

深度矩阵分解推荐算法*

田震¹, 潘腊梅¹, 尹朴¹, 王睿^{1,2}

¹(北京科技大学 计算机与通信工程学院, 北京 100083)

²(北京科技大学 顺德研究生院, 广东 佛山 528300)

通讯作者: 王睿, E-mail: wangrui@ustb.edu.cn



摘要: 协同过滤推荐算法中的矩阵分解因其简单、易于实现,得到了广泛的应用.但是矩阵分解通过简单的线性内积建模用户和物品之间的非线性交互关系,限制了模型的表达能力.为此,He 等人提出了广义矩阵分解模型,通过非线性激活函数和连接权重,将矩阵分解推广到广义矩阵分解,为模型赋予建模用户和物品间的二阶非线性交互关系的能力.但是广义矩阵分解模型是一个浅层模型,并不能很好地建模用户和物品间高阶交互关系,一定程度上可能会影响模型性能.受广义矩阵分解模型启发,提出了深度矩阵分解模型(deep matrix factorization,简称 DMF),在广义矩阵分解模型的基础上引入隐藏层,利用深层神经网络来学习用户和物品间高阶交互关系.深度矩阵分解模型不仅解决了简单内积的线性问题,同时还能够建模用户和物品间的高阶交互,具有很好的表达能力.此外,在 MovieLens 和 Anime 两个数据集上进行了大量丰富的对比实验,验证了模型的可行性和有效性;同时,通过实验确定了模型的最优参数.

关键词: 协同过滤;线性内积;广义矩阵分解;隐藏层;高阶交互

中图法分类号: TP311

中文引用格式: 田震,潘腊梅,尹朴,王睿.深度矩阵分解推荐算法.软件学报,2021,32(12):3917-3928. <http://www.jos.org.cn/1000-9825/6141.htm>

英文引用格式: Tian Z, Pan LM, Yin P, Wang R. Deep matrix factorization recommendation algorithm. Ruan Jian Xue Bao/ Journal of Software, 2021,32(12):3917-3928 (in Chinese). <http://www.jos.org.cn/1000-9825/6141.htm>

Deep Matrix Factorization Recommendation Algorithm

TIAN Zhen¹, PAN La-Mei¹, YIN Pu¹, WANG Rui^{1,2}

¹(School of Computer & Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China)

²(Shunde Graduate School, University of Science and Technology Beijing, Foshan 528300, China)

Abstract: Matrix factorization in collaborative filtering recommendation algorithms is widely used because of its simplicity and facility of implementation, but matrix factorization utilizes a simple linear inner product to model the non-linear interaction between the user and the item, which limits the model's expressive power. He *et al.* proposed a generalized matrix factorization model, which extended the matrix factorization to the generalized matrix factorization through a non-linear activation function and connection weights, and gave the model the ability to model second-order non-linear interactions between users and items. Nevertheless, the generalized matrix factorization model is a shallow model and does not model the high-order interaction between users and items, which may affect the performance of the model to a certain extent. Inspired by the generalized matrix factorization model, this study proposes a deep matrix factorization model, abbreviated as DMF. Based on the generalized matrix factorization model, a hidden layer is introduced, and a deep neural network is used to learn the higher-order interaction between users and items. The deep matrix factorization model, which has a good expression ability, not only solves the linear problem of simple inner product, but also models high-order interactions between users

* 基金项目: 国家自然科学基金(62173158, 61803391); 北京科技大学顺德研究生院科技创新专项资金(BK19CF010, BK20BF012)

Foundation item: National Natural Science Foundation of China (62173158, 61803391); Scientific and Technological Innovation Foundation of Shunde Graduate School, USTB (BK19CF010, BK20BF012)

收稿时间: 2020-03-31; 修改时间: 2020-06-11; 采用时间: 2020-08-28

and items. In addition, a lot of rich comparative experiments are performed on two datasets, MovieLens and Anime, and the results confirm its feasibility and effectiveness. Meanwhile the optimal parameters of the model were determined through experiments.

Key words: collaborative filtering; linear inner product; generalized matrix factorization; hidden layers; high-order interaction

众多协同过滤推荐算法中,矩阵分解(matrix factorization,简称 MF)算法因简单、易于实现,得到了广泛的应用.但是矩阵分解通过简单的线性内积方式无法建模用户和物品之间复杂的非线性关系,限制了模型的推荐性能.虽然许多工作从多个方面对矩阵分解算法进行改进,但是仍然采用线性内积方式建模用户和物品之间的交互,使得模型的性能提升有限.He 等人提出的广义矩阵分解模型(generalized matrix factorization,简称 GMF)^[1]利用神经网络,通过一个激活函数和不全是 1 的连接权重赋予了 MF 非线性学习能力,将 MF 推广到非线性集合,提高了模型的表达能力.GMF 的浅层结构使得它在建模用户和物品的二阶交互关系上有很好的表现,然而却不能很好地捕捉到包含更丰富信息的用户和物品之间的高阶交互关系.随着深度神经网络技术的不断成熟,许多研究人员逐渐意识到深度神经网络强大的非线性学习能力和抽象特征学习能力,并开始将其应用到矩阵分解过程.但是现有研究要么对于协同过滤部分仍采用线性内积,要么只建模了用户和物品的二阶交互关系,忽略了包含更多信息的高阶交互,导致模型性能受限.

针对以上问题,本文提出了一种深度矩阵分解推荐算法,算法在广义矩阵分解算法的基础上,通过非线性内积得到的用户和物品的二阶交互之上引入隐藏层,利用神经网络建模用户和物品之间复杂的高阶非线性交互关系.实验结果表明,提出的深度矩阵分解推荐算法对于模型推荐性能具有显著的提升.

1 相关工作

在众多协同过滤推荐算法中,矩阵分解算法因其可扩展性和易于实现,成为最为普遍和最受欢迎的一种协同推荐算法.通过一组潜在的特征因子,把用户评分矩阵分解为 user-factor matrix 和 item-factor matrix.矩阵分解利用潜在因子向量描述用户和物品,即:将用户和物品映射到共享的潜在向量空间中,并通过映射向量之间的内积来表达用户与物品间的交互关系^[2,3].为了进一步提升推荐性能,许多的工作对矩阵分解算法进行改进,并取得了显著的成果.Steffen 等人对于 BPR-OPT 的学习模型提出了基于随机梯度下降和 bootstrap 抽样的一般学习算法 LearnBPR,并将其用于矩阵分解过程^[4].Badrul 等人^[5]提出了奇异值分解(SVD)来学习特征矩阵,但由 SVD 学习的 MF 模型容易过拟合.之后,Pan 等人提出一种带案例权重的正则化最小二乘优化算法 WR-MF^[6].此外,Mohsen 等人提出了 SocialMF 算法^[7],根据用户信任关系,利用信任用户的潜在特征向量来增强矩阵分解,得到的用户潜在特征向量,从而提升模型推荐性能.但是,这种方法只使用信任关系的表面信息进行推荐,没有考虑社交网络中信任关系是由多种因素形成的,所以郭等人在 SocialMF 的基础上提出了 StrengthMF^[8],将信任关系网络与评分信息联合用于矩阵分解过程,从而优化了推荐结果.同时,孙等人^[9]提出了 SequentialMF 推荐算法,该算法通过对用户时序行为建模发现用户邻居,从而将邻居信息融合到矩阵分解过程.虽然这些算法从不同的方面改进矩阵分解过程,使得模型推荐性能得到改善,但是这些算法的性能仍然受到因采取内积的方式建模用户和物品交互的限制,这种线性乘积的结合方式可能并不足以有效地捕捉交互数据中的非线性结构.即:简单和固定的内积有其局限性,会对模型造成一定的限制.

为了解决上述问题,He 等人在 GMF 中利用神经网络的非线性学习能力,在 MF 之上引入神经网络赋予模型学习用户和物品之间的非线性交互关系的能力,进一步提升了模型的表达能力.但是,GMF 中仅包含嵌入层和输出层,在得到用户和物品潜在特征向量后,通过在向量对应元素相乘的结果上添加一层全连接输出层就输出了结果,只是建模了用户和物品之间的二阶交互,并不能有效地捕捉到包含更丰富信息的用户和物品之间的高阶交互关系.即:GMF 的浅层结构能够建模用户和物品间交互的低层特征,但是并不能有效地捕捉到包含许多丰富信息、更抽象的高阶交互,高阶交互信息的缺失一定程度上会限制模型的推荐性能.

近些年,深度神经网络开始成为学术界和工业界的研究重点.深度神经网络通常拥有更多层次的隐藏层,单层隐藏层能够把输入数据的特征抽象到另一个维度空间,来展现其更抽象化的特征;而多隐藏层能够对输入特

征进行多层次的抽象,最终帮助模型更好的线性划分不同类型的数据.深度神经网络因为其对更高层次、更抽象特征的学习能力,使得越来越多的工作开始尝试将矩阵分解与深度神经网络结合进行推荐任务.文献[10]结合了矩阵分解和深度卷积网络,将通过应用 WMF^[11]得到的潜在因子向量作为训练预测模型的基真值来训练深度卷积神经网络,从而进行音乐的推荐任务.Zhang 等人利用异构网络嵌入和深度学习嵌入方法,从知识库的结构知识、文本知识和视觉知识中自动提取语义表示,然后结合协同过滤中的矩阵分解模型来进行推荐^[12].Wang 等人提出了 CDL 模型^[13],通过对内容信息进行深度表示学习和对评分矩阵进行协同过滤,使得模型具有很好的表达能力.这些方法通过深度神经网络很好地建模用户和物品之间的复杂的非线性高阶交互,赋予了模型学习用户和物品之间更抽象关系的能力,很好地提升了模型的表达能力,但是对于协同过滤部分仍采用 MF,通过简单的线性内积结合用户和物品的特征向量,一定程度上限制了模型的推荐性能.

此外,针对以往的 MF 方法大多只采用显式评分进行个性化推荐,忽略了用户和信息项的隐式反馈的重要性问题,Zhou 等人^[14]提出了基于深度神经网络的矩阵分解模型 HDMF.同时利用显式评分和隐式反馈作为输入,并结合深度神经网络获取潜在因子.同时,应用递归神经网络(RNN)将文本数据转换为辅助因子特征,以促进项目的表达,进一步提升了矩阵分解模型的推荐性能.为了方便、高效地集成各种辅助信息,Yi 等人提出了深度矩阵分解模型 DMF.基于提出的隐式反馈嵌入(IFE)方法和 one-hot 编码,DMF 能够获取融合多种辅助信息用户和物品的向量表示,然后通过多层感知器(MLP)分别对用户和物品向量进行特征转换,最终完成基于矩阵分解的推荐任务^[15].为了充分学习用户和物品的特征表示向量,Ma 等人^[16]提出了偏置深度矩阵分解模型 BD MF,该模型利用深层神经网络强大的表示学习能力,建模用户和项目的抽象特征表示,同时借鉴 BiasedSVD 将偏差引入用户和项目的特征表示中,进而将用户和项目从评分矩阵映射到低维空间.尽管以上 3 个模型结合深度神经网络对矩阵分解算法性能进行了进一步的改善,但是他们的关注重点是利用深度神经网络强大的非线性学习能力帮助模型去学习用户和物品的抽象特征表示,在建模用户和物品的交互关系时,仍然只考虑了用户和物品的二阶交互,导致包含更多抽象信息的高阶交互被忽略.所以,本文提出了深度矩阵分解模型,通过在 GMF 之上添加隐藏层,既能解决矩阵分解简单线性内积对模型的限制问题,又能利用深层神经网络来建模用户和物品之间的非线性高阶交互.

2 深度矩阵分解模型

2.1 问题定义

基于可读性,我们在表 1 中列出了本文使用的重要数学符号.

Table 1 Symbol definition

表 1 符号定义

符号	描述	符号	描述
$U=\{u_1, u_2, u_3, \dots, u_M\}$	用户集合	p_u	用户 u 潜在特征向量
$I=\{i_1, i_2, i_3, \dots, i_N\}$	项目集合	q_i	物品 i 潜在特征向量
$Y_{M \times N}$	用户-物品评分矩阵	W_{L-1}^T	第 $L-1$ 层隐藏层网络链接权重矩阵
M	用户数目	a_{L-1}	第 $L-1$ 层隐藏层网络激活函数
N	物品数目	b_{L-1}	第 $L-1$ 层隐藏层网络偏置向量
$P_{M \times K}$	用户潜在特征矩阵	z_{L-1}	第 $L-1$ 层隐藏层网络输出
$Q_{N \times K}$	物品潜在特征矩阵	\hat{y}_{ui}	用户 u 对物品 i 的预测评分
K	潜在特征维度		

基于这些符号,我们对问题进行了形式化定义.

问题定义. 假设有用户集 $U=\{u_1, u_2, u_3, \dots, u_M\}$, 项目集 $I=\{i_1, i_2, i_3, \dots, i_N\}$, 其中, M 为用户数目, N 为物品数目. $Y_{M \times N}$ 为用户-物品评分矩阵. 现给定用户 $u \in U$ 、物品 $i \in I$, 其中, 用户 u 未与物品 i 产生交互. 我们的目标是根据用户评分矩阵 $Y_{M \times N}$ 中用户 u 的历史评分记录, 建模用户 u 的偏好, 进而预测用户 u 对物品 i 的偏好评分, 再根据预测值对商品排名, 最终为用户推荐前 n 个物品.

2.2 DMF模型

为了学习用户和物品之间复杂的高阶非线性交互关系,本文提出了深度矩阵分解模型,该框架如图 1 所示.

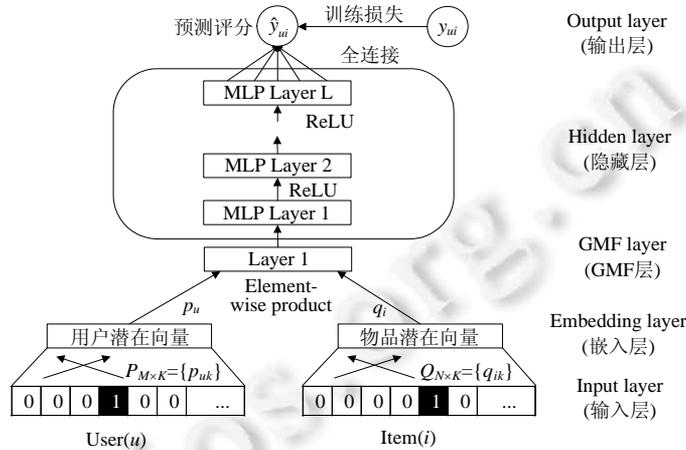


Fig.1 Deep matrix factorization model

图 1 深度矩阵分解模型

该框架主要由 5 个部分组成:Input Layer(输入层),Embedding Layer(嵌入层),GMF Layer(GMF 层),Hidden Layer(隐藏层)和 Output Layer(输出层).输入层获取用户和物品数据;在嵌入层,通过嵌入技术得到用户和物品的潜在特征向量;随后,在 GMF 层通过向量间对应元素相乘建模用户和物品的二阶交互关系;再将结果输入到隐藏层中,利用深度神经网络对结果进行迭代训练,从而建模用户和物品之间的高阶交互关系;最后,在输出层得到用户对物品的评分预测.下面分别对模型各个组成模块进行详细阐述.

- Input Layer(输入层)

对于用户(物品)的原始数据,一般通过编码的方式把它转换为数值数据.通常分为两步,即整数编码和 one-hot 编码.整数编码会给每一个用户和物品都分配一个整数值,比如用 1 表示 user1,2 表示 user2.但是整数之间自然的排序方式会让模型假设用户(物品)之间也具有某种次序关系,从而可能会影响模型的性能.所以,还需要对整数表示进行 one-hot 编码,为每个整数值创建二值表示:除了整数的索引为 1 外,其他都是 0,即 one-hot 编码在任意时候只有一个有效位置.本文将用户编号和物品编号视为两个不同的离散特征进行处理,并且将用户编号和物品编号分开,分别采用 one-hot 编码进行编码.所以输入层用户表示为 001000...,其中,编码长度为用户数目 M ;输入层物品表示为 01000...,其中,编码长度为物品数目 N .

- Embedding Layer(嵌入层)

从输入层得到的 one-hot 编码是高维稀疏的,会严重影响模型训练效率,所以在嵌入层通过 one-hot 编码与潜在特征矩阵相乘将用户(物品)表示映射到低维密致空间,从而得到潜在特征向量.其中, $P_{M \times K}$ 为用户潜在特征矩阵, M 为用户数目, K 为潜在特征维度, $K \ll M$. $Q_{N \times K}$ 为物品潜在特征矩阵, N 为物品数目, K 为潜在特征维度, $K \ll N$.用户的 one-hot 编码与 $P_{M \times K}$ 矩阵相乘,得到 K 维的用户潜在特征向量 p_u .物品的 one-hot 编码与 $Q_{N \times K}$ 矩阵相乘,得到 K 维的物品潜在特征向量 q_i .

- GMF Layer(GMF 层)

将嵌入层得到的潜在特征向量 p_u 和 q_i 对应元素相乘,得到 K 维的向量,如公式(1)所示:

$$s_i = \varphi(p_u, q_i) = p_u \odot q_i \quad (1)$$

虽然通过内积得到用户和物品的二阶交互,但是简单的线性内积不足以建模用户和物品之间复杂的非线性结构,所以再将内积结果输入到一层神经网络中,如公式(2)所示,由此来学习二阶交互的非线性关系,这就是 GMF 模型:

$$z_1 = \phi_1(s_1) = a_1(W_1^T s_1 + b_1) \quad (2)$$

其中, a_1 为激活函数, W_1 和 b_1 别为连接权值和偏置值.

- Hidden Layer(隐藏层)

GMF 层在特征向量 p_u 和 q_i 对应元素相乘得到的 K 维向量之上设置了一层神经网络, 通过激活函数和偏置项赋予了模型建模用户和物品之间的非线性交互关系的能力, 能够有效地拟合用户和物品的非线性二阶交互信息. 但是 GMF 是一个浅层模型, 尽管能够有效地建模用户和物品交互的低层特征, 然而并不能有效地捕捉到包含许多丰富信息更抽象的特征. 比如: 输入用户对音乐的交互信息, 通过一层神经网络训练学习最后捕捉到用户更偏好古典音乐; 接着, 将获取的结果输入到下一层神经网络进行学习, 捕获到用户更倾向于听古典音乐中莫扎特的音乐曲. 所以, 这里在 GMF Layer 的基础上引入更多的隐藏层, 通过深层神经网络能够捕捉更抽象特征的特点, 来建模用户和物品之间的高阶交互关系. 因此, 有:

$$\begin{aligned} z_2 &= \phi_2(z_1) = a_2(W_2^T z_1 + b_2) \\ &\dots \\ z_{L-1} &= \phi_{L-1}(z_{L-2}) = a_{L-1}(W_{L-1}^T z_{L-2} + b_{L-1}) \end{aligned} \quad (3)$$

其中, z_{L-1} 为第 $L-1$ 层网络的输出, a_2, a_{L-1} 为各个隐藏层的激活函数, W_2^T, W_{L-1}^T 和 b_2, b_{L-1} 分别为各个隐藏层的连接权值和偏置项.

- Output Layer(输出层)

得到第 $L-1$ 层神经网络的训练结果 z_{L-1} 后, 再将其输入到一个全连接的隐藏层, 最终得到用户对于物品的预测评分. 如公式(4)所示:

$$\hat{y}_{ui} = \sigma(h^T z_{L-1}) \quad (4)$$

其中, \hat{y}_{ui} 表示用户 u 对物品 i 的预测偏好值, σ 和 h^T 分别为输出层的激活函数和连接权重.

3 实验结果与分析

本节主要工作是在两个真实世界数据集 MovieLens 和 Anime 上训练模型, 利用 NDCG 和 HR 两个性能指标, 通过与对比方法性能指标的对比分析来验证模型的可行性和有效性. 主要围绕一下 4 个方面来进行分析.

- 1) 深度矩阵分解模型推荐性能表现;
- 2) 深度矩阵分解模型中潜在向量维度对模型性能影响;
- 3) 训练集中负样本采样数对模型性能影响;
- 4) 隐藏层层数对模型性能影响.

3.1 实验设置

3.1.1 数据集

本文主要在两个公开数据集 MovieLens^[17]和 Anime^[18]上对模型的推荐性能进行分析验证. 具体数据集特征见表 2.

- MovieLens 数据集是广泛应用于推荐系统的电影数据集, 主要由电影评分数据、用户人口统计学数据以及电影属性数据, 含有多个不同数据大小的版本. 这里, 我们使用 MovieLens 1M 数据集, 记为 MovieLens.MovieLens 1M 数据集含有来自 6 000 名用户对 4000 部电影的 100 万条评分数据, 分值范围为 1~5;
- Anime 是一个公开的推荐数据集, 包含 73 516 名用户对 12 294 部动漫的 7 813 738 条评分数据. 该数据集包含 3 个分类: 用户 id、动漫 id 和用户对动漫的评分, 其中, 评分范围为 1~10. 为了缓解数据集稀疏问题, 对数据集进行过滤处理, 过滤后的数据集包含 8 289 个用户对 6 335 部动漫的 2 964 160 个评分.

这里, 我们采用隐式反馈的数据, 所以要把显式评分数据转化为隐式数据. 即: 对于 MovieLens 数据来说, 将评分矩阵中有评分记录的 1~5 的值记为 1, 没有评分记录的记为 0. 同时选取每个用户最近一次的交互记录作

为测试样本,并和随机抽样的 99 个未交互的样本组合成测试数据集,其余的数据作为训练样本组成训练集.对 Anime 数据集做同样处理.

Table 2 Dataset description

表 2 数据集描述

数据集	用户数	物品数	交互次数	稀疏度
MovieLens	3 706	6 040	1 000 209	95%
Anime	8 289	6 335	2 964 160	94%

3.1.2 评估指标

由于为每个用户预测其对所有未产生交互的物品的偏好过于耗费时间,所以这里遵循一般的策略^[2,19].通过第 3.1.1 节中描述的采样方式,为每个用户采集 1 个正样本和 99 个负样本组成的测试集,并为用户预测其对这 100 个物品的偏好,并按照预测值从大到小排序,生成 Top-10 的排名列表.

推荐性能由命中率 HR(hit ratio)与 NDCG(normalized discounted cumulative gain)^[20]来衡量.

- HR 直观地评估测试物品是否在 Top-10 列表中:若存在于 Top-10 列表中,则为 1;否则为 0.对所有的用户求平均,则命中率代表了命中用户偏好的概率.计算公式如公式(5)所示:

$$HR@K = \frac{\text{number of hits}}{N} \quad (5)$$

其中, K 表示排名列表长度,分子为生成的排名列表中包含测试正样本的用户数,分母为用户的总数.

- NDCG 用来衡量测试样本命中的位置,该值越大,则说明测试物品出现在排名列表中更靠前的位置.为每个用户计算该指标值,并且求取平均.计算公式如公式(6)所示:

$$NDCG@K = \frac{1}{N} \sum \frac{\log_e 2}{\log_e(i+2)} \quad (6)$$

其中, K 表示排名列表的长度, N 为用户总数, i 为排名列表中测试正样本命中的位置, $i \in [1, K]$.可以看出:测试正样本在排名列表中的位置越靠前,即 i 越小,该值越大.

3.1.3 对比方法及参数设置

我们将 DMF 模型与以下 4 个基准方法进行分析对比,从而验证模型的有效性.

- ItemPop 是一种非个性化推荐方法^[21],它将物品与用户交互次数作为物品的流行度,按照物品的流行度产生物品排名列表,并将排名列表中的前 N 项推荐给用户;
- ItemKNN 是一种协同过滤推荐方法^[22],它根据用户与物品的交互记录计算物品间的相似度,再结合待推荐物品与用户已有评分交互的邻居物品来预测用户对推荐物品的预测偏好;
- BPR-MF^[4]是一种具有代表性的基于成对学习的 MF 模型,损失函数关注的是成对排序损失,即最大化正样本与负样本之间的排序^[4,23],而不是减少预测值与真实值之间的误差,正负采样率控制为 1:1;
- GMF 是一种改进后的矩阵分解算法^[1],通过激活函数和权值不全为 1 的连接权重泛化了矩阵分解模型,使得模型能够建模用户和物品之间非线性二阶交互.

以上方法模型和所提出的方法模型 DMF 都基于 Tensorflow 实现.对于基于神经网络的模型,通过均值为 0、方差 0.01 的截断的正态分布来初始化模型参数,并通过交叉熵损失函数和 Adam 梯度下降算法来优化模型.模型批次大小为 256,初始学习率为 0.001,隐藏层的激活函数设置为 ReLU,输出层的激活函数设置为 Sigmoid.

3.2 性能比较和分析

我们从以下 4 个方面进行性能评价和比较分析.

- Result 1:模型 HR@10 和 NDCG@10 性能指标比较

为了回答问题 1,我们比较了提出的 DMF 模型和 ItemPop,ItemKNN,BPR,GMF 这 4 个基准方法的性能.分别在 MovieLens 和 Anime 两个数据集上训练了这 5 个模型,表 3 显示了这 5 个模型在每个数据集中 HR@10 和 NDCG@10 的表现情况.

Table 3 Performance comparison of deep matrix factorization and baselines

表 3 深度矩阵分解与基准方法性能对比

数据集 对比方法	MovieLens		Anime	
	NDCG@10	HR@10	NDCG@10	HR@10
ItemPop	0.268 6	0.459 6	0.246 2	0.413 6
ItemKNN	0.354 1	0.610 5	0.321 5	0.582 1
BPR	0.402 1	0.680 1	0.395 1	0.635 5
GMF	0.423 4	0.701 2	0.471 2	0.739 7
DMF	0.443 2	0.707 1	0.532 1	0.785 3

通过表 3 不难看出:在两个数据集上,HR@10 和 NDCG@10 指标都是 $DMF \gg GMF \gg BPR \gg ItemKNN \gg ItemPop$ 。我们提出的 DMF 模型在两个数据集上的性能表现都优于基准方法模型。

- 相比较于 GMF 模型,其在 Anime 数据集上的 HR@10 和 NDCG@10 指标分别提升了 0.045 6 和 0.060 9;而在 MovieLens 数据集上,两个指标分别提升了 0.005 9 和 0.019 8;
- 相比较于 ItemPop 方法,DMF 在 MovieLens 数据集上两个指标分别提升了 0.247 5 和 0.174 6;而在 Anime 数据集上,HR@10 和 NDCG@10 指标分别提升了 0.371 7 和 0.285 9。

在这些模型中,ItemPop 表现最差.这可能是由于 ItemPop 是根据物品在所有用户中受欢迎程度来推荐的,不能满足用户的个性化需求,无法提供个性化推荐.这表明,非个性化的推荐方法不能有效地用来预测用户偏好.ItemKNN 方法基于用户和物品的历史交互矩阵来计算物品相似度进行物品推荐,虽然能一定程度上满足用户个性化需求;但是由于交互矩阵的稀疏性问题,导致物品相似度的计算偏差大,所以整体影响了最终模型的推荐性能,因此推荐性能虽然相对于 ItemPop 方法有了很大提升,但是性能还是低于 BPR,GMF 和 DMF.同时,DMF 和 GMF 推荐性能强于 BPR,这可能是因为 DMF 和 GMF 有着更灵活的正负采样率,对于每个正训练样本采样更多的合适负样本,更准确建模现实世界用户和物品交互的场景.同时,结果表明:相比较于 BPR 中采用简单线性内积建模用户和物品之间的交互,采用神经网络赋予模型非线性的学习能力的 DMF 和 GMF 具有更突出的表达能力.而 DMF 的 HR@10 和 NDCG@10 指标在两个数据集上表现都优于 GMF,这可能是因为添加的隐藏层将在 GMF 得到的二阶交互映射到更深层次的抽象空间中,从而学习到用户和物品间更高阶交互关系.这表明:通过添加隐藏层,利用深层神经网络学习高阶抽象特征的特点,使得模型能够建模用户和物品之间的高阶非线性交互关系,由此模型具有更突出的表达能力。

- **Result 2:潜在向量维度影响**

为了回答问题 2 潜在向量维度对模型性能的影响,我们比较了潜在向量维度为 10,20,40 和 80 下,ItemKNN, BPR,GMF 和 DMF 这 4 个模型在 MovieLens 和 Anime 两个数据集上的性能表现,如图 2 和图 3 所示.由于 ItemPop 方法性能相对于其他方法表现差距较大,这里不予以比较。

由图 2 和图 3 首先可以看到:随着潜在维度的不断增加,4 个模型的 HR@10 和 NDCG@10 指标不断提升,模型的推荐性能不断改善.这是因为在到达最佳潜在向量维度之前,随着维度的增加,潜在向量可以包含更多的信息量,对于用户和物品特征描述更加全面具体,能够更有效地建模用户和物品的潜在特征,所以模型具有更突出的表现能力,推荐性能得到显著改善.但是维度并不是越大越好,因为当潜在向量维度过大时,会导致模型过拟合,这也是为什么当在 MovieLens 数据集上,潜在向量维度达到 40 后,HR@10 和 NDCG@10 指标开始下降;同样地,在 Anime 数据集上,潜在向量维度达到 40 后,HR@10 和 NDCG@10 指标也开始下降.同时,我们从图 2 和图 3 可以得出结论,DMF 模型在 MovieLens 和 Anime 数据集上最佳潜在向量维度都为 40.其次,可以看到:在 MovieLens 和 Anime 两个数据集中,DMF 模型在不同维度下,HR@10 和 NDCG@10 指标都明显高于其他 3 个模型.而在图 3 中,甚至在潜在维度为 10,DMF 模型性能最差的时候,其 HR@10 和 NDCG@10 指标仍然高于 BPR 模型,这表明我们提出的 DMF 模型的有效性。

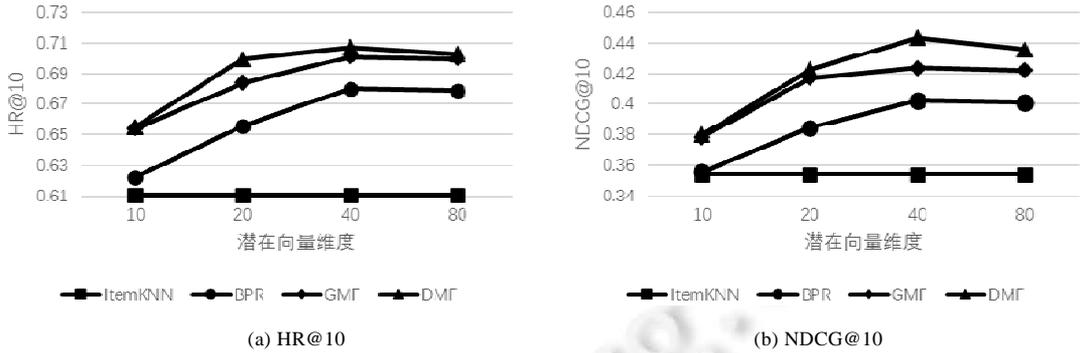


Fig.2 HR@10 and NDCG@10 of models with different potential vector dimension in MovieLens
图 2 MovieLens 数据集下不同潜在向量维度模型的 HR@10 和 NDCG@10

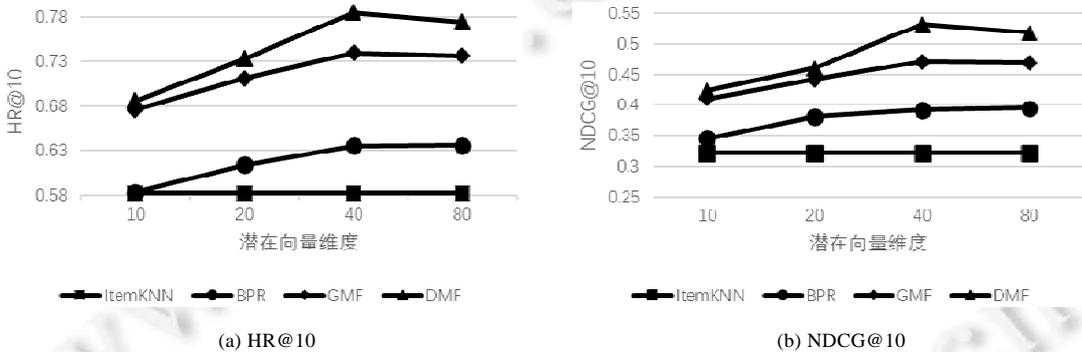


Fig.3 HR@10 and NDCG@10 of models with different potential vector dimension in Anime
图 3 Anime 数据集下不同潜在向量维度模型的 HR@10 和 NDCG@10

• Result 3:负样本采样数影响

BPR 损失函数关注的是成对排序损失,即最大化正样本与负样本之间的排序.构造训练集时,正负样本采样率固定为 1:1 来采集数据样本.而 DMF 和 MLP 采用 Log Loss,其关注重点为逐点损失,一般通过最小化预测值 \hat{y} 和目标值 y 之间的平方误差来更新参数,具有灵活的负样本采样特点.为了探究负样本采样数对模型推荐性能的影响,我们在负样本采样数为 1,3,5,7 和 9 的情况下,分别将 BPR,GMF 和 DMF 这三个模型在 MovieLens 和 Anime 两个数据集上进行了训练,得到了模型的 HR@10 和 NDCG@10 指标表现情况,结果如图 4 和图 5 所示.

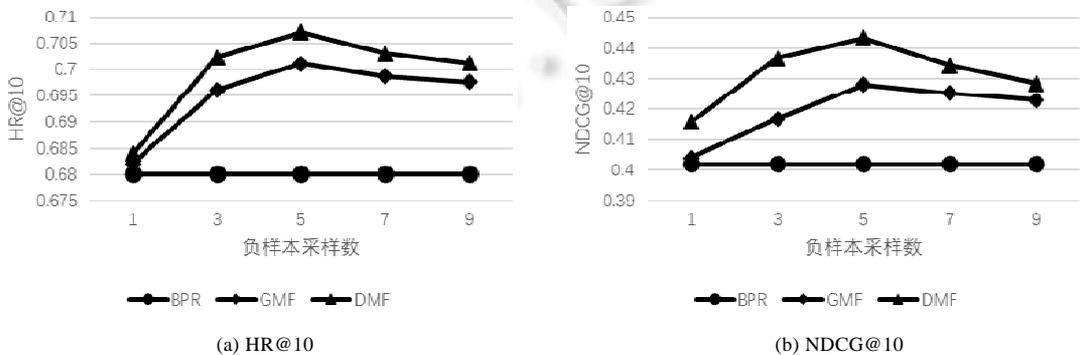


Fig.4 HR@10 and NDCG@10 of models with different number of negative samples collected in MovieLens
图 4 MovieLens 数据集下不同负样本采集数模型的 HR@10 和 NDCG@10

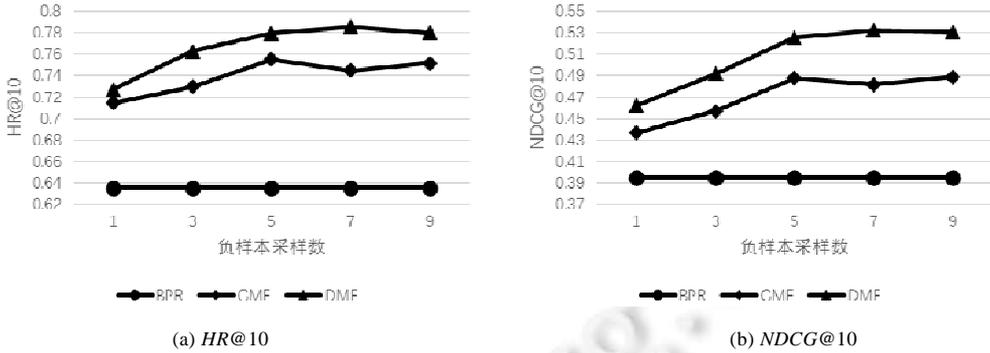


Fig.5 HR@10 and NDCG@10 of models with different number of negative samples collected in Anime
图 5 Anime 数据集下不同负样本采集数模型的 HR@10 和 NDCG@10

首先,由图 4 和图 5 不难看出:拥有更灵活负样本采样的 DMF 和 GMF 模型相比较于 BPR 模型来说,在两个数据集上具有更好的表现;同时,随着负样本采样数的增加,DMF 模型的 HR@10 和 NDCG@10 指标在逐步增加.这表明,灵活的采样使得模型具有更突出的表达能力.这是因为:通过向训练集中添加负样本,使得训练集更符合真实世界用户和物品交互的稀疏场景;此外,通过采集负样本,使得在训练模型过程中更新参数时只更新部分参数,其他参数全部固定,从而减少整个训练过程计算量,并加速了模型的收敛过程.其次还可以看出:当负样本采样数达到某个值之后,随着采样数的增加,模型的推荐性能反而下降了.这是因为负样本采样数过大,导致训练数据集出现正负样本不均衡,从而影响了模型的表现能力.例如:一个训练集中正负样本比为 1:99,如果模型的预测结果全都是负样本,那么模型的准确率就能够达到 99%;同时,损失值会非常小,但是这样的模型是没有实际意义的,模型的推荐效果是很差的.所以,适当地控制负样本的采样数,能够提升模型的推荐性能.对于 DMF 模型来说,在 MovieLens 数据集上的最优负样本采样数是 5,在 Anime 数据集上的最优负样本采样数是 7.

• Result 4:隐藏层数量影响

通过前面的实验与分析,我们不难得出结论:通过向 GMF 模型添加隐藏层,能够建模用户和物品之间更高级的交互关系,采用更深层次的神经网络来学习更抽象的用户和物品特征,使得模型具有更好的表达能力.为了探究隐藏层层数对模型推荐性能影响,我们分别设置了隐藏层层数为 0,4 的对照模型,其中,层数为 0 的模型也就是 GMF 模型,层数为 1 的模型记为 DMF-1,以此类推,并在潜在向量维度为 10,20,40 和 80 情况下,分别利用 MovieLens 和 Anime 两个数据集训练各个模型,得到最终 5 个对照模型在不同维度下的 HR@10 和 NDCG@10 指标的表现,结果见表 4~表 7.

Table 4 HR@10 of models with different layers in MovieLens

表 4 MovieLens 数据集下不同层数的模型的 HR@10

潜在向量维度	GMF	DMF-1	DMF-2	DMF-3	DMF-4
10	0.654 1	0.654 9	0.653 3	0.655 0	0.654 3
20	0.684 1	0.691 3	0.692 2	0.699 7	0.700 0
40	0.701 2	0.702 3	0.705 0	0.707 1	0.703 8
80	0.694 1	0.700	0.701 5	0.703 0	0.696 7

Table 5 NDCG@10 of models with different layers in MovieLens

表 5 MovieLens 数据集下不同层数的模型的 NDCG@10

潜在向量维度	GMF	DMF-1	DMF-2	DMF-3	DMF-4
10	0.378 3	0.378 6	0.379 2	0.379 8	0.419 5
20	0.417 1	0.418 9	0.419 5	0.422 5	0.438 8
40	0.423 4	0.431 4	0.429 4	0.443 2	0.450 7
80	0.422 1	0.422 4	0.423 8	0.435 8	0.436 0

Table 6 HR@10 of models with different layers in Anime**表 6** Anime 数据集下不同层数的模型的 HR@10

潜在向量维度	GMF	DMF-1	DMF-2	DMF-3	DMF-4
10	0.676 2	0.680 8	0.682 3	0.685 9	0.689 2
20	0.711 0	0.730 0	0.731 2	0.732 4	0.734 9
40	0.739 7	0.762 6	0.773 9	0.785 3	0.767 8
80	0.736 5	0.772 3	0.772 9	0.774 4	0.771 1

Table 7 NDCG@10 of models with different layers in MovieLens**表 7** Anime 数据集下不同层数的模型的 NDCG@10

潜在向量维度	GMF	DMF-1	DMF-2	DMF-3	DMF-4
10	0.410 5	0.414 9	0.416 2	0.423 5	0.429 6
20	0.442 2	0.463 7	0.462 2	0.460 6	0.465 2
40	0.471 2	0.500 5	0.496 7	0.532 1	0.492 2
80	0.469 8	0.508 5	0.515 8	0.518 4	0.508 8

由表 4 和表 5 可以看出:在不同潜在向量维度下,随着隐藏层层数从 0 开始不断增加,模型的 HR@10 和 NDCG@10 指标不断提升,模型性能得到明显改善.这是因为单层隐藏层能够将输入数据映射到另一个抽象空间,学习更抽象的特征;而通过增加隐藏层层数,能够学习更丰富更抽象的信息,从而更好地建模用户和物品之间的交互.此外,通过引入隐藏层,利用神经网络模型赋予学习非线性的能力,使得模型具有更突出的表达能力.这表明:通过添加单层隐藏层,利用神经网络建模用户和物品之间复杂的非线性交互,能够提升模型推荐性能.而通过添加更多的隐藏层,利用更深层的神经网络能够学习用户和物品之间更高阶的交互关系,从而进一步提升模型推荐性能.但是,随着隐藏层层数进一步增加,模型 HR@10 和 NDCG@10 指标提升幅度减小,甚至在潜在向量维度为 10,40 和 80 的时候,隐藏层层数从 3 改为 4 后,模型在 MovieLens 数据集上,HR@10 指标出现了下降的情况;同样,在潜在向量维度为 40 和 80 时,隐藏层层数由 3 变为 4 后,模型在 Anime 数据集上,HR@10 指标也出现了下降.这表明对于模型来说,并不是隐藏层层数越多、神经网络越深,模型推荐性能越好.这是因为隐藏层层数过多,会导致模型出现过拟合问题,这样不仅不会提升模型性能,反而会限制模型推荐性能.同时,由于神经网络中过多的隐藏层会使得模型参数指数式增长,增加训练难度导致模型难以收敛.

因此,隐藏层层数的增加使得模型获得更深层神经网络,能够建模用户和物品之间高阶非线性交互关系,但是并不是越多越好,隐藏层层数过多反而会影影响模型性能.这里,我们所提出的模型最优隐藏层层数为 3.此外,通过表 4~表 7 进一步表明,DMF 在 MovieLens 和 Anime 两个数据集上的最优潜在向量维度为 40.

4 总 结

在众多的协同过滤推荐算法中,矩阵分解由于简单、易于实现的特点,获得了广泛的应用.矩阵分解算法利用潜在的特征因子把评分矩阵分解为 user-factor matrix 和 item-factor matrix,将用户和物品用潜在因子向量表示,并且通过向量之间的内积来建模用户与物品间的交互关系.但是这种简单的内积不足以建模用户和物品间的非线性关系.由于神经网络的非线性建模能力,一些工作尝试将矩阵分解和神经网络结合进行推荐任务,虽然推荐性能有一定提升,但是涉及到矩阵分解的关键,即建模用户和物品的交互函数时,仍然采用简单的内积来建模用户和物品之间的线性关系.

GMF 模型中利用神经网络代替矩阵分解中所用的内积,来从数据中学习得到用户和物品的交互函数.通过神经网络结构为内积引入非线性成分,从而提高模型的表达能力.但是 GMF 模型得到的仅仅是用户和物品交互的二阶关系,忽略了可能包含更丰富交互信息的高阶交互.基于此,我们提出了深度矩阵分解模型 DMF,通过在 GMF 之上引入隐藏层,利用更深层的神经网络在非线性的内积所得到的用户与物品间的二阶交互的基础之上,建模用户与物品间的非线性高阶交互,为模型引入用户与物品间的非线性高阶交互信息,从而辅助模型更准确、全面的建模用户和物品间的交互.同时,我们在两组真实世界的数据集上进行了大量的实验,证实了模型的可行性和有效性.本文只考虑了用户和物品的交互信息,没有考虑可能包含更多可以用来建模用户和物品特征的丰

富知识的辅助信息,例如用户年龄、性别、工作和物品类别等,这可能是未来的工作方向.同时,不同种类信息对于推荐结果的贡献程度也是不一样的,如何合理地给不同信息赋予权重,也是值得关注的的问题.

References:

- [1] He XN, Liao LZ, Zhang HW, Nie LQ, Hu X, Chua TS. Neural collaborative filtering. In: Proc. of the 26th Int'l Conf. on World Wide Web. Republic and Canton of Geneva: WWW, 2017. 173–182. [doi: 10.1145/3038912.3052569]
- [2] He XN, Zhang HW, Kan MY, Chua TS. Fast matrix factorization for online recommendation with implicit feedback. In: Proc. of the 39th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. New York: ACM, 2016. 549–558. [doi: 10.1145/2911451.2911489]
- [3] Koren Y. Factorization meets the neighborhood: A multifaceted collaborative filtering model. In: Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM, 2008. 426–434. [doi: 10.1145/1401890.1401944]
- [4] Steffen R, Christoph F, Zeno G, Lars ST. BPR: Bayesian personalized ranking from implicit feedback. In: Proc. of the 25th Conf. on Uncertainty in Artificial Intelligence. Arlington: AUAI, 2009. 452–461.
- [5] Sarwar B, Karypis G, Konstan J, Riedl J. Incremental singular value decomposition algorithms for highly scalable recommender systems. In: Proc. of the 5th Int'l Conf. on Computer and Information Science. 2002. 27–28.
- [6] Pan R, Zhou YH, Cao B, Liu NN, Lukose RJ, Scholz M, Yang Q. One-class collaborative filtering. In: Proc. of the 2008 8th IEEE Int'l Conf. on Data Mining. Piscataway: IEEE, 2008. 502–511. [doi: 10.1109/ICDM.2008.16]
- [7] Jamali M, Ester M. A matrix factorization technique with trust propagation for recommendation in social networks. In: Proc. of the 4th ACM Conf. on Recommender Systems. New York: ACM, 2010. 135–142. [doi: 10.1145/1864708.1864736]
- [8] Guo L, Ma J, Chen ZM. Trust strength aware social recommendation method. *Journal of Computer Research and Development*, 2013,50(9):1805–1813 (in Chinese with English abstract).
- [9] Sun GF, Wu L, Liu Q, Zhu C, Chen EH. Recommendations based on collaborative filtering by exploiting sequential behaviors. *Ruan Jian Xue Bao/Journal of Software*, 2013,24(11):2721–2733 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4478.htm> [doi: 10.3724/SP.J.1001.2013.04478]
- [10] Van den Oord A, Dieleman S, Schrauwen B. Deep content-based music recommendation. In: Proc. of the 26th Int'l Conf. on Neural Information Processing Systems. Nevada: NIPS, 2013. 2643–2651.
- [11] Hu Y, Koren Y, Volinsky C. Collaborative filtering for implicit feedback datasets. In: Proc. of the 2008 8th IEEE Int'l Conf. on Data Mining. Piscataway: IEEE, 2008. 263–272. [doi: 10.1109/ICDM.2008.22]
- [12] Zhang F, Yuan NJ, Lian D, Lian DF, Xie X, Ma WY. Collaborative knowledge base embedding for recommender systems. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM, 2016. 353–362. [doi: 10.1145/2939672.2939673]
- [13] Wang H, Wang NY, Yeung DY. Collaborative deep learning for recommender systems. In: Proc. of the 21th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM, 2015. 1235–1244. [doi: 10.1145/2783258.2783273]
- [14] Zhou J, Wen J, Li S, Zhou W. From content text encoding perspective: A hybrid deep matrix factorization approach for recommender system. In: Proc. of the 2019 Int'l Joint Conf. on Neural Networks (IJCNN). Piscataway: IEEE, 2019. 1–8. [doi: 10.1109/IJCNN.2019.8852443]
- [15] Yi B, Shen X, Liu H, Zhang Z, Zhang W, Liu S. Deep matrix factorization with implicit feedback embedding for recommendation system. *IEEE Trans. on Industrial Informatics*, 2019,15(8):4591–4601. [doi: 10.1109/TII.2010.2893284]
- [16] Ma CS, Li JJ, Pan P, Li GH, Du JB. BDMF: A biased deep matrix factorization model for recommendation. In: Proc. of the 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation. Piscataway: IEEE, 2019. 1039–1045. [doi: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00201]
- [17] MovieLens 1M dataset released Feb. 2003. <https://grouplens.org/datasets/movielens/1m/>
- [18] Anime dataset released December. 2016. <https://www.kaggle.com/CooperUnion/anime-recommendations-database>
- [19] Wakita Y, Oku K, Kawagoe K. Toward fashion-brand recommendation systems using deep-learning: Preliminary analysis. *Int'l Journal of Knowledge Engineering*, 2016,2(3):128–131. [doi: 10.18178/ijke.2016.2.3.066]

- [20] He X, Chen T, Kan MY, Chen X. Trirank: Review-aware explainable recommendation by modeling aspects. In: Proc. of the 24th ACM Int'l on Conf. on Information and Knowledge Management. New York: ACM, 2015. 1661–1670. [doi: 10.1145/2806416.2806504]
- [21] Gueye M, Abdessalem T, Nacke H. A social and popularity-based tag recommender. In: Proc. of the 2014 IEEE 4th Int'l Conf. on Big Data and Cloud Computing. Piscataway: IEEE, 2014. 318–325. [doi: 10.1109/BDCLOUD.2014.44]
- [22] Sarwar B, Karypis G, Konstan J, Riedl J. Item-based collaborative filtering recommendation algorithms. In: Proc. of the 10th Int'l Conf. on World Wide Web. New York: ACM, 2001. 285–295. [doi: 10.1145/371920.372071]
- [23] Strub F, Gaudel R, Mary J. Hybrid recommender system based on autoencoders. In: Proc. of the 1st Workshop on Deep Learning for Recommender Systems. New York: ACM, 2016. 11–16. [doi: 10.1145/2988450.2988456]

附中文参考文献:

- [8] 郭磊,马军,陈竹敏.一种信任关系强度敏感的社会化推荐算法.计算机研究与发展,2013,50(9):1805–1813.
- [9] 孙光福,吴乐,刘淇,朱琛,陈恩红.基于时序行为的协同过滤推荐算法.软件学报,2013,24(11):2721–2733. <http://www.jos.org.cn/1000-9825/4478.htm> [doi: 10.3724/SP.J.1001.2013.04478]



田震(1996—),男,学士,主要研究领域为推荐系统,机器学习.



潘腊梅(1995—),女,硕士,主要研究领域为推荐系统,机器学习.



尹朴(1997—),男,学士,主要研究领域为推荐系统.



王睿(1975—),男,博士,教授,博士生导师,主要研究领域为大数据,物联网,信息融合,边缘智能.