

基于智能推荐的电子商城购物系统设计*

付加星, 孟佳娜, 沈杭春, 武星宇

(大连民族大学计算机科学与工程学院, 辽宁 大连 116600)

摘要: 为提高网络购物的个性化体验, 设计和实现了一个基于智能推荐的电子商城购物系统。为达到用户网页浏览行为的分析与预测, 推荐算法结合了基于用户和基于项的协同过滤算法, 系统构建用户具有偏序结构的关键浏览路径层次图。数据分析结果表明, 改进后的推荐算法有助于提升推荐系统的性能, 从而满足用户个性化需求。

关键词: 智能推荐; 个性化需求; 协同过滤; 相似度; 聚类; 电子商城

中图分类号: TP311

文献标志码: A

文章编号: 1006-8228(2015)06-39-03

Design of E-commerce shopping system based on intelligent recommendation

Fu Jiaxing, Meng Jiana, Shen Hangchun, Wu Xingyu

(School of Computer Science and Engineering, Dalian Nationalities University, Dalian, Liaoning 116600, China)

Abstract: In order to improve the personalized experience of shopping online, an electronic shopping system based on the intelligent recommendation is designed and implemented. In respect of the users' web browsing behavior can be analyzed and forecasted, the recommendation algorithm is combined with the user based and the item based collaborative filtering algorithm. A key browser path hierarchy chart with partial order structure of the user is built. Data analysis results show that the improved recommendation algorithm is helpful to enhance the performance of the recommendation system so that satisfy the users, personalized needs.

Key words: intelligent recommendation; personalized needs; collaborative filtering; similarity; clustering; E-commerce

0 引言

智能推荐系统^[1-2]是 Internet 网上购物过程中解决信息过剩的一种信息过滤技术, 向用户推荐感兴趣的信息和商品(衣物、新闻、书籍、电影等等)。推荐系统首先从用户的浏览史和社会环境等信息得到用户特征, 根据用户特征能够得到一个用户概况, 推荐系统从用户概况等信息推算用户对未看过的商品和信息的喜好程度。

在这个信息大爆炸的时代, 用户数量和项目数量急剧增加, 而可用于评分的项目很少, 系统会消耗大量的时间完成推荐算法, 导致了不能及时响应用户的请求。本文采用一种项目簇偏好的用户聚类方法, 改变了预测评分过程中用户相似性算法, 并考虑用户的兴趣随时间变化的情况, 保证了系统的实时性。为了避免用户在面对网站大量信息, 却无法很快从中获取自己感兴趣的信息的情况, 本文设计了基于 Web 数据挖掘的推荐系统, 根据用户的兴趣、特点和购买行为, 计算出用户可能喜欢的内容或商品, 为不同用户展示不同内容, 满足其个性化需求。

1 系统关键技术

1.1 开发环境

1.1.1 数据库

本系统使用 MySQL 数据库^[3], MySQL 是一种开放源代码

的关系型数据库管理系统, MySQL 数据库系统使用最常用的数据库管理语言——结构化查询语言 (Structured Query Language) 进行数据库管理, 具有数据库的通用性, MySQL 支持标准的结构化查询语言。本系统数据库的数据表是基础, 决定了该系统的可维护与可扩展性, 以用户信息表(db_user)为例, 其字段设计为:

用户信息表 db_user: userId (Id 号、主键)、userName (姓名)、password (密码)、userTime (注册时间)、userEmail (邮箱)、userSex (性别)、userAge (年龄)、userImg (头像)、userPhone (手机)、userCode (邮编)、userAddress (地址)、userWork (职业)、userMoney (工资)、userLike (爱好)、userRight (用户权限、外键)。

1.1.2 Java EE 三大框架

Struts 是开源软件^[4], 使用 Struts 的目的是帮助程序员减少运用 MVC 设计模型来开发 Web 应用的时间。Struts 对 Model, View 和 Controller 都提供了对应的组件。Spring 是一个强大的框架, 解决了许多在 J2EE 开发中常见的问题。Spring 提供了管理业务对象的一致方法。Spring 的架构基础是基于使用 JavaBean 属性的 Inversion of Control 容器。Hibernate 是一个开源代码的对象关系映射框架, 对 JDBC 进行了对象封装, 使得 Java 程序员可以使用对象来操作数据库。Hibernate 可以应用在任何使用 JDBC 的场合, 既可以在 Java 的客户端程序使用, 也

收稿日期: 2015-3-25

*基金项目: 国家自然科学基金资助项目(61202254); 中央高校自主科研基金资助项目(DC201502030202; DC201502030405)

作者简介: 付加星(1992-), 男, 湖南永顺人, 本科在读, 主要研究方向: 数据挖掘。

通讯作者: 孟佳娜(1972-), 女, 吉林四平人, 博士, 教授, 主要研究方向: 自然语言处理, 机器学习。

可以在Servlet/JSP的Web应用中使用。

1.2 推荐算法

1.2.1 协同过滤算法

电子商务推荐系统的应用日益广泛,推荐算法作为推荐系统的核心也得到广泛的研究,协同过滤推荐算法^[5]是目前应用最成功的推荐算法之一。电子商务推荐系统可根据其他用户的评论信息,采用协同过滤技术给目标用户推荐商品。协同过滤算法主要分为基于用户的协同过滤算法和基于项目的协同过滤算法。启发式协同过滤算法主要包含收集用户偏好信息、寻找相似商品或用户和产生推荐三个步骤。

基于用户的协同过滤推荐算法又叫基于内容的推荐^[6],其基本思想是用户选择某个推荐对象,也就是说一些用户对某些推荐对象的评价比较相似,则说明这些用户的兴趣爱好也是相似的。所以协同过滤推荐首先找到与目标用户兴趣爱好相似的最近邻,根据最近邻对推荐对象的评价为用户进行推荐。常用的计算两个用户的相似性公式(1)为余弦相似度公式。

$$\text{sim}(i, j) = \frac{\sum_{c \in I_{ij}} (R_{i,c} - \bar{R}_i)(R_{j,c} - \bar{R}_j)}{\sqrt{\sum_{c \in I_{ij}} (R_{i,c} - \bar{R}_i)^2} \sqrt{\sum_{c \in I_{ij}} (R_{j,c} - \bar{R}_j)^2}} \quad (1)$$

其中, $\text{sim}(i, j)$ 表示用户 i 与用户 j 的相似度, $R_{i,c}$ 表示用户 i 对商品 c 的评分, $R_{j,c}$ 表示用户 j 对商品 c 的评分, \bar{R}_i 是用户 i 的平均评分, \bar{R}_j 是用户 j 的平均评分。

通过计算相似性,用户可以预测该用户对该商品可能的评分,计算方法如公式(2)。

$$\text{pred}(u, p) = \frac{\sum_{i \in \text{ratedItems}(u)} \text{sim}(i, p) * R_{u,i}}{\sum_{i \in \text{ratedItems}(u)} (|\text{sim}(i, p)|)} \quad (2)$$

其中, $\text{pred}(u, p)$ 表示用户 u 对商品 p 的预测评分。

1.2.2 协同过滤算法的改进

虽然传统的协同过滤推荐算法是迄今为止最成功的推荐系统技术之一,且被应用在很多成功的推荐系统中,但是传统的协同过滤推荐存在数据稀疏性、冷启动和可扩展性等问题^[7]。因此,我们对传统协同过滤算法中商品评分的公式(2)作如下改进:

$$\text{pred}(u, p) = \bar{R}_u + \frac{\sum_{i \in \text{ratedItems}(u)} \text{sim}(i, p) * (R_{u,i} - \bar{R}_u)}{\sum_{i \in \text{ratedItems}(u)} (|\text{sim}(i, p)|)} \quad (3)$$

其中, \bar{R}_u 表示用户 u 的平均评分。改进之后的公式(3)相比之前的公式(2)优点体现在:运用了当前用户的平均评分来进行计算,最后的预测评分不可能为0分,其结果更加准确,更具有说服力。

除了以上所述的推荐算法改进外,本系统还将用户评分相似度和用户特征相似度相结合选取用户最近邻居计算用户预测评分,同时将项目评分相似度和项目特征相似度相结合,选取项目最近邻居计算项目预测评分,然后将用户预测评分和项目预测评分相结合产生最终推荐。

2 系统描述

一个完整的个性化智能推荐系统^[8]大概主要包括:收集用

户信息的行为记录模块、分析用户喜好的模型分析模块和推荐算法模块,其中推荐算法模块是推荐系统中最为核心的部分。

本系统在对客户数据透明的情况下,对客户的数据、行为进行分析。当用户登录到系统时,通过系统对身份的验证后,用户信息模块开始收集信息,跟踪客户的行为,收集客户的交易方式、时间、内容和评价的过程和结果的主要信息,数据库进行信息的预处理和归一化,形成客户个性化特征数据库。个性化分析处理模块结合客户个性化数据库进行个性分析,并把结果送到智能重组模块。在智能重组模块,对用户兴趣的分类转化为对项目的分类。对于同一用户,如预测项目所属类别不同,用来预测的邻居用户也不同,也就是邻居用户与待预测的项目在内容上具有一定相似性,从而保证用来预测的邻居用户与当前用户在待预测项目上具有相似的兴趣爱好。智能重组模块根据个性化信息从商品数据库中筛选产品,并推荐给用户。

根据以上分析,系统的总体设计如图1所示:

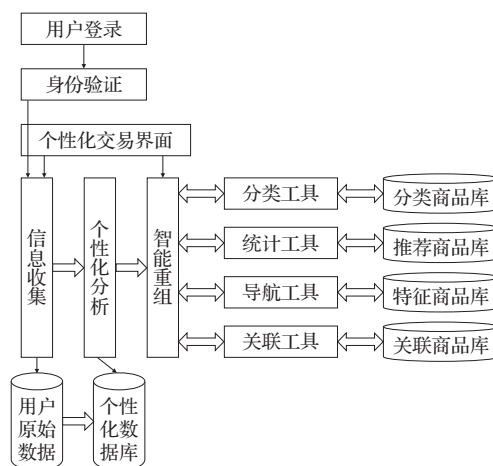


图1 系统总体设计

3 系统实现

3.1 网站页面的设计

图2是网站的主页面,上面的功能主要包括:商品的分类查询、用户注册的最新动态、最活跃的用户排名和商品的简单推荐等。网页是用html、css、JavaScript、jQuery、Ajax等语言相结合进行制作的。

3.2 推荐算法的结果分析

3.2.1 数据集

我们采用了系统虚拟数据(自己模拟的数据)对该算法进行测试,一共提供了232位用户对781个物品的近5000条评分记录。把整个记录数据按照75%和25%的比例划分为训练集和测试集。

3.2.2 度量标准

该系统的度量标准采用统计度量方法中的平均绝对差MAE(Mean Absolute Error)进行度量^[9]。MAE通过计算预测的用户评分与实际的用户评分之间的偏差来度量预测的准确性,MAE越大,推荐质量越高。平均绝对偏差MAE的定义如公式(4)。

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| \quad (4)$$



图2 系统主界面

其中,预测的评分集合表示为 $\{f_1, f_2, \dots, f_n\}$,对应的实际用户评分集合表示为 $\{y_1, y_2, \dots, y_n\}$ 。

3.2.3 实验结果

实验中,对于一个刚刚注册的用户,系统可以对该用户注册的信息进行分析处理,然后与分类的物品进行信息匹配,给该用户推荐与用户信息匹配的一类物品的最热卖商品,或者在用户的集合里找到与该用户A有相同兴趣的用户B,然后推荐B喜爱的物品。同时,传统的协同过滤算法是从头计算到尾,把整个数据矩阵中的数据一一地进行计算,最后得出结果,但是,改进的算法会在中途进行判断,过滤掉一些数据而计算得出结果,比如相似度为负相关的就不需要求其预测评分了,从而大大的提高算法的效率。在推荐的准确性方面我们发现,用户的兴趣爱好是在不断变化的,如果还使用传统的协同过滤算法去为用户进行推荐,那么推荐的准确性就很让人怀疑了。针对这一问题,系统会根据数据矩阵的数据量进行数据筛选。比如,数据矩阵的数据量过大,系统就会自动通过时间来筛选出用户近期留下的数据来进行计算,时间比较久远的数据就会被筛选掉,从而符合了用户近期的兴趣爱好,使推荐的结果更加准确。

通过协同过滤算法的改进和以上实验的设计,得到最终的实验结果如图3所示。

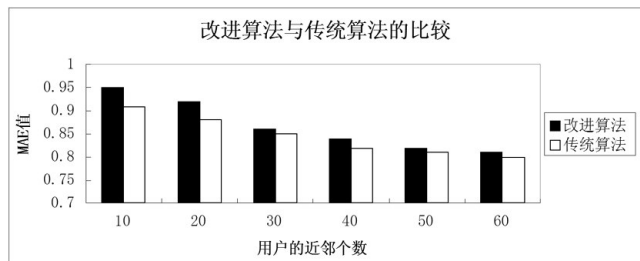


图3 实验的对比结果

实验中,我们设计了不同的用户近邻个数,得到了不同的

MAE值,在不同的近邻数下,改进之后协同过滤推荐算法都得到了更高的MAE值,可见改进之后的协同过滤推荐算法,其推荐结果的准确性得到了很大的提高。同时当近邻数逐步增加后,MAE值逐步下降,可见近邻数增加会增大噪声,对推荐结果起到了反作用,而且近邻数少会提高系统的效率。

4 结束语

本文实现了一个基于智能推荐的电子商城购物系统,该系统为用户快速推荐可能感兴趣的物品,为用户大大的节省了时间,保证了网络数据的有序性、安全性和完整性。系统实现了基于Web数据挖掘的推荐系统,根据用户的兴趣特点和购买行为,计算用户最可能喜欢的内容或商品,并为不同用户展示不同内容,满足其内容的个性化需求。未来在线下推荐和冷启动问题处理方面还需要进一步研究。

参考文献:

- [1] 乔冬春,刘晓燕,付晓东,曹存根.一种基于本体的推荐系统模型[J].计算机工程,2014.40(11):282-287
- [2] 王立才,孟祥武,张玉洁.上下文感知推荐系统[J].软件学报,2012.23(1):120
- [3] 王飞飞,崔洋,贺亚茹.MySQL数据库应用从入门到精通(第2版)[M].中国铁道出版社,2014.
- [4] 李刚.轻量级Java EE企业应用实战[M].电子工业出版社,2013.
- [5] 马宏伟,张光卫,李鹏.协同过滤推荐算法综述[J].小型微型计算机系统,2009.30(7):1282-1288
- [6] 姜书浩,薛福亮.一种利用协同过滤预测和模糊相似性改进的基于内容的推荐方法[J].现代图书情报技术,2014.24(2):41-47
- [7] 郭艳红,郭贵仕.协同过滤系统项目冷启动的混合推荐算法[J].计算机工程,2008.34(23):11-13
- [8] 刘凤霞,孙家蓉.基于商品分类的电子商务推荐系统设计[J].计算机应用与软件,2014.5:37-41
- [9] 刘旭东.B2C网上购物推荐系统的设计与实现[J].计算机应用与软件,2009.26(9):195-197