

协同过滤推荐算法的研究

韩 路^{1,2} (1.沈阳理工大学, 辽宁 沈阳 110159; 2.辽南技师学院, 辽宁 营口 115003)

摘要: 随着互联网的普及, 网络资源的激增, 网上购物的交易方式正在改变着传统的商业模式。为了提供精确而又快速的推荐, 研究者提出了多种推荐算法。本文将针对电子商务发展的需求, 通过协同过滤推荐算法的文献综述, 对传统过滤算法无法适用于用户多兴趣下的推荐问题进行了剖析, 提出了一种基于用户多兴趣的协同过滤推荐改进算法, 分析了基于用户多兴趣的协同过滤推荐算法的电子商务系统。

关键词: 数据挖掘; 协同过滤推荐算法; 电子商务

1 协同过滤推荐算法的研究

目前的电子商务推荐系统中运用的推荐算法主要可分为三大类: 基于内容的推荐算法、基于规则的推荐算法和协同过滤推荐算法。

1.1 基于内容的推荐算法是信息过滤研究的派生和继续

基于内容的推荐系统需要分析资源内容信息, 根据用户兴趣建立用户档案, 用户档案中包含了用户的品位、偏好和需求信息。然后根据资源内容与用户档案之间的相似性向用户提供推荐服务。

1.2 关联规则挖掘技术在零售业得到了广泛的应用, 它可以发现不同商品在销售过程中的潜在相关性

基于规则的推荐技术在评价表上挖掘项目间的关联规则和用户间的关联规则为当前用户进行推荐。而使用用户关联进行推荐时, 用户关联的后件必须是当前用户, 使用用户关联的前件中的用户的共同兴趣模拟当前用户的兴趣, 模拟的可信度就是用户关联的可信度, 以此作为推荐的依据。

1.3 协同过滤的基本概念就是把这种推荐方式变成自动化的流程

协同过滤主要是以属性或兴趣相近的用户经验与建议作为提供个性化推荐的基础。透过协同过滤, 有助于搜集具有类似偏好或属性的用户, 并将其意见提供给同一集群中的用户作为参考, 以满足人们通常在决策之前参考他人意见的心态。

本人认为, 协同过滤技术应包括如下几方面: (1) 一种比对和搜集每个用户兴趣偏好的过程; (2) 它需要许多用户的信息去预测个人的兴趣偏好; (3) 通过对用户之间兴趣偏好相关程度的统计去发展建议那些有相同兴趣偏好的用户。

2 协同过滤推荐现有算法的分类研究与分析

正是因为传统协同过滤推荐算法存在着诸多问题, 研究者们才不断提出改进的协同过滤推荐算法。

2.1 全局数值算法

全局数值算法每生成针对一个用户的推荐项目列表就需要扫描用户评价数据库一遍, 这种方法能随数据的变化而变化, 实现也比较简单, 所以被大量才采用。但是在实践中数据稀疏性难以解决, 面对庞大的用户数据库, 推荐产生也非常耗时, 成为全局数值算法面临的主要挑战。

2.2 基于模型的算法

基于模型的算法只需扫描一遍用户评分数据库就可以完成对所有用户的推荐。优点是建立的模型相对于原始数据集而言小得多, 因此能有效缓解推荐算法的实时性问题。但模型具有滞后效应, 为了保证模型的有效性, 必须周期性的对模型进行更新。而模型的训练代价高, 因此该算法不适合数据更新频

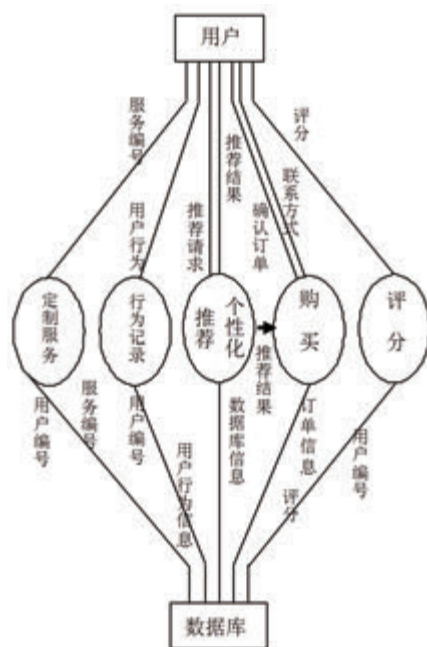
率快的系统。

2.3 组合推荐算法

(1) 协同过滤和基于内容的结合算法。两种算法的结合可以利用基于内容算法的优点, 对项目进行相似度匹配, 尤其当项目尚未得到用户评价的情况下也能推荐给用户, 避免新项目问题; 另一方面利用协同过滤的特点, 当用户数和评价很多时, 协同过滤推荐更准确。(2) 协同过滤和基于关联规则的结合算法。关联规则技术用于协同过滤系统是利用Apriori算法通过挖掘用户的评价记录的关联来进行推荐。该算法往往首先对客户的购买行为进行关联规则挖掘, 并进行单一客户的偏好建模; 然后, 应用协同过滤技术寻找与此客户兴趣相似的客户集, 并从客户集中找出和目标最相似的客户; 最后根据匹配集合求解推荐意见。

2.4 数据流图(DFD1层)主要处理分析

个性化推荐系统主要有以下处理过程: 用户定制服务、行为记录、个性化推荐、购买商品、对商品评分等。第一层数据流程图可如图1所示:



3 结束语

电子商务推荐系统是个新兴的研究与应用领域。随着用户需求水平的提高, 推荐算法与系统的研究在不断发展和完善。文中提出的基于用户多兴趣的协同过滤推荐算法, 正是为了解决现实中存在的用户兴趣问题而产生的。算法中由于对项目进行了分类, 所以跨越项目类别和推荐的新异性在一定程度上可能

不及传统的协同过滤推荐算法。这将在未来的研究中要进一步思考和研究的问题。

[参考文献]

- [1] 陈景艳, 苟娟娟, 著. 电子商务技术基础. 北京: 电子工业出版社, 2003. 9.
- [2] 黄解军, 万幼川. 基于数据挖掘的电子商务策略[J]. 计算机应用与软件, 2004, 21(7): 12-13.
- [3] R. Cooley, B. Mobasher, and J. Srivastava, Data preparation for mining World Wide Web browsing patterns. Knowledge and Information Systems, 1(1): 5-32.