

# 基于增强注意力机制的神经协同过滤



康 雁 卜荣景 李 浩 杨 兵 张亚钊 陈 铁

云南大学软件学院 昆明 650091

**摘 要** 推荐系统是解决信息过载问题的核心。现有的推荐框架研究面临着显式反馈数据稀疏和数据预处理难等问题,特别是对新用户和新项目进行推荐的性能有待进一步提高。随着深度学习的推进,基于深度学习的推荐成为了当前的研究热点,大量的实验证明了深度学习运用于推荐系统的有效性。文中在 NCF 的基础上提出了 EANCF (Neural Collaborative Filtering based on Enhanced-Attention Mechanism),从隐式反馈数据的角度研究了推荐框架,利用最大池化、局部推理以及组合多种不同数据融合方式来考虑数据特征提取;同时,引入注意力机制来为网络合理地分配权重值,减少信息的损失,提升推荐的性能。最后,基于两个大型真实数据集 Movielens-1m 和 Pinterest-20 对 EANCF、NCF 和部分经典算法做了对比实验,并且详细地给出了 EANCF 框架的训练过程。实验结果表明,EANCF 框架确实具有较好的推荐性能,相比于 NCF 框架在 HR@10 和 NDCG@10 上均有显著提升,HR@10 最高提升了 3.53%,NDCG@10 最高提升了 2.47%。

**关键词:**深度学习;协同过滤;隐式反馈;注意力机制

**中图法分类号** TP391

## Neural Collaborative Filtering Based on Enhanced-attention Mechanism

KANG Yan, BU Rong-jing, LI Hao, YANG Bing, ZHANG Ya-chuan and CHEN Tie

School of Software, Yunnan University, Kunming 650091, China

**Abstract** The recommendation system is the core to solve the problem of information overload. The existing research on recommendation framework faces many problems, such as sparse explicit feedback data and difficulty to preprocess data, especially the recommendation performances for new users and new projects need to be further improved. With the advancement of deep learning, recommendation based on deep learning has become a current research hotspot. A large number of experiments have proved the effectiveness of deep learning applied to recommendation system. This paper presents EANCF (Neural Collaborative Filtering based on Enhanced-attention Mechanism) on the basis of NCF. It studies the recommendation framework from the perspective of implicit feedback data, and considers the data feature extraction by means of max-pooling, local inference modeling and combining many different ways of data fusion. Meanwhile, attention mechanism is introduced to reasonably allocate weight value for the network, reduce the loss of information and improve the performance of recommendation. Finally, based on two large real data sets, Movielens-1m and Pinterest-20, comparative experiments are carried out between EANCF and NCF, as well as some classical algorithms, and the training process of EANCF framework is given in detail. The experimental results show that the proposed EANCF framework does have good recommendation performance. Compared with the NCF, both HR@10 and NDCG@10 are significantly improved, with the highest increase of 3.53% for HR@10 and 2.47% for NDCG@10.

**Keywords** Deep learning, Collaborative filtering, Implicit feedback, Attention mechanism

## 1 引言

信息过载 (Information Overload) 是大数据时代需要面对的一个严重问题。如何从这些爆炸式增长的信息里找到每个互联网用户感兴趣的信息,并把这些信息推送给用户,是推荐系统研究的主要内容。推荐系统是互联网时代与大型复杂信息空间交互的一种信息检索工具,通过提供空间的个性化视

图来优先考虑用户可能感兴趣的项目。从 20 世纪 90 年代发展至今,推荐系统已经成为一门独立的学科,在学术界和工业界应用中都取得了很多成果,被广泛应用于电子商务、电影视频、音乐电台和个性化广告等领域。

目前,应用深度学习的推荐系统是最主要的研究方向。传统的深度学习应用仍采用矩阵分解的方式对用户和项目特征之间的交互进行建模,并使用内积结合用户和项目的潜在

到稿日期:2019-09-04 返修日期:2020-01-05 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61762092,61762089);云南省软件工程重点实验室开放基金项目(2017SE204)

This work was supported by the National Natural Science Foundation of China (61762092,61762089) and Yunnan Provincial Key Laboratory of Software Engineering Open Fund Project (2017SE204).

通信作者:康雁(kangyan@ynu.edu.cn)

特征。2017年,He等<sup>[1]</sup>提出的NCF框架强调隐式数据二进制属性的概率,通过形式化用户协同过滤的神经网络建模方法来优化传统的应用方式,用神经结构代替内积从数据中自动学习任意函数,吸引了众多的研究者进行改进。He原团队<sup>[2]</sup>指出了特征向量内积、拼接操作的局限性,在嵌入层上使用外积,得到了一个更有表现力和语义上更合理的二维映射。Bai等<sup>[3]</sup>首次将邻域信息集成到NCF方法中,该模型对隐式推荐任务有极好的性能,但存在模型泛化能力弱、领域信息构建繁琐的弊端。Wang等<sup>[4]</sup>通过利用用户项目交互的高阶连通性,在交互图形结构中对协作信号进行编码。虽然直接使用交互功能可以弥补次优嵌入的不足,但是在实际开发中,交互的规模可以轻易达到数百万甚至更大,使得提取期望的协作信号困难。Jiang<sup>[5]</sup>将图卷积网络引入NCF,把评分信息与多种辅助信息进行融合,并对未观测到的评分矩阵元素进行预测与填充,但导致了复杂度的急剧上升。Xie等<sup>[6]</sup>证明了结合深度神经网络和注意力机制可以有效地提取隐式反馈数据中的潜在特征。但这些研究大多没有考虑到隐式数据特征的提取方式和注意力机制的权重分配问题。

本文提出了一种基于增强注意力机制的神经协同过滤推荐框架EANCF。相比于以往的研究,EANCF从隐式反馈的角度利用最大池化操作,通过组合多种不同数据融合方式获取更多的特征值,使用多层神经网络对用户和数据之间的交互建模,突破了矩阵分解在评分预测任务中的局限性,对矩阵分解的可表达性进行了优化,能够挖掘出用户和项目的深层特征<sup>[7]</sup>;使用并扩展文本分类算法ESIM(Enhanced LSTM for Natural Language Inference)中的局部推理,捕获了差异性<sup>[8]</sup>;同时,引入注意力机制来分配权重值,这有助于有效地提取隐式反馈数据中的潜在特征,提升推荐结果的合理性。

在大型真实数据集MovieLens-1m和Pinterest-20上的实验表明,EANCF模型不仅具有更高的泛化能力,而且推荐结果的正确率更高。

本文的主要贡献有:

1)提出了一种新的神经网络框架EANCF,其利用多种数据融合方式将NCF模型与Attention机制相结合进行建模;

2)通过对用户和物品数据进行最大池化操作、局部推理以及多种数据融合方式的组合操作,获得了更高的相关性特征;

3)实验利用MovieLens-1m和Pinterest-20数据集比较了EANCF框架与NCF框架的3种实现方式的性能,证明了EANCF框架的有效性。

## 2 相关工作

推荐系统算法是当前一个热门的研究领域,不仅能够为用户带来便利,还能够为企业带来巨大的利益。本节将对隐式反馈、协同过滤、注意力推荐3个相关领域进行简单介绍。

### 2.1 隐式反馈

近年来,协同过滤算法在学术界和工业界得到了广泛的研究。协同过滤算法处理的数据按照行为反馈的明确性分为显式反馈数据(如评分等)和隐式反馈数据(如浏览等)<sup>[9]</sup>。随着推荐算法研究的深入,研究者发现很多应用场景缺乏显式反馈数据,因此根据大量的隐式反馈数据来推断用户的偏好

值成为了研究的重点。Yao等<sup>[10]</sup>利用隐式反馈数据构造多层上下文图,然后在多层上下文图中执行基于上下文感知的推荐算法。虽然该方法在真实数据集上被证明是有效的,但是它没有考虑到用户目标的多变性,因此无法获取最实时的偏好值。针对此缺陷,He等<sup>[11]</sup>在使用深度递归神经网络的同时,通过捕捉隐式反馈数据对用户行为进行动态有序的建模。该方法既保留了协同过滤算法带来的准确率,又更好地追踪了用户的实时动态偏好。

### 2.2 协同过滤

1992年,Goldberg等<sup>[12]</sup>首次提出协同过滤这个概念,并提出了第一个协同过滤推荐系统Tapestry。1994年,明尼苏达大学的Resnick等<sup>[13]</sup>研究出第一个自动化推荐系统GroupLens,提出将协同过滤作为推荐系统的重要技术。经过十多年的发展,第一届ACM推荐系统大会于2007年在美国举行。此后,关于推荐系统的研究队伍不断壮大,推荐系统逐渐与更多的研究领域进行融合。例如,京东的Zhao等<sup>[14]</sup>引入强化学习;Pasricha等<sup>[15]</sup>提出了一种序列推荐的算法;Sun等<sup>[16]</sup>将知识图谱作为一种辅助数据源引入。尽管新的推荐算法层出不穷,但协同过滤作为推荐领域应用最为广泛的算法,仍然持续获得了业界众多学者的创新研究。它一般基于矩阵分解,不需要预先获得用户或物品的特征数据,仅依赖于用户的历史数据对目标用户进行建模,从而获得更准确的推荐结果。协同过滤算法尽管非常强大,但不能很好地利用辅助信息,从而导致了冷启动问题<sup>[17]</sup>。为了解决这一问题并提高推荐性能,Zhang等<sup>[18]</sup>提出了一种基于上下文的半监督学习框架TSEL,其在对SVD进行扩充的同时利用Tri-training框架来解决冷启动问题。近年来,学术界更是掀起了将深度学习运用到协同过滤中的热潮。Deng等<sup>[19]</sup>提出了融合动态协同过滤和深度学习的推荐算法,并证明了引入深度学习的方法能够有效考虑动态时间因素并获取任意连续的特征,其性能优于先前基于矩阵分解的模型。Zheng等<sup>[20]</sup>引入卷积神经网络提取用户和物品的特征,然后利用矩阵分解技术对所提取的特征进行预测。但是,这些研究大多只是利用深度学习为一些辅助信息建模。为了更深入地将深度学习应用于推荐系统,He等<sup>[1]</sup>用多层神经网络代替内积给用户和物品进行交互建模,并提出了一种基于神经网络的协同过滤通用框架NCF。

### 2.3 注意力推荐

注意力机制已证明在多元机器学习任务中是有效的,例如图像/视频字幕<sup>[21]</sup>和机器翻译<sup>[22]</sup>。它合理的假设使得人们在需要的时间和地点下能够选择性地关注某一部分感知空间,而不需要立即处理整个信号。注意力机制包含软空间和软时间两种注意力模型。EANCF框架采用软注意力,在总和归一化为1的前提下,学习为一组特征分配注意力权重,较高(较低)的权重表示相应特征对于结束任务是多信息性(低信息性)的。近年来,注意力推荐多次被证明是合理的。Li等<sup>[23]</sup>提出了一种新的神经注意力推荐机制(NARM),其在当前会话中捕获用户的主要目的并将其编码为统一的会话表示形式,最后通过共同学习项目和会话表示以及它们的匹配来训练NARM。测试结果表明,NARM在召回率和MRR(Mean Reciprocal Rank)方面都优于最先进的基线;Guo

等<sup>[24]</sup>在充分利用用户异构的隐式反馈行为的同时提出了一种神经注意力推荐模型,其通过融合用户向量化表示和用户-项目交互的动态时序特征,来提高推荐的性能。

### 3 EANCF 框架

已有的 NCF 框架将某一层的输出当作下一层的输入,由输入层、嵌入层、NCF 层、输出层四大部分组成。其中,嵌入层是一个全连接层,可以将输入层的稀疏表示映射为一个稠密向量,用来描述用户(项目)的潜在向量;NCF 层的每一层都可以通过被定制来发现用户-项目交互的某些潜在结构,最终将潜在向量映射为预测分数传到输出层,这比固定内部函数具有更强大的表达能力。然而,该方法只考虑到了用户的当前顺序行为,没有强调用户行为的主要目的。EANCF 通

过引入注意力机制自动调整权重解决了这个问题。此外,NCF 框架输入层输出的特征向量容易造成过拟合,且用户特征向量和项目特征向量的组合方式过于单一。为了解决过拟合问题,我们分别对用户和项目的特征向量进行一次最大池化操作,在保留主要特征的同时减少参数和计算量,从而提高模型的泛化能力;然后,对这两个特征向量进行加法、乘法、平均值等多种运算组合,得到一个 Attention 融合,并进入 Layer 层进行模型学习;最后,经过一个 Attention 融合得到预测值。

#### 3.1 总体框架

本文所提出的 EANCF 框架主要由输入层、嵌入层、EANCF 层和输出层 4 部分组成,上一层的输出将会作为下一层的输入。EANCF 框架的结构视图如图 1 所示。

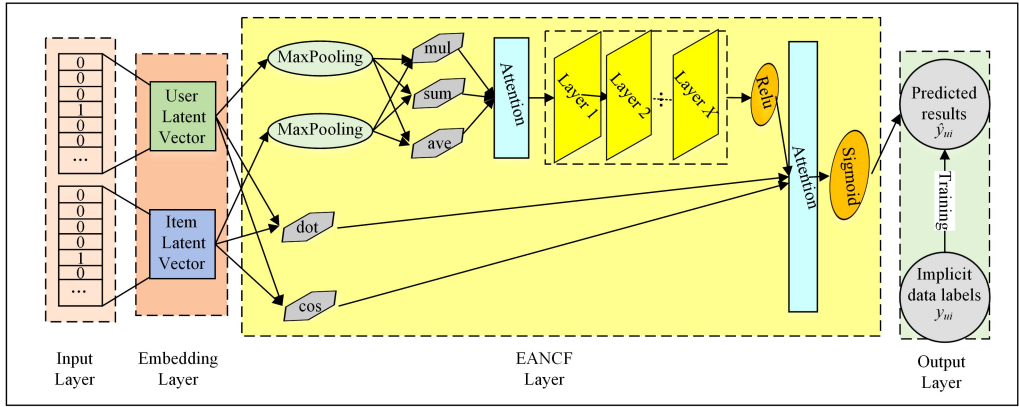


图 1 EANCF 框架的结构视图

Fig. 1 Structure view of EANCF framework

#### 3.2 输入层

实验采用的是电影数据集 Movielens-1m 和 Pinterest-20,主要描述了用户、项目、用户和项目之间的交互以及数据的稀疏度,每个数据集包括上百万条交互信息,提升了学习模型的准确性,而高达 95% 的稀疏度也需要我们对其预先进行过滤处理。在输入层中,首先对用户和项目的 ID、类别等特征值进行独热编码处理,独热编码能够处理非数值属性,所以在处理所选择的电影数据集时的效果要优于神经网络类编码的效果。同时,独热编码在一定程度上扩充了数据特征,相当于对每一维的特征进行了一次归一化处理。编码后,将处理过的用户、项目向量分别作为框架的输入。

#### 3.3 嵌入层

输入层的下一步是一个全连接的嵌入层,它将输入层的数据映射成描述用户和项目在潜在因素模型上下文中的潜在向量。假设  $v_u^i$  和  $v_i^j$  分别为用户  $u$  和项目  $i$  的特征向量,它们的嵌入向量分别为  $f_u$  和  $h_i$ ,则嵌入方法如式(1)所示:

$$f_u = F^T v_u^i; h_i = H^T v_i^j \quad (1)$$

其中,  $F \in R^{M \times K}$  为用户特征的嵌入矩阵;  $H \in R^{N \times K}$  为物品特征的嵌入矩阵;  $K$  为嵌入大小;  $M$  为用户特征个数;  $N$  为项目特征个数。

#### 3.4 EANCF 层

EANCF 层是整个框架的核心部分。该部分首先将嵌入层输出的用户和项目向量作为输入,分别进行一个最大池化操作,得到的输出再两两分别通过相乘、求和、求平均的提取

方式得到一个增强的 Attention 融合。按照融合方式的不同,我们将分别讨论 EAGMF 和 EAMLP 两种模型的不同融合方式。这里我们借鉴了 Chen 等<sup>[8]</sup>提出的 ESIM 模型中的局部推理部分来实现 EANCF 对特征数据的融合。ESIM 主要用于文本的推理和相似度匹配任务,与其他使用复杂网络体系结构的模型不同,它使用了 Tree-LSTM(Tree-structured Long Short-Term Memory networks)、BiLSTM(Bi-directional Long Short-Term Memory)、局部推理建模、推理合成相结合的方式,展现出强大的性能,在斯坦福大学自然语言推理数据集上达到了 88.6% 的准确率。

EAGMF 按照式(2)所示融合方式输出,3 个权重运算组件通过相乘进行融合。

$$P_0^{GMF} = W_1 \circ P_{mul} \otimes W_2 \circ P_{sum} \otimes W_3 \circ P_{avg} \quad (2)$$

其中,  $\circ$  是指元素间相乘;  $P_{mul}$ ,  $P_{sum}$  与  $P_{avg}$  如式(3)一式(5)所示,引入 ESIM 模型的局部推理部分提取每个用户向量和所有项目向量相乘、求和或求平均运算时的推断信息,帮助捕获差异性;  $W_1, W_2$  与  $W_3$  分别表示 3 个运算组件的权重,其将在模型训练过程中通过学习不断更新;  $\otimes$  指采用相乘的方式来提取特征。

$$P_{mul} = \sum_{k=1}^n \sum_{l=1}^m [u_k; i_l; u_k \otimes i_l] \quad (3)$$

$$P_{sum} = \sum_{k=1}^n \sum_{l=1}^m [u_k; i_l; u_k \oplus i_l] \quad (4)$$

$$P_{avg} = \sum_{k=1}^n \sum_{l=1}^m [u_k; i_l; u_k \odot i_l] \quad (5)$$

其中,  $\oplus$  是指采用求和的方式来提取特征;  $\odot$  是指采用求

平均的方式来提取特征。

EAMLP 按照式(6)所示融合方式输出  $P_0^{\text{MLP}}$ , 3 个权重运算组件通过拼接进行融合。

$$P_0^{\text{MLP}} = W_1 \circ P_{\text{mul}} \odot W_2 \circ P_{\text{sum}} \odot W_3 \circ P_{\text{avg}} \quad (6)$$

其中,  $\odot$  是指采用拼接的方式来提取特征。

相比于传统的直接做点积, 这样的融合方法在建模时为不同的用户偏好关系赋予不同的特征权重, 以便有效地捕获长期或短期的兴趣, 损失的关键特征更少。融合完成之后, 将结果  $P_0^{\text{GMF}}$  或  $P_0^{\text{MLP}}$  放入模型学习的全连接层, 其层数具有可控性, 相邻两层之间的计算如式(7)所示:

$$F(P_{x-1}) = P_x \quad (7)$$

其中,  $F$  表示全连接层的计算函数。另外, 我们对输入的用户和项目向量分别进行一个点积和余弦运算以抽取特征向量, 接着利用式(8)将 5 个运算组件学习到的特征进行一个增强的 Attention 融合  $P_{\text{out}}$ , 再通过 sigmoid 激活函数得到最终的预测值  $\hat{y}_{ui}$ , 如式(9)所示:

$$P_{\text{out}} = W_1 \circ P_{\text{dot}} \odot W_2 \circ P_{\text{cos}} \odot W_3 \circ \text{Relu}(P_x) \quad (8)$$

$$\hat{y}_{ui} = \text{sigmoid}(P_{\text{out}}) \quad (9)$$

其中,  $P_{\text{dot}}$  与  $P_{\text{cos}}$  如式(10)和式(11)所示, 引入 ESIM 模型的局部推理部分提取每个用户向量和所有项目向量求点积、求余弦运算时的推断信息, 帮助捕获差异性;  $W_1, W_2$  与  $W_3$  分别表示组件的权重, 并在模型训练过程中通过学习不断更新。  $P_{\text{out}}$  融合对特征数据进行了更深层次的学习, 能捕捉到更多的有用信息, 用  $P_{\text{out}}^{\text{GMF}}$  表示 EAGMF 的融合结果,  $P_{\text{out}}^{\text{MLP}}$  表示 EAMLP 的融合结果。

$$P_{\text{dot}} = \sum_{k=1}^n \sum_{l=1}^m [u_k; i_l; u_k \odot i_l] \quad (10)$$

$$P_{\text{cos}} = \sum_{k=1}^n \sum_{l=1}^m [u_k; i_l; u_k \circ i_l] \quad (11)$$

其中,  $\odot$  指采用点积的方式来提取特征;  $\circ$  指采用余弦的方式来提取特征。

此外, 我们将 EAGMF 和 EAMLP 两种模型进行了组合尝试, 输出  $P'_{\text{out}}$ , 如式(12)所示:

$$P'_{\text{out}} = P_{\text{out}}^{\text{GMF}} + P_{\text{out}}^{\text{MLP}} \quad (12)$$

然而,  $P'_{\text{out}}$  的结果并没有因为再次引入一个 Attention 而提升, 这说明我们不能对特征进行持续多次的融合, 以避免更多重要信息的丢失。

### 3.5 输出层

EANCF 框架的最终目标是让模型学习到更多的数据信息并给出一个预测值  $\hat{y}_{ui}$ 。为了更好地处理隐式反馈数据的 Top-N 推荐, 使最终输出层的预测值和隐式数据标签  $y_{ui}$  之间的损失最小, 从而达到推荐结果最优的目的, 我们选择了逐点对数损失 (point-wise log loss) 下的二值交叉熵损失函数来学习模型的参数  $\theta$ , 函数如式(13)所示。在该方法下, 我们通过 sigmoid 函数转换将预测值控制在 (0, 1) 范围内, 从而更加灵活地控制采样率。

$$L(\theta) = \sum_{(u,i) \in S^+} w_{ui} (y_{ui} \log \hat{y}_{ui} + (1 - y_{ui}) \log (1 - \hat{y}_{ui})) \quad (13)$$

其中,  $S^+$  是负样本集合;  $w_{ui}$  是训练集中样本  $\langle u, i \rangle$  的权重, 用来平衡正负样本的数量。

### 3.6 EANCF 算法

算法 1 概述了 EANCF 框架的训练过程。首先基于数据

集  $R$  构造训练实例, 然后通过反向传播和 Adam 训练框架, 最终得到最优的模型  $Y$ 。

#### 算法 1 Training EANCF

Input: Historical observations  $R = \{x_{ui}^0, x_{ui}^1, \dots, x_{ui}^{n-1}\}$ , number of layers  $L$ , length of embedding  $E$ , attention selection: {concat  $\odot$ , mul  $\otimes$ , average  $\ominus$ }

Output: Predicted model  $Y$

Begin:

1. /\* construct training instances \*/

2.  $D \leftarrow \emptyset$

3. for all historical observation  $t \leftarrow 1$  to  $(n-1)$  in  $R$  do

4.  $\mathbf{f}_u = \mathbf{F}^T \mathbf{v}_u^U$

5.  $\mathbf{h}_i = \mathbf{H}^T \mathbf{v}_i^I$

6. put a pretreated training instance  $\{\mathbf{f}_u, \mathbf{h}_i\}$  into  $D$

7. end for

8. /\* train the model \*/

9. initialize all learnable parameters in EANCF

10. repeat

11. the first Attention fusion:

$$P_0^{\text{GMF}} = W_1 \circ P_{\text{mul}} \otimes W_2 \circ P_{\text{sum}} \otimes W_3 \circ P_{\text{avg}} \text{ for EAGMF;}$$

$$P_0^{\text{MLP}} = W_1 \circ P_{\text{mul}} \odot W_2 \circ P_{\text{sum}} \odot W_3 \circ P_{\text{avg}} \text{ for EAMLP}$$

12. the fusion result enters Layers

13. the second Attention fusion:

$$P_{\text{out}} = W_1 \circ P_{\text{dot}} \odot W_2 \circ P_{\text{cos}} \odot W_3 \circ \text{Relu}(P_x)$$

14. randomly select a batch of instances  $D_b$  from  $D$

15. find  $\theta$  by minimizing

$$L(\theta) = \sum_{(u,i) \in S^+} w_{ui} (y_{ui} \log \hat{y}_{ui} + (1 - y_{ui}) \log (1 - \hat{y}_{ui})) \text{ with } D_b$$

16. end repeat until stopping criteria is met.

End

## 4 实验

### 4.1 数据集

实验采用两个真实的大型可公开访问的数据集 MovieLens-1m 和 Pinterest-20。MovieLens-1m 是一个被广泛用于评估协同过滤算法的电影评级数据集, 它包含 6 000 多个用户, 每个用户至少有 20 个评分。在实验之前, 我们将 MovieLens-1m 转换为隐式数据, 并用  $y_{ui}$  标记, 把评分数据中评分项为 1~5 分的用户-项目对标记为 1, 而未被评分的用户-项目对标记为 0, 从而挖掘 EANCF 从隐式反馈中学习隐式信号的表现。Pinterest-20 是一个用于评估基于内容的图像推荐的隐式反馈数据, 用户通过将图片“pin”在自己的图板上来显示自己对这些图片的偏好。我们使用 leave-one-out 机制来划分测试集和训练集, 测试集为用户最近的一次交互信息, 其余作为训练集。在测试时, 随机抽取 100 个与用户没有进行过交互的项目, 将测试集排列在这 100 个项目。由于这两个数据集大而稀疏, 我们需要提前对其进行过滤。这两个数据集的特征统计如表 1 所列。

表 1 实验评估数据集

Table 1 Experimental evaluation data sets

Dataset	Interaction	Item	User	Sparsity/%
MovieLens-1m	1 000 209	3 706	6 040	95.53
Pinterest-20	1 500 809	9 916	55 187	99.73

## 4.2 实验环境

在深度学习模型中,模型参数的初始化对模型的收敛性和最终表现起到极其重要的作用。为了确定 EANCF 方法的参数,我们为每个用户随机抽取一个交互作为验证数据并调整其超参数,通过优化函数和对数损失来学习 EANCF 模型。训练过程中载入的预训练 EAGMF 和 EAMLP 随机初始化的参数选择满足平均值为 0、标准差为 0.01 的高斯分布,采用自适应估计 Adam 直至模型收敛。使用预先训练的参数对模型初始化后,设置学习率为 0.001,批次大小为 256,使用随机梯度下降(Stochastic Gradient Descent, SGD)进行优化,训练时为每个积极实例进行 4 个消极实例的采样,测试时一个积极实例搭配多个消极实例。考虑到实验主要用于验证所提模型的优化性,且 He 等提出的 NCF<sup>[1]</sup>原模型实验中嵌入层的长度对实验结果的影响并不大,本文不对嵌入层长度变化做深入的实验对比,将其设置为 64。

实验主要是在 GPU 服务器上运行,具体实验环境如表 2 所列。

表 2 实验环境  
Table 2 Experimental environments

Name	Version
OS	Windows Server 2012 R2
Memory	256 GB
CPU	Inter(R) Xeon(R)CPU E5-2680 v2 @ 2.80 GHz
GPU	Tesla K40 m
Number of GPU cards	4
CUDA	8.0
cuDNN	8.0
Tensorflow	1.8.0

## 4.3 评价指标

本文使用式(14)所示的 HR(Hit Ratio)指标和式(15)所示的 NDCG(Normalized Discounted Cumulative Gain)指标来评估 EANCF 模型的性能。HR 衡量的是推荐结果的召回率,即推荐的前  $N$  个项目中有多少能够命中用户实际偏好,该指标值越大越好。

$$HR@K = \frac{\text{Number of Hits}@K}{|T|} \quad (14)$$

其中,分母表示所有的测试集合;分子是每个用户前  $K$  个推荐结果中属于测试集合个数的总和,本文将  $K$  指定为 10。

NDCG 常用作排序的评价指标,具有最好的辨别力。当模型输出推荐结果的 Top- $N$  时,便可以通过 NDCG 来预测这个序列的准确度,该指标值越大越好。

$$NDCG@K = Z_k \sum_{p=1}^K \frac{2^{r_p-1}}{p \log_2(p+1)} \quad (15)$$

其中,  $Z_k$  是归一化系数,表示最好情况下整个  $\sum_{p=1}^K \frac{2^{r_p-1}}{p \log_2(p+1)}$  的倒数;  $r_p$  表示在第  $p$  个位置时的等级关联性,等级关联性规则如式(16)所示:

$$r_p = \begin{cases} 0, & i_p \notin T \\ 1, & i_p \in T \end{cases} \quad (16)$$

其中,  $i$  表示项目;  $i_p$  表示第  $p$  个位置上的项目;  $T$  表示实验测试集。

此外,为了挖掘 EANCF 在隐式反馈数据中学习隐式信

号的表现,我们定义用户对项目的偏好值  $P_{ui}$ 。假设  $u$  和  $i$  分别表示用户和项目,  $n_{ui}$  表示用户  $u$  对项目  $i$  的隐式反馈动作频数,则  $P_{ui}$  如式(17)所示。

$$P_{ui} = \begin{cases} 1, & n_{ui} > 0 \\ 0, & n_{ui} = 0 \end{cases} \quad (17)$$

其中,  $n_{ui} > 0$  不代表用户喜欢该物品,可能是用户无意的浏览或点击;  $n_{ui} = 0$  也不意味着用户不喜欢该项目,可能是用户还没有发现该物品。

## 4.4 结果分析

为了证明 EANCF 模型的有效性,利用 Tensorflow 对 Movielens-1m 和 Pinterest-20 进行仿真实验。HR@10 和 NDCG@10 相对于预测因素数量在 Movielens-1m 上的性能如图 2 所示,HR@10 和 NDCG@10 相对于预测因素数量在 Pinterest-20 上的性能如图 3 所示。

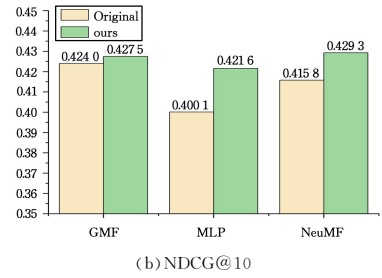
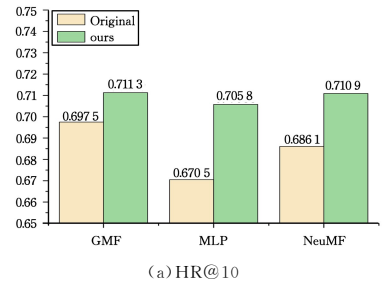


图 2 Movielens-1m 上 Top-10 性能的比较

Fig. 2 Comparison of Top-10 performance on Movielens-1m

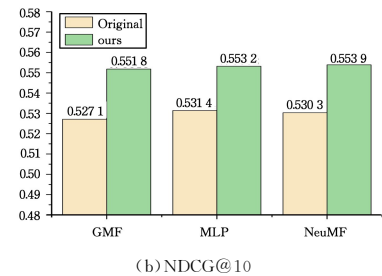
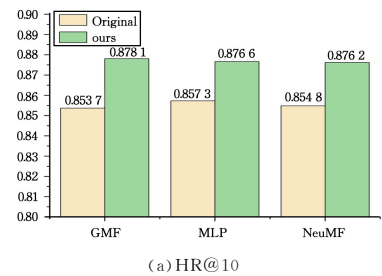


图 3 Pinterest-20 上 Top-10 性能的比较

Fig. 3 Comparison of Top-10 performance on Pinterest-20

从图 2 和图 3 可以明显地看出, EANCF 在两个数据集上

都取得了很好的表现。相对于 Movielens-1m 来说,相比原框架,EANCF 的 HR@10 平均提高了 2.4633%,NDCG@10 平均提高了 1.2833%,性能提升幅度较大。

逐点对数损失的方法能够更有效地验证隐式反馈在推荐性能上的准确性。EANCF 框架在 Movielens-1m 和 Pinterest-20 上每次迭代的损失值和推荐性能分别如图 4、图 5 所示。在现有的实验环境下,一次迭代需要耗时近 5 min,因此为了提高时间利用率,将迭代次数设置为 100。从图 4 可以看出,随着迭代次数的增多,推荐性能逐渐提高,损失逐渐降低,尤其在迭代到第 10 次左右时变化最为明显,这证明了 EANCF 框架在训练初期就已经取得了最

优值,性能极好。从图 5 可以看出,数据集的改变并没有对 EANCF 框架的性能产生较大影响,仍然是迭代到第 10 次左右时所提框架就已经达到了最优值。尽管图 5(a)和图 5(b)中,模型在迭代到接近 40 次时由于过拟合受到框架多层运算的限制,性能出现了一次反弹,但它仍然没有超过前 10 次迭代的性能,这恰恰证明了 EANCF 框架的稳定性。

除了通过以上实验对 EANCF 进行评估和测试之外,我们还对比了 Dacrema 等<sup>[25]</sup>的实验结果来证明 EANCF 框架的有效性。Dacrema 等考虑了以下基线方法,实验评估结果如图 6 和图 7 所示。

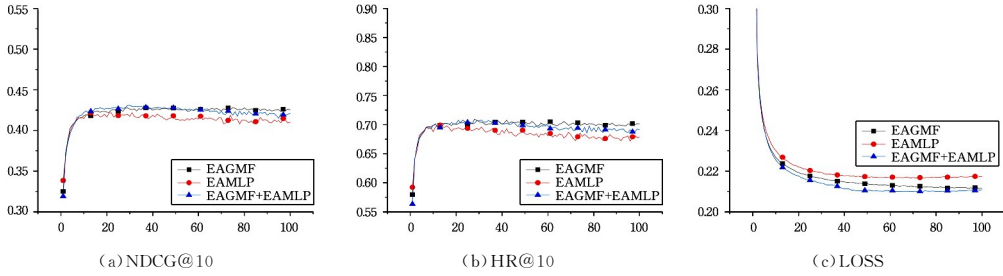


图 4 Movielens-1m 上的实验结果

Fig. 4 Experimental results on Movielens-1m

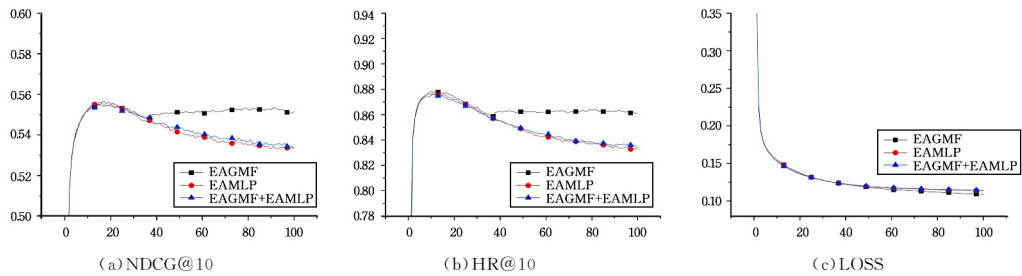


图 5 Pinterest-20 上的实验结果

Fig. 5 Experimental results on Pinterest-20

1) Toppopular:一种非个性化的方法,向每个用户推荐最受欢迎的项目,“最受欢迎”的定义可以通过显式或隐式数据的数量来衡量;

2) ItemKNN:一种传统的协同过滤方法,基于 K 最近邻算法衡量物品之间距离的物品相似度推荐算法;

3) UserKNN:基于相似用户的推荐算法,与 ItemKNN 类似,计算样本点从 ItemKNN 中的物品距离转变成每个用户自身的评级;

4) P<sup>3</sup>α 和 RP<sup>3</sup>β:一种基于图的算法,实现了用户与项目在图上的随机游走,将给用户推荐物品的任务转化为度量用户。

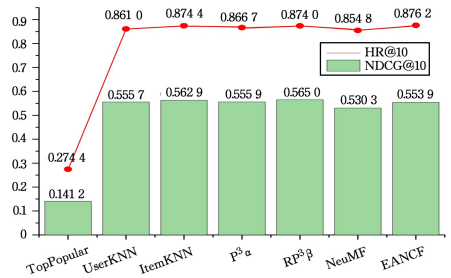


图 7 Pinterest-20 上方法的对比结果

Fig. 7 Comparisons results of methods on Pinterest-20

从图 6 和图 7 可以看出,以 HR@10 和 NDCG@10 为目标,无论是在 Movielens-1m 数据集还是 Pinterest-20 数据集上,EANCF 框架都展现了较好的性能。在 Movielens-1m 数据集上,EANCF 的 HR@10 指标比最好的 NeuMF 提升了 2.48%,NDCG@10 指标比最好的 NeuMF 提升了 1.35%;在 Pinterest-20 数据集上,EANCF 的 HR@10 指标比最好的 ItemKNN 提升了 0.18%,NDCG@10 指标效果略差,我们认为这是由于数据比较稀疏,训练时无法集中于评分较高的项,对这些项无法产生一个很好的拟合。

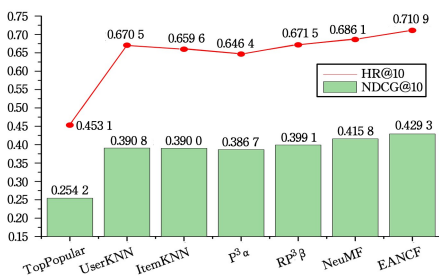


图 6 Movielens-1m 上方法的对比结果

Fig. 6 Comparison results of methods on Movielens-1m

**结束语** 本文将深度学习应用于隐式推荐系统,在原有 NCF 框架的基础上提出了一种新的 EANCF 框架,在数据非隐式而稀疏的情况下,我们对数据进行预处理。此外, EANCF 框架分别对用户和项目向量进行最大池化操作,保留了全局最大信息,引入并扩展了 ESIM 的局部推理方法,通过相乘、求和等多个不同运算组件来对用户和项目向量之间的潜在向量进行提取,帮助捕获差异性,提高隐式关系的获取量,以利于接下来模型的学习;同时,引入注意力机制来合理配置权重,增加推荐的高效性。实验证明,相比于原有的框架, EANCF 优化了数据关系的提取能力,具有较高的召回率和较强的归一化能力。但是, EANCF 也存在不足,协同过滤方法的性能由于用户项目交互数据的稀疏性而受到限制;框架的时间复杂度大;实验设置参数较多,工作量大且参数选取困难。

未来将会继续尝试使用辅助信息来提高性能,对用户评论数据进行情感分析,提取出用户对项目的不同喜好程度作为隐式数据,以此来提高推荐系统的性能。

### 参 考 文 献

- [1] HE X N. Neural collaborative filtering[C]//Proceedings of the 26th International Conference on World Wide Web. 2017:173-182.
- [2] HE X, DU X, WANG X, et al. Outer Product-based Neural Collaborative Filtering [C] // Twenty-Seventh International Joint Conference on Artificial Intelligence. 2018.
- [3] BAI T, WEN J R, ZHANG J, et al. A neural collaborative filtering model with interaction-based neighborhood[C]//Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. ACM, 2017:1979-1982.
- [4] WANG X, HE X, WANG M, et al. Neural graph collaborative filtering[C]//Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval. 2019:165-174.
- [5] JIANG Y. Fusion Information Recommendation Model Based on Graph Convolution and Neural Collaborative Filtering [D]. Changchun: Jilin University, 2018.
- [6] XIE E N, HE L M, WANG X H. Deep Collaborative Filtering Model Based on Attention Mechanism [J]. Journal of Metrology University of China, 2019, 30(2): 219-225, 242.
- [7] PORTEOUS I, ASUNCION A, WELLING M. Bayesian matrix factorization with side information and dirichlet process mixtures[C]//Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence. 2010.
- [8] CHEN Q, ZHU X, LING Z, et al. Enhanced LSTM for Natural Language Inference[J]. arxiv: 1609. 06038, 2016.
- [9] OARD D W, KIM J. Implicit feedback for recommender systems [C] // Proceedings of the AAAI Workshop on Recommender Systems. 1998.
- [10] YAO W L, HE J, HUANG G Y, et al. A Graph-based model for context-aware recommendation using implicit feedback data[J]. World Wide Web, 2015, 18(5): 1351-1371.
- [11] HE M, MORIMOTO Y. Capturing Temporal Dynamics of Implicit Feedbacks for Collaborative Filtering by Using Deep Recurrent Neural Networks[J]. Bulletin of Networking, Computing, Systems, and Software, 2018, 7(1): 33-37.
- [12] GOLDBERG D, NICHOLS D, OKI B M, et al. Using collaborative filtering to weave an information tapestry[J]. Communications of the ACM, 1992, 35(12): 61-70.
- [13] RESNICK P, IACOVOU N, SUCHAK M, et al. GroupLens: an open architecture for collaborative filtering of netnews [C] // Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work. ACM, 1994: 175-186.
- [14] ZHAO X Y, XIA L, ZHANG L, et al. Deep reinforcement learning for page-wise recommendations [C] // Proceedings of the 12th ACM Conference on Recommender Systems. 2018: 95-103.
- [15] PASRICHA R, MCAULEY J. Translation-based factorization machines for sequential recommendation [C] // Proceedings of the 12th ACM Conference on Recommender Systems. 2018: 63-71.
- [16] SUN Z, YANG J, ZHANG J, et al. Recurrent knowledge graph embedding for effective recommendation [C] // Proceedings of the 12th ACM Conference on Recommender Systems. 2018: 297-305.
- [17] DUAN D K, FU X F. Research on User Cold Start Problem in Hybrid Collaborative Filtering Algorithm [J]. CEA, 2017, 53(21): 151-156.
- [18] ZHANG X C. Utilizing Tri-training Algorithm to Solve Cold Start Problem in Recommender System [J]. Computer Science, 2016, 43(12): 108-114.
- [19] DENG C B, YU H Q, FAN G S. Integrating Dynamic Collaborative Filtering and Deep Learning for Recommendation [J]. Computer Science, 2019, 46(8): 28-34.
- [20] ZHENG L. A survey and critique of deep learning on recommender systems [J]. Chicago: University of Illinois, 2016: 31.
- [21] LONG C, ZHANG H, XIAO J, et al. SCA-CNN: Spatial and Channel-Wise Attention in Convolutional Networks for Image Captioning [C] // IEEE Conference on Computer Vision & Pattern Recognition. 2017: 5659-5667.
- [22] BAHDANAU D, CHO K, BENGIO Y. Neural Machine Translation by Jointly Learning to Align and Translate [J]. arxiv: 1409. 0473, 2014.
- [23] LI J, REN P, CHEN Z, et al. Neural attentive session-based recommendation [C] // Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017: 1419-1428.
- [24] GUO X, ZHU J H. Deep Neural Network Recommendation Model Based on User Vectorization Representation and Attention Mechanism [J]. Computer Science, 2019, 46(8): 111-115.
- [25] DACREMA M F, CREMONESI P, JANNACH D. Are we really making much progress? A worrying analysis of recent neural recommendation approaches [C] // Proceedings of the 13th ACM Conference on Recommender Systems. ACM, 2019: 101-109.



**KANG Yan**, born in 1972, Ph.D, post-graduate supervisor, is a member of China Computer Federation. Her main research interests include machine learning and software engineering.