

基于智能多 agent 的推荐系统^①

王卫平 赵明 刘迎意 王选 (中国科学技术大学 管理学院 安徽 合肥 230026)

摘要: 针对传统推荐系统存在的用户评分稀疏性和系统扩展性问题,提出了一种基于智能多 agent 的推荐系统 MASRS。首先采用余弦公式处理用户-项评分矩阵得到用户初始邻居集;然后将用户评分映射到相应项的属性值上,生成用户-属性值偏好矩阵 UPm,并基于此矩阵进行用户相似性度量,得到用户产品推荐集,该方法有效缓解用户评分稀疏性问题;通过智能多 agent 架构推荐系统,使大量复杂计算在线下进行,从而改善系统存在的扩展性问题。同时实验表明新系统在推荐精度上收敛性更好。

关键词: 推荐系统;稀疏性;用户-属性值偏好矩阵;智能多 agent

A Recommendation System Based on Intelligence Multi-Agent

WANG Wei-Ping, ZHAO Ming, LIU Ying-Yi, WANG Xuan

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

Abstract: Traditional recommendation system has the problem of sparse user ratings and system scalability. This paper proposes a recommendation system based on intelligence multi-agent. At first, the cosine similarity measure has been used to handle user-item rating matrix, thus the initial neighbor set for target users can be gained. Then, user ratings have been mapped to relevant item attributes for generating user-attributes value preference matrix UPm of each user. Thus, user similarity can be computed based on UPm and rating sparsity has been alleviated simultaneously. The recommendation system of intelligence multi-agent makes calculating an online processing, and thus improves the system scalability. Experimental results show that the new system achieves a better accuracy in recommended convergence.

Keywords: recommendation system; sparsity; user-attributes value preference matrix; intelligence multi-agent

1 引言

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,其结构也变的越来越复杂,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(Collaborative Filtering, CF),它利用相邻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(Content-Based filtering, CBF)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

是能够实时查找与推荐产品,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,在大数据集的情况下,算法的处理时间将增加到无法容忍的程度,实时处理能力降低。推荐准确度下降,系统性能严重降低^[2-4],同时致使推荐系统也面临严峻的用户评分稀疏性问题^[5,6],这样导致推荐质量下降。对此研究人员陆续提出了解决方法,如文献^[7]中将用户年龄、国籍、性别等个人信息增加作为相似度计算的依据,称为基于人口统计学的过滤方法(demographic filtering)。文献^[8,9]使用奇异值分解(SVD)方法尝试把稀疏的关系矩阵的维度降到更低维,以得到用户的

① 收稿时间:2009-05-19

潜在关系。文献[10]在基于存储和用户建模方法的基础上,提出了一种平滑聚类的方法来克服稀疏性问题。他基于稀疏的用户矩阵,对用户进行推荐。文献[11]提出了一种基于项目评分预测的系统过滤推荐系统的模型,通过计算用户之间的相似性来预测未评分项目,这样就增加了用户共同评分的项目数量,从而有效缓解稀疏性问题。上述方法虽然在缓解稀疏性方面取得了进展,但忽视了推荐质量准确性和推荐系统扩展性的问题。

本文提出了一种基于用户偏好建立的多属性推荐决策评价模型,通过对用户-项目评分矩阵中的用户进行相似性计算比较,得到其初始邻居集,然后构建用户-属性值偏好矩阵,通过该矩阵计算用户与用户最近邻之间的相似性,最后比较分析得到用户 TOP-N 推荐,从而有效缓解用户评分的稀疏性。整个推荐以智能多 agent 系统(MAS, Multi-agent System)为基础建立起一个电子商务推荐系统 MASRS, MAS 技术特点使推荐系统大量计算在线下完成,可以改善推荐系统存在的扩展性问题。实验结果表明新系统模型的有效性。

2 MASRS体系架构

Agent 技术是人工智能技术最新技术之一。Agent 是指一种具有自制能力、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,并能够主动追求目标的能动实体它具有某种程度的“拟人性”。MAS 是由一组独立但又协同工作的多个 agent 构成,通过协同来解决问题。MAS 优于仅由单个 agent 控制的系统,通过 agent 间合作, MAS 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。本文采用 MAS 技术构建推荐系统,在进行用户信息采集、数据智能处理、个性化信息推荐和人性化交互界面的呈现方面,可以显著提高推荐系统智能化程度。系统架构如图 1 所示。

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 MAS 的优势。其分工如下:

① 用户交互界面 agent 负责客户与其他 agent 模块交互,其一直处于活动状态,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

将会通过通信 agent 把用户信息和购买请求分别传给用户 agent 和产品搜索 agent; 产品推荐 agent 则把候选产品的推荐结果发回到用户交互界面 agent 显示,客户做出选择之后,用户界面 agent 负责把用户购物记录保存到相关 agent 进行处理。

② 产品搜索 agent 负责从用户交互 agent 获得用户购买信息(关键字),产品数据库中搜索候选产品。若有满足要求候选产品,产品搜索 agent 就会把结果传递到产品推荐 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。

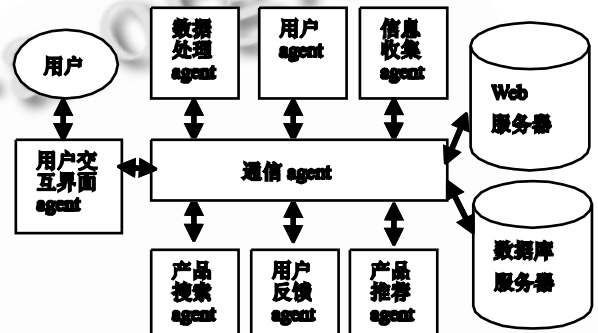


图 1 MASRS 体系架构

③ 产品推荐 agent 是模型最重要的部分,它负责对产品候选集多属性决策评价和产品推荐。首先产品推荐 agent 从产品搜索 agent 和用户 agent 获取候选产品集和相应产品属性用户偏好信息,然后推荐算法进行评价,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,以规范形式显示给用户查看。

④ 数据处理 agent 负责将每次用户购买信息进行处理,并对用户购物记录进行更新,然后通过购物记录生成用户偏好信息模型,根据模型计算用户的相似度,为产品推荐 agent 做数据准备。

⑤ 信息收集 agent 在后台分析 web 服务器和 database 服务器的用户日志,如用户访问的资源、访问次数和访问次序,发掘用户的兴趣点,为产品推荐 Agent 提供数据来源。

⑥ 用户 agent 是用户登录系统时自动生成的,可以从中读取用户个人资料和产品推荐 agent 产生的结果,生成用户记录,记录用户网页链接和访问次数,将用户提交的搜索请求转交给产品搜索 agent,并返回搜索结果。

⑦ 用户反馈 agent 主要收集用户对系统产品所做的评价信息，其可以分为显式反馈和隐式反馈两种方式。显式反馈是指用户直接给出其兴趣偏好的信息，相对隐式反馈更为准确。在本文 MARRS 中，采用显式反馈为主、隐式为辅的综合用户反馈机制。对收集得到的用户评价信息，存储到后台数据库中。

⑧ 通信 agent 用户交互界面 agent、数据库和其他各个 agent 的通信桥梁，协调各 agent 协作和作业调度。

3 MASRS模型的建立和实现

3.1 基于 MAS 推荐系统模型的建立

传统的电子商务网站上，每个产品的信息都保存在数据库的二维表格中，而产品属性则通过二维表格的列表表示。用户在购买某一产品时通常出于对该产品所具有的某些属性值的偏好，因此所购买的产品可以体现用户的兴趣喜好。例如，用户喜欢某首流行歌曲时，可能是由于喜欢歌曲类型，也有因为喜欢该歌唱者。因此，本文借助产品属性值来缓解用户-项评分稀疏性。其实现基本思想是：

① 通过用户-项评分矩阵获取用户最近邻居
在传统推荐系统中，输入数据一般表示为 $m \times n$ 的用户-项评分矩阵 R ， m 是用户数， n 是资源项目数，矩阵元素 $R_{i,j}$ 表示第 i 个用户对第 j 个项目的评分值，一般 $R_{i,j} \in [1,5]$ ，且 $R_{i,j} \in Z$ ，该值表示用户对该项的偏好评分，如表 1 所示。

表 1 用户-项评分矩阵

用户	Item ₁	...	Item _i	...	Item _n
U ₁	0	...	3	...	3
U ₂	5	...	4	...	2
⋮	⋮	...	⋮	...	⋮
U _{m-1}	4	...	2	...	4
U _m	5	...	5	...	4

该表中数据在 MASRS 可以从反馈 agent 中获得，根据用户评分数据，通过计算用户之间的相似性来得到目标用户的初始邻居集，即对任意用户集合 $Set(u) = \{u_1, u_2, \dots, u_m\}$ ， $u \notin S(u)$ 。设用户 u, u_k 的相似度为 $sim(u, u_k)$ ($1 \leq k \leq m$)，采用余弦相似性计算用户 u 与 u_k 的相似性：

$$sim(u, u_k) = \cos(\vec{u}, \vec{u}_k) = \frac{\vec{u} \cdot \vec{u}_k}{\|\vec{u}\| \|\vec{u}_k\|} \quad (1)$$

然后取 $sim(u, u_k)$ 最大的前若干个用户最为用户 u 的初始邻居集 $S(u)$ ，其大小可根据 m 的大小设定合适比例，目的是为了减少后续的计算量，一般在线下完成，有效提高系统实时数据处理能力。

② 用户-属性值偏好矩阵的建立

表 2 产品-属性值表

	attr ₁	Attr ₂	...	Attr _i	...	Attr _k
Item ₁	T _{1,1}	T _{1,2}	...	T _{1,i}	...	T _{1,k}
⋮	⋮	⋮	...	⋮	⋮	⋮
Item _i	T _{i,1}	T _{i,2}	...	T _{i,i}	...	T _{i,k}
⋮	⋮	⋮	⋮	⋮	⋮	⋮

通过用户交互界面 agent 可以获取用户搜索关键字(用户需求产品信息)，在基于 MAS 的推荐系统中，从后台数据库中提取用户 u 所参评的所有产品的信息，每个产品的评分可以看作是对该产品关键属性值的评分。由于不同种类的产品有不同属性值，因此需要对每个属性的值进行统计，分别得到各个属性值的评分集合。为了得到产品属性值信息，本文在创建数据库时设计完成了产品-属性值表，如表 2 所示。表中行表示产品列表，列表示产品属性， $T_{i,j}$ 表示第 i 个产品的第 j 个属性值。由于在不同的属性值集合中，属性的取值也会有所不同(相同的属性在不同产品中具有不同属性值)，因此本模型令 $T[j,t]$ 表示属性 j 的第 t 个属性值，则用户 u 对属性值 $T[j,t]$ 的兴趣偏好度 $P(u, j, t)$ 表示为用户 u 对 $T[j,t]$ 的不同评分的平均值，上述不同评分来源于用户购买产品记录中对产品的评分，于是有：

$$P(u, j, t) = \frac{\sum_{n=1}^N R_n(u, j, t)}{N} \quad (2)$$

其中 N 表示用户 u 对属性 j 的取值 $T[j,t]$ 的评分总次数， $R_n(u, j, t)$ 表示相应的评分值。从而可以得到用户 u 的属性值偏好矩阵 UPM ，如图 2 所示。

$$UPM = \begin{pmatrix} p(u, 1, 1) & p(u, 2, 1) & \dots & p(u, k, 1) \\ p(u, 1, 2) & p(u, 2, 2) & \dots & p(u, k, 2) \\ \vdots & \vdots & \dots & \vdots \\ p(u, 1, N) & p(u, 2, N) & \dots & p(u, k, N) \end{pmatrix}$$

图 2 用户-属性值偏好矩阵

3.2 基于 MAS 推荐系统模型的实现

对于用户 u 和初始邻居集中 $S(u)$ 的任意用户 v ，首先生成各自属性值偏好矩阵 UPM_u 和 UPM_v ，之后找到二

者共同评分的属性集 $R(u \cap v)$ 及评分, 从而在基于用户一属性评分基础上对用户的相似性进行度量, 用户 u 和 v 之间的属性相似性计算公式可表示为:

$$sim_{attr}(u, v) = \frac{\|R(u \cap v)\|}{1 + \sum_{j=1}^{|R(u \cap v)|} \sqrt{p(u, j, t) - p(v, j, t)}} \quad (3)$$

这里, $\|R(u \cap v)\|$ 表示用户 u 和 v 共同评分属性集的大小。同理, 可计算出用户 u 与其他最近邻居集的相似度, 从而得到用户最近邻集合。从最近邻所购买的或给予好评的产品集中删除用户 u 已购买的或给予好评的产品, 剩下的产品即为针对客户 u 的候选推荐集, 以搜索 agent 中记录的关键词(产品信息), 产生产品推荐集 $Set(pro)$ 。此外若推荐产品符合推荐较多, 可以采用下式首先计算用户对候选集产品的预测评分:

$$p_{attr}(u, j) = \bar{r}_i + \frac{\sum_{j=1}^n sim(i, j) \cdot (r_{u, j} - \bar{r}_j)}{\sum_{j=1}^n |sim(i, j)|} \quad (4)$$

其中, \bar{r}_i 和 \bar{r}_j 表示产品 i, j 的平均评分, 且 $i, j \in Set(pro)$, $sim(i, j)$ 则表示产品 i 与相似邻居项目之间的相似性, 可用文献^[12]公式即可计算得到。然后对候选产品集内产品进行预测评分并排序, 选择 TOP-N 作为产品推荐集。

4 仿真实验的结果和分析

4.1 实验建立

基于 MAS 技术, 本文设计建造了一个基于 MAS 的推荐系统, 并将其应用于电脑产品推荐。在系统的搭建过程中, 本文采用 ASP 开发该推荐系统进行实验数据收集, 后台数据库采用 SQL sever2000 组建, 实验环境是 Window XP SP2, 计算机主频为 2100MHz, 内存为 1G。本实验首先在淘宝网随机抓取了 496 种电脑信息。实验有 91 位管理学院研究生注册并进行模拟评分, 评分值采用 5 分制。要求用户至少要对 40 种电脑产品进行评分, 在一周内新系统一共收到 10810 评分数据, 并以此作为实验数据集。实验根据淘宝商铺对电脑产品的描述文件, 依次选取属性项是: 品牌、价格、内存大小、主板类型、CPU 类型和硬盘容量。实验从每个用户的评分数据中随机择取 10 条组成测试集, 其余数据作为训练集。

4.2 评价标准

评价推荐系统推荐质量的度量标准采用统计度

量方法中绝对平均差 MAE(mean absolute error) 进行度量。MAE 通过计算预测的用户评分与时间的用户评分之间的偏差来度量预测的准确性, MAE 越小, 推荐质量越高。对于评分数据、预测的评分集合表示为 $\{p_1, p_2, \dots, p_N\}$ 实际的评分集合表示为 $\{q_1, q_2, \dots, q_N\}$, 则平均绝对值偏差 MAE 定义为:

$$MAE = \frac{\sum_{i=1}^N |p_i - q_i|}{N} \quad (5)$$

4.3 实验结果与分析

实验根据第 3 节提出的模型, 构建了新推荐系统 MASRS, 本文将其模型与传统的过滤算法进行比较, 如基于用户系统过滤推荐算法(UPCC)及基于项的协同过滤推荐算法(IPCC), 使用(4)式对推荐集产品进行预测评分以进行 MAE 测试, 实验结果见图 3。

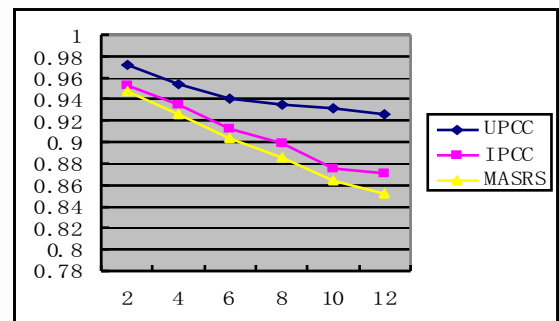


图3 MAE 对比(横轴为用户近邻数目, 纵轴为 MAE 值)

由于 MAE 的值越小, 推荐的精度越高。由图三可以看出, MASRS 在产生推荐结果上要优于 UPCC 和 IPCC。本文模型是基于用户一属性值偏好矩阵进行用户相似性度量的, 从而使某些用户尽管共同评分产品过少但由于在共同属性值具有近似评分而成为最近邻, 然后根据最近邻产生产品推荐集。由此缓解了用户评分数据稀疏性, 提高了推荐质量。

5 结语

本文通过采用 MAS 技术的先进性, 构建了一个基于 MAS 的协同过滤推荐系统, 明确提出利用 MAS 实现推荐智能化, 并详细描述了这个系统的实现原理和其在具体环境下的运行机制。在面对推荐系统不断产生用户评分稀疏性问题和系统扩展性上。本文提出利用用户-项评分矩阵生成用户一属性值偏好矩阵 UPm

方法,然后对用户进行相似性度量,通过剔除和规范,产生产品推荐集 **Set(pro)**,从而缓解评分稀疏性问题,通过 **MAS** 技术优势,将大量计算在线下完成,对比实验证实该系统模型的有效性。同时新系统通过用户反馈 **agent** 可不断收集更多用户购买和评价相关信息,随着用户使用时间的增加推荐质量会逐渐提高,但 **MASRS** 仍然存在“冷启动”问题,这也是该系统今后需要改进的地方。

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. *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. *Proc. of the 14th Conference on Uncertainty in Artificial Intelligence*. 1998.43 – 52.
- 3 Burke RD. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adopted Interaction*, 12(4):331 – 370.
- 4 Herlocker JL, Konstan JA, Terveen LG, et al. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 2004, 22(1):5 – 53.
- 5 Schafer JB, Konstan JA, Riedl J. E-Commerce recommendation applications. *Data Mining and Knowledge Discovery*, 2001,5(1-2):115 – 153.
- 6 Schafer JB, Konstan JA, Riedl J. Recommender system in E-commerce. *Proc. of the ACM Conference on Electronic Commerce: ACM Press*, 1999.158 – 166.
- 7 Pazzani M. A frame for collaborative, content-based and Demographic filtering. *Artificial Intelligence Review*, 1999,13(5-6):393 – 408.
- 8 Billsus D, Pazzani M. Learning collaborative information filters. *Proc. of Int’l Conf on Machine Learning*. San Francisco: Morgan Kaufmann Publishers Inc, 1998. 46 – 54.
- 9 Sarwar B, Karypis G, Konstan J, Riedl J. Application of dimensionality reduction in recommender systems: A case study. *Proc. of the ACM WebKDD Workshop*. 2000.
- 10 Xue GR, Lin CX, Yang Q, et al. Scalable collaborative filtering using cluster-based smoothing. *Proc. of Mini Micro System*, 2004,25(9):1665 – 1670.
- 11 邓爱林,朱扬勇,施伯乐.基于项目评分预测的协同过滤推荐算法. *软件学报*, 2003,14(9):1621 – 1628.
- 12 Karypis G. Evaluation of item-based Top-N recommendation. *ACM Transactions on Information Systems*, 2004,22(1):143 – 177.