

一种基于对话的电子商务推荐系统

薛伟莲, 王蕴慧

(辽宁师范大学 管理学院, 辽宁 大连 116029)

摘 要:提出了一种基于 AIML 的自然语言对话型的智能电子商务推荐系统. 首先通过构建用户模型, 对某类特定用户生成相应的产品推荐列表, 从而完成系统对用户的首次产品推荐; 其次通过引入 AIML 人工智能标记语言, 构建专家知识库, 通过系统与用户之间多次的自然语言交互, 找到用户的兴趣点, 修正首次的产品推荐列表. 这种系统通过交互能力的改善和产品推荐准确性的提升, 可以提高用户的满意度.

关键词:电子商务; AIML; 协同过滤; 推荐系统

中图分类号:TP319 **文献标识码:**A

在 B2C 电子商务领域, 推荐系统可以根据用户的兴趣和爱好推荐顾客可能感兴趣或者满意的商品. 通过这种推荐方式, 可以把用户的潜在需求化为现实的购买力, 从而提高产品销售量. 国内外比较著名的电子商务网站, 如 amazon.com, ebay.com, taobao.com 和 douban.com 等都采用了不同技术的推荐系统. 如何从海量的信息中快速获得有用的信息, 从而提高在线商务活动效率, 成为电子商务领域的一个研究热点.

1 电子商务推荐系统的算法

电子商务推荐系统的算法主要有 4 类: 基于内容的推荐、协同过滤推荐、基于知识的推荐和组合推荐^[1-3].

1.1 基于内容的推荐

基于内容的推荐是根据用户选择和评价的对象, 推荐与其以前喜欢的对象相似的产品. 基于内容推荐的优点是简单有效, 没有新项目得不到推荐的问题, 也没有数据稀疏问题, 但是它无法对新的用户进行推荐, 而且训练分类器需要大量的数据.

1.2 协同过滤推荐

协同过滤推荐算法主要有两类, 基于用户的协同过滤推荐算法和基于项目的协同过滤推荐算法. 基于用户的协同过滤推荐算法的指导思想是如果用户对一些项目的评分比较相似, 则他们对其他项目的评分也比较相似; 而基于项目的协同过滤推荐算法则认为用户对不同项目的评分存在相似性, 当需要估计用户对某个项目的评分时, 可以对用户对该项目的相似项目进行估计. 对于协同过滤算法, 它的优点是无需专业的知识, 可以做到自动推荐, 容易发现用户的兴趣点, 但是它的缺点是有数据稀疏问题, 由于数据稀疏分布, 算法计算复杂度高, 此外它还有无法为新用户推荐产品和新产品得不到推荐的问题.

1.3 基于知识的推荐

基于知识的推荐不是建立在用户需要和偏好基础上的推荐, 而是利用针对特定领域制定规则来进行推理, 从而达到推荐的目的. 基于知识的推荐算法可以将用户的需求映射到产品, 而且非产品属性也纳入了考量范围, 但是对用户和产品的知识的挖掘很困难, 而且适应性差, 算法无法很好地移植到别的产品上.

1.4 组合推荐

组合推荐是将以上 3 种方法加以组合以避免或减少各自单独使用的不足, 目前应用最广泛的是基

收稿日期: 2011-03-20

作者简介: 薛伟莲(1966-), 女, 辽宁营口人, 辽宁师范大学副教授, 大连理工大学博士研究生.

于内容推荐和基于协同过滤推荐的组合. 由于组合的方式不同, 它们的性能和优缺点差异也很大.

基于以上推荐算法构建的推荐系统存在一个共同的问题, 即如果本次推荐的结果准确度不高, 不是用户想要的产品, 那么这种情况下推荐系统没有提供进一步纠正推荐结果、和用户继续交互以挖掘用户需求, 提供最适合用户的产品的功能.

2 基于对话的电子商务推荐系统

提出了一种基于对话的电子商务推荐系统, 该推荐系统有两个主要特点: (1) 能够通过自然语言的方式和用户交互, 根据用户的反馈, 发现用户的兴趣点, 推荐更符合用户需求的产品. (2) 即时回答用户对产品的疑问, 使用户对产品的了解更加全面, 尤其方便用户对新产品的了解, 也可以使新用户快速熟悉产品, 从而解决新用户、新产品的推荐问题.

2.1 系统架构图

图 1 是我们建立的基于对话的电子商务推荐系统的架构图. 图中包括两个用户角色和两个实体对象, 两个用户分别是客户和专家. 客户就是电子商务网站为之服务的顾客, 专家则为系统的知识库提供领域知识. 两个实体对象分别是电子商务网站数据库和网站日志文件. 系统由四个模块组成, 分别是用户界面组件、控制器、用户模型模块和知识库模块. AIML 技术用于维护用户与该系统的自然语言对话, 基于协同过滤推荐算法的用户模型和控制器用于给用户产品推荐.

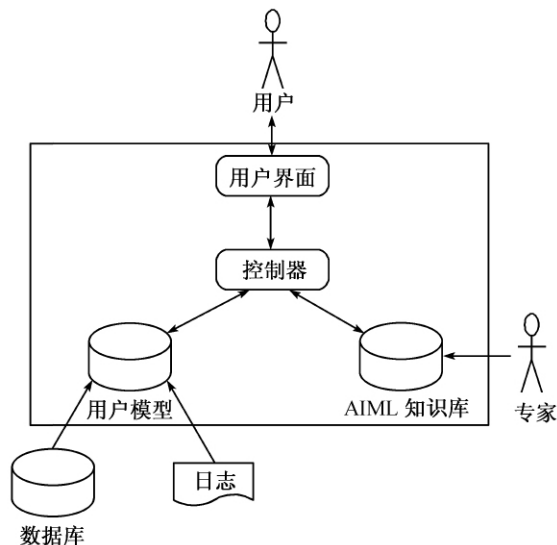


图 1 智能对话型推荐系统架构

2.2 主要模块介绍

2.2.1 系统界面与语音功能 近些年来, 大量的对话系统都是基于文本形式对话的, 然而基于语音的自然语言对话系统在与用户交互时更有效, 更能吸引用户的注意. 本系统实现了一种单向的具有语音发音功能的智能对话型推荐系统, 它可以以文字和语音的形式与用户进行交互.

图 2 是本系统在用户界面展示的虚拟网络人物. 这里我们采用 oddcast 公司的 Sitepal 和 Text-to-speech 这两个产品, Sitepal 是一件易于使用的角色产品, 它的在线界面设计新颖活泼, 并提供了不同风格特点的 2D 和 3D 的人物, 可以满足任何类型的电子商务网站需求, 并且制作也并不复杂. 而 Text-to-speech 将文字转化为语音, 它支持多国语言, 还可以实现准确的嘴唇同步. 两者结合起来, 我们便可以得到一个在网页中可以讲话的虚拟人物, 虚拟人物的语音和视觉动画效果可以大大吸引用户的注意力, 加强用户与系统的交互能力.

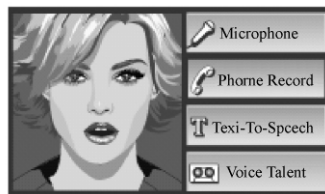


图 2 系统用户界面

2.2.2 基于协同过滤推荐算法的用户建模 电子商务网站搜集的客户数据包括注册数据, 即客户个人的基本社会和自然属性, 服务器端的日志文件和客户在搜索时产生的查询数据以及代理服务器端记录的客户对网站访问时的数据. 这些获取的数据首先要进行消除噪声, 判断缺值数据, 转换数据类型等的预处理, 以过滤掉那些冗余的数据.

用户建模是根据上述得到的客户数据, 将各个用户按照一定的相似度指标进行聚类, 生成 K 个具有相似爱好的用户组. 对于当前与系统交互的用户, 系统首先会根据用户的历史行为与 K 个用户组进行匹配, 找到与该用户最相似的用户组, 然后根据这个用户组的共同爱好给予该用户一定的产品推荐^[4-5].

以上过程可分为 3 个步骤:

(1)表示. 建立一个用户/产品评价矩阵描述用户对产品的评价. 把用户的判断和偏好表示为一个 $m \times n$ 的用户/产品评价矩阵 R , 这里 m 是用户数, n 是产品数, $R=(r_{ij})$, 元素 r_{ij} 表示用户 i 对产品 j 的评价, 在这里, r_{ij} 即可表示用户是否购买商品, 也可表示用户对商品的偏好程度.

(2)聚类. 通过计算所有用户对之间的余弦相似性, 采用 k-means 算法进行聚类. 把用户评分看做 n 维空间上的向量, 用户间的相似性通过向量间的余弦夹角度量, 设用户 i 和用户 j 在 n 维空间上的评分分别表示为向量 \vec{i}, \vec{j} , 则用户 i 和用户 j 之间的相似性为:

$$\text{sim}(i, j) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\| \cdot \|\vec{j}\|}$$

根据以上的用户之间的相似性测量方法, 将 N 个用户划分为 K 个聚类, 即 K 个具有相似爱好的用户组. 使同一聚类中的用户相似度较高, 而不同聚类中的用户相似度较低.

(3)分类. 将待识别的用户与 K 个用户聚类代表数据进行相似度比较, 找到相似度最高的那个聚类 K_i , 将该用户划分到 K_i 中.

将用户模型与知识库相结合, 以自然语言对话的方式经过多次交互之后, 系统可以精确定位用户的兴趣点, 从而挖掘出最适合用户需求的产品, 完成推荐功能.

2.2.3 控制器 控制器模块包括 AIML 解释器和用户模型查询接口. 将两者结合起来, 完成系统与用户的交互和产品信息推荐功能.

原生的 AIML 解释器都是面向英文设计的, 它需要有空格来断字, 并且在运行中会自动大写所有字母, 并在回复中大写首字母. PyAIML 为 Python 提供了对 AIML 的支持, 同时也是开源的模块, 所以可以建立一个完整支持 AIML 规范的平滑支持中文的 AIML 解释器, 从而让其更好的支持中文, 从而利用 AIML 规范中的深层功能.

中文化的原理包括自动空格断字、还原、自动编码转换, 对汉字忽略所有大写及小写转换以及支持单字节编码的汉字. 其中自动空格断字、还原, 这需要修改 PyAIML 的 AIML 读取知识库的接口 AimplParser.py 和人机对话输出接口 Kernel.py 这两个接口来实现. 因为 PyAIML 本身是使用 Unicode 的, 所以在做输入接口时需要将字符转换为 Unicode, 输出时再转换成原始编码. 支持单字节编码的汉字是一个增加兼容性的选择, 可以通过修改 PyAIML 的 PatternMgr.py 文件, 重载 string 模块的 upper 函数来实现.

当用户提问时, 首先通过 AIML 解释器对句子进行解析, 再调用用户模型查询接口来获取用户的购买记录等, 根据用户模型信息, 找到对该用户的产品推荐信息, 通过语音功能推送到系统界面. 整个控制器的工作流程如图 3 所示.

2.2.4 AIML 知识库 由 AIML 描述的知识库可

以表示专家总结的领域知识, 由多个以 AIML 为后缀名的文件组成, 相似领域的话题被放到一个文件中以便知识库的分类整理. 这里的领域知识对应的是电子商务网站中对产品服务的专业解释和功能介绍. 当用户提问有关某一产品的特点时, 系统可以根据知识库中的领域知识给予响应. AIML 具有一套功能完整而强大的标签, 从而在模拟人的智能上有着很好的表现, 它的基本单元是 category, 每个 category 包含一个问题 (stimulus, pattern), 一个回答 (response, template), 和可选上下文 (分别为 that 和 topic), 可以立即使用的 category 模板在 AIML 的库中目前超过 40 000 个. 递归调用标签 <srail> 可以做

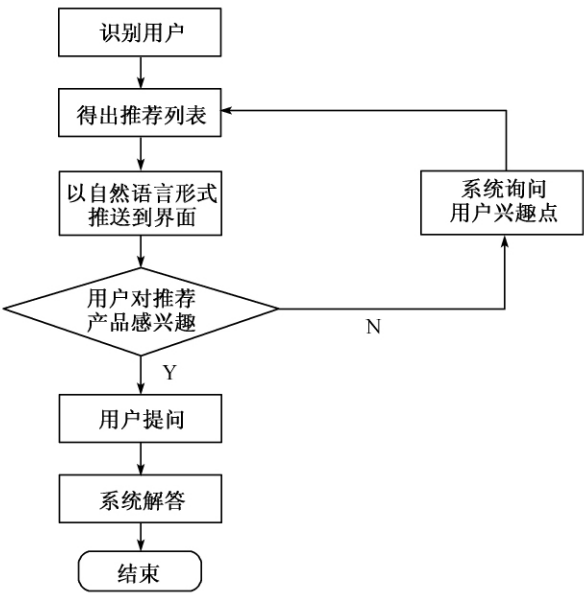


图 3 控制器工作流程图

到同义句转换,使得同一意思的不同问句得到相同的答案,并且还可以进行关键词匹配. <input> 标签使得 AIML 可以有学习的能力,此标签可以使其记住之前的聊天记录. 除了以上几种基本的标签,AIML 还有许多智能的标签^[6].

AIML 是一个基于经验的聊天机器人,这样就需要大量的先验知识才能取得好的对话效果. 然而问话的方式非常多,于是需要加入规则处理,把句式抽取出来形成一定的模板,从而减轻编写的难度和强度. AIML 文件可以进行内联,许多包含特殊领域知识的 AIML 文件可以方便的合并成一个更大的知识库,并且可以通过对聊天记录进行分析,可以得到尚且没有明确回答的问题,并给出建议的模式^[7].

产生系统所需要的知识需要进行知识获取、知识分析和知识提取. 这里知识的来源是产品本身的属性资料,之前用户的购买、咨询以及评级记录,所以知识库中的内容也需要定期进行动态更新. 知识的获取必须要保证是正确的以及合适的,以便进行下阶段的知识设计. 比较常用的知识获取方法为访问专家,除此之外,许多软件工具也可用来实现自动知识获取,如 KAS,KAUS 等.

3 结束语

引入了一种基于 AIML 的领域知识表达方式,提出了结合 AIML 知识库和用户模型的自然语言对话智能推荐系统. 该系统能够弥补之前推荐系统的不足,通过用户和系统之间即时的多次自然语言交互,挖掘用户的需求与兴趣点,从而在此基础上针对用户的特定需求进行精确的产品推荐,解决了新用户和新产品的推荐问题.

系统也存在一些不足,未来的工作将会沿着以下几个方向展开:(1)丰富 AIML 知识库,提高系统自然语言对话的智能性.(2)引入语音识别技术,真正实现双向语音对话,目前系统仅仅支持单向语音功能,不是真正的语音对话,而语音对话的实现,可以大大提高用户与该系统的交互能力,也是该系统智能性的一个重要体现.(3)引入其他推荐算法,完善用户模型,同时也可以建立产品模型,将各种推荐算法组合起来提高系统推荐的准确度.(4)目前 AIML 对中文的支持不如英文成熟,还需要在现有的中文处理算法上进一步完善.

参考文献:

- [1] 黎星星,黄小琴,朱庆生. 电子商务推荐系统研究[J]. 计算机工程与科学,2004(5):7-10.
- [2] 许海玲,吴潇,李晓东,等. 互联网推荐系统比较研究[J]. 软件学报,2009(2):350-362.
- [3] 余力,刘鲁. 电子商务个性化推荐研究[J]. 计算机集成制造系统,2004(10):1306-1312.
- [4] 游文,叶永生. 电子商务推荐系统中的协同过滤推荐[J]. 计算机技术与发展,2006(9):70-72.
- [5] 马宏伟,张光卫,李鹏. 协同过滤推荐算法综述[J]. 小型微型计算机系统,2009(7):1282-1288.
- [6] WALLACE R S. AIML Overview [EB/OL]. <http://www.pandorabots.com/pandora/pics/wallaceaimltutorial.html>,2001.
- [7] 王晓敏. 基于学生模型与 AIML 的智能教学系统的研究[D]. 大连:大连理工大学,2009.

An E-commerce recommender system based on dialogs

XUE Wei-lian, WANG Yun-hui

(The Management College of Liaoning Normal University,Dalian 116029,China)

Abstract: It presents an intelligent E-commerce recommender system with a natural language interface based on AIML. Firstly, the recommendation takes place after matching user with the contributed user models. Then, it introduces AIML to build knowledge base. The system will fix the recommended goods list by chat with user in natural language. User experience will be significantly enhanced through the improvement of interactive ability and recommendation accuracy.

Key words: E-commerce; AIML; collaborative filtering; recommender systems