

基于知识的电子商务智能推荐系统平台设计

刘平峰, 聂规划, 陈冬林

LIU Ping-feng, NIE Gui-hua, CHEN Dong-lin

武汉理工大学 经济学院, 武汉 430070

School of Economics, Wuhan University of Technology, Wuhan 430070, China

E-mail: lpf@mail.whut.edu.cn

LIU Ping-feng, NIE Gui-hua, CHEN Dong-lin. Design of platform for knowledge based intelligent electronic commerce recommender systems. Computer Engineering and Applications, 2007, 43(19): 199-201.

Abstract: In this paper the shortcomings of traditional recommendation techniques are analyzed. Also the characteristics and development of knowledge-based recommendation techniques are elaborated. The basic theory and methodology that should be developed for the knowledge-based intelligent electronic commerce recommender systems are proposed to solve the problems existing in the current knowledge-based recommender systems. The architecture of the knowledge-based intelligent electronic commerce recommender system is designed and the basic theory of it is elaborated.

Key words: electronic commerce; recommender system; intelligent recommendation; knowledge

摘 要: 分析了传统推荐技术存在的不足, 阐述了基于知识的推荐技术的特点及其发展。针对现有基于知识的电子商务推荐系统中存在的不足, 提出了基于知识的电子商务智能推荐需要解决的基本问题, 设计了基于知识的电子商务智能推荐平台的逻辑框架, 并阐述了其工作原理。

关键词: 电子商务; 推荐系统; 智能推荐; 知识

文章编号: 1002-8331(2007)19-0199-03 **文献标识码:** A **中图分类号:** TP183

1 前言

协同过滤技术和基于内容的过滤技术在推荐系统中得到了广泛的应用, 并取得了一定的成功, 但与此同时这两种过滤技术存在的缺陷极大地影响了推荐质量, 甚至会出现推荐失败的情形, 不能推荐客户喜欢的商品或推荐了客户不喜欢的商品, 都会影响客户对网站的信心, 从而造成客户流失, 甚至损害企业形象。协同过滤技术只有在大量用户已对足够多的商品进行了评分的情形下才能取得较好的推荐效果, 不能解决新商品项或新客户造成的冷启动问题, 客户兴趣相关性的假设并不总是成立。如面临以下情形时则会产生推荐失败: 不同的客户基于不同的原因喜欢同样的商品; 同一个客户基于相同的原因喜欢两个截然不同的商品; 一个客户对于某商品具有深层次需求, 不仅仅是整体偏好的问题。基于内容的过滤只能向客户推荐和客户曾经喜欢过的商品相类似的商品, 即存在过于专门化的问题, 不能给客户带来新的惊喜。由于需要基于用户对商品的历史评价来构建有效的用户概要, 基于内容的过滤技术同样存在冷启动的问题。另外商品的特性本身也会对推荐质量产生影响, Konstan 总结了适于过滤推荐系统的产品应该具有的四个特征^[1]: (1) 产品针对特定的群体; (2) 产品价值不高, 人们能容忍推荐的不准确性; (3) 产品相关性不随时间衰减; (4) 产品

没有组合效应(portfolio effect), 即用户已经选择的产品不会影响用户对其它具有相似描述产品的兴趣。象电影、书籍和音乐这样的商品满足这四个特征, MovieLens、Amazon、CDNow 等推荐系统都是基于这样的商品领域。但很多产品并不同时具有这些特征, 如课程的选择、金融服务的选择具有明显的组合效应。

为了克服基于内容的过滤技术和协同过滤技术各自存在的缺陷, 出现了将基于内容的过滤技术和协同过滤技术相结合的混合推荐系统, 如 P-Tango^[2]和 Yoda^[3]。基于内容的过滤技术和协同过滤技术都是基于学习的推荐技术。基于学习的推荐技术都存在冷启动的问题, 只有在用户愿意花时间让系统了解其偏好时才能有较好的工作效果, 对于临时用户则无能为力。另外基于学习的推荐技术一旦经过学习建立起用户概要, 则推荐结果相对于用户偏好的改变具有滞后性。比如说一个肉食者变成了素食者, 他仍然会得到系统关于肉食的推荐, 直到他对肉食和素食新的评分覆盖旧评分的影响为止。一些适应性系统采用窗口技术^[4]或遗忘因子来逐步减弱旧评分对推荐结果的影响, 但与此同时关于用户长期兴趣的信息也会丢失。

基于知识的推荐系统使用用户知识和产品知识, 通过推理什么产品能满足用户需求来产生推荐。这种推荐系统不依赖于用户评分等关于用户偏好的历史数据, 故其不存在冷启动方面

基金项目: 国家自然科学基金(the National Natural Science Foundation of China under Grant No.70572079)。

作者简介: 刘平峰(1972-), 男, 讲师, 博士生, 研究方向: 商务智能、知识管理与知识工程; 聂规划(1957-), 男, 教授, 博导, 研究方向: 信息资源管理、知识管理与知识工程; 陈冬林(1970-), 副教授, 博士, 研究方向: 商务智能, 电子商务解决方案。

的问题。基于知识的推荐系统响应用户的即时需求,当用户偏好发生变化时不需要任何训练。

2 相关研究工作

基于知识的推荐系统包括三类:基于 KDD(Knowledge Discovery in Database)的推荐系统、基于 CBR(Case Based Reasoning)的推荐系统和基于知识推理的推荐系统。严格意义上讲基于 KDD 的推荐系统不属于基于知识的推荐系统,它主要是在传统推荐系统中应用了数据挖掘技术来发现数据中隐含的有价值的知识,并运用发现的知识构建关于用户兴趣的概要模型(User Profile)或向用户作出推荐。基于 CBR 的推荐系统引导用户逐步明确需求,采用 CBR 技术对已有案例进行过滤和选择,根据用户对推荐结果的反馈作出修正与调整,这种系统被 Burke 称为“FindMe”系统^[5]。如 Entrée^[6]就是一个基于 CBR 的推荐饭店的系统。与基于 CBR 的推荐系统不同,基于知识推理的推荐系统包括用户知识、商品知识和功能知识三部分知识,采用统一的知识表示方式来表示用户概要和商品,并采用各种知识推理方法在用户概要和商品之间作出匹配,在此基础上向用户作出推荐。

为了实现知识的共享和重用,语义 Web 研究中引入本体技术,且正在被不断地实践。如果广泛采用语义本体的技术,那么“精细、准确和自动化”的搜索和智能推荐就能够实现。在基于知识的推荐系统中,语义本体被广泛地应用于领域知识、用户概要知识和产品知识的表达。Middleton 等开发的实验系统 Quicksetup 和 Foxtrot^[7]用于学术研究文献的推荐,这两个系统通过监测用户的行为和用户的相关反馈,使用研究文献主题本体来表示用户概要。Ontogator 系统^[8]用于图像推荐,在该系统中,本体被用于支持基于视图的搜索和数据源标注,领域知识、资源标注和信息检索都是基于语义丰富的本体结构而不是简单的关键词分类,其推荐采用了基于规则的方法,应用规则从底层知识库中发现与用户选择图像紧密相关的图像。Marek Hatala 和 Ron Wakkary 研发的 echo 系统^[9]根据博物馆中用户所处的具体场景向用户提供相应的音频对象,该系统研究在抽象语义层而不是内容层来连接环境、交互对象和用户,探索建立场景感知自适应系统的可能性,系统中用户和对象都建立明确的模型,本体作为高层抽象,用户、对象和环境都用本体来标注,对象选择的标准用规则表示,使用规则在音频对象及博物馆环境的语义描述与用户模型之间作出匹配。

但现有的基于知识的推荐系统基本上都是针对具体的应用领域设计,从具体应用的局部角度描述了用户概要知识或商品知识或推荐功能知识的表示与应用,没有形成系统的智能推荐的理论和方法。另外在知识表示的范围上存在局限性。比如没有表示象这样的知识;在不同的系统环境中不同的推荐技术具有不同的推荐质量和效率,Scahfer 提出的元推荐系统^[10]体现了这一思想,但其主要基于数据融合而不是知识推理。

3 基于知识的电子商务智能推荐系统平台设计

3.1 基本原理

与传统的推荐方法不同,基于知识的电子商务智能推荐(Knowledge Based Intelligent Electronic Commerce Recommendation,以下简称 KBIECR)在统一的语义互联环境中获取用户知识和商品知识,通过功能知识的推理或语义匹配向用户

作出推荐,能够根据推荐系统所处的具体环境选择相应的推荐策略,在必要的时候调用传统的推荐技术,结合传统推荐技术的优点采用混合推荐技术向用户作出推荐以获得最佳推荐质量。KBIECR 的实现需要解决以下理论方法问题:

(1)知识的统一表示。基于知识的电子商务智能推荐系统(Knowledge Based Intelligent Electronic Commerce Recommender System,以下简称 KBIECRS)中一切要素都需要采用统一的方式进行描述,具体包括:用户概要、用户所处的场景、商品知识、功能知识、系统组件(包括模型、算法、工具、代理等),以及上述要素之间的关系。除了系统要素外,还有推荐策略以及不同策略下系统要素之间的交互过程也需要进行统一描述。

(2)知识的有效获取。KBIECRS 中的知识包括:商品知识、功能知识、系统组件知识、推荐策略知识、用户概要知识、用户场景知识。这些知识的获取采用何种方法与技术需要明确。

(3)智能推荐系统推理机制与推荐结果的产生。KBIECRS 根据系统中的用户知识和商品知识,利用功能知识进行推理以向用户产生推荐结果,用户知识、商品知识和功能知识的表示决定了选择知识和使用知识的方法以及具体的推理方法。经由推理机推理出的推荐结果可能是有多个商品的推荐列表,也可能为空集。针对这两种不同的情况,需要确定向用户提供合适的推荐商品列表的方法。

(4)推荐系统的适应性。与以往推荐系统适应性研究针对用户偏好不同,KBIECRS 的适应性表现在系统能够根据用户知识和系统信息资源状态选择最合适的推荐算法与工具向用户作出推荐,以取得最佳的推荐质量,称之为推荐策略的自适应性。

(5)基于语义的交互。为推荐系统建立一个语义互联环境,系统要素之间、系统与用户之间都在统一的语义空间中基于语义进行交互。为用户提供基于语义的可视化的交互界面,向用户解释推理结果的产生过程,用户提供基于语义的反馈,系统根据用户反馈对推荐结果进行修正。

(6)知识的规范聚合。KBIECRS 中的知识处于动态的增长过程中,面临越来越多的知识,需要建立起知识的有效管理手段防止知识的无限膨胀,为用户建立知识的逻辑视图,知识的一致性需要得到保障,知识的统一检索机制需要建立。

3.2 平台设计

KBIECRS 平台的逻辑框架如图 1 所示,主要由四层组成,由上至下分别为:语义交互接口层、语义平台和策略执行层、系统功能层、系统存储层。KBIECRS 的基本工作原理是:用户通过语义交互接口与系统进行交互,根据不同的推荐策略,用户及系统要素之间的交互过程不一样。推荐策略的选择可以采用两种方式:一种是事先确定的,即规定只采用哪一种或哪几种推荐方法,称之为“定制选择”;另外一种是先制定好规则或由系统学习得到相应规则,由系统在具体的执行环境中根据规则确定采用什么推荐策略最合适,称之为“自适应选择”。后一种方式较前一种方式提供了更大的灵活性,但由于计算任务的执行会占用系统资源和时间,对系统效率会有一定影响。确定了推荐策略后由推荐策略执行器根据该策略下系统执行过程的定义(该过程被定义为基于知识的工作流,采用 OWL-S 的 Process 模型进行描述)通过语义交互平台调用相应组件完成推荐功能。系统组件之间的交互基于语义,由于共享本体的存在,系统组件对语义有着一致的理解,保证推荐任务无偏差地

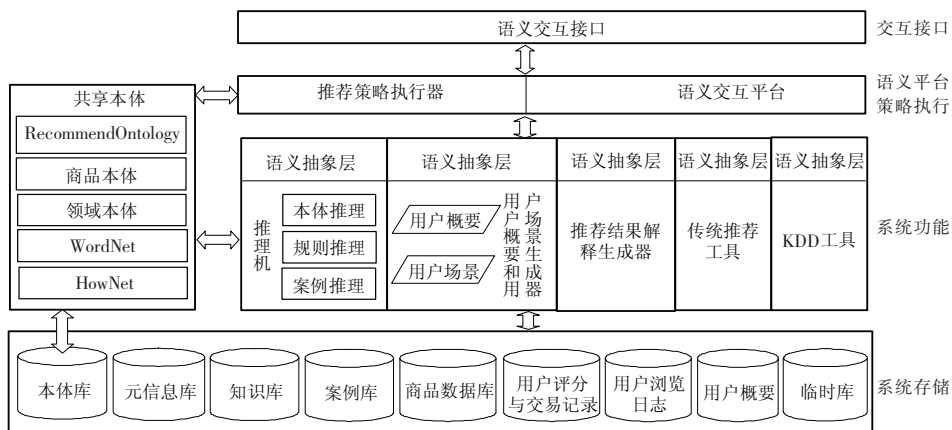


图1 KBIECRS 的逻辑框架图

执行。系统各层的主要功能如下:

(1)语义交互接口层。负责用户与系统之间基于语义进行交互,其实现形式是语义浏览器。语义浏览器可以以图形化的方式显示本体,用户可对本体进行操作来表明自己的需求和偏好,用户对系统的请求被封装成基于语义的形式后提交给系统,系统返回的推荐结果的解释以图形化的方式显示在浏览器中。

(2)语义交互平台与策略执行层。语义交互平台为系统要素之间基于语义的通讯提供了基础设施,负责转发要素之间的语义请求。推荐策略执行器负责根据选定的推荐策略有序地调用相应的系统要素来完成推荐功能。

(3)系统功能层。该层主要包括完成推荐任务所需的组件,各组件的功能为:

①推理机。是基于知识的推荐系统的核心组成部分,主要完成三种形式的推理,即基于本体的推理、基于规则的推理和基于案例的推理。推理机主要用于三个过程,用户知识和商品知识的功能匹配过程,推荐策略的自适应选择过程、用户概要和场景知识的生成过程。

②用户概要和用户场景生成器。用户概要可能是基于评分的形式、或者基于特征的形式、或者基于规则的形式,对于后两种形式,生成器需要调用 KDD 工具的功能。用户场景生成器负责收集与组织和推荐商品相关的用户场景信息。

③推荐结果解释生成器。在有些情况下,推荐系统需要向用户解释推荐结果是如何产生的。由于推理机的推理过程对于用户而言过于抽象,直接将推理过程显示给用户没有意义,推荐结果解释生成器将推理过程转换为状态图后将其发送到用户浏览器端显示,用户能够看见在状态的每一步变化是什么约束和规则导致的结果,从而帮助用户修正约束以获得所需结果。

④传统推荐工具。KBIECRS 是一个元推荐系统,除了可以使用基于知识的推荐技术外,还可以使用基于内容的过滤和协同过滤等传统的推荐技术或结合上述技术的混合推荐技术。在具体的应用中使用何种推荐技术取决于具体的推荐策略。

⑤KDD 工具。主要是指数据挖掘工具和机器学习工具。数据挖掘和机器学习技术被广泛用于推荐过程中,比如用户概要生成过程中用户行为知识的挖掘、文本分类器构建等。KDD 工具挖掘出的知识既可以用于基于知识的推荐技术中,也可以用于传统推荐技术中。将 KDD 工具和传统推荐工具分开设计提供了系统的灵活性。

为了建立语义互联环境,系统功能层的所有组件均构建了语义抽象层,该层负责语义访问与本地调用之间的映射,例如

Web 服务的 OWL-S 本体就是 Web 服务的语义抽象。

(4)系统存储层。该层负责推荐系统各种数据与知识的存储,具体包括:

①本体库。存储共享本体的定义,除了电子商务推荐领域本体外,还包括商品本体、应用领域本体以及支持自然语言处理的通用语义词典 WordNet 和 HowNet。

②知识库。存储推荐系统的功能知识,包括基于规则的知识 and 基于模型的知识。

③商品数据库。存储电子商务网站经营的商品数据。

④用户评分与交易记录。存储用户对商品项的评分以及用户历史交易记录。

⑤用户浏览日志。存储用户浏览网站的历史以便进行行为分析。

⑥用户概要。与用户概要生成器中基于会话的用户概要不同,此处存储的是持久形式的用户概要。

⑦临时库。用于存放推理过程和推荐结果解释生成过程的中间结果知识。

⑧元信息库。存储系统要素的元信息,包括组件的本体描述、知识库与各数据库的本体描述。

4 结语

基于知识的电子商务智能推荐系统能够克服传统推荐系统的不足,本文指出基于知识的电子商务智能推荐系统需要解决的基本理论方法问题,这些理论问题的解决将形成基于知识的电子商务智能推荐的理论方法体系。目前大部分理论方法问题已经得到解决,由于篇幅所限,本文着重设计了平台的逻辑框架,并阐述了其工作原理,今后将陆续行文探讨其它理论方法问题。(收稿日期:2007 年 1 月)

参考文献:

- [1] Konstan J.Recommender system research:perspective and thoughts [C]//Keynote Address,Workshop on Recommender Systems:Algorithms and Evaluation,Berkeley,California,1999.
- [2] Claypool M,Gokhale A,Miranda T.Combining content-based and collaborative filters in an online newspaper [C]//Proceedings of the 22nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'99),Berkeley,Calif, ACM,New York.