

机会发现理论对推荐系统的改进

□ 裴倩 / 西南科技大学经济管理学院 绵阳 621010

王玉平 / 西南科技大学图书馆 绵阳 621010

摘要: 文章分析推荐系统的现状和存在问题, 引入关注动态变化、发现隐性关联和弱信号的机会发现理论, 依次从收集用户信息、掌握用户兴趣变化和匹配资源方面, 探讨机会发现理论对推荐系统的改进。通过改进兴趣模型、建立自适应推荐模块、提供集成推荐等措施, 提升推荐系统效率。

关键词: 机会, 机会发现, 推荐系统

DOI: 10.3772/j.issn.1673—2286.2012.03.011

随着Web 2.0理念的深入, 社交网络、Blog、微博等异军突起, 去中心化催生大量文本、图片、视频等信息, 信息质量良莠不齐, 大量冗余、无用、重复信息的存在导致信息更为琐碎、同质, 使得个人面临的选择越来越多。以搜索引擎和推荐系统为代表的信息过滤机制随之应运而生, 通过统计用户记录, 对用户、信息(或商品)分组的关联和相似度给出推荐。

目前的推荐系统中, 数量较少、小众需求的弱信号往往当作噪声过滤或忽视^[1]。对于人追求新鲜、期待惊喜的本性, 现有的依靠内容、对象关联的过滤机制给出的往往是一些相似的东西, 无法制造惊喜。相比在卓越、当当里接受一些流行热卖的相似图书的推荐, 很多人更愿意逛书店^[2]。同时对于竞争者来讲, 大量同质的信息是没有价值的, 一定是稀有潜在偶然的信息才能支持决策, 比如专利和竞争情报。

日本学者Yukio Ohsawa认为, 在这个瞬息万变的社会中, 应关注

一些新的罕见的事件或状态, 因为这些事件或状态将会对决策的制定产生影响。他将这种事件或状态定义为机会, 并将发现、处理机会的过程定义为机会发现(Chance Discovery)^[3,4]。从信息推荐的目标来讲, 机会发现将弱信号的小众需求和隐性相关资源建立联系, 深入挖掘用户潜在需求和偏好, 在智能、精准的基础上扩大用户视野, 从而改善推荐质量。

1 机会发现

1.1 机会发现的定义

2000年, 日本学者Yukio Ohsawa针对于决策与任务的实际需求差距提出了机会发现, 其主旨是在动态未知预测环境中, 发现对主体决策具有重要影响的事件或状态, 对于那些有利于主体的机遇(opportunity)加以利用, 对那些不利于主体的风险(risk)加以规避^[3,4]。继而, 人工智能、心理学、经济学、系统论和控制论等领域纷纷开

始机会发现在本领域的研究。

机会发现是对机会的重要性产生认识或解释, 特别是机会很少出现并且它的重要性没有被注意到的时候, 提供一种思路 and 手段。通过某个事件的执行来达到更好地支持决策, 其目标是关注机会, 为将来的创新和存活提供手段, 不是预测机会。

1.2 机会发现的工具

1.2.1 双螺旋模型

Ohsawa提出的双螺旋模型^[5](见图1)描述了数据挖掘和人机会的把握两个并行上升分支间的人机交互促进了机会的发现。数据挖掘部分对数据进行统计、分析、清理和转换, 生成可利用的知识, 人通过不断关注外界的动态变化, 理解并预测机会事件, 进而影响决策。人对机会事件关注、分析、判断的过程中, 数据挖掘模块中不断有新的数据产生, 数据挖掘也被反复执行。双螺旋模型将人的智能和数据

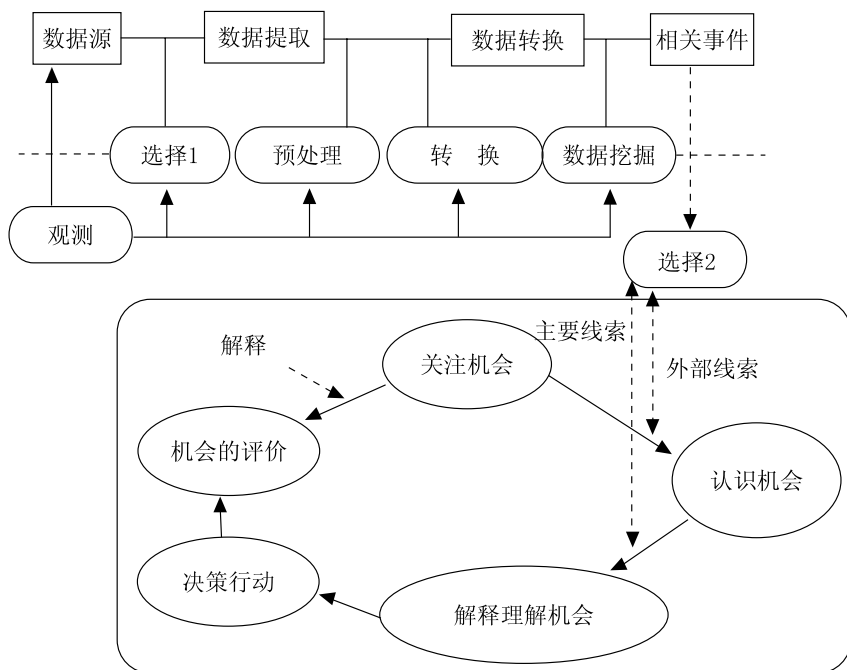


图1 机会发现双螺旋模型

库、数据分析有机结合起来,预测、识别可能带来的风险和机遇。

1.2.2 可视化工具KeyGraph

KeyGraph^[6]通过对高频词、重要词、关键词三种不同类型的节点设置阈值,以结构化的方式进行可视化表示,在展现数据间强连接关系的基础上,进一步挖掘弱连接中的强关系,帮助人们直观发现机会的重要性与关联性。

1.3 隐藏在我们身边的机会发现: 零售店

零售店一直以来都在运用机会发现,简单地说就是售货员在与顾客的交互中,根据顾客的模糊需求,用自己的眼光将适合的物品与顾客建立关联。

从顾客的角度来看: ①浏览商品的过程中,由于商品种类繁多,对

其特质与适应人群不了解,甚至对自己的喜好和需求并不十分清楚; ②与售货员交互过程中,比较自己与商品的适合和喜爱程度,逐渐明确需求; ③试用试穿的过程中,对店内物品熟悉、经验丰富、眼光准确的售货员给出的推荐对顾客来说,是一系列新的信息和机会,通过与自己的需求、适应度相比较,最终支持决策。

从售货员的角度来看: ①确定顾客需求阶段,通过顾客购买记录、喜好以及对顾客的印象和认识,询问并帮助用户明确需求; ②与顾客交互过程,通过推荐的反馈(试用试穿的体会和看法,适合度与不满意的地方等)修改匹配信息,再次推荐; ③挖掘顾客潜在购买需求,推荐相关合适的物品。

从购买结果来看: 顾客不总是购买相似的同质物品,也许连同装饰或组合物品一并购买,甚至额外购买与最初购买意愿毫不相关的

物品,充满意外性和偶然性。顾客往往有新的尝试和选择,售货员则是推动和发现这种机会的人。售货员总是能根据顾客的反馈,挖掘出真正的需求和喜好,给出合适的推荐或恰到好处的建议。研究人员将电子标签传感器应用在顾客行为过程中,借助KeyGraph生成顾客行为可视化图表,帮助售货员准确把握顾客兴趣、关注点与需求,为发现顾客潜在兴趣和需求提供有效途径。

1.4 机会发现的优势及其应用

利用机会发现的理论可以发现动态环境中的异常事件,挖掘出数量少、价值高的隐性知识和关联,为人们发现潜在机会提供可能,因此应用广泛。其中临床风险识别、企业危机管理利用动态环境中的异常事件来监控判断风险;地震预测、专利情报分析利用其少量信息弱关联的处理优势,机会发现在教育学、零售业、竞争优势分析等方面也应用较广^[7,8]。因其独特的优越性,在将用户与其需要的信息建立联系的推荐系统中,机会发现也将具有很大的应用空间。

2 推荐系统的特点和存在的问题

2.1 推荐系统及其特点

推荐系统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,模拟销售人员帮助用户完成购买过程(Resnick & Varian, 1997)^[9],由待推荐的资源、用户和推荐方法组

成,即基于用户特征运用一定方法将资源推送给用户的过程。推荐系统一般由收集用户特征和偏好、对用户信息进行分析建模和推荐算法三大模块构成,目前根据抽取特征不同,有以下四类:

(1) 基于内容的推荐系统: 它将计算得到并推荐给用户一些与该用户已选择过的项目相似的内容。例如,当你在网上购书时,你总是购买与历史相关的书籍,基于内容的推荐引擎就会给你推荐一些热门的历史方面的书籍。基于内容的推荐对内容的依赖性过高,无法处理未进行描述或描述错误的资源,并且只能推荐用户访问过的内容相似的资源,用户将被局限在特定兴趣,而较难发现新兴趣。

(2) 基于协同过滤的推荐系统: 它将推荐给用户一些与该用户品味相似的其他用户喜欢的内容。例如,当你在网上买衣服时,基于协同过滤的推荐引擎会根据你的历史购买记录或是浏览记录,分析出你的穿衣品位,并找到与你品味相似的一些用户,将他们浏览和购买的衣服推荐给你。简单地说就是用户聚类 and 资源聚类的相匹配。这种推荐方法具有推荐新信息的能力,可以发现用户潜在的但自己尚未察觉的兴趣偏好,可以推荐艺术、音乐、电影等难以进行内容特征描述定义的资源。但冷启动和打分的稀疏性会造成推荐的不准确^[10]。

(3) 基于关联的推荐系统: 它将用户和物品直接建立联系,或者建立物品和物品之间的联系,给出推荐。比如,如果你相信UFO的存在,那么你更有可能喜欢百事可乐;如果你有一个MBA学位,那么你更有可能喜欢蓝色的笔^[2]。物品和物品之间,有比如啤酒和尿布的经典

案例。这种方法利用心理行为学的理论研究成果,简单、易用、精准,但是少数的规则无法推广到品类繁多的物品中,局限性大。

(4) 混合推荐系统: 基于内容、协同过滤等推荐算法各有缺陷,因此实际中大多把不同的推荐算法进行结合,使用混合推荐算法,比如亚马逊、淘宝等购物平台。

表1 目前流行的推荐系统

Amazon、 淘宝	电子商务、购物平台
Netflix	在线DVD租赁提供商
Pandora	致力于歌曲推荐的公司
Hunch	电子商务公司,通过分析人帮助用户做决定的网站
Flickr	相片存储、分析平台
Genieo	不断学习你的浏览习惯的个人主页
Hulu	视频推荐网站

总之,推荐系统通过对已掌握的用户的个人信息、行为偏好进行挖掘,将用户的个人喜好与特定的参考特征相比较,试图预测用户对一些未评分项目的喜好程度,给出相关联的信息推荐,使用户轻松发现他本来要找的信息。也许用户并不知道有这样的信息存在,但是需要;也许是他从未涉及的领域,但是很感兴趣。比如,一个人要考英语六级,在他找了一系列六级词汇、听力、阅读的资料之后,系统为他推荐了一些新东方名师的RSS订阅、Tag标签分享等;一个公务员要写很多公文,评论员文章、新华社报道以及一些同行的Blog、圈子、社区等对他来说是很有帮助的;一个从事设计工作的白领,系统为他

推荐了像《杯酒人生》这样的电影和葡萄酒种类及鉴别这样的知识,也许他会感兴趣,也会喜欢。

2.2 推荐系统存在的问题

①目前对用户的分析易放大噪声,导致推荐不准确。用户特征分析中,系统往往将用户浏览、购买记录设置权重较大。事实上,用户购买一个物品受一个时期或一个事件的影响,也可能是帮亲戚朋友查询、购买或用于赠送,那么他的行为就不是他本人的真实需求,预测此行为对现在和将来的趋势和影响就是放大了噪声,造成对用户特征属性分析错误,从而导致推荐不准确。

②推荐易同质化,将用户限定在特定兴趣而较难发现新兴趣。系统的推荐是将与已浏览信息、用户感兴趣的信息和物品关联度、相似性较大的物品推荐给用户,例如Pandora音乐基因组计划,聘请音乐家对每首歌的400种不同属性打分,如果你喜欢一首歌,程序会自动寻找和这首歌的“基因”相同的歌曲。Pandora完全根据一首歌的本身属性和你的喜好记录来判断你喜不喜欢^[2]。这种同质化推荐会限定用户视野,导致你对不熟悉的领域更加陌生,即将“你不知道你不知道”的现象恶化。

③根据信息点推荐信息,无法在宏观结构中深层次发现用户真实需求。目前推荐系统主要依据你正在浏览的信息或物品,从相近品味用户的浏览、购买记录中选择与当下信息或物品匹配度最高的资源给予推荐,并不考虑你的这个需求与其他需求之间的关系之类的问题,因而很难从结构的角理解用户

真实需求,导致推荐的物品杂乱,不成体系,导致用户要找的资源没有找到,接受的信息量反而越来越大,容易迷航。

④随着用户浏览历史的增多,推荐性能趋于平庸。根据频次、相似度来确定用户偏好的推荐系统在用户浏览各种类型网站、寻找各种类型的商品之后无法将用户分类和确定偏好,口味测试将显示没有特点,这时候推荐失效了。

⑤缺乏对用户的持续关注,适应用户需求变化能力差。用户受外界事件和偶然事件的影响较大,易发生兴趣漂移,而系统对用户的兴趣和需求变化的捕捉较为滞后,当系统对用户特征属性进行修改,给出适合此需求的推荐时,也许用户的兴趣与需求已经转移到其他的事物中。

此外,小众冷僻需求难以满足,打分的稀疏性和冷启动也导致推荐不准确,因此,推荐系统的效率、推荐精准性亟需提升。

3 机会发现在推荐系统中的应用分析

3.1 机会发现能更准确地收集用户信息

用户信息分为基本信息和行为信息。鉴于个人隐私安全问题和网络的虚拟性,用户往往只提供简略的个人信息,甚至与实际情况差别很大。如果系统按照用户提供的数据(性别、年龄、职业、地理位置等)结合用户行为给出推荐,结果可能很不准确。用户的行为信息有显隐性之分:显性行为有分享、订阅、加关注等;隐性行为有鼠标悬停、回看等。系统对隐性行为的收

集较困难。机会发现的KeyGraph可视化软件,可以在直观查看用户行为节点、分析用户的行为之间关系的同时,发现那些关联性较弱的行为间的动态变化,进而透过用户信息、行为的表象发现用户的真实信息和意图。

此外,目前的数据分析建立在大量数据的基础上,对数量较少的信息无能为力,一般当作噪声忽略掉^[9]。机会发现具有不确定性和可变性,在事物发展的动态过程中,通过少数潜在隐含相关信息发现机会,预测对未来的影响,尤其关注数据挖掘中那些被忽略清洗的低频潜在信息,从隐性信息、关联中发现可能的机会。因此,机会发现理论可以更准确地收集用户信息。

3.2 机会发现有助于快速掌握用户兴趣变化

在变化迅速、分工细化的今天,用户的兴趣常常发生变化,我们将这些变化叫做兴趣漂移:①受到外界的影响,用户的兴趣可能扩张到新的领域,也可能收缩到更具体的方面。②用户认识问题的阶段和自身改变也会引发兴趣发生戏剧性的变化^[11]。系统往往无法理解一些相互违背的行为,机会发现的双螺旋模型强调在事物动态发展变化中关注机会、预测机会,进而支持决策,有利于发现并快速掌握用户新的兴趣生长点。

其次,一些距离较远、关联性较弱的行为之间的动态关系不易被发现,Keygraph软件模拟人脑发散的思维模式,利用一个个关键词、重要词、低频词等行为节点,将发散的思维与事物联系起来,以及发挥其在弱关联数据中的分析优势,

将思考以可视化的方式展现出来,就可以发现看似关联性不强或毫无关系的事物之间的联系,也可以解释六度分离理论,由此发现用户的潜在隐含兴趣和新的兴趣趋向。

3.3 机会发现有助于匹配资源,更好地满足用户的长尾需求

推荐系统中资源间的关联基于内容及口味相似、功能相关、现实中联系紧密等因素,对于标注不准确或无可用信息标识的资源,很难与其他资源与需求相匹配,以及新资源冷启动之初,怎样对其进行分类、匹配和推荐,是推荐系统的难题。机会发现中对动态事物变化的关注与预测,以及人在关注、分析、判断的过程中,不断利用数据挖掘模块,将人的智能和数据库、数据分析结合起来,编辑定义资源,将其聚类、寻找最近邻。由此,机会发现在推荐系统的应用可用提高资源定义、聚类的准确性,有效解决冷启动问题,有助于资源的聚类和匹配。

在物质种类极大丰富的今天,数量少、需求冷僻、个性化要求较高的长尾需求开始崛起并占领市场^[12],但由于小众用户和小众需求的数量少,与其匹配的资源也较少,加之网络资源标注的随意性和不全面、不准确性,使得本来就较弱的关联性进一步变差,导致推荐系统很难给出合适的推荐。机会发现的关注机会、隐性关联和对数量较少潜在重要信息的分析优势可以更好地发现用户和资源,满足用户的长尾需求。①对用户发现的注重和用户交互,深入分析用户行为、喜好、关注点,找到发散的线索,比如

用户关注建筑,向他推荐不同时期的优秀代表作品、评论和建筑大师等,在交互反馈中不断修正推荐内容,以提高推荐的精准度。②对资源的发现关键在于发现隐性关联,关联的事物可以是时代、风格、人物、观念、绘画、音乐、影像,也就是任意一个发散的节点,机会发现通过隐性关联将看似远离的事物建立联系,找到适合人群,就符合了小众需求推荐的目标。

因此,机会发现因其对事物动态变化和数量较少信息的分析优势,有助于发现表象下的真实意图和原因,快速掌握用户信息和需求变化,挖掘资源,准确聚类,寻找近邻,更好地满足用户的长尾需求,提高推荐效率。

4 机会发现理论对推荐系统的改进

4.1 重视并保持对机会的关注度,积累推荐资源

人们如果过度关注眼前的大众热门信息,就有可能注意不到其他机会事件的发生,从而丧失很多时机和机遇,并忽视一些事情不好的后果,无法及时作出反应。因此,人们应该保持对不同寻常事件的关注,推荐系统也如此。在分析快速增长的新近热门需求的同时,关注专业性较强、层次较高的小众需求,重视低频潜在信息的挖掘、发现;保持对机会事件的敏感性和关注度,不断发现优秀资源来源,积累推荐资源,进一步在技术上成功应用,做好聚类、近邻、隐性关联的分析,加入分类资源库,更好地为用户提供全面、权威、准确、优质的资源。

4.2 建立长短时双层兴趣和口味测试结合的用户兴趣模型

现实中,推荐系统对用户兴趣变化的掌握不够及时准确,因为机器观察到的只是行为,引起兴趣变化的原因却是隐藏的。用户易受大众热门资源或潮流的影响,比如奥运会,即使对体育不感兴趣的人也开始关注赛事,但随着时间的推移则会回归到较稳定的兴趣中去,因此,用户的兴趣可分为短时兴趣和长时稳定兴趣,建立长短期兴趣双层模型。

根据对用户行为的收集和分析,定义用户稳定的兴趣与偏好,放入长期模型中,给予稳定持续的精华资源推荐。在短期模型中,发现、分析用户新近需求,给出推荐,当用户兴趣收缩或转移,可以快速反应,同时将较长时期较稳定的兴趣加入长时兴趣。双层兴趣模型可以更加灵活地适应用户新增短暂需求,同时不遗忘用户长期稳定的兴趣和偏好。如果将用户的兴趣定义为若干个属性,兴趣漂移有可能是某一属性改变,有可能是多个属性发生改变。及时发现这些变化,修改用户的行为喜好,才能给出有针对性的推荐。

同时,在用户使用之初、使用中定期对用户进行有针对性的口味测试,可以更加准确地理解用户的兴趣和偏好,并及时发现新的关注点,有助于修正兴趣模型中的用户属性,进而更改相近邻,及时修改推荐策略。

4.3 建立自适应的推荐模块

目前推荐系统根据用户购买记录 and 系统分析出的需求推荐相关物品,殊不知用户获知了推荐信息,但是没有行动,原因是早就不需要这些、已经拥有这些资源或确定这些资源无法满足需求。在相关资源较少的情况下,系统往往推荐重复的信息,严重影响了推荐质量和用户满意度。

因此推荐系统应建立有记忆、智能的自适应推荐模块,设置关于推荐资源的反馈选项和评论,例如已知道、不需要、感兴趣、非常有用等,用户没有主动评价的情况下,系统可以根据行为模式代替用户评价,同时对已推荐资源进行标注。根据用户反馈和交互信息发现机会,自适应寻找机会事件和机会信息关联的用户和资源,修改兴趣模型,更准确地匹配推荐资源,提升推荐资源与用户行为的自适应性,给出准确、有价值的推荐。

4.4 使用测试性推荐和集成推荐结合的推荐方法

由于弱信号、小众需求与大众物品关联较弱,推荐系统的资源库能力有限,往往很难对小众冷僻需求给出合适的推荐。将与某件事紧密相关或内在隐性相关的其他形式的资源整合,给予集成推荐不失为一种好策略。关于关联较差的事物给予涉及历史事件、人物、著作、影视、音乐、动画、游戏等的知识网络,在帮助解决小众需求的同时,也大大提高了非热门信息(商品)的流行程度。

同时配合推荐与用户兴趣关联较差、距离较远的经典精华资源,关注用户反馈,通过分众分类、标签、评价等测试性推荐方法,发现

用户潜在兴趣,扩大用户视域的同时为用户制造惊喜,从而实现资源主动发现用户的理想。

4.5 提高专业人才在推荐系统中的作用

面对机会有三种情况:①机会没有到达你的认知和发现范围,你对机会一无所知;②机会到达你的视野,你却对它熟视无睹;③机会到达你的视野,你恰好偶然发现了它,意识到它的重要性,并分析可能对未来产生的影响。这说明对某一领域非常懂的人才具有发现蕴含价值的动态变化的眼光。随着算法的改进,推荐效果有了很大提升,但推荐系统仍有其固有缺陷:机器

无法准确理解用户的语义和真实意图;机器抓取并分析用户行为,却无法掌握隐藏在行为下的真实原因。此外数据分析中的阈值设定、不同环境下算法与推荐策略的调整、推荐效果满意度分析等,都需要懂行的人的眼光。

在2011推荐系统高峰论坛上,张栋提出推荐算法的改进对于提升推荐性能的空间相对较小,懂行的人才才是关键(他认为算法只占到10%)^[13]。因此,提高人在推荐系统中的比重,发挥专业人才的独到眼光,才能灵活适应用户需求与变化,在交互中发现机会,在变化中发现机遇,不断调整推荐策略,提升推荐效率。

5 结语

随着推荐系统的广泛流行,人们期待个性化更高、精准、有针对性并伴随着一点点惊喜的推荐,本文仅通过机会发现的理念更好地理解用户的信息需求,掌握用户兴趣漂移,匹配推荐资源,有效建立起物品和用户之间的联系。鉴于机会的不易发现和系统不够智能,以下两个方面的研究尤为重要:①加大懂行的人在系统中的比重,将系统的资源优势 and 人的智能有效结合,形成功能更强大的推荐系统;②进一步改进推荐算法,自适应区分不同情境下的算法权重比例和推荐策略,有效适应用户需求变化,提高推荐质量。

参考文献

- [1] 诸世卓,陈小平. Chance Discovery研究综述[J]. 计算机科学,2004(2):1-5.
- [2] 同入于野. 下一IT增长点:推荐引擎2.0[J]. 新观察,2010(9):20-21.
- [3] OHSAWA Y. The Scope of chance discovery [C]// Proceedings of the Joint JSAI 2002 Workshop on New Frontiers in Artificial Intelligence. Berlin: Springer-Verlag, 2001: 413-418.
- [4] OHSAWA Y, YADA K. Special issue on chance discovery-Discovery of significant events for decision making[J]. Information Science,2009,179(11):1567-1569.
- [5] OHSAWA Y. Modeling the Process of Chance Discovery [M]// OHSAWA Y, MCBURNEY P. Chance Discovery. Heidelberg Springer-Verlag, 2003:2-15.
- [6] OHSAWA Y, BENSON N E, YAEHIDA M. Keygraph: Automatic Indexing By Co-occurrence Graph Based on Building Construction Metaphor [C]// Proceedings of the Advances in Digital Libraries Conference. Seattle, 1998: 12-18.
- [7] 李琳. 机会发现及其在情报学方面的应用[J]. 情报科学,2011(2):173-177.
- [8] 黄晶晶. 机会发现的形式描述与形式建模的研究[D]. 哈尔滨:哈尔滨工程大学,2007.
- [9] RESNICK P, VARIAN H R. Recommender Systems [J]. Communication and the ACM, 1997(4): 56-58.
- [10] 孙小华. 协同过滤系统的稀疏性与冷启动问题研究[D]. 浙江:浙江大学,2005.
- [11] 杨杰. 个性化推荐系统应用及研究[D]. 安徽:中国科学技术大学,2009.
- [12] 安德鲁. 长尾理论[M]. 北京:中信出版社,2006.
- [13] 2011推荐系统高峰论坛[EB/OL]. [2011-05-15]. <http://news.csdn.net/a/20110306/293073.html>.

作者简介

裴倩 (1984-), 女, 硕士研究生, 研究方向: 信息服务与用户。E-mail: peiqian1899@126.com
王玉平 (1958-), 男, 教授, 硕士生导师。E-mail: wangyuping@swust.edu.cn

The Improvement on Recommendation System Based on the Theory of Chance Discovery

Pei Qian / Department of Economy Management, Southwest University of Science and Technology, Mianyang, 621010
Wang Yuping / The Library of Southwest University of Science and Technology, Mianyang, 621010

Abstract: In this paper, current situation and problems of recommendation system are analyzed. Chance discovery theory is introduced to improve recommendation system, for the characteristics of paying more close attention to change, discovering the implicit relationship and weak signal. It is respectively analyzed from the following aspects: to collect user information, and to grasp the changes of users' interest and matching resources. Several measures are benefit to improve the recommendation system, such as to improve the interest model, to establish adaptive recommendation module and to provide integrated recommendation.

Keywords: Chance, Chance discovery, Recommendation system

(收稿日期: 2011-07-01)