

# 基于用户向量化表示和注意力机制的神经网络推荐模型

郭旭<sup>1</sup> 朱敬华<sup>1,2</sup>

(黑龙江大学计算机科学技术学院 哈尔滨 150080)<sup>1</sup>

(黑龙江省数据库与并行计算重点实验室 哈尔滨 150080)<sup>2</sup>

**摘要** 随着互联网应用的蓬勃发展,推荐系统作为解决信息过载的有效手段,成为了工业界与学术界的研究热点。面向用户隐式反馈的传统推荐算法主要基于协同过滤和排序学习等方法,但这些方法未充分利用用户行为中的隐式反馈特征。文中提出了一种基于神经网络的用户向量化表示模型,其能够充分利用用户的异构的隐式反馈行为特征。同时,借鉴机器翻译中的 self-attention 机制,设计了一种神经注意力推荐模型,其融合用户向量化表示和用户-项目交互的动态时序特征以提高推荐系统的性能。在公开数据集上进行对比实验,通过召回率、准确率、NDCG 3 个指标评价推荐性能。结果表明,与其他面向隐式反馈的推荐模型相比,所提推荐模型具有更好的推荐性能,并且对用户行为特征具有很好的泛化能力。

**关键词** 推荐系统,神经网络,注意力机制,隐式反馈,深度学习

**中图分类号** TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.08.018

## Deep Neural Network Recommendation Model Based on User Vectorization Representation and Attention Mechanism

GUO Xu<sup>1</sup> ZHU Jing-hua<sup>1,2</sup>

(School of Computer Science and Technology, Heilongjiang University, Harbin 150080, China)<sup>1</sup>

(Key Laboratory of Database and Parallel Computing of Heilongjiang Province, Harbin 150080, China)<sup>2</sup>

**Abstract** With the rapid development of Internet application, recommendation system, as an effective measure to solve information overloading, has become a research hot topic in industry and academia. Traditional recommendation algorithms for users' implicit feedback are mainly based on collaborative filtering and learning-to-rank method, which do not make full use of the implicit feedback features in users' behaviors. In this paper, a users' vectorization representation model based on neural network was proposed, which can make full use of heterogeneous implicit feedback features of users' behaviors. At the same time, referring to the self-attention mechanism in machine translation, this paper designed a neural attentive recommendation model which integrates the dynamic temporal features of user-item interaction and user vectorization representation, to improve the performance of the recommendation system. The comparison experiment is conducted on two public datasets, and the recommendation performance is evaluated by recall, precision and NDCG. Compared with other recommendation models for implicit feedback, the proposed recommendation model has better recommendation performance and better generalization ability.

**Keywords** Recommendation system, Neural networks, Attention mechanism, Implicit feedback, Deep learning

## 1 引言

随着互联网应用的蓬勃发展和网络数据的爆炸式增长,推荐系统在解决信息过载问题方面发挥着越来越大的作用。传统基于评分的推荐算法中,多数利用显式反馈信息,少数利用隐式反馈信息。现实中,用户对项目的评分、喜爱等显式反馈信息十分稀疏,同时在互联网中却产生了浏览、点击、收藏等大量隐式反馈信息,这些信息包含了用户的项目偏好和行为习惯。如何有效利用各种隐式反馈信息构建推荐系统,以为用户提供更好的推荐体验,是推荐系统研究中一个十分重

要且有实际意义的问题。

传统的面向用户隐式反馈的推荐算法大多是基于项目的协同过滤以及排序学习方法,对隐式反馈中用户的行为特征利用不充分。而且,用户行为数据稀疏会导致模型的训练十分困难。

近年来,神经网络模型在计算机视觉(CV)、自然语言处理(NLP)等机器学习领域显示出了优越的性能。与浅层模型相比,深层模型能更好地从数据中提取抽象特征。基于注意力的循环神经网络(RNN)被证明能够更好地建模序列数据中的长期记忆,而基于注意力的卷积神经网络(CNN)能

能够从输入中识别出与待求解问题最相关的信息,因此注意力机制被广泛应用于多层感知机(MLP)、循环神经网络、卷积神经网络和其他深度学习模型中,推动了计算机视觉、自然语言处理等领域的研究。将注意力机制应用于推荐系统,能够捕获最相关的项目特征,提高推荐质量,增强模型的可解释性。

本文提出了一种基于用户向量化表示和注意力机制的神经网络推荐模型。首先,结合多种用户行为特征构建面向隐式反馈的用户向量化表示模型。该模型基于多层感知机神经网络,以二分类模拟用户与项目的交互,并根据用户的历史行为学习用户的向量化表示模型。其次,为捕获用户行为的时序动态性和用户偏好的权重,本文借鉴 Google 的 self-attention<sup>[1]</sup> 机制,构建了融合用户向量化表达和行为时序特征的注意力网络,以提高推荐模型的性能。

本文的贡献主要包括:

- 1)提出了一种面向隐式反馈的用户向量化模型,对用户的多种隐式反馈行为特征进行建模,学习用户的潜在向量表达。
- 2)提出了一种融合用户向量化表达和注意力机制的推荐模型,利用时序位置编码表达用户行为的时序特征,捕捉用户短期的动态兴趣项目,提高了推荐质量。
- 3)在公开数据集上进行实验,验证了所提推荐模型的推荐性能和隐式反馈行为特征的泛化能力。

## 2 相关工作

### 2.1 面向隐式反馈的推荐

传统的隐式反馈推荐算法主要基于协同过滤和排序学习等机器学习方法。ItemCF 算法<sup>[2]</sup>是一种基于项目的协同过滤推荐算法,根据所有用户的历史偏好计算项目相似性,并向用户推荐与其喜欢的项目相似的项目。Ostuni 等<sup>[3]</sup>提出了一种基于隐式数据的 top-N 推荐方法,该方法借助梯度提升决策树(GBDT)训练得到排序函数。Zimdars 等<sup>[4]</sup>提出了一种基于马尔可夫链的序列推荐模型。Rendle 等<sup>[5]</sup>提出了一种基于贝叶斯后验优化技术的隐式反馈个性化排序框架 BPR,与矩阵分解不同,BPR 通过构建成对排序损失来优化传统的矩阵分解模型。

### 2.2 基于深度神经网络的推荐

AutoRec<sup>[6]</sup>推荐模型基于自动编码器,通过最小化重构误差来更新模型参数。Elkahky 等<sup>[7]</sup>利用深度神经网络抽象用户的行为特征,提出了一种多视角深度神经网络模型。He 等<sup>[8]</sup>将用户项目的交互抽象为二分类问题,利用多层感知机模拟历史交互,并学习用户与未交互项目的潜在交互概率。

循环神经网络(RNN)由于能很好地处理时序信息,被广泛用于序列预测中。Hidasi 等<sup>[9]</sup>将推荐问题近似处理为序列预测问题,利用循环神经网络建模用户与项目的交互行为序列。

### 2.3 基于注意力机制的推荐

Gong 等<sup>[10]</sup>提出了基于注意力机制的卷积神经网络推荐模型,并将其用于微博的 Hashtag 推荐。Li 等<sup>[11]</sup>提出了融合注意力机制的 RNN 模型来捕捉文本的序列特征。Xiao 等<sup>[12]</sup>提出了 AFM 推荐模型,将注意力机制应用于因子分解机。

## 3 用户向量化表示模型

本节介绍如何结合多种用户行为特征来构建面向隐式反馈的用户向量化表示模型。模型的构建过程如图 1 所示,包含两个步骤,即用户行为特征 embedding 和用户向量化表示学习。

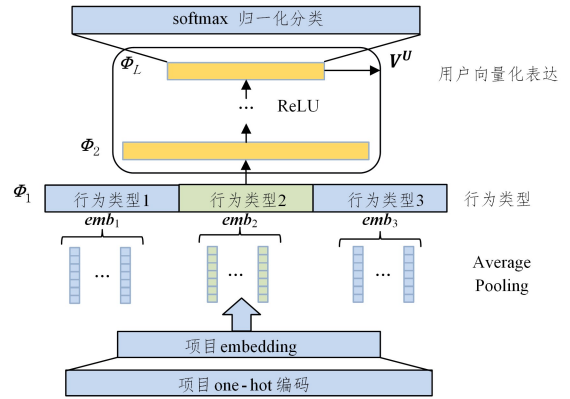


图 1 用户向量化表示的构建模型

Fig. 1 Model of building user vectorization representation

### 3.1 用户行为特征 embedding

给定一个目标用户  $U$  和该用户的历史行为记录中有交互的项目 ID。例如,在用户的电商购物记录中,每个用户都有点击/浏览、收藏、购买交易等行为,我们直观而合理地假设用户与项目的不同交互行为代表着用户不同的行为特征。因此,我们设计了一个用户行为特征的 embedding 层。具体地说,我们首先利用神经网络处理离散特征的通用方法将项目 ID 转化为 one-hot 编码的形式,one-hot 编码向量的维度即为数据集中项目的数量,它是一个高维度的稀疏向量。Embedding 层本质上是一个全连接层,将项目 ID 的 one-hot 编码从高维稀疏向量映射为低维稠密向量,这有利于神经网络对特征的提取和抽象。得到项目的稠密 embedding 向量后,针对向量,我们采用 average pooling 策略分别聚合用户历史行为记录中的不同交互类型的项目,以此作为用户的不同行为特征向量,连接这些行为特征向量形成多层感知机的输入层。

### 3.2 用户的向量化表示

为了更好地将用户行为特征抽象为用户向量,我们把用户与项目的交互预测作为二分类问题,通过全连接的 ReLU 层从用户的行为特征组合向量中学习得到潜在的用户向量化表达。ReLU 函数作为神经元的激活函数,被证明不会导致过饱和;同时它支持稀疏的激活,非常适合稀疏的数据,可使模型不至于过拟合。对于本文模型,具体来说,就是将用户交互的项目作为训练正样本,使用 softmax 输出层得到归一化的分类概率,以交叉熵损失作为目标函数来更新网络参数。提取最后一个 ReLU 层输出向量表示作为用户的向量化表达。对于每个用户,通过对其交互历史记录的学习,可以得到该用户完整的向量化表示。

在模型中, $emb_1, emb_2, \dots, emb_n$  分别代表不同表示行为特征的 embedding 向量,模型的输入层及隐含层分别表示为:

$$z_1 = \Phi_1(emb_1, emb_2, emb_3) = \begin{bmatrix} emb_1 \\ emb_2 \\ emb_3 \end{bmatrix} \quad (1)$$

$$\Phi_2(z_1) = \sigma(W_2 z_1 + b_2) \quad (2)$$

...

$$\Phi_L(z_{L-1}) = \sigma(W_L z_{L-1} + b_L) \quad (3)$$

其中,  $W_i, b_i$  分别代表第  $i$  层的权重矩阵和偏置量,  $\sigma$  代表全连接层的激活函数。在本文中, 我们设置激活函数为 ReLU 函数:

$$\text{ReLU}(x) = \max(0, x) \quad (4)$$

最后一层的激活函数 ReLU 的输出  $z_L$  即为模型所要得到的用户向量化表达  $V^U$ 。

模型的输出为 softmax 的归一化多分类概率, 把用户对项目交互的可能性以归一化概率的形式分类到所有项目上, 结合 softmax 的输出形式, 可以得到在目标项目上的分类概率为:

$$P_i = e^{x_i} / \sum_{k=1}^M e^{x_k} \quad (5)$$

其中,  $x_i$  为 softmax 全连接层在第  $i$  个项目上的输出值,  $M$  为项目集合的大小。

从图 1 中可以看到, ReLU 层输出得到的用户向量  $V^U$  通过 softmax 隐含层的权重矩阵  $W$  可以映射为用户对项目的交互概率评分。与传统的基于评分的矩阵分解方法对照, 可以直观地看出, 权重矩阵的列向量类似于矩阵分解中的项目向量, ReLU 层输出的用户向量  $V^U$  类似于矩阵分解中的用户向量。

#### 4 基于注意力机制的推荐模型

在第 3 节中, softmax 层的归一化分类概率已经可以给出用户对候选项目的粗略排名。但是用户向量化表达模型仅利用了静态历史记录, 不能很好地反映行为的动态时序特性, 也不能捕捉用户对项目的短期兴趣权重。为了提高模型的推荐质量, 我们借鉴 Google 提出的 self-attention 机制, 提出了一种融合用户向量化表达和注意力机制的推荐模型 (UserVec-Att), 进一步提高了推荐质量。UserVec-Att 推荐模型如图 2 所示。

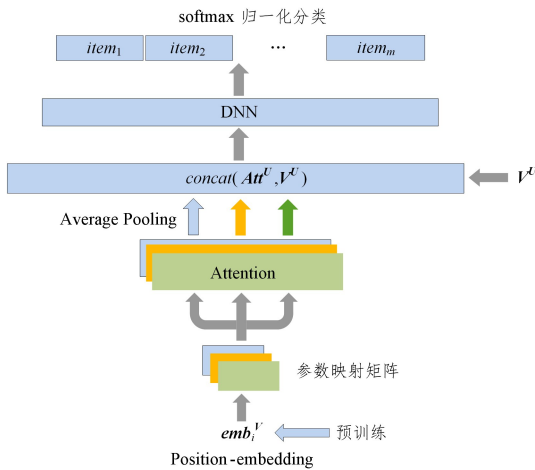


图 2 UserVec-Att 推荐模型

Fig. 2 UserVec-Att recommendation model

UserVec-Att 推荐模型利用向量化表达模型中的项目

embedding 预训练 UserVec-Att 模型的项目 embedding 层, 对用户与项目交互的时间戳做位置 embedding 编码, 并与项目 embedding 向量相加构成新的项目 embedding 向量。通过 Attention 层构建注意力机制, 生成用户的短期动态兴趣项目的向量化表达。融合用户向量表达, 经过模型的 DNN 层输出得到用户交互项目的概率分布, 根据概率值进行推荐。

关于用户与项目交互的时间戳的位置 embedding 编码, 我们采用三角位置编码的方式编码时间戳, 编码公式如下:

$$PE_{2i}(t) = \sin(t/10000^{\frac{2i}{K}}) \quad (6)$$

$$PE_{2i+1}(t) = \cos(t/10000^{\frac{2i}{K}}) \quad (7)$$

其中,  $t$  代表离散的时间戳,  $K$  代表位置 embedding 向量的维度,  $PE_i(t)$  代表位置 embedding 向量第  $i$  维的元素值。融合项目向量与注意力机制的 Attention 函数的定义为:

$$\begin{aligned} \text{Attention}(\text{emb}_i^V, \text{emb}_i^V, \text{emb}_i) \\ = \sum_{i=1}^m \frac{1}{Z} \exp\left(\frac{(\text{emb}_i^V, \text{emb}_i^V)}{\sqrt{D}}\right) \text{emb}_i \end{aligned} \quad (8)$$

其中,  $D$  为  $\text{emb}_i^V$  向量的维度,  $Z$  为归一化因子,  $\text{emb}_i^V$  代表第  $i$  个项目 embedding 向量  $\text{emb}_i$  与位置 embedding 向量  $PE(t)$  相加后得到的新的项目 embedding 向量, 即

$$\text{emb}_i^V = \text{emb}_i + PE(t) \quad (9)$$

为使项目 embedding 向量  $\text{emb}_i^V$  映射在不同的参数空间, 我们使用不同的参数矩阵分别映射  $\text{emb}_i^V$ , 从而得到注意力向量:

$$\text{Att}_{1:H} = \text{Attention}(W_Q X, W_K X, X) \quad (10)$$

其中,  $W_Q, W_K$  分别是  $\text{emb}_i^V$  的不同权重矩阵;  $X$  是由用户行为序列  $\text{emb}_i^V$  构成的矩阵。我们采用平均加权得到的注意力向量, 得到用户  $U$  的短期兴趣项目的向量化表达:

$$\text{Att}^U = \frac{1}{H} \sum_{h=1}^H \text{Att}_h \quad (11)$$

其中,  $H$  是项目序列的长度。我们将用户短期兴趣的向量化表达  $\text{Att}^U$  与第 3 节得到的用户向量  $V^U$  做向量连接, 即  $\text{concat}(\text{Att}^U, V^U)$ , 构成 DNN 网络的输入, 以此捕获用户动态的短期兴趣项目特征与静态的长期历史兴趣项目特征。以 softmax 层作为模型输出, 输出用户对项目的交互概率评分, 并以此作为项目推荐的排序依据。我们以交叉熵损失作为目标函数来更新模型参数, 交叉熵损失函数的定义如下:

$$L(p, q) = - \sum_{i=1}^m p_i \log(q_i) \quad (12)$$

其中,  $q$  是预测的项目概率分布,  $p$  是真实的项目概率分布。

#### 5 实验对比与分析

本节主要介绍对比实验所使用的数据集、推荐评价指标、实验设置以及对比实验结果。对比实验采用的对比算法包括传统的基于协同过滤、排序的方法以及最新的基于神经网络的推荐方法。这些方法共同针对面向用户隐性反馈的推荐问题。

##### 5.1 数据集

本文在 Alibaba<sup>1)</sup> 和 Retailrocket<sup>2)</sup> 数据集上进行了对比实验。

<sup>1)</sup> <https://tianchi.aliyun.com/>

<sup>2)</sup> <https://www.kaggle.com/>

Alibaba 数据集:随机抽取 1 万个淘宝网用户的行为记录,包括点击、收藏、加入购物车和购买 4 种行为,以及行为的时间戳。

Retailrocket 数据集:电子商务网站的用户行为数据,包括 4 个半月内网站访问者的点击、加入购物车、交易行为数据。数据集的统计信息如表 1 所列。

表 1 实验数据的统计概况

Table 1 Statistical overview of experimental dataset

项目类别	统计信息	
	Alibaba	Retailrocket
用户数	10000	1407580
点击/浏览	11550581	2664312
收藏/加入购物车	586120	69332
购买	120205	22457
用户行为数总计	12256906	2756101

## 5.2 实验评价指标

我们采用召回率 (recall)、准确率 (precision) 和 NDCG (Normalized Discounted Cumulative Gain) 3 个指标对推荐性能的进行评价。

召回率:在所有待推荐的商品中被成功推荐的比率。定义如下:

$$recall@k = \frac{|test \cap Top_k|}{|test|} \quad (13)$$

其中,  $Top_k$  是算法给出的长度为  $k$  的推荐列表,  $test$  为测试集中用户有正反馈的项目列表。  $|test|$  代表测试集的规模,  $|test \cap Top_k|$  代表测试集与推荐集有相同项目的集合规模。

准确率:预测准确的商品总数与推荐的所有商品总数的比率。定义如下:

$$precision@k = \frac{|test \cap Top_k|}{|Top_k|} \quad (14)$$

NDCG 排序评价指标:一种排序评价指标,定义如下:

$$NDCG(Q, k) = \frac{1}{|Q|} \sum_{u \in Q} \frac{1}{Z_u} \sum_{p=1}^k \frac{2^{R(u,p)} - 1}{\log(1+p)} \quad (15)$$

其中,  $Q$  为数据集中的用户集合;  $R(u, p)$  为用户  $u$  的排序列表中第  $p$  位的项目关联等级,在推荐系统的评价中,可以用 0/1 处理,即如果  $p$  位置的项目在用户  $u$  的测试集合中,则  $R(u, p) = 1$ ;  $Z_u$  是归一化因子,可使得计算出的 NDCG 值在 0 到 1 之间,且最优的排序 NDCG 值为 1。

利用召回率、准确率和 NDCG 3 个指标来评价推荐系统的性能时,其值越大,代表推荐系统的推荐性能越好。在推荐性能的评价中,我们设置推荐 Top 列表的长度  $k$  分别为 5, 10 和 20,在所有的测试集上测试召回率、准确率、NDCG 3 个指标,并取其平均值作为整体评价指标的最终结果。

## 5.3 实验设置

在实验中,我们把用户有交互的项目作为训练正样本,把用户无交互的项目作为训练负样本。由于数据集中正负样本的比例差距很大,因此我们按比例 1:4 对正负样本进行采样。同时考虑用户购买行为比例的稀疏性,对用户购买行为进行三倍重采样。

在用户向量化模型中,我们设置项目的 embedding 向量为 200 维。考虑到两个数据集的用户数量规模差距较大,我们将用户向量化维度设置为不同的值来考查其对推荐性能的

影响。多层感知机采用常见的塔式结构,即每层的维度缩减为上一层的一半。在模型的训练中,采用高斯分布随机初始化模型参数(均值为 0,标准差为 0.01),通过随机梯度下降调整模型的参数。设置训练次数为 128,学习率为 0.001。

## 5.4 实验结果

将本文的 UserVec-Att 推荐模型和以下 4 种推荐算法进行对比。

1) Item-knn<sup>[2]</sup>:基于项目的标准协同过滤方法。

2) MF-BPR<sup>[5]</sup>:基于贝叶斯个性化排序的推荐算法,该模型通过隐式反馈构建对排序损失。

3) NeuCF<sup>[8]</sup>:一种矩阵分解方法的泛化,基于多层感知机的推荐模型,利用用户和项目特征向量模拟用户与项目的交互行为,用二分类的方法建模隐式反馈行为。

4) GRU-rec<sup>[9]</sup>:基于循环神经网络的推荐模型,建模用户的行为序列,从而得到推荐结果。

各算法在召回率、准确率、NDCG 3 个指标上的对比结果如图 3—图 5 所示。

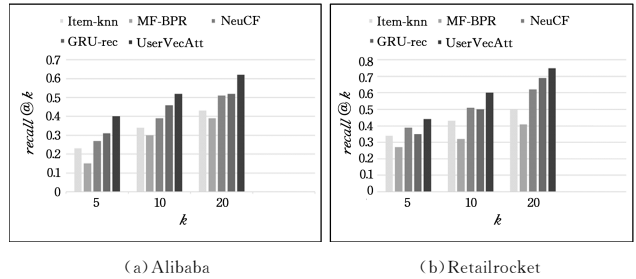


图 3 Alibaba 数据集和 Retailrocket 数据集上不同算法的召回率  
Fig. 3 Recall of different algorithms on Alibaba dataset and Retailrocket dataset

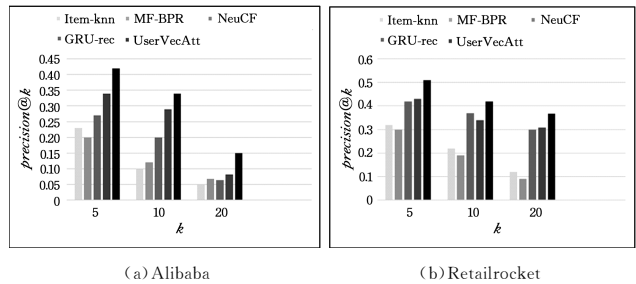


图 4 Alibaba 数据集和 Retailrocket 数据集上不同算法的准确率  
Fig. 4 Precision of different algorithms on Alibaba dataset and Retailrocket dataset

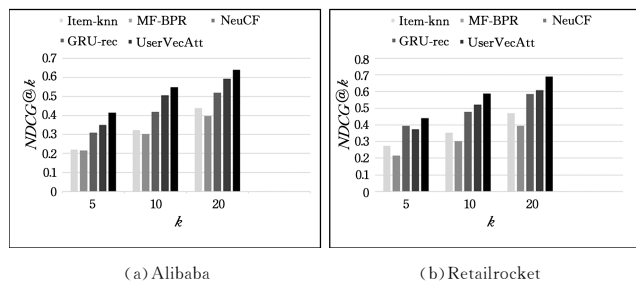


图 5 Alibaba 数据集和 Retailrocket 数据集上不同算法的 NDCG  
Fig. 5 NDCG of different algorithms on Alibaba dataset and Retailrocket dataset

从图 3—图 5 可以看出,随着推荐 Top 列表长度  $k$  的增加,5 种推荐算法的召回率和 NDCG 值增大,准确率减小。本文提出的推荐模型在召回率、准确率、NDCG 3 个指标上均高于对比算法,证明了本文方法具有更好的推荐性能。神经网络推荐模型(NeuCF,GRU-rec 和 UserVec-Att)在隐式反馈推荐问题上,普遍比传统协同过滤与排序方法(Item-knn 与 MF-BPR)有更好的性能。由此可知,神经网络非常适合挖掘用户与项目之间潜在的交互行为以及项目和用户的潜在特征。

考虑到两个数据集在用户数量上的规模差距,在推荐列表长度  $k=20$  的情况下,我们通过设置不同的用户向量化维度大小,对比了本文提出的推荐模型在两组数据集上 NDCG 值的变化情况,对比结果如图 6 所示。

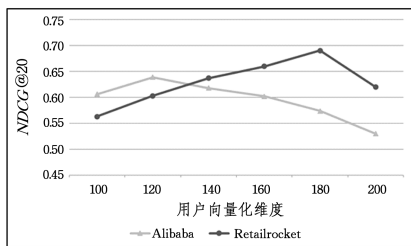


图 6 UserVec-Att 模型在不同的向量维度下 NDCG 的对比

Fig. 6 NDCG of UserVec-Att model under different vector dimensions

由图 6 可知,对于用户数量规模较小的 Alibaba 数据集,用户向量化维度为 120 时,NDCG@20 取得最大值;对于用户数量规模较大的 Retailrocket 数据集,用户向量维度为 180 时,NDCG@20 取得最大值。对于本实验,较大的用户向量维度可能导致模型过拟合,使得推荐模型的效果变差。因此,对于不同的数据规模,选取不同的向量化维度十分重要。

**结束语** 本文基于神经网络设计了一种用户向量化表示模型,通过充分利用用户历史行为中多种异构的隐式反馈数据训练模型,学习得到用户的向量化表示。同时,考虑到用户向量化表示的模型仅仅利用了用户静态的历史行为记录,不能在用户行为的时序层面捕捉用户与项目的交互特征,本文融合用户向量化表示和 self-attention 机制设计了一种神经网络推荐模型。在两个公开数据集上的相关对比实验结果显示了本文方法在召回率、准确率、NDCG 3 个指标上均高于几种对比算法,证明其拥有更好的推荐性能。

未来,我们将研究在用户向量化表示中融入用户的画像属性,以提高推荐模型的性能;同时考虑将用户评论、项目标签等特征作为项目的辅助信息,融合神经网络,增强模型对用户和项目的建模能力。

## 参 考 文 献

[1] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is

all you need[M]// Advances in Neural Information Processing Systems. Bertin: Springer, 2017: 5998-6008.

- [2] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms [C] // Proceedings of the 10th International Conference on World Wide Web. ACM, 2001: 285-295.
- [3] OSTUNI V C, DI NOIA T, DI SCIASCIO E, et al. Top-n recommendations from implicit feedback leveraging linked open data [C] // Proceedings of the 7th ACM Conference on Recommender Systems. ACM, 2013: 85-92.
- [4] ZIMDARS A, CHICKERING D M, MEEK C. Using temporal data for making recommendations [C] // Proceedings of the Seventeenth Conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., 2001: 580-588.
- [5] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian personalized ranking from implicit feedback [C] // Proceedings of the Twenty-fifth Conference on Uncertainty in Artificial Intelligence. AUAI Press, 2009: 452-461.
- [6] SEDHAIN S, MENON A K, SANNER S, et al. AutoRec: Autoencoders meet collaborative filtering [C] // Proceedings of the 24th International Conference on World Wide Web. ACM, 2015: 111-112.
- [7] 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. International World Wide Web Conferences Steering Committee, 2015: 278-288.
- [8] HE X, LIAO L, ZHANG H, et al. Neural collaborative filtering [C] // Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2017: 173-182.
- [9] HIDASI B, KARATZOGLOU A, BALTRUNAS L, et al. Session-based recommendations with recurrent neural networks [J]. arXiv:1511.06939, 2015.
- [10] GONG Y, ZHANG Q. Hashtag Recommendation Using Attention-Based Convolutional Neural Network [C] // Proceedings of International Joint Conference on Artificial Intelligence. AAAI Press, 2016: 2782-2788.
- [11] LI Y, LIU T, JIANG J, et al. Hashtag recommendation with topical attention-based LSTM [C] // Proceedings of the 26th International Conference on Computational Linguistics. ACL, 2016: 3019-3029.
- [12] XIAO J, YE H, HE X, et al. Attentional factorization machines: Learning the weight of feature interactions via attention networks [J]. arXiv:1708.04617, 2017.