

# 服装智能推荐系统在电子商务平台中的应用

司梦楚, 张春明\*

(青岛大学 纺织服装学院, 山东 青岛 266071)

**摘要:**近十年来,中国逐渐进入全民网购的电子商务时代,越来越多的电商平台和产品领域引用推荐系统提高交易效率。服装作为电商品类的一大分支,其推荐系统也展现出巨大的商业潜力和应用前景。为进一步探索电子商务中服装类商品的个性化推荐机制,本文结合产品属性对服装推荐算法及其应用进行研究,着重阐释三种推荐技术的运算步骤及特征,并总结各个算法的优缺点以及目前服装推荐系统研究中存在的不足,为电商平台的服装类产品及品牌营销提供理论与智能算法依据。

**关键词:**推荐系统;个性化推荐;服装;电子商务;网络营销;协同过滤

中图分类号:TS941.2

文献标识码:A

文章编号:2095-414X(2019)02-0033-06

随着互联网的高速发展、电子商务与物流产业的日益成熟,我国电商行业的市场规模也发生了天翻地覆的变化。早在2010年,淘宝网的在线商品数已达到8亿。同年首届双11,在当时网购蔚然成风的情况下,仍创造了一天5000万的销售额。根据近五年中国电子商务研究中心监测数据显示,2012到2017年中国电子商务市场交易规模翻了近4倍(见图1),大多数人对网购已经习以为常。中国网络购物用户规模也呈直线式增长(如图2),5年内网购用户规模翻了一倍,即在中国接近一半人口会使用网购。根据表1显示的各年龄消费者网络购物喜好排行,不难发现服装类产品是网购人群的重要占比类。值得注意的是,服装品类占据了电商B2C销售规模的35%,网购渗透率达到36.9%。

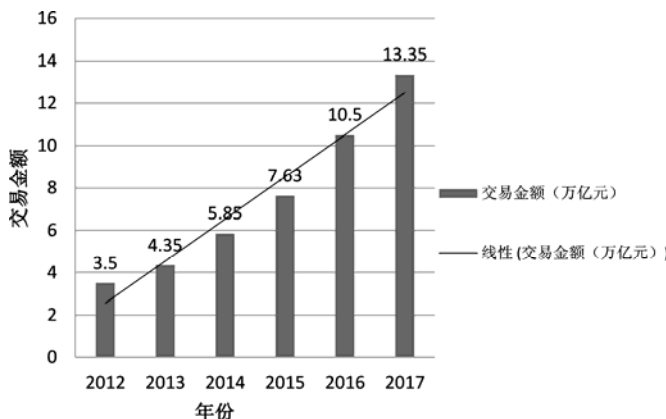


图1 2012-2017年中国电子商务市场交易规模

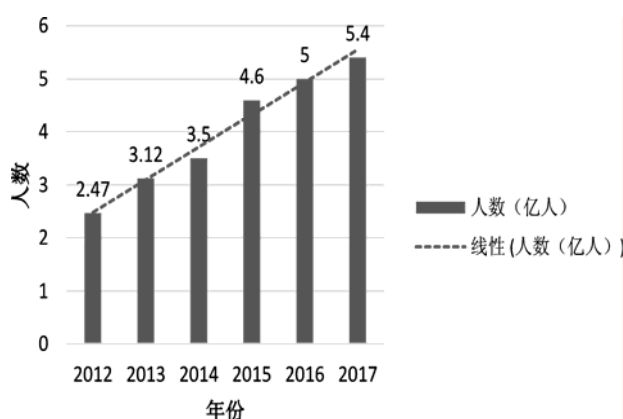


图2 2012-2017年中国网络购物用户规模

商品种类与数量爆炸式增长,可想而知消费者在选购自己心仪商品时会面临信息超载的难题。对于类似C2C、B2C电子商务网站来说,消费者想要从海量信息中筛选出自己想要的商品信息不是易事。但对于B2B、C2M类似的电子商务平台来说,生产商和经营方让自己厂商的产品引起注意更加困难<sup>[2]</sup>。“人找信息”的服务模式已经不适用于信息过剩的现代社会,急需一种依据用户本人信息与历史数据而自动组织和筛选服务模式。推荐系统应运而生,它的做法不止是把消费者作为网站内容的浏览者,同时更是网站内容的制造者。目前,像淘宝、京东这类电商平台已经应用了简单的个性化推荐功能,主要是为了实现商家营销目

\*通讯作者:张春明(1981-),男,副教授,博士,硕士生导师,研究方向:服装文化与品牌营销。

基金项目:中国博士后科学基金项目(2018M632624);中国纺织工业联合会高等教育教学改革项目(2017BKJGX102);青岛大学教学研究与改革项目(2018)。

的,从消费者自身出发的个性化推荐应用少。同时,推荐算法在服装电商行业的应用规模还不成熟。

表1 各年龄消费者网络购物喜好排行<sup>[1]</sup>

喜好排行	90后男性	90后女性	80后男性	80后女性	70后男性
第1位	服装	服装	手机(含配件)	服装	家用电器
第2位	手机(含配件)	鞋/包	电脑硬件/周边	旅游休闲服务	电脑硬件/周边
第3位	电脑硬件/周边	书籍	旅游休闲服务	护理彩妆	旅游休闲服务

## 1 服装个性化推荐系统简介

服装推荐系统是个性化推荐系统应用在服装行业的个例,利用数据挖掘等技术建立服装数据模型和客户模型,经过数据分析,借助电子商务平台适时的将与客户兴趣特点匹配的服装进行展示,帮助顾客完成服装购买过程的协同机制<sup>[3]</sup>。

服装电商领域的商品推荐主要包括:新款服装面向大众化的推荐、热卖服装,爆款服装推荐、商家换季促销特价服装推荐和基于用户浏览记录,历史购买(评分)记录和喜好收藏商品的个性化推荐等。服装作为电商销售平台的销售大类,其推荐系统在推荐原理上与其他商品的推荐类似,但其本身的特殊性,也面临着诸多挑战,具体表现为:

(1) 用户模型不稳定,不易建模。消费者对于服装的喜好会受流行趋势、地域文化、季节变化的影响,甚至是受某明星、某文化的一时影响。并且随着现代生活的丰富,消费者也经常购买功能性的服装。这使得系统很难捕捉用户兴趣,建立的用户模型准确度低。

(2) 与对图书电影、日用消费品和电子类产品的推荐不同,消费者在选购服装产品时会存在求异心理,因此销量好不能等同于值得推荐。服装商品的推荐既要符合流行趋势迎合人们的趋同心理,又要兼顾消费者张扬个性的求异心理。

(3) 推荐系统运算量巨大。从以上两方面不难看出,服装推荐系统运行数据繁多,数据更新快,计算量大,会导致系统维护成本高。这也是限制一些中小型企业和网站投入运行该系统的重要原因<sup>[4]</sup>。

在针对服装的推荐算法上,国外的研究要早于国内。Kudo等通过分析消费者的数据制定用户对服装属性的决策信息表,并将粗糙集理论应用到服装推荐中(2006)<sup>[5]</sup>。Lin等在原有研究的基础上,提出除了用户偏好之外的需求因素的用户模型(2012)<sup>[6]</sup>。Choi等认为隐式评级可以成功地替代协同过滤的显式评级,而协同过滤和序列模式分析(SPA)的混合方法比单独的评级要好,可为用户建立更准确的偏好模型(2012)<sup>[7]</sup>。Geuens等提出了一个新的决策支持框架,根据在线二进制购买数据(稀疏级、采购分配和项目—用户比率)生成建议,以帮助电商公司选择最佳的协同过滤算法(CF)(2018)<sup>[8]</sup>。

国内对于服装个性化推荐建模问题研究较晚,主要是基于算法的改进。吴健春,艾黎,徐兵兵均是用聚类的方法缩小协同过滤时用户群的数量、从而在计算时降低属性空间的维度,来解决信息过载问题<sup>[9-11]</sup>;齐扬关注了服装商品的特殊性,将西服的主要特征分为领型、扣子、面料等不同要素,并对这些特征要素抽取分类,最后用BDEU决策树算法建立用户类别偏好模型(2010)<sup>[12]</sup>;郑充林在协同过滤算法中介入了产品生命周期理论,新算法考虑了时间信息影响的问题,提出了在推荐结果过程中加入由项目生命周期决定的权重信息,克服了传统算法的弊端(2013)<sup>[3]</sup>;陈丹儿等提出的协同过滤改进算法首先将用户历史评分数据映射为用户偏好,然后使用BP神经网络建立偏好模型并预测新项目的评分,形成最近邻后产生推荐(2015)<sup>[13]</sup>;单毓馥等提出的服装推荐方法基于AHP层次分析,该方法量化影响顾客选择主要因素的权重,通过相似顾客的评分来预测其他顾客对该服装的喜好(2016)<sup>[4]</sup>;陈俊虹等调研用户搜索时使用的关键词,开创性构建了标签体系,从而提高用户在购买服装时的精确度(2017)<sup>[2]</sup>;

针对电子商务中服装类商品的个性化推荐问题,国内外学者进行了一些研究,其成果多为根据客户的浏览、点击、收藏行为等进行聚类从而实现产品推荐。然而,结合服装产品特点的智能推荐系统或算法目前却鲜有文献报道。本文以电商平台的服装智能推荐系统为研究对象,结合三种推荐方式以及服装产品种类、款式等提出了新的推荐算法,并对其运算步骤及特征进行了阐释。

## 2 推荐算法在服装智能推荐中的应用

推荐系统除了解决信息过载的问题, 更重要的是激励用户去购买新商品。针对于服装这一类商品, 目前主流采用的算法有以下几种: 基于内容(Content-based, 简称CB)、协同过滤(Collaborative Filtering, 简称CF)和混合推荐。

### 2.1 基于内容的推荐算法

基于内容的推荐不依赖用户对项目的评价, 利用历史浏览、点击、分享、加购和已购买这些项目数据来建立用户偏好模型, 再提取项目的内容属性形成特征矩阵, 最后对比分析用户模型与特征矩阵的吻合程度, 将相关系数高的项目推荐给用户。因此随着用户的浏览记录数据的增加, 这种推荐一般也会越来越准确。项目内容特征提取、用户偏好数据模型建立和相似度匹配是该推荐模型的三大关键。项目特征的表示一般采用向量空间模型, 最常用的是TF-IDF<sup>[14]</sup> (Term Frequency-Inverse Document Frequency) 方法, 其为项目特征分配权重的思想如下:

设项目集为N, 采用k维向量  $\vec{d}_c = (w_{1c}, w_{2c}, \dots, w_{kc})$  和  $\vec{d}_j = (w_{1j}, w_{2j}, \dots, w_{kj})$  分别表示用户c的偏好文档  $\vec{d}_c$  和项目文档  $\vec{d}_j$ , 其中k代表关键词的个数, 表示项目集中包含关键字的项目数,  $f_{ij}$  表示关键词  $k_i$  在文件  $d_j$  中出现的次数。

关键词  $k_i$  在文档  $d_j$  中出现的频率  $TF_{ij}$  定义为:

$$TF_{ij} = \frac{f_{ij}}{\max_j f_{ij}} \tag{1}$$

关键词  $k_i$  在项目集中出现的逆向文档词频  $IDF_i$  定义为:

$$IDF_i = \log \frac{N}{n_i} \tag{2}$$

关键词  $i$  在项目文件  $j$  中的权重为:

$$w_{ij} = TF_{ij} \cdot IDF_i = \frac{f_{ij}}{\max_j f_{ij}} \cdot \log \frac{N}{n_i} \tag{3}$$

用户偏好文档和项目文档的余弦相似度为:

$$\text{sim}(c, d_j) = \cos(d_c, d_j) = \frac{\sum_{i=1}^k w_{ic} w_{ij}}{\sqrt{\sum_{i=1}^k w_{ic}^2} \cdot \sqrt{\sum_{i=1}^k w_{ij}^2}} \tag{4}$$

利用公式 (3) - (4) 计算偏好文档和项目文档的余弦相似度, 算得的相似度越高, 表示用户对该特征越偏好<sup>[15]</sup>。

比如, 在进行款式推荐时首先根据用户的浏览记录等提取用户的兴趣特点建立用户模型, 然后在服装数字模型中寻找与用户模型匹配度较高的服装进行推荐, 其具体的推荐机制的基本原理如表2所示。假设有三种类型的用户A、B、C, 表格中的圆点表示该用户喜欢购买该款式风格的服装。通过匹配用户兴趣模型以及款式的特征矩阵, 系统对款式A的特征提取和款式C的类似。对号表示生成推荐, 根据基于内容的推荐思想, 系统会把款式C推荐给用户A。

表2 基于内容的推荐原理

用户/项目 (款式风格)	款式 A	款式 B	款式 C
	H 型、通勤、职业	A 型、活泼、可爱	H 型、通勤、职业
用户 A	•		✓
用户 B		•	
用户 C		•	•

基于内容的推荐有优点, 如: 没有冷启动问题和稀疏问题; 实现原理简单、直观, 结果易于理解, 不需要具备专业行业知识<sup>[15]</sup>; 可以发现隐藏的“暗信息”, 从而推荐新出现和非流行的项目资源。但也有缺点, 只能适用于资源内容比较容易分析的系统, 对于抽象的内容 (如视频、音乐等) 会由于缺乏特征提取方法

而无法实施；推荐资源的范围也会过于狭窄，只能根据用户历史偏好进行推荐，不能很好地发现用户的新兴趣。

## 2.2 协同过滤推荐算法

协同过滤推荐算法是应用最广泛也是最成功的算法，它的原理是基于这样的假设：生活中当我们接触到新问题或购买新产品时，总会向身边的朋友咨询建议。它的算法基于三点：（1）根据用户历史行为数据分析用户喜好；（2）将具有相似喜好的用户归为一类；（3）综合同类用户对某项目的评分，预测目标用户对该项目的喜好程度。对同一用户群的用户推荐相似的服装时，先根据用户一项目评分矩阵中的用户数据提取相似用户，进而产生推荐。协同过滤推荐算法分为基于用户（User-Based）的协同过滤推荐和基于项目（Item-Based）的协同过滤推荐两种类型。

### 2.2.1 基于用户的协同过滤推荐算法

基于用户的协同过滤推荐算法强调了人与人之间的合作，每个人的行为属性都在某个特定群体的范围之内，相同群体的用户兴趣偏好相似。首先根据用户对项目的评价寻找特定群体范围内的最近邻居集合，根据最近邻居集合对某一项目的评分预测新用户对该项目的喜好。

分为相似度计算、临近选择和产生推荐三步。修正的余弦相似性是衡量用户间相似性的较好方法，公式见式（5）。

$$\text{sim}(U_a, U_b) = \frac{\sum_{k=1}^t (R_{a,k} - \bar{R}_a) \times (R_{b,k} - \bar{R}_b)}{\sqrt{\sum_{k=1}^t (R_{a,k} - \bar{R}_a)^2} \times \sqrt{\sum_{k=1}^t (R_{b,k} - \bar{R}_b)^2}} \quad (5)$$

$R_{a,k}$  和  $R_{b,k}$  分别表示用户  $U_a$  和用户  $U_b$  对项目  $k$  的评分， $\bar{R}_a$  和  $\bar{R}_b$  分别表示用户  $a$  和用户  $b$  的项目评分平均分。最后的计算结果越大，说明用户  $a$  和用户  $b$  的相似性越大。

然后选择最近邻居，用式（5）算得相似性后，选择相似性最大的前  $N$  个用户作为目标用户的最近邻  $S(u)$  集合  $S(u) = \{u_1, u_2, \dots, u_{n-1}, u_n\}$ 。

最后产生推荐，用户  $a$  对项目  $i$  的预测评分  $P_{a,i}$  可用公式（6）表示。

$$P_{a,i} = \bar{R}_a + \frac{\sum_{b \in S(u)} \text{sim}(a,b) \times (R_{b,i} - \bar{R}_b)}{\sum_{b \in S(u)} \text{sim}(a,b)} \quad (6)$$

其中  $\text{sim}(a,b)$  表示用户  $a$  和用户  $b$  的相似性， $R_{b,i}$  表示用户  $b$  对项目  $i$  的评分， $\bar{R}_a$  和  $\bar{R}_b$  分别表示用户  $a$  和用户  $b$  的评分平均分<sup>[6]</sup>。

经过上面的计算，就可以得到目标用户对所以未评分项目的评分，选择评分最大的前  $k$  个项目推荐给用户。

基于用户的协同过滤算法的基本原理如表3所示。假设有三种类型的用户A、B、C，表格中的圆点表示该用户喜欢购买该类型的服装。从表格信息可以推断出，用户A和用户C的兴趣大致类似，他们都喜欢购买运动装和工装。表格中的对号表示将西装推荐给用户A，因为用户A和用户C兴趣类似，用户C喜欢西装，则用户A也可能喜欢西装。

表3 基于用户的协同过滤推荐

用户/项目（种类）	运动装	休闲装	工装	西装
用户 A	•		•	✓
用户 B		•		
用户 C	•		•	•

### 2.2.2 基于项目的协同过滤推荐算法

基于项目的协同过滤推荐算法将项目的特征属性放在特定项目的范围内，假设用户对某一项目有兴

趣, 那与之相似的其他项目也可能引起用户兴趣。根据用户一项目评分矩阵中的服装数据来计算服装之间的相似性, 向顾客提供与以前喜欢的服装相似的服装。大体思想与基于用户的推荐类似, 唯一区别在于提取的是项目的相似度。

基于项目的协同过滤算法也可分为三步, 不同的是计算项目的相似性, 计算公式与式 (5) 类似, 不再赘述。计算完项目的相似性后, 选取相似性最大的前  $N$  个项目作为最近邻集合  $S(i) = \{i_1, i_2, \dots, i_{n-1}, i_n\}$ 。最后一步就是预测目标用户对项目的评分。评分预测计算如公式 (7) 所示。

$$P_{u, i} = \bar{R}_i + \frac{\sum_{j \in S(i)} sim(i, j) \times (R_{u, j} - \bar{R}_j)}{\sum_{j \in S(i)} sim(i, j)} \quad (7)$$

$\bar{R}_i$  和  $\bar{R}_j$  分别表示项目  $i$  和项目  $j$  获得项目评分的平均分,  $sim(i, j)$  表示项目  $i$  和项目的相似性,  $R_{u, j}$  表示用户  $u$  对项目  $j$  的评分, 最后选取评分最高的前  $K$  个项目推荐给目标用户<sup>[17]</sup>。

该算法思想依据假设某用户对某项目产生兴趣, 那么用户也会对与之相似的其他项目产生兴趣, 其基本原理如表 4 所示。表示假设有三种类型的用户 A、B、C, 表格中的原点表示该用户喜欢购买该类型的服装。对于用户 C, 根据所有用户对某种款式的偏好情况, 判断出喜欢款式 1 的用户都喜欢款式 3, 得出款式 1 和 3 相似, 继而推断出用户 C 可能对款式 3 感兴趣。对号✓表示系统会将款式 3 推荐给用户 C。

表 4 基于项目的协同过滤推荐

用户/项目 (款式)	款式 1	款式 2	款式 3
用户 A	•		•
用户 B	•	•	•
用户 C	•		✓

协同过滤系统具有以下优点: 可以对内容识别困难的商品进行项目分析, 比如音乐、电影等; 有挖掘消费者潜在兴趣和偏好的能力, 新用户的个性化学习过程得以加快。但同时也有难以克服的数据稀疏性问题, 以及项目、用户的冷启动问题。这两大问题使得相似度计算的结果与实际相差甚远, 导致推荐质量的下降。随着用户与项目数量的剧增, 传统的算法也将遭遇到严重的扩展性问题。

### 2.3 混合推荐算法

表 5 混合推荐的七种类型<sup>[18]</sup>

类型	处理方式
加权(Weight)	对多种推荐模型进行加权处理。
变换(Switch)	采取不同的推荐方法适应当前情况要求和问题背景。
多项罗列(Mixed)	将多种推荐模型的推荐列表呈现给消费者选择, 不做过多处理。
特征组合(Feature combination)	最后的推荐算法的特征输入采用前面不同推荐数据集的特征。
层叠(Cascade)	首先通过一种推荐模型产生一个初始的推荐列表, 在此基础上, 用第二种推荐模型进一步输出精确度更高的推荐列表。
特征扩充(Feature augmentation)	一种推荐模型计算的的特征向量融入到另一种推荐模型中的特征。
元级别(Meta. level)	将其中一种推荐模型的输出作为初始数据输入到另一种推荐模型中。

每个算法都有一个相对“致命”的缺点, 混合算法的原理是采取组合策略, 使不同的推荐算法“扬长避短”。混合算法最开始是作为一种新思想提出的, 近几年有很多学者针对这一问题进行了研究, 把理论变为实践, 其中应用最多的是把基于内容的推荐与协同过滤算法结合。现实生活中经常会用到组合推荐的形式, 在组合方式上有加权, 变换, 混合, 特征组合, 层叠, 特征扩充和元级别 (见表 5) 7 种组合形式。现阶段, 很多针对于推荐算法提高精确度的新突破都是基于混合这一思想。

## 3 结语

个性化推荐系统目前在用户偏好较稳定的图书、音乐、电影等产品领域的应用较为成熟, 但服装类产

品的智能推荐仍处于摸索阶段。导致服装推荐系统运算数据巨大且准确性不高的原因主要有以下几点:(1) 消费客体易受适用季节、流行生命周期、使用情境等因素的影响;(2) 消费主体购买服装时的动机、态度、价值观、生活方式等心理变量因素较为复杂;(3) 服装类产品的分类和评价方式更加多元化。

本研究梳理总结了了几种常见推荐算法在电商平台服装类产品智能推荐中的应用,着重阐释了三种推荐算法的运行原理,对协同过滤算法进行了优化,计算用户之间的相似度后,采取聚类分析,避免了大量运算,为电商平台的服装类产品及品牌营销提供了理论与智能算法依据。基于内容的推荐没有冷启动的问题,但无法建立抽象内容的特征矩阵,不能发掘用户潜在新兴趣;作为主流算法的协同过滤算法,具备推荐新信息,产生新奇推荐的能力。在资源内容难以分析的情况下,协同过滤是很好的选择。但是当用户评分信息不足10%时,会导致推荐算法性能的下降。还有一种极端的情况,新项目和新用户都面临着第一评价的问题(即冷启动);混合推荐是一种新思想,并没有具体的算法流程,目的是为了避开和改善单一算法的缺点,灵活融合,提高推荐效率。

### 参考文献:

- [1] 唯衣. 2017年消费者网购行为分析,服装稳居首位[EB/OL]. [http://m.sohu.com/a/155530858\\_154452](http://m.sohu.com/a/155530858_154452). 2017-7-8.
- [2] 陈俊虹,刘咏梅. 基于标签的商务女装推荐模型研究[J]. 浙江纺织服装职业技术学院报, 2017, (4): 68-72.
- [3] 郑充林. 协同过滤的服装推荐算法的改进研究[D]. 上海: 东华大学, 2013. 1-5.
- [4] 单毓馥,李丙洋. 电子商务推荐系统中服装推荐问题研究[J]. 毛纺科技, 2016, (5): 66-69.
- [5] Kudo Y, Amano S, Seino T, et al. A simple recommendationsystem based on rough set theory[J]. *KanseiEngineering International*, 2006, 6(3): 19-24.
- [6] Lin Y C, Kawakita Y, Suzuki E, et al. Personalizedclothing-recommendation system based on a modifiedBayesian network[A]. //2012 IEEE/IPSJ 12th InternationalSymposium on Applications and the Internet[C]. IEEE, 2012. 414-417.
- [7] Keunho Choi, DongheeYoo, Gunwoo Kim, et al. A hybrid online-product recommendation system: Combining implicitrating-based collaborative filtering and sequential pattern analysis[J]. *Electronic commerce research and applications*, 2012, 11(4): 309-317.
- [8] StijinGeuens, KristofCoussement, Koen W, et al. A framework for configuring collaborative filtering-based recommendations derived from purchase data[J]. *European journal of operational research*, 2018, 265(1): 208-218.
- [9] 艾黎. 基于商品属性与用户聚类的个性化服装推荐研究[J]. 现代情报, 2015, (9): 165-170.
- [10] 吴健春. 基于协同过滤的服装电子商务个性化推荐系统研究与开发[D]. 天津: 天津工业大学, 2008. 26-48.
- [11] 徐兵兵. 基于粗集和模糊聚类的协同过滤算法[J]. 浙江纺织服装职业技术学院学报, 2010, (2): 56-60.
- [12] 齐扬,朱欣娟. 基于数据挖掘的服装推荐系统研究[J]. 西安工程大学学报, 2010, (4): 438-443.
- [13] 陈丹儿,应玉龙. 基于项目属性和BP神经网络的协同过滤推荐[J]. 信息技术, 2015, (3): 70-73.
- [14] 周军锋,汤显,郭景峰. 一种优化的协同过滤算法[J]. 计算机研究与发展, 2005. 41(10): 27-35.
- [15] 何佳知. 基于内容和协同过滤的混合算法在推荐系统中的应用研究[D]. 上海: 东华大学, 2016. 8-12.
- [16] 汪岭. 基于Hadoop的电子商务推荐系统应用研究[D]. 广州: 广东工业大学, 2015. 8-11.
- [17] 刘建国,周涛,汪秉宏. 个性化推荐系统的研究进展[J]. 自然科学进展, 2009, (1): 2-7.
- [18] 邓仙荣. 基于梯度提升回归算法的O2O推荐模型研究[D]. 淮南: 安徽理工大学, 2016. 18-19.

## Application of Clothing Intelligent Recommendation System in E-commerce Platform

SI Meng-chu, ZHANG Chun-ming

(College of Textiles and Clothing, Qingdao University, Qingdao Shandong 266071, China)

**Abstract:** In order to further explore the personalized recommendation mechanism of clothing products in e-commerce, this paper combines product attributes to study the clothing recommendation algorithm and its application, focuses on the operation steps and characteristics of the three recommended technologies. And at last it has summarized the advantages and disadvantages of each algorithm and the shortcomings of the current clothing recommendation system, and it provides theoretical and intelligent algorithm basis for clothing products and brand marketing of e-commerce platform.

**Key words:** recommendation system; personalized recommendation; clothing; e-commerce; e-marketing; collaborative filtering