

教育资源个性化推荐方法研究与实现

李文欣^{1,2}, 文勇军^{1,2}, 唐立军^{1,2}

(1. 长沙理工大学 物理与电子科学学院, 湖南 长沙 410114;
2. 长沙理工大学 近地空间电磁环境监测与建模湖南省普通高校重点实验室,
湖南 长沙 410114)

摘要:随着教育信息化的快速发展,网络教育资源的开发得到了国内外的高度重视,各类教育资源信息更新速度惊人。为避免教育资源浪费,提高查找效率,及时为用户推荐有效信息资料,运用 Java Web 开发技术,结合网络爬虫技术和预测算法分析方法,设计实现了教育资源个性化推荐系统。系统采用 Java 爬虫技术,获取特定教育资源网站上公开教育资源信息,作为系统研究数据来源;采用矩阵分解模型、多层感知机模型和 NeuMF 预测分析模型,分析处理用户特征值矩阵和教育资源特征值矩阵,关联用户特征信息与教育资源信息,得到用户-教育资源预测值;研究教育资源个性化推荐技术,按照预测值优先级,将符合用户需求的教育资源信息推荐给 Web 用户。经测试,系统运行稳定可靠,资源推荐率大于 80%,资源推荐覆盖率广、推荐性强、实时性好、质量高,可以为用户提供及时的资源推荐服务,大大提高了教育资源的查找效率和资源利用率。

关键词:教育资源;网络爬虫;个性化推荐;预测模型

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2019)06-0018-05

doi:10.3969/j.issn.1673-629X.2019.06.004

Research and Implementation of Personalized Recommendation Method for Educational Resources

LI Wen-xin^{1,2}, WEN Yong-jun^{1,2}, TANG Li-jun^{1,2}

(1. School of Physical & Electric Science, Changsha University of Science & Technology, Changsha 410114, China;
2. Hunan Province Higher Education Key Laboratory of Modeling and Monitoring on the Near-Earth Electromagnetic Environments, Changsha University of Science & Technology, Changsha 410114, China)

Abstract: With the rapid development of education informatization, the development of network education resources has been highly valued at home and abroad. The information updating speed of all kinds of education resources is astonishing. In order to avoid waste of education resources, improve search efficiency, and timely recommend effective information materials for users, we design and implement a personalized recommendation system for educational resources by Java Web development technology in combination with network crawler technology and advance. The system uses Java crawler technology to obtain the information of public educational resources on specific educational resources websites as the data source of the system research, and uses matrix decomposition model, multi-layer perception model and NeuMF predictive analysis model to analyze and process user eigenvalue matrix and education resources eigenvalue matrix and correlate user information and educational resources information for user-education resources predictive value. The personalized recommendation technology of educational resources is studied. According to the priority of predictive value, the education resources information that meets the user needs is recommended to Web users. The system has been tested to be stable and reliable, with a resource recommendation rate of more than 80%. With wide coverage, strong recommendation, great real-time performance and high quality, it can provide users with timely resource recommendation services, greatly improving the search efficiency and resource utilization rate of educational resources.

Key words: educational resources; Web crawler; personalized recommendation; predictive model

收稿日期:2018-08-19

修回日期:2018-12-20

网络出版时间:2019-03-06

基金项目:国家科技支撑计划课题(2014BAH08F04)

作者简介:李文欣(1994-),男,硕士研究生,研究方向为数据挖掘及应用;文勇军,硕士,研究生导师,研究方向为网络信息安全与应用;唐立军,博士生导师,通讯作者,研究方向为信号检测与处理。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190306.0901.012.html>

0 引言

近年来,国内外对个性化推荐技术研究越来越重视,个性化推荐技术广泛应用于在社交网络、电子商务等领域^[1-2]。随着网络教育资源的爆发式增加,教育资源的查找效率越来越低,而教育资源的用户数量越来越大,用户需求迫切与教育资源利用率低的矛盾越来越突出,因此,教育资源的推荐方法研究和推荐系统的实现引起了国内外研究人员的高度重视^[3-4]。文中利用大数据分析技术开展预测算法的研究,构建基于用户和教育资源之间的预测模型,探讨教育资源的推荐方法,设计并实现教育资源个性化推荐系统。

1 教育资源的推荐范围与要求

1.1 教育资源推荐范围

根据教育资源的更新频率和权威性,文中选择参考书、论文、教学资源(课件)和教育发展动态等四类关键资源为研究对象^[5-6],选定四个官方网站公开信息作为本课题研究的数据:选取“中国高校教材图书馆”作为参考书的数据来源,选取“中国知网”作为论文的数据来源,选取“高等教育资讯网”下的“中国高校课件下载中心”作为教学资源(课件)的数据来源,选取“中国教育新闻网”作为教育发展动态的数据来源。

1.2 教育资源推荐要求

一般的推荐系统必须将一定范围内的资源全面、准确、实时地推荐给用户,文中针对教育资源推荐主要考虑推荐覆盖率、推荐准确率、推荐实时性的要求。

(1)推荐覆盖率:推荐的教育资源信息必须涵盖

参考书、论文、教学资源(课件)和教育发展动态等教育资源类型。

(2)推荐准确率:推荐的教育资源信息必须符合用户的特征属性,即必须与用户的实际需求相关联,要求推荐准确率大于80%。

(3)推荐实时性:用户自身特征值的改变,或教育资源的更新,系统能及时响应,自动调整用户-教育资源模型参数,推荐出最新最适合用户的教育资源信息。

2 预测推荐的原理和技术

2.1 预测推荐的原理

(1)矩阵分解模型(GMF)。

GMF模型^[7]是一种在推荐领域中常用来降低维数的技术,将原始矩阵分解为两个或多个矩阵的乘积,用来弥补稀疏矩阵的缺陷,因此可用于推导和完善用户和教育资源特征值信息,能够很好地处理用户和教育资源中的线性关系,如图1所示。为获得用户-教育资源稀疏矩阵中的未知预测值,将用户特征值 u 和教育资源特征值 r 作为参数输入到GMF模型中,形成用户特征值矩阵 p^T 和教育资源特征值矩阵 q ,将用户特征值矩阵与教育资源特征值矩阵进行点乘运算 $p^T * q$,得到用户-教育资源关系矩阵中 Y_{12} 、 Y_{21} 、 Y_{23} 等预测值,表示为:

$$y_{u,r}=f(u,r \mid p_u,q_r)=p_u^Tq_r \tag{1}$$

其中, p_u,q_r 分别表示用户潜在特征和教育资源潜在特征; $y_{u,r}$ 反映了用户-教育资源关联性,表示教育资源对用户需求的符合程度。

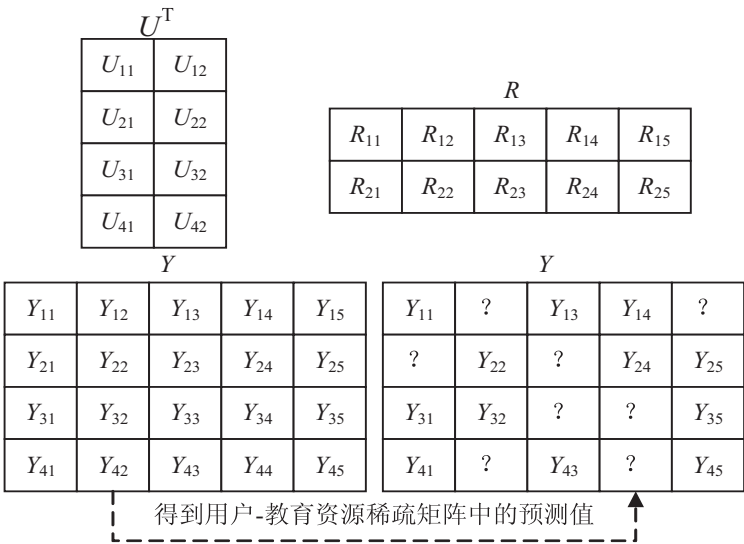


图1 教育资源预测中GMF原理

(2)多层感知机模型(MLP)。
MLP模型^[8]是一种函数近似分类模型,可以用来解决教育资源预测推荐中的用户与教育资源匹配问

题,处理用户和教育资源中的非线性关系。MLP教育资源推荐模型结构为,输入层用于用户特征值和教育资源特征值等参数的输入,由式2可知,ReLU函数在

非线性模型函数中收敛速度维持稳定,利用 ReLU 函数激活输入层,将用户特征值和教育资源特征值等数据映射到隐藏层;由于隐藏层没有规定层数,可以根据各自的需求选择合适的层数,隐藏层对用户-教育资源关联性层层优化,采用链式求导法则中的二分类特性,得到用户-教育资源非线性预测值 $y_{u,r}$,通过输出层输出。

$$\text{ReLU}(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases} \quad (2)$$

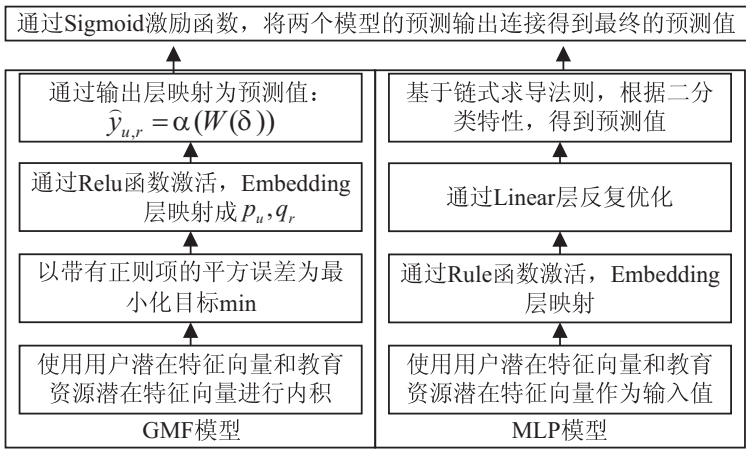


图 2 NeuMF 教育资源推荐模型流程

在 NeuMF 教育资源推荐模型中,线性的用户-教育资源预测值由 GMF 模型输出,用 y^{GMF} 表示;非线性的用户-教育资源预测值由 MLP 模型输出,用 y^{MLP} 表示。由式 3 可知,Sigmoid 函数在特征比较中权值更新准确,使用 Sigmoid 函数激励 y^{GMF} 和 y^{MLP} ,得到式 4 表示的最终教育资源预测值。

$$S(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

$$y_{u,r} = f_{out}(W_{out} \begin{bmatrix} y^{GMF} \\ y^{MLP} \end{bmatrix}) \quad (4)$$

2.3 相关技术

(1)爬虫技术。

爬虫技术^[10],是按照一定规则,自动抓取互联网信息的程序或者脚本,功能上分为数据采集、处理和储存三部分,可以很好地用于教育资源数据获取。实现方法分为分布式爬虫、Python 爬虫和 Java 爬虫等。文中采用 Java 爬虫技术实现教育资源的获取。

教育资源爬虫技术中,设定资源名称、资源类型、资源更新时间等作为爬虫标签,提取数据库中教育资源信息网站库中的 URL,作为爬虫 URL 队列,模拟用户发送访问请求,得到特定网页源代码。通过对网页源代码的解析,根据资源标签找到标签中的资源名称、资源类型和资源更新时间等内容。内容依照教育资源信息表中的资源名称、资源类型、资源来源等字段格式

2.2 预测算法分析与模型

将 NeuMF 预测模型^[9]引入教育资源推荐中,结合上面讨论的线性 GMF 模型和非线性的 MLP 模型,处理用户-教育资源间线性和非线性关系,得到教育资源预测值。

NeuMF 教育资源推荐模型流程如图 2 所示。

化,存入本地 MySQL。

通过 MySQL 定时任务,每天自动启动 Java 爬虫操作,模拟用户请求,对 URL 队列中资源信息进行及时更新,保证教育资源信息爬虫的时效性。同时利用定时任务,定期启动自动删除操作,删除过期的资源数据,保证 MySQL 中教育资源读取效率。

(2)个性化推荐技术。

基于用户-项目特征匹配的个性化推荐技术是一种用户和项目矩阵分解的技术^[11-13],其推荐模式可以直接应用到教育资源推荐。结合教育资源推荐的原理和方法,可以得到教育资源推荐中用户-项目特征匹配推荐模式,教育资源中个性化推荐技术为提取用户 ID、专业领域、学习兴趣、行为等特征值 u 和教育资源 ID、资源名称、资源类型等特征值 r ,作为模型参数,形成用户特征值矩阵 p^T 和教育资源特征值矩阵 q ,经过内积、求导等反复运算,ReLU 函数激活,得到预测值。选取大于预定值 K 的预测值,根据预测值优先级得到最终的推荐教育资源。

3 系统设计与结果分析

3.1 系统架构

采用网络爬虫技术和基于 NeuMF 模型的特征匹配技术,设计实现教育资源个性化推荐系统。系统使用开源的 Java 语言开发设计,采用 Liger UI 框架,结

合 JavaScript 和 CSS 技术,对系统的前端页面进行设计。后台采用 SSM 框架, Spring 中实现业务对象管理, Spring MVC 中的 View 层和 Controller 层响应用户请求, Mybatis 中的 Dao 层作为数据对象的持久化引擎, 封装数据库中用户和教育资源数据^[14]。数据存储采用关系型数据库 MySQL, 其查找速率快和灵活性高等优势为系统性能提供保障。

3.2 系统实现

教育资源个性化推荐系统包含教育资源获取、教育资源信息、推荐资源信息、个人教育资源、学科信息管理等 10 个模块, 下面主要介绍教育资源获取模块和教育资源推荐模块的实现。

1) 教育资源获取模块的实现。

采用爬虫技术中的 Java 爬虫技术, 在特定 URL 页面, 获取需要的教育资源信息。将爬虫获取的数据格式化后存入到本地 MySQL 数据库。

在特定的教育资源网站中, 包含众多公开的教育资源信息, 首先要分析其教育资源信息发布页面的源代码, 找出教育资源相关信息点位置及内容标签结构, 确定正则表达式或标签选择器作为爬取规则, 再利用 Java 语言编写爬虫程序, 获取教育资源信息的标题、作

者、分类属性、链接地址等大数据信息^[15-16]。将获取的数据格式化后存入到本地 MySQL 数据库。实现步骤如下:

- (1) 以 HTTP Web Request 为基类, 创建 DAL 操作类 Request Helper;
 - (2) 通过 Request Helper, 创建实体 Request;
 - (3) 在 Request 中构造请求 HTML, 以 Post 方法提交给 Remote server;
 - (4) 获得 server 302 响应后, Data flow 合并写入 Document;
 - (5) 调用专用文件操作类, 逐条读取文件;
 - (6) 调用 DAL, 写入 MySQL。
- 2) 教育资源推荐模块的实现。

系统从本地 MySQL 数据库中提取数据, 将用户特征值 u 和教育资源特征值 r 进行特征匹配, 建立用户-教育资源特征匹配模型进行分析预测, 得到用户-教育资源间的预测值, 然后将预测值高于预定值 K 的教育资源信息推荐给 Web 用户, 完成教育资源信息推荐。

推荐流程如图 3 所示。

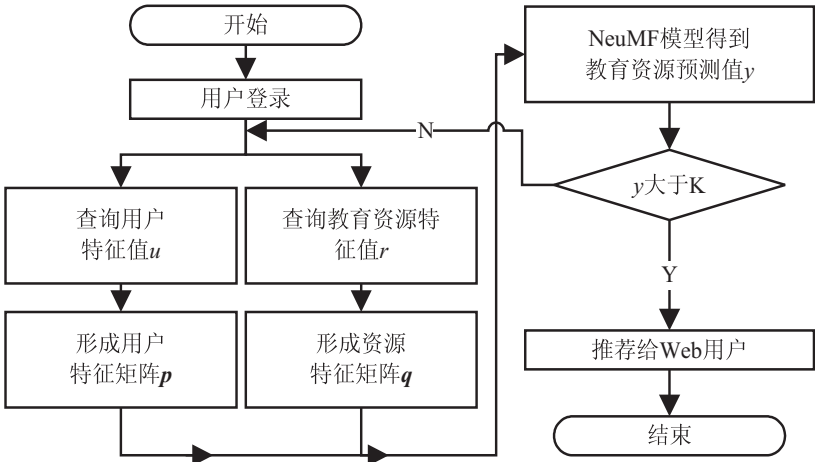


图 3 推荐流程

3.3 系统测试及结果分析

教育资源个性化推荐系统设计完成后, 通过实验室模拟的方式对系统进行实验测试, 包括功能测试和性能测试。测试环境为: 硬件平台为 Dell Inspiron 3420 笔记本; 操作系统为正版 Windows7 64 位; 处理器为 Intel(R) Core(TM) i5-3210M; 内存(RAM) 为 8 G; 数据库为 MySQL 5.7。

(1) 功能测试。

测试方法: 通过模拟系统用户操作, 登录教育资源个性化推荐系统后, 对各个一级模块和二级模块功能进行操作测试。

系统功能测试结果见表 1。

表 1 系统功能测试

序号	一级模块	二级模块	测试结果
1	教育资源管理	教育资源信息	通过
		推荐资源信息	通过
		个人教育资源	通过
2	数据管理	学科信息管理	通过
		学习兴趣管理	通过
3	系统管理	用户管理	通过
		角色管理	通过
		日志查询	通过
4	个人设置	账号信息	通过
		修改密码	通过

表 1 表明,系统一级模块和二级模块等所有功能测试结果均已通过,操作正常,能够正常工作。

(2)性能测试。

测试方法:推荐性能测试,模拟 100 个特征信息互不相同的用户,通过读取 MySQL 数据库中 1 000 份教育资源信息,即参考书、论文、教学资源(课件)和教育发展动态各 250 份,对系统进行预测性能测试。实时性测试,在原基础上改变用户学科信息、学习兴趣、行为等特征值,得到改变后用户特征值的更新结果;改变教育资源,得到更新的教育资源特征值。在相同环境下对系统推荐性能重新测试,启动 10 次系统,查看推荐的教育资源更新情况。

测试结果见表 2 和表 3。

表 2 改变前的推荐结果

	参考书	论文	资源	动态
覆盖范围	√	√	√	√
准确率/%	83.4	93.2	94.5	83.7
最新资源	√	√	√	√

由表 2 可知,系统推荐信息涵盖参考书、论文、教学资源(课件)和教育发展动态等资源类型,推荐准确率均大于 80%,且能推荐最新资源,系统预测推荐性能良好。

表 3 改变后的推荐结果

	参考书	论文	资源	动态
覆盖范围	√	√	√	√
准确率/%	83.8	92.1	95.1	83.2
最新资源	√	√	√	√

由表 3 可知,在改变用户特征值和更新教育资源后,系统能够得到新的推荐教育资源,推荐准确率没有太大波动,且在推荐中包含最新的教育资源,可见系统推荐性能较好,能够及时响应用户-教育资源间的特征改变,达到教育资源推荐要求。

4 结束语

将社交网络、电子商务等领域中应用广泛的个性化技术引入到教育资源推荐中,通过对教育资源用户-项目的特征匹配模型研究,构建了基于用户和教育资源之间的预测模型,得到了教育资源的个性化推荐方法,设计并实现了教育资源个性化推荐系统。从测试结果来看,该教育资源个性化推荐方法实时性强、覆盖率高,推荐效果好,可以推广应用到教育资源推荐

领域。

参考文献:

[1] 李俊薇. 基于协同过滤的校园教育资源网个性化推荐研究[D]. 武汉:华中师范大学,2009.

[2] 文勇军,吴冬冬,王 键,等. Spark 平台下教育资源个性化推荐研究[J]. 智能计算机与应用,2017,7(2):25-30.

[3] 方 超,暴建民,薛四猛. 基于领域特征值的协同过滤个性化推荐方法[J]. 计算机技术与发展,2017,27(11):88-91.

[4] 杨东衡. 基于教育资源的推荐系统研究与实现[D]. 成都:电子科技大学,2017.

[5] SALEHI M,KAMALABADI I N,GHOUSHCHI M B G. An effective recommendation framework for personal learning environments using a learner preference tree and a GA[J]. IEEE Transactions on Learning Technologies,2013,6(4):350-363.

[6] 丘少岷. 基于知识点兴趣模型的教育资源个性化推荐系统[D]. 广州:华南理工大学,2016.

[7] 刘敏娜,赵 蓓. SSI 技术在教学资源平台应用的研究[J]. 计算机技术与发展,2017,27(9):170-174.

[8] 沈筱譔. 在线学习系统中的深度学习推荐算法研究[D]. 武汉:华中师范大学,2017.

[9] 赵建龙. 基于协同过滤推荐技术的学习资源个性化推荐系统研究[D]. 杭州:浙江工业大学,2012.

[10] 乔 峰. 基于模板化网络爬虫技术的 Web 网页信息抽取[D]. 成都:电子科技大学,2012.

[11] KLAŠNJA-MILIĆ EVIC A, VENSE B, IVANOVIĆ M, et al. E-learning personalization based on hybrid recommendation strategy and learning style identification[J]. Computers & Education,2011,56(3):885-899.

[12] AHER S B, LOBO L M R J. Combination of machine learning algorithms for recommendation of courses in e-learning system based on historical data[J]. Knowledge-Based Systems,2013,51:1-14.

[13] 朱柳青. 基于深度学习的课程推荐与学习预测模型研究[D]. 杭州:浙江工商大学,2018.

[14] 于湛麟,李仲秋,任永昌. SSH 框架实现 MVC 架构的电子商务软件平台[J]. 计算机技术与发展,2012,22(10):169-172.

[15] 姜 强,赵 蔚,王朋娇,等. 基于大数据的个性化自适应在线学习分析模型及实现[J]. 中国电化教育,2015(1):85-92.

[16] GARCÍA E, ROMERO C, VENTURA S, et al. A collaborative educational association rule mining tool[J]. The Internet and Higher Education,2011,14(2):77-88.