

文章编号: 1007-130X(2004)05-0007-04

电子商务推荐系统研究^{*}

An Exploration of the Recommender Systems in E-Commerce

黎星星, 黄小琴, 朱庆生

LI Xing-xing HUANG Xiao-qin, ZHU Qing-sheng

(重庆大学计算机学院, 重庆 400044)

(School of Computer Science, Chongqing University, Chongqing 400044, China)

摘要:目前,电子商务网站之间的竞争日益加剧,赢得这场竞争的关键之一是如何更好地了解用户需求以提供更令人满意的服务,于是以AI领域的agent技术为依托,电子商务推荐系统就逐步发展起来。本文首先介绍推荐系统的表现形式,然后分类研究各种主流推荐技术,并在此基础上深入分析和比较这几种推荐技术的异同及其应用范围,最后探讨实际应用效果最好的混合推荐系统技术。

Abstract: E-Commerce companies compete each other nowadays, and one key to winning the competition is to get knowledge about customers' consuming preferences so as to establish better adequate personalized services to satisfy the customers. On this background, recommender systems gradually develops, supported by the agent technologies derived from Artificial Intelligence. The technologies to educe recommendation and their modes of representation are explored in this paper, afterwards similarities and differences among mainstream recommendation technologies and their application ranges are analyzed. At last, a hybrid recommender system technology is deeply investigated.

关键词: 推荐系统; 电子商务; agent

Key words: recommender system; e-commerce; agent

中图分类号: TP311

文献标识码: A

1 引言

Harvard 商学院 Joe Ping 教授在 1993 年的《大规模定制》一书中提出,现代企业应该从大规模生产(以标准化的产品和均匀的市场为特征)向大规模定制(为不同客户的不同需求提供不同的商品)转化^[1]。这就是说,企业至少应提供多种产品以满足多类用户的不同需求,实现大规模定制的方法之一就是推荐系统。

推荐系统泛指任何将个性化推荐作为输出的系统或在大规模可选对象中依据个人喜好引导用户做出选择的系统。然而,电子商务推荐系统则向用户提供商品信息和建议,模拟销售人员帮助用户完成购买过程。

2 推荐技术的分类

应用于电子商务推荐系统的推荐技术可以有多种分类。我们的分类是依据作出推荐所需要的数据

* 收稿日期: 2003-01-22 修订日期: 2003-05-12

作者简介: 黎星星(1978-),男,重庆人,硕士生,研究方向为软件开发环境和Java技术;黄小琴,硕士生,研究方向为网络信息系统和Java技术;朱庆生,教授,研究方向为多媒体传输与压缩和计算机网络。

通讯地址: 400044 重庆市重庆大学计算机学院 Tel: (023)65102506; E-mail: jeanxingxing@yahoo.com.cn

Address: School of Computer Science, Chongqing University, Chongqing 400044, P. R. China

和这些数据的使用方法。包括(1) 后台数据: 推荐开始前系统所具有的信息;(2) 输入数据: 为得出推荐与系统交互的信息;(3) 综合后台数据和输入数据而作出推荐的算法。

鉴于此, 我们可以列出如表 1 所示的五种推荐技术。其中, I 是商品集合, U 是已知其兴趣的用户集合, u 是需要进行推荐的个人用户, i 是能预测个人用户兴趣的商品。

表 1 推荐技术

推荐技术	后台数据	输入	处理
协作过滤推荐	用户集合 U 对商品集合 I 的评价	个人用户 u 对商品集合中部分商品的评价	从用户集合中识别出与个体用户 u 兴趣相似的用户, 以这些用户对具体商品 i 的评价为依据作出推断
基于内容的推荐	商品集合 I 中商品的特征描述	个人用户 u 对商品集合中商品的评价	产生一个与个人用户 u 的评价相应的分类器, 并将其应用于商品 i
基于人口统计的推荐	用户集合 U 的人口统计信息, 及这些用户对商品集合 I 中商品的评价	个人用户 u 的人口统计信息	从用户集合中识别出与个体用户 u 的人口统计信息相似的用户, 以这些用户对具体商品 i 的评价为依据作出推断
基于效用的推荐	商品集合 I 中商品的特征描述	个人用户 u 的兴趣效用函数	将该函数应用于商品并决定商品 i 的等级
基于知识的推荐	商品集合 I 中商品的特征描述, 及这些商品如何满足一位需要的知识。	关于个人用户的需要或兴趣的描述	推断出商品 i 与个人用户需要之间的匹配

(1) 协作推荐。协作推荐的核心思想是认为用户会倾向于购买具有相似意向的用户群所购买的商品。因此, 它在预测某个用户的商品购买倾向时是根据一个用户群的情况而决定的。

在协作推荐系统中, 用户描述的典型方法是采用以商品及其评价为分量的向量来表示, 向量将随着用户与系统交互时间的增加而不断增大。有的系统采用基于时间的评价折扣来解决用户兴趣的转移问题。评价可以是二值的(喜欢/不喜欢), 也可以用实数表示, 用以表示喜欢的程度。一些有名的协作推荐系统还可能是基于记忆的, 即直接使用相似度来比较用户; 也可能是基于模式的, 模式由评价历史数据导出并用于预测, 它使用了包括神经网络、潜在语义索引、贝叶斯网在内的多种学习技术。

协作推荐技术的最大优点是完全不需要任何被推荐商品的机器表示, 因而能在因用户品位变化而导致对商品偏好发生改变的复杂场合下高效工作(因为用户的品位都会变化, 总有人兴趣相投)。Schafer 等称之为“人与人的相关性推荐”。

(2) 基于人口统计的推荐: 旨在根据个人特征对用户分类, 并基于人口统计信息作出推荐。早期的 Grundy 是通过交互式的对话来收集个人信息, 用户的反应与一个人工创建的模式库相匹配。另外, 还有一些系统是采用机器学习来得到一个基于人口统计信息的分类器。

基于人口统计的推荐系统与协作推荐似乎相似, 但实际使用的数据完全不同, 其优点在于不需要用户评价历史数据。

(3) 基于内容的推荐: 是信息过滤的派生和继续, 是用相关特征来定义所要推荐的商品。例如, NewsWeeder 等新闻过滤的文本推荐系统就是采用文本中的单词作为文本的特征。系统通过学习用户已评价过的商品的特征来获得对用户兴趣的描述。

用户描述类型的产生取决于系统所采用的学习算法, 判定树、神经网络、基于向量的表示等技术都可应用于其中。与协作推荐相同, 基于内容推荐的用户描述也是长效(长期存在)型的, 它将随着系统对用户偏好的学习而不断更新。Schafer 等称之为“商品与商品的相关性推荐”。

(4) 基于效用的推荐: 与基于知识的推荐一样, 并不概括出用户的长效型描述, 而是基于用户需要和可选集之间匹配的评估之上, 通过计算商品对用户的效用来作出推荐。当然, 焦点问题是如何为每一位用户创建出合适的效用函数。

用户描述就是系统为用户创建的效用函数, 采用受限制满足技术来确定最佳匹配。它的优点是能在效用函数中考虑如卖主的可靠性、产品的可获得性等非产品因素。

(5) 基于知识的推荐: 通过推断用户的需要来作出推荐。这种推荐系统具有特定商品满足特定用户需要的知识, 并由此推导出用户需要与某一推荐物品的相互关系。

用户描述可以是支持这种推导的任何知识结构。在 Google 这种最简单的情况中, 用户描述仅仅是用户所构造的查询语句。其它系统还可能采取更详细的用户需求描述方法, 如引入事例推理技术。系统使用的知识也可以有多种形式, 如

Google 使用 Web 页面间的链接信息来推断流行度和权威性。其实,象 Persona-Logic 这种基于效用的推荐系统也是使用了关于功能的知识来计算效用函数的。

3 推荐技术的比较

我们需要先介绍一下推荐系统中常遇见的“装料问题”。它涉及两个既相关又彼此区别的问题:

(1) 新用户问题: 推荐系统中对用户的分类是依据目标用户与其它用户的比较, 这种比较主要基于不断累积的用户评价, 所以作出较少评价的用户群(如新用户)将难以分类。

(2) 新商品问题: 与新用户问题相似, 一个未获得足够评价的新商品也不易被推荐出去。它常出现在不断增加新商品、但用户仅对其中一小部分做出评价的情况下。这也被称为早期评价者问题, 因为最早对新商品作出评价的用户并不能由此得到任何好处。这就需要推荐系统引入一些新机制来激励用户多作评价。

协作推荐依赖于用户间评价的重叠, 故当用户评价较少时就会出现“评价稀少”问题, 即只有较少用户对同一件商品作出评价。特别是在新系统的开发初期, 会有大量商品出现, 如果用户基群不够大, 用户间能共享的商品评价将会很少。此问题可以通过基于模式的方法得到缓解, 如用单值分解法对评价比较空间进行降维。这说明, 协作推荐技术最适于用户兴趣密度相对较高而商品数量相对较少且恒定的情况。

另外, 协作推荐适合于用户兴趣明显的情况, 而不适合于“灰色绵羊”情况, 即一位用户游离于不同用户群体之间。

基于人口统计的推荐同样存在“灰色绵羊”情况, 但它没有“新用户”问题, 因为此系统不需要用户对商品的评价。但是, 搜集必需的个人统计信息却是这种推荐系统所面临的问题。随着个人隐私问题越来越敏感, 用户可能拒绝对系统透露或虚构用来预测用户偏好的、极为关键的个人信息。所以, 基于人口统计的推荐可能会逐渐减少。

基于内容的推荐需要积累足够数量的评价才能构建出一个可靠的分类器。

与协作推荐相似, 基于内容的推荐也因与所推荐的商品紧密相关而存在不足。例如, 基于内

容的电影推荐系统必须依赖于为每一部电影所编写的作者名称、情节概述等描述资料, 因为电影本身并不能被系统所理解。所以, 这些作为系统数据的描述语言就起着举足轻重的作用了。

协作推荐和基于内容推荐都存在着“部长职务效应”问题。一个理想的推荐系统不会建议用户购买已经持有的股票, 推荐已看过的电影。但是, 对于一个新系统, 因为那些与用户已有商品极其相似的其它商品可能表现出对用户有用的新价值或新特征, 故会被推荐, 可这些商品却不值得购买。基于内容的 DailyLearner 推荐系统采用了相似度上限的方法, 滤掉了与用户已看过新闻相似的其它新闻。

基于效用和基于知识的推荐系统都没有“装料问题”和“评价稀少”问题, 因为他们的推荐并不是基于用户数据的累积。基于效用的推荐需要系统构建一个能涵盖所考虑商品全部特征的效用函数。它的优势是能综合考虑决定商品价值的各种特征因素, 如递送计划、担保条款等等, 而不仅仅是产品具体的性能特点。这些非产品因素有着特殊的作用, 比如产品递送的速度就对急于使用的用户显得至关重要。因此, 基于效用的系统会让用户表达出所有的考虑因素, 以作出推荐。

用户要构建一个完整的兴趣效用函数, 就必须衡量每一个因素的权重。这就为用户与系统间的交互增加了额外的负担。有的系统是将价格、质量、递送时间等因素显式地列出供用户挑选, 有的系统(如 Persona-Logic)则通过一系列问卷调查推导用户兴趣函数。当然, 一个完整的显式效用函数会对有特殊购买需求的专业用户有很大帮助, 但对一位具备较少相关知识的临时用户就会增加麻烦。这说明基于效用的推荐系统不适合于临时浏览用户。

基于知识推荐的重点和难点是需要获取知识。它会涉及三种知识:

(1) 目录知识: 即所推荐商品及其特征的知识。例如, Entrée 食品推荐系统知道泰式烹饪是亚洲烹饪风格的一种。

(2) 功能知识: 系统应当能将用户的需求和可能满足该需求的商品的特征进行比较。例如, Entrée 食品推荐系统知道拥有平静海滨景色的饭店能满足一位用户关于浪漫晚餐的需求。

(3) 用户知识: 为了提供优质的推荐服务, 系统必须要有一些关于用户的知识。这可能是一般

的统计信息或是特殊的需求信息。

基于知识推荐的优势在于它对用户的要求较少,因此适用于用户临时随机浏览的情况。虽然它不象协作推荐那样能为每位用户在用户群中找到合适的位置,但只要所依据的知识允许,它作出的推荐就能为多数用户所用。因此,它也没有新系统刚开始时常有的低质量推荐问题。

表2总结了我们的五类推荐系统,指出了各自的优势与不足。

表2 推荐方法的比较

技术	优势	缺陷
协作过滤推荐	A. 能跨类型推荐 B. 不需要域知识 C. 自适应性好, 随时间推移, 推荐质量会提高 D. 充分的隐式反馈	I. 新用户装料问题 J. 新商品装料问题 K. “灰色绵羊”问题 L. 推荐质量依赖于大规模历史数据 M. 稳定性与可塑性问题
基于内容的推荐	B, C, D	I, L, M
基于人口统计的推荐	A, B, C	I, K, L, M N. 必须收集人口统计信息
基于效用的推荐	E. 不存在装料问题 F. 对用户偏好的变化敏感 G. 能涵盖非产品性能因素	O. 用户必须输入效用函数 P. 静态推荐能力(不能学习)
基于知识的推荐	E, F, G H. 能比较、匹配用户需求和产品特征	P Q. 需要知识引擎

协作推荐和基于人口统计推荐都具有跨类型推荐的独特优势,能向用户介绍一些新商品。若能事先获得一些关联知识,基于知识的推荐系统也能做到这一点。

所有基于学习的推荐(协作、基于内容和基于人口统计)都存在着“装料”问题。在系统中,用户描述模型一旦建立,除非用户给出明显的评价作为提示,否则用户兴趣就不易更新。

对于协作推荐和基于内容推荐,“装料”问题还有边界效应,不能使系统充分体现全部的优势。所以,基于学习的推荐更适合于愿意投入时间和精力使自己的兴趣能为系统所知的用户。然而,基于效用和基于知识的推荐就不存在这种问题,因为它们并不依赖描述用户兴趣的历史数据。

4 混合推荐系统

混合推荐系统整合两种或更多推荐技术以取得更好的实际效果。最常见的作法是将协作过滤推荐与其它某一种推荐技术相结合,以克服“装

料”问题。常见的混合推荐方法有:

(1) 权重型: 由多个推荐方法的计算结果组合而成;

(2) 转换型: 根据具体情形在几个推荐方法间转换;

(3) 合并型: 几个推荐方法的结果一起呈现;

(4) 特征组合: 不同推荐数据源的特征信息被一个推荐算法所使用;

(5) 瀑布型: 后一个推荐方法优化前一个推荐方法;

(6) 特征递增型: 后一个推荐方法的输入包含了前一个推荐方法的输出;

(7) 元层次型: 前一个推荐方法所得出的模型作为后一个推荐方法的输入。

5 结束语

通过推荐系统实现个性化服务已成为电子商务应用的一项新兴技术,这也有助于网站内容和结构自适应性的实现。同时,人们可以将已长期研究的系统的人性化、智能化的经验总结和实现方法等应用于推荐系统之中。我们相信,推荐系统技术的应用将使电子商务 Web 站点更人性化、更个性化、更能符合每个用户的喜好。

参考文献:

[1] Schafer J B, Konstan J, Riedl J. Electronic Commerce Recommender Applications[J] . Journal of Data Mining and Knowledge Discovery, 2001, 5(1-2): 115-152.

[2] Schafer J B, Konstan J, Riedl J. Recommender Systems in E-Commerce[A] . Proc of ACM Conf on Electronic Commerce[C] . 1999. 158-166.

[3] Marko Balabanovii. Recommender Systems: Interfaces and Architectures[EB/OL] . <http://web.engr.oregonstate.edu/herlock/rsw2001/final/whos.who.norattendess/1balabanovic.pdf> 2001-12.

[4] 许敏,邱玉辉.电子商务中推荐系统存在的问题及其对策研究[J] . 计算机科学, 2001, 28(4): 122-124.

[5] Good N, Schafer J B, Konstan J. Combining Collaborative Filtering with Personal Agents for Better Recommendations[A] . AAAI-99 [C] . 1999. 439-446.

[6] Sarwar B M, Karypis G, Konstan J A, et al. Item-Based Collaborative Filtering Recommendation Algorithms[A] . WWW01[C] . 2001.

[7] Sarwar B M, Karypis G, Konstan J A, et al. Application of Dimensionality Reduction in Recommender System—A Case Study[A] . ACM WebKDD Web Mining for E-Commerce Workshop[C] . 2000.

[8] David G, David N, Brian M, et al. Using Collaborative Filtering to Weave an Information Tapestry[J] . Communications of the ACM, 1992, 35(12): 61-70.