

# Overview and Prospect of Personalized Recommendation Algorithm

Xinxin Li

Dalian University of Foreign Languages, Dalian Liaoning  
Email: 15042037186@163.com

Received: Jun. 19<sup>th</sup>, 2019; accepted: Jul. 2<sup>nd</sup>, 2019; published: Jul. 9<sup>th</sup>, 2019

---

## Abstract

In recent years, the word “information overload” frequently appears in people’s vision, it has become a hot word in the field of computer, and it is also an important problem that researchers urgently need to solve. In order to solve the problem of information overload, researchers in the field of computer constantly optimize the personalized recommendation algorithm, strive to reduce the difficulty of information retrieval for users, to provide users with the best personalized recommendation results. This paper gives a brief overview of the personalized recommendation methods which are widely used and common. Combined with the experience of using personalized recommendation algorithm to generate results in daily life, the author puts forward expectations for the development of personalized recommendation algorithm in the future.

## Keywords

Personalized Recommendation, Collaborative Filtering, Hybrid Recommendation

---

# 个性化推荐算法概述与展望

李鑫欣

大连外国语大学, 辽宁 大连  
Email: 15042037186@163.com

收稿日期: 2019年6月19日; 录用日期: 2019年7月2日; 发布日期: 2019年7月9日

---

## 摘要

近年来, “信息过载”一词频繁出现在人们的视野中, 它成为了计算机相关领域中的热门词汇, 同时它也是研究人员急待解决的重要问题。为解决信息超载的问题, 计算机领域研究人员不断优化个性化推荐

算法, 力求降低用户的信息检索难度, 为用户提供最优的个性化推荐结果。本文对于应用范围较广、较为常见的个性化推荐方法做出简要的概述, 并结合日常生活中使用个性化推荐算法生成结果的经历, 对未来个性化推荐算法的发展提出期望。

## 关键词

个性化推荐, 协同过滤, 混合推荐

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

近年来, 随着网络信息技术的不断发展与进步, 网民数量激增, 网络覆盖率日益增高。根据中国互联网络信息中心(CNNIC)发布的第 43 次《中国互联网发展状况统计报告》显示, 截止 2018 年 12 月, 我国网民规模为 8.29 亿, 全年新增网民 5653 万, 互联网普及率达 59.6%, 较 2017 年底提升 3.8 个百分点[1]。互联网行业发展迅速, 购物、出行、医疗、教育等生产生活中的方方面面都有着“互联网+”的影子。在这个网络化、信息化的时代, 网络技术已经在潜移默化地影响着、改变着人们的生活方式和思想观念。网络信息呈指数级增长, 海量信息在方便人们生活的同时也增加了用户检索目标信息的难度。基于信息超载的情况, 个性化推荐算法应运而生。个性化推荐系统通过挖掘用户在网络上留下的“信息足迹”, 采集并分析用户的网络行为与消费偏好, 根据不同的推荐算法将精准的、契合度高的内容推荐给用户。个性化推荐算法的产生与发展极大地便利了人们的生产与生活, 对于用户而言, 不用再为在海量的信息中检索需要的内容而苦恼, 对于商家而言能够更好地分析用户行为, 提高竞争力与实现经济效益的最大化增长。个性化推荐系统源于人们的生活, 同时也为在更好地服务于人们的生活。

## 2. 个性化推荐方法概述

### 2.1. 协同过滤推荐

协同过滤推荐(collaborative filtering recommendation)是由 Goldberg 等人于 1992 年提出来的[2], 主要思想是通过现有的用户群以往的意见和行为, 对当前用户最有可能感兴趣的物品进行预测。协同过滤技术一般分为两大类: 基于记忆的技术与基于模型的技术。传统的基于用户的技术是基于记忆的, 通过内存中已保存的用户原始行为数据可以快速地、直接地生成推荐结果。而基于模型的方法会首先离线处理现有的原始数据, 通过使用“学习过”的训练模型来预测、提供推荐结果。在理论上, 基于记忆的方法数据量更大, 推荐的结果精确度更高。协同过滤技术是电子商务推荐系统中最广泛使用的、最成功的推荐算法, 但是该算法在发展过程中会受到冷启动、稀疏性、可拓展性等相关制约因素的影响[3]。随着人工智能领域学习热潮的兴起, 研究者在传统的推荐算法基础上融合了人工智能技术, 有效地利用数据, 缓解了传统协同过滤算法中的冷启动等问题[4]。对协同过滤算法的特点与发展现状做简要地概述如下。

#### 2.1.1. 协同过滤算法的优点

##### 推荐结果新颖:

协同过滤算法基于搜索相似用户的个人兴趣偏好进行个性化推荐, 不需要分析用户画像与物品画像

的相似性，推荐结果新颖程度更高、利于发现用户潜在的个人兴趣偏好，个性化推荐结果更具多样化特点。

### 2.1.2. 协同过滤算法的局限性

#### 冷启动：

当系统中的新用户或者新物品(物品指推荐系统为用户推荐所有内容的总称)出现时，由于系统中没有新用户与的相关评分、个人偏好信息和新物品的被评分信息，使用协同过滤推荐算法会导致初期的推荐系统准确度不高，推荐内容不完善。

#### 数据稀疏性：

当存在系统中的用户对于物品的评价数据非常少或随着系统规模的不断扩大、物品的不断增加用户的评价矩阵内容变得非常稀疏时会导致无法为用户找到近邻，推荐结果的准确度大幅下降。

#### 可扩展性问题：

协同过滤推荐算法的思想是基于近邻的搜索方法生成推荐结果。随着网站中用户与物品的不断增加，使得用户一项目评分矩阵成为高维矩阵，算法的计算复杂度极具增高[5]，由此产生了协同过滤推荐算法的可扩展性问题。

### 2.1.3. 发展现状

为解决传统的协同过滤推荐算法的不足之处，研究者提出了矩阵因子分解模型。由于用户的喜好是不断发展变化的，所以要充分考虑时间因素对于推荐结果的影响。矩阵因子分解的通过将评分分解为不同的项，可以分别处理不同方面的时序影响[6]，提高了推荐结果的精准度。

## 2.2. 基于内容的推荐

基于内容推荐(content-based filtering recommendation)是指通过掌握的物品特征的描述和描述了用户历史兴趣的记录[7]，确定最能匹配用户喜爱的物品并推荐给用户。与协同过滤算法相比，基于内容的推荐不需要掌握巨大的用户群和评分记录，即使只有一个用户，也能够为其生成个性化推荐列表。典型的对多值特征物品进行相似度度量的方法会用到 Dice 系数[8]。常用的基于内容相似度检索方法有通过最近邻、相关性反馈——Rocchio 方法、基于概率模型的方法、显示决策模型等。通过研究人员对于不同相似度检索方法的对比研究发现贝叶斯和 Rocchio 算法在许多领域表现良好。基于内容的推荐多应用与电子邮件或新闻中，通过从文档中提取或者自动从文字描述中抽取关键词来生成关键词列表。IBM 的 FileNetP8 产品就是基于内容的全文检索服务产品。利用全文检索，用户可以在 P8 Object Store 中搜索那些内容包含特定单词或者短语的文档、附件、和字符串形式的属性等。对基于内容的推荐算法的特点与发展现状做简要地概述如下。

### 2.2.1. 基于内容推荐算法的优点

#### 用户独立性高：

相比于协同过滤算法，基于内容的推荐算法仅使用当前用户提供的评分数据构建个人信息集，用户独立性程度高。

#### 易于解决冷启动问题：

对于系统中出现的新物品，在没有任何用户评分的情况下，也可以进行推荐。

### 2.2.2. 基于内容算法的局限性

#### 语义处理难度大：

基于内容的推荐在处理文本信息是经常会遇见一词多意、同义词等问题，增强了算法研究的难度，对于提供精准的个性化推荐精准度是巨大的挑战。

#### **推荐结果新颖度低：**

基于内容的推荐算法生成结果严格衡量用户个人喜好与物品信息匹配程度，虽然推荐结果专业化程度高，但是也很难发现新颖的、惊喜度高的物品。

### **2.2.3. 发展现状**

近年来，语义技术的快速发展和维基百科等开放知识源的普及，极大地推动了基于内容推荐系统的发展[9]。研究者将自然语言处理、语义技术等其它深度分析与基于内容的推荐算法结合在一起把基于用户及物品基本特征的表示从基于关键词的级别提升至基于概念的级别，形成了一种基于内容推荐算法的新的发展方向——基于内容的语义感知推荐系统(semantics-aware content-based recommender system)。

### **2.3. 混合推荐**

为克服协同过滤算法、基于内容算法等其他算法的局限性、提高个性化推荐结果的精准度，研究者将两种以上算法和模型的优点结合在一起，提出一种新的个性化推荐算法——混合推荐(hybrid recommendation)算法来提高推荐结果的准确程度。三种基本的混合设计为：1) 整体式混合设计(包括特征组合混合方案、特征补充混合方案)；2) 并行式混合设计(包括交叉式混合、加权式混合、切换式混合)；3) 流水线混合设计(串联混合、分级混合)。三种基本的混合设计各有利弊，可根据实际情况和对比分析结果选择合适的混合设计方式。在实际应用过程中，基于内容的推荐算法与基于协同过滤的推荐算法相结合成为混合推荐系统的主流[10]。现在大多数的个性化推荐系统都是综合多种推荐算法而构建混合推荐系统。在未来，混合推荐系统将会被更广泛地应用。

## **3. 个性化推荐算法应用实例**

### **3.1. UGC 标签系统——豆瓣网站**

UGC (user generated content, 用户生成的内容)标签系统是很多 Web2.0 网站的重要组成部分。该系统是由用户通过网站为物品打标签在描述物品的语义的同时标记用户自身的兴趣爱好的标签系统。豆瓣网站是中国著名的评论和社交网站，同时也是中国个性化推荐领域的领军企业之一[11]。以豆瓣的读书网站为例，在网站主页会显示一系列信息显示热门标签名称，用户可根据自己的兴趣爱好选择合适的标签查看相关书籍；同时在每一本书籍的详细页面会显示读过该书的用户标注的标签，以供浏览此书的用户能够快速了解书籍类型。在豆瓣读书网站，通过用户为书籍打标签来扩充书籍的语义解释范围。在最终的个性化推荐结果里，豆瓣利用标签将用户推荐结果做了聚类，显示了对不同标签下用户的推荐结果，从而增加推荐的多样性和可解释性[12]。用户既是标签的创作者，也是标签推荐系统的服务对象。

### **3.2. 利用用户行为数据——购物网站**

用户浏览网站时会留下大量的“信息足迹”。通过挖掘用户产生的商品评价以及搜索、浏览商品信息情况，了解用户的兴趣和需求，从而为其个性化推荐他们可能感兴趣的内容。最著名的就是啤酒和尿不湿的例子。从设计人员的角度看，这一成功的实践案例说明了用户行文数据中蕴含着大量的、并不是那么显而易见的规律与信息。个性化推荐算法被广泛地应用于各大电商平台。网站通过分析用户留下的显示信息(包括收藏、交易记录)以及隐式信息(多次搜索、浏览的商品类别)，挖掘用户的兴趣与需求，从而推测用户消费偏好，为其推荐同类商品或相关商品，提高用户体验度进而提高电商平台的经济效益。

### 3.3. 基于内容的推荐算法——今日头条

今日头条是一款新闻资讯类 APP (Application)。与传统的新闻客户端不一样，它们自身并不产生新闻，而是新闻的搬运工，通过分析读者的阅读偏向，向用户推荐个性化的新闻资讯[13]。今日头条通过用户的手机地理位置信息、关注人信息及标签、实时搜索内容、网站访问足迹(包括点赞、评论等操作)信息，为用户贴上多种多样的标签，使用该应用的最常用算法之一的基于内容的推荐算法，为用户提供符合用户喜好的相关信息，提高软件的使用率与信息的点击率。

## 4. 个性化推荐算法发展趋势展望

个性化推荐算法发展至今已经有十多年的历史了，科研人员不断致力于探索更加高效的推荐算法。但是现在的个性化推荐技术仍有不完善之处。从用户角度来看，对个性化推荐系统有着更多的期待。

### 4.1. 多角度考虑用户行为信息

在分析用户行为的过程中，应多角度考虑用户的浏览行为数据，不仅要考虑哪些内容或商品是用户喜欢的，还要考虑哪些是用户不喜欢或者不关注的，通过双向信息相结合才能使推荐算法更准确。以电商销售平台为例，系统通过分析用户的隐式、显示的喜好信息，为用户提供了相似的商品内容，接下来要对了解用户对于推荐内容的反馈，推荐的商品用户是否有点击浏览的“行为痕迹”。一些被推荐的内容出现几次不被用户点击或者查看的情况下，是否有必要考虑此类商品不再出现，并加入到用户“弱喜爱”的队列中，当用户主动搜索此商品时再加入推荐队列中。个性化推荐算法在考虑用户喜欢什么的同时，也要关注用户不喜欢什么，从多角度分析用户的行为数据才能促使推荐算法更加成熟，推荐结果更加准确。

### 4.2. 细化商品类别

无论是新闻资讯网站的文章还是电商平台销售的物品，在此处可以把它们看作商品来分析。网站在不断尝试各种优秀的推荐算法的同时，还应该尽可能地细化自身的商品类别，更加关注商品间的关联强弱程度。对于商品内部类别、属性的细化分类与外部与其它商品的关联关系的优化有助于更加精确地分析用户行为信息，提高个性化推荐算法的准确度。

### 4.3. 推荐结果更注重时效性

在根据用户搜索的内容个性化地为其推荐最新、最热的内容的同时，也要考虑用户搜索时间点的问题，要为用户推荐符合时间要求的内容。例如，当用户使用某个学习软件多次查找考研相关资料后，该软件的个性化推荐系统根据此用户以往的历史浏览数据分析出他可能是近期处于备考阶段，因此给其推荐一些考研相关的主题，如“你考研的时候是怎么查资料的？”、“2020 考研数学该怎么复习？”、“考研复试如何准备个人简历”等，但是偶尔也会出现“如何准备 2019 年考研”这一类的与用户预期时间问题有差距的话题。在搜索的内容对时间有较严格的要求但又未在搜索栏中的情况下，研究人员应着重考虑根据暗含着的时间信息将具有时效性的内容推荐给用户，提高个性化推荐的准确度。

### 4.4. 推荐结果质量更高

在经过采集、分析用户的浏览和操作网页的信息后，为其推荐符合其偏好的内容或物品的同时，也要考虑推荐内容的质量问题。例如，用户在某购物网站中浏览或收藏过护肤品相关的内容，现在的个性化推荐算法的实时性很强，很快就会根据用户行为推荐出许多护肤品品牌。也许前一分钟你还在搜索着

各种各样的护肤品，后一分钟当你刷新首页后就会惊奇地发现网站已经为你提供了各种品牌的护肤品。在追求快速分析与推荐内容与物品的同时，要对为用户提供的信息的质量进行严格的把控。要为用户推荐高质量的内容，就护肤品而言，综合该用户收藏过的相关品牌、销量较高、购买过的用户评价较高、售后服务较好等综合因素为其推荐更符合用户要求、更有可能被用户购买的相关商品。

#### 4.5. 推荐内容更多样化

推荐内容的多样化不仅体现在符合用户多样化的兴趣需求的多样化，也体现在推荐系统本身所提供的内容是多样化的，是并不局限于用户某一段时间内显示的兴趣方向特征的。用户的兴趣具有广泛性的特点，为提高个性化推荐内容的质量，推荐列表应最大程度上覆盖用户不同的兴趣领域，推荐内容应具有多样性的特点。尽管用户的兴趣在较长的时间跨度是不一样的，但具体用户访问推荐系统的某一刻，其兴趣往往是单一的，那么如果推荐列表智能覆盖用户的一个兴趣点，而这个兴趣点不是用户这个时刻的兴趣点[14]，推荐列表就不会让用户满意。用户的兴趣需求与关注内容是不断发展变化的，在每一次浏览网页或互联网产品页面时用户想要收获的不仅仅是符合自己兴趣的个性化推荐内容，同时也期待能够有新颖的、精彩的领域等着自己去探索。若一味地“享受”单一的内容推荐，会让人的视野只局限于固定大小的范围，不利于开阔视野及全面的发展。在用户在浏览某网站无搜索需求的情况下，如果个性化推荐算法只为用户提供了通过分析历史数据推荐出来的符合以往搜索内容的兴趣内容，有可能会出现用户浏览内容过于单一、降低用户体验感与满意度，不利于网站的长期发展。网站只有不断地迎合、拓展用户的需求，才能提高用户与网站的交互频率，实时挖掘用户潜藏行为偏好数据，做到“比你更了解你自己”。

#### 4.6. 预测结果更准确

预测准确度度量一个推荐系统或者推荐算法预测用户行为的能力。以电商平台为例，提高预测系统的准确度不仅要考虑与当前商品相关性较强的商品，还有分析有购买此商品的用户的消费行为，在购买此商品后购买的于此商品相关的其它商品有哪些。

### 5. 结语

本文对几类个性化推荐算法进行基本的概论，从一个用户角度提出了对于个性化推荐算法未来发展的期望。个性化推荐算法发展至今已经有十几年的历史了，在科研人员的不断努力下达到了非常高的水平。个性化推荐算法的应用范围不断扩大，包括电子商务网站、娱乐网站、社交网站等许多方面。作为一个不断创新的领域，人们对个性化推荐算法的期望也在不断提高。在不断创新发展的过程中，个性化推荐系统会更好服务于人们的生活，提高用户生活质量和互联网产品的效益。

### 致 谢

感谢大连外国语大学软件学院关菁华老师对我的指导与帮助。

### 基金项目

2018年大连外国语大学大学生创新创业训练计划项目资金支持(项目编号: 201810172071)。

### 参考文献

- [1] CNNIC 发布第 43 次《中国互联网发展状况统计报告》[EB/OL].  
[http://cnnic.cn/gywm/xwzx/rdxw/20172017\\_7056/201902/t20190228\\_70643.htm](http://cnnic.cn/gywm/xwzx/rdxw/20172017_7056/201902/t20190228_70643.htm), 2019-02-28.

- [2] 张志威. 个性化推荐算法研究综述[J]. 信息与电脑(理论版), 2018(17): 27-29.
- [3] 冷亚军, 黎忠雪. 个性化推荐及其相关技术分析[J]. 内蒙古科技与经济, 2019(5): 58-60.
- [4] 王永贵, 尚庚. 融合注意力机制的深度协同过滤推荐算法[J/OL]. 计算机工程与应用, 1-8. <http://kns.cnki.net/kcms/detail/11.2127.TP.20190417.1209.018.html>, 2019-05-19.
- [5] 冷亚军, 陆青, 梁昌勇. 协同过滤推荐技术综述[J]. 模式识别与人工智能, 2014, 27(8): 720-734.
- [6] 李忠武, 王辉, 魏再超. 基于推荐系统时间敏感的因子模型算法研究[J]. 电子商务, 2017(9): 55-56, 89.
- [7] 黄明波. 基于 Slope One 算法的增量音乐推荐系统的设计与实现[D]: [硕士学位论文]. 重庆: 重庆大学, 2016.
- [8] 黄琼. 网络图书资源个性化推荐算法研究[D]: [硕士学位论文]. 成都: 西南交通大学, 2014.
- [9] 弗朗西斯科·里奇, 力奥·罗卡奇等. 推荐你系统: 技术、评估及高效算法[M]. 李艳民, 等, 译. 北京: 机械工业出版社, 2018.
- [10] 孙光浩, 刘丹青, 李梦云. 个性化推荐算法综述[J]. 软件, 2017, 38(7): 70-78.
- [11] 吴幸良. 基于图模型的个性化标签推荐技术研究[D]: [硕士学位论文]. 重庆: 重庆大学, 2014.
- [12] 项亮. 推荐系统实战[M]. 北京: 人民邮电出版社, 2012.
- [13] 郝涛. 大数据技术下移动新闻客户端的传播模式研究——以人民日报、腾讯新闻、今日头条为例[J]. 新媒体研究, 2019, 5(5): 12-15.
- [14] 陈芸. 基于协同过滤的上下文感知推荐算法的研究[D]: [硕士学位论文]. 武汉: 武汉理工大学, 2014.

### 知网检索的两种方式:

1. 打开知网首页: <http://cnki.net/>, 点击页面中“外文资源总库 CNKI SCHOLAR”, 跳转至: <http://scholar.cnki.net/new>, 搜索框内直接输入文章标题, 即可查询;  
或点击“高级检索”, 下拉列表框选择: [ISSN], 输入期刊 ISSN: 2163-145X, 即可查询。
2. 通过知网首页 <http://cnki.net/>顶部“旧版入口”进入知网旧版: <http://www.cnki.net/old/>, 左侧选择“国际文献总库”进入, 搜索框直接输入文章标题, 即可查询。

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [hjdm@hanspub.org](mailto:hjdm@hanspub.org)