基于用户浏览轨迹的商品推荐

郭俊霞 许文生 卢 罡
(北京化工大学信息科学与技术学院 北京 100029)

摘 要 随着电子商务的迅速发展,推荐系统在这些网站中得到了广泛的应用。目前应用最广泛的个性化推荐算法是协同过滤推荐算法,但是该方法存在稀疏矩阵与冷启动问题。根据用户浏览记录推荐商品是缓解这些问题的一个重要研究方向,这些方法根据用户在电子商务网站的访问日志,提取出用户的浏览路径序列,即用户浏览轨迹,为用户推荐偏爱商品。目前,通过分析用户浏览路径为用户推荐商品的方法主要依据用户浏览轨迹模式匹配或者从用户浏览轨迹中商品与下一个商品关系的角度进行考虑。而本研究从浏览轨迹中被浏览商品与最终被购买商品关系的角度出发,并以此为基础建立用户浏览轨迹偏爱模型,挖掘用户偏爱,为用户推荐商品。实验表明,所提方法能够在一定程度上解决因为新用户缺少历史购买及评分记录而引起的新用户冷启动问题,提高了推荐方法的准确度与召回率。

关键词 个性化推荐,浏览轨迹,冷启动
中图法分类号 TP274 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.12.041

Recommending Commodities Based on User-browsing Tracks

GUO Jun-xia XU Wen-sheng LU Gang
(College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract With the rapid development of E-commerce, recommendation system has been widely used in the Websites. Currently the collaborative filtering recommendation algorithm is the most widely used, however, this kind of methods has sparse matrix and cold-start problems. In order to solve or at least improve these problems, methods based on users' browsing records were proposed. These methods extract every user's browsing path sequence called user browsing tracks from the users' access log, and then recommend preference commodities for the user based on the analyzing result of the tracks. By now, the most methods that recommend commodities for users through analysis browsing path are based on sequence pattern matching or the view of the relationship between commodity and the next browsed commodity. We considered from the view of the relationship between browsing commodities and eventually bought commodities, establishing the user browsing tracks preference model based on this, mining users' preference, and recommending products for new users. Experiments show that our method plays a certain role in solving the problem of cold-start for new users and enhancing the accuracy and recall rate of the recommendation system in E-commerce.

Keywords Personalized recommendation, Browsing tracks, Cold-start

1 引言

随着电子商务的迅速发展,用户数量以及商品种类迅速增长,这使得用户在寻找喜欢的商品时耗费大量的时间与精力。为了帮助用户快速有效地找到喜欢的商品,提高用户体验,电子商务网站通常针对每位用户进行个性化推荐,使用户可以更加快捷地找到自己偏爱的商品,当前的个性化推荐系统大致可以分为以下几种:1) 基于内容的推荐算法[1]。这种方法主要依据商品属性推荐商品,适合于推荐新商品,但不能针对用户进行个性化推荐。2) 基于知识的推荐算法。基于知识的推荐算法需要用到当前用户额外信息和商品专业详细特征。这种方法需要用户对商品特征有着深入的专业理解,同时对用户与系统交互有更高的要求。3) 协同过滤推荐算法[2]。协同过滤算法依据用户购买商品后反馈的用户评分为用户推荐商品,这种方法能够跨越不同种类的商品进行推荐,是目前针对用户个性化推荐使用最广泛的算法。然而这种算法存在冷启动问题[3,4],包括新用户的冷启动问题、新商品的冷启动问题,以及新用户新商品的冷启动问题,导致对于新用户或新商品无法进行有效推荐。

在电子商务中,用户已浏览的商品在一定程度上反映了用户的爱好,可以用来作为向用户推荐商品的重要参考依据。例如,文献[5,6]对用户浏览的商品、频度、内容以及轨迹等行为进行挖掘,发现用户行为中包含用户爱好的模式,从而获得用户偏爱,提取用户的兴趣特征,再根据用户的兴趣特征为用户提供个性化推荐服务。这种方法由于不依赖于用户的评分记录以及用户和商品的特征信息,对于解决传统推荐方法中的冷启动问题具有重要意义。

本文提出了一个根据用户浏览轨迹向用户推荐商品的方

到稿日期:2016-01-20 返修日期:2016-04-25 本文受中央高校基本科研业务费(YS1404)资助。
郭俊霞(1977—),女,博士,讲师,CCF会员,主要研究方向为广义信息提取、信息聚合工具、Web用户行为分析等;许文生(1989—),男,硕士生,主要研究方向为商品推荐,E-mail:xuwenshg@163.com;卢 罡(1981—),男,博士,讲师,主要研究方向为复杂网络与社会计算、高性能计算
大型复杂网络技术资助。

法。该方法依据某电商网站的浏览日志挖掘用户浏览记录与用户偏爱的关系，建立浏览轨迹-用户偏爱模型，并根据此模型向用户推荐其可能喜欢的商品。

本文第2节介绍目前关于解决冷启动问题的研究进展；第3节介绍所提出的方法模型以及算法的详细设计；第4节通过实验来评估所提算法的精确度、召回率以及算法的效率。

2 相关研究

目前，已经有一些用来解决协同过滤算法冷启动问题的方法被提出，例如：众数法、信息熵法[7]、相似度改良法[8-10]等。文献[8]使用隐马尔可夫模型(HMM)代替简单的相似模型来度量用户相似性，提高了最近邻推荐的准确性，解决了实时性推荐和数据空间的可扩展问题。基于计算商品语义相似度的协同过滤推荐算法[9]，改进了商品相似度计算的精确度，并减轻了商品的冷启动问题。基于内容的社会标签推荐方法[10]提出了一种新的基于内容的社会标签排序方法，用来推荐描述中不包含的语义标记。该方法通过实证获得被量化的词与词之间的语义关系，构造加权的tag-有向图，执行这个改进的基于图的排名算法来完善推荐的每个候选标记评分。这些方法使用一般用户的评分来预测新用户或者对新商品的评分信息，在很大程度上以牺牲用户个性化需求为代价，因此预测评分的准确度也不是很高。

还有一些文献提出了使用多种推荐方法相混合[11,12]或概率统计模型[13]。文献[11,12]综合使用协同过滤和基于内容的推荐方法，在原始评分矩阵基础之上，结合用户特征或商品信息等内容，集成协同过滤方法和基于内容推荐方法的优点向用户推荐商品。文献[13]将用户或项目内容信息初始化为一个概率分布，代替协同过滤推荐方法中的评分概率分布，然后使用Hofmann的EM迭代算法完成推荐。这些方法在一定程度上解决了协同过滤方法存在的冷启动问题，使得推荐效果有所提高，但是也使得推荐方法更加复杂且对用户或商品内容信息有更高的要求。

另外，还有方法使用协同过滤与机器学习相结合的方法，减少了评分矩阵的稀疏性，从而改善了冷启动问题。文献[14]提出了一种基于协作马尔科夫模型的用户行为分析系统和分布式数据处理方法。该方法首先根据用户浏览轨迹对商品进行聚类，然后运用Slope One算法为用户推荐相应商品。文献[15]提出了用户关系与偏爱聚类算法，该方法通过商品聚类压缩稀疏矩阵来减少矩阵维度，然后结合聚类结果和社交网络中的用户标签挖掘用户偏爱信息，从而在缺乏用户反馈记录或评分的情况下，仍然可以向用户提供个性化推荐。文献[16]提出了一种使用关联规则和聚类技术来解决冷启动问题的方法，该方法使用关联规则技术处理用户稀疏问题，对新商品进行聚类，使得用户对商品的稀疏评分转化为用户对商品类的较为稠密的评分。文献[17]提出了一种双聚类与融合的方法(Bi-cluster and Fusion，BiFu)。该方法使用的双聚类方法减少了评分矩阵的维度，采用平滑与融合技术，克服了数据稀疏与评分多样性问题。文献[18]通过注入有限数目的智能体，模拟学习新用户和新项目的兴趣概况信息，从而对新用户和新项目给予适当的预测评分。这些方法不仅提高了推荐的精确度而且改善了冷启动问题，但是这些算法需要较多的用户偏爱信息以及商品属性信息，所以只适用于特定的数据集。

引入用户浏览行为中的用户浏览轨迹内容信息作为数据依据的方法[19-28]也可以用来改善冷启动问题。这类算法可分为几种：序列模式匹配方法[19-23]、基于蚁群算法的方法[24-26]、基于马尔科夫模型的方法[27,28]。序列模式匹配方法是将用户浏览轨迹与历史浏览轨迹进行模式匹配，向用户直接推荐高度匹配的历史轨迹中的商品，或者将用户浏览轨迹作为用户相似度比较的维度结合到系统过滤等推荐方法中，这种方法最明显的缺点就是时间复杂度较高。结合蚁群算法和用户浏览轨迹的方法，将用户看作蚂蚁，将商品看作食物，将用户选择商品的过程看作蚂蚁觅食的过程，模拟蚁群的行为，为用户推荐最优的路径，直到用户找到偏爱商品。这种方法从用户浏览轨迹中被浏览商品之间的关系的角度出发，缩短了用户寻找路径的长度，但不能使用户寻找路径最短。基于马尔科夫的方法依据Markov模型进行推荐，Markov模型包括状态空间和状态转移概率矩阵，构建Markov模型首先需要初始化状态空间，然后估计状态转移概率并填充状态转移概率矩阵，估计状态转移概率使用的方法有最大释然估计、贝叶斯条件概率方法等。这种方法只考虑了路径中相邻的状态对，而忽略了浏览时间和浏览路径中状态的整体顺序性。

本文从用户浏览轨迹中被浏览商品与最终被购买商品间的关系的角度出发，提出一个依据用户浏览轨迹与购买记录为用户推荐商品的方法。该方法不是考虑浏览路径中商品与下一个商品之间的关系，而是从浏览路径中商品与最终被购买商品之间关系的角度进行考虑，结合商品特征属性构建用户浏览轨迹-偏爱推荐模型。该推荐模型不必依赖用户评分数据，而是以用户浏览轨迹和商品本身特征属性作为数据基础，统计用户浏览轨迹中商品与最终被购买商品之间的关系，并依据此模型和当前用户浏览记录向用户推荐商品，以改善新用户冷启动问题。

3 用户浏览轨迹-偏爱模型

3.1 偏爱模型概述

用户在选购商品时会先浏览一些商品，这些被浏览的商品按顺序逐渐趋向于用户想要购买的商品。从用户开始浏览这些商品，直到放弃购买，或者最终购买某一个或者几个商品，形成用户浏览轨迹和购买记录。用户浏览轨迹表示用户在最终购买商品之前浏览过的商品序列，最终购买的商品不作为该序列的最后一项列出，则该序列表示为：$List=\{p_1,p_2,\cdots,p_n\}$。用户浏览轨迹与购买记录定义为：$Track=\langle U,p_{buy},List\rangle$，其中$U$表示一个用户，$p_{buy}$表示用户最后购买的商品。

商品作为向用户推荐的对象，其本身有一些属性，这些属性反映了商品本身所具有的一些特征，本文以标签的形式描述商品的这些属性特征。商品定义为：$p=\langle PID,TAG\rangle$，其中，$TAG=\{tag_1,tag_2,\cdots,tag_n\}$。$PID$表示商品的编号，唯一确定一个商品；$tag_i$表示商品的第$i$个属性标签，表示商品的某个属性特征。在浏览轨迹中，这些特征与最终被购买商品特征的相似程度表现出了某种变化趋势，即商品特征趋势，是用户偏爱的展现形式。

商品之间的相似程度，即商品相似度，从一定程度上反映了商品与商品之间的相关程度。喜欢类似物品的用户可能有相同或者相似的口味和偏好，他们会选择特征相似的商品。在传统协同过滤方法中，推荐系统通过用户对商品的评分相似程度计算商品相似度，在缺乏用户评分的情况下，本文根据商品特征标签属性计算商品相似度。每个商品都包含一些标签，如“健康”、“实用”等，两个商品可能会包含相同标签，那么

这些相同标签就是这两个商品的相似之处，这些相同标签占两个商品所包含所有标签的比值就是两个商品标签属性的相似程度，即商品相似度，用 $Sim(p_i,p_j)$ 表示，计算方法如式(1)所示：

$$Sim(p_i,p_j)=\frac{|Tags(p_i)\cap Tags(p_j)|}{|Tags(p_i)\cup Tags(p_j)|} \quad (1)$$

其中，$Tags(p)$ 表示 p 包含的所有标签。对于 p_i 和 p_j，它们所包含的标签的交集中元素的个数与并集中元素的个数的比值表示 p_i 与 p_j 的相似程度。

基于以上定义，本文提出的用户浏览轨迹-偏爱推荐模型如图 1 所示。模型从网站访问日志中提取出用户浏览轨迹，从被浏览商品与被购买商品关系的角度出发，结合商品浏览轨迹中商品的特征趋势，计算出商品间的购买转移关系，构建用户偏爱推荐模型。当新用户在线浏览商品时，根据购买转移关系，向用户实时推荐商品。

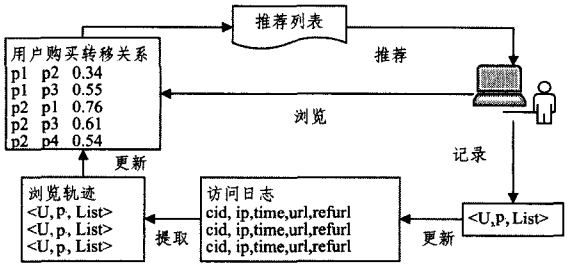

图 1　用户浏览轨迹-偏爱推荐模型

用户浏览轨迹反映了用户偏爱，通过分析用户浏览轨迹中的商品序列得到购买转移概率，即用户在浏览一个商品 p_i 时，最终购买了商品 p_{buy} 的概率。这构成了商品 p_i 和 p_{buy} 之间的购买转移关系，记为 $p_i->p_{buy}$。

模型中用户购买转移关系的计算方法将在 3.2 节详细介绍；使用模型中购买转移关系向用户推荐商品的具体过程将在 3.3 节详细介绍。

3.2　购买转移概率

商品间的购买转移概率 $P(p_i->p_{buy})$ 与 p_i 和 p_{buy} 在浏览轨迹上的距离有关，也与该转移关系出现后存在的时间长短有关。

3.2.1　浏览距离因素

浏览距离是指在浏览轨迹中，从 p_i 到达 p_{buy} 的跳数，用 d 表示。本文通过实验对浏览轨迹中浏览距离与购买概率和商品的特征趋势进行了统计，结果如图 2 所示。

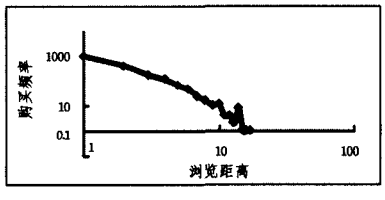

图 2　购买概率与浏览距离统计分布图

结果显示，不同浏览距离的商品被购买的概率近似呈幂律分布。这说明大部分情况下，用户只浏览很少的几个商品就做出了购买的选择，而很少有用户在浏览了许多商品后才最终购买，同时还说明浏览距离越小的商品最终被购买的概率越大。

商品特征趋势如图 3 所示，可以看出，商品特征趋势同样近似呈幂率分布，这说明在浏览轨迹中，商品特征趋势同最终被购买的商品是越来越接近的。

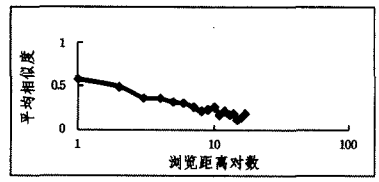

图 3　商品特征变化趋势图

用户浏览轨迹中，购买概率是浏览距离为 k 的商品被购买的条件概率，特征趋势是商品特征与最终被购买商品的相似度变化趋势。购买概率分布和商品特征趋势共同反映了浏览距离对最终被购买商品的影响力，即浏览距离 k 的影响力，用 β_k 表示，β_k 的计算方法如式(2)所示。

$$\beta_k=\frac{\sum_C Sim_k(p_i,p_{buy})*\sum_C S_k}{|C_k|^2} \quad (2)$$

其中，C 是浏览轨迹集合，$|C_k|$ 表示集合中长度不小于 k 的浏览轨迹条数。$Sim_k(p_i,p_{buy})$ 表示在一条浏览轨迹中，浏览距离为 k 的商品与最终被购买商品的相似度，S_k 表示在一条浏览轨迹中，商品浏览距离是否为 k 且被购买，若是则为 1，否则为 0。

3.2.2　时间衰减因素

文献[29]指出，数据过时是影响数据质量的重要因素。衰减因子[30]改变了数据统计权重，减小了时效性对于数据计算准确性的影响。用户浏览轨迹的影响会随时间衰减，因此通过引入浏览轨迹的衰减度概念，考虑了用户浏览轨迹的时效性。用户浏览轨迹的衰减度用 ρ 表示，其值越大表示衰减越快，$\rho=0$ 时表示不衰减，则用户浏览轨迹在出现 t 天后的衰减程度为 $(1-\rho)^t$。

3.2.3　基于浏览距离和时间衰减的购买转移概率

通过引入浏览距离影响力，将一条浏览轨迹中不同浏览距离商品到被购买商品的转移关系分配合适的影响因子，使得离散的对应关系具有整体性。通过引入衰减度，对较早时间的浏览轨迹进行衰减，充分考虑历史记录的时效性。

依据真实的浏览轨迹与购买记录数据集，结合浏览顺序因素和时间衰减因素，本文提出浏览轨迹-购买转移概率计算方法，如式(3)所示：

$$P(p_i->p_{buy})=\frac{\sum_C((1-\rho)^t\sum_{k=1}^{r}S_{k,p_i->p_{buy}}*\beta_k)}{\sum_C((1-\rho)^t\sum_{k=1}^{r}S_{k,p_i}*\beta_k)} \quad (3)$$

其中，$P(p_i->p_{buy})$ 表示购买转移关系 $p_i->p_{buy}$ 的概率，C 为浏览轨迹集合，r 表示最大的浏览轨迹长度，β_k 表示浏览距离的影响力，$S_{k,p_i->p_{buy}}$ 表示在浏览轨迹中，浏览距离为 k 的购买转移关系 $p_i->p_{buy}$ 是否存在，若是则为 1，否则为 0。

3.3　应用模型推荐

通过计算购买转移概率，得出所有商品的购买转移关系，构成购买转移关系模型。当新用户浏览商品时，根据用户当前浏览轨迹，从模型中找出满足条件的商品推荐给当前用户，具体过程如下所示。

输入：推荐模型，List=$\{p_1,p_2,\cdots,p_n\}$，recLength

输出：推荐列表 rec

推荐过程：

1)取得 List 中对应商品的 β_k；

2)建立长度为 recLength 的数组 rec[]；

3)foreach p_i in List：

　　计算 $P_k(p_i->p_{buying})=P(p_i->p_{buying})*\beta_k$；

　　将 p_{buying} 按 $P_k(p_i->p_{buying})$ 值从大到小的顺序插入 rec[]中；

4)推荐 rec 数组中的商品。

其中，*List* 是当前用户浏览轨迹，*recLength* 为推荐列表中的商品数量，p_{buying} 为推荐模型中的候选推荐商品，$P_k(p_i->p_{buying})$ 表示在浏览轨迹 *List* 中的浏览距离为 k 的 p_i 购买 p_{buying} 的推荐分。

在利用模型向用户进行推荐时，本文充分利用当前用户的整个浏览轨迹，将浏览距离影响力作为因子，计算出新的转移概率即推荐分作为推荐依据，把推荐分最高的 *recLength* 个商品放到推荐列表 *rec* 中，推荐给用户。

4 实验

4.1 实验数据

本文的实验数据是某电商网站 2013 年 3 月份的访问日志，格式为：

{cid, ip_addr, time, request, status, http_referer, user_agent}

其中，cid 是用户客户端的 cookie_id，ip_addr 是客户端的 ip 地址，time 是访问时间，request 是请求的网址，status 是服务器返回的状态，http_referer 表示从哪个网址链接过来的。日志中包含了用户访问网站时浏览的商品记录和最终购买的商品记录。

通过对日志的分析，提取出 1811 条用户浏览轨迹与购买记录，这些记录包括用户的访问时间、ip 地址、用户浏览轨迹、用户最终购买的商品。其中独立 ip 地址 1462 条，包含商品 685 件。

4.2 实验设计

本文用较早时间浏览轨迹记录的 60%作为训练集，较晚时间浏览记录的 40%作为测试集，并进行了多次计算以排除实验结果的随机性。根据图 2 中实验数据统计可知，用户最倾向于选择购买自己最近浏览的商品，当用户浏览轨迹中的商品距离最终购买的商品的距离越远时，最终被购买的几率越小，当浏览距离大于 10 时，被用户购买的频度已接近 0。因此本文将使用接近用户购买商品的一段长度为 10 的浏览轨迹作为依据进行实验。

本文采用研究中常用的几个指标作为验证标准：精确度、召回率、F_1-*measure*。

在实验数据中，经计算，本文的算法给出一个推荐列表，表示要推荐给用户的项目列表，用 *RL* 表示；推荐列表中包含项目的数目称为推荐长度，用 *recLength* 表示；应该推荐给用户的项目列表用 *TL* 表示。项目列表 *RL* 与 *TL* 的交集是被正确推荐的项目，这些项目的数量表示为 *TP*，项目列表 *RL* 中不应该被推荐而被错误推荐的项目的数量表示为 *FP*，表 *TL* 中应该被推荐而未被推荐的项目数量表示为 *FN*，具体如表 1 所列。

表 1 应该推荐与实际该推荐列表的关系

	应该推荐(TL)	不该推荐
实际推荐(RL)	TP	FP
没有推荐	FN	

精确度 P 反映了项目列表 *RL* 中被正确推荐项目的数量与算法推荐的项目列表 *RL* 中项目数量的比例，如式(4)所示：

$$P=\frac{TP}{(TP+FP)} \tag{4}$$

召回率 R 反映了项目列表 *RL* 中被正确推荐项目的数量与应该被推荐的项目列表 *TL* 中项目数量的比例，如式(5)所示：

$$R=\frac{TP}{(TP+FN)} \tag{5}$$

F_1-*measure* 为根据准确率 P 和召回率 R 二者给出的一个综合的评价指标，如式(6)所示：

$$F_1\text{-}measure=\frac{2*P*R}{P+R} \tag{6}$$

4.3 实验结果分析

推荐长度是推荐效果的重要影响因素；衰减度是所提算法思想的重要参考因素，所以必须明确该因素对推荐效果的影响；算法时间复杂度是方法应用中用户体验评价的重要方面；所提算法的优势要通过与其他相关算法的对比才能体现。所以实验主要包括推荐长度对推荐效果的影响、衰减度对推荐效果的影响、时间复杂度分析以及所提方法与蚁群算法的比较。

4.3.1 推荐长度对推荐效果的影响

图 4 示出随着推荐长度的增加，精确度、召回率以及 F_1-*measure* 的变化曲线，横轴表示推荐长度，纵轴表示精确度、召回率、F_1-*measure*。从图 4 可以看出，召回率随着推荐列表长度的增长而缓慢地增长；相反，推荐精确度与 F_1-*measure* 则会随着推荐长度的增长而下降。当推荐长度为 2 时，推荐精确度与 F_1-*measure* 最大，推荐精确度为 0.74，F_1-*measure* 值为 0.61，召回率为 0.64，增长缓慢，所以当推荐长度为 2 时，推荐结果最好。

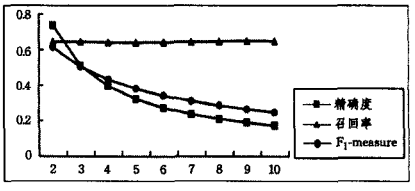

图 4 推荐长度对推荐效果的影响

4.3.2 衰减度对推荐效果的影响

在不考虑浏览轨迹衰减因素的情况时，推荐长度为 2 时推荐效果最好，但考虑衰减度下，不同推荐长度对推荐效果的影响尚未可知。所以本文也对推荐长度为 4 时，衰减度对推荐效果的影响进行了实验。图 5、图 6 分别表示推荐长度为 2 和 4 时，随着衰减度增加，精确度、召回率以及 F_1-*measure* 的变化情况。横轴表示每条轨迹随时间的衰减程度，纵轴表示精确度、召回率以及 F_1-*measure*。

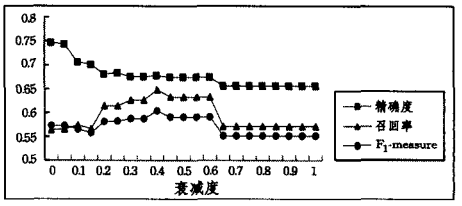

图 5 推荐长度为 2 时，衰减度对推荐效果的影响情况

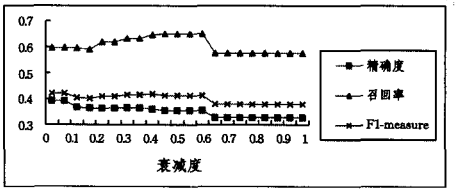

图 6 推荐长度为 4 时，衰减度对推荐效果的影响情况

从图5可以看出，随着衰减度的增加，虽然精确度稍微下降，但是召回率和 F_1-*measure* 呈现出先上升后下降的趋势，当衰减度为0.45左右时，F_1-*measure* 最大，综合推荐效果最佳。

从图6可以看出，随着衰减度的增加，精确度、召回率和 F_1-*measure* 总体上呈现先上升后下降的趋势，当衰减度为0.4时，综合效果较好。

4.3.3 算法的时间复杂度分析

所提算法主要分为建立模型和推荐商品两个部分。在建立模型部分，程序需要首先扫描一遍数据以统计商品特征趋势，然后第二次扫描数据，结合商品特征趋势和时间衰减因素，计算购买转移概率。该过程的执行时间取决于浏览轨迹与购买记录的规模，属于线性增长，理论上时间复杂度为 $O(n)$。在推荐部分，算法根据当前用户的浏览轨迹，参考商品特征趋势计算推荐分，将推荐分最高的商品推荐给用户，其过程取决于商品规模和推荐长度。而所提方法避免了对所有商品排序，而是采用顺序插入的方法。由于推荐长度一般较短，可以看作常数，因此执行时间属于线性增长，理论上时间复杂度为 $O(n)$。

本文在PC机上对算法时间复杂度进行了多次测试并取平均值，以排除偶然因素干扰，测试环境如表2所列。

表2 测试机配置表

配置项目	配置参数
处理器	Intel Core i5-3470 CPU @ 3.2 GHz 双核
运行内存	4.00 GB
操作系统	Window 7 旗舰版 64-Bit
编程语言	Java SE JRE 8 64-Bit

建立模型是所提算法的主要组成部分，所以要对建立模型算法的时间复杂度进行分析，测试不同浏览轨迹规模对执行时间的影响，观察随着数据规模的增长，执行时间的变化情况，结果如图7所示。

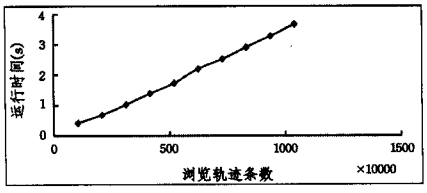

图7 浏览轨迹规模对建立模型时间的影响分析

从图7可以看出，随着浏览轨迹条数的增加，时间复杂度呈线性增长，与理论分析相吻合。另外对于上千万条的用户浏览轨迹，也能在4s之内运行完成。

推荐部分的执行时间直接影响到用户体验，执行效率是必须考虑的方面，所以要对推荐部分算法的时间复杂度进行分析，测试不同商品规模对执行时间的影响，观察随着数据规模的增长，执行时间的变化情况，结果如图8所示。

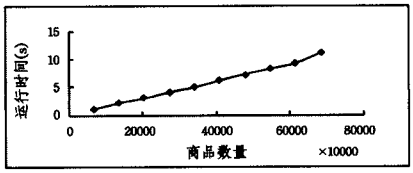

图8 商品规模对推荐时间的影响分析

从图8可以看出，随着商品规模扩大，推荐时间复杂度呈线性增长趋势，对于几亿商品的规模，也能在几秒内完成推荐过程。

通过图7、图8可以看出，所提算法的执行时间不会因为数据集规模增长徒然增高，时间复杂度为 $O(n)$，能够满足运行较大数据集的需求。

4.3.4 与蚁群算法的比较

将文献[24]中所述蚁群算法应用到本文的数据集，并与所提算法进行比较。

图9—图11分别是所提算法与蚁群算法在精确度、召回率以及 F_1-*measure* 3个推荐效果指标上的比较。图中横轴表示推荐长度，纵轴分别是两种算法的精确度、召回率以及 F_1-*measure*。

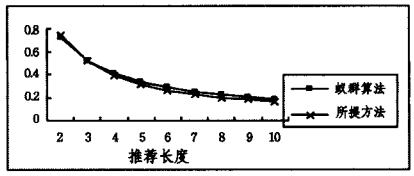

图9 本文算法与蚁群算法的精确度比较

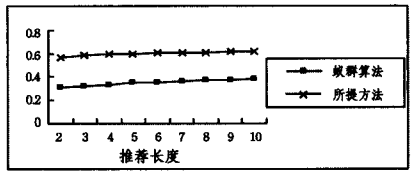

图10 所提算法与蚁群算法的召回率比较

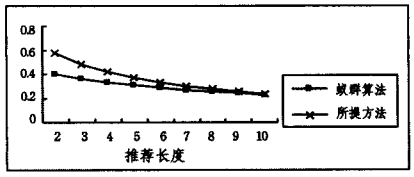

图11 所提算法与蚁群算法的 F_1-*measure* 率比较

从图9可以看出，蚁群算法和所提算法在精确度方面基本持平，当推荐长度不同时，两种算法精确度基本一致。

从图10可以看出，所提算法在推荐长度不同时，召回率明显高于蚁群算法。

从图11可以看出，所提算法在推荐长度不同时，F_1-*measure* 明显高于蚁群算法，虽然随着推荐长度增长差距有所减小，但整体上，所提算法的优势依然比较明显。

通过以上实验结果可以看出，虽然精确度没有明显提升，但是召回率和 F_1-*measure* 值提升明显。这说明所提算法总体上要优于蚁群算法。

结束语 本文提出的用户浏览轨迹-偏爱推荐模型，相比蚁群算法不仅有较好的推荐效果，而且缓解了新用户的冷启动问题，对解决新用户的冷启动问题起到了积极的作用。

参 考 文 献

[1] Pazzani M J, Billsus D. Content-Based Recommendation Systems [M] // The Adaptive Web. Springer Berlin Heidelberg, 2007: 325-341

[2] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms[C] // Proceedings of International Conference on World Wide Web. 2001: 285-295

[3] Schein A I, Popescul A, Ungar L H, et al. Methods and metrics for cold-start recommendations[C] // Proceedings of the 25th

Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2002: 253-260

[4] Lam X N, Vu T, Le T D, et al. Addressing cold-start problem in recommendation systems[C]// Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication. ACM, 2008: 208-211

[5] Song Y. Personalized Search Engine Based on the User Behavior Analysis[J]. New Century Library, 2013

[6] Ye Y. Research on Interest Model of User Behavior[D]. Shanghai: East China University, 2012(in Chinese)
叶彧. 基于用户行为的兴趣模型的研究[D]. 上海: 东华大学, 2012

[7] Sun X H. Research on sparsity and cold start of collaborative filtering system[D]. Hangzhou: Zhejiang University, 2005(in Chinese)
孙小华. 协同过滤系统的稀疏性与冷启动问题研究[D]. 杭州: 浙江大学, 2005

[8] Huang G Q, Zhao Y M. Approach to collaborative filtering recommendation based on HMM[J]. Journal of Computer Applications, 2008, 28(6): 1601-1604(in Chinese)
黄光球, 赵永梅. 基于 HMM 模型的协同过滤推荐方法[J]. 计算机应用, 2008, 28(6): 1601-1604

[9] Juan B. Collaborative filtering recommendation algorithm based on semantic similarity of item[C]// IEEE Fifth International Conference on Advanced Computational Intelligence. 2012: 452-454

[10] Fan M, Zhou Q, Zheng T F. Content-Based Semantic Tag Ranking for Recommendation[C]// 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology. IEEE Computer Society, 2012: 292-296

[11] Yao L, Sheng Q Z, Segev A, et al. Recommending Web Services via Combining Collaborative Filtering with Content-Based Features[C]// 2013 IEEE 20th International Conference on Web Services. IEEE, 2013: 42-49

[12] Popescul A, Ungar L H, Pennock D M, et al. Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments[C]// Proceedings of the 17th Conference on Univertointy in Artificial Intelligence. 2013: 437-444

[13] Chu W, Park S T. Personalized recommendation on dynamic content using predictive bilinear models[C]// Proceedings of the 18th International Conference on World Wide Web. ACM, 2009: 691-700

[14] Wu T, He H, Gu X, et al. An intelligent network user behavior analysis system based on collaborative Markov model and distributed data processing[C]// 2013 IEEE 17th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, 2013: 221-228

[15] Gong M, Xu Z, Xu L, et al. Recommending Web Service Based on User Relationships and Preferences[C]// IEEE International Conference on Web Services. 2013: 380-386

[16] Sobhanam H, Mariappan A K. Addressing cold start problem in recommender systems using association rules and clustering technique[C]// International Conference on Computer Communication & Informatics. IEEE, 2013: 1-5

[17] Zhang D, Hsu C, Chen M, et al. Cold-Start Recommendation Using Bi-Clustering and Fusion for Large-Scale Social Recommender Systems[J]. IEEE Transactions on Emerging Topics in Computing, 2014, 2(2): 239-250

[18] Park S T, Pennock D, Madani O, et al. Na07ve filterbots for robust cold-start recommendations[C]// Proc of Acm Sigkdd Int'l Conf. 2006: 699-705

[19] Che G Y, Zhang L, Zhang L X. User Browsing Behavior Extraction and Analysis Based on Sequence Pattern[J]. Computer Technology & Development, 2012, 22(9): 9-12(in Chinese)
车高营, 张磊, 张禄旭. 基于序列模式的用户浏览行为提取与分析[J]. 计算机技术与发展, 2012, 22(9): 9-12

[20] Yap G E, Li X L, Yu P S. Effective next-items recommendation via personalized sequential pattern mining[C]// Proceedings of the 17th International Conference on Database Systems for Advanced Applications-Volume Part II. Springer-Verlag, 2012: 48-64

[21] Li Y, Niu Z, Chen W, et al. Combining collaborative filtering and sequential pattern mining for recommendation in e-learning environment[C]// International Conference on Advances in Web-based Learning. Springer-Verlag, 2011: 305-313

[22] Choi K, Yoo D, Kim G, et al. A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis[J]. Electronic Commerce Research & Applications, 2012, 11(4): 309-317

[23] Chen W, Niu Z, Zhao X, et al. A hybrid recommendation algorithm adapted in e-learning environments[J]. World Wide Web-internet & Web Information Systems, 2014, 17(2): 1-14

[24] Zhou Y N, Zheng H S. Ant Collaborative Filtering Based on Options of Browsing Path: Used for M-commerce Personalized Recommendation System[C]// Systems Engineering Society of China. 2012(in Chinese)
周玉妮, 郑会颂. 基于浏览路径选择的蚁群推荐算法: 用于移动商务个性化推荐系统[C]// 中国系统工程学会学术年会. 2012

[25] Khan S, Baig A R, Shahzad W. A novel ant colony optimization based single path hierarchical classification algorithm for predicting gene ontology[J]. Applied Soft Computing, 2014, 16(3): 34-49

[26] Mohanthy R, Naik V, Mubeen A. Software Reliability Prediction by Using Ant Colony Optimization Technique[C]// Fourth International Conference on Communication Systems & Network Technologies. IEEE Computer Society, 2014: 496-500

[27] Singh L K, Vinod G, Tripathi A K. Approach for parameter estimation in Markov model of software reliability for early prediction: A case study[J]. Iet Software, 2015, 9(3): 65-75

[28] Sampathkumar H, Chen X W, Luo B. Mining Adverse Drug Reactions from online healthcare forums using Hidden Markov Model[J]. Bmc Medical Informatics & Decision Making, 2014, 14(1): 1-18

[29] Li M H, Li J Z, Cheng S Y. Uncertain rule based method for evaluating data currency[J]. Journal of Software, 2014, 25(Suppl. (2)): 147-156(in Chinese)
李默涵, 李建中, 程思瑶. 一种基于不确定规则的数据时效性判定方法[J]. 软件学报, 2014, 25(Suppl. (2)): 147-156

[30] Yang C, Wu A R. Method of evaluation data freshness based on reduction-factor[J]. Computer Engineering and Design, 2010, 31(3): 684-686(in Chinese)
杨超, 吴爱荣. 基于衰减因子的评价数据时效性处理方法[J]. 计算机工程与设计, 2010, 31(3): 684-686