

# 基于 Word2Vec 和深度神经网络的协同推荐模型

金楠<sup>1</sup>, 王瑞琴<sup>1,2</sup>, 陆悦聪<sup>1</sup>

(1. 湖州师范学院, 信息工程学院, 浙江 湖州 313000; 2. 浙江省现代农业资源智慧管理与应用研究重点实验室, 浙江 湖州 313000)

**摘要:** 针对目前推荐系统中广泛存在的数据稀疏性、预测评分不准确和神经网络模型复杂度较高等问题, 提出深度神经网络模型 ProfileDNN. 模型借鉴自然语言处理中 Word2Vec 表征学习方法预训练物品的嵌入向量, 然后利用物品的嵌入向量构建物品和用户画像, 最后基于深度神经网络模型学习用户对物品的预测评分. 基于 3 个公共数据集的对比实验表明, ProfileDNN 模型相比同类模型, 其复杂度更低, 且推荐准确率提升最高达 1.1%.

**关键词:** 推荐系统; 嵌入向量; Word2Vec; 表征学习; 深度神经网络

**中图分类号:** TP391.3

**文献标志码:** A

**文章编号:** 1009-1734(0000)00-0000-00

## 0 引言

随着网络的迅速发展, 人类从信息匮乏时代走向了信息过载时代. 信息消费者在面对海量信息时无法轻易获取自身感兴趣的信息. 信息生产者将信息展示给合适的用户也变得越来越困难. 而个性化推荐系统是解决信息过载问题的有效方法之一. 它是根据用户的历史信息帮助用户找到喜欢的物品, 探索用户喜欢的新事物<sup>[1]</sup>. 对商家而言, 利用推荐系统能为用户提供个性化服务, 与用户建立亲密关系, 从而提高用户的信任度.

协同过滤是推荐系统中较流行且应用较广泛的推荐算法之一<sup>[2]</sup>. 协同过滤算法主要分为基于邻域的协同过滤推荐算法和基于模型的协同过滤推荐算法. 基于邻域的协同过滤推荐算法主要分为基于用户的协同过滤算法和基于物品的协同过滤算法, 其分别根据用户相似度和物品相似度为用户推荐物品. 其精度主要取决于用户和物品的数量, 在数据稀疏时精度将受到极大的影响<sup>[3-4]</sup>. 基于模型的协同过滤是目前最流行的协同过滤类型, 包括矩阵分解算法、神经网络算法等. 矩阵分解应用广泛, 包含矩阵因式分解 (Matrix Factorization, MF)<sup>[5]</sup>、概率矩阵分解 (Probabilistic matrix factorization, PMF)、奇异值分解 (Singular Value Decomposition, SVD) 等. 矩阵分解算法可解释性差, 易受稀疏数据的影响<sup>[6]</sup>.

Hash 编码和独热编码是推荐系统中常用的编码方式. Hash 编码因受嵌入冲突问题的困扰, 不可避免地会损害模型的性能<sup>[7]</sup>. 而独热编码在用户和物品数量很多时, 其特征空间会变得非常大<sup>[8]</sup>, 稀疏问题也愈发严重. 此外, 独热编码在向量空间中拥有相同的距离<sup>[9]</sup>, 无法体现向量之间的相似性. 而聚类技术、降维技术、图论技术和模糊技术是缓解数据稀疏性问题的常用技术<sup>[10]</sup>. Word2Vec 作为降维技术之一, 可在百万数量级的词典和上亿的数据集上进行高效的训练, 得到低维的、相似性的向量, 从而有效降低模型的复杂度. 随着深度学习的发展, 推荐系统中关于神经网络的研究日益增多, 从典型的自动编码器、受限波斯曼机到卷积神经网络, 再到深度神经网络, 均表现出优异的性能. 其中, 深度神经网络凭借其强大的非线性拟合能力和特征提取能力, 受到越来越多学者们的广泛关注.

本文结合 Word2Vec 模型和深度神经网络, 提出 ProfileDNN 模型. 该模型利用 Word2Vec 训练物品

**收稿日期:** 2021-09-09

**基金项目:** 浙江省自然科学基金项目 (LQ12F02008).

**通信作者:** 王瑞琴, 博士, 教授, 研究方向: 自然语言处理和推荐系统. E-mail: 02291@zjhu.edu.cn

的嵌入向量缓解数据的稀疏性问题,并通过向量之间的相似性为用户构建画像,还通过只训练高评分物品的方式,在不降低预测精度的同时节省约 46% 的训练时间.为获取用户与物品画像之间的高阶非线性交互,该模型使用深度神经网络进行充分训练,提高了评分预测的准确性.

本文的主要贡献为:

(1) 将自然语言处理模型 Word2Vec 用于个性化推荐系统,通过只训练高评分物品的方式减少了物品嵌入向量的训练时间.

(2) 充分利用物品嵌入向量之间的相似性,为用户构建画像,解决了 Word2Vec 模型在推荐系统领域无法训练得到用户画像的缺陷.

(3) 通过 Word2Vec 预训练得到用户和物品的画像,使得神经网络输入层维度较低且固定,不易受用户数量、物品数量和评分数量的影响,从而降低了模型的复杂度.

(4) 基于 3 个真实数据集的实验表明,本文模型能有效提高推荐系统预测评分的准确性.

## 1 研究背景

### 1.1 学习在推荐系统上的应用

表征学习允许机器学习和深度学习模型学习使用特征的同时,也学习如何提取特征,避免了手动提取特征的麻烦. Kang 等提出采用深度哈希嵌入方式来表示物品,该方式有效降低了向量维度,其性能优于传统哈希方法,略低于独热编码<sup>[7]</sup>. 熊回香等通过 Word2Vec 算法模型生成候选相关学者集和候选相关论文集进行推荐<sup>[11]</sup>. Vasile 等利用尺寸信息来帮助计算低维的项目嵌入,使模型更加灵活<sup>[12]</sup>. Zhao 等提出一种基于嵌入的框架,用于学习和传输电子商务中所有类型 ID 的低维表示<sup>[13]</sup>. 此外, Radosavljevic 等将表征学习应用于 APP 推荐<sup>[14]</sup>, Kenthapadi 等将表征学习应用于工作推荐<sup>[15]</sup>,均取得了不错的效果. 表征学习在特征提取和降低维度上的优异表现,使其在推荐系统中受到越来越广泛的应用.

词嵌入是自然语言处理中语言模型与表征学习技术的统称. 它把一个维数为所有词数量的高维空间嵌入一个低维的连续向量空间,以发掘单词之间的语义和语法关系<sup>[16]</sup>. 谷歌开源的 Word2Vec 采用的模型包含 CBOW 模型和 Skip-Gram 模型. 这两种模型在数据集上高效训练后得到的词向量可用于计算词与词之间的相似性. 在大量文本的训练中, Word2Vec 方法比许多经典方法拥有更好的表现.

目前, Word2Vec 方法已应用扩展到自然语言处理以外的邻域. Barken 和 Koenigstein 提出将 Item2Vec 方法用于协同过滤推荐,该方法将物品转换为嵌入向量,即使在用户信息不可用的情况下也可推断出项目对项目的关系<sup>[17]</sup>. Item2Vec 能够很好地利用 Word2Vec 模型训练得到物品之间的相似性特点,并通过计算获得物品的相似度矩阵,从而将其用于推荐,相比 SVD 算法其表现更为优异. 但 Item2Vec 方法无法获得用户向量,也不能用于评分预测. Grbovic 等首次介绍了 User2Vec 模型,该模型通过将用户视为一个“全局上下文”来同时学习项目和用户的表示,但其本质还是计算用户与物品之间的相似度,无法预测用户对物品的偏好<sup>[18]</sup>.

### 1.2 基于深度学习技术的推荐方法

深度学习具有强大的表征能力,在计算机视觉和语音识别等领域取得了出色的表现,且在自然语言处理和推荐系统中起着越来越重要的作用<sup>[19]</sup>. 一般观点认为,相比传统算法,深度学习在推荐系统中的表现更为优异<sup>[20]</sup>. 作为深度学习的一部分,深度神经网络(Deep Neural Networks, DNN)能够学习二阶特征和高阶特征交互<sup>[21]</sup>,且有足够的隐藏层和隐藏单元,可在一定光滑性假设下逼近任意函数,从而达到任意精度<sup>[22]</sup>. Huang 等提出深度结构化语义模型,利用 DNN 实现文本语义相似度的计算<sup>[23]</sup>. Elkahky 等提出多视角深度学习推荐框架,通过 4 个 DNN 分别提取用户偏好和 3 个不同领域的物品信息,然后将其映射到同一空间来计算相似度<sup>[24]</sup>. Wang 等开发了一种用于文章推荐的动态关注深度模型,该模型跨多个深度神经网络,解决了编辑为最终用户选择新闻文章时没显式选择标准的问题<sup>[25]</sup>. Ji 等<sup>[26]</sup>使用深度神经网络探索市场中各种异构数据之间的关系,发现用户的兴趣,并为用户推荐需要的物品. Alashkar 等<sup>[27]</sup>通过深度神经网络识别输入的面部

特征,为最终的化妆风格推荐提供帮助. Wang 等结合交叉网络和深度神经网络,提出 DCN 模型,有效解决了大规模稀疏数据特征工程困难等问题<sup>[28]</sup>. Guo 等提出的 DeepFM 模型使用深度神经网络学习用户和物品特征之间的交互,充分挖掘用户和物品特征之间存在的关联性,提高了点击率和预测的准确性<sup>[29]</sup>.

鉴于深度学习在推荐系统中表现出的强大生命力,本文在表征学习的基础上,充分发挥 DNN 神经网络的表征能力,以提高模型评分预测的准确性.

## 2 ProfileDNN 模型介绍

ProfileDNN 模型是通过 Word2Vec 模型中的 Skip-Gram 模型训练获得每个物品的嵌入向量来表示物品,并使用负采样方法进行优化. 但 Skip-Gram 模型无法训练获取用户的嵌入向量,因此本文利用物品嵌入向量之间存在的相似性,构建用户画像来表示用户,最后使用深度神经网络训练预测评分. 模型架构如图 1 所示.

### 2.1 Skip-Gram 模型和负采样优化

Word2Vec 主要分为两种模式:CBOW 模型和 Skip-Gram 模型. 其中, CBOW 模型可从原始句子中推断出目标词,更适用于小型数据集; Skip-Gram 模型可从目标词推断出原始句子,更适合于大型语料库. 鉴于推荐系统庞大的数据量,本文使用 Skip-Gram 模型训练物品的嵌入向量. 对物品序列  $[\tau_{w_1}, \tau_{w_2}, \tau_{w_3}, \dots, \tau_{w_i}]$ , Skip-Gram 模型的目标函数为:

$$\frac{1}{K} \sum_{i=1}^K \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{i+j} | w_i), \quad (1)$$

其中,  $c$  为上下文窗口的大小,推荐系统物品在无辅助信息的前提下不存在上下文关系,这里的  $c$  值可设为正无穷,表示在用户评分的物品序列中物品皆存在上下文关系.  $K$  为单词序列的长度. softmax 激活函数为:

$$p(w_j | w_i) = \frac{\exp(u_{w_i}^T v_{w_j})}{\sum_{k=j-c}^{k=j+c} \exp(u_i^T v_k)}, \quad (2)$$

其中,  $u_{w_i}$  为物品  $i$  作为目标训练时的嵌入向量,  $v_{w_j}$  为物品  $j$  作为上下文内容训练时的嵌入向量.

当词库的数据量十分庞大时,使用 softmax 激活函数会使每次计算都非常昂贵,在实际训练中难以应用. 噪声对比估计(Noise Contrastive Estimation, NCE)可将数据与噪声区分开,然后使用逻辑回归来估算 softmax 的对数概率<sup>[30]</sup>. 因此,本文使用负采样方法代替 softmax 激活函数. 与传统方法每次更新所有权重不同,负采样一次仅更新一个训练样本的部分权重,这大大减少了梯度在下降过程中的计算量<sup>[17]</sup>. 分层的 softmax 为:

$$p(w_j | w_i) = \sigma(u_{w_i}^T v_{w_j}) \prod_{k=1}^N \sigma(-u_{w_i}^T v_{w_k}), \quad (3)$$

其中,  $\sigma(x) = 1/(1 + \exp(-x))$ ,  $N$  为每个正样本对应的负样本数量.

将带有负采样的 Skip-Gram 应用于推荐系统中嵌入训练物品信息. 在自然语言处理中,数据集包含多个句子,句子中的单词之间存在上下文关系, Word2Vec 模型可通过学习多个句子中单词的上下文关系获得词向量. 相似的,推荐系统数据集包含多个用户的评分序列,同一用户评分序列中的物品之间存在相关性,因此可将用户的评分序列等同于句子,物品等同于单词,使用 Word2Vec 模型训练学习物品向量. 模型使用负采

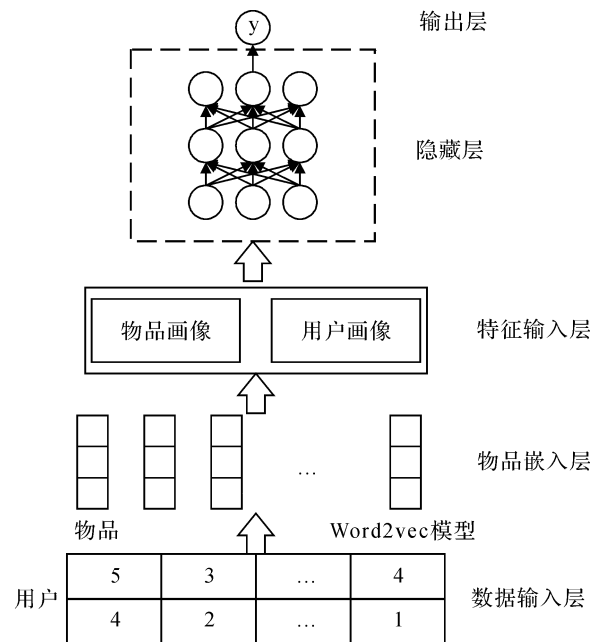


图 1 ProfileDNN 模型示意图

Fig. 1 ProfileDNN model

样方法进行优化,同一用户评分序列中的任意两个物品皆为正样本,这增加了训练样本的数量,即使稀疏的数据集也能得到充分的训练. Skip-Gram 模型的目标函数为:

$$\frac{1}{T} \sum_{t=1}^T \sum_{j \neq i} \log p(w_{t+j} | w_t) (R_{ui} > R_{uj}), \quad (4)$$

其中,  $R_{ui}$  为用户对物品的评分,  $R_{ui}$  为所有评分的平均值.

## 2.2 画像构建

利用 Word2Vec 模型训练所得的物品嵌入向量构建用户和物品的画像,并将其作为深度神经网络的输入. 基于用户偏好研究在心理学、行为学、经济学、情报学等领域的关注度,分析用户的偏好信息,以发现用户的真实信息需求,从而提供符合用户意愿和需求的个性化推荐.

利用物品嵌入向量之间存在的相似性,构建用户画像  $U$  来表示用户偏好. 假设用户最高评分的物品序列为  $[item_1, item_2, item_3, \dots, item_n]$ , 对应的物品嵌入向量序列为  $[v_1, v_2, v_3, \dots, v_n]$ , 用户积极画像  $U_{pos}$  表示用户的偏好信息,可用最高评分物品序列的平均值计算,其公式为:

$$U_{pos} = \frac{1}{N} \sum_{n=1}^N v_n, \quad (5)$$

其中,  $N$  为最高评分物品的数量,  $v_n$  为最高评分物品所对应的嵌入向量.

同理可以构建用户消极画像  $U_{neg}$ . 假设用户最低评分序列为  $[item_1, item_2, item_3, \dots, item_m]$ , 对应的物品嵌入向量序列为  $[v_1, v_2, v_3, \dots, v_m]$ . 用户消极画像  $U_{neg}$  为:

$$U_{neg} = \frac{1}{M} \sum_{m=1}^M v_m, \quad (6)$$

其中,  $M$  为最低评分物品的数量,  $v_m$  为最低评分物品所对应的嵌入向量.

用户平均画像  $U_{avg}$  为用户评分均值. 假设用户评分序列为  $[r_1, r_2, r_3, \dots, r_k]$ , 则  $U_{avg}$  的计算公式为:

$$U_{avg} = \frac{1}{k} \sum_{k=1}^K r_k, \quad (7)$$

其中,  $K$  为用户评分序列的评分数量.

最终,将  $U_{pos}$ 、 $U_{neg}$  和  $U_{avg}$  拼接构成用户画像  $U$ , 其计算公式为:

$$U = [U_{pos}, U_{neg}, U_{avg}]. \quad (8)$$

在物品画像的构建上,本文选择将物品对应的嵌入向量  $v$  与物品平均画像  $I_{avg}$  进行拼接. 物品画像  $I_{avg}$  为物品被用户评分序列  $[r_1, r_2, r_3, \dots, r_t]$  的均值,其计算公式为:

$$I_{avg} = \frac{1}{t} \sum_{t=1}^T r_t, \quad (9)$$

其中,  $T$  为物品评分序列的评分数量. 物品画像  $I$  的计算公式为:

$$I = [I_{avg}, v]. \quad (10)$$

## 2.3 DNN 网络结构

深度神经网络 DNN 按不同的位置划分,可分为输入层、隐藏层和输出层,层与层之间是全连接的. 第一层是输入层,由用户画像  $U$  和物品画像  $I$  拼接构成. 中间部分是多个隐藏层,神经元的个数随层数的增加而递减,并使用 sigmoid 函数激活. 最后一层为输出层,包含一个神经元,表示用户对物品的预测评分. 输入层  $x_0$  为:

$$x_0 = [U, I]. \quad (11)$$

模型使用 3 层隐藏层学习用户画像和物品画像的高阶非线性交互,神经元逐层递减,以减少模型训练的复杂度. 隐藏层第 1 层和第  $i$  层的输出向量分别为:

$$h_1 = \sigma(W_1 x_0 + b_1), \quad (12)$$

$$h_i = \sigma(W_i h_{i-1} + b_i), \quad (13)$$

其中,  $h_i$  为第  $i$  个隐藏层的输出向量,  $W_i$  为权值矩阵,  $b_i$  为偏置项,  $\sigma$  为 sigmoid 激活函数.

输出层仅有一个神经元,采用全连接的方式得到模型的输出,其公式为:

$$\hat{y} = W_n h_{n-1} + b_n, \quad (14)$$

其中, $\hat{y}$  为用户对物品评分的预测, $n$  为深度神经网络隐藏层的个数.

#### 2.4 模型训练

采用 Skip-Gram 模型训练物品嵌入向量时,将单个用户评分的全部物品视为正样本,未评分的物品视为负样本,每个正样本随机选取 5 个负样本,模型训练 100 次.神经网络部分采用批次训练的方式,批次大小为 128,同时每次训练打乱训练集的顺序.模型损失函数为:

$$L = \frac{1}{m} \sum_{i=1}^m |\hat{y}_i - y_i|, \quad (15)$$

其中, $L$  为损失函数, $m$  为评分总数, $\hat{y}_i$  为 ProfileDNN 模型预测所得评分, $y_i$  为真实评分.

模型使用标准梯度下降法(Gradient Descent, GD)训练,网络层神经元的个数随层数的增加而不断减少,并使用 dropout 方法以防止模型过拟合,模型训练 1 000 次.参数设置如表 1 所示.

表 1 参数设置

Tab. 1 Parameters

参数	参数描述	FilmTrust	Movielen 100 k	Movielen 1 M
$c$	上下文距离	$+\infty$	$+\infty$	$+\infty$
$N$	负采样个数	5	5	5
min_count	最小次数	1	1	1
batch_size	批次数量	128	128	128
embedding_size	向量维度	100	100	200
epoch	迭代次数	1 000	1 000	1 000
keep_prob	保留概率	0.5	0.7	0.8
lr	学习率	0.01	0.01	0.01
optimizer	优化器	GD	GD	GD

### 3 实验结果与分析

在 3 个公开数据集的基础上对本文模型和现有模型进行对比分析,并通过实验验证本文模型的有效性.下面给出数据集的描述、实验设置、性能评价指标和实验结果的分析.

#### 3.1 实验数据集

本次实验选取的数据集为:FilmTrust<sup>[31]</sup>、MovieLens 1M 和 MovieLens 100 k<sup>[32]</sup>,如表 2 所示.MovieLens 1 M 包含 6 040 个用户对 3 706 部电影的 1 000 209 个显式评分,评分密度为 4.47%. MovieLens 100 k 包含 943 个用户对 1 682 部电影的 100 000 个显式评分,评分密度为 6.30%. FilmTrust 包含 1 508 个用户对 2 071 部电影的 35 497 个显式评分,评分密度为 1.14%. 3 个数据集评分都较稀疏,其中 FilmTrust 数据集评分最不密集,平均每个用户仅对 1.1% 的物品进行评分.数据集划分时将 90% 的数据作为训练集,10% 的数据作为测试集.

表 2 数据集统计信息

Tab. 2 Statistics of datasets

属性	FilmTrust	MovieLens 100 k	MovieLens 1 M
用户数	1 508	943	6 040
物品数	2 071	1 682	3 706
评分数	35 497	100 000	1 000 209
评分范围	[0.5, 4.0]	[1, 5]	[1, 5]
评分密度/%	1.14	6.30	4.47

### 3.2 对比模型

ProfileDNN 在数据选择上只使用用户对物品的评分信息,没有使用用户和物品的特征信息,也没有考虑时间因素. 本文将该方法与本文复现的同类型的协同过滤算法、SVD 算法、DNN 神经网络等模型进行比较.

**ItemCF**<sup>[33]</sup>: 基于物品的协同过滤推荐计算物品之间相似度,根据用户历史喜欢的物品向用户推荐相似度高的物品.

**PMF**<sup>[34]</sup>: 概率矩阵分解是推荐系统的基础算法之一. 传统的协同过滤算法既不能处理非常大的数据集,也不能轻松处理稀疏的数据集,而 PMF 模型在大型、稀疏且不平衡的数据集上具有很好的表现.

**DNN**<sup>[35]</sup>: 相比单层神经网络,深度神经网络可以模拟世界中更加复杂的情形,采用拼接用户 ID 将物品 ID 的 one-hot 编码方式作为深度神经网络的输入,使用深度神经网络训练获得结果.

**SVD**<sup>[36]</sup>: 奇异值分解推荐算法在预测准确性、稳定性上具有明显的优势. SVD 算法采用对评分矩阵分解并点积的方式计算预测评分.

**NeuMF**<sup>[35]</sup>: NeuMF 结合传统的矩阵分解算法和多层感知机,可同时抽取低维和高维的特征,具有优异的推荐效果.

### 3.3 性能评估指标

均方根误差 (Mean Absolute Error, MAE) 和均方根误差 (Root Mean Square Error, RMSE) 是衡量变量精度的两个最常用的指标. 本文以这两个指标作为衡量模型好坏的标准. MAE 表示绝对误差的平均值,可反映预测值与真实评价之间的偏差,偏差越小则预测精度越高,模型的推荐效果越好<sup>[37]</sup>. 其计算公式为:

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |\hat{y}_i - y_i|, \quad (16)$$

其中,  $\hat{y}_i$  为 ProfileDNN 模型预测所得评分,  $y_i$  为真实评分.

RMSE 也可测量误差的平均大小,表示预测值与实际观测之间平方差异平均值的平方根. 其计算公式为:

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (\hat{y}_i - y_i)^2}. \quad (17)$$

浮点运算数 (floating-point operations, FLOPs)<sup>[38]</sup> 用来衡量算法或模型的复杂度. 其计算公式为:

$$\text{FLOPs} = (2I - 1)O, \quad (18)$$

其中,  $I$  为输入的维度,  $O$  为输出的维度.

### 3.4 实验结果分析

#### 3.4.1 不同模型的推荐性能对比

ProfileDNN 模型与对比模型在 3 组数据集上运行后得到的 RMSE 和 MAE 如表 3 和表 4 所示. ItemCF 算法、PMF 算法和 SVD 算法基于评分矩阵进行推荐,受数据稀疏性的影响,其评分准确率较低. DNN 模型使用独热编码表示用户物品,容易高维稀疏. ProfileDNN 通过画像方式降低维度,以提升评分准确率. 相比结合矩阵分解和 DNN 的 NeuMF 模型, ProfileDNN 模型的评分准确率有所提升,最高提升 1.1%.

表 3 ProfileDNN 模型与对比模型的 RMSE 值

Tab. 3 RMSE of ProfileDNN model and comparison models

Models	FilmTrust	MovieLens 100 k	MovieLens 1 M
ItemCF	0.882 4	0.969 7	0.955 6
PMF	0.823 1	0.933 5	0.897 1
SVD	0.822 8	0.922 3	0.907 5
DNN	0.824 2	0.929 6	0.880 0
NeuMF	0.810 6	0.906 2	0.859 3
ProfileDNN	0.804 7	0.896 0	0.855 0

表 4 ProfileDNN 模型与对比模型的 MAE 值

Tab. 4 MAE of Profile DNN model and comparison models

Models	FilmTrust	MovieLens 100 k	MovieLens 1 M
ItemCF	0.657 3	0.749 1	0.731 3
PMF	0.640 5	0.753 2	0.735 6
SVD	0.634 6	0.746 7	0.721 8
DNN	0.648 2	0.725 0	0.692 8
NeuMF	0.632 6	0.712 3	0.687 6
ProfileDNN	0.623 3	0.703 7	0.676 4

表 5 为各个神经网络模型的复杂度对比.从表 5 可以看出,相比使用独热编码的 DNN 模型和 NeuMF 模型,ProfileDNN 模型的 FLOPs 值和参数数量最小,复杂度最低.这是因为在数据庞大且稀疏时,独热编码表示的方式会使输入层的维度较大,而通过 Word2Vec 模型预训练得到的用户和物品画像维度是固定且低维的,输入层的复杂度相对较低,从而降低了模型训练的复杂度.

表 5 神经网络模型复杂度对比

Tab. 5 Comparison of the complexity of neural network models

Models	FLOPs	Params
DNN	155 928	78 049
NeuMF	1 748 674	875 481
ProfileDNN	40 856	20 513

### 3.4.2 维度对 ProfileDNN 模型推荐结果的影响

表 6 为当嵌入向量维度为 100 和 200 时,ProfileDNN 模型在 3 个数据集上的表现.实验表明,FilmTrust 和 MovieLens 数据集的结果差异并不明显.由于 FilmTrust 和 MovieLen 100 k 数据集较小,本文选取 100 作为嵌入向量维度.MovieLen 1 M 选取 200 作为嵌入向量维度.

表 6 Embedding size 为 100 和 200 时 ProfileDNN 模型表现

Tab. 6 Performance of ProfileDNN for embedding size = 100 and embedding size = 200

数据集	Embedding size=100		Embedding size=200	
	MAE	RMSE	MAE	RMSE
FilmTrust	0.623 3	0.804 7	0.626 1	0.806 2
MovieLen100k	0.703 7	0.896 0	0.712 1	0.902 2
MovieLen 1M	0.679 6	0.857 1	0.676 4	0.855 0

### 3.4.3 迭代次数对 ProfileDNN 模型推荐结果的影响

ProfileDNN 模型在 FilmTrust、MovieLen 100 k 和 MovieLen 1 M 3 个数据集上的 RMSE 和 MAE 值随迭代次数变化的曲线如图 2 和图 3 所示.从图 2 和图 3 可以看出,在使用 dropout 方法后,模型迭代 200 次可以得到理想的结果.在随后的 800 次迭代中模型基本保持稳定,RMSE 和 MAE 值没有较大的变化.

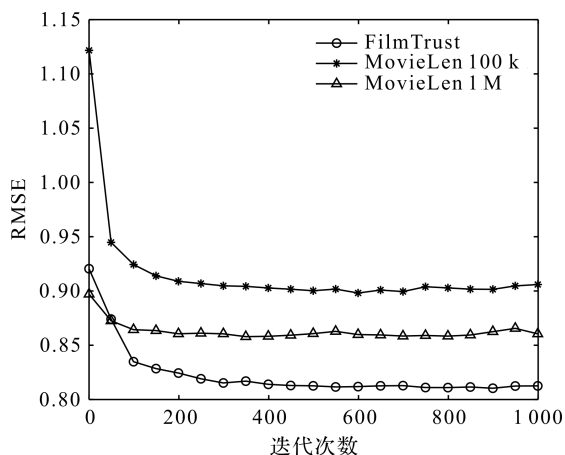


图 2 迭代次数对 RMSE 的影响

Fig. 2 Effect of iterations in term of RMSE

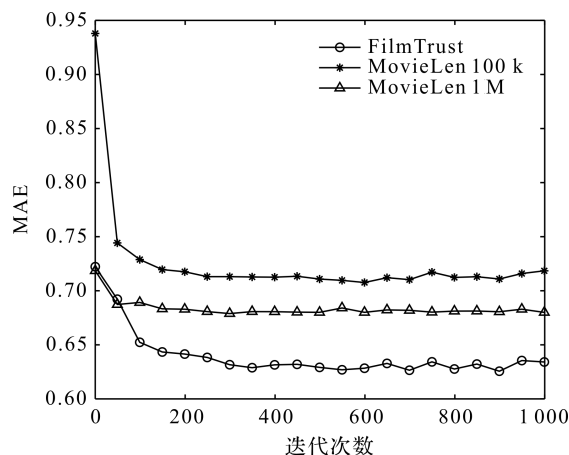


图 3 迭代次数对 MAE 的影响

Fig. 3 Effect of iterations in term of MAE

#### 3.4.4 保留概率对 ProfileDNN 模型推荐结果的影响

Dropout<sup>[39]</sup>是指神经元保留在网络中的比率,是解决神经网络过度拟合的有效方法,可以提高模型的鲁棒性.如图 4 和图 5 所示,当保留概率为 1 时,模型在测试集上的 MAE 和 RMSE 值均较高.这是因为当保留概率为 1 时,模型没有使用 dropout 方法保留所有神经元,导致出现过拟合问题,而在使用 dropout 方法后,模型的推荐效果显著提升.

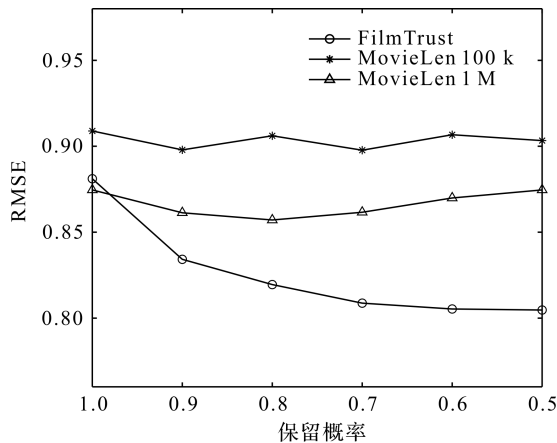


图 4 保留概率对 RMSE 的影响

Fig. 4 Effect of dropout in term of RMSE

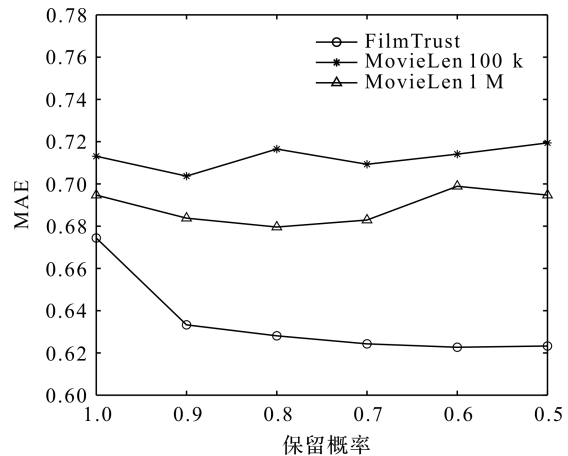


图 5 保留概率对 MAE 的影响

Fig. 5 Effect of dropout in term of MAE

#### 3.4.5 模型在稀疏数据上的稳定性测试

为验证模型在稀疏数据上的稳定性,本文将 MovieLens 100 k 数据集按不同比例(训练集与测试集的比例为 9 : 1, 7 : 3, 1 : 1, 3 : 7)划分进行对比实验,结果如表 7 所示.在数据稀疏时各模型的 RMSE 值均有一定上升,但本文模型的 RMSE 值在 4 种不同划分数据集上仍优于对比模型.实验证明,本文模型在稀疏数据集上仍能有效提取特征,保证了评分预测的准确性.

表 7 MovieLens 100 k 数据集上不同数据划分对各模型 RMSE 值的影响

Tab. 7 The impact of different data partitions on various models on the MovieLens 100 k

Models	训练集与测试集的比例			
	9 : 1	7 : 3	1 : 1	3 : 7
ItemCF	0.969 7	1.117 8	1.118 6	1.139 2
DNN	0.929 6	0.932 5	0.944 5	0.968 0
NeuMF	0.906 2	0.938 1	0.944 7	0.960 6
ProfileDNN	0.896 0	0.924 1	0.940 3	0.959 5

## 4 结语

随着互联网的发展和科技的进步,推荐系统得到了人们的广泛关注,人们对推荐系统的要求也逐渐提高.针对推荐系统中的数据稀疏问题,本文借鉴自然语言处理方法 Word2Vec 学习每个物品的嵌入向量,这种方法可以有效地将物品的表示方法从高维稀疏的独热编码转变成低维稠密的嵌入向量,并充分利用物品嵌入向量之间存在的相似性,构建用户画像表示用户.本文提出的模型使用深度神经网络训练的方式提升评分预测的准确性.实验结果表明,本模型可有效解决传统方法中存在的稀疏性问题,能显著提高推荐系统评分预测的准确性.

本文在构造用户画像时尝试过不同的构造方法,不同方式的结果差异也较大,最终选取效果最为优异的方法作为本文画像的构造方式.未来的研究将继续探索更优秀的构造用户画像方法,以使模型拥有更佳的表现.



---

**参考文献:**

- [ 1 ] 黄立威,江碧涛,吕守业,等. 基于深度学习的推荐系统研究综述[J]. 计算机学报,2018,41(7):1 619 - 1 647.
- [ 2 ] ZHOU M, HE F, GU M, et al. Translation - based model checking for PLC programs[C]//2009 33rd Annual IEEE International Computer Software and Applications Conference, 2009:553 - 562.
- [ 3 ] LINDEN G, SMITH B YORK J. Amazon. com recommendations: Item - to - item collaborative filtering[J]. IEEE Internet computing, 2003, 7(1):76 - 80.
- [ 4 ] 翁小兰,王志坚. 协同过滤推荐算法研究进展[J]. 计算机工程与应用,2018,54(1):25 - 31.
- [ 5 ] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems[J]. Computer, 2009, 42(8):30 - 37.
- [ 6 ] KHAN Z, ILTAF N, AFZAL H, et al. Enriching non - negative matrix factorization with contextual embeddings for recommender systems[J]. Neurocomputing, 2020, 380:246 - 258.
- [ 7 ] KANG W C, CHEMG D Z, YAO T, et al. Learning to embed categorical features without embedding tables for recommendation[J]. 2020.
- [ 8 ] 梁杰,陈嘉豪,张雪芹,等. 基于独热编码和卷积神经网络的异常检测[J]. 清华大学学报(自然科学版), 2019, 59(7):523 - 529.
- [ 9 ] ZHAO K, LI Y, SHUAI Z, et al. Learning and transferring ids representation in e - commerce[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018:1 031 - 1 039.
- [10] 陈金晨. 缓解数据稀疏性问题的传统技术研究综述[J]. 数据通信, 2021(2):32 - 35.
- [11] 熊回香,李跃艳. 基于 Word2vec 的学者推荐与跨语言论文推荐模型研究[J]. 情报科学, 2019, 37(12):19 - 26.
- [12] VASILE F, SMIRNOVA E, CONNEAU A. Meta - prod2vec: Product embeddings using side - information for recommendation[C]//Proceedings of the 10th ACM Conference on Recommender Systems, 2016:225 - 232.
- [13] ZHAO K, LI Y, SHUAI Z, et al. Learning and transferring ids representation in e - commerce[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018:1 031 - 1 039.
- [14] RADOSAVLJEVIC V, GRBOVIC M, DJURIC N, et al. Smartphone app categorization for interest targeting in advertising marketplace[C]//Proceedings of the 25th International Conference Companion on World Wide Web, 2016:93 - 94.
- [15] KENTHAPADI K, LE B, VENKATARAMAN G. Personalized job recommendation system at linkedin: Practical challenges and lessons learned[C]//Proceedings of the Eleventh ACM Conference on Recommender Systems, 2017:346 - 347.
- [16] 张晓娟. 利用嵌入方法实现个性化查询重构[J]. 情报学报, 2018, 37(6):621 - 630.
- [17] BARKAN O, KOENIGSTEIN N. Item2vec: Neural item embedding for collaborative filtering[C]//2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), IEEE, 2016:1 - 6.
- [18] GRBOVIC M, RADOSAVLJEVIC V, DJURIC N, et al. E - commerce in your inbox: Product recommendations at scale [C]//Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining, 2015:1 809 - 1 818.
- [19] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7 553):436 - 444.
- [20] SINGHAL A, SINHA P, PANT R. Use of deep learning in modern recommendation system; A Summary of recent works[J]. International Journal of Computer Applications, 2017, 180(7):17 - 22.
- [21] XU J, HE X, LI H. Deep learning for matching in search and recommendation[C]//The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018:1 365 - 1 368.
- [22] VEIT A, WILBER M J, BELONGIE S. Residual networks behave like ensembles of relatively shallow networks[J]. Advances in neural information processing systems, 2016, 29:550 - 558.
- [23] HUANG P S, HE X, GAO J, et al. Learning deep structured semantic models for web search using clickthrough data [C]//Proceedings of the 22nd ACM international conference on Information & Knowledge Management, 2013:2 333 - 2 338.
- [24] ELKAHKY A M, SONG Y, HE X. A multi - view deep learning approach for cross domain user modeling in recommendation systems[C]//Proceedings of the 24th international conference on world wide web, 2015:278 - 288.
- [25] WANG X, YU L, REN K, et al. Dynamic attention deep model for article recommendation by learning human editors'

- demonstration[C]//Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining. 2017;2 051 – 2 059.
- [26] JI H, SONG L, JIANG L, et al. ReCom: An efficient resistive accelerator for compressed deep neural networks[C]//2018 Design, Automation & Test in Europe Conference & Exhibition (DATE). IEEE, 2018;237 – 240.
- [27] ALASHKAR T, JIANG S, WANG S, et al. Examples – rules guided deep neural network for makeup recommendation [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2017;941 – 947.
- [28] WANG R, FU B, FU G, et al. Deep & cross network for ad click predictions[M]//Proceedings of the ADKDD'17. 2017;1 – 7.
- [29] GUO H, TANG R, YE Y, et al. DeepFM: A factorization – machine based neural network for CTR prediction[C]//Twenty – Sixth International Joint Conference on Artificial Intelligence. 2017;1 725 – 1 731.
- [30] GUTMANN M U, HYVARINEN A. Noise – contrastive estimation of unnormalized statistical models, with applications to natural image statistics[J]. The journal of machine learning research, 2012, 13(1):307 – 361.
- [31] GOLBECK J, HENDLER J. Filmtrust: Movie recommendations using trust in web – based social networks[C]//Proceedings of the IEEE Consumer Communications and Networking Conference. 2006, 96(1):282 – 286.
- [32] HARPER F M, KONSTAN J A. The movielens datasets: History and context[J]. Acm transactions on interactive intelligent systems (tiis), 2015, 5(4):1 – 19.
- [33] LINDEN G, SMITH B, YORK J. Amazon. com recommendations: Item – to – item collaborative filtering[J]. IEEE Internet computing, 2003, 7(1):76 – 80.
- [34] SALAKHUTDINOV R. Probabilistic Matrix Factorization[C]// Proceedings of the 20th International Conference on Neural Information Processing System. 2007;1 257 – 1 264.
- [35] HE X, LIAO L, ZHANG H, et al. Neural collaborative filtering[C]//Proceedings of the 26th international conference on world wide web. 2017;173 – 182.
- [36] ZHENG S, DING C, NIE F. Regularized Singularvalue decomposition and application to recommender system[J]. 2018. <https://arxiv.org/abs/1804.05090>.
- [37] 郭筑城, 黄旭, 曾孟佳. 一种基于 LPP – BP 神经网络的季节性  $P_{2.5}$  预测方法[J]. 湖州师范学院学报, 2022, 44(2):47 – 55.
- [38] MOLCHANOV P, TYREE S, KARRAS T, et al. Pruning convolutional neural networks for resource efficient transfer learning[C]//International Conference on Learning Representations. 2017;13 – 17.
- [39] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The journal of machine learning research, 2014, 15(1):1 929 – 1 958.

## Collaborative Recommendation Model Based on Word2vec and Deep Neural Network

JIN Nan<sup>1</sup>, WANG Ruiqin<sup>1,2</sup>, LU Yuecong<sup>1</sup>

(1. School of Information Engineering, Huzhou University, Huzhou 313000, China)

(2. Zhejiang Province Key Laboratory of Smart Management & Application of Modern Agricultural Resources, Huzhou 313000, China)

**Abstract:** Score prediction has always been the core issue in the research process of recommender systems. In view of the widespread data sparsity problem in the current recommendation system, the inaccurate prediction score and the high complexity of the neural network model, a deep neural network model ProfileDNN is proposed. The model draws on the Word2Vec characterization learning method in natural language processing to pre – train the embedding vector of the item, then uses the item embedding vector to construct the item and user portrait, and finally learns the user's predicted score for the item based on the deep neural network model. Comparative experimental studies on three public data sets show that the ProfileDNN model is less complex than similar models, and the recommendation accuracy rate is improved by up to 1.1%.

**Keywords:** recommended system; embedding; Word2vec; representation learning; deep neural network

[责任编辑 高俊娥]