

融合特征选择和交叉网络的增强推荐模型^①

师欣雨^{1,2}, 林珊玲^{1,2}, 刘珂^{1,2}, 林坚普^{1,2}, 吕珊红^{1,2}, 林志贤^{1,2,3}, 郭太良^{2,3}

¹(福州大学 先进制造学院, 泉州 362251)

²(闽都创新实验室 (中国福建光电信息科学与技术创新实验室), 福州 350108)

³(福州大学 物理与信息工程学院, 福州 350116)

通信作者: 林志贤, E-mail: lzx2005000@163.com



摘要: 针对目前大多数推荐模型在特征交互时, 存在忽视特征重要程度使得推荐模型准确率不高的问题, 为此本文提出融合特征选择和交叉网络的增强推荐模型. 该模型采用 SENet 网络在特征交互前过滤不重要的特征, 使其挖掘到更有价值的交互信息. 在此基础上, 进一步使用并行的交叉网络和深度神经网络, 以捕捉显式特征交互和隐式特征交互. 同时, 在交叉网络中引入低秩技术, 将权重向量改进为低秩矩阵, 在保证模型性能的同时, 降低模型的训练成本. 该模型在 MovieLens-1M、Criteo 数据集上与其他推荐模型进行了对比实验, 实验结果表明所提推荐模型在 AUC 指标上明显优于其他模型, 证明了所提推荐模型的有效性.

关键词: 推荐算法; 深度学习; SENet 网络; 特征交互; 低秩矩阵

引用格式: 师欣雨, 林珊玲, 刘珂, 林坚普, 吕珊红, 林志贤, 郭太良. 融合特征选择和交叉网络的增强推荐模型. 计算机系统应用, 2024, 33(12): 97-105. <http://www.c-s-a.org.cn/1003-3254/9706.html>

Enhanced Recommendation Model Integrating Feature Selection and Cross Network

SHI Xin-Yu^{1,2}, LIN Shan-Ling^{1,2}, LIU Ke^{1,2}, LIN Jian-Pu^{1,2}, LYU Shan-Hong^{1,2}, LIN Zhi-Xian^{1,2,3}, GUO Tai-Liang^{2,3}

¹(School of Advanced Manufacturing, Fuzhou University, Quanzhou 362251, China)

²(Fujian Science & Technology Innovation Laboratory for Optoelectronic Information of China, Fuzhou 350108, China)

³(College of Physics and Information Engineering, Fuzhou University, Fuzhou 350116, China)

Abstract: Most current recommendation models often overlook the importance of features during feature interactions, leading to low accuracy. To address this issue, an enhanced recommendation model combining feature selection and the cross network is proposed. The SENet network is employed to filter out unimportant features before feature interaction, enabling the extraction of more valuable interaction information. On this basis, parallel cross network and deep neural network are utilized to capture explicit and implicit feature interactions. Additionally, low-rank techniques are introduced in the cross network, transforming weight vectors into low-rank matrices to maintain model performance and reduce model training costs. Comparative experiments on the datasets of MovieLens-1M and Criteo demonstrate that the proposed recommendation model is significantly superior to other models in terms of AUC metrics, which proves the effectiveness of the proposed recommendation model.

Key words: recommendation algorithm; deep learning; SENet network; feature interaction; low-rank matrix

随着互联网的飞速发展, 人们每天都要面对大量的信息, 在享受获取信息便利的同时, 也伴随着信息选择困难、信息焦虑等问题^[1]. 用户要从海量的数据中选

出自己喜欢的内容需要花费很大精力, 智能推荐技术应运而生. 推荐算法能够基于大量的用户数据, 从中快速发现用户需求, 并推荐给用户. 其目前已被广泛应用

① 基金项目: 国家重点研发计划 (2021YFB3600603)

收稿时间: 2024-05-27; 修改时间: 2024-06-26; 采用时间: 2024-07-04; csa 在线出版时间: 2024-10-25

CNKI 网络首发时间: 2024-10-25

于个性化阅读、视频推荐、电子商务等各个领域^[2]。

传统的推荐算法包括基于内容的推荐、基于协同过滤的推荐和混合推荐^[3]。因为其未将用户和物品的属性考虑在内^[4]，只对用户和物品建立相似度矩阵进行分析，使得推荐结果种类单一。传统的推荐算法通常无法捕捉用户和物品之间复杂的非线性关系且推荐的物品集中在流行物品上，缺乏多样性，不能满足用户的长尾需求，这一问题大大限制了推荐算法的效果。深度学习的出现使得传统推荐算法的数据稀疏、推荐单一性等问题得到改善^[5]。Google Play 提出^[6]著名的深广 (Wide & Deep learning) 模型，其将非线性模型 DNN (deep neural network) 和浅层模型 LR (logistic regression) 精巧融合^[7]，赋予了模型记忆能力和泛化能力，模型的复杂程度低，便于训练和推理。Wang 等人^[8]将 Wide 侧的 LR 换成了 cross，提出 DCN (deep & cross network) 模型，在元素级别上实现自动的高阶交互。Guo 等人^[9]为了让模型实现端到端的训练并自动学习低阶交互特征，提出 DeepFM (factorization-machine based neural network) 模型，即在 Wide & Deep 模型上将 Wide 部分替换为因式分解机 FM (factorization machine)。He 等人^[10]提出的 NFM (neural factorization machine) 模型在嵌入层和多层神经网络之间加入特征交叉池化层。Xiao 等人^[11]提出的 AFM (attentional factorization machine) 模型在特征交叉层和输出层间加入注意力网络，通过计算二阶交叉特征的权重系数来学习每个特征交互的重要度。Chen 等人^[12]为解决隐层共享不足和输入层过度共享的问题提出了改进模型 EDCN (enhanced deep & cross network)。

特征交互在推荐模型中是最重要的一步，但是上述大部分模型的特征交互挖掘仍不够，有些模型无法有效地自动学习有限阶的特征交互，还有些模型则将所有单独的特征和特征组合等同对待，进而影响了最终的推荐效果。因此，为了深入挖掘用户和物品之间的特征交互，本文提出了一种融合特征选择和交叉网络的增强推荐模型 (enhanced recommendation model integrating feature selection and cross network, SECNet)。主要工作贡献如下。

(1) 针对特征交互时只使用内积或哈达玛积计算，这种忽略特征重要程度使得推荐模型准确率不高的问题，提出组合内积和哈达玛积进行双线性特征交叉，从而提高模型对特征间复杂关系的学习能力和推荐准确率。

(2) 通过融合特征交叉层来提高模型对特征之间

交互信息的捕捉能力，进而更准确地表达显式和隐式的特征交互，从而提高推荐算法的性能；同时在交叉网络中引入低秩技术，旨在保证模型性能的同时降低模型的训练成本。

(3) 在 MovieLens-1M 和 Criteo 数据集上进行对比实验和消融实验，结果表明 AUC 分别达到 81.618% 和 78.211%，较 DCN 基准模型提高了 1.054% 和 2.213%，验证本文提出的模型效果较好，并分析模型改进的有效性。

1 相关理论

1.1 DCN 模型

随着深度学习技术的研究更加深入，推荐系统领域也逐渐迎来了革新。传统的推荐算法虽然在一定程度上能够满足需求，但伴随数据量的急剧增加和用户行为的复杂性^[13]，其效果逐渐受到限制。而基于深度学习的推荐模型，通过更深层次的特征学习，使其更好地挖掘数据之间的潜在关联，从而提升推荐的准确性和个性化程度。DCN 模型^[14]是一种经典的并行架构融合网络，由交叉网络 (cross network) 和深度神经网络 (deep network) 两者独立计算^[15]，输出部分由两者融合。交叉网络^[16]用于特征的自动化交叉编码，深度神经网络用于特征的隐式交叉编码，与 Wide & Deep 和 DeepFM 模型类似，都由嵌入层进行预处理开始，输入到两个网络中，最后将两个并行网络的输出进行简单拼接通过激活函数作为模型的输出：

$$y = \text{Sigmoid}\left(\left[x_{L_1}^T, h_{L_2}^T\right] W_{\text{logits}}\right) \quad (1)$$

其中， $x_{L_1}^T$ 和 $h_{L_2}^T$ 分别是交叉网络和深度神经网络的输出。

DCN 模型首先对用户性别、年龄等类别特征^[17]进行编码和嵌入处理，然后将其与交互时间等连续特征进行拼接组合。其不需要特征工程来获得高阶的交叉特征，因而拥有更高的计算效率并且能够提取到更高阶的交叉特征，也因此作为一个基准模型^[18-20]被广泛应用。

1.2 SENet 网络

SENet 网络最早在计算机视觉领域中应用。其核心思想是在卷积神经网络的基础上，通过建模不同特征通道之间的关系，增强重要特征的信号，同时抑制不重要的特征通道，从而更好地捕捉和利用关键特征，提高神经网络的表征能力，进而提升模型性能。近年来，

一些研究将 SENet 网络引入推荐模型^[21,22], 用于提前筛选重要特征, 赋予其更大的权重, 使组合特征表达具有差异性, 从而让推荐更加符合用户的实际需求. 基于这一研究路线, 本文设计了一个 SECNet 模型, 将 SENet 引入推荐系统, 有效缓解了特征交互不足和忽略组合特征差异性的问题, 提高了推荐效果.

2 SECNet 模型

针对目前大多数推荐模型在特征交互时, 存在忽视特征重要程度使得推荐模型准确率不高的问题, 本文提出一种融合特征选择和交叉网络的增强推荐模型

SECNet. 如图 1 所示. 该模型的输入层以用户属性和商品属性作为输入, 对连续型特征进行归一化直接输入模型的主体部分, 对类别型特征进行 Label Encoder 编码再通过嵌入层进入模型的主体部分. 将原始输入向量与经过 SENet 网络后加入权重的特征向量一起送入双线性特征交互层, 然后将其得到的结果进行拼接送入并行的深度神经网络和交叉网络, 两部分共享输入. 其次通过桥接模块对深度神经网络和交叉网络的输出进行拼接, 最后将桥接模块的输出与深度神经网络和交叉网络的输出进行线性整合和激活函数作用后, 输出最终预测概率.

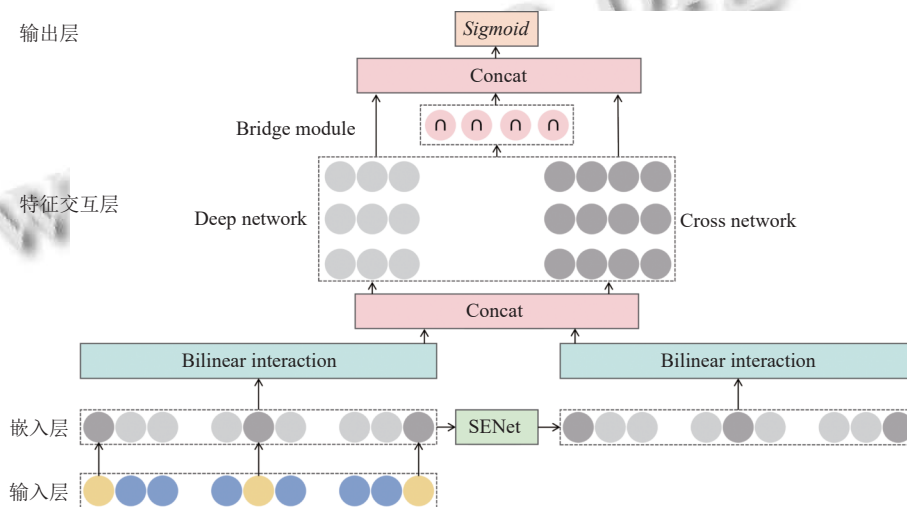


图 1 SECNet 模型结构图

2.1 模型输入部分

在推荐系统中, 模型输入的特征一般分为连续型特征和类别型特征^[23]. 通过将其映射到相同的低维特征空间中进行特征交叉.

连续型特征直接映射到低维嵌入空间中, 每个连续型特征对应一个特征向量. 而类别型特征, 则需对其使用 Label Encoder 编码处理为数值型. Label Encoder 编码通过为每个类别分配一个唯一的整数标签来表示该类别^[20], 从 0 开始递增.

为了进一步将这些整数标签转换为稠密向量表示, 需要在类别型编码层后添加嵌入层. 每个整数标签被映射为一个高维的实数向量, 这样就可以在嵌入空间中表示类别之间的相似性和差异性. 这种表示能够捕捉到类别之间的复杂关系, 尤其在高维空间中可以更好地区分不同的类别. 即使在训练过程中没有见过的类别, 通过嵌入层也能够得到一个合理的向量表示, 而

不是简单的将其归为某个整数标签, 这使得模型对未知类别的泛化能力更强.

2.2 模型网络部分

2.2.1 特征选择层

将嵌入层输出的低维稠密向量通过 SENet 网络进行特征选择, 该网络的作用与注意力机制相同, 通过放大重要特征, 抑制无效噪声, 在特征交互前过滤不重要特征. SENet 层如图 2 所示.

SENet 网络里面的具体计算过程, 主要分为 3 个步骤, 分别是 Squeeze, Excitation, Re-weight^[22].

Squeeze 阶段: 通过将每个经过嵌入处理后的低维稠密向量进行压缩, 并求其均值, 以表示相关的特征信息. 与计算机视觉中求其最大值不同, 推荐算法使用平均值来表示每个特征都有其存在的意义. 假设经过嵌入层的特征向量有 f 个, 每个特征向量的维度是 k 维, 求这些元素的均值, 经嵌入层后每个样本的特征向量 $E = [e_1,$

$\dots, e_f]$ 转变为经 Squeeze 阶段压缩后的向量 $Z = [z_1, \dots, z_f]$, 具体的计算公式如下:

$$z_i = F_{sq}(e_i) = \frac{1}{k} \sum_{t=1}^k e_i^{(t)} \quad (2)$$

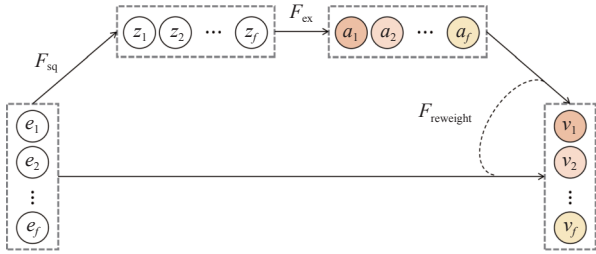


图2 SENet 层

Excitation 阶段: 利用 Squeeze 阶段计算得到的 Z 去学习每个特征对应的权重, 通过两个全连接层进行特征交互, 来判断输入的特征的影响程度, 权重向量用 $A = [a_1, \dots, a_f]$ 来表示. 计算公式如下:

$$A = F_{ex}(Z) = \sigma_2(W_2\sigma_1(W_1Z)) \quad (3)$$

其中, $A \in R^f$, $W_1 \in R^{f \times \frac{f}{r}}$, $W_2 \in R^{\frac{f}{r} \times f}$, r 是 reduction ratio 即压缩比例, f 是特征向量的个数, σ_1 和 σ_2 表示激活函数, SENet 模型使用 ReLU 激活函数.

Re-weight 阶段: 将 Excitation 中得到的权重大小与原来的 E 相乘, 对特征进行重要性加权. 公式如下:

$$V = F_{reweight}(A, E) = [a_1 \times e_1, \dots, a_f \times e_f] = [v_1, \dots, v_f] \quad (4)$$

其中, $a_i \in R$, $e_i \in R^k$, $v_i \in R^k$, k 为特征向量的维度. a_i 的大小决定了特征的重要程度.

2.2.2 双线性特征交互层

双线性特征交互通过引入一个额外的参数矩阵 W 来学习特征交互, 同时结合内积和哈达玛积的优势, 实现全局与局部的特征交互, 使模型更加精确地捕捉特征间的复杂关系, 挖掘到更丰富的交互信息, 双线性特征交互层如图 3 所示.

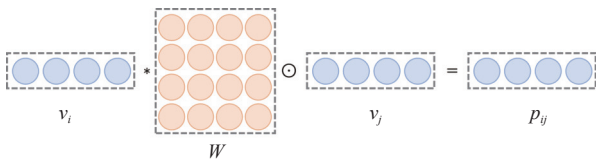


图3 双线性特征交互层

参数矩阵 W 有 3 种计算方式:

(1) Field-all type

$$p_{ij} = v_i \cdot W \odot v_j \quad (5)$$

所有的交叉特征使用同一个权重矩阵 W , $W \in R^{k \times k}$, k 为权重矩阵的维度.

(2) Field-each type

$$p_{ij} = v_i \cdot W_i \odot v_j \quad (6)$$

每个特征维护一个权重矩阵 W_i , $W_i \in R^{k \times k}$, $W = [W_1, W_2, \dots, W_i, \dots, W_f] \in R^{f \times k \times k}$, f 表示特征的个数.

(3) Field-interaction type

$$p_{ij} = v_i \cdot W_{ij} \odot v_j \quad (7)$$

每一组的交叉特征维护一个权重矩阵 W_{ij} , $W_{ij} \in R^{k \times k}$, 将特征向量两两交叉, 交叉特征对共有 $f \times (f-1)/2$ 个, 权重也有 $f \times (f-1)/2$ 个.

将通过 SENet 网络进行加权后的特征向量 V , 输入双线性特征交互层得到向量 $P = [p_1, \dots, p_{f(f-1)/2}]$, 同时将原始输入向量, 输入双线性特征交互层得到向量 $Q = [q_1, \dots, q_{f(f-1)/2}]$, 将两部分输出向量进行 Concat 拼接得到向量 $C = [c_1, \dots, c_{f(f-1)}]$, 将 C 送入并行的深度神经网络和融入低秩技术的交叉网络.

2.2.3 深度神经网络和交叉网络并联层

深度神经网络用来捕获非线性高阶特征, 是一个全连接的多层前馈神经网络. 其正向传播过程用计算公式可以表示为:

$$h_{l+1} = \sigma(W_l h_l + b_l) \quad (8)$$

其中, l 表示隐含层的层数, h_l 表示第 l 层的输出, h_{l+1} 同理, W_l 表示第 l 层的权重, b_l 表示第 l 层的偏置项 σ 表示激活函数, SENet 模型使用 ReLU 激活函数.

交叉网络是由多个交叉层堆叠而成的结构, 用来提取线性交叉组合特征. 第 $l+1$ 层输出向量是由第 l 层输出向量和原始输入向量共同计算得到的. 其每一层的计算公式如下:

$$x_{l+1} = x_0 x_l^T w_l + b_l + x_l \quad (9)$$

其中, x_0 表示初始输入, l 表示交叉层数, x_l 表示交叉层中第 l 层的输出, x_{l+1} 表示交叉层中第 $l+1$ 层的输出, w_l 是第 l 层的权重向量, b_l 是偏置项.

由于交叉网络的跨网络代表性有限, 多项式类仅由输入大小来表征, 这在很大程度上限制了随机交叉的灵活性. 鉴于此, 将交叉层计算中的权重向量 w_l 改进为权重矩阵 W_l 来解决这个问题. 改进后的公式如下:

$$x_{l+1} = x_0 \odot (W_l \times x_l + b_l) + x_l \quad (10)$$

特征向量会先在交叉前对 x_i 进行线性映射, 然后与初始的输入特征 x_0 做哈达玛积. 特征向量中的每个元素与交叉向量的所有元素进行特征交互, 增强了模型显式特征交互的能力.

同时, 为了使模型在保证性能的同时减少计算成本, SECNet 将低秩技术^[24]运用到交叉网络中. 由于低秩矩阵的行和列可以通过其他行或列来表示, 因而包含了大量冗余信息. 模型可以利用这些信息来提取相关特征. 在计算公式中, 将权重矩阵 W_i 近似为两个低秩矩阵 U_i 和 V_i :

$$x_{i+1} = x_0 \odot (U_i (V_i^T x_i) + b_i) + x_i \quad (11)$$

桥接模块 (bridge module) 通过一种函数将两个网络的信息进行组合, 然后输出一组新的特征. 该模型采用的交叉组合方式为拼接+全连接层 (Concatenation+FC), 利用带有激活函数的线性层进行输出, 是一种数据合并的操作, 通过在特定维度上连接数据, 使新的数据结构包含了原始数据的信息, 以便提供更丰富的信息给后续的层, 公式为:

$$f_L = \text{ReLU}(w^T [x_L, h_L] + b) \quad (12)$$

其中, w^T 是全连接层的权重矩阵, b 是偏置项, $[x_L, h_L]$ 是交叉网络和深度神经网络拼接后的输出向量.

并行的深度神经网络和交叉网络, 两部分共享输入, 并且通过桥接模块对深度神经网络和交叉网络的输出进行拼接.

2.3 模型输出部分

SECNet 模型使用各部分的输出层进行线性整合并利用 Sigmoid 函数输出最后的预测结果, 公式可以表示为:

$$\hat{y} = \text{Sigmoid}(w^T (x_L, h_L, f_L) + b) \quad (13)$$

其中, x_L 是交叉网络的输出, h_L 是深度神经网络的输出, f_L 是桥接模块的输出, w^T 表示每个输出的权重, b 是偏置项.

3 实验结果与分析

3.1 实验设置

3.1.1 数据集

本实验所使用的数据集为 MovieLens-1M 数据集和 Criteo 数据集. MovieLens 数据集由 GroupLens Research 研究项目收集并整理^[25]. 此数据集涵盖了用

户性别、年龄等用户信息和电影编号、类别等电影信息. Criteo 数据集是由 Criteo 广告公司提供的 4500 万条用户点击数据^[26], 包含 13 个连续型特征和 26 个类别型特征, 受硬件设备影响, 选取数据集的前 100 万条作为实验数据. 对于 MovieLens-1M 数据集, 不考虑文本特征, 因此删除数据集中的“Title”特征. 本实验根据“Rating”字段新添加一条类别特征记作“Label”, 将“Rating=5”的样本的“Label”特征设置为 1, 即评分为 5 分的记作 1, 其他样本记作 0. 由于 Criteo 数据集存在大量的数据缺失, 统一将连续型特征缺失值补充为 0, 类别型特征缺失值补充为-1. 将数据集分为训练集 (80%) 和测试集 (20%). 在训练集中再次划分出验证集 (80% 用于训练, 20% 用于验证).

3.1.2 模型评估

实验评价指标是 AUC (area under the ROC curve)^[27]. AUC 表示的是 ROC 曲线下的面积^[28], 用来衡量不同模型的预测性能, 取值范围为 [0, 1]. AUC 指标为模型能正确地将正样本排在负样本前面的概率. 在推荐系统中, AUC 越大说明越容易被点击的商品排得越靠前, 推荐效果越好. 该评价指标不依赖于样本是否均衡, 在不均衡的样本数据中, 依然可以合理预测, 所以选择其作为本实验的评价指标之一. AUC 计算公式如下所示:

$$AUC = \frac{\sum I(P_{\text{正样本}}, P_{\text{负样本}})}{M \times N} \times 100\% \quad (14)$$

$$I(P_{\text{正样本}}, P_{\text{负样本}}) = \begin{cases} 1, & P_{\text{正样本}} > P_{\text{负样本}} \\ 0.5, & P_{\text{正样本}} = P_{\text{负样本}} \\ 0, & P_{\text{正样本}} < P_{\text{负样本}} \end{cases} \quad (15)$$

其中, M 是正样本个数, N 是负样本个数, $M \times N$ 表示总的样本对数.

对数损失 LogLoss (logarithmic loss)^[13,29] 用于评估模型预测真实值与预测值之间的差距, 取值范围为 [0, 1]. 在推荐算法中以此来衡量模型的预测准确率, LogLoss 值越小表示模型的预测与真实情况的吻合程度越高. 设 $y_i \in \{0, 1\}$ 为样本真实值, 为样本是正例的概率. 对数损失函数计算公式如下所示:

$$\text{LogLoss} = -\frac{1}{N} \sum_{i=1}^N (y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)) \quad (16)$$

3.1.3 对比模型

(1) FM^[30]: 使用因子分解机, 对特征两两交互, 考虑了特征之间的非线性关系.

(2) DeepFM^[9]: 使用因子分解机 (FM) 捕获低阶交

叉特征, 神经网络捕获高阶交叉特征, 最终两者并
联预测输出。

(3) AFM^[11]: 在 FM 模型上加入一层注意力机制, 通过对每对特征的交互引入注意力权重, 用以区分不同特征重要性对测试结果的影响。

(4) DCN^[8]: 由交叉网络和神经网络并列组成, 弥补 FM 模型表层结构限制, 将参数共享机制从一层拓展到多层。

(5) NFM^[10]: 使用 Bi-Interaction 层进行二阶特征交互, 增强了模型特征交叉的能力。

(6) xDeepFM^[31]: 与 DCN 模型结构类似, 在模型中引入 CIN 网络, 将比特级别的特征交互增强到向量级别。

(7) EDCN^[12]: 在 DCN 模型的每一层并行结构中穿插使用了“桥接模块 (bridge module)”和“调节模块 (regulation module)”, 以此来捕获 cross 和 deep 之间的分层交互信号, 加强了信息的融合。

(8) DCN V2^[24]: 优化了交叉层的建模方式, 利用更灵活的非线性变换来增强特征交互的表达能力。

3.1.4 实现环境及超参数设置

本实验的操作系统为 Windows 11, 处理器为 11th Gen Intel(R) Core(TM) i5-11300H@3.10 GHz, 运行内存为 16 GB, 软件平台是 PyCharm 2022.3.3 专业版, 深度学习计算框架是 TensorFlow 2.6.2^[32]。模型主要参数设置如表 1 所示。

表 1 模型参数设置

超参数	MovieLens-1M	Criteo
batch_size	256	256
学习率	0.001	0.001
优化器	Adam	Adam
epoch	10	10
MLP参数	[256, 128, 64]	[256, 128, 64]
dropout	0.2	0.2
l2	0.0001	0.001
reduction_ratio	9	7
low_rank	16	16
cross_num	5	5
bilinear_type	interaction	interaction

3.2 实验结果

为验证本文提出的 SECNet 模型的有效性, 在两个数据集上对所有模型进行实验, 实验结果如表 2 所示。

由表 2 不同模型在两个数据集上的实验结果可知, 本文所提出的 SECNet 模型的 AUC 和 LogLoss 在两个

数据集上均有提升和优化。相较于基准模型 DCN, 在 MovieLens-1M 和 Criteo 数据集上 AUC 分别有 1.054% 和 2.213% 的提升, LogLoss 上分别有 0.823% 和 1.701% 的优化。结果表明, 将融合特征选择和交叉网络的增强推荐模型在 MovieLens-1M 和 Criteo 数据集上进行预测是真实可靠的, 并优于其他推荐模型。一方面, 通过融合交叉网络和特征选择层, 模型能够更准确地捕捉特征之间的交互信息。使模型能够识别和利用特征之间复杂的关系, 在减少噪声和冗余的同时, 确保模型专注于重要的特征, 从而显著提升模型的性能和泛化能力。另一方面, 引入桥接模块和低秩矩阵进一步优化了模型的结构, 桥接模块建立有效的连接, 使信息传递更加顺畅, 从而提升模型的整体表现。低秩矩阵通过捕捉数据中的冗余信息, 提高了模型的实用性和可扩展性, 能够更好地适应不同应用场景的需求。

表 2 不同模型在两个数据集上的实验结果 (%)

Model	MovieLens-1M		Criteo	
	AUC	LogLoss	AUC	LogLoss
FM	79.522	43.307	75.866	48.077
DeepFM	80.858	42.118	75.661	48.284
AFM	78.603	43.888	75.064	48.743
DCN	80.564	42.240	75.998	48.071
NFM	79.899	42.585	75.292	48.746
xDeepFM	81.089	41.829	75.837	48.195
EDCN	80.761	42.891	75.093	48.734
DCN V2	80.691	42.177	76.049	48.124
SECNet	81.618	41.417	78.211	46.370

3.3 消融分析

为了分别验证 SENet 网络和双线性特征交互层、低秩技术和 bridge 模块在 SECNet 模型中对模型性能的影响, 本节通过消融实验对比来验证各个模块在本文 SECNet 模型中的有效性。实验结果由表 3 所示。

表 3 消融实验分析 (%)

Model	MovieLens-1M		Criteo	
	AUC	LogLoss	AUC	LogLoss
DCN	80.564	42.240	75.998	48.071
SECNet ^a	81.484	41.614	76.477	47.761
SECNet ^b	80.691	42.177	76.049	48.124
SECNet ^c	81.492	41.544	76.578	47.685
SECNet	81.618	41.417	78.211	46.370

表 3 中的第 1 个模型是经典模型 DCN, SECNet 模型是在此基础上进行改进的, SECNet^a 是只在基准模型上引入 SENet 网络和双线性特征交互层的模型, SECNet^b 表示的是将权重向量改进为低秩矩阵的模型,

SECNet^c是在SECNet基础上去除bridge模块的模型。由表可知,SENet网络和双线性特征交互层可以动态学习特征的重要性,通过放大重要的特征,抑制无效的噪声,使其在特征交互前过滤不重要的特征,SENet网络和双线性特征交互层对模型是有效的;引入低秩技术减少计算成本的方法,使得模型性能的提升明显;这表明SECNet模型进一步提高了特征交叉的能力,使模型性能得到进一步提升且每个组件都是有必要的。

3.4 重要超参数影响分析

为了研究超参数对SECNet模型性能的影响并找到最优值,本节设计了不同超参数的实验。

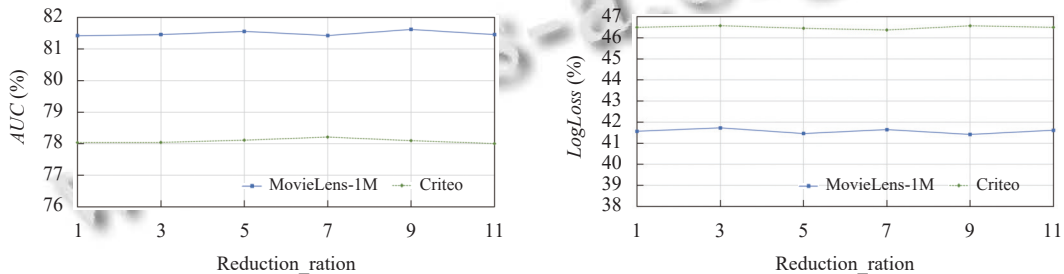


图4 Reduction_ratio对模型性能的影响

(2) 交叉网络中权重矩阵的秩对模型性能的影响

其他超参数设置与上述相同,MovieLens-1M数据集reduction_ratio为9,Criteo数据集reduction_ratio为7,实验结果如图5所示。在交叉网络中,权重矩阵的秩决定了参数数量,同时也影响了通过交叉层传递的低频信号的比例。在MovieLens-1M数据集中low_rank

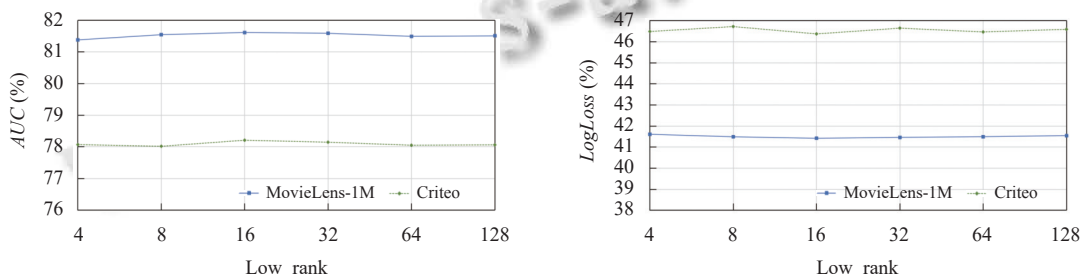


图5 Low_rank对模型性能的影响

(3) 交叉层中层深对模型性能的影响

其他超参数设置与上述相同,low_rank为16,实验结果如图6所示。随着层深的增加,交叉层捕获的高阶特征交叉组合数量也随之增加。交叉层深过大会影响模型的性能。由图可知,在5层时达到最佳效果。在超

(1) SENet中网络压缩程度对模型性能的影响

超参数设置为low_rank为16,cross_num为5,双线性模块中权重矩阵的计算方式选择interaction,实验结果如图4所示。Reduction_ratio是SENet网络在Excitation阶段对网络压缩程度的调节参数。由图4可知,在MovieLens-1M数据集中reduction_ratio当设置为9时,模型效果达到最优,在超过9时效果下降。在Criteo数据集中reduction_ratio为7时效果最好,7之后效果下降并振荡。这是因为过度压缩会导致信息丢失,而过低时模型会保留过多的冗余和噪声信息,导致效果振荡且不稳定。

设置为16时,模型效果达到最优。在Criteo数据集中low_rank为16时效果最好,16之后效果下降。由图5可知,适中的low_rank有助于保留和传递有用的低频信号,提升模型的泛化能力,而过高的low_rank可能引入过多的高频噪声,干扰模型的信号传递,导致性能下降。

过5层后,模型的性能开始下降。是因为层数过深会导致模型的复杂度增加,从而使模型出现了过拟合的现象。

(4) 权重矩阵计算方式对模型性能的影响

其他超参数设置与上述相同,cross_num为5,实验结果如图7所示。由图7可知,两数据集上双线性特

征交叉模块中权重矩阵的计算方式均选择 interaction 效果最好. 相比于其他计算方式, interaction 能够考虑

到特征之间的双线性关系, 从而更准确地建模特征交互, 提高模型的表达能力.

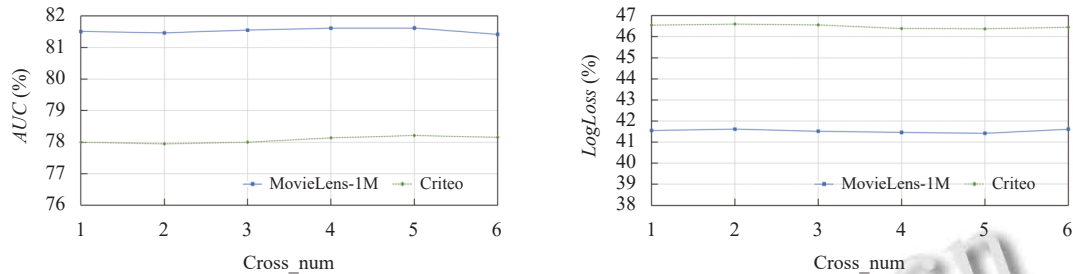


图6 Cross_num 对模型性能的影响

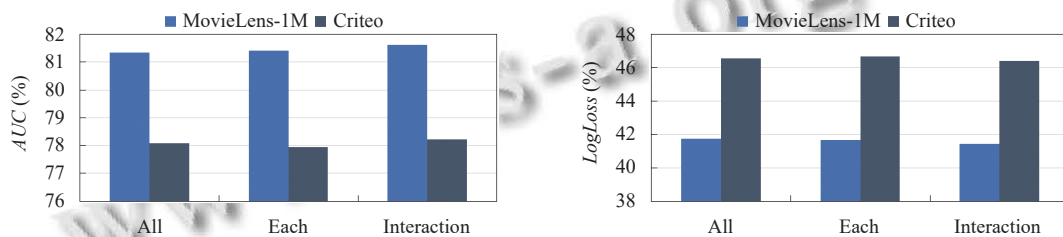


图7 权重矩阵计算方式对模型性能的影响

4 结束语

本文针对智能商品推荐问题, 提出一种基于深度学习的混合推荐模型 SECNet. 该模型首先将嵌入层输出的特征向量和经过 SENet 网络加权后的特征向量一起送入双线性特征交互层, 然后将得到的结果进行拼接送入并行的深度神经网络和交叉网络, 其次通过桥接模块对深度神经网络和交叉网络的输出进行拼接, 最后将桥接模块的输出与深度神经网络和交叉网络的输出进行线性整合和激活函数作用后, 产生最后的预估结果. 在真实数据集 MovieLens-1M 和随机采样的数据集 Criteo 上进行实验, 发现 SECNet 模型较于其他模型指标均有提升, 验证了模型的有效性.

该模型通过挖掘用户与商品间的交互特征, 放大重要的特征, 抑制无效的噪声, 有效提高了对重要信息的提取, 且推荐具有多样性, 解决了推荐问题单一的问题. 但本算法较依赖用户历史交互记录, 在冷启动问题上还有待进一步优化. 下一步工作是将传统推荐算法与基于深度学习的推荐算法进一步融合, 期待解决冷启动问题.

参考文献

1 翟梅. 个性化新闻推荐系统研究综述及探讨. 计算机与现

代化, 2024(4): 12–20. [doi: 10.3969/j.issn.1006-2475.2024.04.003]

2 温民伟, 梅红岩, 袁凤源, 等. 多任务推荐算法研究综述. 计算机科学与探索, 2024, 18(2): 363–377. [doi: 10.3778/j.issn.1673-9418.2303014]

3 毛骞, 谢维成, 乔逸天, 等. 推荐系统冷启动问题解决方法研究综述. 计算机科学与探索, 2024, 18(5): 1197–1210. [doi: 10.3778/j.issn.1673-9418.2308044]

4 苏湛, 陈学谦, 艾均, 等. 基于用户相似性选择及标签距离的推荐算法. 应用科学学报, 2023, 41(6): 940–957. [doi: 10.3969/j.issn.0255-8297.2023.06.003]

5 沈鑫科, 李勇, 陈建伟, 等. 融合协同知识图谱和图卷积网络的推荐算法. 计算机技术与发展, 2024, 34(1): 150–157. [doi: 10.3969/j.issn.1673-629X.2024.01.022]

6 Cheng HT, Koc L, Harmsen J, *et al.* Wide & Deep learning for recommender systems. Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. Boston: ACM, 2016. 7–10.

7 侯娜, 邵新慧. 基于 AP-XDeepFM 模型的广告点击率预测. 计算机应用与软件, 2022, 39(12): 108–113, 131. [doi: 10.3969/j.issn.1000-386x.2022.12.016]

8 Wang RX, Fu B, Fu G, *et al.* Deep & Cross network for ad click predictions. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax: ACM, 2017. 12.

- 9 Guo HF, Tang RM, Ye YM, *et al.* DeepFM: A factorization-machine based neural network for CTR prediction. Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne: AAAI Press, 2017. 1725–1731.
- 10 He XN, Chua TS. Neural factorization machines for sparse predictive analytics. Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Shinjuku: ACM, 2017. 355–364.
- 11 Xiao J, Ye H, He XN, *et al.* Attentional factorization machines: Learning the weight of feature interactions via attention networks. Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne: AAAI Press, 2017. 3119–3125.
- 12 Chen B, Wang YC, Liu ZR, *et al.* Enhancing explicit and implicit feature interactions via information sharing for parallel deep CTR models. Proceedings of the 30th ACM International Conference on Information & Knowledge Management. Singapore: ACM, 2021. 3757–3766.
- 13 白雪, 王霞光, 金继鑫, 等. 基于细粒度特征交互选择网络的农产品推荐算法. 计算机系统应用, 2024, 33(5): 271–279. [doi: 10.15888/j.cnki.csa.009519]
- 14 杨哲, 葛洪伟, 李婷. 特征融合与分发的多专家并行推荐算法框架. 浙江大学学报(工学版), 2023, 57(7): 1317–1325. [doi: 10.3785/j.issn.1008-973X.2023.07.006]
- 15 崔少国, 张岗, 王奥迪. 基于感知注意力的深度交叉网络推荐模型. 计算机与现代化, 2023(7): 54–60. [doi: 10.3969/j.issn.1006-2475.2023.07.010]
- 16 Gao TH, Gao S, Xu J, *et al.* DDRCN: Deep deterministic policy gradient recommendation framework fused with deep cross networks. Applied Sciences, 2023, 13(4): 2555. [doi: 10.3390/app13042555]
- 17 陈彬, 张荣梅, 张琦. DCFM: 基于深度学习的混合推荐模型. 计算机工程与应用, 2021, 57(3): 150–155. [doi: 10.3778/j.issn.1002-8331.1910-0424]
- 18 王盛民. 基于深度表征学习的弱偏差推荐系统研究与应用 [硕士学位论文]. 北京: 北京邮电大学, 2023. [doi: 10.26969/d.cnki.gbydu.2023.000162]
- 19 陈实秋. 基于推荐算法的种植推荐系统研究 [硕士学位论文]. 武汉: 武汉轻工大学, 2023. [doi: 10.27776/d.cnki.gwhgy.2023.000195]
- 20 唐震. 基于 DCN-CatBoost 的音乐混合推荐算法研究与应用 [硕士学位论文]. 南京: 南京邮电大学, 2021. [doi: 10.27251/d.cnki.gnjdc.2021.000047]
- 21 何李杰, 高茂庭. 基于交叉注意力的点击率预测模型. 计算机工程与应用, 1–9. <http://kns.cnki.net/kcms/detail/11.2127.TP.20240416.1000.012.html>. (2024-06-27).
- 22 童瑞嘉. 基于特征交互的电信套餐推荐算法研究与应用 [硕士学位论文]. 上海: 东华大学, 2023. [doi: 10.27012/d.cnki.gdhuu.2023.001338]
- 23 阿荣. 基于混合推荐的电影推荐系统的设计与实现 [硕士学位论文]. 北京: 北京邮电大学, 2023. [doi: 10.26969/d.cnki.gbydu.2023.001034]
- 24 Wang RX, Shivanna R, Cheng D, *et al.* DCN V2: Improved Deep & Cross network and practical lessons for web-scale learning to rank systems. Proceedings of the 2021 Web Conference 2021. Ljubljana: ACM, 2021. 1785–1797.
- 25 崔少国, 独潇, 杨泽田. 多注意力机制融合低高阶特征的神经推荐算法. 计算机工程与应用, 2023, 59(8): 192–199. [doi: 10.3778/j.issn.1002-8331.2112-0570]
- 26 陈育康, 龙慧云, 吴云, 等. 增强型高阶注意力因子分解机点击率预测模型. 计算机工程与应用, 2022, 58(24): 158–165. [doi: 10.3778/j.issn.1002-8331.2106-0001]
- 27 余鹏, 刘星雨, 程颢, 等. 在线课程推荐系统综述. 计算机工程与应用, 2023, 59(22): 1–14. [doi: 10.3778/j.issn.1002-8331.2305-0162]
- 28 张文奇, 王海瑞, 朱贵富. 融合异构图神经网络的时间卷积知识追踪方法. 小型微型计算机系统, 2024: 1–8. <http://kns.cnki.net/kcms/detail/21.1106.tp.20230915.1115.020.html>. (2023-09-18).
- 29 孙伟智, 欧鸥, 杜雪垒, 等. 多层结合的多阶显式交互的融合推荐模型. 计算机应用研究, 2022, 39(5): 1349–1356. [doi: 10.19734/j.issn.1001-3695.2021.10.0447]
- 30 Rendle S. Factorization machines. Proceedings of the 2010 IEEE International Conference on Data Mining. Sydney: IEEE, 2010. 995–1000. [doi: 10.1109/ICDM.2010.127]
- 31 Lian JX, Zhou XH, Zhang FZ, *et al.* xDeepFM: Combining explicit and implicit feature interactions for recommender systems. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London: ACM, 2018. 1754–1763.
- 32 于佳俊. 基于深度学习的个性化门店推荐系统的设计与实现 [硕士学位论文]. 上海: 华东师范大学, 2022. [doi: 10.27149/d.cnki.ghdsu.2022.004850]

(校对责编: 张重毅)