

文章编号: 2095-2163(2021)07-0228-06

中图分类号: TP393.01

文献标志码: A

# 推荐系统综述

赵 岩, 刘宏伟

(哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001)

**摘要:** 推荐系统是信息过滤系统领域的一个重要研究方向。随着信息技术的发展,推荐系统在提升用户体验和增加企业效益等方面发挥着越来越重要的作用。主流的推荐系统大多基于矩阵分解模型和深度学习模型,近年来又提出了基于记忆网络和集成学习的推荐系统为用户精确地推荐物品。本文将对基于矩阵分解、基于深度学习、基于记忆网络和基于集成学习的推荐系统进行分析 and 总结,展望未来的研究方向。

**关键词:** 推荐系统; 矩阵分解; 神经网络; 记忆网络; 集成学习

## A survey on recommender systems

ZHAO Yan, LIU Hongwei

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

**[Abstract]** Recommender systems are an important research direction in the field of information filtering systems. With the development of information technology, recommender systems play an increasingly important role in improving user experience and increasing enterprise revenue. Most of the mainstream recommendation systems are based on matrix factorization models and deep learning models. In recent years, researchers have proposed recommender systems based on memory networks and ensemble learning to accurately recommend items for users. This paper summarizes and introduces matrix factorization based, deep learning based, memory network based, and ensemble learning-based recommender systems, and discusses future research directions.

**[Key words]** recommender system; matrix factorization; neural network; memory network; ensemble learning

## 0 引言

随着信息技术的飞速发展,人们在购物和浏览信息方面有了更多的选择。但众多的选择也让用户很难从中挑选出符合自己偏好的物品或信息,从而降低了用户体验。根据 Netflix 中的数据,如果用户无法在 60~90 s 内找到自己偏好的电影,其将失去观看电影的偏好。研究者们对推荐系统进行了大量的研究和广泛的部署,以便帮助用户快速、准确地选择物品。

推荐系统是一种信息过滤系统<sup>[1]</sup>,通过学习和预测用户的偏好,向用户推荐符合用户偏好的物品。目前,推荐系统在电子商务、在线观看和社交网络等许多领域中得到了广泛的应用,并为许多行业带来了巨大的效益。以电商行业为例,推荐系统为淘宝购物平台带来了 50% 的收益和网络流量;为著名购物平台 Amazon 带来了 35% 的销售额。2017 年,中国人工智能协会和罗兰贝格共同发布了《中国人工智能创新应用白皮书》,预计推荐系统将为零售业带来约 4 200 亿元的降本增值。另外,由清华大学

全球工业研究院等在 2019 年发布的《工业智能白皮书》指出,推荐系统能够创造用户价值,具有不断扩散的用户基础,具有广阔的发展空间。

推荐系统已经得到了广泛的研究和部署,其目的是通过诸如购买等交互数据向用户推荐物品<sup>[2]</sup>。基于推荐机制,推荐系统被划分为以下 3 类:(1) 基于内容的推荐系统<sup>[3]</sup>,(2) 基于协同过滤的推荐系统<sup>[4]</sup>,(3) 融合以上两者的混合型推荐系统<sup>[5]</sup>。基于内容的推荐系统通过分析用户的画像和物品的特征,从而为用户做出推荐。这类推荐系统可以很好地解决冷启动问题,但只能为用户推荐与其交互过的物品相似物品,在推荐的多样性方面存在一定的局限性。与之不同的是,基于协同过滤的推荐系统,对目标用户以及和目标用户相似的用户历史行为进行分析,从而根据用户的偏好为用户推荐物品<sup>[6]</sup>。与基于内容的推荐系统相比,这种基于协同过滤的推荐系统一般可以得到更为精确的推荐,并且能够更好地应对多种场景、应用多种先进技术,从而得到更深入的研究和更广泛的应用。

当前主流基于协同过滤的推荐系统,大多采用

**作者简介:** 赵 岩(1991-),男,博士研究生,主要研究方向:推荐系统、人工智能、云计算;刘宏伟(1971-),男,博士,教授,博士生导师,主要研究方向:容错计算、移动计算、云计算。

收稿日期: 2021-03-24

哈尔滨工业大学主办 ◆ 科技创新与应用

矩阵分解模型和深度学习模型。这些模型通常首先学习用户嵌入和物品嵌入,然后根据用户嵌入和物品嵌入做出推荐。具体而言,基于矩阵分解的推荐系统利用线性运算(如用户嵌入与物品嵌入间点积等),从交互数据中学习用户的偏好和物品的特征。而基于深度学习的推荐系统,通常采用多层神经网络以实现更为精确的推荐。此外,近年来也出现了基于记忆网络和集成学习等先进集成学习方法的推荐系统。基于以上背景,本文将对基于矩阵分解、基于深度学习、基于记忆网络和基于集成学习的推荐系统做详细的介绍和分析。

## 1 基于矩阵分解的推荐系统

本节通过两部分介绍基于矩阵分解的推荐系统。即不依赖辅助信息的基于矩阵分解的推荐系统和依赖辅助信息的基于矩阵分解推荐系统,见表1。

表1 基于矩阵分解的推荐系统

Tab. 1 Matrix factorization based recommender systems

分类	代表性工作	应用的技术或主要思想
不依赖辅助信息	文献[7]	奇异值分解
	文献[8]	概率论
	文献[9]	个性化偏置
	文献[10]	权重分解
	文献[13]	贝叶斯优化和个性化排序
依赖辅助信息	文献[15]	社交约束
	文献[16]	社交约束
	文献[17]	贝叶斯优化

### 1.1 不依赖辅助信息基于矩阵分解的推荐系统

奇异值分解(Singular Value Decomposition, SVD)<sup>[6]</sup>是一种用于矩阵分解的数学方法。受奇异值分解的启发,Simon Funk提出了FunkSVD<sup>[7]</sup>为用户推荐物品。FunkSVD首先将一个评分矩阵分解为两个低阶矩阵,分别包含用户的嵌入和物品的嵌入。与原版的奇异值分解方法相比,FunkSVD不仅时间复杂度较低,而且可以适用于用户与物品交互矩阵中常见的、具有缺失值的矩阵。之后,研究者们基于矩阵分解方法提出了更为先进的推荐模型。

例如:Salakhutdinov提出了概率矩阵分解(Probabilistic Matrix Factorization, PMF)<sup>[8]</sup>,利用概率理论提高推荐精确度。PMF用高斯分布初始化用户的嵌入和物品的嵌入,根据这些初始化好的嵌入学习用户偏好和物品特征,最后用这些学习到的用户偏好和物品特征进行推荐。此外,文献[8]中的工作将PMF扩展为约束PMF,即假设具有相似评分物品的用户可能具有相似的偏好。其实验结果表明,这种约束PMF在Netflix数据集上更加有效,表现良好。

此外,文献[9]还在原有的矩阵分解模型中增加了偏置向量,从而在推荐时也考虑了目标用户和目标物品的个性化特征。该方法首先计算了数据集的总体平均评分以及目标用户和目标物品的评分偏差,然后将这3个值相加,作为矩阵分解过程的正则项。通过这种方法,预测的评分能够针对目标用户和目标物品进行调整,从而达到更加准确。之后,文献[9-10]改进了矩阵分解模型,使其能够通过购买、浏览和点击等隐性行为为用户推荐物品。与评分等显性行为相比,隐性行为更常见,也更容易收集。因此,应该对依靠隐性行为的推荐系统进行深入研究。然而,这是一项具有挑战性的任务,因为隐性行为并不能明确地反映用户的偏好。例如,用户可能会在观看一个视频后发现对该视频不感兴趣。为了解决这个问题,文献[10-11]都提出了加权矩阵分解模型。该模型为每个交互分配一个权重,以表示目标用户对目标物品的偏好。此外,文献[10]的工作还采用了负采样(negative sampling)的方法,利用没有观察到的用户-物品交互,有效地训练矩阵分解模型。这种负采样的方法,有利于提高隐性行为推荐的准确度,并在文献[11-12]等推荐模型中被广泛采用。

以上提到的推荐系统都是元素级(element-wise)推荐系统。即通过最小化真实标签或值与相应预测值的差异训练推荐模型。与这些元素级推荐系统不同,文献[13]提出了贝叶斯个性化排名方法(Bayesian Personalised Ranking, BPR)。该方法通过元组级(pair-wise)的方法学习用户偏好,从而实现精准推荐。该方法假设用户对交互过的物品比对自己遗漏的物品有更多的偏好,先对<目标用户、交互物品、遗漏物品>的元组进行采样,然后通过贝叶斯优化方法<sup>[14]</sup>,利用这些元组训练矩阵分解模型。训练目标使得学习到的用户,对其交互物品的偏好大于其对遗漏物品的偏好。实验证明,BPR对于训练推荐模型是有效的,但在训练过程中引入了更多的时间复杂性。

### 1.2 依赖辅助信息基于矩阵分解的推荐系统

虽然上述研究增强了矩阵分解模型学习用户偏好和物品特征的能力,但数据稀疏性问题一直影响着推荐的精确度。为此,一些研究者提出采用辅助信息来解决这一数据稀疏性问题。例如,文献[5]提出了一种带有社交约束的推荐系统,利用用户之间的社交网络信息来提高推荐的精确度。文中假设,在社交网络中具有相似历史评分的好友用户会对物品有相似的偏好。基于这一假设,此工作利用用户与其好友之间的嵌入差异的 $L_2$ 范数作为正则

项,更有效地训练矩阵分解模型。此外,还将这些正则项与目标用户及其对应好友的相似度相乘。通过这种方法,与目标用户有更多相似偏好的好友将对正则项做出更多贡献。

除了利用相似偏好的好友的嵌入作为正则项之外,文献[16]还利用不同偏好好友的嵌入和物品社交关系,进一步提高推荐的精确度。文中假设,学习到的目标用户的嵌入和不相似好友的嵌入是多样化的,因此应该最大化两者之间的差异。此外,利用物品之间的评分相似性以便表示物品之间的隐性社会关系。然后,与用户社会关系类似,这种隐性的物品社会关系也被作为正则项,以便更有效地训练矩阵分解模型。

除社交约束外,一些推荐系统还加入了如用户画像和物品属性等其它辅助信息。例如,文献[17]提出了一种基于贝叶斯优化的矩阵分解模型,以便较好地利用这些辅助信息。作者首先用向量表示给定的辅助信息,然后将这些向量纳入到对应用户或物品嵌入的先验条件中,之后采用 Gibbs 采样器,根据这些用辅助信息增强的嵌入进行最终的预测。文献[18]对这一方法进行了进一步的改进,采用了层次结构,可以更充分地利用辅助信息,更好地辅助推荐。

虽然经过不断的努力,但基于矩阵分解的推荐系统普遍受到线性运算和浅层结构的限制,因此难以准确学习非线性和复杂的用户与物品关系,对进一步提高推荐精确度存在一定难度。

## 2 基于深度学习的推荐系统

与基于矩阵分解的推荐系统相比,基于深度学习的推荐系统更有效。这种有效性部分来自于其非线性操作、深层结构以及适应不同推荐场景的高灵活性。本文中,将介绍两部分基于深度学习的推荐系统:(1)不依赖辅助信息的基于深度学习的推荐系统;(2)依赖辅助信息的基于深度学习的推荐系统。

表 2 基于深度学习的推荐系统

Tab. 2 Deep learning based recommender systems

分类	代表性工作	应用的技术或主要思想
不依赖辅助信息	文献[21]	自动编码器
	文献[22]	深度-广度学习
	文献[23]	双翼多层感知器
	文献[24]	深度-广度学习
	文献[25]	卷积神经网络
依赖辅助信息	文献[26]	卷积神经网络
	文献[27]	循环神经网络
	文献[28]	图神经网络

### 2.1 不依赖辅助信息的基于深度学习的推荐系统

文献[19]中提出了基于自动编码器的推荐系统(AutoRec),用无监督的自动编码器为用户推荐物品。该方法给定一个评分矩阵,将行(或列)作为基于用户(或基于物品)的自动编码器,输入学习隐含状态,然后重建这些行(或列)。通过这种方法,缺失的评分就由重构后的行(或列)进行预测。文献[19]中的方法,证明了深度结构有利于实现更高的推荐精确度。此外,该方法在文献[20-21]中得到了增强,使其更加稳定和有效。在此之后,研究者们提出了基于多层感知器(Multi-layer Perception, MLP)的推荐系统,并取得了较好的推荐精确度。例如,文献[22]中提出了一种基于广度和深度学习的推荐系统,采用多层感知器学习用户和物品的复杂、非线性特征。此外,除了这个深层结构外,该方法还采用了一种广层结构记忆原始输入的特征,将深层结构和广层结构相互融合以便相互加强,从而实现精确推荐的目的。

文献[23]提出了深度矩阵分解模型(Deep Matrix Factorisation, DMF),采用双翼多层感知器结构,为用户精确地推荐物品。该模型采用两个多层感知器,分别精确地学习用户的偏好和物品的特征。然后,将这些学习到的用户偏好和物品特征通过余弦相似度函数进行融合,从而预测用户偏好和物品特征的匹配程度。通过这种方式,在更准确地学习到用户偏好和物品特征的情况下,深度矩阵分解模型有望实现更高的推荐精确度。

文献[24]提出了神经矩阵分解模型(Neural Matrix Factorization, NeuMF),以便进一步提高推荐的精确度。该模型采用全连接的神经网络层,代替线性点积操作,以便对原有的矩阵分解模型进行增强。然后,将该增强的矩阵分解模型的输出向量与多层感知器模型输出的向量进行融合,以便更好地模拟复杂的用户与物品关系,从而为用户提供准确的推荐。

### 2.2 依赖辅助信息基于深度学习的矩阵分解模型

与基于矩阵分解的推荐系统类似,基于深度学习的推荐系统也采用了辅助信息,以提高重组精确度。此外,基于深度学习的推荐系统灵活地构建了多种类型的神经网络结构。如,卷积神经网络(Convolutional Neural Network, CNN)、循环神经网络(Recurrent Neural Network, RNN)和图神经网络(Graph Neural Network, GNN)等,以便有效地利用不同类型的辅助信息。

例如:卷积矩阵分解(Convolutional Matrix Factorization, ConvMF)<sup>[25]</sup>,通过卷积神经网络,利用用户对物品的评论信息,缓解了数据稀疏性问题,从而提高了推荐精确度。该方法首先通过卷积神经网络,将用户对物品的评论嵌入其中,然后利用这些评论的嵌入,很好地初始化相应物品的嵌入。最后,采用基于概率的矩阵分解模型,根据这些初始化好的物品嵌入和随机初始化的用户嵌入进行推荐。

文献[26]提出的深度合作神经网络(Deep Cooperative Neural Networks, DeepCoNN)利用评论信息初始化相应用户的嵌入,然后将这些初始化的用户和物品的嵌入进行串联,再送入全连接层预测用户对物品的偏好。与卷积矩阵分解相比,深度合作神经网络更好地利用了评论信息,以便精确地学习用户偏好,可以实现更高的推荐精确度。

循环神经网络已经被推荐系统广泛地利用以便进行精确的序列推荐。例如,文献[27]提出了一种基于会话的推荐系统。该推荐系统采用循环神经网络,以便捕获会话中物品之间的顺序依赖关系。循环神经网络最初是为了学习序列数据中嵌入的特征而提出的,因此适用于从会话中的物品序列中学习其依赖性,从而实现基于会话的推荐。近些年来,研究者们又提出了许多其它基于循环神经网络的用于会话场景的推荐系统。

此外,研究者们也提出了基于图神经网络的推荐系统,以便更有效地利用物品之间的转换以进一步提高基于会话的推荐的精确度。例如,文献[28]中将每个会话建模为一个有向子图,将所有子图合成为包含所有会话信息的完整会话图,利用门控图神经网络捕获物品之间的转换和依赖关系,从而实现基于会话的精准推荐。

### 3 基于记忆网络的推荐系统

记忆网络<sup>[29]</sup>是一种有效的机器学习模型。该模型最初在文献[29]中提出,并在文献[30]中进一步增强为端到端的模型。记忆网络的有效性已经在自然语言处理、图像处理和模式识别等多种领域得到了广泛的验证。

近年来,研究者们挖掘了记忆网络在用于推荐领域的潜力。例如文献[31]设计了一种双翼记忆网络以及分级注意机制,在类似推特(Twitter)的应用中执行提及推荐。该工作采用了两个记忆网络分别存储作者和候选用户的推特历史,利用一个词级编码器和一个句子级编码器,分别从作者和候选用

户对应的推特历史中,协同嵌入作者和候选用户的偏好。值得注意的是,这两个编码器都依赖于注意力机制。最后,利用学习到的用户偏好,通过判断哪些候选用户与作者相匹配,以便进行提及推荐。后来,文献[32]中的工作进一步改进了基于记忆的提及推荐。此外,研究者们也通过记忆网络,提高了针对类似推特应用的哈希标签推荐的精确度。

后来,Dong等人<sup>[33]</sup>利用记忆网络有效地解决了推荐中长期存在的冷启动问题。该工作采用了特征特定的记忆网络和任务特定的记忆网络,以个性化的方式,辅助初始化用户的嵌入,并指导更有效的预测。通过利用特征特定记忆网络,记忆现有用户的资料 and 嵌入,可以在参考现有用户嵌入的同时,根据新用户的资料初始化其嵌入,从而更准确地反映这些新用户的偏好。至于记忆用户快速梯度的任务特定记忆网络,则是通过学习所有用户的交互,以便较好地初始化预测参数,从而为新用户精确地推荐物品。记忆网络已被广泛应用于基于序列的推荐系统。例如,文献[34]提出,先将历史会话存储在记忆网络中,然后通过端到端的记忆网络,利用历史会话中的物品的偏好,对当前的会话进行精准的推荐。

### 4 基于集成学习的推荐系统

集成学习方法融合多个独立的机器学习模型,用来实现更好的学习效果。根据集成策略,集成学习可以分为4大类:Bagging、Boosting、Stacking和混合专家模型(Mixture of Experts, MoE)。近年来,研究者们提出了基于Bagging和混合专家系统的推荐系统。

Bagging的目的:是通过对多个利用采样得到的数据训练过的独立模型,对其结果取平均值。通过这种方法,Bagging生成的集成模型比任意的单独模型有更低的偏差,从而提高学习性能。Bagging在推荐中的有效性已经得到了验证。例如,文献[35]提出了一种基于Bagging的矩阵分解模型,以提高推荐的精确度。该模型先对数据集中的多个训练数据子集进行有放回的采样,然后利用这些子集分别训练多个矩阵分解模型,最后对这些模型的预测值进行平均,从而用于推荐。

此外,混合专家系统是另一种有效的集合学习方法。该方法可以通过衡量不同专家模型的权重,有选择地融合多个专家模型,以达到更好的学习性能。一般来说,混合专家系统包含3部分:

(1)多个专家模型。其中每个专家模型都是一

个原子模型,专门针对特定类型的输入数据学习;

(2)门控网络。用于计算反映每个专家模型关于输入的专业程度的门控权重;

(3)融合模块。用于通过门控权重,融合多个专家模型的输出。

由于混合专家系统的有效性,已被应用于推荐系统,以便实现更高的推荐精确度。其中,文献[36]在最早的尝试中,提出了混合专家系统以顺序策略和联合策略,集成不同类型的协同过滤的推荐模型。在顺序策略中,前面协同过滤推荐模型的预测结果被反馈到后面的方法中,以获得更精确的预测结果。在联合策略中,所有的协同过滤模型,都以原始交互矩阵为输入,通过投票获得最终的预测结果。

最近,一些推荐系统更好地利用了混合专家系统的潜力,进一步地提高了推荐的精确度。这些工作在面对不同类型的交互时,采用门控网络为不同的专家模型,动态地分配不同的权重。通过这种方式,每个专家模型都被训练成专门针对某类交互的专家,通过针对每个交互的特点,为相应的专家模型分配更高的权重,为此交互做出更为精确的预测。通过这种方式,利用专家模型向用户推荐物品有望进一步提高推荐的精确度。

## 5 展望

现有的工作为对推荐系统做进一步研究打下了坚实的基础。本节列举了在推荐系统领域有潜力的研究方向和一些开放问题供读者参考。

### 5.1 流式推荐系统

尽管推荐系统在改善用户体验、提高企业效益等方面起到了重要作用,但是现有的推荐系统一般都需要离线训练,即定期使用大量历史数据对推荐模型进行训练,无法有效处理广泛存在的数据流,也很难为用户在流式场景中推荐物品。伴随着网络应用的发展和普及,用户与物品互动产生源源不断的数据流,近期的交互数据包含了用户的近期偏好,因此更能准确地反映出用户当前对物品的偏好程度。但传统的离线推荐系统在训练时会出现延迟,不能及时学习用户的近期偏好,从而很难对用户进行流式推荐。为了在流式环境中为用户提供精确的推荐,应该尝试使用新的数据来对其进行培训。

### 5.2 多行为推荐系统

虽然研究者们已提出多种推荐系统,但大多可以被称为单行为推荐系统。因这些推荐系统利用单一行为类型(如购买)的交互数据流为用户推荐商

品。直观地讲,在有多种行为类型(如购买、添加到购物车和浏览)的交互情况下,应将这些多行为交互纳入考虑之中,以实现精确的流式推荐。纳入更充分的多行为交互,有利于解决长期存在的数据稀疏性问题,从而在面对多行为交互的数据流时,有助于更准确的流推荐。

### 5.3 可解释推荐系统

现存的推荐系统大多是高度不可解释的,因此,做出可解释的推荐是一项艰巨的任务。可解释性推荐系统的重要意义主要有两个方面:一是向用户做出可解释的预测,让用户理解推荐背后的因素(即为什么推荐这个物品/服务);二是集中在对从业者的可解释性上,探究模型参数的权重和活跃度,以了解更多关于模型的信息。

## 6 结束语

推荐系统在人们的日常生活中发挥了越来越重要的作用。本文对基于矩阵分解、基于深度学习、基于记忆网络和基于集成学习的推荐系统进行了全面的介绍。充分分析了这几类推荐系统的特点,并对其进行了比较。希望这篇综述文章可以使读者对推荐系统有充分的了解并且为未来的工作提供灵感。

## 参考文献

- [1] 张时俊,王永恒. 基于矩阵分解的个性化推荐系统研究[J]. 中文信息学报, 2017, 31(3): 134-139.
- [2] HE X, LIAO L, ZHANG H, et al. Neural collaborative filtering [C] // Proceedings of the International Conference on World Wide Web 2017, Perth: ACM, 2017: 173-182.
- [3] LOPS P, DEGEMMIS M, SEMERARO G. Content-based recommender systems: State of the Art and Trends [G] // Recommender Systems Handbook. 2011: 73-105.
- [4] XUE H, DAI X, ZHANG J, et al. Deep Matrix factorization models for recommender systems [C] // Proceedings of the International Joint Conference on Artificial Intelligence 2017. Melbourne: [s.n.], 2017. 3203-3209.
- [5] CHEN W, NIU Z, ZHAO X, et al. A hybrid recommendation algorithm adapted in e-learning environments [J]. World Wide Web, 2014, 17(2): 271-284.
- [6] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian personalized ranking from implicit feedback [C] // Proceeding of the conference on Uncertainty Artificial Intelligence 2009. Montreal: AUAI, 2009:452-461.
- [7] SIMON F. Funksvd [EB/OL]. [2020-12-14]. <https://sifter.org/simon/journal/20061211.html>.
- [8] SALAKHUTDINOV R, MNIH A. Probabilistic matrix factorization [C] // Proceedings of the Annual Conference on Neural Information Processing Systems 2007, Vancouver: [s.n.], 2007: 1257-1264.