

基于模糊兴趣集和智能 Agent 的推荐系统模型研究

王卫平, 赵 明

(中国科学技术大学 管理学院, 安徽 合肥 230026)

摘 要: 提出了一种结合用户模糊兴趣和智能多 Agent 的电子商务推荐系统模型, 利用智能 Agent 获取用户对产品属性的喜好度和相应权重, 依此建立起清晰的用户兴趣模型。在智能多 Agent 技术下建立一种多属性综合评价机制, 采用逼近理想解法(Topsis 法)对候选产品进行实时处理, 并根据排序结果产生 Top-N 推荐。

关键词: 推荐系统; 智能 Agent; 模糊兴趣集; 多属性决策; 逼近理想解法

中图分类号: TP311

文献标识码: A

Model of recommender system based on fuzzy interest set and intelligence agent

WANG Wei Ping, ZHAO Ming

(School of Management, University of Science & Technology of China, Hefei 230026, China)

Abstract: The paper proposes a model of electronic business recommendation system based on users' fuzzy interest set and intelligence multi-agent. By using intelligence agent, user preferences and its weights for product attributes are obtained, and a clear model of user interest established. The model also applies multi-agent to build a comprehensive multi-attributes evaluation mechanism. The topsis method has been used to perform real-time ranking. After that, the Top-N are recommended in accordance with the results.

Key words: recommender system; intelligent agent; fuzzy interest set; multi-attributes decision; topsis method

随着互联网普及和电子商务的发展,越来越多的人愿意网上购物。电子商务系统在为用户提供越来越多选择的同时,其结构也变得越来越复杂,用户常会迷失在大量的商品信息空间中,无法找到自己满意的产品。推荐系统的出现为用户提供了解决这类问题的工具。目前有很多不同种类的推荐系统,其中应用最成功的是协同过滤推荐技术 CF(Collaborative Filtering),它利用相邻邻居用户之间的相似度做出推荐,其他主要技术还有基于内容的推荐 CBF(Content-Based Filtering)和混合推荐(Hybrid method)等^[1]。现有推荐系统在使用时,最重要的特点是能够实时查找、推荐产品,但随着用户和产品的增加,推荐系统算法的计算量会呈几何数增长,在大数据量的情况下,算法的处理时间可能延长到无法容忍的程度,导致处理能力低下、推荐准确度降低、系统性能严重下降^[2-4]。近年来,为解决这类问题,学者们进行

了大量研究,其中以 Agent 技术为基础的推荐系统,以其在解决扩展性问题的优势,得到了不少研究者的重视^[5-7]。参考文献[5-6]以智能多 Agent 系统为基础,建立一个基于偏好的评价模型,并在个性化的购物过程中,提出了一种基于 Agent 的决策分析模型。参考文献[7]提出了以智能 Agent 为基础,融合了案例推理和多 Agent 技术的电子商务推荐模型,建立了 1 种多属性的综合评价机制。但此类推荐系统存在的问题是:没有或很少说明此类推荐系统如何获取用户兴趣或偏好,使得以智能 Agent 为基础的推荐系统存在很大局限性,影响了此类推荐系统的现实应用。

本文提出了一种基于用户模糊兴趣度建立的多属性推荐决策评价模型,通过隐性方式收集用户的兴趣,并用智能多 Agent 建立起 1 个电子商务推荐系统,为用户提供个性化服务。

1 智能 Agent 技术和 Topsis 方法

1.1 智能 Agent 技术

Agent 技术是人工智能技术最新技术之一。Agent 是一种具有自制能力、智能和目标驱动属性,能够通过各种学习、推理等方法感知和适应复杂的动态环境,能够主动追求目标的能动实体,具有某种程度的“拟人性”,其本意就是要代替人工完成某些工作。现在普遍的观点认为 Agent 是 1 个实体,能够针对特定的目标求解问题,具有自主性、交互性、适应性、社会性、反应性、可通信性、学习性和自成长性等特性,因而利用 Agent 来构建个性化推荐系统具有先天性的优势。

多 Agent 系统是由多个 Agent 组成的系统,其基本思想是将具有不同目标的多个 Agent 对其目标、资源等进行合理的安排,以协调各自行为,达到最大程度地实现各自的目标。每 1 个 Agent 又是 1 个具有目标、知识和能力的自治计算实体,多个 Agent 协调合作形成问题的求解网络。多 Agent 系统优于仅由单个 Agent 控制的系统,通过 Agent 间合作,多 Agent 系统不仅改善了每个 Agent 的基本能力,而且可以解决单个 Agent 无法解决的复杂问题。

采用多 Agent 技术可以克服现有电子商务推荐系统存在的局限性。Agent 技术的反应性、主动性和自治性,使系统可快速反应不断变化的环境并及时做出行动。Agent 技术的社会性保证了系统内部及系统与用户或外部系统之间的交互和协同能力。

1.2 Topsis 法介绍

Topsis 法是管理决策分析中对多目标方案进行排序处理的常用方法,是一种有效的多属性评价决策方法^[8]。它通过构造多属性问题的理想解和负理想解,并以其为基准,作为评价各可行方案的依据。例如用户 C 要购买具有 A_1, A_2 2 个属性的产品,设其均为效益型指标,现有此类产品 p_1 和 p_2 ,为了从中挑选用户满意的产品,将产品 $p_i(i=1, 2)$ 用集合记为 (A_{i1}, A_{i2}) ,且令 $A^*_{i1} = \max_{1 \leq i \leq 2} \{A_{i1}\}$, $A^*_{i2} = \max_{1 \leq i \leq 2} \{A_{i2}\}$; $A^-_{i1} = \max_{1 \leq i \leq 2} \{A_{i1}\}$ 及 $A^-_{i2} = \max_{1 \leq i \leq 2} \{A_{i2}\}$,则此问题最理想产品属性应为 A^*_{i1} 和 A^*_{i2} ; 最不理想产品属性应为 A^-_{i1} 和 A^-_{i2} 。设产品 p_i 对应的到最理想产品和最不理想产品的距离分别为:

$$S_i^* = \sqrt{\sum_{j=1}^2 (A_{ij} - A^*_{ij})^2}, S_i^- = \sqrt{\sum_{j=1}^2 (A_{ij} - A^-_{ij})^2} \quad (1)$$

产品 p_i 与到最理想产品和最不理想产品的相对贴近度表示为:

$$C_i^* = S_i^- / (S_i^- + S_i^*), 0 \leq C_i^* \leq 1 \quad (i=1, 2) \quad (2)$$

于是便得到产品的偏好排序,某候选推荐产品贴近最理想产品而远离最不理想产品时,应将其优先推荐给用户 C。

2 基于模糊兴趣集的推荐系统模型

2.1 用户模糊兴趣引入

由于知道推荐系统的目标对象是用户,因而其本质都依赖于用户的兴趣。但传统的个性化电子商务推荐系统,其推荐过程是基于集合统计的原则,很少从用户兴趣的本质考虑推荐模型。推荐过程大多数都是推荐某种商品,而对为什么推荐这个商品,用户喜欢什么知之甚少,这样的推荐难以做到真正的个性化。

传统的推荐系统面临的棘手问题是对商品描述过于简单,商品属性描述表达不力。以音乐推荐(流行,古典,爵士,民谣,摇滚,欧美,韩日)论域为例,1 部音乐可能对应其中多项,每一项的侧重都有所不同。通常的做法是用 0 或 1 进行表示,但人们的喜好很难用喜欢或者不喜欢进行决断,其本身是个模糊的概念,因此建立 1 个用户模糊兴趣模型很重要。

2.2 用户模糊兴趣集的建立

2.2.1 产品属性隶属函数的建立

用户模糊兴趣集产生的第一步即是产品属性隶属函数,采用模糊数即 $[0, 1]$ 区间值对产品属性进行描述,从而使产品属性可以同时具有几个不同类型的值。这样提高了属性描述的精确度^[9]。新模型采用德尔菲法获得产品属性隶属度。设产品 U 属性集为 $A = \{A_1, A_2, \dots, A_i, \dots, A_n\}$,让每位专家分别对每个属性 A_i 给出一个隶属度估计值,记第 k 位给出的估计值为 $S_k(A_i)(i=1, 2 \dots n; k=1, 2 \dots m)$,于是隶属度可表示为:

$$S(A_i) = \frac{1}{m} \sum_{k=1}^m S_k(A_i) \quad (3)$$

但德尔菲法存在缺陷,即无法表示一些随时间改变而变化的量(比如新旧程度等)。于是对此类属性采用函数表示:

$$S_i(A_i) = F(A_i) \quad (4)$$

于是,得到产品 U 属性的隶属度:

$$S = \begin{cases} S(A_i) & A_i \text{ 可以用德尔菲法表示} \\ S_i(A_i) & A_i \text{ 不可以用德尔菲法表示} \end{cases} \quad (5)$$

2.2.2 模糊兴趣集的建立和探讨

兴趣度就是用户对某一产品或某一类产品的兴趣的强弱。它反应了用户对某一产品或某一类产品的喜好程度。一般可采用 2 种方式来计算和表示用户兴趣度并建立模糊兴趣集:显式方式和隐式方式。显式方式让用户参与对产品和产品属性进行兴趣度评分,无疑这种收集的用户兴趣度方式是比较准确,但它要求用户参与,增加了用户的负担,所以很难被接受。本研究摒弃了这种建立用户兴趣集的方式,而采用另一种隐式方式收集客户兴趣度,建立模糊兴趣集。因为用户的兴趣会反应在其购买行为上,因此用户购买记录集会反应用户的兴趣。因此直接使用此集合构造 1 个用户兴趣集。

设 $U = \{u_1, u_2, \dots, u_m\}$ 为客户购买记录集,

其中 $u_i=(S_{i1}, S_{i2}, \dots, S_{ik}, \dots, S_{im}), S_{ik}$ 即(3)式的结果,表示产品 u_i 对属性 A_i 的隶属度。于是用户兴趣集可表示为: $I_k=\{I_1, I_2, \dots, I_k, \dots, I_m\}$, 其中 I_k 表示该客户对第 k 项属性的兴趣度:

$$I_k = \frac{\sum_{i=1}^m (S_{ik})^2}{m} \quad (k=1, 2, \dots, m) \quad (6)$$

但通过分析产品集和用户的选择行为可知, 依赖用户兴趣度去描述用户兴趣是不清晰的, 用户的兴趣应该包括对产品的喜好程度、产品属性的权重 2 个方面进行准确描述, 但(6)式模型显然无法直接刻画其 2 个方面, 因此模型并不完善, 需要对其进行扩展^[10]。通过对用户喜好程度和产品属性权重属性进行探讨, 寻求建立完善的模糊兴趣模型。

(1) 产品属性喜好度的获取

高阶属性喜好度实际可以看成 1 棵树, 产品喜好度的获取将从最高层向最底层逐级获取。用 I_L 表示商品的属性喜好度, 但实际计算采取从底层到高层方式。

(2) 产品属性权重的获取

由于用户偏好的差异性, 导致对属性喜好程度不同, 也使对不同的属性有不同的重视度, 而这种重视度就是产品属性的权重。因此, 为了建立准确的模糊用户模糊兴趣集, 有必要针对每个属性给出不同的权重。

为获取有效的权重, 初期赋值再次采用德尔菲法取得产品属性权重, 具体如下: 设产品 U 属性集为 $A=\{A_1, A_2, \dots, A_i, \dots, A_n\}$, 请 Q 位专家, 让每一个专家分别对 A_i 给出权重估计值, 记第 p 位专家给出估计值为 $W_p(A_i)$ ($i=1, 2, \dots, n; p=1, 2, \dots, Q$), 得最后的产品属性权重为:

$$I_Q(A_i) = \frac{1}{Q} \sum_{p=1}^Q W_p(A_i) \quad (7)$$

德尔菲法产生权重有一定的局限性。它不能针对每个人的偏好有所不同, 难以实现真正的个性化, 但由于推荐系统前期工作时缺乏用户的有效数据, 因此可以利用此种方式获得权重, 推荐系统正式运行之后, 便可以获得大量的客户购物信息, 此时可以通过智能 Agent 取得属性权重。

综合以上内容, 利用 $\{I_L, I_Q(A_i)\}$ 表示用户模糊兴趣集。 $I_L, I_Q(A_i)$ 分别表示产品属性喜好度和产品属性权重。在这个模型中, 对于每个属性和量的描述都是模糊的, 是 1 个 $[0, 1]$ 之间的数, 用户的兴趣会包括多个产品属性, 也可能分布在多个层次中。此外本研究认定用户兴趣集具有时效性, 即认定在一定的时间内用户的兴趣是稳定的, 兴趣集会随着用户信息的完善而渐趋稳定。

2.3 推荐系统模型

本文设计的推荐系统为电子商务网站注册用户提供个性化的推荐服务。系统的工作流程如图 1 所示。(1)

用户提交购买请求并指定需求时, 推荐系统从产品库中搜索产生候选产品集。(2)找到目标用户的模糊兴趣集, 从中提取出用户兴趣度, 并与产品属性匹配。(3)利用 topsis 方法对候选产品集进行评价, 获得产品的偏好排序, 并从中选择评价最高的若干产品进行 Top- N 推荐。(4)记录用户购买信息, 将其保存到用户模糊兴趣集中, 并对其进行自适应性学习。

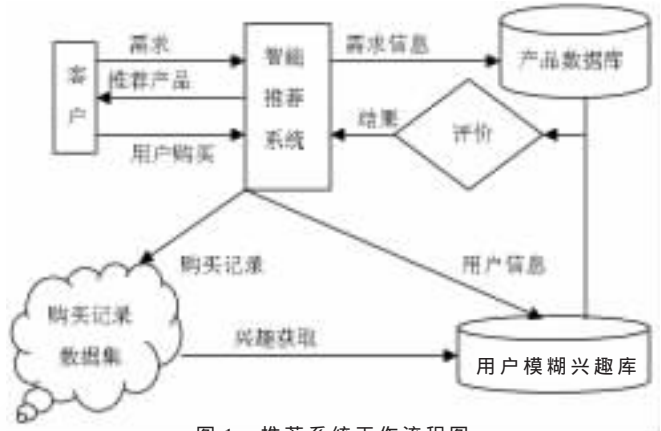


图 1 推荐系统工作流程图

3 结合模糊兴趣度和智能 Agent 推荐系统模型实现模型

在研究提出的推荐模型中, 用多个 Agent 通过协同合作完成系统推荐任务, 如图 2 所示。整个推荐系统内部模块结构可划分为产品管理 Agent、兴趣获取 Agent、评价推荐 Agent 和用户交互界面 Agent 4 个部分, 由这 4 部分共同分工协作, 完成系统推荐任务。

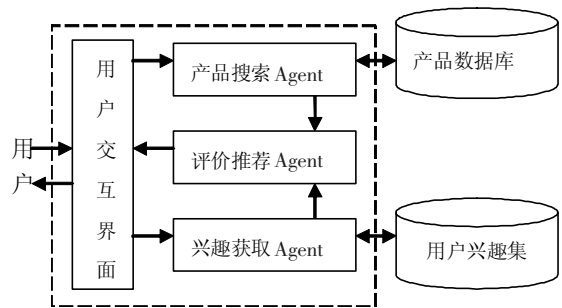


图 2 推荐系统系统结构

上述模块划分使各个 Agent 功能相对独立, 各 Agent 之间可并发运行, 从而提高了推荐系统的运行速度提高了推荐的性能。解决并发问题是多 Agent 系统的优势, 4 个 Agent 分工如下:

(1) 用户交互界面 Agent 负责客户与其他 Agent 模块交互。由于一直处于活动状态, 便可以实时处理用户的购物信息。当用户登录系统后, 用户交互界面 Agent 将会把用户信息和购买请求分别传给兴趣获取 Agent 和产品管理 Agent, 而评价推荐 Agent 则把候选产品的推荐结果发回到用户交互界面 Agent 显示, 客户作出选择之后, 界面 Agent 负责把用户购物记录保存到相应用户的

记录集中。

(2) 产品搜索 Agent 负责从用户交互 Agent 获得用户购买, 产品数据库中搜索候选产品。若满足要求, 产品搜索 Agent 就会把结果传递到评价推荐 Agent; 反之, 则会把信息反馈到用户交互界面 Agent, 并给用户推荐网站热销产品。

(3) 兴趣获取 Agent, 其主要负责对用户模糊兴趣集的管理和维护。首先, 兴趣获取 Agent 从用户交互界面 Agent 获得用户信息和购买信息, 然后从用户模糊兴趣集调取相应数据传递到评价推荐 Agent, 并将实时处理; 当用户购买完毕时, 通过用户交互 Agent 将购买记录传递到兴趣获取 Agent, 用户将对购物信息进行规范化, 利用(6)、(7)式提取用户的模糊兴趣, 并将结果保存到客户模糊兴趣集中, 为了提高系统运行效率, 一般选择离线处理。

(4) 评价推荐 Agent 是模型最重要的部分, 它负责对产品候选集进行多属性决策评价和产品推荐。首先评价推荐 Agent 从产品搜索 Agent、兴趣获取 Agent 获取候选产品集和相应产品属性模糊兴趣集, 然后利用 topsis 方法进行评价。最后从高到低对候选产品进行排序, 选择 Top-N 进行推荐, 并将结果传递给用户交互界面 Agent, 以规范形式显示便于用户查看。其数学模型如下:

记有 n 个产品和 m 个属性, 用 Y_{ij} 表示第 i 个产品第 j 个属性的喜好度, 得某顾客规范评价矩阵 A 为:

$$A = \begin{pmatrix} Y_{11} & \cdots & Y_{1n} \\ \vdots & Y_{ij} & \vdots \\ Y_{m1} & \cdots & Y_{mn} \end{pmatrix} \quad (8)$$

从上面矩阵构成规范决策矩阵, 其中元素 Z_{ij} 为:

$$Z_{ij} = Y_{mj} / \sqrt{\sum_{i=1}^n Y_{ij}^2}, \text{ 其中 } i=1, \dots, m; j=1, \dots, n$$

步骤 1: 由兴趣集获取 Agent 产品的属性权重, 并将其赋予相应 $w_j, w_j (j=1, \dots, n)$ 是第 j 个属性的权值, 则其构成加权的规范决策矩阵, 其中的元素 $x_{ij} = w_j Z_{ij}, i=1, \dots, m; j=1, \dots, n$ 。

步骤 2: 确定最理想产品和最不理想产品属性值大小:

$$x^* = \{(\max_i x_{ij} | j \in J), (\min_i x_{ij} | j \in J') | i=1, \dots, m\} \\ = \{x_1^*, x_2^*, \dots, x_m^*\} \quad (9)$$

$$x^- = \{(\min_i x_{ij} | j \in J), (\max_i x_{ij} | j \in J') | i=1, \dots, m\} \\ = \{x_1^-, x_2^-, \dots, x_m^-\} \quad (10)$$

上式是效益型目标的集, J' 是成本型目标的集, 比如产品中价格一些指标即属于成本型指标集, 属于逆向指标。而功能和效果即属于效益型目标的集, 属于正向指标。

步骤 3: 计算距离, 每个候选产品到最理想产品和

最不理想产品的距离为:

$$S_i^* = \sqrt{\sum_{j=1}^m (x_{ij} - x_j^*)^2}, i=1, \dots, m \quad (11)$$

$$S_i^- = \sqrt{\sum_{j=1}^m (x_{ij} - x_j^-)^2}, i=1, \dots, m \quad (12)$$

步骤 4: 计算每个候选产品与最理想产品的相对接近度。将(11)、(12)式结果代入(13)式中:

$$C_i^* = S_i^- / (S_i^- + S_i^*), 0 \leq C_i^* \leq 1, i=1, \dots, m \quad (13)$$

步骤 5: 排列产品的优先顺序, 进行产品推荐。由(13)式中获取 C_i^* 的值, 然后大小顺序排列对 C_i^* 进行排列, 选取值排在前列的进行 Top-N 推荐, 并将结果传递到用户交互界面 Agent。

本文提出了以模糊兴趣集和智能 Agent 为基础的推荐模型, 通过在用户兴趣中引入模糊兴趣集, 解决了长期以来以 Agent 为基础的推荐系统无法通过隐式收集用户对产品属性评价的问题。在评价推荐过程中, 采用 topsis 方法对候选产品进行多属性评价, 建立起 1 个高效、扩展性强的推荐系统, 并可将其应用于电影、音乐、书籍及其他频繁购买产品推荐上。

参考文献

- [1] ADOMAVICIUS G, TUZNILIN A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions [J]. IEEE transactions on knowledge and data engineering, 2005, 17(6): 734-742.
- [2] BREESE J, HECKERMAN D, KADIE C. Empirical analysis of predictive algorithms for collaborative filtering [A]. In Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, 1998: 43-52.
- [3] BURKE R D. Hybrid recommender systems: Survey and experiments [J]. User modeling and User-Adopted Interaction. 12(4): 331-370.
- [4] HERLOCKER J L, KONSTAN J A, TERVEEN LG, et al. Evaluating collaborative filtering recommender systems [J]. ACM Transaction on Information Systems, 2004, 22(1): 5-53.
- [5] 闵君, 邓晓. 智能导购 Agent 系统的研究 [J]. 控制与决策, 2003, 18(4): 497-499.
- [6] 谭雪清, 曾子明, 周洞汝. 个性化电子商务购物模型的研究 [J]. 微计算机信息, 2007, 23(12): 147-148.
- [7] 曾子明. 基于 Agent 和 CBR 的电子商务推荐模型的研究 [J]. 现代情报, 2008(3): 209-213.
- [8] 陈珽. 决策分析 [M]. 北京: 科学出版社, 1987.
- [9] 杨纶标, 高英仪. 模糊数学原理和应用 [M]. 广州: 华南理工大学出版社, 2005.
- [10] 唐灿, 朱征宇. 基于模糊兴趣模型的个性化推荐算法 [J]. 计算机工程与应用, 2006, 9: 166-168.

(收稿日期: 2009-05-09)