DOI:10.16707/j.cnki.fjpc.2018.11.001

基于隐式反馈的协同过滤算法研究综述

李 改，邹小青
（顺德职业技术学院电子与信息工程学院 广东 顺德 528333）

【摘 要】基于隐式反馈的协同过滤算法是信息推荐系统中广泛运用的核心技术，近年来在国内外得到了深入研究。以往有关协同过滤的研究综述主要侧重于传统的协同过滤算法，有关基于隐式反馈的协同过滤算法的研究综述较少。文中对基于隐式反馈的协同过滤算法的相关研究进行全面总结，首先介绍基于隐式反馈的协同过滤算法的简介及其所面临的挑战，接着详细介绍当前各类基于隐式反馈的协同过滤算法的研究现状，最后给出基于隐式反馈的协同过滤算法需要进一步解决的问题和可能的发展方向。文中详细介绍基于隐式反馈的协同过滤算法的知识框架，理清了基于隐式反馈的协同过滤算法的研究脉络，为后续研究提供参考。相信该研究工作对推进个性化信息服务的发展具有重大意义。

【关键词】推荐系统；协同排序；个性化服务；单类协同过滤；隐式数据

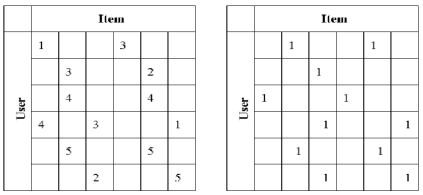
1 引言

随着互联网，特别是电子商务的蓬勃发展，互联网上的信息总量呈现出爆炸性增长。大量无用、冗余的信息严重干扰、妨碍网民高效快捷的获取和分析正确、有价值的信息。在大数据时代，如何在浩瀚的信息海洋中高效的、快捷的获取有用的个性化信息成为广大网民的迫切需求[1][2][3]。

传统的信息系统，例如搜索引擎，它可以根据输入关键字来搜索用户所需要的信息，从而检索到部分高质量、专指性强的信息。例如利用搜索引擎可以检索到学术期刊、论文或者商品。因此，传统的信息系统可以部分解决“信息过载”的问题。但是，对于相同的用户，搜索引擎返回的结果都是一样的，因此传统的信息系统缺乏个性化。为了解决该问题，信息推荐系统应运而生。信息推荐系统的核心思想是：首先通过信息系统收集和分析用户的各种信息特征，接着运用各种机器学习方法来学习用户的个性化兴趣和行为模式，最后根据学习分析得到的用户的兴趣和行为模式，为该用户推荐他所需要的个性化商品或服务[3-17]。信息推荐系统在互联网，特别是电子商务领域得到了广泛应用，如：豆瓣电影（https://movie.douban.com）为用户推荐各种其可能喜欢的电影；淘宝网（www.taobao.com）为用户推荐各种其可能喜欢的商品，如书籍、音像、电器、服装等；世纪佳缘（www.love21cn.com）、百合网（www.baihe.com）等为用户推荐男/女朋友；谷歌（https://www.google.com）、百度（www.Baidu.com）、雅虎（https://www.yahoo.com）等为用户提供个性化的新闻推荐和搜索服务。

基于协同过滤的推荐算法是在信息推荐系统中运用最广泛、最成功的信息推荐算法[3][4]。协同过滤算法的核心思想是：首先分析某个用户的兴趣模式；接着在用户群中找到与该用户兴趣模式相似的用户群组，综合这些相似用户群组对某一推荐对象（如：电影）的评分，把该评分作为信息推荐系统自动形成的该指定用户对此推荐对象（如：电影）的预测评分；最后把预测评分高的推荐对象（如：电影）推荐给该用户。

近年来随着协同过滤算法在互联网工业界的广泛应用，国内外学术界对其进行了广泛研究。按照所处理的数据类型的不同，协同过滤算法主要分为两大类：一类是处理明确的偏好数据，如：显式评分数据（Netflix Prize 比赛数据集是 1–5 分）[4][18-20]。另一类则是处理隐式反馈数据[21-26]，如：对网页点击与否（图 1 给出了显式评分（左）和隐式反馈（右）数据的示意图）。在真实的应用环境中，用户隐式反馈数据存在更广泛、更易获取，譬如用户是否看过某个电影，用户是否听过某首歌曲等等。这类数据无需用户特意提供明确的评分，所以更易获取。由于隐式反馈数据中负例不确定，仅有正例可明确区分，所以学术界也把仅基于隐式反馈的协同过滤问题称为单类协同过滤（OCCF）问题[21]。基于隐式反馈的协同过滤的核心任务是：运用各种机器学习方法学习、分析这些隐式反馈数据，学习用户的行为和兴趣模式，从而来针对特定用户的偏好对推荐对象集按该用户的喜好程度排序。考虑到在以电子商务交易系统为代表的信息推荐系统中，推荐算法所处理的数据大多是隐式数据；同时，在学术界，对基于隐式反馈的协同过滤算法的研究已成为推荐系统研究领域的一大研究热点，得到了众多学者的广泛研究。本文将总结基于隐式反馈的协同过滤算法的研究现状，并提出基于隐式反馈的协同过滤算法所面临的新问题和可能的发展方向，为后续研究提供参考。相信该研究工作既具有极大的学术价值，同时在也对推进个性化信息服务的发展具有重大意义。



（a）显式评分 （b）隐式反馈

图 1 传统协同过滤所处理的数据类型示意图

本文具体内容安排如下：第 1 节为本文的引言部分；第 2 节介绍基于隐式反馈的协同过滤算法的简介及其所面临的挑战；第 3 节详细介绍当前各类基于隐式反馈的协同过滤算法的研究现状。第 4 节给出基于隐式反馈的协同过滤算法需要进一步解决的问题和可能的发展方向，以及本文的总结。

基金项目：国家自然科学基金项目（61370186,61003140,61033010）；广东省自然科学基金项目（2016A030310018,2015A030310336）；
广东省科技计划项目（2014A010103040, 2014B01011600l）；广州市科技计划项目（2014J4100032,201510010203）；
教育部和中国移动联合研究基金（MCM2012105l）；2017 广东省教育厅“创新强校工程”特色创新类项目(2018- KJZX037)资助。

2 基于隐式反馈的协同过滤算法简介及其所面临的挑战

2.1 基于隐式反馈的协同过滤算法简介

由于基于显式评分的协同过滤算法处理的是显式评分数据，系统在收集这些显式评分数据时会提示用户输入对不同推荐对象的喜好/厌恶程度的评分，这些都需要用户的主动完全配合才能完成，因此增加了数据收集的难度和不可靠性。而基于隐式反馈的协同过滤算法所处理的隐式反馈数据不需要用户的刻意参与，隐式反馈数据只需从记录用户行为的系统日志记录中就可以获取，因此隐式反馈数据的获取更加容易，基于隐式反馈的协同过滤信息推荐系统也得到了更广泛的应用，如：网上购物（亚马逊：http://www.amazon.com，淘宝：http://www.taobao.com），在线新闻（Google 新闻系统：http://news.google.com），以及在线社交网络（http://www.facebook.com 和 http://www.twitter.com）等等。随着电子商务、在线新闻和在线社交网络的快速发展，近年来基于隐式反馈的协同过滤算法得到了工业界和学术界的广泛关注和深入研究。

当前对基于隐式反馈的协同过滤算法的研究主要分为三类，一类是基于排序学习思想的单类协同排序算法[24-28]；一类是基于评分预测的单类协同过滤算法[21-23]（基于评分预测的单类协同过滤算法又分为两类：一类是二维的基于评分预测的单类协同过滤算法；一类是三维的基于评分预测的单类协同过滤算法[29-31]；）；一类是三维的基于评分预测的单类协同过滤算法[29-31]；另一类是融合显/隐式反馈的协同过滤算法[32-33]。

2.2 基于隐式反馈的协同过滤算法所面临的挑战

目前学术界和工业界普遍认为基于隐式反馈的协同过滤算法的研究主要面临四大挑战：噪声数据问题，稀疏性问题，冷启动问题和可扩展性问题。

（1）噪声数据问题

在显式评分数据和隐式反馈数据中均存在噪声数据，但相比于给出明确评分的显式评分数据，隐式反馈产生的隐式反馈数据更容易包含噪声[26]。如我们可能看到某个用户点击了某个网页，但当用户看过这个网页后，该用户对该网页的内容可能并不感兴趣，这就将在隐式反馈数据中产生噪声数据。还譬如记录用户隐式反馈的数据库日志程序也可能出错，因而记录了错误的信息，这也会在隐式反馈数据中产生噪声数据。隐式反馈数据中的噪声数据对基于评分预测的单类协同过滤算法的影响相对较小，对基于排序学习的单类协同过滤算法的影响会更大。

例如用户 u 可能并没有点击过商品 nis，但是由于某些误操作（如：用户错误点击，或系统日志记录出错），记录用户 u 点击过商品 nis，那么对于基于评分预测的单类协同过滤算法来说，只会产生一个噪声数据点（u，nis）。但基于排序学习的单类协同过滤算法（譬如 BPRMF）来说，此时如果 neg 表示缺失数据点，那么大量的训练数据对（nis，neg）u 将被产生，但实际上应该是大量的训练数据对（pos，nis）u 被产生（pos 表示正例数据点），连锁反应式的错误配对将导致大量的噪声数据产生，而训练数据中的噪声数据将给训练过程带来巨大波动，从而导致基于排序学习的单类协同过滤算法（譬如 BPRMF）的不准确性。目前有关噪声数据对单类协同过滤算法影响的研究还比较少。

（2）稀疏性问题

稀疏性问题指的是，在协同过滤算法所采用的显式评分数据集和隐式反馈数据集中，均存在大量的评分数据缺失[21-22]。例如在包含显式评分数据的 movielens 数据集中，71567 个用户对 10681 个推荐对象仅有 10000054 个评分，稀疏性达到了 98.69%；在包含隐式反馈数据的 Last.fm 数据集中，148111 个用户对 1631028 个艺术家标注了 24296858 个标注点，稀疏性达到了 99.98%。在网页搜索和电子商务的实际应用中，考虑到隐式反馈数据的广泛存在，用户数量和推荐对象数量均极其巨大，以及相比于显式评分数据集隐式反馈数据中的信息量较小，稀疏性问题对基于隐式反馈的协同过滤算法的影响更大。因此研究更具抗稀疏性的基于隐式反馈的协同过滤算法具有重大的学术价值和应用价值。

（3）冷启动问题

冷启动问题指的是，在采用协同过滤技术的推荐系统中当出现一个新的用户，或添加一个新的推荐对象时，由于没有足够的历史评分数据，无法给新用户提供高质量的推荐服务，或者无法高质量的把新推荐对象推荐给用户。实际的信息推荐系统中，一般采用的是对新用户进行随机推荐或热门推荐；而对新推荐对象则是随机推荐给用户。近年来，学术界也提出了一些解决冷启动问题的方法。如 Kim 等人采用聚类的方法来对历史评分数据少的推荐对象进行推荐[34]；Basilico 等人则是采用综合基于内容的推荐方法来解决新推荐对象的冷启动问题[35]；Li 等人通过利用用户的上下文信息来解决新用户的冷启动问题[36]；Wang 和 Blei 通过利用推荐对象的文本信息来解决新推荐对象的冷启动问题[37]；Wang 和 Chen 等人通过利用推荐对象的文本信息以及推荐对象之间的社交网络信息来解决新推荐对象的冷启动问题[38]；Purushotham 等人通过利用推荐对象的文本信息以及用户之间的社交网络信息来解决新用户和新推荐对象的冷启动问题[39]。在现实世界的信息推荐系统中，往往都会存在大量的新用户和新推荐对象，如何给新用户推荐服务或商品，以及如何把新服务或商品推荐给用户，是基于隐式反馈的协同过滤算法的一大挑战。因此研究基于隐式反馈的协同过滤算法的冷启动问题同样具有重大的学术价值和应用价值。。

（4）可扩展性问题

可扩展性问题指的是，在互联网上实际的信息推荐系统中的用户数量和推荐对象的数量往往是数以百万计、千万计、甚至是上亿数量级的。显然，在这种情况下，要遍历所有用户或者所有推荐对象来计算最近邻，从而有效的进行推荐的代价是非常大的，计算量和计算时间也往往是无法承受的。如何有效的解决可扩展性问题也是当前协同过滤所面临的主要挑战，也是协同过滤研究的热点问题之一[34][40]。最简单的办法则是采用随机抽取的原则，或通过聚类来缩减查找的范围。Amatriain 等人采用专家的评分集合来代替群体的用户评分集合来对用户进行推荐[41]。由于用户的数量和推荐对象的数量是无法约束的，因此研究时间复杂度与用户数量或推荐对象数量线性相关的算法具有重大的学术价值和实际应用价值。

噪声数据问题，稀疏性问题，冷启动问题和可扩展性问题是基于隐式反馈的协同过滤算法所面临的四大挑战，也是主要的研究方向。学术界提出了大量的解决方法。

3 基于隐式反馈的协同过滤算法的研究现状

当前对基于隐式反馈的协同过滤算法的研究主要分为三类，一类是基于评分预测的单类协同过滤算法[21-23]（基于评分预测的单类协同过滤算法又分为两类：一类是二维的基于评分预测的单类协同过滤算法；一类是三维的基于评分预测的单类协

同过滤算法[29-31]；）；一类是基于排序学习思想的单类协同排序算法[24-28]；另一类是融合显/隐式反馈的协同过滤算法[32-33]。本节将从解决基于隐式反馈的协同过滤算法所面临的上述四大挑战的角度来介绍各类基于隐式反馈的协同过滤算法的研究现状。

3.1 基于评分预测的单类协同过滤算法的研究现状

按处理数据的维度的不同，基于评分预测的单类协同过滤算法分为两类：一类是二维的基于评分预测的单类协同过滤算法；另一类是三维的基于评分预测的单类协同过滤算法。二维的基于评分预测的单类协同过滤算法所处理的隐式数据存在高度稀疏性和不平衡性问题，不平衡性问题又会导致冷启动问题。为了解决这些问题，学者们提出了一系列方法。如 Scholkopf 等人提出运用分类的思想来解决二维的 OCCF 问题[42]。潘等人提出了加权的 ALS 算法[21][22]，该算法的核心思想是：把所有的缺失数据均当成负例数据，对这些缺失数据（负例数据）进行加权和抽样，并给这些缺失数据赋予一定的权重。Hu 等人提出了一种与潘等人提出的 ALS 算法近似的新的 OCCF 模型，称为改进的因子分解模型[23]，其核心思想是：对隐式反馈数据集中的正例数据和负例数据分别分配一个变化的信任权值。Sindhwani 等人提出了 ldNMF 模型，该模型中引入了辅助矩阵作为辅助数据，并运用非负矩阵分解技术对集合矩阵进行分解建模[43]。Zhang 等提出运用奇异值分解（SVD）技术来解决二维的基于评分预测的单类协同过滤问题[44]。以上对二维的基于评分预测的单类协同过滤算法的研究都是通过某种方法在一定程度上缓解数据稀疏性和不平衡性的影响，进而提高二维的基于评分预测的单类协同过滤算法的性能，但是不能解决新用户和新对象的冷启动问题。当前为了进一步缓解隐式反馈数据的稀疏性和不平衡性对二维的基于评分预测的单类协同过滤算法的影响及解决新用户和新对象的冷启动问题，以期进一步提高二维的基于评分预测的单类协同过滤算法的性能，除了已有的隐式反馈数据信息（隐式反馈数据矩阵）外，研究者们还提出了许多模型去进一步利用额外的数据信息：如 Li 等人提出在传统的 OC-CF 模型的基础上融入用户的上下文信息[36]以进一步提高算法的性能，该模型能够进一步缓解隐式反馈数据的稀疏性和不平衡性对二维的基于评分预测的单类协同过滤算法的影响，同时也可解决新用户的冷启动问题。Kaya 等人提出在传统的最近邻模型上融入属于特定领域的社交网络信息[45]，该模型能够进一步缓解隐式反馈数据的稀疏性和不平衡性的影响，从而进一步提高单类协同过滤推荐算法的性能，同时也可解决属于某个特定社交网络的新用户的冷启动问题。Wang 等人提出了 CTR 模型[38]，该模型在传统的矩阵分解模型上融入推荐对象的文本描述或内容信息，该模型可进一步缓解隐式反馈数据的稀疏性和不平衡性的影响，从而进一步提高单类协同过滤推荐算法的性能，同时也可解决新对象的冷启动问题。Ding 等人提出了 CSTR 模型[46]，该模型在传统的矩阵分解模型上融入推荐对象的文本描述或内容信息和推荐对象的社交网络信息，该模型可进一步缓解隐式反馈数据的稀疏性和不平衡性的影响，从而进一步提高单类协同过滤推荐算法的性能，同时也可解决新对象的冷启动问题。Purushotham 等人提出了 CTR with SMF 模型[39]，该模型在传统的矩阵分解模型上融入推荐对象的文本描述或内容信息和用户的社交网络信息，该模型可进一步缓解隐式反馈数据的稀疏性和不平衡性的影响，从而进一步提高单类协同过滤推荐算法的性能，同时也可解决新用户和新对象的冷启动问题，但是用户的社交网络信息并不是存在于所有的信息推荐系统中，故影响了该模型的广泛使用。文献中的实验结果显示：利用上述额外的数据信息能进一步有效缓解二维的基于评分预测的单类协同过滤算法所面临的隐式反馈数据的高度不平衡性和高度稀疏性问题，从而可以进一步提高二维的基于评分预测的单类协同过滤算法的性能；同时上述额外的数据信息也可在一定条件下解决新用户或新对象的冷启动问题。尽管 CTR with SMF 模型既能够在很大程度上有效缓解二维的基于评分预测的单类协同过滤算法所面临的隐式反馈数据的高度不平衡性和高度稀疏性问题，又能够同时解决新用户和新对象的冷启动问题，但由于使用该模型的特殊条件限制了该模型的广泛使用。这就需要我们研究更具通用性的新模型，使得这个新模型既能有效缓解二维的基于评分预测的单类协同过滤算法所面临的隐式反馈数据的高度不平衡性和高度稀疏性问题，又能够同时解决新用户和新对象的冷启动问题。

从评分预测和三维的角度来研究 OCCF 问题的算法又称为三维的单类协同过滤算法。与二维的单类协同过滤问题相比较，三维协同过滤问题所面临的困难更加显著，具体描述如下：[29]：

1、三维协同过滤中三个维度的对象之间的关系更加复杂。这个复杂关系不仅存在于各个三维对象的对象值之间，也存在于单个对象的对象值内部。例如对于广告投放而言，我们关心的是对象（用户、商品、广告）之间的三维关系。也即，对于一个给定的用户和这个用户浏览/购买的商品，我们的目的是预测这个用户是否或有多喜欢某个广告。所以我们需要构建一个统一的模型去模拟这三类对象和他们之间的复杂关系。

2、以二维协同过滤相比，三维协同过滤所面对的隐式反馈数据的高度不平衡性和高度稀疏性问题更加显著。每个用户只会对少量的照片贴上标签；当某个用户浏览/购买的某个商品时，一般也只投放少量的广告信息；当某个用户针对某个关键字检索查询时，搜索系统一般也只会返回少量高度相关的查询项。

目前学术界有关三维单类协同过滤算法的研究还很少：Sun 等人提出了 CubeSVD 模型，该模型通过在高维矩阵分解中引入 SVD 模型来研究三维协同过滤问题[29]；文献[47]扩展了 Hofmann 模型，使其适用于三维协同过滤问题；其它有关三维单类协同过滤算法的研究，大多仍采用二维的单类协同过滤算法的的研究思路，通过收集、提取用户的偏好信息来产生推荐[48]。

三维单类协同过滤（立方体填补）在互联网中得到了广泛应用，如：个性化广告投放、个性化网页搜索、社会化书签和社会化标注等。因此提出性能更好的三维单类协同过滤算法，以期更好地解决三维协同过滤问题所面临的挑战，具有极大的学术价值和实际应用价值。

3.2 基于排序学习的单类协同过滤算法的研究现状

在信息检索领域，通过排序学习来提高检索算法的性能得到了广泛关注和快速发展。近年来出现了一些学者提出运用信息检索领域排序学习的思想来研究 OCCF 问题。其中最经典的是 Rendle 等人提出的 BPR 模型，该模型运用基于贝叶斯（Bayesian）成对损失函数从排序学习的角度研究 OCCF 问题。为了可否 BPR 模型的不足，学者们又提出了一些 BPR 模型的扩展模型[49-51]。如 Pan 等人提出了 GBPR 模型，该模型放弃了 BPR 模型中有关用户之间相互独立的假设，一个用户的偏好受

与其相似的同一个兴趣组的其他用户的影响。Du 等人提出了 UGPMF 模型，该模型的核心思想是在 BPR 模型中融入用户信息(如用户的社交网络信息、标签信息等)以进一步提高单类协同排序算法的性能。Yang 等人提出了 CCF 模型用于新闻推荐[52],该模型在扩展 BPR 模型的基础上引入额外的用户选择的上下文信息以进一步提高排序算法的性能。Kanagal 等人提出了 TF 模型用于电子商务中的商品推荐[53],该模型在扩展 BPR 模型的基础上运用分类学原理来学习用户的购物行为(TF 模型需要额外的用户购物行为的分类信息)，以进一步提高协同排序算法的性能。Tang 等人提出了 BPR-CMF 模型[54],该模型在 BPR 模型的基础上融合多领域数据信息,以进一步提高二维单类协同排序算法的性能。以上所述的 BPR 的扩展模型大都是在传统的 BPR 模型中引入额外的数据信息(社交网络、文本描述、标签等)，以进一步有效缓解 OCCF 所面临的数据的高度不平衡性和数据的高度稀疏性问题，进而提高单类协同排序算法的性能。GBPR 模型是上述 BPR 扩展模型中唯一不需要外部辅助信息的单类排序推荐模型。

当前出现了通过直接优化排序学习的评价指标来进一步提高二维单类协同排序算法的性能的研究。如 Shi 等人提出了 CLiMF 模型[24],该模型通过直接优化评价指标 MRR 来建立新的协同排序模型，在直接优化评价指标 MRR 的基础上来对推荐对象进行排序。由于 CLiMF 模型会过度拟合 MRR 评价指标,而 MRR 评价指标主要适用于评价给用户推荐 3-5 个最重要的推荐对象的算法,因此影响了该算法的通用性。

3.3 融合显/隐式反馈的协同过滤算法的研究现状

当前单类协同过滤算法的另一研究方向是融合显/隐式反馈的协同过滤算法,据查资料,该类算法有 Koren 等人提出了 SVD++算法[55],Li 等人提出了 MERR SVD++[32],Liu 等人提出了 Co_Rating[33],该算法是基于评分预测的融合显/隐式反馈的协同过滤算法。

4 结论与展望

基于隐式反馈的协同过滤算法研究仍有待进一步深入与广泛地展开,下一步研究工作主要包括:

(1) 未来可以研究具体优化某个评价标准 (如:ERR、NDCG、AUC 和 MAP)的三维单类协同排序算法,因为目前据全面查寻相关资料,还没有发现研究优化某个具体评价标准的三维单类协同排序算法,相信这方面的研究一定会产生重大的学术价值和实际应用价值。

(2)本文所介绍的各类算法的并行化也是值得进一步研究的重点问题。可以研究设计本文所介绍的各类算法在大数据集下的分布式/并行化实现。譬如运用 OpenMP、CUDA、MPI、MapReduce 这四种分布式/并行计算方式在大数据集上实现本文所介绍的各类算法。

(3)基于隐式反馈的协同过滤算法在互联网上的信息推荐领域得到了广泛的应用。本文所介绍的各类算法在互联网领域的实际应用研究,特别是工程上的实现问题也是未来研究的一大重点课题。

参考文献：

[1]印鉴,王智圣,李琪,苏伟杰. 基于大规模隐式反馈的个性化推荐[J].软件学报,2014,25(9) :1953-1966.

[2]张燕平,张顺,钱付兰,张以文. 基于用户声誉的鲁棒协同推荐算法[J]. 自动化学报, 2015, 41(5): 1004-1011.

[3]Adomavicius G, Tuzhilin A. Toward the next generationof recommendersystems: A survey of the state-of-the-art and possible extensions [J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6): 734-749.

[4]吴金龙.Netflix Prize 中的协同过滤算法(D).北京:北京大学,2010: 1-100.

[5]罗辛,欧阳元新,熊璋,袁满.通过相似度支持度优化基于 K 近邻的协同过滤算法[J].计算机学报, 2010, 33 (8) :1437-1445.

[6]陈健,印鉴.基于影响集的协作过滤推荐算法[J].软件学报,2007,18 (7) :1685-1694.

[7]项亮. 推荐系统实践 [M]. 北京: 人民邮电出版社, 2012:1-200.

[8]李聪, 骆志刚. 用于鲁棒协同推荐的元信息增强变分贝叶斯矩阵分解模型 [J]. 自动化学报, 2011, 37(9): 1067-1076.

[9]Mehta B, Hofmann T, Nejdl W. Robust collaborative filtering [C]. In: Proceedings of the 2007 ACM Conference on Recommender Systems. New York, USA: ACM, 2007: 49-56.

[10]Gunes I, Kaleli C, Bilge A, Polat H. Shilling attacks against recommender systems: a comprehensive survey [J]. Artificial Intelligence Review, 2014, 42(4): 767-799.

[11]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, 19(1): 1-15.

[12]Li R H, Yu J X, Huang X, Cheng H. Robust reputation-based ranking on bipartite rating networks [C]. In: Proceedings of the 2012 SDM International Conference on Data Mining. Hong Kong: SIAM, 2012: 612-623.

[13]Koren Y, Robert B, Chris V. Matrix factorization techniques for recommender systems [J]. Computer, 2009, 42(8): 30-37.

[14]Huber P J, Ronchetti E M. Robust Statistics (Second Edition) [M]. New Jersey: John Wiley and Sons, 2009. 149-199.

[15]Mobasher B, Burke R, Bhaumik R, Williams C. Toward trustworthy recommender systems: an analysis of attack models and algorithm robustness [J]. ACM Transactions on Internet Technology, 2007, 7(4): 1-40.

[16]http://zh.wikipedia.org/wiki/长尾.

[17]安德森.长尾理论 2.0[M].北京:中信出版社,2009:200-300.

[18]Netflix. https://www.netflix.com/global.

[19]Bell R, Koren Y, Volinsky C.The bellkor 2008 solution to the netflix prize [R]. NewYork:2008: 1-10.

[20]Bell R, Koren Y, Volinsky C.The bellkor solution to the netflix prize [R]. NewYork:2007:1-10.

[21]Pan R, Zhou Y H, Cao B, Liu N N, Lukose R, Scholz M, Yang Q. One-class collaborative filtering [C]. In: Proceedings of the IEEE International Conference on Data Mining. Pisa, Italy: IEEE, 2008: 502-511.

[22]Pan R, Scholz M. Mind the gaps: weighting the unknown in large-scale one-class collaborative filtering [C]. In: proceedings of the 22nd International Conference on Knowledge Discovery and Data Mining. Paris, France: ACM, 2009: 667-676.

[23]Hu Y, Koren Y, Volinsky C. Collaborative filtering for implicit feedback datasets [C]. In: Proceedings of the IEEE International Conference on Data Mining. Pisa, Italy: IEEE, 2008: 263-272.

[24]Shi Y, Karatzoglou A, Baltrunas L. CLiMF: Collaborative Less-Is-More Filtering [C]. In: Proceedings of the Twenty-third International Conference on Artificial Intelligence. Beijing, China: ACM, 2013: 3077-3081.

[25]Rendle S, Freudenthaler C, Gantner Z, Schmidt-Thieme L. BPR: Bayesian personalized ranking from implicit feedback [C]. In: Proceedings

of the 22nd International Conference onUncertainty in Artificial Intelligence. Montreal, Canada, 2009: 452−461.

[26]**李改,李磊. 鲁棒的单类协同排序算法**[J]. **自动化学报**, 2015, 41 (2): 405−418.

[27]Li G, Ou W H. Pairwise Probabilistic Matrix Factorization for Implicit Feedback Collaborative Filtering [J]. Neurocomputing, 204, 2016: 17−25.

[28]Li G, Wang L Y, Ou W H. Robust Personalized Ranking from Implicit Feedback [J]. International Journal of Pattern Recognition and Artificial Intelligence, 30(1), 2016: 1−28.

[29]Sun J T, Zeng H J, Liu H, Lu Y, Chen Z. Cubesvd: a novel approach to personalized web search [C]. In: Proceedings of the 16th International Conference on World Wide Web. Chiba, Japan: ACM, 2005: 382−390.

[30]**李改,潘嵘,李磊**. CubeALS_**新的三维协同过滤推荐算法**[J]. **计算机科学与探索**, 2012, 6(2): 156−164.

[31]Fang X M and Pan R. Fast DTT −− A Near Linear Algorithm for Decomposing a Tensor into Factor Tensors [C]. In: Proceedings of the 20th International Conference on Knowledge Discovery and Data Mining. Las Vegas: ACM, 2014: 967−976.

[32]Li G and Chen Q. Exploiting Explicit and Implicit Feedback for Personalized Ranking [J]. Mathematical Problem in Engineering, 2016: 1−11.

[33]Liu N N, Xiang E W, Zhao M, and Yang Q. Unifying explicit and implicit feedback for collaborative filtering [C]. In: Proceedings of the 19th International Conference on Information and Knowledge Management (CIKM '10), Toronto, Canada: ACM, 2010: 1445 ⊢1448.

[34]Kim B M, Li Q, Kim J W, Kim J. A New Collaborative Filtering System Addressing Three Problems [C]. In: Proceedings of the 8th Pacific Rim International Conference on Artificial Intelligence. AuckLand, New Zealand: ACM, 2004: 495−504.

[35]Basilico J, Hofmann T. Unifying Collaborative and content−based Filtering [C]. In: Proceedings of the 3th International Conference on Machine Learning. Banff, Alberta, Canada: ACM, 2004: 112−118.

[36]Li Y N, Zhai C X, Hu J, Chen Y. Improving one−class collaborative filtering by incorporating rich user information [C]. In: Proceedings of the 19th ACM Intenational Conference on Information and Knowledge Management. New York, USA: ACM, 2010: 959−968.

[37]Wang C, Blei D M. Collaborative topic modeling for recommending scientific articles [C]. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Diego, CA: ACM, 2011: 448−456.

[38]Wang H, Chen B Y, Li W J.Collaborative Topic Regression with Social Regularization for Tag Recommendation [C]. In: Proceedings of the Twenty−third International Conference on Artificial Intelligence. Beijing, China: ACM, 2013: 2719−2725.

[39]Purushotham S, Liu Y, Kuo C. Collaborative topic regression with social matrix factorization for recommendation systems [C]. In: Proceedings of the 29th ACM Intenational Conference on Machine Learing. Edinburgh, Scotland, UK: ACM, 2012: 1255−1265.

[40]Xue G R, Lin C X, Yang Q, Xi W S, Zeng H J, Yu Y. Scalable Collaborative Filtering Using Cluster−based Smoothing [C]. In: proceedings of the 19nd International Conferenceon Research and Development in Information Retrieval. Salvador, Brazil: ACM, 2005: 114−121.

[41]Amatriain X, Laithia N, Pujol J M. The Wisdom of the few [C]. In: proceedings of the 23nd International Conferenceon Research and Development in Information Retrieval. Boston, Massachusetts, USA: ACM, 2009: 532−539.

[42]Scholkopf B, Platt J, Shawe−Taylor J, Smola A, Williamson R. Estimating the support of a high−dimensional distribution [J]. Neural Computation, 2001, 13(7): 1443−1471.

[43]Sindhwani V, Bucak S S, Hu J, Mojsilovi A. One−class matrix completion with low−density factorizations [C]. In: Proceedings of the IEEE International Conference on Data Mining. Washington, DC, USA: IEEE, 2010: 1055−1060.

[44]Zhang S, Wang W H, Ford J, Makedon F, Pearlman J. Using singular value decomposition approximation for collaborative filtering [C]. In: Proceedings of the 7th IEEE International Conferenceon on E−Commerce. München, German: IEEE, 2005: 257−264.

[45]Kaya H, Alpaslan F. Using social networks to solve data sparsity problem in one−class collaborative filtering [C]. In: Proceedings of the 7th IEEE Intenational Conference on Information Technology. Ankara, Turkey: IEEE, 2010: 249−252.

[46]Ding X T, Jin X M, Li Y J, Li L H.Celebrity Recommendation with Collaborative Social Topic Regression [C]. In: Proceedings of the Twenty−third International Conference on Artificial Intelligence. Beijing, China: ACM, 2013: 2612−2618.

[47]Lathauwer L D, Moor B D, Vandewalle J. A multilinear singular value decomposition [J]. SIAM Journal on Matrix Analysis and Applications. 2000, 21(4): 1253−1278.

[48]Haveliwala T H. Topic−sensitive pagerank [C]. In: Proceedings of the 16th International Conference on World Wide Web. Honolulu, Hawaii, USA: ACM, 2002: 517−526.

[49]Pan W K, Chen L. GBPR: Group Preference based Bayesian Personalized Ranking for One−Class Collaborative Filtering [C]. In: Proceedings of the Twenty−third International Conference on Artificial Intelligence. Beijing, China: ACM,2013: 3007−3011.

[50]Rendle S, Schmidt−Thieme L. Pairwise interaction tensor factorization for personalized tag recommendation [C]. In: Proceedings of the 3rd ACM international conference on web search and data mining. New York, USA: ACM, 2010: 81−90.

[51]Du L, Li X, Shen Y.User graph regularized pairwise matrix factorization for item recommendation [C]. In: Proceedings of the 7th International Conference on advanced data mining and applications. Berlin, German: ACM, 2011: 372−385.

[52]Yang S, Long B, Alexander J, Zha H, Zheng Z. Collaborative competitive filtering: learning recommender using context of user choice [C]. In: Proceedings of the 34th ACM international conference on research and development in information retrieval. Beijing, China: ACM, 2011: 295−304.

[53]Kanagal B, Ahmed A, Pandey S, Josifovski V, Yuan J, Garcia−Pueyo L. Supercharging recommender systems using taxonomies for learning user purchase behavior [C]. In: Proceedings of the VLDB Endowment. Istanbul, Turkey: ACM, 2012: 956−967.

[54]Tang J, Yan J, Ji L, Zhang M, Guo S D, Liu N, Wang X F, Chen Z. Collaborative users' brand preference mining across multiple domains from implicit feedbacks [C]. In: Proceedings of the Twenty−Fifth AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI, 2011: 477−482.

[55]KOREN Y. Factorization Meets the Neighborhood: a Multifaceted Collaborative Filtering Model [C]. In: Proceedings of the 25th International Conference on Knowledge Discovery and Data Mining. Las Vegas: ACM, 2008: 426−434.

(**下转第 9 页**)

开发推出的一款通过网络快速发送语音短信、视频、图片和文字的手机聊天软件，可支持多人群聊。用户可以通过微信与好友进行形式上更加丰富的类似于短信、彩信等方式的联系。截至2018年6月我国微信朋友圈的使用率为86.9%[1]。

三、社交媒体的主要特征

1、互动性强

互动与参与是社交媒体的显著特色。人人都能参与信息的生产和传递，例如用户在其微信朋友圈所发的个人信息很快就能被好友浏览、点赞、评论和分享，又如美国国家档案馆Facebook主页日均活跃用户16617个，You Tube上的视频月均观看量392495次[4]。

2、内容越来越丰富

从二十世纪八十年代BBS进行的文字交流开始，信息载体形式越来越多样化，图形图片、音乐、视频等逐一上场。贴吧、论坛、空间、群、朋友圈、直播间等多种形式并存。尤其近年来的媒介融合飞速发展，促进了在传播内容方面更丰富化和多元化，传播速度也更加快捷，信息传播范围更加广泛。

3、关注的人群众多

多种社交媒体应用拥有的粉丝量众多。2018年微博拥有用户33741万人，电子邮件用户30556万人[1]。2016年近一半的网络社交用户通过社交应用获取新闻，84.4%的用户通过社交应用观看网络电视节目[2]。

四、社交媒体的分类

中国互联网络信息中心在2016年中国社交应用用户行为研究报告中，将社交应用产品进行了如下分类。(1)即时通信工具(QQ、微信等)；(2)综合社交应用(QQ空间、微信朋友圈、新浪微博等)；(3)垂直社交应用(婚恋社交、社区社交、职场社交等)[5]。

凯度(KANTAR)在2017凯度中国社交媒体影响报告中，将社交媒体分为九类，即微信、微博、交友类社交媒体、通讯类社交媒体、论坛类社交媒体、生活类社交媒体、带有社交评论功能的新闻类媒体、带有社交评论功能的电商类媒体、带有社交评论功能的视频或网络平台[6]。

五、社交媒体的发展新趋势

1、社交媒体移动化，进一步加快了社交应用的广泛传播

截至2018年6月我国网民人数已经达到8.02亿，网民通过手机入网的比例达到98.3%[1]，移动互联网已经成为用户最重要的信息交流平台。通过手机可以轻松上网查看所关注的APP信息，2018年我国微博用户规模已经达到3.37亿。移动社交类APP运营商竞争激烈，出产多种社交应用APP，用户规模也不断攀升。

2、社交媒体应用的全民化，促进了传统媒体和新媒体的进一步融合

社交等新媒体的蓬勃发展，用户量上亿规模，带动了传统媒体的改革，使得传统媒体积极拥抱社交网络。企业和政府充分利用微博或微信公众号、在线社区、百科或者其他互联网平台媒体来进行营销、公共关系和客户服务维护，扩大传播范围，增强舆论声势。

3、短视频异军突起

随着手机网速、流量资费等不断优化，网民娱乐的碎片化，短视频现在也像文字、图片一样快速流转开来。通过短视频，大众年轻人可以轻松自在地表达自我，很快便得到了互联网用户尤其是年轻人的广泛关注。短视频(诸如抖音等)社交应用快速兴起，2018年全国使用短视频应用的网民规模有74.1%[1]。

4、社交媒体智能化新发展

2016年Skype、Line、Facebook的Messenger都引入了聊天机器人程序，BBC、《纽约时报》等传统媒体机构，也开始以机器人对话的方式将新闻资讯呈现给大众。大数据、无人机、机器人、人工智能、虚拟现实等高新技术研发成果不断应用到传媒业，使“智媒”成为未来媒体发展的一种主要趋向[7]。

参考文献：

[1]中国互联网络信息中心.第42次《中国互联网络发展状况统计报告》[EB/OL].http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201808/P020180820630889299840.pdf，2018-08-20.

[2]中国互联网络信息中心.第41次《中国互联网络发展状况统计报告》[EB/OL].http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201803/P020180305409870339136.pdf，2018-03-05.

[3]谭天. 新媒体概论(第二版)[M]，暨南大学出版社，2016-11-01.

[4]蔡明娜.美国国家档案馆社交媒体应用现状分析与启示[J]，浙江档案，2018-8-31.

[5]中国互联网络信息中心. 2016年中国社交应用用户行为研究报告[EB/OL].http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/sqbg/201712/P020180103485975797840.pdf

[6]搜狐. 2017凯度中国社交媒体影响报告[EB/OL].https://www.sohu.com/a/147589678_291947.

[7]苏涛，彭兰. “智媒”时代的消融与重塑——2017年新媒体研究综述[J]，国际新闻界，2018-01.

作者简介：

吴春颖(1977-)，女(汉族)，河北省固安县人，副教授，硕士，主要研究方向为计算机应用、计算机信息安全(775251252@qq.com)。

(上接第5页)

作者简介：

李改(1981-)，男(汉族)，湖北省松滋市，副教授，博士，主要研究方向为数据挖掘、推荐系统和大数据处理，E-mail：ligai999@126.com；邹小青(1964-)，男(汉族)，江西省赣州市，讲师，本科，主要研究方向为系统集成和数据库。