

基于用户关联行为的个性化搜索系统设计^①

刘声田¹ 卢守东² 刘忠强³

(1.山东广播电视大学 计算机与通信学院 山东 济南 250014;2.广西财经学院 计算机与信息管理系 广西 南宁 530003;3.山东广播电视大学 继续教育学院 山东 济南 250014)

摘要: 设计了应用于网上商店的个性化信息搜索系统,该系统根据查询条件和用户信息记录分析用户兴趣,把初步的兴趣商品结果提交给用户,并将用户对结果的使用情况回馈给数据库。系统可用于定位消费需求、挖掘用户的隐含表达;创新点是提出了用户关联行为方案,为系统设计了关联行为的模型。

关键词: 搜索系统;用户行为;网上商店;兴趣模型;向量空间模型

Design of a Personalized Search System Based on Related Actions of Users

LIU Sheng-Tian¹, LU Shou-Dong², LIU Zhong-Qiang³

(1. School of Computer & Communication, Shandong TV University, Jinan 250014, China;

2. Department of Computer & Information Management, Guangxi University of Finance and Economics, Nanning 530003, China;3. School of Continuing Education, Shandong TV University, Jinan 250014, China)

Abstract: This paper designs a personalized information search system(named PISS) applied to online shopping. According to users' searching issue and information database, PISS provides some initial interest merchandise. After analyzing the performance of the merchandise, the system records the result and returns it to the database. PISS can be used in finding consumers' needs and mining the data of connotative expression. The innovation of this paper is providing solution to related actions of users and designing the interesting model.

Keywords: searching system; user's action; online shopping; interesting model;vector space model

随着网上购物方式被越来越多的消费者接受,利用 Internet 宣传和销售商品已成为商业活动的重要手段,专门的网上商店也初具规模并日益成熟。根据 Hallerman 调查报告显示,88%的美国成人使用互联网搜索引擎查找主题信息,其中 51%购物之前进行 Internet 搜索^[1]。

在 Kotler 和 Keller 提出的消费者购物模型中,消费者购物过程由“问题识别”、“信息搜集”、“评价抉择”、“购买决定”和“购物行动”5个有序环节组成^[2],这种购物模型也适用于网上购物过程。对比传

统购物方式,网上购物的优点是:第一,商品价格较低。网上商品免去了物流、储藏、管理等费用,同等利润下销售成本较低;第二,客观影响因素较少。网上购物过程可以不受地域、时间、天气以及消费者身体状态等因素影响,也较少会遭遇影响购物过程的突发事件;第三,沟通方式更加便捷。搜索工具可以帮助消费者快速查询满足指定条件的商品,便于消费者快速、准确了解商品参数、出售者、价格以及其它消费者评价等信息,这些信息还有利于买卖双方通过电话、网络等方式即时议价、交流沟通。网上购物这种

① 基金项目:山东省教育厅高等学校优秀青年教师国内访问学者基金;教育部高等学校青年骨干教师国内访问学者基金

收稿时间:2009-06-11

新的购物方式,随着支付手段、买卖保证体系构建和完善,正在吸引更多的消费者;许多厂商看好此前景,纷纷推出自己的网上商店,比较著名的有联想网上商城、当当购书网、淘宝网等。

消费者在网上商店购物通常会使用该商店的购物系统。购物系统一般都会设置产品介绍栏目,满足消费者“问题识别”的需求,且这种介绍性栏目的实现难度不大;为消费者提供“信息搜集”服务则是现在研究的热点,特别是个性化的信息搜索。近年来新出现的技术可以用于改善消费者的“评价抉择”,如界面定制、商品对比等。

本文讨论的搜索系统使用搜索模型推理用户兴趣,为用户提供个性化商品信息搜索,目的是优化搜索结果,简化搜索过程,提高搜索效率;同时,还简单介绍了新技术在系统中的应用和特点。

通常的网上购物基本流程为:查找和选择所需商品;会员登录或注册;填写收货人信息与付款方式、配送方式等;提交订单;等待取货。在购物过程中,最耗费消费者时间的是第1步,即查找和选择所需商品。在网上商店中查找和选择商品的难易程度和效率,影响了消费心理,进而改变消费者的意愿。为了把用户所需求的商品信息准确提取出来,我们设计了针对网上商店(特别是大型的、综合的网上商店)的个性化搜索系统 **PISS (Personalized Information Search System)**,该系统能较准确地定位消费需求,并对隐含表达的需求进行挖掘。

1 个性化搜索系统

PISS 的工作原理是先从消费者输入的需求信息中提取关键词;然后把关键词映射到语义树,进行关联搜索,得到消费者搜索要求和隐含要求;接下来根据消费者对系统所提供的搜索结果行为推定用户兴趣、商品关注度。

从结构上看,**PISS** 由接口模块、信息处理模块和行为分析模块3部分组成,如图1所示。下面分别介绍每个组成部分的功能和设计思想。

1.1 接口模块

接口模块包括用户接口和管理员接口两部分。用户接口用于接收消费者输入的查询条件,并把查询条件传送给信息处理模块。查询条件可以是文本信息、列表选择信息、按钮提交信息或者特殊请求等信息;

用户接口以 **Web** 形式提供,是消费者浏览的网上商店一级页面、二级页面的组成部分。

管理员接口主要用于维护商品语义树、商品数据库和会员数据库,还可用来响应用户的特殊请求;同时,也可以用于管理员向用户直接推荐商品。管理员接口以单独 **Web** 页面提供,需要身份验证。

1.2 信息处理模块

信息处理模块的功能是先依据用户输入的查询条件提取商品关键字,然后定位到具体的、用户可能感兴趣的商品,并根据当前数据库信息推理。模块目标是为消费者提供出满足一定条件的、按照一定规则排序的商品列表。

信息处理模块包括查询处理、查询定位和需求推理3个过程。

1.2.1 查询处理

查询处理的作用是提取查询条件中的关键字。对于接收的文本查询条件,在处理时要进行分词操作,获取其中的名词或名词词组;对于列表选择信息,需要映射到指定的商品类关键字;对于按钮提交信息,进行相应的信息合并或者求交集的处理。查询处理的结果是一组跟商品名或者商品属性有关的关键字,如商品类、商品型号、生产厂商或者产地等商品信息。

1.2.2 查询定位

查询定位的过程是根据上一步骤的关键字及其组合,通过搜索商品语义树,准确地对应到网上商店的全部可能的、完整的商品信息,并把所有的这些商品信息暂时保存,为需求推理过程做准备。

1.2.3 需求推理

需求推理是对暂时保存的商品信息进行归类、筛选和排序的过程,其归类的依据、筛选的条件和排序的原则均需要推理判定。推理得到的结果是一个(组)最可能需要的商品(集),以及附加结果——跟该商品(集)有密切关系的其它相关商品。

从接收用户查询到向消费者发送需求推理的结果是可循环过程,允许消费者进行重新选择或者在原有的搜集结果之上二次搜索。

1.3 行为分析模块

本文的行为指的是消费者利用搜索系统在网上商店搜索商品以及对搜索结果和附加搜索结果的操作。模型中“交易状态”部分用于辅助购买行为的判定,用来标记是否成交。对于不愿意进行网上支付贷款的

消费者, 交易状态为“订购”、“收货”、“付款”、“退货/成交”4部分,“退货”状态超过预定时间则认定为交易成功(成交)。

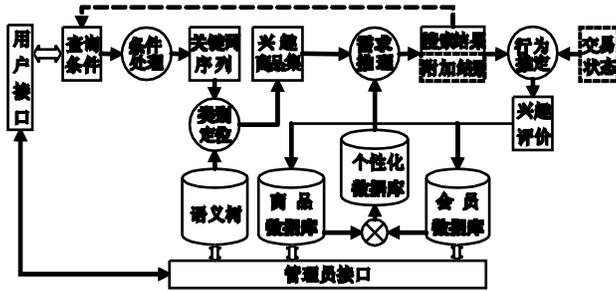


图1 个性化搜索系统图

1.3.1 行为分析假设

对用户的行为分析通常基于以下假设:

- ① 用户行为在一定程度上反映出用户的兴趣;
- ② 多数用户的共同兴趣也代表个人兴趣;
- ③ 不同行为影响兴趣的程度不同, 每个行为均有一个行为值。

同时, 我们遵循目前广泛认可的“使用网页表示了用户对网页主题有兴趣”的假设, 并且约定兴趣浓度与用户系列行为有关系, 这种关系在我们的模型中被定义为关联关系。

1.3.2 行为分析方案

PISS 的行为分析模块采用关联行为分析策略, 我们设计的方案有三个特点: 一是将具有直接关联的两个用户行为配对为关联关系, 并为每个关联关系设定一个正数数值; 二是把用户行为分若干等级, 不同等级的行为对应不同程度的商品兴趣影响度; 三是考虑了商品兴趣漂移特性, 兴趣度随时间变化逐渐衰减。

1.3.3 兴趣模型

主要由网页预处理模块、兴趣模型生成模块以及更新模块三个部分组成, 模型体系结构如图2所示。网页预处理模块调集用户在一定时间段的访问记录, 提取页面中的特征词, 产生VSM(网页空间向量模型); 根据VSM和用户行为的关联产生短期兴趣, 根据采用终态行为的积累计算长期兴趣; 最后结合短期兴趣和长期兴趣计算获得用户综合兴趣度。综合兴趣计算表达式为:

$$U = \alpha \times U^{cur} + (1 - \alpha) \times U^{per} \quad 0 < \alpha < 1 \quad (1)$$

其中, α 代表短期兴趣在综合兴趣模型中比例因子,

U^{cur} 是短期兴趣, 短期兴趣代表的是消费者意向, 主要用途是向消费者推荐相关信息; U^{per} 是长期兴趣, 代表了用户已经满足的需求, 用于向管理者反馈用户实际需求。 α 的取值反映了兴趣模型对短期兴趣的倾向。

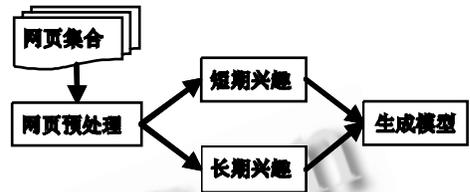


图2 用户兴趣模型体系结构

长期兴趣表达的是消费者稳定的需求, 可以根据网上商店历史订购和销售情况确定; 短期兴趣代表了消费者近期的关注, 需要统计消费者在网上商店的网页行为来评估, 具体方法是统计网页特征项。为了反映兴趣的时效性, 长期兴趣和短期兴趣需要定时更新, 方法是统计指定时期内的消费者行为。

1.3.4 网页特征项

网页特征项是网页关键词的抽象, 统计特征项信息对于匹配用户搜索关键词、为用户提供搜索结果具有重要意义。特征项获取借鉴了tf-idf表示方法^[3], 设计的特征项表达式如下:

$$\omega(t_k, p_i) = X + Y \quad (2)$$

$$X = \frac{tf(t_k, p_i) \times \log\left(\frac{m}{m_{t_k}} + 0.01\right)}{\sqrt{\sum_{\substack{t_k \in T \\ p_i \in P}} tf(t_k, p_i) \times \log\left(\frac{m}{m_{t_k}} + 0.01\right)}} \quad (3)$$

$$Y = \ln\left(\sum_{a_i \in A} pref(a_i)\right) \quad (4)$$

表达式(2)由两个部分组成, 第一部分是tf*idf表达式, T是特征项集合, P是网页集合, m是网页数量, m_{t_k} 是具有特征项 t_k 的网页数量; 第二部分是用户行为统计, A是用户行为集合, $pref(a_i)$ 是用户行为值统计表达式, 具体计算方法为:

$$pref(a_i) = 1 + \sum_{\substack{a_i \in A \\ \langle a_i, a_j \rangle \in R_a}} \left(\frac{\tau_{a_i} \times num_{a_i} - \tau_{a_i} \times (num_{a_i} - num_{a_j}) \times d^{\langle a_i, a_j \rangle}}{const} \right) \quad (5)$$

表达式(5)中的 R_a 是关联关系集合, τ_{a_i} 是行为 a_i 的行为权值, num_{a_i} 是行为 a_i 的发生次数, $d^{\langle a_i, a_j \rangle}$ 是

行为 a_i 和 a_j 的关联度。 $const$ 是调整参数, 特别当 $num_{a_i} < num_{a_j}$ 时, $const$ 取负值。表达式(2)的含义是统计行为 a_i 的总行为值并按关联度比例减去所有未产生的关联行为的行为总值, 采用减法表达可以减缓兴趣模型的增长速度。

2 辅助技术

在 PISS 中可以采用 CSS、Ajax 和 Agent 等辅助技术, 以改良系统可用性, 提高搜索精度和效率。

2.1 CSS 技术

CSS 标准由 W3C 组织制定, 用于控制网页样式。用 CSS 设计网上商店的网页样式, 有利于网站内容维护以及订制个性化购物页面。在搜索系统中, 利用 CSS 样式表突出显示特定类别、特定条件或者具有特定属性的商品, 直接引起消费者的关注、点击和购买行为, 从而为用户提供最便捷的搜索方式。

2.2 Ajax 技术^[4]

Ajax 技术采用 XML 封装和传输数据, 可使程序异步执行。使用 Ajax 技术, 对于搜索过程有以下帮助: 第一, 提供关键词建议。消费者在输入查询条件过程中, Ajax 程序可自动根据用户输入的一部分关键词来检索相关建议词, 以下拉列表形式把建议词提供给消费者; 第二, 可以取消网页中不必要的刷新和跳转; 第三, 利用异步传输和按需加载的方式, 满足消费者快速浏览的需求。

2.3 Agent 技术

Agent 具有自主性、交互性、反应性和主动性的特点^[5]。PISS 中引入 Agent 技术, 有利于协助、代替管理员, 根据预定规则和所处的状态自动地维护商品数据库和会员数据库, 从而减轻了管理员的负担, 避免人工操作带来的失误。

3 小结

本文设计了应用于网上商店的搜索系统。该系统具有的特点是: 能够依据用户的查询条件推荐兴趣商品, 能够把用户行为反馈到搜索系统; 创新点是提出了基于关联的行为思想并设计了用户兴趣模型, 模型中采用了行为值累加后去对数的方案, 可以减弱恶意行为对模型的影响; 不足在于兴趣模型中的行为权值、行为关联度调整等需要根据不同的网上商店应用进行调整。

网上商店的搜索系统对于改变当前完全被动的网上商品销售状况有着重要意义。在搜索系统中设计的成功的推理方案, 有助于引发连续销售效应; 搜索系统中采用的新技术, 在提高搜索精度和效率的同时, 可以提供个性化的用户界面, 如网页色调、背景、音乐等, 有利于改善消费心理。

今后工作重点是将 PISS 系统思想应用到具体网上商店站点设计与开发中, 在实践中探索不同应用中的综合兴趣比例因子的合理取值以及模型中的参数取值。

参考文献

- 1 Hallerman, D. Search marketing: players and problems. eMarketer. [2007-04-20] http://www.emarketer.com/Reports/All/Search_marketing_players_apr06.aspx
- 2 Kotler P, Keller KL. Marketing management (12 ed.). Upper Saddle River, NJ: Pearson Prentice Hall, 2006.
- 3 李峰, 裴军, 游之洋. 基于隐式反馈的自适应用户兴趣模型. 计算机工程与应用, 2008, 44(9): 76-79.
- 4 Johnson D, White A. 张祖良, 等译. AJAX 企业级开发: strategies for building high performance applications. 北京: 人民邮电出版社, 2008.
- 5 何炎祥, 陈莘萌. Agent 和多 Agent 系统设计与应用. 武汉: 武汉大学出版社, 2001.