

# 基于大数据技术的电子商务个性化信息 自动推荐算法应用\*

李加军

(广东财经大学华商学院 数据科学学院, 广东 广州 511300)

**摘要:** 当前个性化推荐算法应用因未构建增量更新推荐模型,对电子商务平台的大数据进行处理,导致推荐结果不能及时更新,严重影响用户体验。因此,提出基于大数据技术的电子商务个性化信息自动推荐算法。首先对平台中用户与项目的相异度进行计算,利用计算后的相异度矩阵构建数据增量更新模型;然后,将该模型作为个性化推荐模型,引入IU-UserCF推荐算法,提高预测准确性;最后以电商物流为应用对象,将所提方法引入应用进行探索。实验对比结果表明,所提方法在电子商务应用中有着更加出色的推荐效果。

**关键词:** 电子商务;个性化推荐算法;相异度模型;大数据技术;IU-UserCF

中图分类号:TP29 文献标识码:A 文章编号:1003-7241(2021)10-0038-05

## Application of Personalized Information Automatic Recommendation Algorithm in E-commerce Based on Big Data Technology

LI Jia-jun

(Huashang College Guangdong University of Finance and Economics, School of Data Science, Guangzhou 511300 China)

**Abstract:** At present, the application of personalized recommendation algorithm does not build incremental update recommendation model to process the big data of e-commerce platform, resulting in the recommendation results can not be updated in time, which seriously affects the user experience. Therefore, an automatic recommendation algorithm for personalized information of e-commerce based on big data technology is proposed. Firstly, the dissimilarity degree of users and projects in the platform is calculated, and the incremental updating model of data is constructed by using the difference degree matrix after calculation; then, the model is taken as the personalized recommendation model, and the iu-usercf recommendation algorithm is introduced to improve the prediction accuracy; finally, taking e-commerce logistics as the application object, the proposed method is introduced into the application for exploration. Experimental results show that the proposed method has better recommendation effect in e-commerce applications.

**Key words:** E-commerce; personalized recommendation algorithm; dissimilarity model; big data technology; IU UserCF

### 1 引言

目前,电子商务领域不断发展<sup>[1]</sup>,网上购物已经是当下新潮,电商用户呈现爆炸式增长,提供的商品也是种类

繁杂,因此,每时每刻都会产生大量的数据。为了解决“数据过载”的现象,个性化自动推荐技术已经在电子商务中成为了重要的应对手段,且被广泛使用<sup>[2]</sup>。

当前个性化自动推荐技术忽视了数据处理的难点,造成个性化自动推荐结果不能实时更新,影响用户的体验。针对这一问题,Guan Y<sup>[3]</sup>等人采用堆叠式自动编码网络将多视点信息映射到统一的潜在空间,增加认知层

\*基金项目:广东财经大学华商学院校内导师制科研项目(编号2020HSD S28);广东财经大学华商学院校级质量工程电子商务特色专业(编号HS2019ZLGC05)

来描述消费者的异质认知风格,并引入集成模块来反映多视点潜在表征的交互作用。虽然该方法推荐的结果覆盖更广泛,但未使用数据增量更新模型对用户和项目进行相异度计算,所以推荐结果不够精准。李宇琦<sup>[4]</sup>等人提出了基于网络表示学习的个性化商品推荐的方法,首先,通过历史记录获取项目的顺序信息,再将项目映射到低维空间中,并利用低维向量对相似度进行评估,计算出用户的偏好,完成个性化自动推荐。但该算法筛选时间耗时过长、效率较低,难以被广泛地应用。

因此,提出一种基于大数据技术的电子商务个性化信息自动推荐算法。将用户与项目相异度进行分析计算,并从推荐更新增量着手,构建数据增量更新模型,对含有变化和增加的数据进行局部更新。提出引入IU-UserCF的个性化自动推荐算法,最后,以电商物流作为应用对象,将所提算法运用其中。实验对比结果表明,该方法在实际应用当中,拥有高效处理海量数据及增量更新的能力,而且预测精准度较高。

## 2 构建数据增量更新推荐模型

当前的个性化推荐算法,通常是每隔一段时间将用户现有评分数据进行相异度计算<sup>[5-6]</sup>,但是该方法需要运算数据量过大,计算量也过多。因此,提出构建数据增量更新推荐模型。先对用户和项目的相异度进行评分向量计算,将得到的相异度矩阵作为模型构建核心,使发生变化的局部数据通过模型实时更新。该模型可以对大数据进行有效筛选处理、减少工作量、提高效率。

### 2.1 项目相异度

首先,将同一用户对不同项目的评价,通过评分相减计算得出项目的相异度,衡量项目相异的指标。

将 $x$ 设为评分集, $i$ 和 $j$ 为任意两个项目,同一用户 $U$ 对其评分分别为 $U_i$ 和 $U_j$ ,评分后的用户设为 $S_i(x)$ , $count(S_i(x))$ 为用户数量,项目 $i$ 和 $j$ 项目的相异度计算如公式(1)所示:

$$dev = \sum \frac{U_i - U_j}{count(S_i(x))} \quad (1)$$

由公式(1)得知,如果用户对 $i$ 和 $j$ 的评分结果十分接近的话,则项目的相异度非常小,相反则大。项目的相异度计算是从用户评分维度上进行分析的结果,如果从项目评分维度上进行分析的话将得到用户的相异度计算。

### 2.2 用户相异度

利用不同用户对相同项目的评分相减提取绝对值,

得出用户的相异度。通过该计算可以用来衡量用户偏好的相异程度。

将 $X$ 设为用户评分集, $A$ 和 $B$ 为任意两个用户,同一项目 $I$ 分别由用户 $A$ 和 $B$ 进行评分,用户 $A$ 的评分为 $A_i$ ,用户 $B$ 的评分为 $B_i$ ,同时被 $A$ 和 $B$ 评过的项目为 $S_2(x)$ , $count(S_2(x))$ 为项目数量,用户 $A$ 和用户 $B$ 的相异度计算如公式(2)所示:

$$dev_1 = \sum \frac{A_i - B_i}{count(S_2(x))} \quad (2)$$

由公式(2)得知,如果用户 $A$ 和 $B$ 用户在同一项目的评分结果都十分接近的话,则用户的相异度很小,相反则大。

### 2.3 数据增量更新模型构建

当用户的相异度计算出来后,就可以按照参数构建数据增量更新推荐模型,将有变化的项目评分进行向量计算,使变化数据通过模型实时更新。如图1所示为数据增量更新推荐流程图。

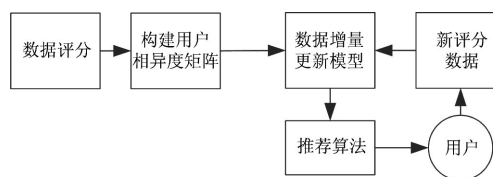


图1 增量更新推荐流程

#### 2.3.1 输入数据

用户的评分数据集代表输入的数据量,而且必须保证每行只有一个数据集。将用户的ID设为第一列;项目的ID设为第二列;用户对项目的评分设为第三列。该数据集其实是一个 $m \times n$ 的矩阵,设为 $R$ ,其中,用户的数量为 $m$ ,项目的数量为 $n$ ,用户 $m$ 对 $n$ 项目的评分如公式(3)所示:

$$R_{mn} = \begin{cases} r_{11}, r_{12}, \dots, r_{1n} \\ r_{21}, r_{22}, \dots, r_{2n} \\ \vdots \\ r_{m1}, r_{m2}, \dots, r_{mn} \end{cases} \quad (3)$$

#### 2.3.2 构建用户相异度矩阵

基于 $R_{mn}$ ,利用公式(2)计算两个用户之间的相异度,并记录两位用户同时评分过的项目数量,相异度矩阵 $D$ 由公式(4)计算得出<sup>[7]</sup>:

$$D = \begin{cases} --, (dev_{12}, count_{12}), \dots, (dev_{1m}, count_{1m}) \\ (dev_{12}, count_{12}), --, (dev_{2m}, count_{2m}) \\ (dev_{m1}, count_{m2}), (dev_{n1}, count_{n2}), \dots, -- \end{cases} \quad (4)$$

其中,用户的相异度为 $D$ ,两位用户同时评分过的项目数量为 $count$ ,对比等式如公式(5)所示:

$$(dev_{m1}, count_{m2}) = (dev_{n1}, count_{n2}) \quad (5)$$

所以 $D$ 是一个对称的相异度矩阵。

### 2.3.3 数据增量更新策略

由上述计算得知,相异度矩阵可以对新增数据进行局部在线更新,而且每次更新时只使用一个评分向量,不需要与其他向量相互运算,从而节省运算时间。具体更新方法如下所示。

设项目 $I$ 过运算后,又新增若干条用户评分,若 $N_1$ 为该项目的新增用户, $N_2$ 为旧用户的评分集合,选取 $A_1$ 和 $B_1$ 为任意两位用户,用户 $A_1$ 和 $B_1$ 对项目 $I$ 的评分为 $a_1$ 和 $b_1$ ,并且 $a_1 \in N_1, b_1 \in (N_1 * N_2)$ ,那么, $A_1$ 与 $B_1$ 之间的相异度更新方法如公式(6)所示:

$$dev_{A_1 B_1} = dev_1 + |b_1 - a_1| \quad (6)$$

其中, $dev_{A_1 B_1}$ 为旧用户的相异度。

因为增量更新模型一次只需要处理一个向量,不需要运用所有的评分数据进行更新,所以具有工作量少,处理速度快等特点。

## 3 引入IU-User的个性化自动推荐算法

IU-UserCF是在数据增量更新模型基础上提出的个性化自动推荐算法。首先,选取同时出现次数较多的用户、相异度较小的用户作为最近邻居<sup>[8]</sup>,然后根据邻居的评分数据向用户推荐更加符合的项目。该模型的个性化自动推荐算法在应用中有着良好的适用性和极高的预测精准度。引入IU-User的个性化自动推荐算法流程如图2所示。

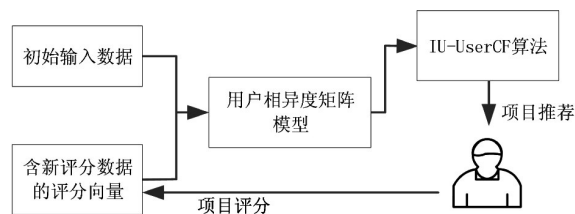


图2 基于IU-User算法的个性化自动推荐流程

首先利用相异度计算方法找到与用户 $u$ 兴趣偏好临近的用户,并将临近用户出现次数限定在5次以内,然后通过阈值计算构成最近的邻居集合 $nn$ ,计算如公式(7)所示:

$$(dev < 1count > 5) = u \in nn \quad (7)$$

通过对用户 $u$ 的最近邻居 $nn$ 中的项目评分计算得到用户 $u$ 对于项目 $K$ 的预测打分 $P$ ,由此得到用户的相应推荐,计算方法如公式(8)所示:

$$P = \frac{\sum_{nn} C * R}{\sum_{nn} C * u} \quad (8)$$

$C$ 为出现相同的次数,在公式中作为权重的存在,邻居用户 $u$ 对项目 $K$ 的评分为 $R$ 。

## 4 基于个性化自动推荐算法的电商物流应用

近年来,电子商业如日方升,越来越多的商家通过网络渠道将商品进行售卖,网上购物已经是不可缺少的营销模式。因此,物流行业也得到了前所未有的发展,人们提供的物流服务也越来越多。但是很多用户因为“数据过载”的问题,导致无法匹配到更符合自己的物流服务。而且,一些优质的中小物流企业因为得不到优秀的推荐导致客源少、使用效率低、成本开销大等问题,阻碍了企业的发展。因此,具备大数据处理能力的个性化自动推荐算法服务系统可以准确、智能地为用户提供精准物流服务体验。

### 4.1 需求分析

所提算法能够实时从项目评价数据库中筛选出用户偏好的评分数据,并通过这些数据提供符合用户的物流服务。具体内容如下所示:

- (1) 对相关的物流服务的评分数据进行调取与匹配;
- (2) 将符合用户的物流服务进行在线匹配与推荐;
- (3) 利用算法实现个性化自动推荐技术的管理。

### 4.2 电商物流业务流程分析

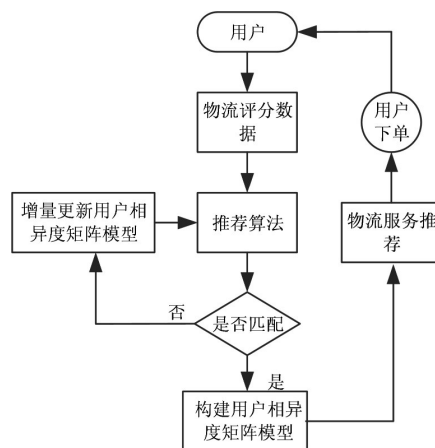


图3 基于所提算法的物流业务流程图

首先,利用用户相异度矩阵在交易评分数据中获取用户对物流服务的评分数据,然后将该数据进行数据增量模型构建。当用户在平台成功下单后,后台就会利用



个性化自动推荐算法向用户进行合适的物流服务,当用户确认商品签收并对本次物流服务作出评价时,再次采用增量更新的方式,对用户的评分数据进行局部实时更新,该业务流程如图3所示。

### 4.3 电商物流服务数据分析

物流服务的流程分析是为了反映出数据在电子商务后台中的流动和处理情况。通常是利用流程图进行描述,如图4所示。

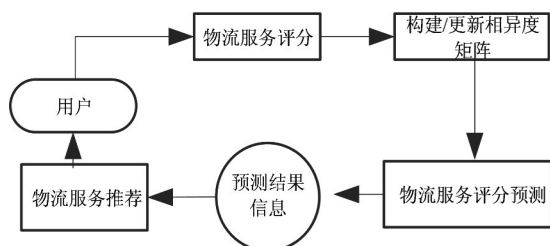


图4 物流服务数据流程图

从图4中可以看出,数据流程剔除具体的组织结构、人员等物理要素。可以生动描述数据的流动与处理等使用情况。

## 4.4 构建电商物流个性化自动推荐服务系统

### 4.4.1 系统功能分类模块

经过上述流程分析可知,要想构建电商物流个性化自动推荐服务系统必须要具备的系统的主要功能模块有用户行为记录模块、增量更新模型管理模块和推荐算法模块,如图5所示。

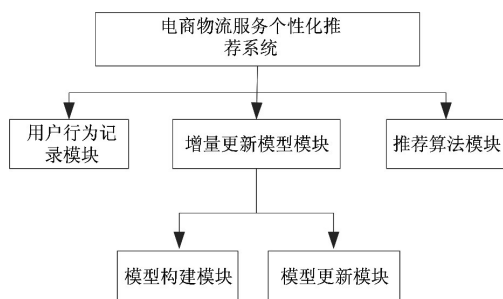


图5 个性化自动推荐系统功能模块结构图

(1) 用户行为记录模块:主要管理用户对物流服务的评分数据等行为。用于推荐模型中的数据构建及更新;

(2) 增量更新模型模块:主要用于系统初始阶段运行,当模型利用相异度矩阵构建完成后就可以利用更新模块对产生的新评分数据进行评分向量并进行局部更新;

(3) 推荐算法模块:主要采用IU-UserCF个性化自动推荐算法对物流服务进行评分计算,将符合要求的物流服务推荐给用户。

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

### 4.4.2 实现自动化推荐系统应用

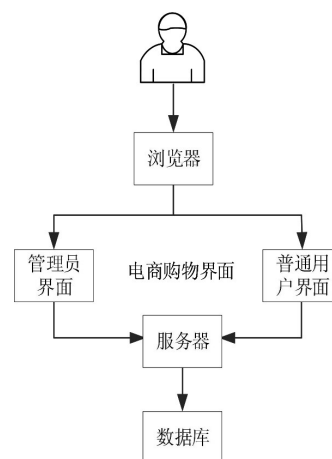


图6 电商物流个性化自动推荐服务系统

由图6所示,系统的界面主要集合在电商购物界面,其中包含了普通用户界面和管理员界面。普通用户界面主要在用户成功下单后自动推荐符合条件的物流服务;管理员界面主要向管理员提供推荐系统数据的维护功能。然后将服务器设置在云平台上,利用数据增量更新模型对海量数据进行局部在线更新。

## 5 实验分析

为了验证所提方法的有效性,分别将文献[3]、文献[4]与所提方法的适用性、应用效率和预测精准度进行对比。

实验采用 Windows7 系统;SQLserver2008 数据库;数据集选取 MovieLens100K。其数据集的稀疏度计算方法如公式(9)所示:

$$1 - \frac{60000}{(821 \times 1256)} = 0.9418 \quad (9)$$

### 5.1 适用性

为了验证所提方法具有更优秀的适用性,实验选取相同的出现次数,利用三种方法对不同的用户数据集进行匹配度对比,结果如表1所示。

表1 个性化自动推荐算法应用匹配度对比结果

用户数据集	出现次数	匹配度%		
		文献[3]	文献[4]	所提方法
数据集1	10	51.2%	73.4%	88.6%
数据集2	10	62.1%	85.2%	89.4%
数据集3	10	59.5%	76.2%	91.3%

从表1对比结果可以看出,文献[3]方法平均匹配度为64.3%,可以有效地向用户提供个性化自动推荐服务;文献[4]方法推荐匹配度高于文献[3],平均匹配度为78.3%;

而所提方法在提供个性化自动推荐时,利用了数据增量更新模型对用户的相异度进行了评估,计算出用户的偏好。因此,该方法平均匹配度高达90.4%,明显高于其他两种方法,证明了所提方法在实践中拥有更好的适用性。

## 5.2 应用效率

在个性化自动推荐过程中,如果没有及时筛选符合用户条件的项目,就会大大影响用户的体验感。

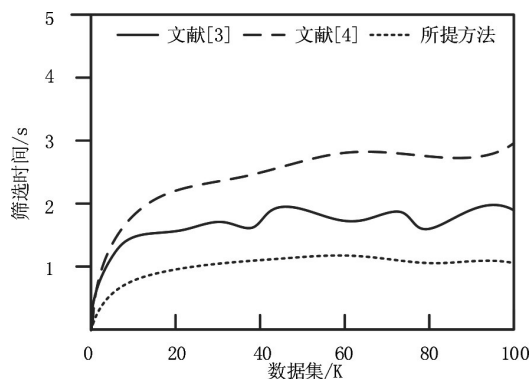


图7 项目筛选时间对比结果

图7是三种方法为项目筛选的时间对比结果,由图可知,文献[3]方法的筛选时间虽然用时较短,但未使用增量更新模型对网络中的大数据进行局部更新,所以稳定性较差;文献[4]方法筛选时间随着数据集的增加也不断增加,所以在数据量较大的情况下,该方法不能快速筛选出符合用户条件的项目;而所提方法使用增量更新模型,引入了IU-UserCF的个性化自动推荐算法,以相异度较小的用户作为近邻,并根据其评分数据对用户进行推荐,就算在数据量较大时也可以快速筛选出符合用户条件的项目。实验证明,该方法在缩短了筛选时间的同时也拥有良好的数据处理能力与稳定性。

## 5.3 预测精准度

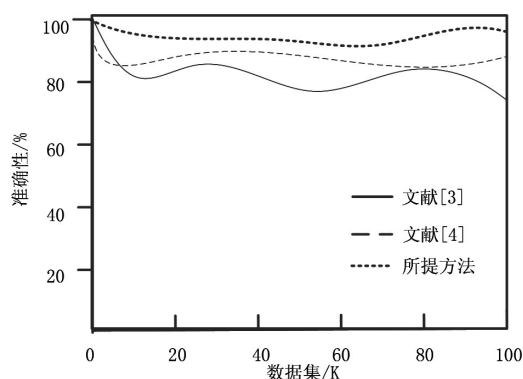


图8 项目预测精准度对比结果

由图8的实验结果可以看出,文献[3]方法因为没有对用户的相异度进行计算,所以随着数据集的增加准确性

呈现逐渐下降的趋势;文献[4]方法虽然准确性优于文献[3]方法,但随着数据集的增加准确性也呈现逐渐下降的趋势;然而所提方法构建了数据增量更新模型,将有变化的项目评分数据通过模型进行局部在线更新。虽然准确性随着数据集的增加有所下降,但仍保持在90%以上。以上结果表明,该方法不仅可以对数据的增量进行实时更新,而且具备准确性非常高的推荐预测。

## 6 结束语

针对当前个性化自动推荐算法在电子商业中应用的不足,提出所提方法。通过对用户和项目的相异度进行计算,得到相异度矩阵,构建了数据增量更新模型。当数据量较大时,利用该模型将增加的数据实时更新,提高数据的处理能力。基于此,引入IU-UserCF根据邻居评分数据向用户提供更加符合条件的项目,提高推荐算法的应用效率。并以电商物流为推荐对象,构建了更全面化的个性化自动推荐系统。实验对比结果表明,所提方法在实践中有着更加出色的推荐效果。但是该方法只是借鉴了大数据中的显性的评分数据,而且用户的偏好可能随着时间的变化而变化,因此在接下来的研究中,会结合大数据中的隐式数据构建实时性更强的个性化自动推荐。

## 参考文献:

- [1] 司梦楚,季同同,张春明.服装智能推荐系统在电商平台中的应用[J].江南大学学报(自然科学版),2019,4(6):498-503.
- [2] 邹坤.无线网络的电子商务商品自动推荐系统[J].现代电子技术,2020,43(7):82-85.
- [3] Guan Y, Wei Q, Chen G. Deep learning based personalized recommendation with multi-view information integration[J]. Decision Support Systems, 2019, 118(5): 58-69.
- [4] 李宇琦,陈维政,闫宏飞,等.基于网络表示学习的个性化商品推荐[J].计算机学报,2019,15(8):1767-1778.
- [5] 耿秀丽,潘亚虹.考虑用户体验的产品服务系统模块重要度判定方法[J].计算机集成制造系统,2020,265(5):149-157.
- [6] 秦星红,苏强,洪志生.考虑顾客期望与质量成本的网购物物流服务供应链的竞争合作策略研究[J].管理工程学报,2019,33(3):136-146.
- [7] 张瑞典,钱晓东.用余弦相似度修正评分的协同过滤推荐算法[J].计算机工程与科学,2020,306(6):149-158.
- [8] 王宇琛,王宝亮,侯永宏.融合协同过滤与上下文信息的Bandits推荐算法[J].计算机科学与探索,2019,13(3):361-373.

作者简介:李加军(1982-),男,硕士,讲师,研究方向:电子商务、大