文章编号：1007-757X(2021)10-0205-04

基于关联规则算法的电子商务商品推荐系统设计与实现

宋倩

（咸阳职业技术学院 财经学院，陕西 咸阳 712000）

摘　要：互联网时代给人们的消费带来了便利，但琳琅满目的商品也为用户带来了选择困难，在没有明确需求的情况下，如何为消费者推荐存在潜在商机的商品是电商急需解决的难题。为了提升商品推荐的精确度，设计了基于关联规则算法的电子商务商品推荐系统，对 FP_Growth 算法进行优化改进，提出了一种更高效的 CTE-MARM 挖掘算法，构建关联规则库，并与用户兴趣商品链联合分析，将具备强关联关系的商品按照用户兴趣度高低选取 TOP-N 进行推荐，经过测试验证实际命中率较高，提升分析效率的同时也为商家后续营销决策提供有力的数据支撑。

关键词：数据挖掘；关联规则；CTE-MARM 算法

中图分类号：TP311.13　　**文献标志码**：A

Design and Implementation of E-commerce Commodity Recommendation System Based on Association Rule Algorithm

SONG Qian

(Department of Finance and Economics, Xianyang Vocational and Technical College, Xianyang 712000, China)

Abstract: The Internet era has brought convenience to people's consumption activities, but a wide range of goods has also brought difficulties for users to choose. In the absence of clear demand, how to recommend goods with potential business opportunities to consumers is an urgent problem for e-commerce. In order to improve the accuracy of the recommendation, this paper designs an electronic commerce recommendation system based on association rules algorithm, optimizes the algorithm of FP_Growth. This paper proposes a more efficient mining algorithm of CTE-MARM, constructs library association rules, by associating with the user interest commodity chains analysis. The algorithm has strong correlation of goods in accordance with the user interest degree of discretion, provides TOP-N recommendations. Through tests, it is verified that the actual percentage is higher, enhances the analysis efficiency, also provides powerful data for merchants subsequent marketing decision support.

Key words: data mining; association rules; CTE-MARM algorithm

0　引言

随着信息技术的发展，人们在享受科技带来便利的同时也受到了信息过载的困扰，对于电子商务领域同样如此。如何帮助用户获得满意商品是电商企业获取利润和提升自身信誉的关键，电子商务商品推荐系统应运而生，将数据挖掘技术应用于用户日常购物活动的场景之中，利用关联挖掘算法分析历史数据来实现潜在商机预测，既为用户节约了寻找感兴趣商品的时间，也为商家提升了销量及用户忠诚度。

1　需求分析

1.1　功能需求

电子商务的商品推荐系统主要是根据收集的用户浏览行为以及历史消费记录分析其兴趣偏好、挖掘预测潜在购买商机进行推荐。其中最关键的是个性化以及实时性。系统的主要功能体现在以下几个方面。

（1）数据采集：提取相关记录及行为数据。

（2）数据预处理：剔除无用数据、确保数据完整。

（3）用户兴趣分析：构建用户兴趣模型，分析积累用户兴趣商品库。

（4）关联规则库：根据挖掘算法构建关联规则库。

（5）商品推荐：推荐用户感兴趣的商品。

1.2　关键参数

（1）置信度：降低关联规则中“规则爆炸”情况，提升算法精准率。

（2）时效度：在实际场景中，人们的购物习惯是在不断变化的，距现在时间越近的相关浏览记录或购买记录越能代表当前的需求偏重点。

（3）兴趣度：在电子商务领域可以反映用户兴趣的因素有很多，包括购买、浏览、收藏、评分、评论等。兴趣度体现为将多种商务行为经过算法策略得出的兴趣程度的权值[1-2]。

2　相关技术

2.1　数据挖掘

数据挖掘是一门新兴的技术，是多学科综合形成的产物，指的是从海量不完整的、存在脏数据的、比较模糊的数据

作者简介：宋倩(1977-)，女，硕士，讲师，研究方向：电子商务、物流管理。

集中抽取出未知的有潜在价值的、有意义的模式或规律的计算过程，主要包括数据清洗、数据集成、数据选择、数据转换、数据挖掘、数据评估及数据展示。

（1）清洗与集成：数据质量是保证挖掘出来的知识可靠性的基础，需要清除重复数据、不完整数据、脏数据，并将多个数据源的数据集成到一起完成后续操作。

（2）选择与转换：选择将进行数据挖掘的目标，针对不同数据类型进行统一化处理，消减特征维数，降低不必要的计算。

（3）挖掘与评估：运用相关的聚类、分类算法进行数据计算，从挖掘结果中根据一定的评估标准选出有意义的知识。

2.2 关联规则

关联反映的其实是事物之间的依赖关系，其中两个或多个属性之间的取值如果呈现规律，则认为有关联关系，根据其中一项属性值即可预测其他属性值。在数据挖掘领域中基于关联规则挖掘的研究是其中的重要研究方向。关联规则挖掘的基本概念如下。

（1）数据项与数据项集：设 $I=\{i_1,i_2,\cdots i_m\}$ 是 m 个不同项的集合，则每个 $i_k(k=1,2,\cdots,m)$ 代表数据项，I 为数据项集。

（2）事务：事务 T 是数据项集的非空子集，每个事务与唯一标识符对应，记为 TID。多个事务构成事务集 D。

（3）支持度：假设 X 为数据项集，A 为事务集 D 中所有事务数量之和；B 为 D 中包含 X 的数量之和，则 X 的支持度为 $\text{support}(X)=\frac{B}{A}$。

（4）关联规则：$X\Rightarrow Y$ 用的形式表示，其中 $X\subset I, Y\subset I$ 且 $X\cap Y=\varnothing$，表示如果 X 项集在某一事务中出现，则 Y 也会出现。

（5）关联规则置信度：置信度指的是包含 X 和 Y 的事务数与包含 X 的事务数的比值，置信度越高，关联规则的可靠性越好。计算式为式(1)。

$$\text{confidence}(X\Rightarrow Y)=\frac{|\{T:X\cup Y\subseteq T,T\in D\}|}{|\{T:X\subseteq T,T\in D\}|} \quad (1)$$

（6）最小支持度和最小置信度：最小支持度用来过滤出现频率低的项集，最小置信度用来剔除可靠性低的关联规则。

3 关键算法

3.1 FP_Growth 算法

FP_Growth 算法采用模式增长的方式来发现频繁项集，首先建立一棵频繁模式数 FP_tree，存放事务集的所有频繁项集，然后将树中压缩后的事务集划分为一组条件事务集，每个事务集关联一个频繁项，分别挖掘每个条件事务集。该算法可以明显压缩被搜索的事务集[3-4]。

3.2 CTE-MARM 算法

由于 FP_tree 频繁项集查找时存在节点多、递归调用次数多等问题，本文针对电子商务商品推荐的个性化应用性问题，对 FP_Growth 算法做出优化与改进，提出一种基于 FP_Growth 算法的约束事务扩展多层关联规则挖掘算法 CTE-MARM(Constraint Transaction Extension—Multi-level Association Rule Mining)，以此提升挖掘效率、减少冗余规则。主要改进项如下。

（1）对每条事务基于 K 层次约束扩充，将每个事务项的前 $k-1$ 个祖先项添加到当前事务，之后剔除重复项，既约束数量的扩展又可以保证发现关联规则。

（2）创建 FP_tree 时对每个节点添加两个域：Condition-Memory 用来存放结点前缀路径上的结点、IsVisited 用来判断当前结点是否被遍历，避免多次回溯。

（3）增加风险度阈值指标，确保事务约束扩展层次 k 选值合理。

4 用户兴趣模型

用户兴趣模型是电子商务商品推荐系统的核心，是确保推荐质量的关键模块。首先，构建商品－用户行为特征矩阵。

	$good_1$	…	$good_j$	…	$good_n$
点击次数 cc	W_{c1}	…	w_{cj}	…	w_{cn}
访问次数 rc	W_{r1}	…	w_{rj}	…	w_{rn}
停留时间 st	W_{s1}	…	w_{sj}	…	w_{sn}
用户评价 el	W_{e1}	…	w_{ej}	…	w_{en}

然后，向量空间模型采用三元组〈用户，商品集，兴趣集〉的形式，〈$user_i$，goods，interests〉，其中，

$user_i$ 为具体用户；

goods 为商品集合，goods＝〈$good_1$，$good_2$，…，$good_j$，…$good_n$〉；

interests 为兴趣度集合，interests＝〈IR_{i1}，IR_{i2}，…，IR_{ij}，…IR_{in}〉；

IR_{ij} 为用户 i 对商品 j 的兴趣度。

$$I_{ij}(\text{cc},\text{rc},\text{st},\text{el})=w_{\text{cc}}\alpha_1+w_{\text{rc}}\alpha_2+w_{\text{st}}\alpha_3+w_{\text{el}}\alpha_4,\ (\textstyle\sum_{i=1}^{4}\alpha_i=1)$$

（cc：鼠标点击次数；rc：重复访问页面次数；st：停留时间；el：用户对商品评价分值；w：相关权重；α：加权参数）

通过三元组记录用户感兴趣的商品集，按 IR_{ij} 降序排列，采用 TOP-N 策略，选取 IR 值前 N 个商品形成“用户－兴趣商品链”。另外，还需要考虑其他数据挖掘出的关联规则，利用改进的 CTE-MARM 算法进行关联挖掘，如果有用户－兴趣商品链表中的 $good_i$ 存在强关联规则的 $good_k$，则将 $good_k$ 也添加到三元组中[5-6]。模型构建流程如图 1 所示。

5 电子商务商品推荐系统设计

5.1 开发环境

数据库：MySQL

编译环境：JDK＋Tomcat

开发工具：MyEclipse

开发语言：Java、JavaScript、MySQL

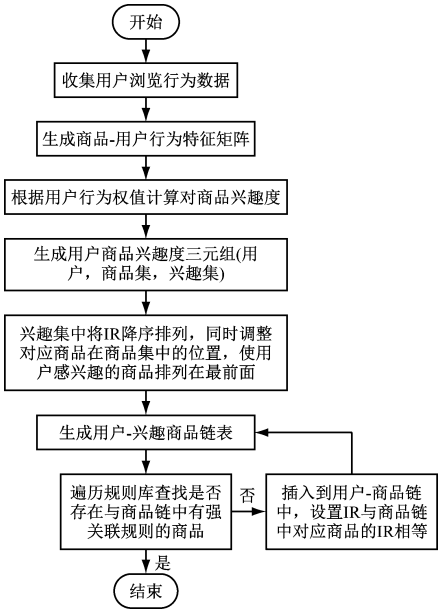

图 1　用户兴趣模型构建流程

5.2　总体设计

基于关联规则算法，本文将电子商务商品推荐系统功能划分为数据采集及清洗、用户兴趣分析、构建关联规则库、TOP-N 推荐 4 个部分，整体架构设计如图 2 所示。

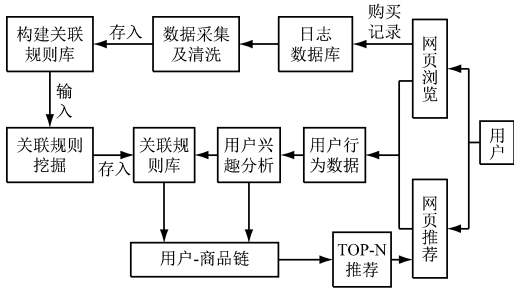

图 2　系统整体架构图

(1) 数据采集及清洗模块：数据采集包括两类数据，一类是已购买记录，一类是各种商务行为数据。这两类数据是关联规则挖掘以及用户兴趣度分析的基础。另一方面，由于原始数据包含很多杂乱、模糊、残缺的脏数据，为提升挖掘效率需要对原始数据进行清洗，通过检测空订单、检测交易记录中商品是否存在，剔除交易记录中无关字段等方法对数据完成简化处理[7-8]。

(2) 用户兴趣分析模块：负责建立和更新兴趣模型，前台自动获取用户行为，存于相应数据库表，按照前文阐述的计算兴趣度算法构建用户－兴趣商品链。与关联规则库联合，查找强关联商品，作为 TOP-N 推荐模块的输入。

(3) 构建关联规则库模块：采用前文介绍的 CTE-MARM 算法根据具体需求来挖掘指定层次间的关联，约束层次 k 的值设置为 2，针对事务交易表按置信度高低挖掘出规则存于规则表。

(4) TOP－N 推荐模块：在关联规则库中查询与用户兴趣模型中存在强关联关系的商品进行网页可视化展示。

5.3　数据库设计

鉴于电子商务商品推荐系统的行为分析及商品推荐功能，构建数据库 E－R 模型，如图 3 所示。

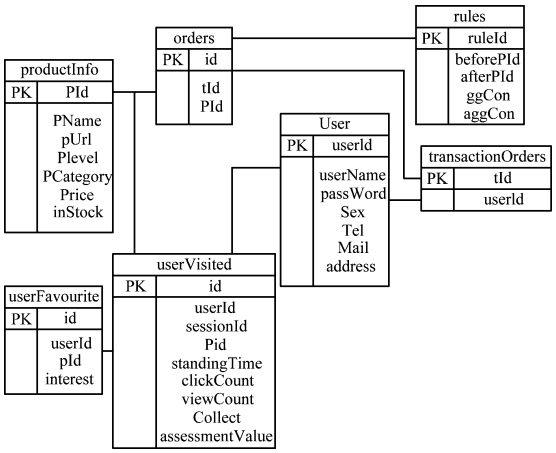

图 3　数据库 E-R 图

其中，核心数据库表如下。

(1) 用户信息表 user：主要包括用户 ID、注册名称、密码、性别、联系方式、地址。

(2) 商品信息表 productInfo：主要包括商品 ID、名称、链接、层次、类别、价格、库存。

(3) 事务交易表 transactionOrders：主要包括事务交易 ID、用户 ID。

(4) 购物记录表 orders：主要包括唯一标识 ID、事务交易 ID、商品 ID。

(5) 用户行为表 userVisited：主要包括唯一标识 ID、用户 ID、商品 ID、点击次数、重复访问次数、浏览时间、是否收藏、评价得分。

(6) 关联规则表 rules：主要包括唯一标识 ID、规则前件商品 ID、规则后件商品 ID、商品间置信度、类别与商品间置信度。

(7) 用户兴趣表 userFavourite：主要包括唯一标识 ID、用户 ID、商品 ID、兴趣度。

6　系统测试与验证

为验证本文构建模型及推荐商品的准确性，将 3 000 个用户的购买信息作为测试数据，通过人工统计的结果、人工分析的用户习惯以及系统计算结果进行对比，每测试一次后，将数据重新交叉组合，共测试 5 次，得到结果如图 4 所示。

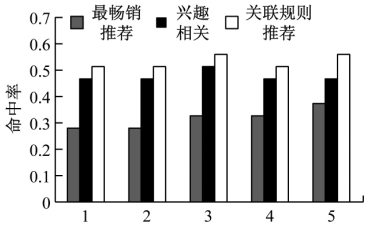

图 4　人工分析与系统算法结果命中率分析

结果表明，人工分析的最畅销商品命中率为 31.8%，人工分析用户兴趣推荐的命中率为 51.1%，而基于关联规则算法构建的电子商务商品推荐系统推荐的商品命中率为 55.8%，相对来说系统算法更为有效，命中率比较高，符合最

初设计预期[9]。

7 总结

本文利用优化改进的 CTE-MARM 算法提升挖掘效率，构建关联规则库，并通过用户兴趣模型实现用户－兴趣商品链分析，两者联合起来为用户提供 TOP-N 商品推荐，经验证具有较高实用性。但在多层关联挖掘算法优化方面还需进一步探索，在商品推荐命中率提升方面还需要考虑更多的影响因素与技术手段。

参考文献

[1] 林穗，郑志豪. 基于关联规则的客户行为建模与商品推荐研究[J]. 广东工业大学学报，2018，35(3)：90-94.

[2] 郝海涛，马元元. 基于加权关联规则挖掘算法的电子商务商品推荐系统研究[J]. 现代电子技术，2016，39(15)：133-136.

[3] 张勇杰，杨鹏飞，段群，等. 基于关联规则的商品智能推荐算法[J]. 现代计算机(专业版)，2016(10)：25-27.

[4] 刘枚莲，刘同存，肖吉军. 基于双向关联规则的网络消费者偏好挖掘研究[J]. 微电子学与计算机，2013，30(3)：20-26.

[5] 王宝军. 基于商品标签化与关联规则的电子商务商品关联推荐研究[J]. 商场现代化，2020(11)：8-10.

[6] 黄玲，余霞. 基于云平台的电子商务商品智能推荐系统[J]. 现代电子技术，2020，43(5)：183-186.

[7] 邓灵斌，申慧. 电子商务平台商品推荐信息特性对消费者购买意愿的影响实证研究[J]. 南华大学学报(社会科学版)，2019，20(2)：60-65.

[8] 姚剑，余炎，黄诗盛，等. 基于个性化导购的商品智能动态推荐系统[J]. 价值工程，2017，36(35)：199-201.

[9] 游运，万常选，陈煌烨. 考虑对象关联关系的多样化商品推荐方法[J]. 计算机工程与应用，2018，54(7)：70-76.

(收稿日期：2020.11.15)

(上接第 204 页)

提供最精准的故障信息。系统采用模拟动态波形图的方式降低对联动更新的影响[9]。中间节点的文件中记录了版本信息，一旦检测到关联信息变化，则更新逻辑模型。

3.3 可视化展示

系统支持两种故障仿真展示方式：自动方式是根据中间节点文件记录的时刻定时仿真，手动方式可以按照时间逐条仿真。在仿真过程中，逻辑图与故障录波波形图有所关联，系统支持逻辑图的缩放和位置移动，故障录波波形图支持横向、纵向缩放、波形复位。并在逻辑仿真存在异常时发出告警，以便采取后续处理措施。

3.4 版本兼容

系统具有高兼容性，对于非标准化编码的情况也同样适用，可识别不同型号装置的录波文件。系统版本管理采用自适应模式，根据装置版本分类存储，利用图形化界面创建索引，实现与逻辑模型的相互映射并在模型数据库中保存，升级装置版本时通过逻辑图及映射关系的增量配置即可快速实现，从而使系统的复用度和兼容性得以有效提升[10]。

4 总结

为了实现可视化变电站故障分析，提高实时性，构建了变电站故障过程可视化分析系统，单独装置在平台之上，操作简单直观、通用性强，采用加密 G 语言描述逻辑图文件，兼顾安全性与系统灵活性，增强系统可扩展性。采用 SQLite 数据库存储中间节点文件与逻辑图模型的映射关系，通过通信服务模块管理装置的中间节点文件，支持可视化展示逻辑图、故障波形曲线，支持定时刷新，支持自适应版本升级。系统既可以减少变电站的故障分析工作量，又可以提升故障处理的时效。

参考文献

[1] 李兴美，尹文兆，段兰，等. 智能变电站过程层故障诊断及定位技术研究[J]. 科技创新与应用，2018(32)：145-146.

[2] 余成波，曾亮，张林. 基于 OTSU 和区域生长的电气设备多点故障分割[J]. 红外技术，2018，40(10)：1008-1012.

[3] 单志鹏，张美勇. 电力系统变电站实时故障诊断技术[J]. 科技与创新，2018(17)：145.

[4] 邹辉，黄福珍. 基于 FAsT-Match 算法的电力设备红外图像分割[J]. 红外技术，2016，38(1)：21-27.

[5] Huda A S N，Taib S，Ghazali K H，*et al*. A new thermographic NDT for condition monitoring of electrical components using ANN with confidence level analysis[J]. ISA Transactions，2014，53(3)：717-724.

[6] 刘清泉，郝晓光，刘勇，等. 智能变电站物理回路故障诊断方案设计及实现[J]. 供用电，2018，35(2)：79-84.

[7] 徐蔚波，刘颖，章浩伟. 基于区域生长的图像分割研究进展[J]. 北京生物医学工程，2017，36(3)：317-322.

[8] 沈萍萍，余勤. 基于离散余弦变换的非局部均值图像去噪算法[J]. 计算机工程与设计，2017，38(1)：183-186.

[9] 张文军，王晓忠，易善军，等. 智能变电站监测端口系统设计及其实现[J]. 微型电脑应用，2020，36(12)：155-157.

[10] 方鸣，他悦蓉. 变电站继电保护常见故障分析与对策探讨[J]. 冶金管理，2020(19)：41-42.

(收稿日期：2020.10.15)