青岛农业大学

毕业论文(设计)开题报告

|  |  |  |  |
| --- | --- | --- | --- |
| 学生姓名 | 于子源 | 学 号 | 20170300497 |
| 专业年级及班级 | 软件1703 | 指导教师及职称 |  |
| 毕业论文（设计）题目 | 基于协同过滤算法的推荐系统 | | |
| 文献综述（选题研究意义、国内外研究现状、主要参考文献等，不少于1000字）   1. 选题研究意义   现在的社会是一个高速发展的社会，科技发达，信息流通，人们之间的交流越来越密切，生活也越来越方便，大数据就是这个高科技时代的产物。阿里巴巴创办人马云来台演讲中就提到，未来的时代将不是IT时代，而是DT的时代，DT就是Data Technology数据科技，显示大数据对于阿里巴巴集团来说举足轻重。  有人把数据比喻为蕴藏能量的煤矿。煤炭按照性质有焦煤、无烟煤、肥煤、贫煤等分类，而露天煤矿、深山煤矿的挖掘成本又不一样。与此类似，大数据并不在“大”，而在于“有用”。价值含量、挖掘成本比数量更为重要。对于很多行业而言，如何利用这些大规模数据是赢得竞争的关键。  为了解决“信息过载”的问题，已经有无数科学家和工程师提出了很多天才的解决方案，其中代表性的解决方案是分类目录和搜索引擎。而这两种解决方案分别催生了雅虎和谷歌两家互联网领域的著名公司。雅虎凭借分类目录起家，将著名的网站分门别类，从而方便用户根据类别查找网站。但是随着互联网规模的不断扩大，分类目录也只能覆盖少量的热门网站，越来越不能满足用户的需求因此催发了搜索引擎的诞生，以谷歌为代表的搜索引擎可以让用户通过关键词找到自己需要的信息。但是，搜索引擎需要用户主动提供准确的关键词来寻找信息，因此不能解决用户的很多其他需求，比如当用户无法主动提供准确描述自己需求的关键词时，搜索引擎就无能为力了。另一方面，搜索引擎根据关键字检索的信息是千篇一律的，不能满足各种用户在不同阶段不同背景下的个性化需求。由于搜索引擎存在这样的缺陷，推荐系统应运而生。  推荐系统也是一种帮助用户快速发现有用信息的工具，和搜索引擎不同的是，推荐系统不需要用户提供明确的需求，而是通过分析用户的历史行为为用户的兴趣建模，从而主动推荐能够满足用户兴趣和需求的信息。搜索引擎满足了用户有明确目的时的主动查找需求，而推荐系统能够在用户没有明确目的时帮助他们发现感兴趣的新内容。从物品的角度出发，推荐系统可以更好地发掘物品的长尾。互联网条件下，电子商务网站往往能出售比传统零售店更丰富的商品。虽然其中绝大多数是不热门的商品，但其数量极其庞大，因此这些长尾商品的销售额将是一个不可小觑的数字。如果要通过发掘长尾提高销售额，就必须充分研究用户的兴趣，而这正是个性化推荐系统主要解决的问题。推荐系统通过发掘用户行为，找到用户的个性化需求，从而将长尾物品准确地推荐给需要它的用户，帮助用户发现那些他们感兴趣但很难发现的物品。   1. 国内外研究现状 2. 国外研究现状   个性化推荐技术是推荐系统的核心，其研究最早开始于国外。1992年，Goldberg等人在一篇学术论文中提出了协同过滤算法(CF，Collaborative,Filtering),并基于此构建了首个个性化推荐系统一Tapestry邮件过滤系统，Tapestry根据用户下载的新闻内容来计算用户之间的兴趣相似度，进而利用这种相似度为用户推荐相关的新闻。协同过滤推荐算法的提出对当时整个推荐系统的研究具有划时代的意义。因此Tapestry被学术界认为是第一个真正意义上的推荐系统。  1997年3月，针对个性化推荐系统，美国计算机协会ACM进行了专题报告，首次探讨了个性化推荐服务在电子商务中的应用，并且.Resnick、Vafian提出了电子商务推荐系统的定义。同一年，明尼苏达大学的研究人员创建的在线电影推荐系统MovieLens，是协同过滤技术发展中一个开创性的进步，MovieLens通过利用用户对电影感兴趣的程度与协同过滤推荐技术向用户推荐其可能感兴趣且没看过的电影。从1999年开始，ACM设立推荐系统年会(ACM recomme-nder systems)，并且每年召开一次电子商务讨论会，个性化推荐研究开始成为了关注的焦点。  2001年，著名电子商务网站亚马逊将推荐系统运用到其网站中，个性化推荐开始从学术研究向实际应用中迈进。亚马逊的推荐系统深入到了其各类产品中，其中最主要的应用是个性化商品推荐列表和相关商品推荐列表。个性化推荐列表采用的是基于物品的协同过滤推荐算法，该算法给用户推荐那些和他们之前喜欢的物品相似的物品。亚马逊推荐系统的应用是学术研究和工业实践相结合的典范，之后越来越多的研究者和企业工程师投入到了个性化推荐系统的实践中。其中视频网站Netflix尤为出名,它是利用用户对电影的评分信息为用户推荐电影，并于2006年开始举办著名的Netflix Prize推荐系统比赛，以征集效率更高的算法。该项赛事对推荐系统的发展起到了重要的推动作用。2007年，Google根据用户最近搜索记录去挖掘用户的兴趣爱好，从而依据关键字在在线广告AdWorks中提供个性化推荐服务。  与此同时，ACM把推荐系统年会列为计算机领域顶级会议(ACM recommender systems）和数据分析领域高水平会议之中，推荐算法的文章数目每年都在增加。推荐系统研究的知名团队包括:明尼苏达大学GroupLens小组、密西根大学、卡内基梅隆大学、微软研究院等。  2. 国内研究现状  直至2000年左右，推荐系统才引起国内的关注，并逐渐成为计算机领域研究的热点。1999年，清华大学路海明等人提出基于多代理技术的混合智能个性化推荐服务。2000年，北京大学余锦风等人提出了个性化定制服务。2001年，南京大学研发了个性化信息检索智能系统DOLTRL—  Agent。2003年邓艾琳等人的《基于项目评分预测的协同过滤推荐算法》﹔2004年余力等人的《电子商务个性化推荐研究》﹔2007年彭玉等人的《基于属性相似的Item-based协同过滤算法》﹔2009年许海玲等人的《互联网推荐系统比较研究》﹔2013年孟祥武等人的《移动推荐系统及其应用》;  2015年张玉洁等人的《组推荐系统及其应用》。这些优秀的系统或者论文标志着个性化推荐技术的理论研究在学术界逐渐丰富起来。  近几年，随着互联网电商的快速发展，以及Amazon个性化推荐系统的应用成功，国内电子商务网站也纷纷构建其推荐系统。2006年，当当网开始提供个性化推荐服务，用以向客户推荐书籍。2008年，淘宝网推出了个性化推荐系统，用于帮助用户从大量的商品中找到符合自己偏好的商品。2011年，百度推出了个性化推荐首页，根据用户的行为向其推荐符合需求的信息。2014年，阿里巴巴开始举办“天猫”推荐算法大赛，吸引了国内外众多研究者的参加，促进了个性化推荐系统的发展   1. 主要参考文献   [1]Bawden D,Robinson L. The dark side of information: overload, anxiety and other paradoxes and pathologies.Journal of information science,2009,35(2):180-191.  [2]Adomavicius G，Tuzhilin A.Context-aware recommender systems.In:Recommender Systems Handbook[M]. 2011: 217-253.  [3] Resnick P, Varian H R.Recommender systems. Communications of the ACM,1997,40(3):56-58.  [4]Wang Guoxia, Liu Heping.Survey of personalized recommendation system. Computer Engineering and Applications, 2012,48 (7):66-76.  [5]克里斯·安德森长尾理论[M]﹒北京:中信出版社，2006.  [6]Goldberg D,Nichols D, Oki B M, et al.Using collaborative filtering to weave an information tapestry.Communications of the ACM, 1992,35(12):61-70.  [7] Konstan J A, Miller B N, Maltz D, et al. GroupLens: applying collaborative filtering to Usenet news. Communications of the ACM,1997,40(3):77-87.  [8] Schafer J B, Konstan J, Riedi J.Recommender systems in e-commerce[CV/Proceedings of the 1st ACM conference on Electronic commerce.ACM,1999:158-166.  [9]Linden G, Smith B,York J.Amazon.com recommendations: Item-to-item collaborative filtering.Internet Computing,IEEE, 2003,7(1):76-80.  [10]Available: www.netflix.com.  [11]Celma 0,Lamere P. Music recommendation tutorial[J].ISMIR. Vienna, Austria,2007.  [12]孟祥武，胡勋，王立才，张玉洁.移动推荐系统及其应用[J].软件学报，2013,24(1):91108.  [13]Meng Xiang-Wu, Wang F, Shi Yan-Cui, Zhang Yu-Jie. Mobile user requirements acquisition techniques and their applications.Journal of Software, 2014,25(3):439-456 (in Chinese).  [14]张玉洁，杜雨露，孟祥武．组推荐系统及其应用[J].计算机学报，2015，Vo1.38:在线出版号 No.89.  [15]Adomavicius G Tuzhilin A.Toward the next generation of recommender systems:A survey of the state-of-the-art and possibleextensions[J].IEEE Trans on Knowledge and Data Engingeeri-ng 2005,17,6:734-749. | | | |
|  | | | |

|  |  |
| --- | --- |
| 研究方案**（**研究目的、内容、方法、预期成果、条件保障等）  **1.基于Mahout开源平台构建传统的协同过滤推荐系统，并对其进行研究。**  **2.分析传统协同过滤推荐系统中的不足，提出改进措施，并基于Python平台将改进后的算法编程实现。**  **3.采用网络公开数据集活某零售商品牌专有销售记录进行实验，并将其改进后算法的实验结果与传统算法的实验结果进行对比，根据对比分析得出结论并采用迭代的方式不断优化。** | |
| 进程计划（各研究环节的时间安排、实施进度、完成程度等）  **2020.09-2020.12深入查阅与课题有关的资料，研究协同过滤推荐算法，包括相似度计算、最近邻**  **计算，并在研究过程中对课题进行调研和分析，准备开题报告。**  **2020.12-2021.02**  **研究矩阵分解技术，包括矩阵奇异值分解、隐语义模型。学习使用开源软件Mahout，并能用其实现传统的协同过滤推荐算法。学习Python语言编程。**  **2021.02-2021.03掌握Python编程以及Python第三方库的使用，并能初步编写SVD奇异值分解和**  **隐语义模型。**  **2021.03-2021.04将矩阵分解技术和协同过滤推荐算法进行融合，并在数据集上对改进的算法进行**  **验证，并与传统的协同过滤推荐算法在各种情况下进行性能比较，对实验结果给出详细的分析。**  **2021.04-2021.05撰写毕业论文，准备答辩。** | |
| 指导教师意见    指导教师签字： | |
| 毕业论文（设计）工作组意见 | |
| 毕业论文（设计）工作组成员签名 |  |
| 论证地点: 论证日期:20 年 月 日 | |