

文章编号:1671-6833(2020)06-0026-07

基于改进 Wide&Deep 交互特征提取的移动 APP 转化率预估

孙晓燕, 聂鑫, 暴琳, 陈杨

(中国矿业大学 信息与控制工程学院, 江苏 徐州 221008)

摘 要: 针对移动 APP 广告转化率预估特征维度高而稀疏、特征间高度交互使得转化率精准预估面临挑战等问题,提出一种融合场感知分解机(field-aware factorized machine, FFM)和深度卷积神经网络的改进 Wide&Deep 模型,以有效获取高维度稀疏特征的低阶和高阶交互关系,实现特征自动高效组合,提高移动 APP 广告转化率预估精度。在给出算法框架的基础上,针对稀疏数据的嵌入,提出了基于宽度模块 FFM 挖掘低阶特征交互关系的特征组合算法,然后,根据 FFM 所提取隐含特征向量,进一步给出了基于深度模块多层卷积神经网络提取高阶交互关系的特征提取策略,最后,将宽度和深度模块分别获取的特征组合用于转化率预估。所提算法在腾讯移动 APP 广告转化率预估中的应用表明了该方法在提高预估精度上的有效性。

关键词: 转化率预估; 特征交互; 场感知分解机; Wide&Deep; 移动 APP

中图分类号: TP391 **文献标志码:** A **doi:**10.13705/j.issn.1671-6833.2020.06.005

0 引言

在计算广告领域衡量广告投放效果时,通常通过曝光、点击和转化等环节来衡量。大多数广告系统受广告效果数据回流的限制,往往将曝光或点击率(click through rate, CTR)作为广告投放效果的衡量标准^[1-2]。而实质上,广告的转化数据(conversion rate, CVR)在衡量广告投放效果、广告费用结算和广告投放排序中有重要作用。因此,转化率预估(conversion rate prediction, CRP)已成为计算广告领域的重要研究内容之一^[3]。移动 APP 转化率预估问题具有维度高、量大、稀疏、特征间高度交互等特点^[4-5],如何使基于大量高维度异构的结构型数据自动获取特征间的交互关系,并对其进行有效融合和抽取,成为实现移动 APP 精准转化率预估的关键问题。

特征抽取主要分为两类,一是基于人工经验的手工特征提取策略,二是基于学习机制的特征自动获取方法^[6-7]。手工特征提取是一项繁杂的工作,需要专家熟悉相关的领域知识,而基于用户经验的手工特征提取,极难充分发掘特征之间的交互关系,存在大量有效特征组合无法被专家识

别的问题。此外,手工特征提取可能导致特征维度爆炸,如常用的 Poly-2 模型^[8-9],该模型采用直接对二阶特征组合建模来学习特征之间的内在交互关系,这种方式提取的特征数量与原数据所有特征数量乘积相关,易引起特征爆炸,且此类特征组合方式尚未考虑高阶特征交互关系。

基于学习机制的特征自动获取又可分为显式特征交互关系提取,如基于贪心算法和搜索算法的研究^[10-11];基于树模型的半显式^[12]特征交互关系提取;基于深度神经网络(deep neural networks, DNNs)的隐式特征交互关系提取^[13]。Bowers 等^[14]利用 GBDT 和逻辑回归的组合树模型实现了半显式特征交互关系提取和特征组合。针对个性化推荐中用户和物品等数据的稀疏特性,Rendle^[4, 15]提出了基于分解机(factorized machine, FM)稀疏数据的特征提取,实现一阶、二阶特征之间的组合,但无法发掘特征之间的高阶交互关系。针对此问题,Blondel 等^[16]提出了高阶分解机模型,以学习特征之间的高阶交互关系,但是该方法会引入无效特征组合。Xiao 等^[17]通过在 FM 中引入注意力机制,证明了无效特征交互关系对模型性能的影响。针对点击率预估问题,Juan 等^[5]在

FM 模型基础上,引入场感知 (field-aware) 属性,通过引入辅助向量,考虑不同领域特征之间的关系,提高了模型表现力。FM 和 FFM 算法在识别用户行为数据特征间低阶交互关系方面具有较好的性能,但是却难以深入挖掘出高阶交互关系,从而影响了更多有效特征组合的识别和基于上述特征的模型构建。

深度学习作为一种先进的非线性模型在特征组合挖掘方面具有很大的优势。深度卷积神经网络 (deep convolutional neural networks, DCNN) 具有较强的隐式提取及表示高阶特征的能力^[18-20]。Zhang 等^[21]提出了解决多域类别型数据的深度学习模型 (factorized-based neural network, FNN) 以获得用户行为特征之间的高阶交互关系。Qu 等^[6]针对用户响应预估问题,提出了一种基于乘积算子的深度神经网络 (product-based neural network, PNN) 方法,通过在嵌入层和神经网络层之间引入乘积操作,实现对用户行为数据高阶交互关系的识别。显然,基于深度学习机制的 FNN 和 PNN 均可有效发现用户行为数据特征间的高阶交互关系却忽略了低阶交互关系对特征组合的影响。为了同时兼顾特征间的高阶和低阶交互关系,Cheng 等^[3]提出了一种 Wide&Deep 神经网络模型,并成功应用于 Google Play 产品推荐中,但是该模型仍然首先基于人工提取的基础特征作为宽度部分的信息,当数据非常稀疏时,利用人工特征提取和合成的 Wide&Deep 将难以奏效。

综上所述,针对用户行为数据的稀疏性、特征的高度交互和难以提取性,FM、FFM、Deep Learning、Wide&Deep 等能够自动获取特征低阶和高阶交互关系,并进行特征抽取,进而实现点击率预估和个性化推荐。但是已有研究尚未有效构建同时自动获取用户行为数据的低阶和高阶交互关系,没有获取更加有效的特征组合模型,亦没有将相关模型应用于移动 APP 广告转化率预估中。为此,本文提出一种基于 FFM 和深度卷积神经网络的改进 Wide&Deep 用户行为特征交互关系自动提取和组合的算法即 DCNFFM,并用于移动 APP 广告转化率预估中。

1 基于 FFM 和 CNN 的改进 Wide&Deep 算法

1.1 算法框架

本文研究的问题具有数据类型多样和稀疏的特点,因此,在进行转化率预估时需首先进行数据

归一化和预处理。然后,基于上述数据,设计特征交互关系抽取和特征组合模型。最后,基于提取特征,构建转化率预估单元。针对 Google 所提 Wide&Deep 模型的不足和本文所研究问题的特性,提出如图 1 所示的融合数据预处理和特征交互自动提取模块的改进 Wide&Deep 广告转化率预估模型。

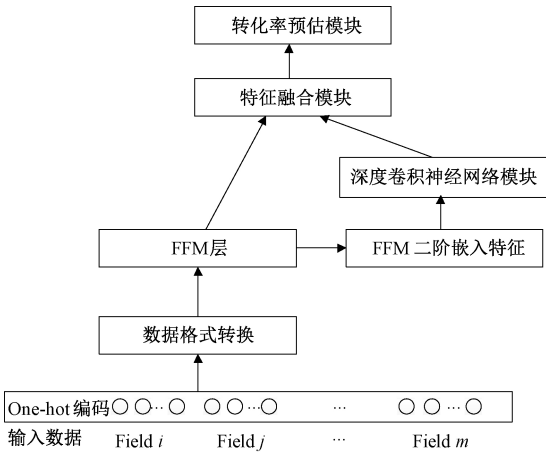


图 1 改进的 Wide&Deep 模型结构

Figure 1 Improved Wide&Deep model structure

与 Google 所提用于移动 APP 广告推荐的 Wide&Deep 模型相比较,笔者所提算法的不同之处在于如下 3 个部分:①数据预处理模块不同,不对原始数据进行任何人为特征聚类或抽取,充分结合 FFM 特性,在数据输入端直接根据原始数据所提供的域 (Field) 进行特征的 One-hot 编码,再转换为 FFM 数据格式^[14]。②Wide 模块部分,充分利用 FFM 在特征高阶交互关系上的突出性能,设计了基于 FFM 自动提取特征低阶复杂交互关系的宽度模块,以弥补传统 Wide&Deep 模块中需要人工提取特征宽度模块的不足。③深度模块部分,采用深度卷积神经网络模块,其输入为 FFM 所提取低阶交互关系特征的二阶嵌入 (各特征的隐含向量表示),以更好地融合低阶特征,既避免了再次进行嵌入操作的复杂性,又自然地特征转化为深度卷积神经网络易于处理的矩阵模块。关于输出模块部分,转化率预估本质为回归问题,因此,各类回归模型皆可适用,如逻辑回归、支持向量机或多种模型的融合等。下面给出所提算法核心部分的说明。

1.2 深度卷积神经网络模块

本文提出基于 FFM 二阶交互特征信息的深度模型,模块结构如图 2 所示。首先,深度模块的输入采用 FFM 进行二阶特征交互关系提取后的每个特征的隐含向量 $\mathbf{V}_i = [v_{i1}, v_{i2}, \dots, v_{ik}]^T$ 作为

深度卷积神经网络的输入^[14]。这样既可以充分利用 FFM 提取特征的隐含向量表示,减少了传统 Wide&Deep 中深度模块还需要再次进行编码表示操作的复杂性,又将二阶特征间的交互关系自然地融入到深度神经网络中。这里的隐含向量是对特征的一种自然表示,可类比于图像的特征,对

于 n 个特征,则其隐含向量维度为 $n \times k$,如共有 196 个元素,进行变换后成为 14×14 的矩阵,输入深度卷积神经网络中,自然合理。在图 2 中,深度卷积网络共 7 层,由两部分组成,即 2 层卷积神经网络和 3 层前馈全连接神经网络,这里全连接神经网络的输出单元数设为 2。

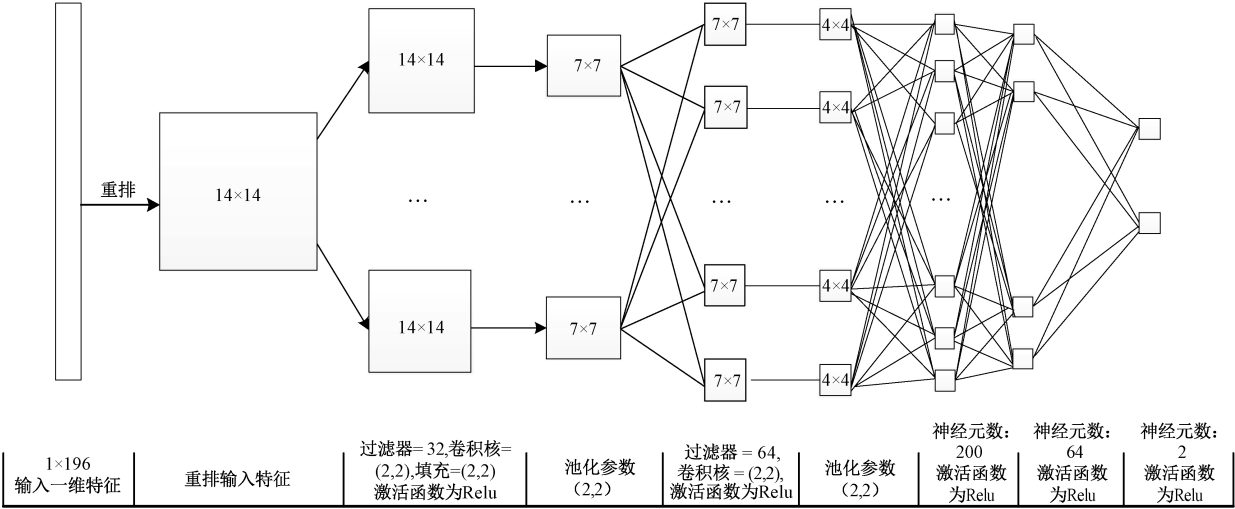


图 2 深度卷积神经模块
Figure 2 Deep convolutional neural networks module

1.3 基于逻辑回归的特征融合

为了实现对 Wide 部分 FFM 模型和 Deep 部分深度卷积神经网络的参数训练和特征融合,分别将两部分模型的交互特征输入至逻辑回归单元,利用输入样本的转化率作为监督信息,实现参数最优化。

对于第 p 个样本,FFM 模块提取的一阶特征为 $X_1^p = \sum_{i=1}^n w_i x_i^p$,二阶特征为 $X_2^p = \sum_{i=1}^n \sum_{j=i+1}^n \langle V_{i,j}^p, V_{j,i}^p \rangle x_i^p x_j^p$,深度神经网络部分提取的具有高阶交互关系的特征为全连接神经网络的二维输出,简记为 $X_3^p = \delta(W_{\text{deep},1}, x^p)$, $X_4^p = \delta(W_{\text{deep},2}, x^p)$,其中, $\delta(\cdot)$ 表示深度模块输入输出函数关系。将上述 4 类具有不同阶交互关系的特征再次进行交叉整合,作为逻辑回归单元的输入,实现对低阶和高阶交互特征的融合,进而可获得预估转化率为:

$$p(\hat{y} = 1 | x^p) = \frac{1}{1 + e^{-f(x^p)}} \quad (1)$$

式中: $f(x^p) = [X_1^p + X_3^p, X_2^p + X_4^p]^T$ 。

宽度和深度模块采取联合训练的思想,即通过逻辑回归将拟合误差同时反馈给浅层 FFM 模型和深层 CNN 模型中,采用 Adam 作为优化器,基于式(2)所示的交叉熵损失函数,通过随机梯度下降法获得各模块各部分的最优权值。

$$\log loss = -\frac{1}{N} \sum_{p=1}^N (y_p \log \hat{y}_p + (1-y_p) \log (1-\hat{y}_p)) \quad (2)$$

式中: N 表示测试集样本数量; y_p 表示测试集中第 p 个样本的真实转化率; \hat{y}_p 表示第 p 个样本的预估转化率。

1.4 算法流程

输入: 训练样本集; 特征数量 n ; field 数量 m ; Wide&Deep 各模块的结构参数;
输出: 移动 APP 转化率预估值 \hat{y}_p ;
开始:

- 1. 随机初始化深度卷积神经网络参数,对训练样本 One-hot 编码,进行数据格式转换;
- 2. 采集的数据输入 FFM 模块,提取数据低阶和高阶交互特征;
- 3. 计算 FFM 的低阶和高阶特征交互项 $X_{\text{fe1}}^p = \sum_{i=1}^n w_i x_i^p$ 以及 $X_2^p = \sum_{i=1}^n \sum_{j=i+1}^n \langle V_{i,j}^p, V_{j,i}^p \rangle x_i^p x_j^p$;
- 4. 基于 FFM 所得各特征的隐含向量矩阵

$$V = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1k} \\ V_{21} & V_{22} & \cdots & V_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ V_{n1} & V_{n2} & \cdots & V_{nk} \end{bmatrix},$$
 获得卷积神经

网络的输出 $X_3^p = \delta(W_{\text{deep},1}, x^p)$, $X_4^p = \delta(W_{\text{deep},2}, x^p)$;

5. 基于 FFM 和深度卷积神经网络的输出,利用式(1)计算逻辑回归输出 \hat{y}_p ;
6. 利用式(2)计算交叉熵损失函数;
7. 利用随机梯度下降法更新 FFM 和深度卷积神经网络的参数;
8. 重复步骤 2~7,直至收敛满足停止条件。

2 算法在腾讯移动 APP 广告转化率预估中的应用

2.1 实验背景

为验证本文所提算法性能,将其应用于腾讯移动 APP 广告转化率预估中,该问题以移动 APP 广告为研究对象,在给定广告、用户和上下文情况下,通过特征提取和模型构建,预测广告被点击后发生激活的概率。腾讯公司原始数据集共提供了 25 个高度交互特征。根据所提供的数据,并考虑广告主数据回流的滞后特性,这里设定某月第 17 d 到第 29 d 的数据为训练数据,共计 34 548 872 条;第 30 d 数据为测试数据,共有 3 364 044 条。

2.2 实验设置

为了说明该算法解决移动 APP 广告转化率预估的有效性,共进行 4 组实验,分别设置如下:①算法参数对结果的影响,通过实验,给出正则化策略、卷积神经网络隐含层数量和激活函数的选择。②基于原始特征和 Xgboost 特征约减后,DCNN 解决该类问题的性能,以及基于 Xgboost 特征重要度排序的特征抽取对预测结果的影响。③通过比较,说明基于 FFM 模块提取稀疏数据特征的 DCNN 的有效性。④算法整体性能,通过与 BP 神经网络、深度卷积神经网络、FM、FFM、深度全连接神经网络、Google 所提 Wide&Deep 以及 Wide&Cross 算法的比较进行说明。模型最优参数通过网络搜索法获得,所比较算法的模型参数设置如下。

Xgboost 参数设置为学习率 0.05,最大树深度 5,样本随机采样比例 0.7,列采样比例 0.7,目标函数采用二分类逻辑回归。BP 神经网络参数设置为两隐层网络,隐层神经元数目分别为 64 和 512,隐层单元激活函数为 Relu,输出层使用 softmax 单元,采用随机梯度下降学习策略,其中,学习率为 0.01,损失函数采用式(2)交叉熵。深度卷积神经网络:将 25 个特征变换为 5×5 矩阵作为卷积神经网络输入,第 1 卷积层的卷积核个数为 32,第 2 层为 16。两层卷积核大小均为 (2,2),激活函数采用 Relu。池化层采用最大池化操作。

后接两个全连接隐含层,隐含层神经元个数分别为 256 和 68。在本文所提 Wide&Deep 模块中,DCNN 输入为来自 FFM 的隐含特征向量,共 196 个,变换为 14×14 矩阵作为网络输入,其余参数基于实验结果,将在 2.3 节给出。

算法性能指标:这里通过预测精度、AUC、损失函数和计算时间说明所比较算法的性能,其中,精度值和 AUC 的值越大,损失函数和计算时间值越小,说明算法性能越好。

2.3 模型超参数的选择

本文分析了深度卷积神经网络的正则化方法、神经网络激活函数、卷积神经网络层数、全连接层神经元个数对实验结果的影响,结果如表 1 所示。

表 1 DCNN 模型超参数实验结果				
Table 1 Experimental results of DCNN parameters				
超参数	超参数取值	精度	AUC	损失函数
正则化	L_1	0.745	0.788	0.549
	L_2	0.723	0.784	0.558
	none	0.718	0.782	0.560
激活函数	tanh	0.722	0.784	0.557
	Relu	0.745	0.788	0.549
	sigmoid	0.719	0.783	0.558
网络层数	1 层	0.719	0.784	0.558
	2 层	0.745	0.788	0.549
	3 层	0.733	0.785	0.554
神经元个数	100 个	0.717	0.781	0.559
	200 个	0.745	0.788	0.549
	300 个	0.725	0.784	0.554

由表 1 可以看出,①正则化操作可有效提高算法的各类性能指标,而采用 L_1 范数时性能最优;②激活函数中 Relu 优于 tanh,两者均明显优于 sigmoid 函数;③卷积层的层数并非越多越好,在本文中,两层卷积时效果最佳;④全连接层神经元数量在 200 个左右时,性能最优。

综上,本文 Wide&Deep 模块中,DCNN 超参数设置: L_1 范数正则化、两卷积层,第 1 层卷积核规模为 32,第 2 层为 64;两层卷积核大小均为 (2,2),激活函数为 Relu 函数;两层中的池化操作均采用最大池化方法。两卷积层后为含两隐层全连接前馈神经网络,两隐层神经元个数分别为 200 和 64。

2.4 基于原始特征及特征约减后的深度模块性能

这里首先研究基于原始特征的深度学习性能,通过与特征抽取中典型的算法 Xgboost、BP 神经网络

络(复杂)进行对比,说明 DCNN 对于本问题原始特征的处理能力,各算法的参数设置如 2.2 节和 2.3 节所述。3 类算法的性能指标如表 2 所示。

表 2 基于原始特征的转化率预估

Table 2 Conversion rate estimation based on original feature				
模型	精度	AUC	损失函数值	训练时间/s
Xgboost	0.690	0.743	0.588	2 075.840
BP	0.653	0.713	0.620	1 335.471
DCNN	0.643	0.686	0.639	2 075.840

由表 2 可知:①在未对原始数据进行任何特征提取的情况下,Xgboost 的预测精度、AUC 以及损失函数均最优,其运行时间高于 BP 神经网络;②BP 神经网络的预测性能与 Xgboost 较相似,算法运行时间较少;③与 Xgboost 和 BP 神经网络相比较,DCNN 算法整体性能均较差,精度明显劣于上述两种方法,运行时间却远大于比较算法。综上所述,对于本文所研究移动 APP 广告转化率预估,在未对特征进行任何预处理情况下,DCNN 算法性能难以满足要求。

本文进一步采用 Xgboost 方法,对属性重要度进行排序,然后选择重要度高的特征(建模时去掉重要度最低的 5 个特征)对上述 3 种方法进行比较,结果如表 3 所示。

表 3 基于重要特征的算法性能分析

Table 3 Performance analysis of algorithms based on important features				
模型	精度	AUC	损失函数值	训练时间/s
Xgboost	0.688	0.737	0.593	1 758.265
BP	0.661	0.725	0.607	1 362.392
DCNN	0.629	0.683	0.648	2 130.745

比较表 3 和表 2 数据可知,基于 Xgboost 重要度排序,在删除重要性较小的特征之后 3 类算法的整体性能都有所下降,而计算时间却有所提升,显然,采用传统的自动特征处理也无法满足本文所研究问题的性能要求。其原因在于,上述实验均直接将原始数据输入到模型中,模型并不能很好地提取特征之间的高阶交互关系,影响了模型的整体性能。

2.5 基于 FFM 提取特征的深度模块性能

2.4 节实验表明,从数据中直接提取特征的深度学习模块在本问题上性能较差,为了进一步验证本文所提基于 FFM 模块直接提取特征的深度卷积神经网络(FFM+DCNN)的有效性,分别和直接使用原始数据特征输入卷积神经网络、对原

始稀疏特征进行嵌入处理之后输入到卷积神经网络中(所采用的嵌入处理方式与 Wide&Deep 模型使用的方法一致)进行比较。实验结果如表 4 所示。

表 4 基于 FFM 提取特征的 DCNN 有效性

Table 4 Performance of DCNN with FFM feature extraction				
模型	精度	AUC	损失函数值	训练时间/s
原始数据	0.643	0.686	0.639	2 075.840
嵌入编码	0.653	0.713	0.628	2 356.276
FFM+DCNN	0.694	0.728	0.605	2 575.840

从表 4 实验结果可以看出:①本文基于 FFM 提取低阶特征,再将其作为深度模块输入进行高阶特征提取的方法,可使得转化率预估精度明显提升,比直接输入原始数据提高了 7.9%,比传统嵌入编码方式提高了 6.3%;与表 1 和表 2 中其他方法相比,算法性能也最优;②直接输入原始数据时,耗时最少,但是,精度等性能指标皆相对最差;③本文所提特征交互提取方法耗时相对增加,比直接输入原始数据多用时 500 s,比嵌入编码方式多用时 219.564 s,均在可接受范围之内。

综上所述,本文所提基于 FFM 提取低阶特征的深度卷积神经网络高阶特征抽取在可接受耗时下能够获得更好的性能,表明低阶特征抽取模型 FFM 可以从稀疏输入数据中提取出更好的特征供深度卷积神经网络使用,提高转化率预估精度。

2.6 基于改进 Wide&Deep 模型的转化率预估

为了说明本文所提算法(DCNFFM)的整体性能,将其与 BP 神经网络、DCNN、因子分解机 FM、场感知因子分解机 FFM,含有 FFM 的深度全连接神经网络(Relu 单元组成)(DNNFFM)、Google 公司所提 Wide&Deep 以及另一种 Wide&Cross^[12]模型进行比较。对于 Google 公司所提 Wide&Deep 以及 Wide&Cross 均采用原文中所给参数,这里不再赘述。算法 5 次运行后的平均性能指标值如表 5 所示。

由表 5 可知:①在预测精度、AUC,损失函数 3 个性能指标方面,本文所提算法明显优于所比较其他算法,与 Google 所提 Wide&Deep 以及 Wide&Cross 相比,其精度分别提高约 2.8%和 2.3%;②加入了 FFM 的特征提取模块,可有效改善深度神经网络在转化率预估上的模型性能,DNNFFM 和 DCNFFM 的整体性能在所有算法中表现优良,比单独 DCNN 预测精度分别提高了

表 5 基于不同特征提取模型的转化率预测

Table 5 Conversion rate prediction based on different feature extraction models

模型	精度	AUC	损失函数值	训练时间/s
BP	0.631	0.689	0.751	2 536.212
DCNN	0.653	0.713	0.620	2 356.276
FM	0.682	0.731	0.598	869.135
FFM	0.698	0.742	0.586	1 208.568
DNNFFM	0.728	0.769	0.569	2 942.759
DCNFFM	0.746	0.788	0.549	2 786.007
Wide&Deep	0.718	0.759	0.571	2 369.487
Wide&Cross	0.723	0.768	0.559	2 469.134

8.9%和10.7%;③在算法计算复杂度方面,本文算法较高,但低于DNNFFM,说明将FFM与深度学习进行融合时,选择深度卷积神经网络比传统深度前馈神经网络可在提高模型性能的同时,降低计算复杂度。

3 结论

提出了一种改进的Wide&Deep模型,即宽度模块采用FFM实现一阶和二阶特征交互关系的提取,然后基于所提取二阶交互关系,构建深度卷积神经网络模型,进一步获取特征间的高阶交互关系。采用所提模型,构建移动APP广告转化率预估时,可以避免对原始数据进行前期手工特征抽取,并通过直接利用FFM二阶交互特征隐含向量,减少传统Wide&Deep需进行多次嵌入操作的复杂性。

将所提算法(DCNFFM)应用于腾讯移动APP广告转化率预估中,并通过与FM、FFM、BP神经网络、深度全连接神经网络(Relu)(DNNFFM)、深度神经网络(DCNN)、Wide&Deep模型、Wide&Cross模型比较,说明所提算法可在可接受的计算复杂度下,获取更高的预估精度。

参考文献:

[1] YAO Y, ZHAO W X, WANG Y, et al. Version-aware rating prediction for mobile app recommendation [J]. ACM transactions on information systems, 2017, 35 (4):38.

[2] CHAPELLE O, MANAVOGLU E, ROSALES R. Simple and scalable response prediction for display advertising [J]. ACM transactions on intelligent systems and technology, 2015, 5(4):1-34.

[3] CHENG H T, ISPIR M, ANIL R, et al. Wide & deep learning for recommender systems [C]//Proceedings of the 1st Workshop on Deep Learning for Recommender

Systems-DLRS 2016. New York:ACM, 2016:7-10.

[4] RENDLE S. Factorization machines [C]//2020 IEEE International Conference on Data Mining. New York: IEEE, 2011:995-1000.

[5] JUAN Y, ZHUANG Y, CHIN W S, et al. Field-aware factorization machines for CTR prediction [C]//Proceedings of the 10th ACM Conference on Recommender Systems. New York:ACM, 2016:43-50.

[6] QU Y R, CAI H, REN K, et al. Product-based neural networks for user response prediction [C]//IEEE International Conference on Data Mining. New York: IEEE, 2016: 1149-1154.

[7] WANG R X, FU B, FU G, et al. Deep & cross network for ad click predictions [C]//Proceedings of the ADK-DD'17. New York:ACM, 2017:1-7.

[8] CHANG Y W, HSIEH C J, CHANG K W, et al. Training and testing low-degree polynomial data mappings via linear SVM [J]. Journal of machine learning research, 2010, 11(11):1471-1490.

[9] KASAP Ö Y, TUNGA M A. A polynomial modeling based algorithm in top-N recommendation [J]. Expert systems with applications, 2017, 79:313-321.

[10] 孙晓燕,朱利霞,陈杨. 基于可能性条件偏好网络的交互式遗传算法及其应用 [J]. 郑州大学学报(工学版), 2017, 38(6):1-5.

[11] SUN X Y, GONG D W, JIN Y C, et al. A new surrogate-assisted interactive genetic algorithm with weighted semisupervised learning [J]. IEEE transactions on cybernetics, 2013, 43(2):685-698.

[12] ZHU J, SHAN Y, MAO J C, et al. Deep embedding forest: forest-based serving with deep embedding features [C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York:ACM, 2017:1703-1711.

[13] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553):436-444.

[14] BOWERS S, BOWERS S, BOWERS S, et al. Practical lessons from predicting clicks on ads at facebook: eighth international workshop on data mining for online advertising [C]//Eighth International Workshop on Data Mining for Online Advertising. New York: ACM, 2014:1-9.

[15] RENDLE S. Factorization machines with libFM [J]. ACM transactions on intelligent systems and technology, 2012, 3(3):1-22.

[16] BLONDEL M, FUJINO A, UEDA N, et al. Higher-order factorization machines [C]//International Conference on Neural Information Processing Systems. Massachusetts:MIT, 2016:3359-3367.

[17] XIAO J, YE H, HE X N, et al. Attentional factorization machines: learning the weight of feature interactions via attention networks [C] // Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Melboerne: IJCAI, 2017: 3119–3125.

[18] HINTON G E, DENG L, YU D, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups [J]. IEEE signal processing magazine, 2012, 29(6): 82–97.

[19] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2015-04-10) [2020-02-04]. <https://arxiv.org/abs/1409.1556>.

[20] NIU Y, LU Z, WEN J, et al. Multi-modal multi-scale deep learning for large-scale image annotation [J]. IEEE transactions on image processing, 2019, 28(4): 1720–1731.

[21] ZHANG W N, DU T M, WANG J. Deep learning over multi-field categorical data [M] // Lecture Notes in Computer Science. Cham; Springer International Publishing, 2016: 45–57.

The Mobile APP Conversion Rate Prediction Based on Improved Wide&Deep of Interactive Feature Extraction

SUN Xiaoyan, NIE Xin, BAO Lin, CHEN Yang

(School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221008, China)

Abstract: The conversion rate prediction (CRP) of mobile APP advertising was challenging due to the high-dimension, sparsity, and high interactions. Motivated by this an improved Wide&Deep model was proposed by fusing field-aware factorized machine (FFM) and deep convolutional neural network (DCNN) to effectively and automatically obtain the lower-order and higher-order interactions of high-dimensional sparse features, so as to realize the automatic and efficient combination of features and improve the accuracy of CRP. The framework of the proposed algorithm was first delivered, and a feature combination algorithm based on the width module FFM to extract the interactive relations of lower-order features was presented for the embedded sparse data. The extraction of the higher-order interactive features based on a DCNN was further given by fusing the latent features obtained by the FFM. Finally, the interactive feature combinations obtained by width and depth modules were integrated for the CRP. The application of the proposed algorithm in predicting the conversion rate of Tencent’s mobile APP advertisements demonstrated the effectiveness of the method.

Key words: conversion rate prediction; feature interaction; field-aware factorized machine; Wide&Deep; mobile APP