集合，就可以对目标用户的兴趣进行预测，产生推荐结果。依据推荐的不同目的进行不同形式的推荐，较常见的推荐形式有 Top-N 推荐和关联推荐^[1]。Top-N 推荐是针对个体用户产生，对每个用户产生不一样的结果，例如：通过对目标用户的最近邻用户感兴趣的网络资源进行统计，出现频率高且不在目标用户已浏览范围内的作为推荐结果推荐给目标用户；关联规则推荐是通过对最近邻用户的记录进行关联规则发现来给出推荐结果。

(2)基于项目的协同过滤

基于用户的协同推荐算法随着用户数量的增多，计算规模变得越来越大，推荐效率也会随之降低。针对该问题 Sarwar 等在 2001 年提出了基于项目的协同过滤推荐算法(Item-based Collaborative Filtering Algorithms)^[27]。与基于用户的方法类似，基于项目方法需要同样的三个步骤：

①得到用户-项目的评分数据。

②针对项的最近邻搜索，即对项进行相似度计算。在基于项目的方法中，要对资源 A 和资源 B 进行项相似性计算，通常分为两步：1) 找出同时对资源 A 和资源 B 打过分的组合；2) 对这些组合进行相似度计算，常用的算法包括：余弦相似性、皮尔森相关系数、修正的余弦相似性和条件概率等^[50]。

③产生推荐。

相对于基于用户的推荐方法，基于项目的推荐方法最大的改进在于提高了协同过滤方法的扩展性及性能。基于项目的协同过滤不考虑用户间的差别，所以推荐的精度比较差；但是不需要用户的历史资料来进行用户识别，所以用户之间的相似性可以通过计算项之间的相似性来代替，而由于项目具有较好的稳定性，所以工作量最大的相似性计算部分可以离线完成，降低了在线计算量，提高了推荐效率^[49]。

协同过滤推荐系统的优点是：不依赖于资源的内容，不仅适用于文本领域，也适用于其他领域，如音乐、电影、视频等等；有推荐新信息的能力；可以发现内容上完全不相关的信息，用户对推荐信息的内容事先是预料不到的；协同过滤可以发现用户潜在的但自己尚未发现的兴趣偏好；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^[28]。

协同过滤推荐系统的缺点是：稀疏性问题，当每个用户都只对很少的项目

给出评价时,整个用户-项目矩阵变得非常的稀疏,这样协同过滤推荐系统得不到足够的用户评分来进行预测,致使用户之间的相似度计算不准确,得到的邻居用户不可靠^[29];冷启动问题包括新资源问题和新用户问题,当一个新的资源加入到系统中时,因为该资源没有用户的评价信息,所以无法被推荐给用户,同时如果一个用户没有评价任何资源,跟其他用户的评价项交集太少,他也不能找到最近邻而获得推荐^[30]。

2.2.2 基于内容的推荐

基于内容的推荐技术来源于搜索领域的相关成果,主要用于文档的推荐。其主要思想是抽取项目的内容信息,根据项目的内容特征,产生与用户兴趣相似的推荐项目。所谓“内容”,就是指的是文本和非文本(如图片、音视频等)的特征词。

基于内容的推荐过程如下:①抽取项目的内容特征,得到项目的特征词向量;②根据用户的个人信息,获取用户的兴趣偏好;③寻找项目特征与用户兴趣相匹配的项目,产生推荐^[31]。

一般用特征词向量来描述项目的内容,其维数是所有项目中特征词的总个数。一个项目的特征词向量的每一维的值是该特征词在项目中出现的次数,然后通过特征词向量计算规范化的 TFIDF(Term Frequency/Inverter Document Frequency)向量^[39]。其中每个 TFIDF 向量都表示一个项目,项目之间的相似度就表示为对应两个 TFIDF 向量距离的倒数,使用余弦相似度等计算方法,计算用户个人档案中的项目的 TFIDF 向量与用户为选择的项目的 TFIDF 向量的相似度,然后按相似度从大到小取前 N 个项目进行推荐^[36]。

基于内容的推荐的优点是:①不需要其他用户的数据,没有冷开始和稀疏问题;②建模和商品间的相似性计算可以脱机进行,具有很快的推荐响应时间;③通过列出推荐项目的特征可以解释为什么推荐那些项目^[34]。

基于内容的推荐系统的缺点是:①一般只能用于文本领域,只能对文本内容进行浅层的分析,不能区分资源在层次化结构上体现出的特征差异^[32];②“过度规范”问题,用户只能得到和当前用户兴趣最相似的信息,用户获取的资源被限定在用户以前所阅读或评价的资源范畴内,不能为用户发现新的感兴趣的资源^[33];③仅考虑了当前用户自己的行为,不能显式的得到其他用户的意见。

2.2.3 基于混合的推荐

混合推荐的一个重要原则就是通过混合后避免或弥补各自推荐技术的弱点。目前研究最多的是基于内容的推荐和协同过滤推荐的混合^{[34][35]}。根据采用的组合思路,将组合思路分成三类:后融合、前融合和中融合^[36]。

①后融合:融合两种或两种以上的推荐方法各自产生的推荐结果。如使用协同过滤方法和基于内容的方法分别得到推荐列表,融合列表的结果决定最后推荐的对象。

②中融合：以某一种推荐方法为框架，融合另一种推荐方法。如以基于内容的方法为框架，融合协同过滤的方法，或者以协同过滤的方法为框架，融合基于内容的方法。

③前融合：直接融合各种推荐方法。如将基于内容的推荐方法和协同过滤的推荐方法整合到一个统一的框架模型下。

此外，在采用组合推荐技术的时候，必须注意组合策略。这些策略与被组合的推荐技术有关。

2.3 基于标签的推荐系统模型

为了解决标签推荐系统中现存的一些问题，本文提出了如下基于标签的推荐系统模型，如图 2-3 所示。

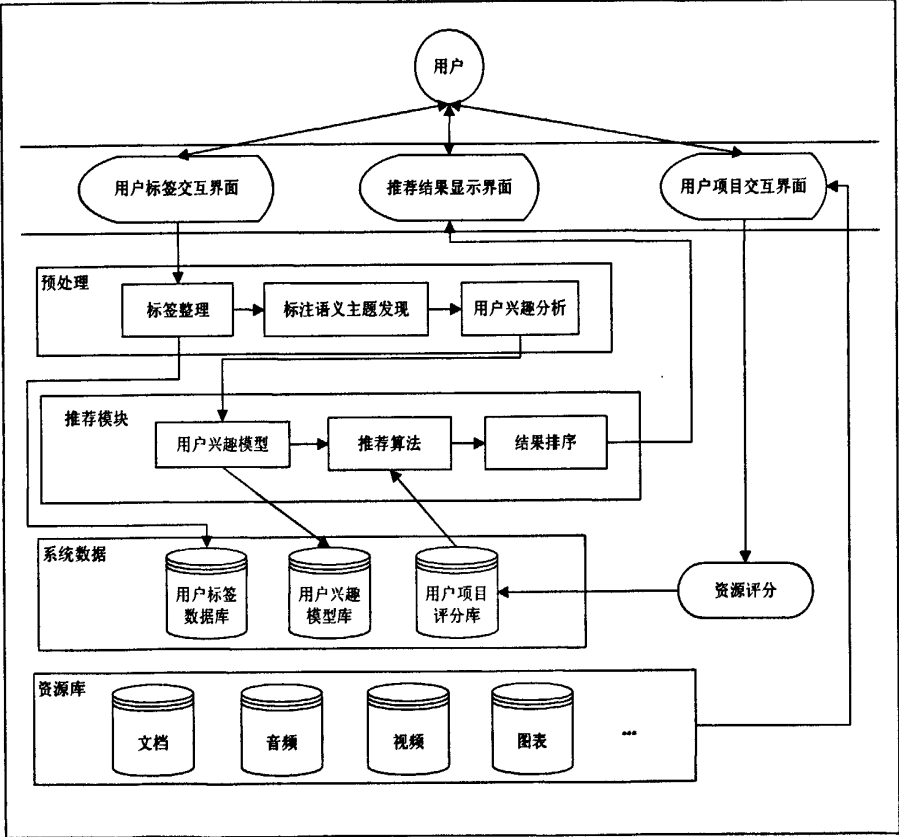


图 2-3 基于标签的推荐系统模型

构建这样一个标签推荐系统模型主要基于三点考虑：①在标签整理和推荐算法的模块研究标注语义主题发现，可以解决传统标签推荐系统中的标签语义模糊问题。②独立的用户多兴趣模型可以根据实际应用领域构建细致的模型，并进行实时的更新。③将资源评分也纳入模型中可以更好的描述所推荐项目资源的质量。④独立的用户标签数据库用于存储用户和标签交互信息，用户兴趣模型数据库用于存储用户兴趣模型，可以对标签信息和用户兴趣模型进行离线整理，提高系统的效率。

由图 2-3 可以看出, 本文的标签推荐系统模型主要由四部分组成: 人机交互模块、预处理模块、推荐模块和系统数据库模块。下面简单介绍下标签推荐系统模型各个模块:

(1) 人机交互模块

人机交互层用于实现用户与系统间的交互, 主要分为用户标签交互界面、用户项目交互界面以及推荐结果显示界面三类。其中, 用户标签交互界面是用于用户使用标签浏览或标注资源, 负责接收用户标注或浏览等数据, 传递给预处理模块; 用户项目交互界面用于用户为项目评分, 负责接收用户评分数据; 推荐显示界面接收推荐模块返回的推荐结果进行显示。

(2) 预处理模块

预处理模块是一个中间层, 负责接收人机交互模块传递的数据并对数据进行相关分析和预处理, 包括“标签整理”“标注语义主题发现”“用户兴趣分析”三个子模块。其中标注语义主题发现是本文研究的重点之一, 通过语义主题发现消除标签的语义问题。

(3) 推荐模块

推荐模块是整个模型的核心, 它由用户兴趣模型、推荐算法和推荐结果优化三部分构成。其中用户兴趣模型的构建和推荐算法也是本文研究的重点, 本文首先在语义主题发现的基础上构建了三级的用户多兴趣模型, 描述用户多方面的兴趣偏好, 然后结合项目评分, 通过协同过滤方法将最近邻的高质量项目资源推荐给用户。推荐结果优化是将推荐的项目资源按照语义相关程度进行排序, 它与人机交互模块的推荐结果显示界面联系密切。

(4) 系统数据库模块

系统数据库模块是整个模型的基础层, 为模型提供数据服务, 主要由用户标签数据库、用户兴趣模型数据库和用户项目评分库组成。资源库包含推荐系统中信息内容, 是推荐系统推荐; 用户标签数据库存储整理后的用户标注信息, 用户兴趣模型数据库存储用户兴趣模型并根据用户操作周期更新, 用户项目评分库存储用户对项目的评分信息, 用于对项目资源质量进行评估。

在本文提出的基于标签的推荐系统模型的各个组件中, 其中最重要的是标注语义主题发现、用户兴趣模型和推荐算法着三个部分, 这三部分直接影响到整个系统的推荐效果, 因此在后面的章节, 本文将在这三个部分做重点研究。

2.4 本章小结

本章主要阐述了推荐技术和标签系统的一些背景知识, 介绍了社会化标签和标签系统的定义、特点、用途和应用。介绍了几种主流的个性化推荐技术方法如基于内容的推荐、基于协同过滤的推荐和其他推荐。最后提出了基于标签的推荐系统模型, 并简单介绍了各个模块。为后面的阐述做了知识上的准备。

第三章 基于大众标注的语义主题发现

由于标签的社会性和随意性，如果没有较好的语义描述，必然存在着语义模糊等问题。因此，本章从分析标注中用户、资源和标签的潜在语义关系出发，引入PLSA模型，并对其进行扩展，将用户标注映射到具有明确意义的语义主题上，主要研究了基于大众语义主题发现方法，并通过实验验证了该方法的有效性。

3.1 语义主题发现模型

3.1.1 语义主题分析

已有研究中发现，标签通常是语义相关的，如果他们多次用来标记相同或相关的资源，那么使用这些语义相关标签的用户可能有相似的兴趣。如果某些资源被许多有着相似兴趣的用户标注的话，那么这些资源在语义上也是相关的，而这种语义相关可以从多种角度去发现，例如，一组由志趣相投的用户经常使用的标签，在语义上是相关。相关资源通常是由语义相关的标签多次标记，如果用户在他们的标注中分享很多资源，那么他们可能具有相似的兴趣偏好。这种语义相关链体现在用户，资源和社会标签共同出现的不同频率，共同出现的这些频率体现在他们的潜在语义。如图3-1所示。

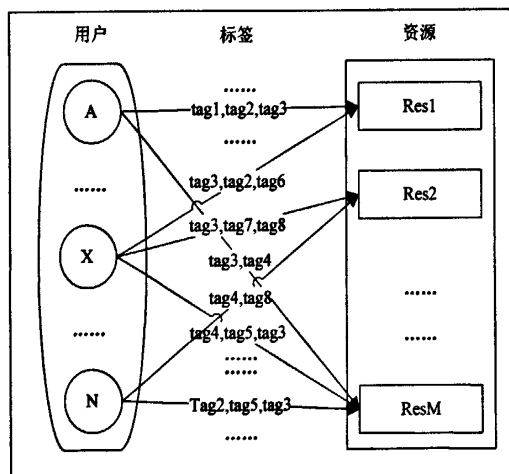


图 3-1 用户、标签、资源语义相关链

显然，在标签系统中，一个标注至少包含以下三个元素：用户、资源、标签，其形式化描述如下：

U：表示用户的集合

R：表示项目资源的集合

T：表示标签集合

对于 $u_i \in U$ 、 $t_n \in T$ 、 $r_j \in R$ ，三元组 (u_i, r_j, t_n) 表示用户 u_i 使用标签 t_n 标注了项目资源 r_j ，本文将一个三元组 (u_i, r_j, t_n) 的语义看做多维的向量，其中每一维代表了一个意义明确的特殊语义，每个三元组都可以被映射到多维向量上，即具有多种语义，将所有的三元组都映射到一个多维的向量空间，这个向量空间被

称之为“潜在语义空间”，而潜在语义空间的每一维都表示了一类意义明确的“语义主题”。如图3-2所示。

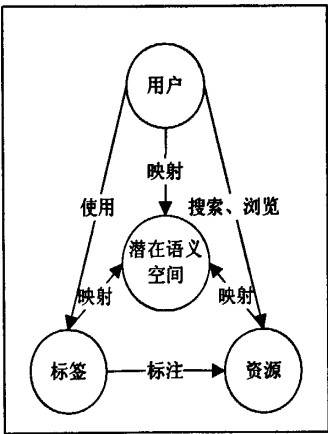


图 3-2 用户、资源和标签与潜在语义的关系

当一个用户使用标签标注项目，或通过标签浏览、搜索资源时，在他的脑海中，这些标签是带着某个意义明确的语义的，即都是针对某个特定语义主题的。如果不能明确用户标签所代表的语义，那么对于用户的真正意图也就无从而知了。如图 3-3 所示。

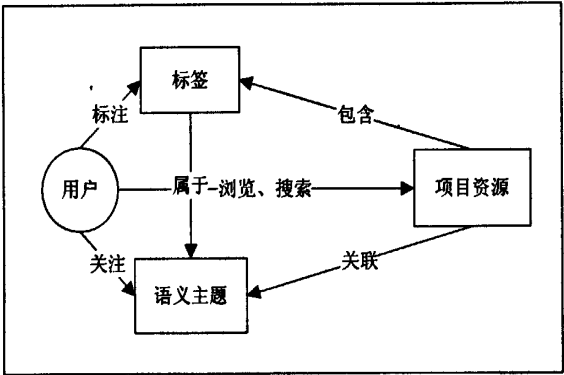


图 3-3 标注语义主题

标签的语义问题对于相同兴趣的用户群识别也有着重要影响，通过语义主题发现，标签的语义问题得到了消除。如图 3-4 所示，用户 Ua 和用户 Ub 都拥有相同的标签“Apple”，用户 Uc 有标签“Micro”，如果不对标签语义进行分析，很可能会误认为用户 Ua 和用户 Ub 有着相同的兴趣爱好，但是事实上用户 Ua 的标签是属于语义主题“水果”，用户 Ub 的标签是属于语义主题“计算机”，这两个用户所关注的语义主题是不相同的。而用户 Uc 的标签“Micro”属于语义主题“计算机”，那么很显然，虽然用户 Ub 和用户 Uc 的标签不相同，但是他们对相同语义主题“计算机”有着共同的兴趣。

	水果	计算机	语义主题n
Ua	Apple				
Ub			Apple		
Uc			Micro		

图 3-4 标签语义对于用户群识别的影响

因此基于大众标注的语义主题发现,使用语义主题解决标签语义模糊问题,对于标签系统的信息组织,为用户提供更好的搜索和推荐服务都有着很重要的意义。

3.1.2 PLSA 扩展模型

对于这种潜在语义所对应的语义主题,可以通过 PLSA 模型来计算。PLSA 模型起源于自然语言处理研究,由 Hofmann 在 1999 年提出。PLSA 算法是以概率为基础的算法,其核心思想是示象模型(aspect model)。该模型使得一不可见的隐性变量集 $z_k \in \{z_1, z_2, \dots, z_k\}$ 与每一次观测值相关^[37]。这里的观测值为一特定文本 d_i 中某词 w_j 的出现概率,可得词-文的联合概率模型如(3-1)所示。

$$\begin{aligned}
 P(d_i, w_j) &= P(w_j | d_i) \\
 &= \sum_{k=1}^K P(w_j | z_k) P(z_k | d_i)
 \end{aligned}
 \tag{3-1}$$

显然,上式须用所有可能的 z_k 去模拟示象模型,其原理如图 3.4 所示。考虑到潜在变量的基数 K 通常比文本或词的数目小的多,将图 3-5(1)转化为图 3-5(2),则式中的联合概率如式(3-2)所示。

$$P(W_j | d_i) = \sum_{k=1}^K P(z_k) P(w_j | z_k) P(d_i | z_k)
 \tag{3-2}$$

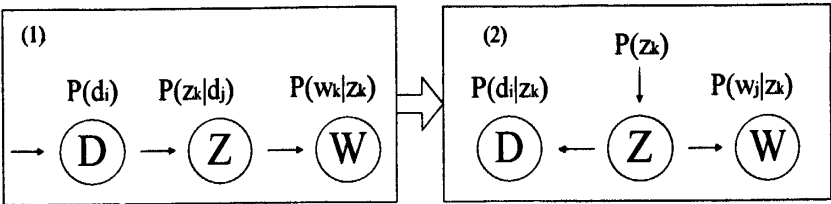


图 3-5 PLSA 算法示意图

PLSA 将隐性变量与可视共现数据对 D-W 联系起来建立统计模型,使用概率密度函数作为文和词之间潜在语义关联的呈现方式;基于多项式分布和条件分布的混合共现概率建模,建立层次结构达到降维的效果。PLSA 模型被广泛运用于信息检索、文本分类上,其降维可去除部分噪声,与语言无关、自动处理同义词、多义词的能力^[38]使之尤其适合于本文要解决的问题。

受 PLSA 模型研究的启发, 本文通过研究在标签系统中用户、资源和标签共同出现的频率发现标注中所隐含的语义关系, 发现用户语义主题。为了适应在标签系统中的应用, 本文将 PLSA 模型扩展为三元模型。

假设潜在语义空间是一个 K 维向量空间, 即 $Z = (z_1, z_2, \dots, z_K)$, 每一维代表了一类意义明确的语义主题, $P = \{P(z_1), P(z_2), \dots, P(z_K)\}$ 表示用户对语义主题的偏好程度, 且 P 满足 $\sum_{k=1}^K P(z_k) = 1$ 。

三元组 (u_i, r_j, t_n) 表示用户 u_i , 使用标签 t_n 标注资源 r_j , 利用概率分布去尽可能地逼近特定的潜在语义分布。然而, 就一个标签系统而言, 同一个用户只能对同一个资源使用同一个标签一次, 所以, 本文将这个三元模型分为两部分, 一个是由用户和标签组成, 一个是由标签和项目资源组成。用标签将用户和项目资源联系起来, 再分别利用两者之间的共现关系反应其中隐含的语义关系。因为所对应的潜在语义空间是由用户、标签和资源共同组成的, 所以两部分所对应的语义主题也是相同的。其模型如图 3-6 示。

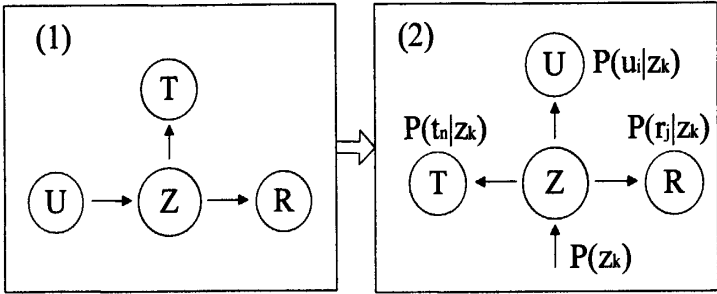


图 3-6 三元 PLSA 模型

3.2 基于大众标注的语义主题发现方法

3.2.1 语义主题发现基本流程

基于大众标注的语义主题发现算法的基本思想如下: 针对标签系统用户生成标签的随意性, 引入概率潜在语义技术(PLSA)并对其进行扩展, 通过改进的 PLSA 模型生成了标签及资源的潜在语义, 发现语义主题, 消除标签的同义性、一词多义等语义模糊问题; 同时得出标签与各语义主题的关系和用户对各语义主题的关注度, 从而为用户在标签系统中的浏览、搜索和推荐提供了良好基础。

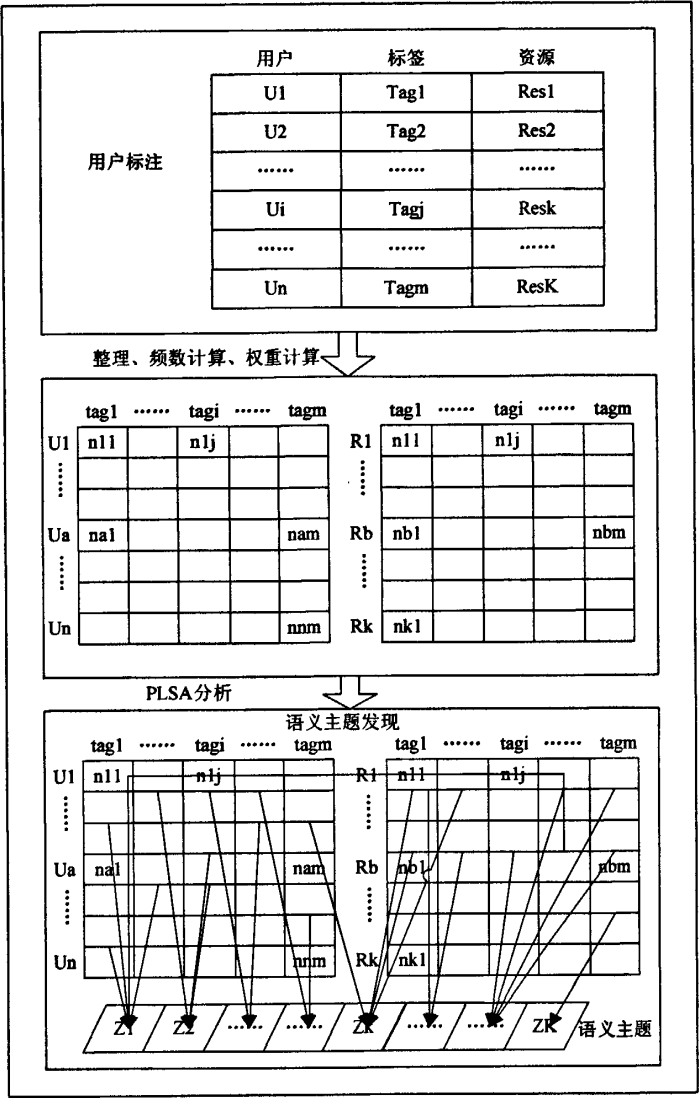


图 3-7 语义主题发现流程图

3.2.2 语义主题发现方法描述

定义变量如下：

U：表示用户的集合

R：表示项目资源的集合

T：表示标签集合

(u_i, r_j, t_n) ：表示用户 u_i ，使用标签 t_n 标注资源 r_j

假设 $Z = (z_1, z_2, \dots, z_K)$ 表示 K 个语义主题集合，根据本章第二小节所示的三元 PLSA 模型，则用户 u_i 使用 t_n 的概率表示如式(3-3)。

$$P(u_i, t_n) = \sum_{k=1}^K p(z_k) p(u_i | z_k) p(t_n | z_k) \quad (3-3)$$

项目资源 r_j 被标注上标签 t_n 的概率表示如式(3-4)。

$$P(r_j, t_n) = \sum_{k=1}^K p(z_k) p(r_j | z_k) p(t_n | z_k) \quad (3-4)$$

对于已知的数据集, 利用中采用的最大似然法(Maximum likelihood)^[39]来确定各参数, 对于三元组 (u_i, r_j, t_n) , 目标函数 Γ 如式(3-5)。

$$\Gamma = \sum_n \left[\alpha \sum_i n(u_i, t_n) \log P(u_i, t_n) + (1-\alpha) \sum_j n(r_j, t_n) \log P(r_j, t_n) \right] \quad (3-5)$$

其中权值 $n(u_i, t_n)$ 表示 u_i 和 t_n 共同出现的次数, $n(r_j, t_n)$ 表示 u_i 和 r_j 共同出现的次数, α 是系数。

对于上述目标函数求最大值, 通过 EM(Expectation Maximization)算法来完成。其中, 在最大化过程中要满足的约束条件如式(3-5)所示。

$$\sum p(u|z)=1, \sum p(r|z)=1, \sum p(t|z)=1 \quad (3-5)$$

在 E-step 中计算潜在语义 z_k 的后验概率如式(3-6) (3-7)。

$$P(z_k | u_i, t_n) = \frac{P(t_n | z_k) P(z_k | u_i)}{P(t_n | u_i)} \quad (3-6)$$

$$P(z_k | r_j, t_n) = \frac{P(t_n | z_k) P(z_k | r_j)}{P(t_n | r_j)} \quad (3-7)$$

在 M-step 中更新各参数值如式(3-8)-(3-10)。

$$P(z_k | u_i) \propto \sum n(u_i, t_n) P(z_k | u_i, t_n) \quad (3-8)$$

$$P(z_k | r_j) \propto \sum_j n(r_j, t_n) P(z_k | r_j, t_n) \quad (3-9)$$

$$P(t_n | z_k) \propto \alpha \sum_i n(u_i, t_n) P(z_k | u_i, t_n) + (1-\alpha) \sum_j n(r_j, t_n) P(z_k | r_j, t_n) \quad (3-10)$$

通过不断迭代最终得到目标函数的近似最大似然解, 得到模型参数集, 根据贝叶斯概率公式, 可以计算出 u_i, r_j, t_n 与语义主题 z_k 的相关程度, $P(z_k | u_i)$ 表示用户 u_i 对语义主题 z_k 的关注度, $P(z_k | u_i, t_n)$ 表示的是对于用户 u_i 来说, 标签 t_n 属于语义主题 z_k 的概率 P_{u_i} , 所以用户标注就可以用其在各语义主题的概率分布来表示。令 $U = \{u_1, u_2, \dots, u_n\}$ 表示 n 个用户的集合, $T = \{t_1, t_2, \dots, t_m\}$ 表示 m 个标签的集合, 用户在各语义主题的概率分布如图 3-8 示。

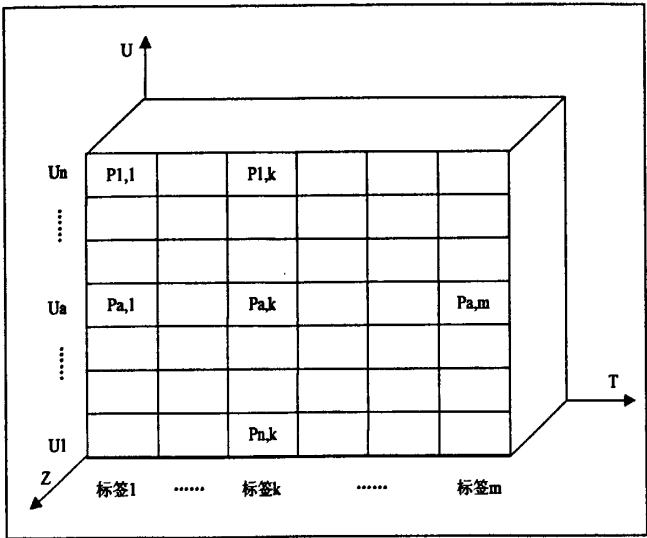


图 3-8 用户标签分布在各语义主题的概率

3.2.3 基于语义主题的标签聚类

应用上述改进的PLSA模型分析，发现语义主题，得到用户标注在各语义主题的分布。将用户标签聚类到个语义主题下较好的解释了标签语义。对于用户来说，一个标签可能有多重语义，即属于多个语义主题。根据用户标签与各语义主题的相关度，方法如下：

输入：“用户-资源-标签”矩阵，聚类阈值 γ 。

输出：聚类结果即用户各语义主题下的标签集 T_z 。

Step1：选择“用户-资源-标签”矩阵，根据改进的PLSA模型分析，得到语义主题集 Z 。

Step2：选择语义主题集 Z 中一个语义主题 z_k ，选择用户的一个标签 t_n ，计算该标签与语义主题 z_k 的相关度 $P_{u,t} = p(z_k | u, t_n)$ 。如果 $P_{u,t} > \gamma$ ，将标签 t_n 分配到语义主题 z_k 下。

Step3：选择用户其他的标签，重复Step2，直到计算用户所有标签与语义主题 z_k 的相关度，形成的标签集 T_{z_k} 为用户在语义主题 z_k 下的标签集。

Step4：选择语义主题集 Z 中其他语义主题，重复Step2、Step3。

Step5：迭代结束，输出最终的聚类结果 $T_z = \{T_{z_1}, T_{z_2}, \dots, T_{z_K}\}$ 。

3.3 实验验证

为验证该方法的有效性，本节在 MovieLens 数据集上对该方法展开实验，验证该方法对标签语义模糊问题的消除效果。本节采用的数据集是 GroupLens 小组开放的 MovieLens10M100K 数据集。该数据集中每个用户至少对 20 部电影参加过评分。由于其中有多达 100,000 个标签数据。

该数据集中有三个 DAT 文件：movies.dat、ratings.dat 和 tags.dat。其中 movies.dat 储存了电影的 ID 对应的名称和电影类别，ratings.dat 储存了用户对

电影的评分和评分时间, tags.dat: 储存了用户对电影标记标签信息和标记时间, 具体数据格式如表 3-1 所示。

表 3-1 MovieLens10M100K 数据集具体数据格式

文件	格式	说明
movies.dat	MovieID::Title::Genres	电影 ID, 名称, 类型
ratings.dat	UserID::MovieID::Rating::Timestamp	用户 ID, 电影 ID, 评分, 时间戳
tags.dat	UserID::MovieID::Tag::Timestamp	用户 ID, 电影 ID, 标签, 时间戳

由于 MovieLens 数据集数据量较大, 考虑到实验硬件环境和时间因素的限制, 本实验选取 MovieLens 数据集的 20%的数据作为实验数据。使用 tags.dat 数据文件, 将实验数据导入数据库, 生成 UserTags 表。UserTags 表的一条记录表示特定用户给特定电影标记标签的记录, 数据库的表定义如 3-2 所示。

表 3-2 UserTags 表

UserTags		
字段名	类型	描述
ID	Char	主键, 自增 ID
UserID	Char	用户 ID
MovieID	Char	电影 ID
Tag	Char	用户对电影标注的标签

根据方法描述, 经过扩展 PLSA 模型计算潜在语义, 发现语义主题。经过多次迭代循环, 在不同潜在语义空间维数下记录下目标函数值, 对于本文实验数据将维数置为 55 时结果较好。对于每个语义主题, 将标签按照 $p(z_k | t_n)$ 从高到低排列, 记录下各语义主题下的标签集, 可以得到语义主题的直观标签表示, 如下给出其中 6 个语义主题的前 10 个标签, 如表 3-3 所示。

表 3-3 潜在语义空间与标签

潜在语义空间	标签
语义主题 1	World War II,holocaust,true story,history,Civil War, military, biography, anti-war, politics
语义主题 2	Psychology,Serial killer, crime, murder, drugs, cannibalism, dystopia, mystery, memory
语义主题 3	Adventure,fantasy,time travel,sci-fi,future,magic,Johnny Depp, Sturn Award, religion, archaeology
语义主题 4	Action,superhero,martial arts,James Bond, forceful, Mafia, revenge, trrorism, Quentin Taratino, violence
语义主题 5	Disney,Pixar,Animation,cartoon,fairy tale, children, funny, cute, kids, family
语义主题 6	comedy,Jim Carry,parody,funny,black comedy,Steve Carell, satire, hilarious, whimsical, Bill Murray, humorous
....

通过表 3-3, 可以很明显地观察到 PLSA 模型能够在一定程度上将具有代表

性的标签显示出来，每个维表示的都是特殊的明确的语义信息。例如上表“语义主题 3”表示的是冒险类影片，其该语义主题下的标签集也包含了此类影片的主要信息，如“Sturn Award”是此类影片的重要的奖项。而“语义主题 5”表示的是动画类影片，其该语义主题下的标签集包含了影片出品公司，语义主题和评价等方面，

熵可以表示信息的不确定程度，通过计算信息熵得到标签的模糊性，标签熵计算如式(3-11)所示。

$$\text{entropy}(t_n) = - \sum_{k=1}^K p(z_k | t_n) \log p(z_k | t_n) \tag{3-11}$$

对于实验数据进行计算，得出表 3-4 所示的部分数据。

表 3-4 标签与熵

标签	熵	标签	熵
DVD	3.62	train	1.24
owned	3.05	health care	0.42
based on a book	2.98	World War II	0.35
seen more than once	2.73	weather	0.02
classic	2.68	Pixar	0

通过上表可以看出标签“DVD”是 MovieLens 数据集上使用的语义最模糊的标签，该标签基本在各语义主题上都分布的很均匀，标签“Pixar”具有较低的熵，它在这个数据集的意义是很明确的，而标签“train”由于本身具有多义性，可以理解为“培训”，也可以认为是“火车”，对于这三个标签分别计算与语义主题的条件概率，记录下标签在不同语义主题下的相关度，结果如图 3-9 所示。

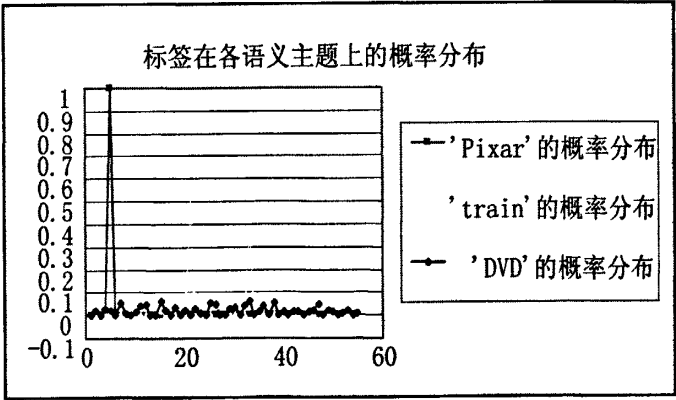


图 3-9 标签在潜在语义空间的概率分布

通过上图可以看出，标签“DVD”在各语义主题的分布较均匀，而标签“Pixar”仅仅与“语义主题5”潜有较大的相关性，标签“train”则与“语义主题21”和“语义主题33”均有较强的相关性。

通过实验证明本章提出的基于大众标注的语义主题发现方法，可以较好的发

现隐含在标注中的潜在语义，进而解决了标注的语义模糊问题。

3.4 本章小结

作为标签系统的最基本的组成元素，标签体现了用户的兴趣爱好，然而由于标签的语义模糊性，导致标签在推荐中的作用不能充分发挥，所以本章在详细分析大众标注的潜在语义关系的基础上，分析了语义主题与用户的关系，以及语义主题发现在标签系统的作用。通过语义主题发现，体现了用户标注背后所隐含的潜在语义，使得用户通过标签浏览，搜索，或者为用户推荐资源有了良好的语义基础。

总结基于的语义主题发现方法的优势主要有以下几个方面：

(1)基于大众标注的语义主题发现方法通过扩展的 PLSA 模型，分析用户、资源和标签出现的频率，绕开了自然语言的限制，得到了标注的潜在语义，并以概率的方式映射在各语义主题上，在用户语义主题范围内理解用户标签语义，解决了标签语义模糊问题，消除了标签的语义歧义。

(2)通过语义主题发现，发现用户标签与各语义主题的关系，同时得到用户对各语义主题的兴趣，为用户个性化服务提供了良好的基础。

总之，基于大众标注的语义主题发现方法较好的解决了标签系统中标签语义模糊问题，对于为用户提供更高质量的个性化服务有着很大的作用。

第四章 基于标签的个性化推荐方法

本章在语义主题研究的基础上，构建了基于语义主题的用户多兴趣模型，并提出了相应的更新策略。该模型较好的描述了用户多方面兴趣及其偏好程度，基于该模型的推荐提高了推荐精度，结合传统的项目评分，保证了推荐项目的质量。

4.1 基于语义主题的用户多兴趣建模

4.1.1 用户兴趣建模分析

用户兴趣模型最常采用的就是向量空间模型表示方法。这种方法表示和实现非常直观而且容易，应用于后续资源匹配阶段的向量计算非常方便。但是该方法无法准确地反映用户兴趣的语义信息，标签的出现将资源分类，准确、全面的描述资源的特征，如第二章的介绍，标签直接反应用户的兴趣偏好，是用户兴趣点的最佳描述，其特点是具有极强的个性。然而这种特质对于协同推荐有着很重大的影响。

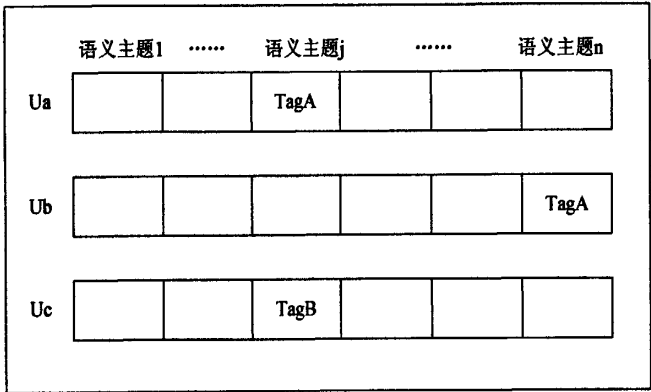


图 4-1 标签语义模糊

如图4-1所示，用户Ua和用户Ub有着相同的标签“TagA”，用户Uc有着与用户Ua不同的标签“TagB”。Ua的标签“TagA”和Ub的标签“TagA”虽然表面看起来是相同的，但是Ua的标签“TagA”所对应的语义主题是“语义主题j”，而Ub的标签“TagA”所对应的语义主题则是“语义主题n”，两者具有完全不同的意义。相反，表面看起来不相同的Uc的标签“TagB”却和Ua的标签“TagA”对应着相同的语义主题“语义主题j”。在这种情况下，如果没有考虑到这点，依然采用传统的协同过滤算法来进行推荐，那么系统就会把本身兴趣不一样的用户Ub当做相似用户，把用户Ub标注的项目资源推荐给Ua，而真正兴趣相同的Uc却没有被推荐。这样的推荐结果显然是不能使人满意的。

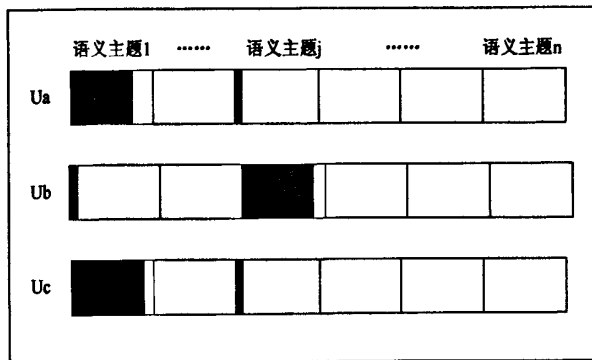


图 4-2 用户多兴趣

传统的协同过滤算法是基于单一的兴趣模型上计算的。但是，在现实生活中事实并不是这样。两个用户都会有相同和不同的兴趣偏好，有着相同和不同的擅长方面。如仅对财经信息感兴趣且擅长的人，为他人去推荐体育信息显然是不合适的，同样，希望获得体育信息推荐的人也不会找不了解体育信息的人去推荐。如图 4-2 所示，用户 Ua 和 Uc 显然在“语义主题 1”方面的兴趣度较高，对于“语义主题 j”方面兴趣度较低，而用户 Ub 刚好相反，如果不考虑用户对不同语义的兴趣度，将 Ub 在“语义主题 1”上的项目资源推荐给 Ua 显然是不合常理的。因此，比较两个用户的相似性是有前提限制的，即必须考虑不同用户不同资源的差异性，充分考虑到用户对不同语义主题的兴趣、擅长的差异性，这样的推荐才有足够的可信度。

因此在不同语义主题上描述用户的兴趣度，在不同的兴趣点上分别进行推荐是本文研究的重点之一。由于社会化标签是基于大众智慧的一种描述资源、分类资源工具，并且反应了用户的兴趣偏好，而语义主题是社会化标签在语义上的体现，因此，在理解标签语义的前提下，将语义主题用于用户用户兴趣建模非常适合。

4.1.2 基于语义主题的用户多兴趣建模

用户的兴趣特点是多样的，并且是动态变化的。仅仅将用户兴趣划分为“感兴趣”和“不感兴趣”两大类，就不能细致、充分的描述用户多个方面的多种兴趣特征，也不能及时跟踪用户兴趣的变化。针对这个问题，本章提出了一种基于语义主题的用户五元组多兴趣表示模型和一种动态的用户兴趣模型更新策略。该模型能够描述用户多方面的兴趣主题，以及用户对各兴趣主题的偏好程度，而意义明确的兴趣主题也避免了标签的语义模糊问题带来的干扰，更新策略用以捕捉用户兴趣的变化更新，提高了模型表达精度。

在标签系统中，一个用户可以使用多个标签，标注多个项目资源；一个标签可以被多个用户使用，可以用来标注多个项目资源；一个项目资源可以被多个用户标注，可以被多个标签标注。结合第三章的描述分析，用户、标签和项目资源之间都具有一定的语义关系，通过扩展的 PLSA 模型将其映射到潜在语义空间，利用其概率分布可以很好的描述用户对各语义主题的兴趣。因为每个

语义主题都是意义明确的且不交叉，较好的解决了标签语义模糊问题。因此，本章将用户感兴趣的语义主题作为用户兴趣主题，与该语义主题密切相关的标签作为用户该兴趣主题下的特征标签。构建了“用户-兴趣主题-特征标签”的三级管理方式，如图 4-3 所示。

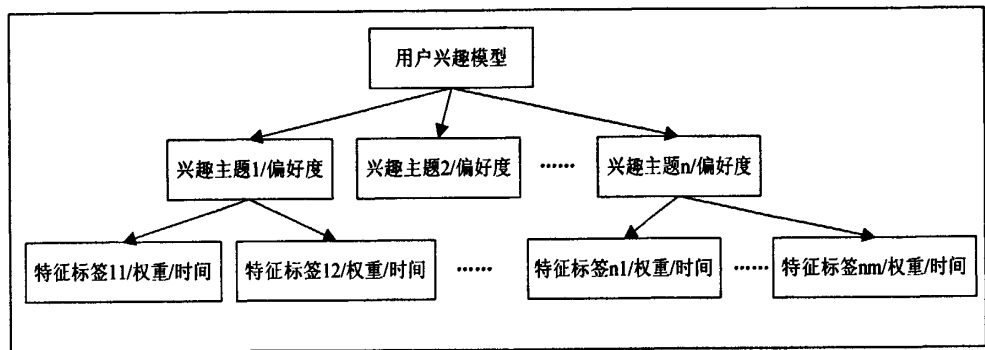


图 4-3 用户兴趣表示示意图

设用户有个不同领域的爱好，用户多兴趣模型 U 可采用如下五元组表示如式(4-1)

$$U = \{Z, T, W, H, P\} \tag{4-1}$$

其中， Z 为用户感兴趣的主题集合； T 是对应兴趣主题的特征标签集合即与对应兴趣主题密切相关的标签集合； W 是兴趣主题中相应特征标签的对应权重，初始值为与对应兴趣主题的相关程度，如对于用户 u_i ，标签 t_n 与语义主题 z_k 相关度表示为 $p(z_k | u_i, t_n)$ ； T 为特征标签权值变化更新时间，初始值为该标签被用户第一次使用的时间，主要用来更新用户兴趣变化； P 为该用户对兴趣主题的偏好程度，如用户 u_i 有 k 个兴趣主题，该用户对这 k 个兴趣主题的偏好程度表示为 $P_{u_i} = \{p(z_1 | u_i), p(z_2 | u_i), \dots, p(z_k | u_i)\}$ 。其中 $p(z_k | u_i)$ 表示用户 u_i 在与语义主题 z_k 的相关程度，即用户对兴趣主题 z_k 的偏好程度。

4.1.3 用户兴趣模型更新策略

用户的兴趣随着时间的推移而动态变化，为了保证推荐的有效性和实时性，所以用户兴趣模型也要随着用户兴趣的变化而进行相应的调整变化。从系统运行效率来考虑，如果用户兴趣模型过大，推荐效率就会随之降低，所以应及时更新用户不感兴趣的语义主题，调整添加用户新兴趣特征。

本章提出的用户多兴趣模型更新策略主要通过调整用户的兴趣主题和特征标签权重来体现用户兴趣的变化，其示意图如 4-4 所示。

(1)用户兴趣主题更新方法：当用户在通过标签浏览、搜索资源或标注资源时，系统检验该标签所对应的语义主题，同时对该用户现有的兴趣模型进行查询，如果用户兴趣模型中已经包含了该语义主题，则对该兴趣主题的权重进行修改。如果用户兴趣模型中不包括该语义主题，系统先检测用户兴趣模型的存储空间，如果有空间，则直接生成一个新的兴趣主题，对该兴趣主题下的特征标签进行聚类。如果空间已满，则先删除一个用户长期不关注的且用户偏好度最低兴趣主题，然后再生成新的兴趣主

题。

(2)特征标签权重更新方法: 对于标签系统来说, 一个最新标注的标签往往更能反应用户的兴趣点, 因此, 对标签时间因素加以分析和利用, 采用式(4-2)对标签的权值进行调整, 更新特征标签权重。

$$w'(T)=\frac{\alpha}{\alpha+(h-H)}+w(T) \tag{4-2}$$

其中 α 是调整系数, h 是当前时间, H 为兴趣主题中标签上一次修改或创建时间。 $w'(T)$ 是 $w(T)$ 更新后的值。调整系数 α 的值越大, 特征标签权重随时间降低的速度也就越快。的取值可以根据具体应用的需要来确定。

为了能够更加准备的理解用户意图, 系统还应该提供手动的用户兴趣模型更新功能, 让用户主动的编辑兴趣主题, 添加相应特征标签。

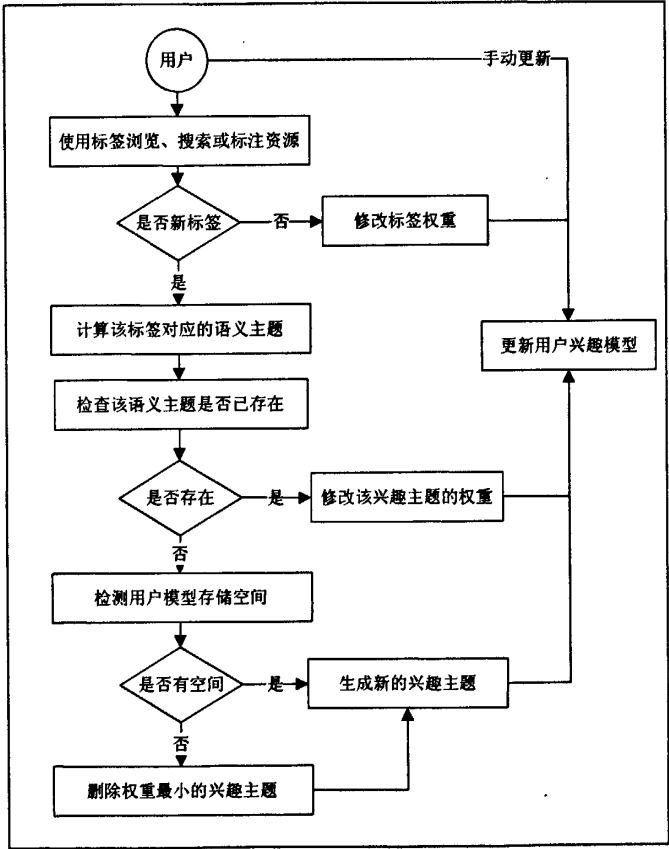


图 4-4 用户兴趣模型更新示意图

4.2 基于用户多兴趣模型和评分的协同过滤推荐算法

上节介绍了基于语义主题的用户多兴趣模型构建, 描述了用户的多方面兴趣偏好, 但是仅仅依靠标签语义是无法描述项目资源的质量。一个语义主题范围内可能有很多不同项目资源, 这些项目资源质量参差不齐, 显然用户希望系统推荐的资源是高质量的。本节在用户兴趣模型的基础上, 结合用户的项目评

分，使用用户兴趣模型预测用户兴趣偏好，结合项目评分为用户推荐高质量的项目资源，对于提高推荐的质量和精度有着很重要的意义。下面主要详细介绍这种算法的推荐流程和详细描述。

4.2.1 算法描述

本节提出的基于用户多兴趣模型和评分的协同过滤基本思想是：在用户多兴趣模型的研究基础上，结合项目评分，先为用户进行语义主题推荐，发现用户的潜在兴趣点(标签)；然后在所推荐标签的范围内为用户推荐项目资源，选择评分最高的项目资源作为推荐对象。

针对用户的相似度，提出了用户-语义主题的相关性矩阵，如图 4-5 所示。具体如下：

令 $U = \{u_1, u_2 \cdots u_n\}$ 代表 n 个用户的集合， $Z = \{z_1, z_2 \cdots z_m\}$ 代表 m 个语义主题， $p_{i,k}$ 表示用户 u_i 与语义主题 z_k 的相关度，从而得到一个 $n \times m$ 的矩阵，及用户-语义主题的相关矩阵。其中 u_a 为当前用户， z_k 为需要预测的语义主题，即需要预测的目标用户的兴趣， $p_{a,k}$ 为当前用户 u_a 与语义主题 z_k 的相关度。

	语义主题1	语义主题k	语义主题m
U1	P1,1		P1,k		
⋮					
Ua	Pa,1		Pa,k		Pa,m
⋮					
Un			Pn,k		

图 4-5 用户-语义主题相关性矩阵

基于用户多兴趣模型和评分的协同过滤推荐流程如图 4-6 所示：

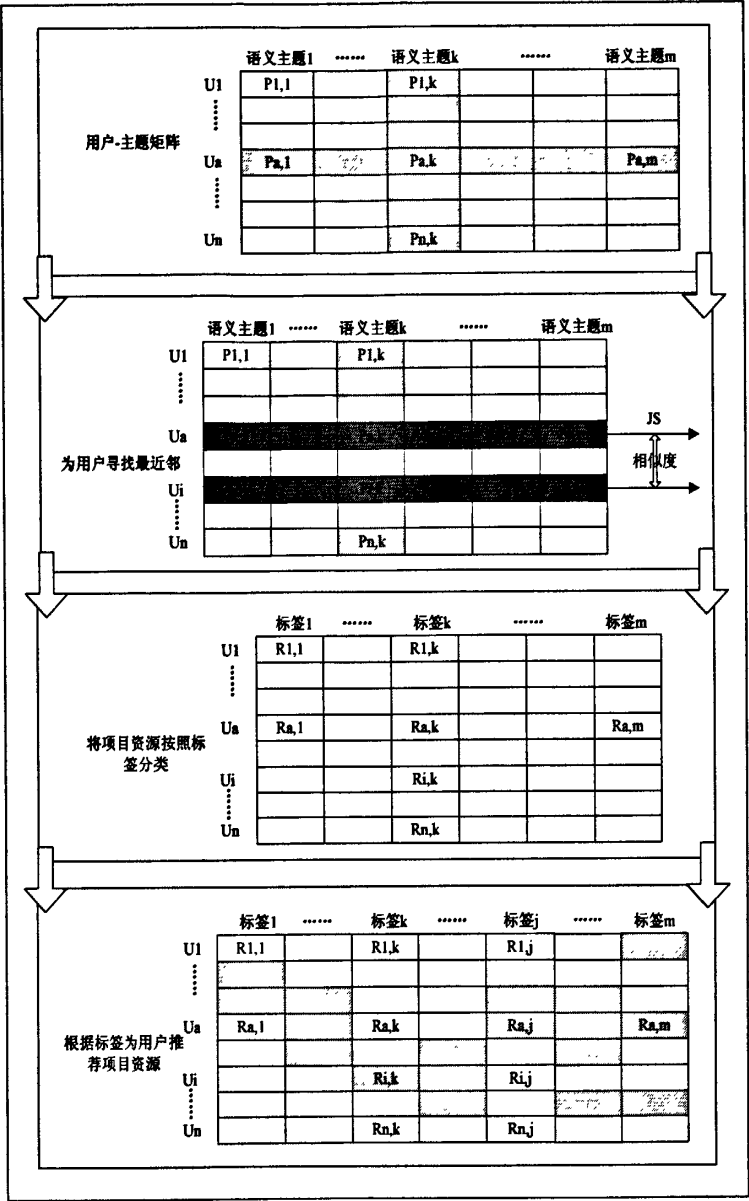


图 4-6 基于用户多兴趣模型和评分的协同过滤算法流程

由图 4-6 所示，基于用户多兴趣模型和评分的协同过滤方法主要分两步：

1. 为用户推荐标签。基于用户-语义主题相关性矩阵发现用户的兴趣，采用最近邻居的方法，选取相关性居前的邻居用户的兴趣主题中特征标签推荐给目标用户。

2. 基于各个推荐的标签生成预测。选择在各个推荐的标签对应评分最高的项目集合中为用户生成项目资源预测和推荐。

基于标签的协同过滤推荐算法描述如下：

输入：用户-项目评分矩阵和用户-语义主题相关性矩阵

输出：为目标用户 u_i 生成的项目资源推荐

(1) 推荐标签阶段：

Step1: 基于用户-语义主题相关矩阵，取用户 u_i 和用户 u_j 均有概率分布的

语义主题的交集, 获得语义主题的集合 $Z_{i,j}$ 。

Step2: 在集合 $Z_{i,j}$ 中, 计算目标用户 u_i 与用户 u_j 之间的用户-语义主题相关度向量的相似度, 本文采用 JS 相似度算法^[43], 该算法在概率分布中计算相似度有着更好的准确性^[44], 则用户 u_i 和用户 u_j 的相似度可以用式(4-2)计算。

$$\begin{aligned} D_{js}(u_i, u_j) &= \text{distance}(UZ_i, UZ_j) \\ &= \frac{1}{2} [D_{KL}(u_i \parallel \text{avg}(u_i, u_j)) + D_{KL}(u_j \parallel \text{avg}(u_i, u_j))] \end{aligned} \quad (4-2)$$

Step3: 选择另一个用户-语义主题相关向量, 重复 Step1、Step2, 直到获得用户 u_i 和所有用户的相似度集合 $S = \{D_{js}(i,1), D_{js}(i,2), \dots, D_{js}(i,m)\}$ 。

Step4: 采用 Top-N 方法, 从集合 S 中选取 N 个相似度最高的用户最为目标用户的最近邻居, 得到最近邻居集合 $U_i = \{u_1, u_2, \dots, u_n\}$ 。

Step5: 选择最近邻居用户 u_j 的用户-语义主题向量, 选择相关度最高的几个语义主题作为候选推荐语义主题集合 Z'_i 。

Step6: 选择其他的最近邻居的用户-语义主题向量, 重复 Step5, 直到获取所有最近邻居的候选推荐语义主题集合。

Step7: 采用并集合并所有的候选推荐语义主题。合并的时候遇到相同的语义主题, 合并的同时保留与之相似度最高的用户信息, 即兴趣度最高的, 在这个语义主题下最具权威的用户, 将最终得到的集合 Z_i 。

Step8: 选择 Z_i 中对应的且不存在与目标用户兴趣模型中的用户特征标签组成集合 T_i 推荐给目标用户 u_i 。

(2)推荐项目阶段:

Step1: 选择给目标用户 u_i 推荐的特征标签集合 T_i 中其中一个特征标签 $T_{i,j}$ 和该特征标签所对应的最近邻居 u_j 所标注的项目资源集合, 选择用户评分最高的 N 个项目资源, 得到集合 R_j 。

Step2: 在集合 R_j 中, 选择一个目标用户未标注的项目资源 r_1 , 计算 r_1 在最近邻居 u_j 中所对应语义主题 Z_j 的相关度 $S_{i,j}$ 。计算公式如式(4-3)所示。

$$S_{i,j}^1 = P(z_j | r_1, u_j) \quad (4-3)$$

Step3: 选择集合 R_j 中其他的目标用户未标注的项目资源, 重复 Step2, 得到项目资源与该语义主题的相关度集合 $S_{i,j}$, 采用 Top-N 的方法, 取相关度最高的 N 的项目资源作为候选推荐项目。

Step4: 选择 Z_i 中所有的语义主题, 重复 Step2, Step3, 得到所有推荐项目的并集 R_i 推荐给目标用户 u_i 。

4.2.2 算法分析

标签系统中的社会化标签的特殊性是在于它体现了用户的兴趣点和项目的特征, 它作为一个桥梁将用户和项目紧密联系起来。然而考虑到标签本身缺乏

语义性，存在着语义模糊等问题，所以本章在第三章研究的基础上，构建了基于语义主题的用户多兴趣模型，在描述用户标签语义的同时，描述了用户多兴趣及对各兴趣的偏好程度，结合评分，为用户推荐高质量的项目资源，将基于用户的协同推荐和项目评分结合起来，使两种方法的优势都得到发挥，而且形成互补。归纳基于用户多兴趣主题和评分协同过滤推荐方法的优势主要有以下几个方面：

(1)有效减小计算的规模，提高了推荐效率。用标签将项目资源分类，一个标签所对应的项目集要远远小于整个项目集，减少了计算的次数；通过第三章介绍的潜在语义分析可知，语义主题是在一个有限维的潜在语义空间，因而通过用户多兴趣模型进行用户相似度计算时降低了计算维数。这两方面结合提高了本文的推荐方法的计算效率。

(2)发现用户潜在兴趣，有效避免标签语义模糊及单一兴趣模型对推荐准确性的影响。本章提出的推荐方法首先构建用户多兴趣模型并获得用户多个兴趣主题以及对每个兴趣主题的偏好程度，对于每个兴趣主题用与之语义相关的特征标签描述。再将每个兴趣主题作为一个推荐模型产生推荐，得到了用户的已知的和潜在的兴趣点，使用特征标签将项目分类，使得推荐的项目可以涵盖用户的多个兴趣，这样的推荐更有针对性，与用户的期望更吻合。同时兴趣模型的更新策略使得推荐的实时性和有效性得到保证。

(3)结合评分的推荐保证推荐质量。传统基于标签的推荐没有考虑到项目资源本身的质量问题，本文提出的方法在对项目资源进行推荐时，将评分作为项目资源质量好坏最直观的表现，选择对应标签范围内评分最高的项目资源进行推荐，使得对目标用户所推荐的项目资源质量得到保证。

总之，本章基于标签的个性化推荐方法可以有效的弥补在标签系统中传统推荐方法的不足，对于改善推荐效率和准确性有很大的帮助。

4.3 实验验证

评价推荐算法有很多指标，通过它们来度量或评价一个推荐算法的优劣。本文使用了推荐命中率指标 HR 指标和命中排序 ARHR 作为评价算法的依据^[45]。命中率(HitRate, HR)即评估为用户推荐的 N 个项目中命中测试数据集中评分较高的项目的次数。ARHR(Average Reciprocal Hit Rank)用于评估命中项目在推荐排序集合中的位置。

HR 的计算公式如式(4-4)所示。

$$HitRate = \frac{hit}{N} \quad (4-4)$$

hit 表示推荐系统推荐的项目集中命中的项目个数，N 表示项目集中项目的总个数。可见 HR 值越大，说明推荐系统推荐的项目命中就越多，推荐的效果就越好，推荐系统的可信度就越高。

ARHR 的计算公式如式(4-5)所示。

$$ARHR = \frac{1}{N} \sum_{i=1}^{hit} \frac{1}{pos_i} \tag{4-5}$$

pos 表示命中项目在推荐集中的位置。hit 表示命中的项目个数，N 表示项目集中项目的总个数。ARHR 越大，说明命中的项目在项目集中位置越靠前，推荐时优先推荐这些命中项目，推荐效果越好。

在第三章实验的基础上，随机选取 20% 的数据做测试数据，剩下 80% 做训练数据。随机选取 10 个用户，在 MovieLens 训练数据集上，使用本文提出的用户兴趣模型和评分的协同过滤算法和传统的基于标签协同推荐算法分别对他们生产项目推荐。在最终推荐的项目中，如果其中的项目用户实际评分是 5 分或对其进行标注的称为一次命中。分别记录这 10 个用户的预测和推荐，进行 HR 和 ARHR 评估。取平均值使用柱状图表示，进行对比。分别记录 Top-5 和 Top-10 下 HR 和 ARHR 指标评价结果如图 4-7 和 4-8 所示。

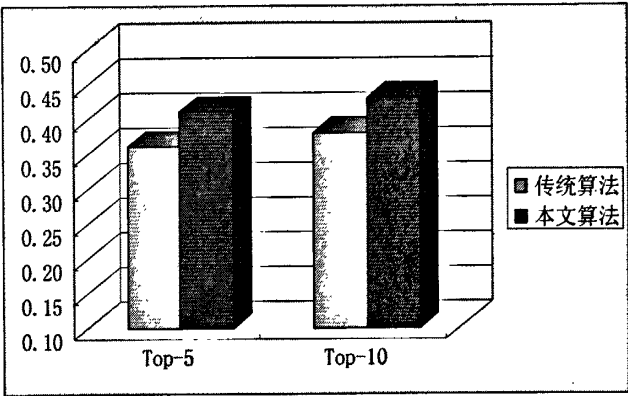


图 4-7 HR 评价指标对比图

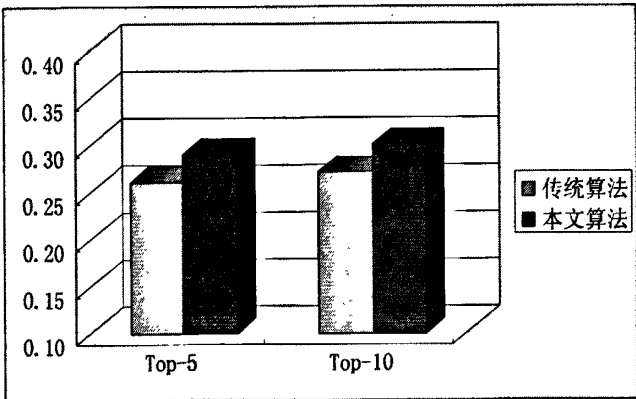


图 4-8 ARHR 指标评价对比图

由图 4-7 和图 4-8 可以看出，本文提出的改进的协同过滤算法的 HR 和 ARHR 的值要大于传统的基于社会化标签的协同过滤算法。由于 HR 和 ARHR 的值越大越好，值越大项目推荐的命中率就越高，推荐的项目的质量就越高，

就越能满足用户的兴趣需要。所以，本文提出的改进的协同过滤算法在推荐项目的质量上要明显好于传统的基于社会化标签的协同过滤算法。

4.4 本章小结

本章首先针对传统用户兴趣模型中的问题，分析了标签系统中用户兴趣模型的特点，构建了基于语义主题的用户多兴趣模型，提出了模型的更新策略。在此基础上，提出了改进的基于用户多兴趣模型和评分的协同过滤算法，详细阐述了改进算法的推荐过程和算法描述，并对这种改进算法加以详细的分析。最后在第三章实验的基础上，对本章推荐方法与传统方法进行了对比实验，实验证明了本章方法要优于传统方法。

第五章 原型系统实现

5.1 背景和目的

网上购物的盛行给图书销售带来了新的机遇，网上书店的出现使得人们可以方便快捷的用户查询到自己所关注的图书信息,实现在线购买。据某知名图书销售网站公布，仅 2010 年该网站的净销售额就达到了 342 亿美元，同比 2009 年增长 40%。人们消费观念的不断转变，网上购物的方式被越来越多的人认可，网购用户的持续增长为图书网上销售带来了更加广阔的发展前景。面对浩如烟海的书籍信息，推荐系统根据用户操作判断用户需求，进行主动推荐，提高了系统的服务能力，给网上图书销售带来巨大的效益。随着互联网的发展，标签技术的出现为个性化推荐服务带来了至关重要的数据源，可以更加准确的获取用户的特征，为用户推荐其“量身定做”的信息。然而随着应用的不断深入，人们发现基于标签的推荐也存在着某些问题：(1)语义模糊的标签难以描述用户的真实意图使得推荐精度不高。(2)海量标签降低了推荐效率。(3)推荐的图书质量得不到保证。

为了解决上述问题，本章结合本文第三章和第四章介绍的方法，设计并开发了基于标签的图书推荐系统，介绍了推荐系统的体系结构和推荐流程，并对其进行了比较完整的设计，实现了系统总体框架中的大部分核心功能，实现了个性化推荐服务，并通过实验验证了系统的部分性能与实现效果。系统使用的主要数据均来自豆瓣读书。

5.2 系统体系结构及工作流程

本章对基于标签的图书推荐系统进行了完整的系统设计，并实现了系统总体框架中的部分核心子系统，具有良好的运行效果。本系统在体系结构设计上划分为三个主要的模块：人机交互模块，中间模块和推荐模块，如图 5-1 所示。

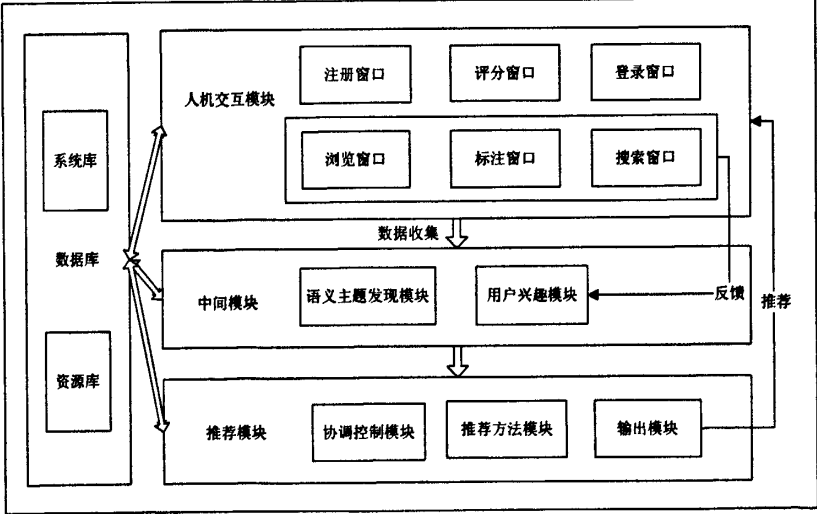


图 5-1 基于标签的图书推荐系统架构图

(1) 人机交互模块

人机交互模块是系统在对用户推荐时提供方便的接口,其中包括注册窗口、登录窗口、评分窗口、浏览窗口、标注窗口和搜索窗口,通过这些窗口记录用户操作,为中间模块和数据库提供基础数据,浏览窗口接受输出模块推荐的图书展示给用户,并通过标注窗口获取用户反馈信息。

(2) 中间模块

中间模块通过人机交互模块接受用户数据,完成语义主题挖掘,并分析用户兴趣,构建用户兴趣模型。该模块包括两个主要的子模块:①语义主题发现模块。本模块一般为离线完成,整理用户标注信息,通过分析标注的潜在语义,挖掘用户语义主题,其过程如本文第三章所述。②用户兴趣模块。包括用户兴趣构建和更新两部分,用户兴趣构建接收语义主题发现模块提供的用户偏好主题,对用户兴趣模型三级管理,如本文4.2.2节描述。更新模块根据用户行为进行实时更新,策略如4.2.3节所示。

(3) 推荐模块

推荐模块是整个检索系统的核心,提供对用户的个性化推荐服务,包括协调控制模块、推荐方法模块和输出模块三个子模块。协调控制模块包括对用户操作记录的分析,在特定情况下触发相应的功能事件,根据不同用户选择不同的推荐方法。推荐方法模块根据协调控制模块调用不同的推荐算法,其中包括了本文提出的协同推荐方法,当某个目标用户活动时,对该用户启动该模块,返回N个推荐给目标用户的图书。返回的图书由输出模块展示给用户。

图 5-2 描述了基于社会化标签的图书推荐系统的工作流程。

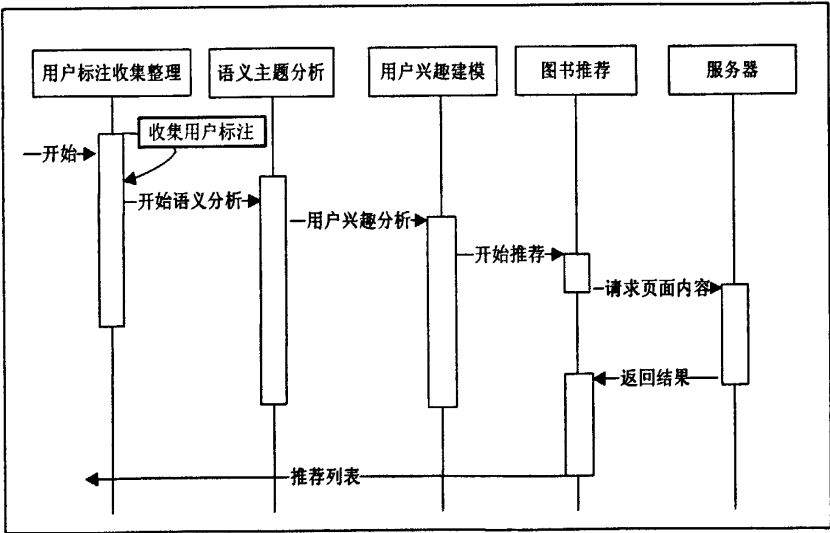


图 5-2 基于标签的图书推荐系统工作流程

第一步,人机交互模块收集用户的原始数据,包括标签、标注时间等,这些原始数据是对用户进行推荐的数据来源,进过预处理后存储到数据库中,并交给中间模块进行语义分析。

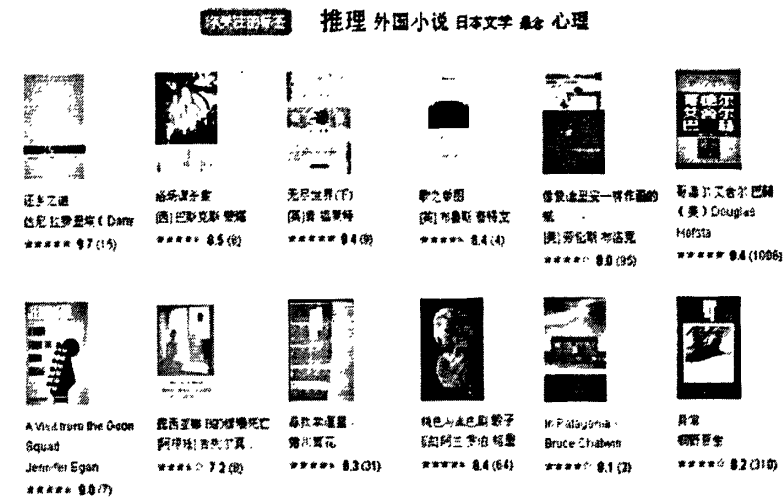


图 5-4 系统推荐页面

用户点击推荐的图书图片可以看到该图书详细信息介绍，对于该图书用户可以对其进行浏览、评价、评分和标注等。这些操作会被记录在数据库中，用来更新用户模型。如图 5-5 所示。

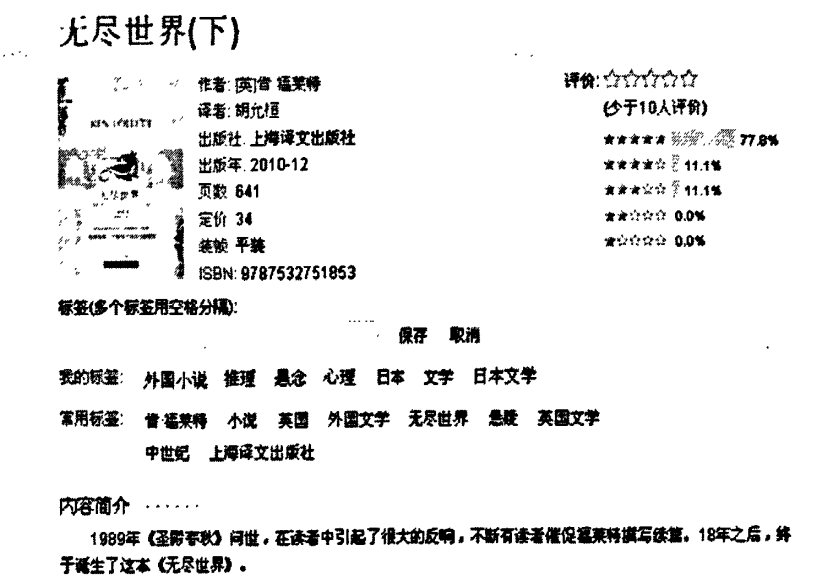


图 5-5 用户系统交互界面

5.4 实验与分析

为了验证本系统的推荐效果，挑选了两个不同兴趣爱好的用户使用本系统，其中一个用户比较偏好文学类的图书，而另一位用户则对科技类的图书比较感兴趣，首先建立一个账户，邀请对偏好文学类图书的用户使用本系统，两周后邀请偏好科技类图书的用户使用同一个账号继续使用本系统两周，通过对这两

个用户在四周内的连续跟踪观察，其兴趣的变化统计直方图如图 5-6 所示：

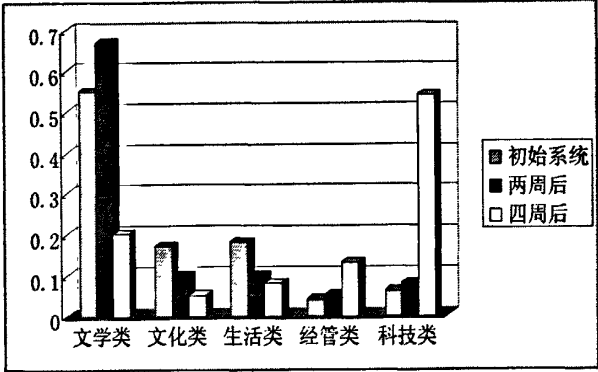


图 5-6 兴趣变化统计直方图

从图 5-7 可以看出，随着用户浏览内容的变化，用户不同类别的兴趣度也随之变化。从图形反映的情况来看，一周后，用户的兴趣比较集中于文学类，而两周之后，用户的兴趣则偏向于科技类。由此可以看出，基于本文推荐方法的系统能基本反映用户的兴趣爱好及其变化情况，达到比较好的效果。

为了更好地验证本文提出方法对个性化信息推荐的效果的影响，挑选 5 名用户分别在本系统上和普通推荐系统上进行连续两周使用，并邀请用户对系统推荐的信息进行评价，使用正确率作为评价依据，即系统推荐且用户感兴趣的图书数量与系统推荐图书总数的百分比。分别记录 5 名用户对系统推荐的评价，其推荐图书的正确率对比如图 5-7 所示。

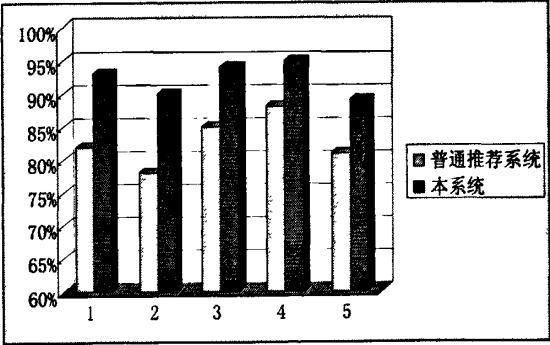


图 5-7 用户评价系统推荐正确率对比图

从图 5-7 可以看出，在本系统上，由于进行了用户语义主题挖掘并构建了用户多兴趣模型，个性化信息推荐效果明显要优于普通的推荐系统，

5.5 本章小结

本章以图书推荐为切入点，结合本文第3章和第4章的研究，设计并开发了一个简单的图书个性化推荐系统，介绍了开发该系统的背景和目的，给出了系统的体系结构，介绍了系统的工作流程，实现了系统的核心功能，介绍了系统的具体运行流程。

第六章 结论与展望

6.1 结论

目前互联网上的信息资源呈现爆炸式的增长,导致了“信息过载”等问题的出现,个性化推荐技术在一定程度上满足了人们不同兴趣爱好的个性化需求,标签的出现使得人们可以自由的对网络资源进行标注,标签将用户和资源联系起来,描述了资源的特征,体现了用户的兴趣偏好,为个性化推荐服务带来了至关重要的数据源。然而由于标签本身的社会性,存在着语义模糊问题,降低了基于标签的推荐精度,极大的限制了标签在推荐上的应用,同时用户和标注的急剧膨胀降低了推荐效率。针对上述问题,本文做了以下工作:

(1)对大众标注的语义主题发现进行了研究。通过对大众标注潜在语义的研究,提出了基于大众标注的语义主题发现方法,消除了标签语义模糊问题。通过分析标注中用户、资源和标签的潜在语义关系,引入 PLSA 模型并对其进行扩展,通过语义主题发现方法获得用户标注的潜在语义,将用户标注映射到具有明确意义的语义主题上,该方法消除了标签语义模糊问题,同时发现用户对各语义主题的关注度。

(2)对用户兴趣模型进行了研究。在对语义主题研究的基础上,构建了三级的基于语义主题的用户多兴趣模型,并提出了更新策略捕捉用户兴趣变化,保证推荐的实时性和有效性。该模型较好的描述了用户兴趣特征,基于该模型的个性化推荐提高了推荐精度。

(3)对基于用户兴趣模型和评分的协同推荐算法进行了研究。在用户兴趣模型研究的基础上,提出了基于用户多兴趣模型和评分的混合协同推荐算法,解决了标签语义在推荐中的问题,提高了推荐的精度和效率,结合项目评分,保证了推荐的项目质量。

(4)最后,以图书推荐为切入点,设计并实现了一个简单的基于标签的图书个性化推荐系统,并通过实验验证了系统的部分性能与实现效果。

6.2 展望

本文在标签系统的个性化推荐上进行了研究,提出了基于大众标注的语义主题发现方法和基于用户兴趣和评分的协同过滤推荐算法,并通过实验验证了其有效性。然而由于时间和条件的限制,对于标签系统的个性化推荐研究还有待于不断完善,今后需要做进一步的研究,包括:

(1)本文提出的基于大众标注的语义主题发现方法,未考虑其更新方法,如何使用一种简单有效的方法,更新语义主题值得在以后的研究中考虑。

(2)语义主题虽然描述了用户标签的语义,但是不足以反映标签所表示的资源属性,因此在未来的研究中将结合本体对其进行扩展。

(3)虽然通过语义主题发现可以较好的解决标签语义模糊问题,但是对于一

个标签系统来说，如何提供一种机制使用户使用标签更规范，使标签准确表征项目特征和用户兴趣的同时，尽可能的减小标签语义的重复，发挥它更大的作用，也是值得深究的问题之一。

参考文献

- [1]石静.基于混合模式的个性化推荐系统的应用研究[D]. 硕士学位论文.2010.
- [2]杨杰.个性化推荐系统应用及研究[D].硕士学位论文.2009.
- [3]David Goldberg,David Nichols,Brian M. Oki,Douglas Terry.Using collaborative filtering to weave an information tapestry[J]. Communications of the ACM - Special issue on information filtering. 1992.12(35). ACM New York, NY, USA.
- [4]魏建良,朱庆华.基于社会化标注的个性化推荐研究进展[J].情报学报. 2010.4(29): 625 - 633.
- [5]Niwa S,Doi T,Honiden S.Web Page Recommender System Based on Folksonomy Mining for ITNG'06 Submissions [C] // Proceeding ITNG '06 Proceedings of the Third International Conference on Information Technology: New Generations . IEEE Computer Society Washington, DC, USA .2006:388- 393.
- [6]Ae-Ttie Ji, Cheol Yeon, Heung-Nam Kim ,Geun-Sik Jo.Collaborative Tagging in Recommender Systems[J]. AI 2007: Advances in Artificial Intelligence.Lecture Notes in Computer Science, 2007.4830:377-386.
- [7]Diederich J,Iofciu T.Finding Communities of Practice from User Profiles Based on Folksonomies [C] //Proceedings of the 1 st International Workshop on Building TechnologyEnhanced Learning Solutions for Communities of Practice. Crete. Greece. 2006.
- [8]Sasaki A,Miyata T,Inazumi Y,et al. A Study of a Growing Meta-data Model for Content Recommendation on Folksonomies [J] .IEIC Technical Report,2006, 150 (106): 67-72.
- [9]Yeung C A,Gibbins N,Shadbolt N . A Study of User Profile Generation from Folksonomies [C]//Workshop on Social Web and Knowledge Management at the 17th International Conference on World Wide Web.Beijing,China,2008.
- [10] Karen H. L. Tso-Sutter, Leandro Balby Marinho, Lars Schmidt-Thieme.Tag-aware Recommender Systems by Fusion of Collaborative Filtering Algorithms [C] //Proceeding SAC '08 Proceedings of the 2008 ACM symposium on Applied computing.New York:ACM,2008:1995- 1999.
- [11]Yanfei Xu,Liang Zhang. Personalized Information Service Based on Social Bookmarking [C] // Tanaka K,Xing C,Goh D H.al.Digital Libraries:Implementing Strategies and Sharing Experiences.Berlin:Springer,2005:475- 476.
- [12]Yanfei Xu, Liang Zhang,Wei Liu.Cubic Analysis of Social Bookmarking for Personalized Recommendation [C] // Tanaka K,Yu Y,Zhang Y,et al. Frontiers of WWW Research and Development - APWeb 2006. Berlin:Springer,2006:733- 738.
- [13]Cameron Marlow, Mor Naaman, Danah Boyd, Marc Davis. HT06, tagging

- paper, taxonomy, Flickr, academic article, to read [C] //Wiil U K,Nurnberg P J,Rubart J.Proceedings of the seventeenth Conference on Hypertext and Hypermedia.New York:ACM,2006:31-40 .
- [14] Elke Michlmayr, Steve Cayzer.Learning User Profiles from Tagging Data and Leveraging them for Personal (ized) Information Access[C]//Workshop on Tagging and Metadata for Social Information Organization at the 16th International Conference of World Wide Web.Banf,Canada 2007.
- [15]Zi-Ke Zhang, Tao Zhou,Yi-Cheng Zhang. Personalized recommendation via integrated diffusion on user-item-tag tripartite graphs[J]. Physica A: Statistical Mechanics and its Applications. 2010.1(389): 179-186.
- [16]Szomszor M,Cattuto C,Alani H,et al.Folksonomies the Semantic Web and Movie Recommendation[C]//Proceedings of 4th European Semantic Web Conference.Bridging the Gap between Semantic Web and Web 2.0. Innsbruck. Austria.2007.
- [17] Anon Plangprasopchok, Kristina Lerman.Exploiting Social Annotation for Automatic Resource Discovery [C] //Nambiar U,Nie Z.Proceedings of AAAI workshop on Information Integration from the Web.Menlo Park:The AAAI Press, 2007: 86-91.
- [18]Eck D,Lamere P, Bertin- Mahieux T,et al.Automatic Generation of Social Tags for Music Recommendation[C]//Twenty- First Annual Conference on Neural Information Processing Systems Conference.Vancouver.Canada.2007.
- [19]刘向红,宋 文,姚朋.基于标签的 Folksonomy 机制研究—以 CiteUlike 为例[J].图书馆理论与实践.2010(5).29-33.
- [20]易明.基于 Web 挖掘的个性化信息推荐[M].第一版.北京:科学出版社, 2010: 136-139.
- [21]Scott A. Golder and Bernardo A. Huberman. The structure of collaborative tagging systems[J]. Journal of Information Sciences 32(2):198–208.
- [22]George W. Furnas, Caterina Fake, Luis von Ahn, Joshua Schachter, Scott Golder, Kevin Fox, Marc Davis, Cameron Marlow, Mor Naaman.Why do tagging systems work?[C]// Proceeding CHI EA '06 CHI '06 extended abstracts on Human factors in computing systems.ACM New York, NY, USA.2006.
- [23]黄建年, 侯汉清. Tag 分类基本问题探究[J].信息系统.2008.3(31).461-465.
- [24]Grigory Begelman.Automated Tag Clustering: Improving search and exploration in the tag space[C]//In Proc. of the Collaborative Web Tagging Workshop at WWW'06,2006.
- [25]Resnick P,Iacovou N,Suchak M,et al.Grouplens:An open architecture for

collaborative filtering of netnews[C]//Proc.of the ACM CSCW'94 Conf.on Computer- Supported Cooperative Work.Chapel Hill:ACM,1994:175- 186.

[26]常富洋,林鸿飞,许侃. 基于用户向量扩展的协同推荐方法[J]. 情报学报. 2010. 4:688-694.

[27]Sarwar B, Karypis G, Konstan J, Riedl J. Analysis of recommendation algorithms for e-commerce[C]//Proceedings of 2nd ACM conference on Electronic Commerce, Minneapolis, Minnesota, USA: ACM, 2000: 158-167.

[28]郭红艳. 推荐系统的协同过滤算法与应用研究[D].博士学位论文.2008.

[29]宣照国,苗静,党延忠. 基于扩展邻居的协同过滤算法[J].情报学报.2010. 29(3). 443-448.

[30]李聪,梁昌勇. 基于属性值偏好矩阵的协同过滤推荐算法[J].情报学报.2008. 27(6). 884-890.

[31]曾春,邢春晓,周立柱. 基于内容过滤的个性化搜索算法[J]. 软件学报. 2003. 14(5):999-1004.

[32]岑咏华,甘利人,丁晟春. 基于内容的 Web 个性化推荐技术研究[J]. 图书情报工作. 2003.8.30-34.

[33]崔春生,吴祈宗. 基于 Vague 集的内容推荐算法研究[J]. 2010. 27(6). 2109-2111.

[34]Michael J. Pazzani. A Framework for Collaborative, Content-Based and Demographic Filtering[J]. Artificial Intelligence Review, 1999,13(5-6): 393-408.

[35]Marko Balabanović, Yoav Shoham. Fab: Content-Based, collaborative recommendation[J]. Communications of the ACM, 1997,40(3):66-72.

[36]许海玲,吴潇,李晓东,阎保平.互联网推荐系统比较研究[J].软件学报. February 2009.2(20):350-362.

[37]Thomas Hofmann.Probabilistic latent semantic analysis[C]//Uncertainty in Artificial Intelligence, Berkeley:ACM,1999.289-296.

[38]唐朝辉.潜在语义分析在互联网数据挖掘中的应用研究[D].硕士学位论文.2009.

[39]Paul Resnick ,Hal R. Varian. Recommender systems[J].Communications of the ACM. 1997.3(40) .New York, NY, USA.

[40]J. Ben Schafer, Joseph Konstan, John Riedi. Recommender systems in e-commerce [C]//Proceeding EC '99 Proceedings of the 1st ACM conference on Electronic commerce . ACM New York, NY, USA 1999.

[41]J. Ben Schafer, Joseph A. Konstan and John Riedl.E-Commerce Recommendation Applications[J].Data Mining and Knowledge Discovery.2001. 5.115-153.

[42]刘颖. 电子商务推荐系统研究——新用户和用户兴趣变化问题研究[D]. 硕

士学位论文.中国科技大学.2007.

[43]Lillian Lee.On the Effectiveness of the Skew Divergence for Statistical Language Analysis[C]//Artificial Intelligence and Statistics.2001:65-77.

[44]Takeharu Eda, Masatoshi Yoshikawa,Toshio Uchiyama,Tadasu Uchiyama.The Effectiveness of Latent Semantic Analysis for Building Up a Bottom-up Taxonomy from Folksonomy Tags[J]. Computer Science.2009.12(4):421-440.

[45]刘建国,周涛,郭强,汪秉宏.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):1-10.

[46]李以正.基于用户反馈与协同过滤的情报检索系统的建立[J].情报杂志.2007.2.129-134.

[47]余小高,余小鹏.基于隐式评分的推荐系统研究[J].计算机应用.2009.6. 1585-1589.

[48]曾小波.基于协同过滤的系统的研究[D].硕士论文.2010.

[49]陶剑文,潘红艳.基于相似项目与用户评分的协同推荐算法[J].情报学报. 2008.27(2).199-204.

[50]陈逸,于洪.一种基于相同评分矩阵的协同过滤补值算法[J]. 计算机应用研究.2009.26(12).4513-4519.

[51]Junyan Zhu, Can Wang, Xiaofei He, Jiajun Bu, Chun Chen, Shujie Shang, Mingcheng Qu, and Gang Lu. Tag-Oriented Document Summarization[C]// WWW 2009, Madrid, Spain.1195-1196.

[52] Heung-Nam Kim*, Ae-Ttie Ji, Inay Ha, Geun-Sik Jo. Collaborative filtering based on collaborative tagging for enhancing the quality of recommendation[J]. Electronic Commerce Research and Applications .2010.9. 73-83

攻读硕士学位期间发表的论文和参加的科研项目

1. 蒋翠清,张玉,丁勇.基于 PLSA 的大众标注潜在语义发现.现代图书情报技术. 2010(10). 28-32.

作者：[张玉](#)
学位授予单位：[合肥工业大学](#)

本文读者也读过(10条)

1. [夏宁霞](#), [苏一丹](#), [覃华](#), [张敏](#), [XIA Ning-xia](#), [SU Yi-dan](#), [QIN Hua](#), [ZHANG Min](#) [社会化标签系统中个性化的用户建模方法](#)[期刊论文]-[计算机应用](#)2011, 31(6)
2. [赵英海](#), [查正军](#), [李珊珊](#), [吴秀清](#), [ZHAO Yinghai](#), [ZHA Zhengjun](#), [LI Shanshan](#), [WU Xiuqing](#) [基于视觉内容语义相关度的图像标签自动排序方法](#)[期刊论文]-[中国科学技术大学学报](#)2011, 41(2)
3. [李欢](#) [银联商务内蒙古分公司银行卡收单业务竞争战略研究](#)[学位论文]2011
4. [李微娜](#) [面向用户偏好的MADM方法在个性化推荐中的应用](#)[学位论文]2011
5. [陈永利](#) [新时代信托股份有限公司风险管理研究](#)[学位论文]2011
6. [李坤元](#) [基于证据推理的家用轿车综合评价研究](#)[学位论文]2011
7. [于长伟](#) [控车床热误差补偿建模研究](#)[学位论文]2011
8. [冯瑞](#) [混合动力城市客车参数优化及能量回馈研究](#)[学位论文]2011
9. [王洪成](#) [电磁阀式连续可变阻尼减振控制系统设计与开发](#)[学位论文]2011
10. [王鑫](#) [数控机床热误差测试及建模理论与方法的研究](#)[学位论文]2011

本文链接：http://d.g.wanfangdata.com.cn/Thesis_Y1887099.aspx