

美家 365 网上商城智能推荐系统研究与设计

戚永军¹ 翟智平²

(1. 北华航天工业学院 网管中心, 河北 廊坊 065000; 2. 北华航天工业学院 图书馆, 河北 廊坊 065000)

摘 要: 智能推荐系统作为一种信息过滤的重要手段, 是当前解决信息超载问题的重要途径, 能有效解决电子商务网站中用户找不到满意商品, 商家找不到正确促销对象的问题。本文介绍了美家 365 网上商城智能推荐系统的功能需求和系统架构, 探讨了数据搜集、数据清理、数据预处理的过程, 不仅满足了用户提出的需求, 还对类似系统的开发有一定的参考价值。

关键词: 网上商城; 智能推荐; 研究; 设计

中图分类号: TP393

文献标识码: A

文章编号: 1673-7938(2013)02-0001-03

0 引言

香河家具城以家具品种品牌多、产品规格全、质量信誉好, 被国内业界人士和新闻媒体誉为“中国最具发展潜力和活力的家具市场”。近年来, 电子商务的兴起, 传统实体店的经营模式受到极大挑战, 为适应潮流、拓宽销售渠道, 香河家具城推出了家具电子商务平台——美家 365 网上商城。本文为该平台设计了一套智能推荐系统, 以便提高销售额, 发现潜在客户, 增加商品促销响应率。

1 功能需求

智能推荐系统应具备如下所示的功能:

(1) 分析用户的购买习惯提供个性化服务。通过大量数据挖掘, 分析用户消费习惯, 有针对性提供其感兴趣的商品, 同一页面不同用户打开, 显示不同商品列表, 实现个性化服务。

(2) 访问页面时自动推荐感兴趣的商品。用户在打开网站或使用搜索功能时, 能呈现该用户最感兴趣的商品, 商品按关注度降序排列。

(3) 购买时自动推荐有价值的商品。在用户付款前后, 自动推荐其在购买过程中没有想到但确实需要的商品。

(4) 将用户消费习惯分类, 发现用户特征, 为广告和销售提供支持。

(5) 只推荐本站商品, 不爬取其他网站商品

信息。

2 系统设计

2.1 系统结构

该系统由数据搜集、数据清理、数据预处理、商品推荐等模块组成, 系统结构如图 1 所示。

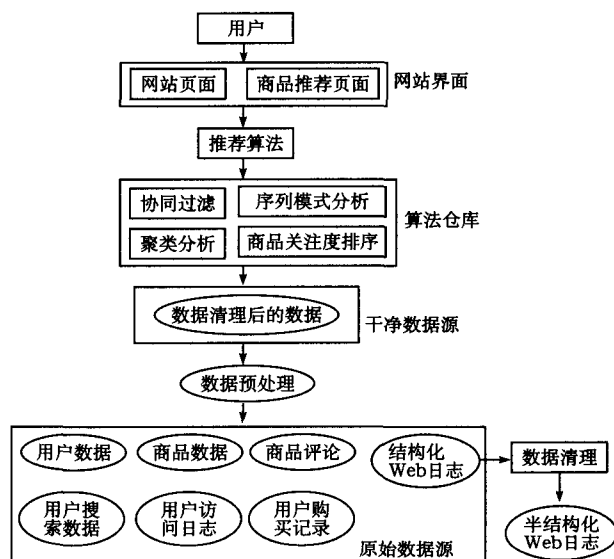


图 1 智能推荐系统结构图

2.2 数据搜集

数据搜集是将用户访问网站的各种信息记录下来供数据挖掘使用, 数据存入数据库中。数据表结构及其关系如图 2 所示。

用户信息表: 记录注册用户的信息, 为商品推荐的协同过滤算法提供依据。尽量要求用户提供真实、详细的个人信息, 这些信息越准确、越详细, 商品推荐的质量越高。

基金项目: 廊坊市科技局科技支撑计划项目(2012011011)

收稿日期: 2013-01-08

作者简介: 戚永军(1976-), 男, 硕士, 工程师, 河北唐山人, 从事计算机网络技术和软件开发研究。

商品评论表:记录用户对商品的评论信息,该表中商品的评级可分为 1-5 分,分值越高说明对商品的关注度越高,同时分值也为协同过滤算法提供依据。

用户购买记录表:记录用户购买商品的信息,购买某一商品表示对该商品关注度较高。

搜索记录表:该表记录用户搜索商品的相关信息,通过搜索关键字可推测商品的关注度。

用户浏览日志表:记录用户点击网站的动作,其中“当前页面地址”、“浏览时长”、“下一个页面地址”等信息可为序列模式分析、聚类分析提供依据,有助于挖掘用户的习惯,提供个性化服务。

商品表:记录商品的细节,为推荐页面提供数据。

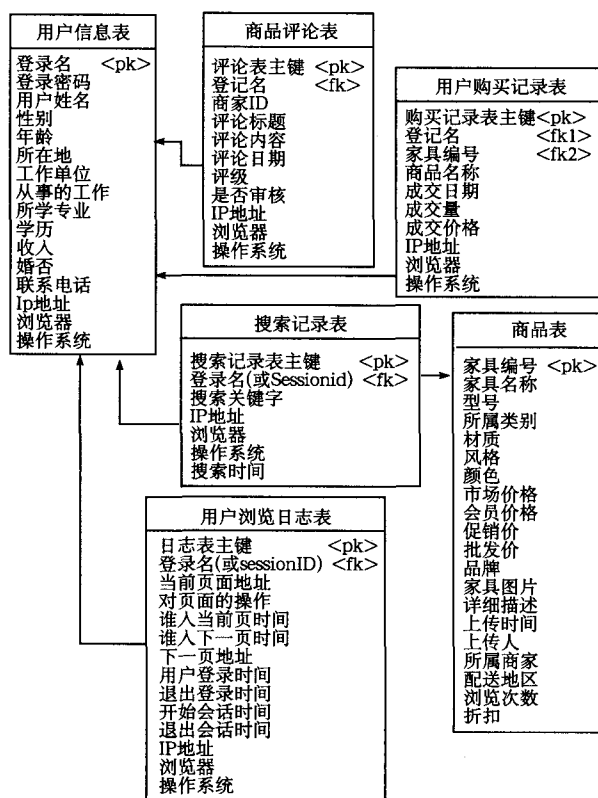


图 2 原始数据源

2.3 数据清理

本系统数据大部分由网站程序采集。此外,服务器会自动采集一些信息存入 Web 日志中。由于 Web 日志是以文本的形式存储,是半结构化数据且存在许多无关信息,使用前需进行预处理。其步骤包括:清理数据、识别用户或会话、补充路径。

(1)清理数据:是删除不相关的记录,主要包括以下几类:

删除 URI 中后缀名为 gif、jpg 等图片或音视频

相关的记录。

删除请求样式表的记录。

删除访问错误的页面,即属性 Status 中代码值小于 200 或大于 299 的日志记录。

常见页面请求方法包括 GET、POST 和 HEAD,由于只有 GET 方法反映了用户的访问行为,所以只保留 GET 方式的记录。

保留 URI 中含有 id 参数的记录,由于该网站的商品页面含有 id 参数。

(2)识别用户或会话:获取 Web 日志中的 IP 地址、浏览器、操作系统,分别和用户表、商品评论表、用户购买记录表、用户浏览日志表中的记录比对,识别出注册用户,无法识别的保留 Session id 值。

(3)补充路径:由于浏览器存在缓存,一些用户的操作并不能记录到 Web 日志中,从而导致日志中的记录路径信息不完整,通过分析将操作路径信息补充完整。

2.4 数据预处理

该模块是系统的核心,包括一系列算法,最终生成商品推荐结果集。

(1)商品关注度的几项指标

商品关注度是商品推荐的主要指标,生成商品推荐页面时按该用户对商品的关注度降序取出。表 1 是优先级从高到低的影响商品关注度的几项指标。

表 1 商品关注度的几项指标

| 指标名称 | 关注度 |
|------------|-----------------------------|
| 是否购买 | 购买该商品越多关注度越高。 |
| 所在地域 | 用户和商家的地域越近,商品关注度越高。 |
| 商品评分 | 对商品评分越高,关注度越高。 |
| 浏览次数 | 浏览商品次数越多,关注度越高。 |
| 搜索商品 | 搜索商品次数越多,搜索关键字和商品月接近,关注度越高。 |
| 浏览时长 | 浏览该商品次数越多,关注度越高。 |
| 相似用户的商品关注度 | 和其他用户越相似,和该用户商品关注度越接近。 |

(2)协同过滤算法

协同过滤推荐技术是推荐系统中最为成功的技术之一,它始于 20 世纪 90 年代。其基本思想是:找到与当前用户 c_{cur} 相似的其他用户 c_j ,计算对象 s 对于用户的效用值 $u(c_j, s)$,利用效用值对所有 s 进行排序或者加权等操作,找到最适合 c_{cur} 的对象 s^* 。其思想在日常生活中也有体现,例如,对一个人 M 不太了解,但对 M 的朋友 F 很熟悉,可通过 F 可推

断出 M 的一些信息。在本系统中可以根据用户信息表、商品评论表的信息推算出和其他用户的相似度,进而查询出相似用户的商品关注度。

(3) 序列模式分析

序列模式是从序列数据库中发现相对时间或者其他顺序所出现的高频率子序列。目的是通过在带有交易时间属性的交易数据库中发现频繁项目序列以发现一时间段客户的购买活动规律。本系统中可根据用户购买记录表、用户搜索记录表、用户浏览日志表、web 日志表的相关信息推算出用户对商品的感兴趣程度。

(4) 聚类分析

聚类是将一个数据集中在某些方面相似的数据成员进行分类组织的过程。一个聚类就是一些数据实例的集合,这个集合中的元素彼此相似,但它们都与其他聚类中的元素不同。本系统通过聚类算法将用户分组从而定位一组用户对某类商品的关注度。

经过以上算法,得出商品推荐结果集,存入预处理结果表中,该表结构比较简单分别为:表主键、用户登录名或 Session id、商品表主键、关注度。

2.5 商品推荐

用户访问该网站时取 Session id 和登录名,推算

出与其相似的用户,从预处理结果表中,按照商品的关注度从高到低,取一定数量的商品,生成商品推荐页面呈现出来。

3 结 语

本文讨论了美家 365 网站商品智能推荐系统的系统需求和框架设计,该系统使网站具备一定的智能性,不仅能提供个性化服务,还能有效解决用户找不到满意商品,商家找不到正确促销对象的问题,防止顾客流失。系统不仅能满足美家 365 网站的业务需求,还对类似的系统开发有一定的参考价值。

参考文献:

- [1] 夏明波,王晓川.序列模式挖掘算法[J].计算机技术与发展,2006,16(4):4-6.
- [2] 李燕,冯博琴,鲁晓锋.Web 日志挖掘中的数据预处理技术[J].计算机工程,2009,35(22):44-46.
- [3] 李晓明.搜索引擎:原理、技术与系统[M].北京:科学出版社,2005.
- [4] 邱哲,符滔滔.开发自己的搜索引擎——Lucene 2.0 + Heritrix[M].北京:人民邮电出版社,2007.
- [5] 许海玲,吴潇,李晓东.互联网推荐系统比较研究[J].软件学报,2009,20(2):350-359.
- [6] 张晗,潘正运,张燕玲.旅游服务智能推荐系统的研究与设计[J].微计算机信息,2006,22(5):170-172.

Intelligent Recommendation System in Meijia 365 Online Shopping Mall

QI Yong-jun¹ ZHAI Zhi-ping²

(1. Administration Center of Networks, North China Institute of Aerospace Engineering, Langfang 065000, China;

2. Library, North China Institute of Aerospace Engineering, Langfang 065000, China)

Abstract: Intelligent recommendation system, which is a important means of information filter, is an important way to solve the problem of information overload. It can solve the issues of electronic commerce effectively. For example, users can not find satisfaction goods or businesses, nor can they find the right promotional objects. In the paper, the intelligent recommendation system of Meijia 365 online shopping mall is introduced from functional requirements to system architecture. Also the data collection process, data cleaning, data preprocessing are discussed. In short, the system not only meets the demand of the users, but also has certain reference value for the similar system.

Key words: online shopping mall; intelligent recommended; study; design