

DOI:10.16136/j.joel.2023.06.0375

# 点击率的深度交叉注意力预估模型

赵佰亭<sup>1</sup>, 梁润<sup>1</sup>, 贾晓芬<sup>1,2\*</sup>

(1. 安徽理工大学 电气与信息工程学院, 安徽 淮南 232001; 2. 安徽理工大学 省部共建深部煤矿采动响应与灾害防控国家重点实验室, 安徽 淮南 232001)

**摘要:**为了区分不同高阶交叉特征的重要程度并剔除冗余交叉特征,提高点击率的预估精度,提出了一种深度交叉注意力预估网络(deep cross attention prediction network, DAPN)模型。它将具有稀疏高维特征的输入信息表示为低维稠密向量后,分别送入因子分解机(factorization machine, FM)和深度交叉注意力层(deep cross attention, DCA)。FM通过一阶特征和二阶特征交叉挖掘训练数据中从未出现或者很少出现的低阶交叉特征。DCA层利用缩放点积注意力机制(dot-product attention,  $DP\_Att$ )设计交叉注意力层,用于区分高阶交叉特征的重要程度,并设计深度神经网络(deep neural network, DNN)实现对高阶交叉特征建模。仿真表明,DAPN在MovieLens-1m和Avazu等公开数据集上均具有良好的预估性能,它使用并行结构能同时有效地学习低阶和高阶交叉特征,提高预估精度。

**关键词:**点击率预估;注意力机制;深度神经网络(DNN);特征交叉

**中图分类号:**TP391 **文献标识码:**A **文章编号:**1005-0086(2023)06-0586-06

## Prediction model of deep cross attention based on click through rate

ZHAO Baiting<sup>1</sup>, LIANG Run<sup>1</sup>, JIA Xiaofen<sup>1,2\*</sup>

(1. Institute of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan, Anhui 232001, China; 2. State Key Laboratory of Mining Response and Disaster Prevention and Control in Deep Coal Mines, Anhui University of Science and Technology, Hainan, Anhui 232001, China)

**Abstract:**In order to distinguish the importance of different high-order crossover features and eliminate redundant cross features, and improve the accuracy of click through rate, a deep cross attention prediction network (DAPN) algorithm is proposed. DAPN represents the input information with sparse high-dimensional features as low-dimensional dense vectors and sends them to the factorization machine (FM) and deep cross attention (DCA) layer respectively. FM uses first-order features and second-order features to cross mine low-order cross features that never appear or rarely appear in training data. DCA layer uses scaled dot-product attention mechanism ( $DP\_Att$ ) to design cross attention layer, distinguish the importance of high-order cross features, and design the deep neural network (DNN) to model the high-order cross features. Simulation results show that DAPN has good prediction performance on public data sets, such as Movielens-1m and Avazu. DAPN uses parallel structure to effectively learn both low-order and high-order crossover features and improve estimation accuracy.

**Key words:**click through rate estimation; attention mechanism; deep neural network (DNN); feature crossover

## 0 引言

随着计算机技术的高速发展,互联网产生大

量信息,但也出现了“信息过载”<sup>[1]</sup>的问题。如何从海量信息中筛选出符合用户需求的信息是一个热点话题,个性化推荐<sup>[2]</sup>变得非常重要。点击率

\* E-mail:jxfzbt2008@163.com

收稿日期:2022-05-20 修订日期:2022-06-06

基金项目:国家自然科学基金面上项目(52174141)、安徽省自然科学基金面上项目(2108085ME158)、安徽高校协同创新项目(GXXT-2020-54)和安徽省重点研究与开发计划(202004a07020043)资助项目

预估是推荐系统的重要组成部分,目的是预测用户对给定的物品做出响应的概率,例如预测用户点击广告或购买商品的概率等。

早期解决点击率预估问题只是简单的使用线性模型对不同特征加权求和,每一个特征都是单独使用并没有考虑特征之间的相互影响,例如逻辑回归(logistic regression, LR)<sup>[3]</sup>、SGD(stochastic gradient descent, SGD)回归<sup>[4]</sup>等,LR 模型的优点是模型简单、解释性强,缺点是只能做简单的特征加权无法做特征交叉。TAO 等<sup>[5]</sup>使用因子分解机(factorization machine, FM)模型进行点击率预估,FM 考虑了二阶特征组合,模型的性能要优于线性模型。FM 模型对类别型特征使用 one-hot 编码处理,但对数值型特征的处理较差,只学习了二阶交叉特征,并没有考虑高阶特征交叉,损失了特征间的部分关联信息。

随着深度学习的快速发展,将深度学习应用到点击率预估领域是一个研究趋势。学者开始利用深度神经网络(deep neural network, DNN)对更具表现力的高阶交叉特征进行建模。如, KIM 等<sup>[6]</sup>使用具有记忆能力的 LR 模型和具有泛华能力的 DNN 模型相结合的 Wide&Deep 模型实现对输入数据的分析。华为诺亚方舟实验室的 GUO 等<sup>[7]</sup>将进行特征交叉的 FM 替换为 Wide&Deep 模型中的线性部分,提出 DeepFM(deep factorization machines, DeepFM)模型。之后为加强模型的特征交叉能力,能进行特征交叉并线性增加网络复杂度的 DCN(deep & cross network, DCN)模型<sup>[8]</sup>、保留 Cross 的优势并引入 Field vector-wise 思想的 xDeepFM(extreme deep factorization machines, xDeepFM)模型<sup>[9]</sup>等被相继提出。受 NLP(natural language processing, NLP)领域 Transformer<sup>[10]</sup>的启发,刘等<sup>[11]</sup>利用注意力机制实现 Bit-wise 级别的特征融合,利用 Attention 机制对组合特征进行赋权,消除了暴力组合方式带来冗余特征的影响。FENG 等<sup>[12]</sup>也直接在 DSIN(deep session interest network)中应用这种多头自注意力机制进行点击率预估。但是仅仅通过叠加注意力机制实现复杂的特征交叉可能限制模型的性能,只使用 DNN 对高阶特征进行隐式的特征交叉可能会引入冗余信息,冗余信息带来的噪声会增加模型训练的难度,降低在特征交叉方面的学习效率。

为了解决模型中存在的上述问题,实现点击率的高精度预估,本文基于 DeepFM 模型,提出了深度交叉注意力预估网络(deep cross attention prediction network, DAPN)。它在 Cross Network

架构中融入缩放点积注意力机制(dot-product attention, DP\_Att),形成交叉注意力层结合多层感知机设计一种深度交叉注意力(deep cross attention, DCA)结构,实现对高阶非线性交叉特征的有效学习。同时, DAPN 利用 FM 将低阶特征进行两两交叉建模,充分学习一阶和二阶交叉特征。DAPN 利用并行结构共享相同的输入,在减少手动设计特征工程成本的同时提高模型性能,得到更好的预估效果。

## 1 相关工作

### 1.1 缩放 DP\_Att

注意力机制可以使模型在预测的时候,针对不同特征分配权重参数,协助模型捕捉和预测任务相关的重要信息。缩放 DP\_Att 的高度优化并行矩阵乘法具有计算速度快、计算参数量少、存储空间小的优点。缩放 DP\_Att 能够区分特征交叉的重要程度,对模型特征交叉有重要意义。假设将向量键值对  $\langle k, v \rangle$  和编码向量  $q$  通过缩放 DP\_Att 映射为另一个向量,计算过程如式(1)所示:

$$DP\_Att(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (1)$$

式中,  $Q, K, V$  分别代表编码向量  $q$  和多个键值对  $\langle k, v \rangle$  组成的矩阵,  $d$  是向量的维度。 $Q$  和  $K$  做点积计算后再除以缩放因子  $\sqrt{d_k}$ , 经过 Softmax 函数后再和  $V$  相乘得到结果。

### 1.2 Cross Network

Cross Network 是 DCN 模型中的网络结构,用来高效地学习组合特征,DCN 模型结构如图 1 所示。Cross Network 由交叉层(cross layer)组成,交叉层的输出如式(2)所示。每个交叉层的输出

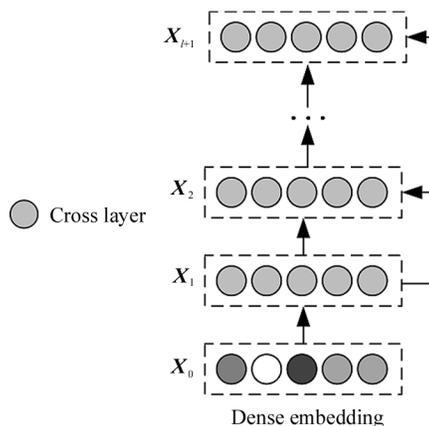


图 1 Cross Network 结构

Fig. 1 Cross Network structure

都是特征交叉函数  $f$  的结果与上一层输出结果的和。映射函数  $f$  的功能是拟合  $X_{l+1}$  和  $X_l$  的残差,如式(3)所示:

$$\mathbf{X}_{l+1} = \mathbf{X}_0 \mathbf{X}_l^T \mathbf{W}_l + \mathbf{b}_l + \mathbf{X}_l = f(\mathbf{X}_l, \mathbf{W}_l, \mathbf{b}_l) + \mathbf{X}_l, \quad (2)$$

$$f(\mathbf{X}_l, \mathbf{W}_l, \mathbf{b}_l) = \mathbf{X}_{l+1} - \mathbf{X}_l, \quad (3)$$

式中,  $\mathbf{X}_0 \in R^{dn}$  为 Cross Network 的输入,  $\mathbf{X}_{l+1}, \mathbf{X}_l \in R^{dn}$  分别代表第  $l+1, l$  层交叉层的输出,  $\mathbf{W}_l, \mathbf{b}_l \in R^{dn}$  分别为第  $l$  层的权重参数和偏差,  $n$  为所有输入特征

的数量,  $d$  为向量的维度。

## 2 DAPN 点击率预估模型

DAPN 模型结构如图 2 所示,包括输入层(input layer)、FM 层、DCA 层、输出层(output layer)。输入层负责输入数据处理、编码和降维;FM 层负责对一阶特征和二阶交叉特征进行建模;DCA 层负责对高阶交叉特征进行建模学习;输出层负责将经 FM 层得到的低阶特征信息和经 DCA 层得到的高阶特

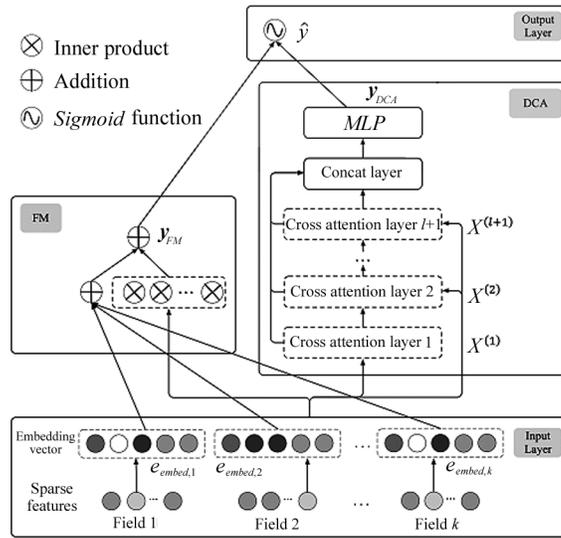


图 2 DAPN 模型结构

Fig. 2 DAPN model structure

征信息加和,最后通过 Sigmoid 函数得到输出结果。

### 2.1 输入层

输入层是模型的第一阶段,负责将输入的稀疏高维特征表示为低维稠密的 Embedding 向量,作为 FM 层和 DCA 层的共享输入,结构如图 2 Input Layer 部分所示。点击率预估模型中输入特征大多是稀疏的类别特征,通常处理方法是先将相同类型的特征划分为同一个域(field),再使用 one-hot 编码将类别型特征转换成高维稀疏的二进制特征(sparse feature)。为了降低特征维数采用 Embedding 编码将高维稀疏特征向量转化为低维稠密的实数 Embedding 向量。同一个域的所有特征向量共享同一个编码矩阵,其数学模型为,

$$\mathbf{e}_{embed,j} = \mathbf{W}_{embed,j} \mathbf{x}_j, \quad (4)$$

式中,  $\mathbf{e}_{embed,j} \in R^{dm}$  ( $j=1, 2, \dots, k$ ) 为第  $j$  个域所对应的 Embedding 向量,  $\mathbf{x}_j \in R^{vm}$  为第  $j$  个域中所有特征的 one-hot 向量,  $\mathbf{W}_{embed,j} \in R^{dv}$  为第  $j$  个域对应的 Embedding 编码矩阵,  $v$  为单词数量,  $d$  为向量维度,  $m$

为第  $j$  个域中特征向量的数量。输入层 Embedding 编码后的输出是所有域的 Embedding 向量的集合,如式(5)所示:

$$\mathbf{E} = [\mathbf{e}_{embed,1}, \mathbf{e}_{embed,2}, \dots, \mathbf{e}_{embed,k}], \quad (5)$$

式中,  $\mathbf{E} \in R^{dn}$  为所有域的 Embedding 向量的集合。

### 2.2 FM 层

FM 层是一个因子分解机,用于学习特征交叉,FM 层的输入是所有特征的 Embedding 向量。FM 除了使用特征间的线性关系作为一阶特征交叉,还利用两两特征向量内积作为二阶特征交叉,以便更好地学习训练数据中从未出现或者很少出现的特征交叉。为减少特征交叉的权重数量,利用矩阵分解将特征交叉权重矩阵分解成两个向量,FM 结构如图 2 中 FM 部分所示,其数学模型为:

$$\mathbf{y}_{FM} = \mathbf{w}_0 + \sum_{i=1}^n \mathbf{w}_i \mathbf{x}_{embed,i} + \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{V}_i, \mathbf{V}_j \rangle \mathbf{x}_{embed,i} \mathbf{x}_{embed,j}, \quad (6)$$

式中,  $w_0 \in R^d$  为模型的全局偏差,  $w_i \in R^d$  为第  $i$  个特征的权重,  $n$  为所有特征的数量,  $V_i, V_j$  为第  $i$  个特征和第  $j$  个特征的潜在特征向量,  $\langle a, b \rangle$  表示对向量  $a, b$  做内积操作,  $\langle V_i, V_j \rangle$  为第  $i$  个特征和第  $j$  个特征的交叉权重参数。

### 2.3 DCA 层

DCA 层结构如图 2 的 DCA 部分所示, 为了通过深化注意力网络探索更多的显性高阶特征交叉, 利用缩放 *DP\_Att* 点积注意力机制设计交叉注意力层。第  $l+1$  层交叉注意力层结构如图 3 所示, 计算过程如式(8)、(9)所示。第  $l$  层的输出作为第  $l+1$  层的输入经缩放 *DP\_Att* 得到特征向量  $Z^{(l+1)}$ , 再与输入层的 Embedding 向量做内积计算, 最后通过平均池化(average pooling)<sup>[13]</sup> 降维得到第  $l+1$  层的输出结果  $X^{(l+1)}$ 。在深度特征交叉之后使用连接层将所有交叉注意力层的输出拼接起来, 再输入到多层感知机中进一步探索隐式特征交叉。DCA 层的输入和 FM 的输入相同都是所有特征的 Embedding 向量, 假设 DCA 层具有  $l+1$  层, 交叉注意力网络的输出的数学模型为:

$$y_{DCA} = MLP(Concat(X^{(1)}, X^{(2)}, \dots, X^{(l+1)})), \quad (7)$$

$$Z^{(l+1)} = DP\_Att(X^{(l)}W^Q, X^{(l)}W^K, X^{(l)}W^V), \quad (8)$$

$$X^{(l+1)} = AvePooling(\sum_d (Z^{(l+1)} \otimes E)), \quad (9)$$

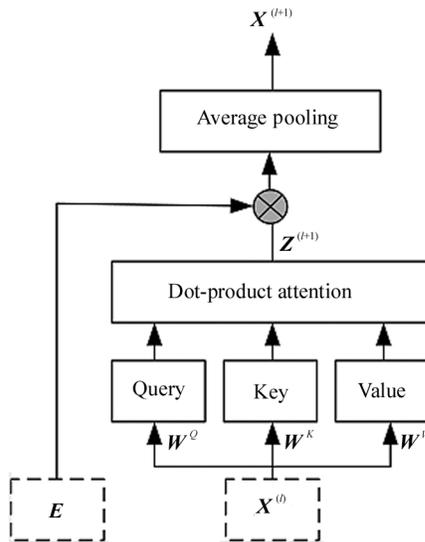


图 3 交叉注意力层

Fig. 3 Cross attention layer

式中,  $X^{(i)} \in R^{dn}$  ( $i=1, 2, \dots, l+1$ ) 为第  $i$  层交叉注意力层的输出,  $Z^{(i)} \in R^{dn}$  ( $i=1, 2, \dots, l+1$ ) 为第  $i$  层交叉注意力层中缩放 *DP\_Att* 的输出,  $E \in R^{dn}$  为输入 Embedding 特征向量,  $W^Q \in R^{d \cdot dq}$  为查询变换矩

阵,  $W^K \in R^{d \cdot dk}$  为关键字变换矩阵,  $W^V \in R^{d \cdot dv}$  为值变换矩阵, 设置  $dq=dk=dv$ 。

### 2.4 输出层

将经过 FM 层得到的低阶特征信息  $y_{FM}$  和经过 DCA 层输出的高阶组合特征信息  $y_{DCA}$  加和, 然后输入到 *Sigmoid* 函数计算输出结果, 模型输出层表达式如式(10)所示。*Sigmoid* 函数表达式如式(11)所示:

$$\hat{y} = Sigmoid(y_{FM} + y_{DCA}), \quad (10)$$

$$Sigmoid(Y) = \frac{1}{(1 + \exp(-Y))}, \quad (11)$$

式中,  $y_{FM}$  为 FM 层的输出,  $y_{DCA}$  为 DCA 层的输出。

### 2.5 模型损失函数

点击率预估问题是一个二分类问题, 为了更好地学习模型权重参数, 使用 *Logloss* 作为模型的损失函数。*Logloss* 又称为交叉熵损失函数, 用来衡量模型预测值和实际值之间的差距, *Logloss* 损失函数如下:

$$Logloss = -\frac{1}{N} \sum_{i=1}^N [y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log (1 - \hat{y}^{(i)})], \quad (12)$$

式中,  $\hat{y}$  表示模型预测用户点击概率,  $y$  表示用户真实点击结果为 0 或 1,  $N$  为训练样本总数。

### 2.6 空间复杂度

对于多层交叉注意力层, 假设  $l+1$  表示层数,  $d$  表示向量的维度, 交叉注意力层隐藏层神经元数量为  $c$ , 则  $l+1$  层交叉注意力层的结果是  $X^{(1)}, X^{(2)}, X^{(3)}, \dots, X^{(l+1)}$ , 交叉注意力层为单隐藏层全连接网络, 对应的参数量为  $dc(l+1)$ 。可知交叉注意力层网络空间复杂度是维度  $d$  的线性函数, 相比于模型中 DNN 部分, 交叉注意力层所引入的复杂度微不足道, 所以 DAPN 模型和 DeepFM 模型相比, 复杂度几乎不变。

## 3 实验结果与分析

### 3.1 实验设置

仿真实验在 Windows 10 操作系统下进行, 编程语言是 Python3。模型使用 Pytorch 深度学习框架搭建, 使用 3080Ti 显卡进行加速训练。使用 Adam<sup>[14]</sup> 作为模型优化器, 模型超参数设置为: batch\_size 为 4 096, 学习速率为  $1 \times 10^{-3}$ , embedding\_size 为 16, 训练次数为 100。

### 3.2 数据集和评价指标

采用 MovieLens-1m 和 Avazu 两个公共数据集进行实验。MovieLens-1m 数据集包含 6 000 个用户

对 4 000 部电影的 100 万条用户数据,Avazu 数据集包含 4 000 万条真实点击日志数据。将数据集随机打乱,按照 9 : 1 的比例划分为训练数据集和测试数据集。

使用 ROC 曲线下的面积 (area under ROC curve, *AUC*) 和 *Logloss* 作为模型的评价指标, *Logloss* 的值越小代表模型的精度越高, *AUC* 的值越大代表点击率预估模型的性能越好。

### 3.3 模型关键模块分析

DAPN 模型主要可分为两大部分:第一部分是包括二阶特征交叉和线性模型部分的 FM 层;第二部分是学习高阶交叉特征的 DCA 层, DCA 层又可分为多层交叉注意力部分和多层感知机 *MLP* 部分。为证明模型结构的有效性,开展消融实验,关键模块分析结果如表 1 所示。model1: 去掉 FM 层; model2: 去掉 DCA 层; model3: 去掉 Cross Attention Layer。

表 1 DAPN 关键组件分析  
Tab. 1 DAPN key component analysis

Model	MovieLens-1m		Avazu	
	<i>AUC</i>	<i>Logloss</i>	<i>AUC</i>	<i>Logloss</i>
DAPN	<b>0.8116</b>	<b>0.5209</b>	<b>0.7798</b>	<b>0.3741</b>
model1	0.8089	0.5224	0.7719	0.3783
model2	0.8040	0.5409	0.7700	0.3793
model3	0.8066	0.5321	0.7755	0.3742

由表 1 可知, DAPN 模型在去除不同结构以后, 模型性能都有所下降, 证明了 DCA 层和 FM 层对模型性能具有提升效果。

### 3.4 模型对比

选择 LR<sup>[3]</sup>、DCN<sup>[9]</sup>、DeepFM<sup>[7]</sup>、PNN<sup>[15]</sup>、DCAP<sup>[14]</sup> 等 5 种点击率预估模型作为基础线与模型 DAPN 进行对比。为保证实验结果的稳定性和可靠性, 最后的结果取 10 次训练结果的平均值。各模型在 MovieLens-1m 和 Avazu 数据集的 *AUC* 指标和 *Logloss* 损失如表 2 所示。

由表 2 可知, 只考虑特征之间的线性关系的 LR 模型相比其他模型表现最差, 说明特征之间是有相互影响关系的, 适当对特征进行交叉组合能提高模型性能。在两个公开数据集上, DAPN 模型的 *AUC* 指标分别比 LR 模型提高了 0.16% 和 0.25%, *Logloss* 指标提高了 0.62% 和 1.6%。

表 2 各模型在两个公开数据集上的表现

Tab. 2 Performance of each model on two public datasets

Model	MovieLens-1m		Avazu	
	<i>AUC</i>	<i>Logloss</i>	<i>AUC</i>	<i>Logloss</i>
LR <sup>[3]</sup>	0.7923	0.5399	0.7594	0.3793
DCN <sup>[9]</sup>	0.8041	0.5261	0.7734	0.3754
DeepFM <sup>[7]</sup>	0.8073	0.5323	0.7765	0.3783
PNN <sup>[15]</sup>	0.8102	0.5242	0.7779	0.3804
DCAP <sup>[14]</sup>	0.8082	0.5259	0.7789	0.3757
DAPN	<b>0.8116</b>	<b>0.5209</b>	<b>0.7798</b>	<b>0.3741</b>

由图 4 和图 5 可以直观地看出, 同时学习高阶特征和低阶特征能够提高模型预测性能。DCPN 同时考虑低阶特征组合和高阶特征组合, 与其他点击率预估模型相比, 在 MovieLens-1m 数据集和 Avazu 数据集上的各性能指标更好。与只学习高阶交叉特征的 PNN 模型和 DCAP 模型相比, 在 *AUC* 指标上分别提高了 0.16% 和 0.25%、0.42% 和 0.95%, 在 *Logloss* 指标上分别提高了 0.62% 和 1.6%、0.12% 和 0.43%。与相同结构同时学习低阶特征和高阶特

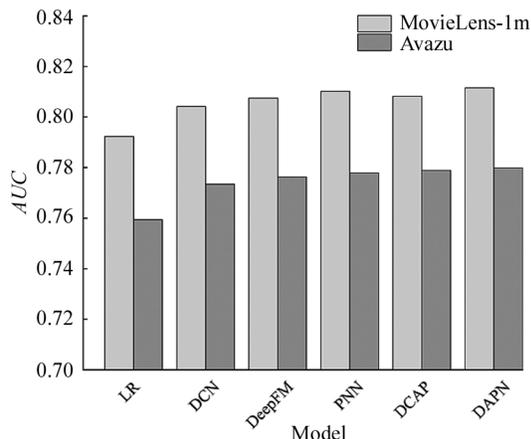


图 4 模型 *AUC* 指标对比

Fig. 4 Comparison for *AUC* index of models

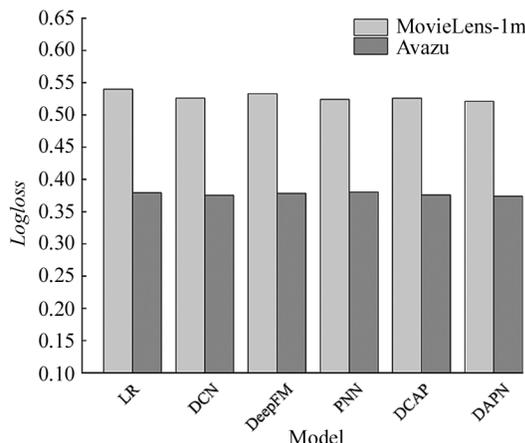


图 5 模型 *Logloss* 指标对比

Fig. 5 Comparison for *Logloss* index of models

征的 DCN 模型和 DeepFM 模型相比,在 AUC 指标上分别提高了 0.93%、0.53% 和 0.82%、0.47%,在 Logloss 指标上分别提高了 1.0%、2.1% 和 0.35%、1.1%。

## 4 结论

提出了一种 DAPN 模型。模型的核心部分是使用 DCA 层建模高阶交叉特征,将每一层的输出连接,保留所有的特征信息。利用 FM 对低阶特征信息进行学习,最后融合低阶交叉特征和高阶交叉特征两部分构建模型,进行点击率预估。通过减少不同模型关键组件进行消融实验,结果表明:同时利用低阶和高阶特征学习模型,能获取更好的性能和更高的点击率预估准确率。在 MovieLens-1m 和 Avazu 数据集上进行验证实验,结果表明:DAPN 模型的性能比其他点击率预估模型更加优异,预估效果更好。下一阶段工作将研究模型计算复杂度低的变体,研究新的特征工程对现有模型进行优化和改进,以提高点击率预估效果。

### 参考文献:

- [1] TANG H, FAN S, TANG F, et al. Recommendation algorithm combining knowledge graph and attention mechanism[J]. Computer Engineering and Applications, 2022, 58(5):94-110.  
唐宏, 范森, 唐帆, 等. 融合知识图谱与注意力机制的推荐算法[J]. 计算机工程与应用, 2022, 58(5):94-110.
- [2] WANG X, TANG J. Deep learning-based personalized paper recommender system[J]. Journal of Chinese Information Processing, 2018, 32(4):114-119.  
王妍, 唐杰. 基于深度学习的论文个性化推荐算法[J]. 中文信息学报, 2018, 32(4):114-119.
- [3] NUSINOVICI S, THAM Y C, YAN M Y C, et al. Logistic regression was as good as machine learning for predicting major chronic diseases[J]. Journal of Clinical Epidemiology, 2020, 122:56-69.
- [4] AMIRI M M, GVNDVZ D. Machine learning at the wireless edge: distributed stochastic gradient descent over-the-air [J]. IEEE Transactions on Signal Processing, 2020, 68: 2155-2169.
- [5] TAO Z, WANG X, HE X, et al. HoAFM: A high-order attentive factorization machine for CTR prediction[J]. Information Processing & Management, 2020, 57(6):102076.
- [6] KIM M, LEE S, KIM J. A wide & deep learning sharing input data for regression analysis[C]//2020 IEEE International Conference on Big Data and Smart Computing (BigComp), February 19-22, 2022, Busan, Korea (South). New York: IEEE, 2020: 8-12.
- [7] GUO H, TANG R, YE Y, et al. DeepFM: An end-to-end wide & deep learning framework for CTR prediction [EB/OL]. (2018-04-12) [2022-05-20]. <https://arxiv.org/abs/1804.04950>.
- [8] WANG R, FU B, FU G, et al. Deep & cross net-work for ad click predictions[C]//Proceedings of the ADKDD'17. August 14, 2017, Halifax NS, Canada. New York: Association for Computing Machinery, 2017:1-7.
- [9] LIAN J, ZHOU X, ZHANG F, et al. xDeepFM: Combining explicit and implicit feature interaction for recommender systems[C]//24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, July 19-23, 2018, London, United Kingdom. New York: Association for Computing Machinery, 2018:1754-1763.
- [10] JIA X, YANG S R, GUAN D Y. Semantics-aware transformer for 3D reconstruction from binocular images[J]. Optoelectronics Letters, 2022, 18(5):293-299.
- [11] LIU Y L, GUO M, MA M. CRNN with multi-scale attention feature fusion for sound event detection [J]. Journal of Optoelectronics • Laser, 2021, 32(12):1271-1277.  
刘亚灵, 郭敏, 马苗. 基于多尺度注意力特征融合 CRNN 的声音事件检测[J]. 光电子 • 激光, 2021, 32(12):1271-1277.
- [12] FENG Y, LV F, SHEN W, et al. Deep session interest network for click-through rate prediction [EB/OL]. (2019-05-16) [2022-05-20]. <https://arxiv.org/abs/1905.06482>.
- [13] WANG S H, KHAN M A, GOVINDARAJ V, et al. Deep rank-based average pooling network for COVID-19 recognition[J]. Computers, Materials & Continua, 2022, 70(2):2797-2813.
- [14] CHEN Z K, ZHONG F T, CHEN Z M, et al. DCAP: Deep cross attentional product network for user response prediction[C]//30th ACM International Conference on Information & Knowledge Management, November 1-5, 2021, Queensland, Virtual Event, Australia. New York: Association for Computing Machinery, 2021:221-230.
- [15] QU Y, FANG B, ZHANG W, et al. Product-based neural networks for user response prediction over multi-field categorical data [J]. ACM Transactions on Information Systems, 2018, 37(1):1-35.

### 作者简介:

贾晓芬 (1978—), 女, 博士, 教授, 主要从事深度学习, 图像处理方面的研究。