

# 基于场矩阵分解机和 CNN 的点击率预测模型<sup>①</sup>



王志格, 李汪根, 夏义春, 高 坤, 束 阳, 葛英奎

(安徽师范大学 计算机与信息学院, 芜湖 241002)

通信作者: 李汪根, E-mail: [xchen@ahnu.edu.cn](mailto:xchen@ahnu.edu.cn)

**摘 要:** 点击率预测是在线广告和推荐系统的基本任务之一. 主流模型通常通过对高阶和低阶特征进行特征交互建模来提升性能和泛化能力. 然而, 许多模型往往仅学习每个特征的固定表示, 而忽视了特征在不同上下文中的重要性, 并且一些模型结构过于简单. 因此, 本文提出了特征细化卷积神经网络融合场矩阵分解机 (FRCNN-F) 模型, 以解决这些问题. 首先, 在特征细化网络 (FRNet) 中融合了卷积神经网络的特征生成模块, 利用其在局部模式下重新组合生成新特征的优势, 提升了重要特征选择能力. 其次, 设计了场矩阵分解机, 使模型能够感知上下文并通过不同场的交互进行显示建模, 从而增加了子模型的组合方式. 最后, 通过在 Frappe 和 MovieLens 两个公开数据集上对比实验, 实验结果表明, FRCNN-F 模型相比基线 FRNet 在 AUC 得分分别提升了 0.32% 和 0.40%, 交叉熵损失函数 *Logloss* 分别降低了 1.50% 和 1.11%. 该研究对于实现广告的精准投放和个性化推荐具有实际应用的价值.

**关键词:** 点击率预测; 特征交互; 特征细化网络; 卷积神经网络; 场矩阵分解机

引用格式: 王志格, 李汪根, 夏义春, 高坤, 束阳, 葛英奎. 基于场矩阵分解机和 CNN 的点击率预测模型. 计算机系统应用, 2024, 33(1): 87-98. <http://www.c-s-a.org.cn/1003-3254/9353.html>

## Click-through Rate Prediction Model Based on Field-matrixed Factorization Machines and CNN

WANG Zhi-Ge, LI Wang-Gen, XIA Yi-Chun, GAO Kun, SHU Yang, GE Ying-Kui

(School of Computer and Information, Anhui Normal University, Wuhu 241002, China)

**Abstract:** Predicting click-through rate (CTR) is a fundamental task in online advertising and recommendation systems. Mainstream models often enhance performance and generalization by modeling interactions between high-order and low-order features. However, many models only learn fixed representations of each feature, neglecting the importance of features in different contexts and having overly simplistic model structures. To address these issues, this study proposes the feature refinement convolutional neural network-fusion matrix factorization (FRCNN-F) model. Firstly, the study integrates the feature generation module of convolutional neural networks into the feature refinement network (FRNet), leveraging its ability to generate new features by recombining local patterns to enhance important feature selection. Secondly, the study designs the fusion matrix factorization mechanism to enable the model to perceive context and model displays through interactions across different scenarios, thereby enhancing the combination of submodels. Finally, through comparative experiments on the publicly available datasets Frappe and MovieLens, the results demonstrate that the FRCNN-F model outperforms the baseline FRNet, with improvements of 0.32% and 0.40% in AUC scores and reductions of 1.50% and 1.11% in cross-entropy loss (*Logloss*) respectively. This research has practical applications in achieving precise advertising and personalized recommendations.

**Key words:** click-through rate prediction; feature interaction; feature refinement network; convolutional neural network (CNN); field-matrixed factorization machine

① 基金项目: 国家自然科学基金 (61976006)

收稿时间: 2023-07-03; 修改时间: 2023-08-08; 采用时间: 2023-08-11; csa 在线出版时间: 2023-11-24

CNKI 网络首发时间: 2023-11-28

点击率 (click-through rate, CTR)<sup>[1]</sup>是指在当前显示次数下内容被点击的占比,旨在估计用户点击项目的概率。CTR 预测在广告领域和推荐系统中得到广泛应用,其核心目标是准确预测用户点击行为,以优化广告投放和内容推荐策略。在众多应用领域中,CTR 预测的价值显著。在广告优化和投放方面,广告主可以根据 CTR 预测结果调整广告创意、定向和投放时间,从而提高广告的点击率和转化率。广告排名和展示上,CTR 预测决定了广告在搜索结果页面、网站页面或移动应用中的排名顺序,优先展示可能带来更高点击率的广告。个性化推荐方面,CTR 预测根据用户历史点击行为,为用户提供个性化的广告推荐,提升用户体验等。因此,准确的 CTR 预测不仅能提升用户体验,还能带来巨大的商业价值,近年来备受关注。特征交互建模在近年来取得了巨大进展。根据最新发展情况,它们被简单地分为两种 CTR 预测方法:(1)用于捕获低阶特征交互方法,如因子分解机 (factorization machine, FM)<sup>[2]</sup>的方法;(2)用于捕获高阶特征交互的方法,如广度与深度神经模型 (Wide&Deep)<sup>[3]</sup>、深度因子分解机模型 (DeepFM)<sup>[4]</sup>和自适应交互模型 (AutoInt)<sup>[5]</sup>等。

Renderle<sup>[2]</sup>提出的 FM 模型通过内积建模二阶交叉特征,而 NFM<sup>[6]</sup>模型使用双线性交互层通过哈达玛积进行特征交叉,并结合 DNN<sup>[7]</sup>网络捕捉高阶非线性特征交叉。深度压缩交互网络 (xDeepFM)<sup>[8]</sup>模型通过显式和隐式组合的方式在特征字段上生成特征交互。基于卷积神经网络的点击率预估模型增强了特征空间中的学习交互,并与其他模型进行融合<sup>[9]</sup>,具备自动识别能力。

尽管现有模型结构日趋精细,性能也有所提升,但仍存在一个遗留问题:大多数模型仅关注每个特征的固定表示,忽视了在不同上下文中每个特征的不同重要性。例如,在{男性,网吧,打游戏,早上}和{男性,网吧,打游戏,晚上}这两种情况下,对用户进行预测时,应考虑某些特征在不同情境下的影响差异。因此,解决当前问题需要增加模型对特征细化的灵活性,更好地捕捉每个特征在不同上下文中的不同重要性。

自注意力机制<sup>[10]</sup>是 CTR 预测中常用且理想的方法之一,它通过学习特征之间的交叉特征重要性权重得分来捕捉交叉特征关系。然而,自注意力机制仅通过对权重归一化来捕捉同一实例中特征的相对重要性,忽略了不同实例之间特征重要性的差异。例如,在{男性,网吧,打游戏,早上}和{男性,网吧,打游戏,晚上}

的情况下,自注意力机制只能学习到“男性”特征相似的表示,而相较于特征“早上”和“晚上”与“网吧”和“打游戏”相比,“男性”的权重得分可能会非常小。然而,在这种情况下,“男性”用户的行为仍然可能在早上和晚上发生显著变化。因此,一个理想的特征细化卷积神经网络模块应该能够识别重要的跨实例上下文信息,并在不同的上下文中学习到显著不同的表示。

为了进一步研究上述问题<sup>[2,10-12]</sup>,本研究设计了一种名为 FRCNN-F 的 CTR 预测模型,这项研究创新性地当场矩阵分解机和卷积神经网络相结合,不仅旨在学习每个特征的固定表示,还为了进一步学习在不同上下文的重要性,该模型充分利用了卷积神经网络在重新生成特征方面的优势,以增强模型对特征重要性的选择能力。同时,引入了场矩阵交互项模块,进一步提升了模型对特征的提取能力。通过这种综合方法,达到了提升 CTR 模型预测能力的目的,从而在广告投放、排名和展示等应用领域实现更精准的预测。如图 1 所示,FRCNN-F 由两个重要部分组成:(1)特征细化卷积神经网络 (feature refinement convolutional neural network, FRCNN) 模块:该模块结合了卷积神经网络和自注意力机制。通过将局部特征重新生成新特征,并与上下文提取结果做哈达玛积,从而能够更好地实现上下文感知表征学习,以捕捉重要性特征的重组。(2)场矩阵分解机 (field-matrixed factorization machine, FmFM) 交互项模块:该模块引入了共享嵌入矩阵,以支持不同特征采用不同的嵌入向量维度。通过将 FmFM 的交互项作为单独的模块,并利用共享嵌入矩阵来学习不同特征场之间的重要信息,从而增加了数据输入的灵活性。由于 FmFM 交互项具有矩阵的自由度和可训练性,相比其他因子分解机,它具有更强的预测能力<sup>[12]</sup>。如图 1 所示,FRCNN-F 与现有 CTR 预测方法正交,因此可以方便地应用于许多现有方法中,以提高其性能。

本文设计 FRCNN-F 模型时,主要受到两个方面的启发。一方面,受场感知因子机 (FFM) 和场感知线性特征交互机 (FwFM) 的启发,通过嵌入向量进行二阶交叉,以捕捉特征之间的低阶交互信息。为了充分利用特征之间的交互关系,引入了场和特征权重的概念,以建模特征之间的差异性和重要性。然而,这种方法需要仔细设计和选择场,增加了特征工程的复杂性,并且处理稀疏特征会增加模型的计算和内存开销。因此,在本文

中, 尝试将 FmFM 的交互项单独作为一个模块融合到特征细化卷积神经网络模块中, 并通过后续实验验证

了该设计的有效性. 在去除 FmFM 后, FRCNN-F 的性能明显低于原始模型.

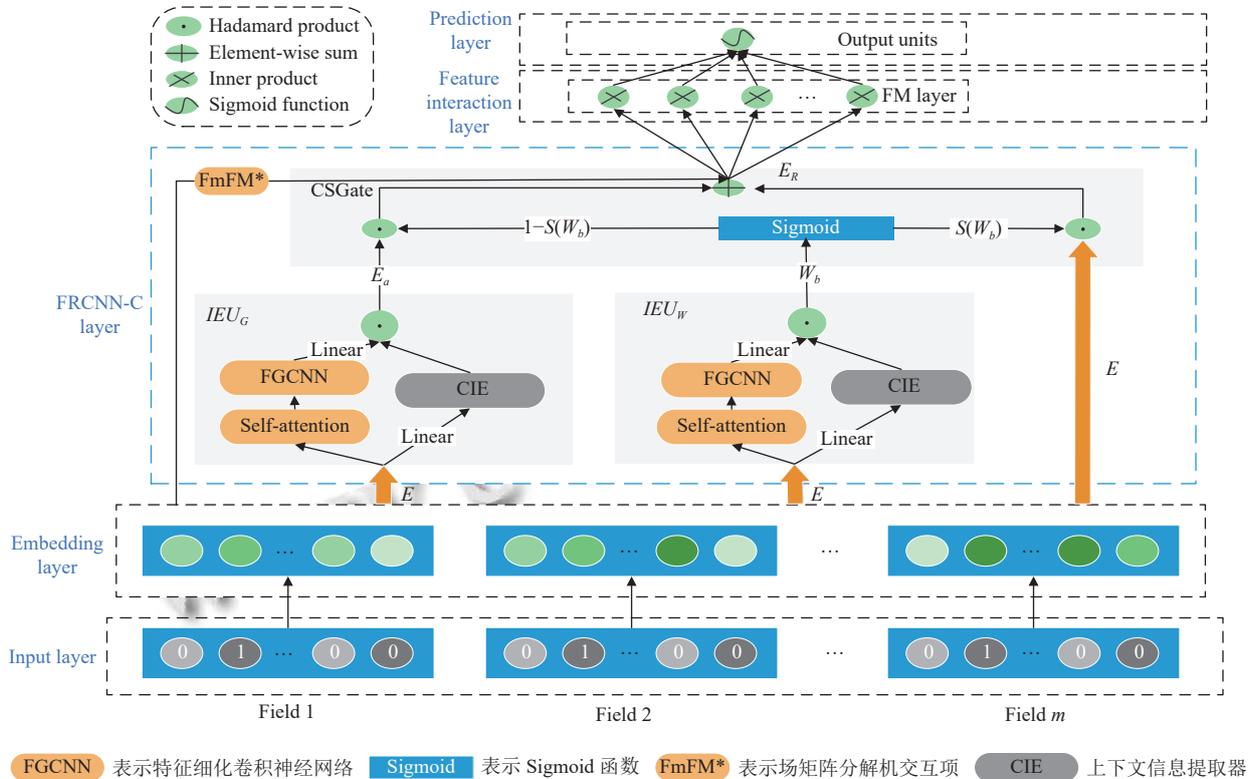


图1 基于场矩阵分解机和卷积神经网络的CTR预测模型

另一方面, 受到 Wide&Deep 和 DeepFM 的双塔结构的启发, 本文设计了一种“深度模型+浅层模型+浅层模型”的结构. 该结构将嵌入层的 Embedding 向量分别输入到两个不同的子模型中. 一个子模型在 FRCNN 中利用局部模式重新组合并生成新特征, 进行高阶特征交叉探索; 另一个子模型通过 FmFM 的交叉项进行低阶特征交叉探索. 最后, 将这两个结果拼接到一起作为 FM 的输入, 再进行低阶特征的探索, 并提供给最终的 CTR 预测. 通过后续实验, 本文还对各种不同的融合结构进行了对比实验, 验证了该设计的有效性. 接下来, 在本文方法中将按照由底向上的顺序介绍 FRCNN-F 模型各组成部分.

## 1 相关工作

许多 CTR 预测方法通过特征交互建模进行特征表示学习, 取得了巨大成功. CTR 预测方法可分为两类: 旨在捕获低阶特征交互方法<sup>[2]</sup>和旨在捕获高阶特征交互方法<sup>[3-10]</sup>. FM<sup>[2]</sup>是一种广泛使用的传统 CTR 预测

方法, 已有很多基于它的相关工作<sup>[2,13-26]</sup>.

然而, 这些方法无法捕捉高阶特征交互. 为解决此问题, 提出了捕获高阶特征的多种方法. Wide&Deep<sup>[3]</sup>联合训练广度线性单元和多层感知 (MLP), 结合记忆和泛化的优势. DeepFM<sup>[4]</sup>将 Wide&Deep<sup>[3]</sup>的大部分替换为 FM, 减轻特征工程的人工特征工程. 在 DeepFM 的基础上, xDeepFM<sup>[8]</sup>设计了一种新的压缩交互网络 (CIN), 明确建模高阶特征交互. AutoInt<sup>[5]</sup>使用堆叠的多头自注意层对特征交互进行建模. 特征交互图神经网络 (Fi-GNN)<sup>[21]</sup>使用图结构来对特征交互进行建模. FmFM<sup>[12]</sup>使用共享嵌入矩阵来对特征交互进行建模. 除了对特征交互建模外, 交叉网络 (XcrossNet)<sup>[27]</sup>和自动离散化 (AutoDis)<sup>[28]</sup>模型还设计了各种结构来学习连续型特征的特征嵌入. 直观地说, 上述方法仅学习每个特征的固定表示, 未考虑每个特征在不同上下文中的不同重要性, 导致性能较差. 最近的一些 CTR 预测方法<sup>[29]</sup>试图学习特征表示的向量级权重, 以解决固定特征表示问题. 因子化模型 (IFM)<sup>[29]</sup>和双因子化模型

(DIFM)<sup>[30]</sup>分别提出因子估计网络 (FEN) 和双重 FEN, 通过学习不同特征表示的向量级权重来改进 FM. 类似地, 双线性特征交互网络 (FiBiNET)<sup>[31]</sup>使用挤压和激励网络 (SENET)<sup>[32]</sup>通过重新加权原始特征来提取信息特征. 然而, 仅将向量级权重分配给同一特征会导致同一特征的学习表示具有严格的线性关系. 然而, 仅产生线性变换来改进固定特征表示是不合理的, 因为它们不够灵活, 无法捕捉不同上下文下每个特征的不同重要性. 最近, EGate<sup>[33]</sup>为每个特征应用了一个独立的 MLP 来学习元素级权重. FGCNN 通过局部特征生成, CNN 用来学习邻居特征交互, 而 MLP 用于全局的特征交互. 尽管如此, 相同特征的表示仍然是固定的, 因为它只转换表示空间. FRNet<sup>[34]</sup>通过非线性上下文感知特征表示学习建模, 具有一定的先进性, 但并不完美.

本文的方法与现有的方法相关, 但最大的不同在于本文既保留了生成更灵活的非线性上下文感知特征表示的优势, 又结合了卷积神经网络特征生成和场矩阵分解机各自的优势, 以实现更好的全局特征交互和原始特征不同特征场间的交互.

## 2 本文方法

本文将详细介绍论文提出的 FRCNN-F 模型, 如图 1 所示. 将按照由底向上的顺序逐一介绍模型的组成, 包括: 输入层和嵌入层、FRCNN-F 层、特征交互层和预测层. 最后, 通过实验分析, 验证了 FRCNN-F 模型具有更好的性能.

### 2.1 输入层和嵌入层

首先将用户所有的特征字段表示为稀疏向量, 这是所有字段的串联, 如式 (1) 所示:

$$X = [x_1; x_2; \dots; x_m] \quad (1)$$

其中,  $m$  是特征字段总数,  $x_i$  ( $1 \leq i \leq m$ ) 表示第  $i$  个特征字段.

然后, 将输入层高维稀疏的特征映射到低维稠密空间, 具体如式 (2) 所示:

$$e_i = V_i x_i \quad (2)$$

其中,  $V_i$  表示第  $i$  个字段嵌入矩阵.  $x_i$  为第  $i$  个特征字段表示, 如果  $x_i$  是类别特征, 则  $x_i$  为 one-hot 向量; 如果  $x_i$  为数值特征, 则  $x_i$  为标量值. 因此, 得到一个嵌入矩阵  $E \in \mathbb{R}^{m \times d}$ , 其中  $k$  表示嵌入维度. 然后, 将嵌入矩阵  $E$  送入到 FRCNN-F 网络中.

## 2.2 FRCNN-F 层

本文提出的 FRCNN-F 主要由两个重要部分组成.

(1) FRCNN 模块: 该模块通过信息提取单元 (IEU) 生成重要性特征, 并通过互补选择门 (CSGate) 自适应地将原始特征和互补特征进行元素级权重集成. 这样的集成能够更好地捕获上下文交叉特征, 以指导上下文感知特征细化的过程.

(2) 该模块主要通过不同特征场之间的相互交互来捕获不同场特征之间的重要性. 通过使用 FmFM 交互项模块, 模型能够有效地对特征场进行建模, 从而提高对特征交叉的建模能力.

### 2.2.1 FRCNN 模块

为了进一步研究特征细化网络特征提取的思想, 本文设计了 FRCNN 模块, 该模块主要由 IEU 和 CSGate 两部分组成. IEU 由 3 个基本组成部分构成: 第一, 特征生成单元: 在自注意力卷积下, 该单元捕获明确的特征交叉关系, 用于识别和生成新的重要特征. 通过自注意力机制, 模型能够聚焦于重要的特征关系, 从而提取更具表征力的特征. 第二, 上下文信息提取器 (CIE): 该组件旨在对不同的上下文信息进行编码. 它能够捕捉到数据中的上下文依赖关系, 并将这些信息整合到特征提取过程中. 通过 CIE, 模型能够更好地利用上下文信息来增强特征的表示能力. 第三, 集成单元: 该单元将自注意力特征重组生成的卷积单元与输出的上下文信息单元做哈达玛积. 通过将两者进行元素级的加权集成, 模型能够更好地融合特征生成和上下文信息, 以获取更丰富的特征表示. 本文使用两个 IEU 单元的目的是更好地生成互补的特征表示. 通过多个 IEU 单元的组合, 可以更充分地提取重要的上下文信息, 并进一步优化特征细化过程.

综上所述, IEU 单元主要通过特征生成单元、上下文信息提取器 (CIE) 和集成单元等模块的组合来实现特征的细化和上下文信息的提取. 这些模块的协同作用有助于提高特征表示的能力和模型的性能.

特征生成单元: 首先通过计算相关的特征加权和来生成新的表示. 自注意单元为:

$$Att(Q, K, V) = \text{Softmax}(QK^T)V \quad (3)$$

$$Q = W_i^{(Q)}E, K = W_i^{(K)}E, V = W_i^{(V)}E \quad (4)$$

其中,  $Q, K, V$  是输入矩阵  $E$  映射的 3 个不同矩阵:  $W_i^{(Q)}$ ,

$W_i^{(K)}, W_i^{(V)} \in \mathbb{R}^{d_k \times d}$  是转换矩阵, 将矩阵  $Q$  和矩阵  $K$  进行点积, 然后再进行 *Softmax* 来获得值  $V$  上的关注矩阵. 最后, 本文将输出矩阵维数变换为与输入相同的维数. 为了实现从重要性特征中生成新的重要特征这一

目标, 卷积特征生成模块设计了一个合适的神经网络结构来识别有用的特征交互, 然后自动生成新的重要特征. 特征生成的 CNN 和重组结构, 即卷积层、池化层和重组层, 如图 2 所示.

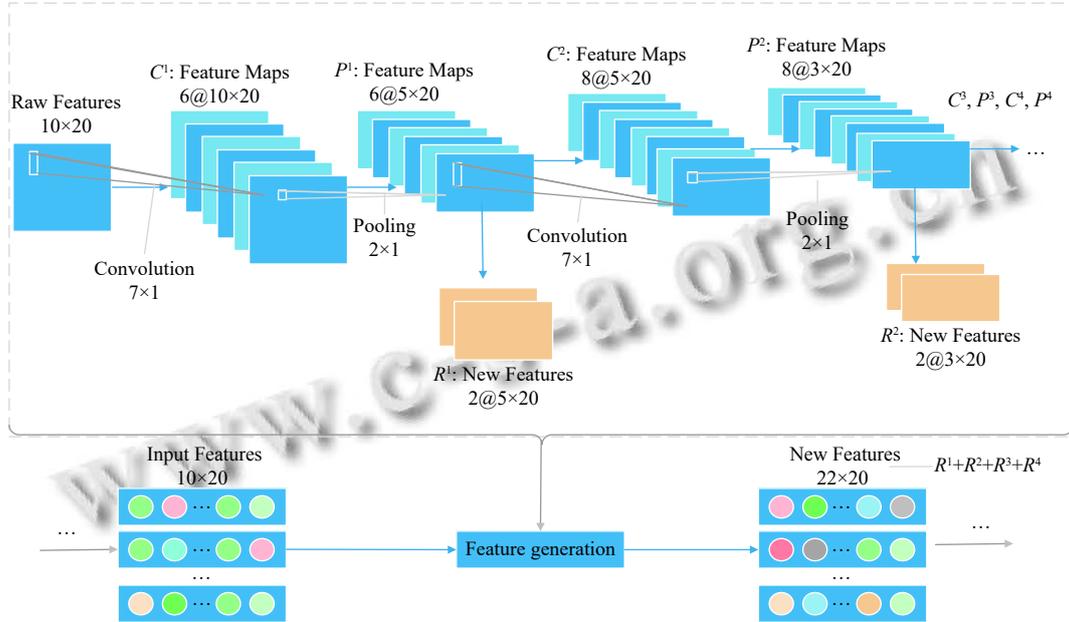


图 2 卷积神经网络特征生成模块 (图中的超参数是 Frappe 数据集上 FGCNN 的最佳设置)

FRCNN 卷积层: 卷积特征输入的嵌入矩阵为  $E \in \mathbb{R}^{f \times d}$ , 其中  $f$  是字段数,  $d$  是嵌入大小. 为了方便, 将嵌入矩阵重塑为  $E^1 \in \mathbb{R}^{f \times d \times 1}$  作为第 1 个卷积层的输入矩阵. 为了捕捉相邻特征的相互作用, 通过将权重矩阵  $W^1 \in \mathbb{R}^{h^1 \times 1 \times 1 \times m_c^1}$  与非线性激活函数卷积来获取第 1 个卷积层输出 (其中  $h^1$  是第 1 卷积层权重矩阵的高度,  $m_c^1$  是第 1 卷积层中特征图的数量). 假设第 1 个卷积层的输出表示为  $C^1 \in \mathbb{R}^{f \times d \times m_c^1}$ , 本文可以将卷积层公式化如下:

$$C_{a,b,i}^1 = \tanh \left( \sum_{m=1}^1 \sum_{j=1}^{h^1} E_{a+j-1,b,m}^1 W_{j,1,1,i}^1 \right) \quad (5)$$

其中,  $C_{a,b,i}^1$  表示第 1 卷积层中的第  $i$  个特征图,  $a$ 、 $b$  是第  $i$  个特征图的行和列索引,  $m$  表示卷积层中特征图数量,  $j$  表示卷积层权重矩阵的高度. 特别提示的是, 式 (5) 不包括在实际执行中的填充.

FRCNN 池化层: 在第 1 个卷积层后, 应用最大池化来捕获最重要的特征交互并减少参数的数量. 池化层的高度为  $h_a$  (宽度为 1). 第 1 个池化层输出为:

$$P_{a,b,i}^1 = \max(C_{a,h_a,b,i}^1, \dots, C_{a,h_a+h_b-1,b,i}^1) \quad (6)$$

第  $i$  个池化层结果将是第  $i+1$  个卷积层的输入:  $E^{i+1} = P^i$ .  $P^i \in \mathbb{R}^{(f/h_a^i) \times d \times m_c^i}$  包含相邻特征的模式. 为了保证后面全局非邻居特征交互不被忽略, 本文设计了一个完全连接层来重组局部特征, 并生成新的重要特征.

FRCNN 重组层: 在第  $i$  次卷积层和池化层之后, 本文通过一个全连接层来重组局部相邻特征并生成重要新特征, 将权重矩阵表示为  $W_r^i \in \mathbb{R}^{(f/h_a^i m_r^i) \times (f/h_a^i m_r^i)}$  和偏置  $B_r^i \in \mathbb{R}^{(f/h_a^i m_r^i)}$ , 其中  $m_c^i$  是第  $i$  卷积层特征图数量,  $m_r^i$  是第  $i$  重组层新的特征图数量. 因此, 在第  $i$  个重组层中, 产生了  $f/h_a^i m_r^i$  个特征.

$$R^i = \tanh(P^i \cdot W_r^i + B_r^i) \quad (7)$$

FRCNN 连接层: 可以通过执行多次 CNN 和重组结构来生成新的特征. 假设存在  $n_c$  个卷积层、池化层和重组层, 并且特征场  $N_i = f/h_a^i m_r^i$  由表示为  $R^i$  的第  $i$  轮生成. 通过特征生成的总体新特征  $\mathfrak{X} \in \mathbb{R}^{N \times d}$  ( $N$  为所有特征场  $N_i$  的总和) 被形式化为:

$$\mathfrak{X} = (R^1, R^2, \dots, R^{n_c}) \quad (8)$$

然后, 将总体的新特征进行连接后通过 Linear 函数处理, 与上下文信息提取器的结果做哈达玛积运算。

上下文信息提取器 (CIE): 该部分需要提取每个实例中的所有隐式包含的上下文特征. 因为需要确保所有特征都由于上下文信息一般是相对简单的, 所以选择 MLP 提取上下文信息是一个简单而有效的选择. 具体步骤如图 3 所示. 本文首先将原始特征通过 Linear 函数展平表示得到  $E_{con}$  作为 MLP 的输入。

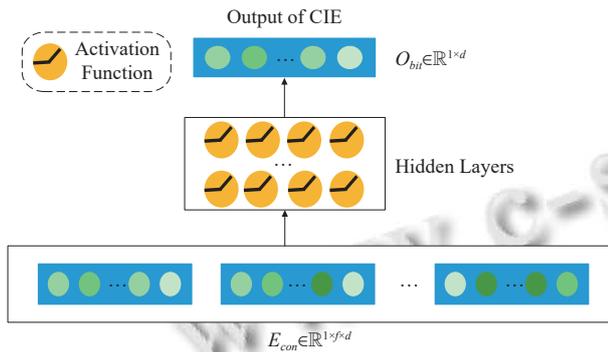


图 3 CIE 结构图

然后, 隐藏层计算如式 (9) 所示:

$$L_i = \text{ReLU}(W_{i-1}L_{i-1} + b_{i-1}) \quad (9)$$

其中,  $L_i$  和  $L_{i-1}$  分别是第  $i$  和第  $i-1$  隐藏层,  $W_{i-1}$  是第  $i-1$  层的权重矩阵,  $b_{i-1}$  是第  $i-1$  层可学习参数. 通过  $\text{ReLU}$  作为激活函数, 最后一层隐藏层的维度就是嵌入维度大小  $d$ . 最后一层输出上下文信息向量  $O_{bit} \in \mathbb{R}^{1 \times d}$ , 它可以表示特定实例中的所有上下文信息, 压缩  $E_{con}$  中的所有信息. 因为不同实例包含的特征是不同的, 所以实例对应的上下文信息特征是唯一的。

集成单元: 在获得上下文信息  $O_{bit}$  后, 本文直接使用  $O_{bit}$  来加权特征表示  $O_{vec}$ . 具体的计算如式 (10) 所示:

$$O_{IEU} = O_{vec} \odot O_{bit} \quad (10)$$

其中,  $\odot$  表示哈达玛积, 也就是元素对应相乘. 通过自注意力特征重组卷积单元重要性权重去捕获跨特征关系, 让  $O_{bit}$  每个特征都能感知上下文信息. 式 (10) 确保每个特征在不同情况下可以具有显著不同的表示。

如图 1 所示, 本文划分了两个独立的  $IEU$ :  $IEU_G$  生成互补特征;  $IEU_W$  学习元素级权重. 得到的结果  $E_a$ ,  $W_b \in \mathbb{R}^{f \times d}$ , 本文将在后面的内容介绍关于它们的细节。

在互补选择门中, 为更好地利用元素级权重从原始和互补特征中选择重要特征信息, 本文设计融合这

种新的门机制来控制信息流.  $E_a$  (互补特征表示)、 $W_b$  (权重矩阵)、 $E$  (原始特征表示) 进行 3 个通道输入 CSGate. 从 CSGate 输出的是上下文感知特征表示:

$$E_R = \underbrace{E \odot S(W_b)}_{\text{重要特征}} + \underbrace{E_a \odot (1 - S(W_b))}_{\text{互补特征}} \quad (11)$$

其中,  $S$  函数表示 Sigmoid 函数, 用于将结果限制在  $[0, 1]$  范围之间. 式 (11) 包含两部分特征: 重要特征和互补特征. 两个部分由选择门连接。

重要特征: 基于选择门结构, 通过该特征从元素级原始特征中选择重要特征信息. 具体来说, 通过  $S(W_b)$  元素的概率在  $[0, 1]$  之间来衡量特征重要性. 因此, 可以很好地学习非线性上下文感知特征表示. 与之前大多模型相比, 学习元素级权重矩阵  $S(W_b)$  具有两个优点: 第一, 设计通过特征重组卷积单元和元素级权重矩阵可以比先前的线性变换更灵活、更细致. 第二, 它同时包含跨特征关系和上下文信息, 这使得能够更好地学习到相同特征下不同特征实例的上下文感知表示。

互补特征: 基于选择门结构, 通过该特征进一步增强上下文特征的互补信息的表达能力. 现有的大多方法没有去考虑未选择的信息, 只考虑了原始特征的权重分配. 然而, 本文认为, 当前未选择的特征仍可能以不同的方式对 CTR 预测有帮助, 所以本文利用  $1 - S(W_b)$  来表示上下文特征的另一部分. 特别是利用门机制来实现重要特征和互补特征的自适应平衡。

总而言之, FRCNN 主要分为 3 个部分来生成上下文感知特征表示: 首先利用  $IEU_G$  生成互补特征. 其次利用  $IEU_W$  学习元素级权重. 最后利用 CSGate 集成原始特征和互补特征表示的上下文感知特征交互。

## 2.2.2 FmFM 交互项模块

围绕上述模型, 为更好地提取低阶非线性特征交互, 本文设计了一个新模型作为本文一个子模型, 将场与场的相互作用表示为矩阵. 为每个特征学习一个嵌入向量. 通过定义一个矩阵  $M_{F(i), F(j)}$  来表示场  $F(i)$  和场  $F(j)$  之间相互作用。

$$\mathfrak{J}_{FmFM}((w, v), x) = w_0 + \sum_{i=1}^m x_i w_i + \sum_{i=1}^m \sum_{j=i+1}^m x_i x_j \langle v_i M_{F(i), F(j)}, v_j \rangle \quad (12)$$

其中,  $v_i, v_j$  是特征  $i$  和特征  $j$  的嵌入向量。

FmFM 是 FwFM 的扩展, 因为它使用二维矩阵对

不同场对进行交互,而不是像 FwFM 中用标量权重. 利用这些矩阵,可以将嵌入空间中的特征转移到  $n-1$  空间; 本文将这些矩阵命名为场矩阵. 图 4 中演示了特征对  $\langle v_i, v_j \rangle$  和  $\langle v_i, v_k \rangle$  相互计算, 其中  $i, j, k$  来自 3 个不同场.

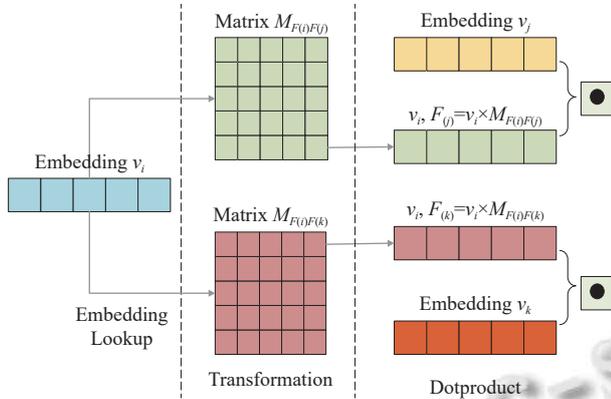


图 4 FmFM 交互项计算演示

模型计算大致可分为 3 步: 首先通过嵌入查找, 从嵌入表中找到 3 个特征嵌入向量  $v_i, v_j$  和  $v_k$ ; 然后通过转换, 分别将  $v_i$  和矩阵  $M_{F(i),F(j)}, M_{F(i),F(k)}$  计算, 获取两组对应的中间向量  $v_i, F(j)$  和  $v_i, F(k)$ ; 最后通过点积, 获得最后的交互项的结果.

总体来看, 场矩阵分解机的交互项在模型中的引入具有以下几个关键效果: 首先, 通过引入共享嵌入矩阵, 交互项增强了嵌入层的灵活性, 支持不同嵌入向量维度的输入. 其次, 场矩阵分解机交互项的引入允许模型实现特征之间的高阶交叉, 从而能够更好地捕捉特征之间的复杂关系. 此外, 这种引入还有助于模型学习更具区分度的上下文特征表示, 进一步提升了模型在特征学习方面的能力.

### 2.3 特征交互层与预测层

本文中在该层主要选择了 FM 作为本文特征交互模型. 在本文的基线模型中进行实验表明, 从模型性能和效能等综合因素出发, FM 相对来说是最优的选择. 如式 (13) 所示:

$$\mathcal{J}_{FM}((w, v), x) = w_0 + \sum_{i=1}^m x_i w_i + \sum_{i=1}^m \sum_{j=i+1}^m x_i x_j \langle v_i, v_j \rangle \quad (13)$$

其中,  $\langle v_i, v_j \rangle$  是  $k$  维向量  $v_i$  和  $v_j$  的点积.

在预测层中, 本文使用的是交叉熵损失函数, 具体的计算如式 (14).

$$\text{Logloss} = -\frac{1}{N} \sum_{j=1}^N (y_j \log(\hat{y}_j) + (1 - y_j) \log(1 - \hat{y}_j)) \quad (14)$$

其中,  $N$  表示送入网络训练的样本总数,  $y_j$  和  $\hat{y}_j$  分别表示用户的真实和估计点击率, 在训练过程中只要最小化损失即可.

## 3 实验

### 3.1 准备工作

为了有效进行实验比较, 实验部分的数据集使用 NFM<sup>[6]</sup> 和 AFM<sup>[14]</sup> 论文中提供的数据集: Frappe 和 MovieLens 数据集. 其中 Frappe 数据集是来自一个上下文感知的应用程序发现工具. 该数据集由 Baltrunas 等人创建, 包含了 96203 条用户在不同上下文情况下的应用程序使用日志. 实验使用的数据集共包括 668953 条用户对电影的打标签数据, 其中共有 23743 个电影项, 17045 个用户, 以及 49657 个不同的标签.

同样, MovieLens 数据集中的每条用户电影标签数据 (即用户 ID, 电影 ID 和标签) 都分别使用 one-hot 转换为特征向量, Frappe 数据集总共生产了 5382 个特征, MovieLens 数据集总共获得 90445 个特征. 表 1 总结了最终实验数据集的情况:

表 1 数据集统计信息

Dataset	#Train	#Valid	#Test	#Fields	#Features
Frappe	202027	57722	28860	10	5382
MovieLens	1404801	401372	200686	3	90445

本文采用 AUC (ROC 曲线下面积) 和 Logloss (二进制交叉熵损失) 作为评估指标. AUC 是随机给定一个阳性样本和一个阴性样本, 用一个分类器进行分类和预测, 来衡量阳性样本的得分高于阴性样本的得分的概率. AUC 越高, 性能越好. Logloss 主要是为了衡量预估点击率和实际点击率的拟合程度. Logloss 越小越好. 总的来说, AUC 和 Logloss 都很好地体现了模型的综合性能.

### 3.2 实验细节

用 PyTorch 实现了本文的所有方法, 所有模型保证相同的实验环境, 本文的实验环境如表 2 所列.

表 2 实验环境

实验环境	实验内容
操作系统	Ubuntu 19.1
编程语言	Python 3.8
GPU	NVIDIA RTX3080ti
内存	32 GB
开发环境	CUDA 11.4 + PyTorch 1.9

为了解决低配置机器内存问题, 引入 LMDB 缓存数据库, 减轻读写压力, 让低配置机器更好的运行. 同

时,本文所有模型都采用 Adam 优化器优化交叉熵损失来学习的.训练过程中,如果给定指标在 4 个连续时期内停止改进,则通过 ReduceLR-OnPlateau 调度器将学习率降低 10 倍.默认的学习率为 0.001.当验证集上的 AUC 停止改善时,本文使用提前停止来避免过拟合情况. BatchSize 大小设置为 4096. Frappe 和 MovieLens 的嵌入大小设置为 20. 根据先前的工作,本文对设计 MLP 的模型采用相同的结构(即 3 层, 400-400-400), 以进行公平比较. 除非有特别说明, 否则所有的激活函数均为 ReLU, 并将 Dropout 设置为 0.5. 在 FRCNN 中, CIE 中 MLP 的维度设置为 128. 对于其他模型设置, 本文采用原论文最佳设置. 为了确保实验公平可信, 本文通过改变随机种子来运行实验 5 次, 并报告平均结果. 在实验环节中将提出的 FRNet-FGCNN 应用于 FM. 在基线论文中通过比较发现应用于 FM 的方法比高阶或集成方法具有更少的参数、更快的训练速度以及相对较好的性能. 因此, 本文选择应用于 FM. 将提出的 FRCNN-F 与其他主流模型相比较, 包括 FM<sup>[2]</sup>、IFM<sup>[29]</sup>、DIFM<sup>[30]</sup>、NFM<sup>[6]</sup>、IPNN<sup>[15]</sup>、FINT<sup>[16]</sup>、FiBiNET<sup>[31]</sup>、DeepFM<sup>[4]</sup>、xDeepFM<sup>[8]</sup>、AutoInt+<sup>[5]</sup>、DESTINE<sup>[25]</sup>、DCN-V2<sup>[26]</sup>、AutoInt-APG<sup>[35]</sup>、FM<sub>FRNet</sub><sup>[34]</sup>, 同时也尝试各种不同的 FRNet 改进方案.

### 3.3 实验结果及分析

#### 3.3.1 模型性能比较

在本节中, 通过表 3 对 Frappe 和 MovieLens 测试集上各个模型的总体性能进行了总结. 从表 3 中可以观察到, FRNet 作为基线模型在性能表现上相当出众, 但由于其相对简单的模型结构, 难以适应复杂的数据. 然而, 通过实验可以看出, 本文提出的模型相比主流模型具有更好的效果. 这也在一定程度上说明, 引入卷积神经网络和场矩阵分解机交互项是一个正确的选择. 在对表 3 中各个模型进行比较后, 本文提出的 FRCNN-F 模型表现最佳. 在 Frappe 数据集下, FRCNN-F 模型的 AUC 值比 DESTINE、DCN-V2 和 AutoInt-APG 模型分别高 0.40%、0.38% 和 0.21%, FRCNN-F 模型的 AUC 值比 FRNet 高出 0.32%, 在 MovieLens 数据集下, FRCNN-F 模型的 AUC 值比 DESTINE、DCN-V2 和 AutoInt-APG 模型分别高 0.90%、0.77% 和 0.53%, FRCNN-F 模型的 AUC 值比 FRNet 高出 0.40%. 同时, Frappe 数据集下提出的 FRCNN-F 模型的 Logloss 值比 FRNet 低 1.50%, FRCNN-F 模型比 DESTINE、DCN-V2 模型的 Logloss 值分别低 2.09%、1.34%, 比

AutoInt-APG 稍微高 0.03%. 在 MovieLens 数据集下提出的 FRCNN-F 模型的 Logloss 值比 FRNet 低 1.11%, FRCNN-F 模型比 DESTINE、DCN-V2 和 AutoInt-APG 模型的 Logloss 值分别低 2.43%、2.21% 和 1.13%.

表 3 当前流行的 CTR 模型在 Frappe 和 MovieLens 真实数据集下的表现

Model name	Frappe		MovieLens	
	AUC	Logloss	AUC	Logloss
FM <sup>[2]</sup>	0.9708	0.1934	0.9391	0.2856
IFM <sup>[29]</sup>	0.9765	0.1896	0.9471	0.2853
DIFM <sup>[30]</sup>	0.9788	0.1860	0.9490	0.2459
NFM <sup>[6]</sup>	0.9746	0.1915	0.9437	0.2945
IPNN <sup>[15]</sup>	0.9791	0.1759	0.9490	0.2785
FINT <sup>[16]</sup>	0.9791	0.1921	0.9498	0.2674
FiBiNET <sup>[31]</sup>	0.9787	0.1867	0.9471	0.2630
DeepFM <sup>[4]</sup>	0.9789	0.1770	0.9465	0.3079
xDeepFM <sup>[8]</sup>	0.9792	0.1889	0.9480	0.2889
AutoInt+ <sup>[5]</sup>	0.9786	0.1890	0.9501	0.2813
DESTINE <sup>[25]</sup>	0.9790	0.1733	0.9608	0.2450
DCN-V2 <sup>[26]</sup>	0.9792	0.1658	0.9621	0.2428
AutoInt-APG <sup>[35]</sup>	0.9809	0.1521	0.9645	0.2320
FM <sub>FRNet</sub> <sup>[34]</sup>	0.9798	0.1674	0.9658	0.2318
FM <sub>FRCNN-F</sub>	<b>0.9830</b>	<b>0.1524</b>	<b>0.9698</b>	<b>0.2207</b>

#### 3.3.2 纵向比较各个改进模型的表现能力

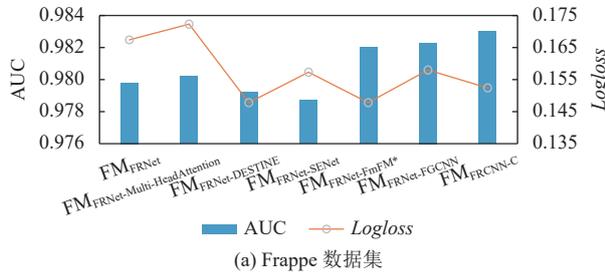
在本节中, 针对 FRNet 模型进行了不同的改进尝试, 并通过实验结果在表 4 和图 5 中进行了展示.

表 4 FM<sub>FRNet</sub> 模型的改进方案在 Frappe 和 MovieLens 真实数据集下的表现

Model Name	Frappe		MovieLens	
	AUC	Logloss	AUC	Logloss
FM <sub>FRNet</sub>	0.9798	0.1674	0.9658	0.2318
FM <sub>FRNet-Multi-Head Attention</sub>	<b>0.9802</b>	0.1723	<b>0.9679</b>	0.2458
FM <sub>FRNet-DESTINE</sub>	0.9792	<b>0.1478</b>	<b>0.9659</b>	0.2440
FM <sub>FRNet-SENet</sub>	0.9787	<b>0.1573</b>	0.9617	0.2333
FM <sub>FRNet-FmFM*</sub>	<b>0.9820</b>	<b>0.1478</b>	<b>0.9693</b>	<b>0.2308</b>
FM <sub>FRNet-FGCNN</sub>	<b>0.9823</b>	<b>0.1580</b>	<b>0.9671</b>	<b>0.2311</b>
FM <sub>FRCNN-F</sub>	<b>0.9830</b>	<b>0.1524</b>	<b>0.9698</b>	<b>0.2207</b>

首先, 进行了融合多头注意力方法的改进, 将注意力机制改为多头注意力机制. 通过多个层次的多头自注意力网络, 按顺序组合建模, 捕获更完整的上下文信息. 实验结果显示, 在 Frappe 数据集上, 融入多头注意力网络模型的 AUC 得分提高了 0.04%, 但 Logloss 升高了 0.49%. 在 MovieLens 数据集上, 融入多头注意力网络模型的 AUC 得分提高了 0.21%, 但 Logloss 升高了 1.40%. 其次, 进行了融合解耦注意力网络的改进,

通过解耦注意力网络替换注意力机制,实现特征交互白化,更准确地学习特征重要性.实验结果显示,在 Frappe 数据集上,融入解耦注意力网络模型的 AUC 得



分降低了 0.06%, 但 *Logloss* 降低了 1.96%. 在 MovieLens 数据集上,融入解耦注意力网络模型的 AUC 得分提高了 0.01%, 但 *Logloss* 升高了 1.22%.

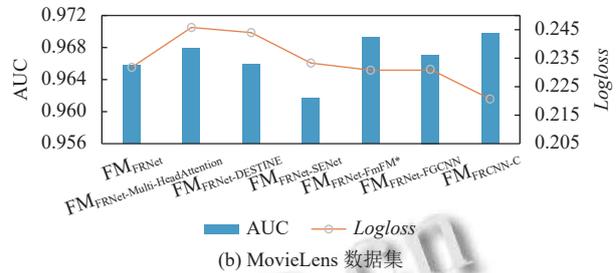


图5 改进的不同模型在两个真实数据集下的性能对比

另外,进行了融合挤压激励网络方法的改进,通过挤压激励网络来动态学习特征重要性.实验结果显示,在 Frappe 数据集上,融入挤压激励网络模型的 AUC 得分降低了 0.11%, 但 *Logloss* 降低了 1.01%. 在 MovieLens 数据集上,融入挤压激励网络模型的 AUC 得分降低了 0.41%, 且 *Logloss* 升高了 0.15%. 另外两个改进包括融合场矩阵分解机模型和卷积神经网络模型.融合场矩阵分解机模型的实验结果显示,在 Frappe 数据集上, AUC 得分升高了 0.22%, *Logloss* 降低了 1.96%; 在 MovieLens 数据集上, AUC 得分升高了 0.35%, *Logloss* 降低了 0.10%. 融合卷积神经网络模型的实验结果显示,在 Frappe 数据集上, AUC 得分升高了 0.25%, *Logloss* 降低了 0.94%; 在 MovieLens 数据集上, AUC 得分升高了 0.13%, *Logloss* 降低了 0.07%.

综合表 4 和图 5 的实验结果可以得出结论,本文提出的 FRCNN-F 模型在性能表现上最佳.

### 3.3.3 超参数实验

在本节中,针对 FRCNN-C 模型中的卷积特征生成模型部分,有两个关键的超参数需要进行分析,即卷积核的高度和用于特征生成新特征核的数量.通过在 Frappe 和 MovieLens 真实数据集上进行实验,具体实验数据如图 6 所示.

首先是卷积核的高度.卷积核的高度与卷积层的感知范围相关,高度越大涉及的特征就越多,但需要优化的参数也越多.实验结果显示,在 Frappe 数据集中,随着卷积核高度的增加,性能先上升后下降.尽管 AUC 得分有所提高,但 *Logloss* 也增加了.在 MovieLens 数据集中,随着卷积核高度的增加,性能先略有下降,然后再提升.实验表明,随着高度的增加,卷积核能涉及

到更多的特征,学习到更高阶的特征交互,从而提高性能.然而,由于有用的特征交互通常是稀疏的,较大的高度可能导致难以有效地学习它们,进而导致性能下降.

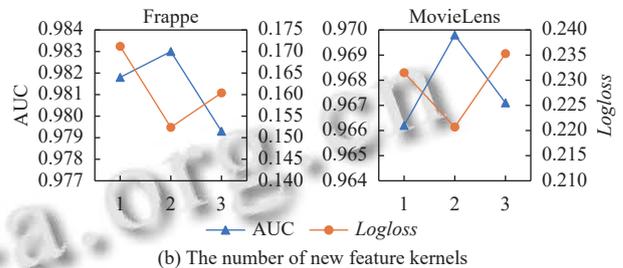
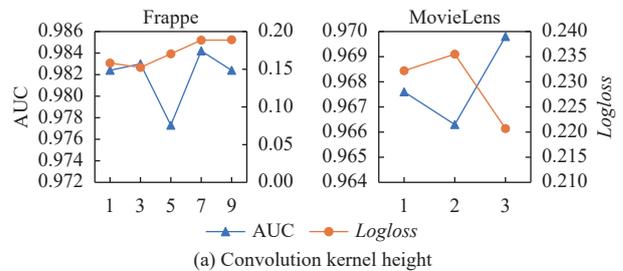


图6 卷积核的高度和新特征的核数的参数实验

其次是用于特征生成新特征核的数量.本文研究了生成特征的数量对 FRCNN-F 模型性能的影响.在不同的重组层中,使用相同数量的核来生成新特征.如图 6 所示,随着生成更多特征,性能先逐渐提升后降低.这表明有效的特征交互是有限的,当生成太多特征时,额外的新特征可能会带来噪声,增加学习的难度,导致性能下降.

通过实验结果的分析,可以得出结论,卷积核的高度和用于特征生成新特征核的数量对 FRCNN-F 模型的性能有一定的影响.在选择这些超参数时,需要平衡感知范围、参数数量和特征交互的稀疏性,以获得最佳的性能表现.

### 3.3.4 可解释性推荐

一个好的推荐系统不仅需要提供好的推荐结果,还需要提供很好的解释性.因此,在本文中我们以 Frappe 数据集为例,将介绍 FRCNN-F 如何解释推荐结果.

本文对数据中不同特征场之间相关性很感兴趣.因此,通过计算特征场在整个数据中的注意力得分来衡量特征场之间的相关性.图 7 展示了不同的特征之间的相关性热图.

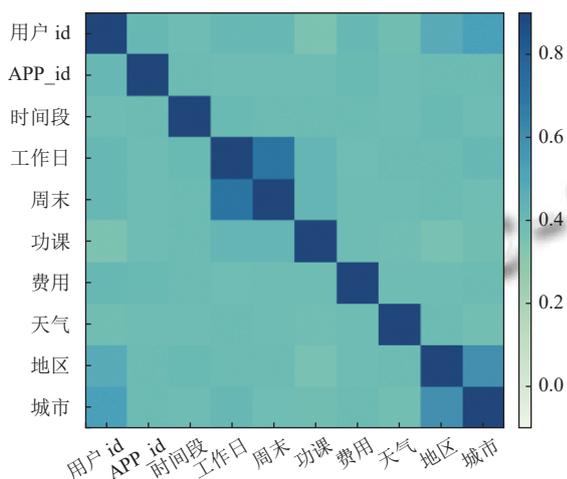


图 7 邻域特征注意力热图

从邻域特征注意力热图可以发现, <用户 id、APP\_id、时间段、功课、费用、天气>它们相互之间有着较弱的相关性.通过热图分析,它们主要和单个特征场自身有着强相关性.其次,通过整个数据中不同特征场的平均注意力得分,本文可以清楚地了解到捕捉低阶特征场和整体上下文信息是本文研究的重要方向.此外,还可以观察到<工作日、周末>、<地区、城市>存在着一定的联系,这体现了我们邻域推荐规则的可解释性.

### 3.3.5 消融实验

在本节中,本文在 Frappe 和 MovieLens 数据集上进行了实验,来证明 FRCNN-F 中每个组件都起着至关重要的作用.如表 5 所列,通过使用依次移除 FRCNN-F 中某一个组件来观察对模型的影响.为了更好地了解每个组件的相对重要性,本文将 FRCNN-F 设置成本文的基本模型(即 BASE),然后逐步移除组件.

**NO-FmFM\***: 移除了场矩阵分解机的交互项,通过自注意力卷积神经网络模块和上下文提取器的结果进行哈达玛积,并通过互补选择门自适应地集成原始特征和互补特征的元素级权重,以捕获上下文交叉特征.实验结果显示, Frappe 和 MovieLens 数据集中的 AUC

得分分别下降了 0.07% 和 0.27%, Logloss 上升了 0.56% 和 1.04%.

表 5 FRCNN-F 依次去除不同组件的性能

Model name	Frappe		MovieLens	
	AUC	Logloss	AUC	Logloss
BASE	0.9830	0.1524	0.9698	0.2207
NO-FmFM*	0.9823	0.1580	0.9671	0.2311
NO-FGCNN	0.9798	0.1674	0.9658	0.2318
NO-CSGate	0.9788	0.1852	0.9536	0.2653
NO-CIE	0.9776	0.1867	0.9522	0.2668
FM	0.9708	0.1934	0.9391	0.2856

**NO-FGCNN**: 移除了卷积神经网络的重要特征生成.随着进一步的移除,模型变为本文的基线模型: FRNet.实验结果显示, Frappe 和 MovieLens 数据集中的 AUC 得分分别下降了 0.25% 和 0.94%, Logloss 上升了 0.13% 和 0.07%.

**NO-CSGate**: 移除了互补选择门单元.随着进一步的移除,通过自注意力模块和上下文提取器的结果进行哈达玛积,仅从原始特征中提取上下文特征.实验结果显示, Frappe 和 MovieLens 数据集中的 AUC 得分分别下降了 0.10% 和 1.22%, Logloss 上升了 1.78% 和 3.35%.

**NO-CIE**: 移除了上下文提取器.随着进一步移除,仅通过自注意力模块对重要特征进行提取.通过实验显示,在 Frappe 和 MovieLens 数据集中, AUC 得分分别下降了 0.12% 和 0.14%, Logloss 上升了 0.15% 和 0.15%.

**FM**: 移除了整个 IEU 模块,只保留了 FM.可以观察到明显的下降,实验结果在两个数据集中显示了本文 IEU 模块的重要性.

综合表 5 实验分析可知,各个模块的组件对应本文模型性能都必不可少,说明本文方法的合理性.

## 4 结论与展望

本文提出了一种新颖的特征组合方式,利用注意力单元进行重要性特征提取,并结合卷积神经网络特征生成,以生成新的重要性特征.然后结合上下文提取单元和互补选择门单元,以更好地整合原始特征和上下文感知的互补特征表示的特征交互.同时,在特征细化的卷积神经网络模块中加入场矩阵分解机,以实现更好的场对之间的特征交互.本文提出的模型名为 FRCNN-F,优势在于在进一步细化上下文感知的同时,

保留了场对之间的特征交互,从而提升性能.该模型不仅具备 FRNet 算法提升上下文表征学习的特点,而且与其他子模型兼容性良好.实验结果表明,与基线模型相比,本文提出的新模型在两个真实数据集上性能始终更好,但在计算复杂度方面仍存在一些不足之处.未来的工作方向有两个.

1) 对于互补选择门与互补特征的连接方式,希望能够寻找更好的方法来提高解释性.

2) 在不降低模型性能的前提下,希望能够找到新的可靠子模型,同时尽可能降低计算复杂度.

### 参考文献

- 1 Zhang WN, Qin JR, Guo W, *et al.* Deep learning for click-through rate estimation. Proceedings of the 30th International Joint Conference on Artificial Intelligence. Montreal: IJCAI, 2021. 4695–4703.
- 2 Rendle S. Factorization machines. Proceedings of the 2010 IEEE International Conference on Data Mining. Sydney: IEEE, 2010. 995–1000.
- 3 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.
- 4 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.
- 5 Song WP, Shi CC, Xiao ZP, *et al.* AutoInt: Automatic feature interaction learning via self-attentive neural networks. Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing: ACM, 2019. 1161–1170.
- 6 He X, 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.
- 7 Covington P, Adams J, Sargin E. Deep neural networks for YouTube recommendations. Proceedings of the 10th ACM Conference on Recommender Systems. Boston: ACM, 2016. 191–198.
- 8 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.
- 9 刘梦娟, 曾贵川, 岳威, 等. 基于融合结构的在线广告点击率预测模型. 计算机学报, 2019, 42(7): 1570–1587.
- 10 Shaw P, Uszkoreit J, Vaswani A. Self-attention with relative position representations. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). New Orleans: ACL, 2018. 464–468.
- 11 Liu B, Tang RM, Chen YZ, *et al.* Feature generation by convolutional neural network for click-through rate prediction. Proceedings of the 2019 World Wide Web Conference. San Francisco: ACM, 2019. 1119–1129.
- 12 Sun Y, Pan JW, Zhang A, *et al.* FM2: Field-matrixed factorization machines for recommender systems. Proceedings of the 2021 Web Conference. Ljubljana: ACM, 2021. 2828–2837.
- 13 Pan JW, Xu J, Ruiz AL, *et al.* Field-weighted factorization machines for click-through rate prediction in display advertising. Proceedings of the 2018 World Wide Web Conference. Lyon: International World Wide Web Conferences Steering Committee, 2018. 1349–1357.
- 14 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: IJCAI, 2017. 3119–3125.
- 15 Qu YR, Cai H, Ren K, *et al.* Product-based neural networks for user response prediction. Proceedings of the 16th IEEE International Conference on Data Mining (ICDM). Barcelona: IEEE, 2016. 1149–1154.
- 16 Zhao ZS, Yang S, Liu GH, *et al.* FINT: Field-aware interaction neural network for CTR prediction. arXiv: 2107.01999, 2021.
- 17 Shan Y, Hoens TR, Jian J, *et al.* Deep crossing: Web-scale modeling without manually crafted combinatorial features. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016. 255–262.
- 18 Liu Q, Yu F, Wu S, *et al.* A convolutional click prediction model. Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. Melbourne: ACM, 2015. 1743–1746.
- 19 Juan Y, Zhuang Y, Chin WS, *et al.* Field-aware factorization machines for CTR prediction. Proceedings of the 10th ACM Conference on Recommender Systems. Boston: ACM, 2016.

- 43–50.
- 20 Li ZY, Cheng W, Chen Y, *et al.* Interpretable click-through rate prediction through hierarchical attention. Proceedings of the 13th International Conference on Web Search and Data Mining. Houston: ACM, 2020. 313–321.
- 21 熊中敏, 舒贵文, 郭怀宇. 融合用户偏好的图神经网络推荐模型. 计算机科学, 2022, 49(6): 165–171.
- 22 Li ZK, Cui ZY, Wu S, *et al.* Fi-GNN: Modeling feature interactions via graph neural networks for CTR prediction. Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing: ACM, 2019. 539–548.
- 23 Lang L, Zhu ZL, Liu XY, *et al.* Architecture and operation adaptive network for online recommendations. Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. ACM, 2021. 3139–3149.
- 24 Li F, Yan BC, Long QQ, *et al.* Explicit semantic cross feature learning via pre-trained graph neural networks for CTR prediction. Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2021. 2161–2165.
- 25 Xu YC, Zhu YQ, Yu F, *et al.* Disentangled self-attentive neural networks for click-through rate prediction. Proceedings of the 30th ACM International Conference on Information & Knowledge Management. ACM, 2021. 3553–3557.
- 26 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. Ljubljana: ACM, 2021. 1785–1797.
- 27 Leng JX, Liu Y, Wang ZH, *et al.* CrossNet: Detecting objects as crosses. IEEE Transactions on Multimedia, 2022, 24: 861–875. [doi: [10.1109/TMM.2021.3060278](https://doi.org/10.1109/TMM.2021.3060278)]
- 28 Guo HF, Chen B, Tang RM, *et al.* An embedding learning framework for numerical features in CTR prediction. Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. ACM, 2021. 2910–2918.
- 29 Yu YT, Wang Z, Yuan B. An input-aware factorization machine for sparse prediction. Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao: AAAI Press, 2019. 1466–1472.
- 30 Lu WT, Yu YT, Chang YZ, *et al.* A dual input-aware factorization machine for CTR prediction. Proceedings of the 29th International Joint Conference on Artificial Intelligence. Yokohama: IJCAI, 2020. 3139–3145.
- 31 Huang TW, Zhang ZQ, Zhang JL. FiBiNET: Combining feature importance and bilinear feature interaction for click-through rate prediction. Proceedings of the 13th ACM Conference on Recommender Systems. Copenhagen: ACM, 2019. 169–177.
- 32 Hu J, Shen L, Sun G, *et al.* Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2017. 7132–7141.
- 33 Huang T, She Q, Wang Z, *et al.* GateNet: Gating-enhanced deep network for click-through rate prediction. arXiv: 2007.03519, 2020.
- 34 Wang FY, Wang YX, Li DS, *et al.* Enhancing CTR prediction with context-aware feature representation learning. Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. Madrid: ACM, 2022. 343–352.
- 35 Yan BC, Wang PJ, Zhang K, *et al.* APG: Adaptive parameter generation network for click-through rate prediction. Proceedings of the 32nd International Conference on Neural Information Processing Systems. New Orleans: NeurIPS, 2022. 24740–24752.

(校对责编: 牛欣悦)