计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2021,30(6):231−237 [doi: 10.15888/j.cnki.csa.007958]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于改进降噪自动编码器的点击率预测①

刘　勐[1], 王洪波[1,2,3], 王富豪[1], 李亚峰[1]

[1](复旦大学 工程与应用技术研究院, 上海 200433)
[2](复旦大学 上海智能机器人工程技术研究中心, 上海 200433)
[3](复旦大学 智能机器人教育部研究中心, 上海 200433)
通讯作者: 王洪波, E-mail: wanghongbo@fudan.edu.cn

摘　要: 点击率 (CTR) 预测是个性化广告和推荐系统中的一项基本任务. 针对提升点击率预测效果和处理冷启动问题, 本文中提出了一种基于改进降噪自动编码器的点击率预测模型 ADVAE (ADditional Variational AutoEncoder), 该模型在输入数据加入高斯随机噪声, 利用改进的降噪自动编码器生成新的嵌入特征, 然后分别进行低阶和高阶的特征交互来预测用户点击行为. 该方法可以在数据稀疏以及系统冷启动情况下, 更深层地学习特征嵌入与交叉之间的关系. 该模型关注特征域之间的交互, 动态修复低频数据的特征嵌入, 具有更强的鲁棒性. 此外, 该方法可以动态应用到其他深度学习模型, 具有更高的灵活性. 实验结果表明, 该方法在点击率预测和系统冷启动问题上的性能表现均优于现有方法.

关键词: 点击率预测; 特征交互; 降噪自动编码器; 冷启动

引用格式: 刘勐,王洪波,王富豪,李亚峰.基于改进降噪自动编码器的点击率预测.计算机系统应用,2021,30(6):231–237. http://www.c-s-a.org.cn/1003-3254/7958.html

Click-Through Rate Prediction Based on Improved Denoising Autoencoder

LIU Meng[1], WANG Hong-Bo[1,2,3], WANG Fu-Hao[1], LI Ya-Feng[1]

[1](Academy for Engineering and Technology, Fudan University, Shanghai 200433, China)
[2](Shanghai Engineering Research Center of AI & Robotics, Fudan University, Shanghai 200433, China)
[3](Engineering Research Center of AI & Robotics, Fudan University, Ministry of Education, Shanghai 200433, China)

Abstract: Click-Through Rate (CTR) prediction is a fundamental task in personalized advertising and recommendation systems. This study proposes a model named ADditional Variational AutoEncoder (ADVAE) based on an improved denoising autoencoderto improve CTR prediction and cold-start. It adds random Gaussian noise to the input data and generates new embedded features by the improved denoising autoencoder. Then, multi-level features interact to predict the users' clicking. This method can learn the relationship between feature embedding and interactions in data sparse and cold-start situations. In addition, it has strong robustness since it focuses on the interaction in feature domains and dynamically repairs feature embedding of low-frequency data. Besides, this method can be dynamically applied to other deep learning models, with high flexibility. The results show that the proposed approach outperforms its counter parts in terms of CTR prediction and cold-start.

Key words: CTR prediction; feature interaction; denoising autoencoder; cold-start

信息爆炸与大数据技术的普及, 促进了个性化推荐技术的快速发展[1]. 作为广告推荐系统一项重要任务, 点击率预测 (CTR) 对于许多互联网公司来说都是必不可少的. 例如, YouTube 每天的视频播放时间已超

① 基金项目: 河北省重点研究计划 (20371801D)
Foundation item: Major Research Program of Hebei Province (20371801D)
收稿时间: 2020-10-11; 修改时间: 2020-11-05, 2020-11-17; 采用时间: 2020-11-24; csa 在线出版时间: 2021-06-01

过 10 亿小时. 其推荐系统需要根据用户需求、兴趣等, 通过推荐算法通过从海量数据中高效准确的预测用户感兴趣项目, 并将结果以个性化列表的形式推荐给用户[2].

用户与项目数据的特征通常是离散和稀疏的, 因此 CTR 预测任务的关键挑战是如何有效学习特征之间的交互来建模这种数据. 过去, 已有许多学者提出相关算法来解决此问题, 如逻辑回归 (LR)[3]、基于树的模型[4]、贝叶斯模型[5]、基于张量的模型[6] 和基于因子分解的模型[7,8] 等. 这些模型通过学习成对特征之间的低阶交互提高特征表示能力, 但同时会带来了与任务无关的特征交互组合. 近年来, 依靠强大特征表示能力, 深度学习在计算机视觉[9,10] 和自然语言处理[11,12] 等许多领域开始大放光彩. 最近几年深度学习模型也开始逐步在推荐领域得到应用, 例如基于神经网络的因子分解模型 (FNN)[13], 基于注意力机制的因式化机模型 (AFM)[14], Wide & Deep[15], DeepFM[16] 等. 当前绝大多数解决点击率预测问题的深度学习算法大致分为 3 个步骤. 首先, 利用嵌入表示模型将用户和项目高维稀疏特征映射为低维稠密向量. 然后, 对得到的嵌入向量使用内积、外积或者哈达玛积等运算获得特征交叉表示. 最后, 基于隐向量使用多层感知器 (MLP) 预测用户对于项目的评分或者偏好.

在实际应用中, 由于相当大比例的用户和项目属性通常是离散和稀疏的, CTR 模型会使用嵌入操作来处理输入数据. 但是, 常见的嵌入表示模型在处理数据集出现频率较低的样本时, 很难学习到合适的特征表征, 在系统冷启动时性能表现较差. MLP 在深度学习模型中起到了基于 bit-wise 层级的特征交互和非线性变换的作用, 但在 vector-wise 层次的特征交互上表现较差. 同时, 随着 MLP 的深度和宽度的增加, 在增加学习能力的同时也增加了参数量和过拟合的风险.

基于上述问题, 本文提出了一种基于改进降噪自动编码器 (DAE)[17] 的点击率预测模型 ADVAE (ADditional Variational AutoEncoder). 该方法通过添加噪声来生成嵌入信息来学习稀疏和高维输入特征的稠密低维表示, 提高了模型在 bit-wise 和 vector-wise 层次的特征交互能力, 改善由于数据稀疏性引起的特征不平衡问题. 同时, ADVAE 模块使得模型即使在数据样本特征稀疏甚至缺失情况下, 也可以产生有效的嵌入表示, 有效缓解了推荐系统中常见的冷启动问题. 同时, 该模型的 ADVAE 模块可以针对不同任务动态地应用于到其他模型, 具有很强的灵活性.

1 点击率预测

点击率预测利用用户与项目之间的二元关系, 基于用户历史行为记录或者相似性关系帮助发现用户可能感兴趣的项目, 对用户的点击行为进行预测.

1.1 点击率预测模型

点击率预测模型主要包含传统模型和深度学习模型两种, 其中传统的推荐方法主要分为以下 3 种[18]: 基于内容的推荐 (content-based recommendtion)[19]、协同过滤推荐 (collabortive filtering recommendation)[20] 和混合推荐 (hybrid recommendation)[21], 深度学习通过组合低阶特征形成稠密的高阶语义信息, 从而自动发现数据的分布式特征表示, 解决了传统机器学习手工设计特征的问题. 常见的深度学习模型除了包含上述 3 种传统方法外还有基于社交网络[22]、场景感知[23] 等方法的推荐模型.

1.2 卷积神经网络

卷积神经网络在推荐系统中有着较为广泛的应用, 主要用于从图像、文本、音频等信息中提取数据的隐藏特征. 相比较多层感知机, 卷积神经网络使用权重共享结构降低模型复杂度, 有着更好的泛化能力.

在目前的 CTR 模型中, ConvNCF[24] 模型应用 CNN 来改进 NCF, NCF 使用外积而不是点积来建模用户-项目交互模式. 此外, CCPM[25] 使用对齐方式对相邻字段执行卷积, 学习多个卷积层的相邻特征间的依存关系. FGCNN[26] 使用卷积层来代替传统的交互方式, 使用多层卷积生成新的嵌入向量. 上述卷积神经网络均在点击率预测任务上取得了不错表现, 但是忽略了特征嵌入和交互的关系.

1.3 降噪自动编码器

自编码器 (Auto Encoder, AE)[27] 通过一个编码器和一个解码器来重构输入数据, 学习数据的隐层表示. 但是如果仅仅通过最小化输入输出的误差来对模型训练, 自编码器常会学到一个恒等函数. 为解决这个问题, 降噪自编码器[17] 通过在自动编码器的输入数据中加入噪声得到, 这样降噪自编码器在重构输入数据时, 就被迫去除这种噪声来学习到更加鲁棒的输入数据的表达, 降噪自编码器通过这种方式提升了泛化能力以及在稀疏数据下的表现.

2 ADVAE

考虑到数据的稀疏性和不平衡性, 本文提出了一种改进的降噪自动编码器 ADVAE . 如图 1 所示, 该模型通过特征嵌入和特征交互合并为统一操作动态获取嵌入向量的方式提高在点击率预测任务上的表现, 主要包括普通 Embedding模块、ADVAE Embedding 模块、低阶特征交叉模块以及高阶特征交叉模块. 其中, ADVAE 模块解决了特征稀疏嵌入的问题, 生成的特征可以与原始特征嵌入向量融合, 为其他分类模型灵活地提供更为丰富的特征输入, 是该模型的关键部分.

2.1 输入层与嵌入层

大多数点击率模型的数据输入采用 one-hot 的形式, 使用嵌入操作将高维稀疏数据映射为低维特征向量. 假设 user 与 item 的输入数据表示为:

$$I_{\mathrm{n}}=\left[S_1;S_2;S_3;S_4;\cdots;S_f\right] \tag{1}$$

其中, s_i表示在第 i 个域的输入数据的 one-hot 表示, f 表示输入数据域的总数, 在不同的数据集上会有所变化.在域 $i(1\le i\le f)$中,嵌入操作后的向量表示为$e_i\in\Re^{f\times D}$, 其中 D 是嵌入向量的维度. 因此, 每个输入可以表示为矩阵$E=\left(e_1;e_2;e_3;\cdots;e_f\right)$, 其中$E\in\Re^{f\times D}$.

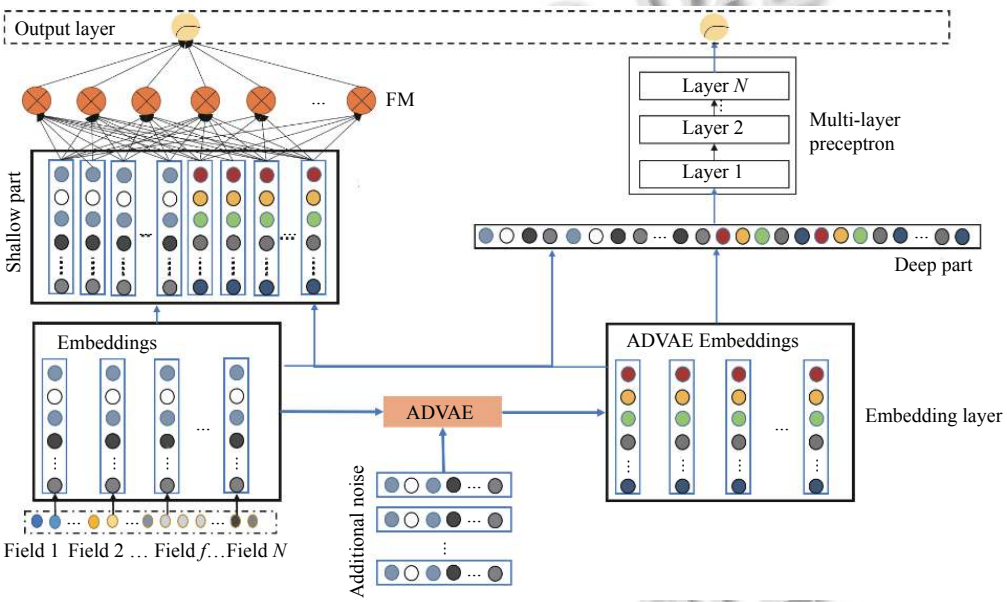


图 1 ADVAE 网络结构图

2.2 随机噪声

在受模型复杂度、训练集数据量以及数据噪音等问题的影响下, 通过编码器得到的初始模型往往存在过拟合的风险. 在本文中, 输入数据的部分使用噪声替代, 这种加入同源随机噪声的方式一定程度上减轻了训练数据与测试数据的差异性, 可以提高模型的有效性和鲁棒性, 增加模型的泛化能力. 其中, 对于每个噪声向量$e_i\in\Re^{1\times D}$使用高斯分布进行随机初始化, 并将普通 Embedding 模块得到的向量与噪声向量拼接作为 ADVAE 模块卷积操作的输入.

$$E'=concat(E,N)=\left(\left[e_1;e_2;e_3;\cdots;e_f\right],\left[e_{f+1};e_{f+2};\cdots;e_D\right]\right) \tag{2}$$

其中, $concat$ 表示矩阵级联, $E\in\Re^{f\times D}$是嵌入层的输出, $N\in\Re^{(D-f)\times D}$是噪声矩阵.

2.3 ADVAE 模块

如图 2 所示, ADAVE 模块主要包含 3 个卷积层和 3 个转置卷积层. 卷积操作不仅可以学习到 pair-wise 层次的特征交互, 而且学习到多个域之间的交互. 在对隐向量进行上采样的过程中, 转置卷积的引入, 一方面还原原始特征信息, 另一方面过滤掉了原始特征中对任务不相关特征, 使得模型更关注于与任务相关的特征, 上采样过程中相关特征则会被赋予更大的权重,起到了数据特征在位置维度上的注意力机制作用. 同时, ADAVE 模块进行的全局特征交互是 bit-wise 层次的, 而不仅是 pair-wise 层次的. 此外, 添加噪声输入可以克服嵌入向量稀疏和数据不平衡的缺点, 同时, 这些额外

的噪声即可以捕获内部关系, 也可以生成有利于交互的新向量.

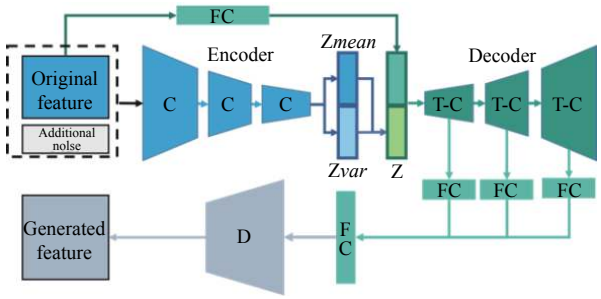


图 2 ADVAE 模块

编码器由三层卷积层构成, 每经过一层卷积层, 卷积核的宽度减少一倍, 编码器的输入为嵌入向量与噪声向量组成的特征矩阵$E' \in \Re^{1\times D\times D}$, 卷积输出为$X_0 = \Re^{C\times N\times N}$. 在卷积层最后使用两层全连接网络得到隐向量的均值和方差, 计算过程如下所示.

$$Z_{\text{mean}} \equiv \mu = \varphi(W_m X_o + b_m) \tag{3}$$

$$Z_{\log} \equiv \log\sigma^2 = \varphi(W_l X_o + b_l) \tag{4}$$

其中, X_o表示输入, W_m和W_l是两层全连接网络的权重矩阵, b_m和b_l是偏差, φ是激活函数, 则隐变量$Z \sim (\mu, diag(\sigma)) \in \Re^{N\times N}$.

$$Z_e = \varphi(W_e E + b_e) \tag{5}$$

$$Z = concat(Z_e, Z_c) \tag{6}$$

解码器使用与编码器同数量的转置卷积层, 通过对隐变量的上采样还原原始特征矩阵. 同时, 在编码器部分加入了一个判别器, 帮助提高 ADVAE 对点击率预测任务的特征学习.

2.4 损失函数

该模型的损失函数共由 4 部分组成, 分别是编码器重构损失L_M, 判别器的分类交叉熵损失L_B、MLP 分类器的交叉熵损失L_C以及编码器隐变量的 KL 损失 KL, 其中:

$$L = \lambda_1 L_M + \lambda_2 L_B + \lambda_3 KL + \lambda_4 L_C \tag{7}$$

降噪自动编码器一种具有降噪功能的特征提取器, 目的是将一个包含噪声的输入数据转化为一个干净的数据输出. 在损失函数中, 重构损失和编码器隐变量的 KL 损失保证了 ADVAE 模型尽可能在引入噪声条件下仍能还原出原始数据的输入分布. 其中, 重构损失使用均方误差损失, 以误差的平方和作为损失, 其函数易于求导, 保证模型对原始数据的生成能力; KL 损失函数通过计算 KL 散度估计两个分布的相似度, 对编码后隐变量进行优化, 保证模型的编码能力.

$$L_M = \frac{\sum_i^N \left(\left\|X_i - X_{f_i}\right\|^2\right)}{N} \tag{8}$$

$$KL = -KL\left(q_\phi\left(z_j|\mathrm{x_j}\right) \| \, p\left(Z_j\right)\right) \tag{9}$$

模型在对点击结果预测上使用了两部分的损失, 一是使用解码器多层级特征作为预测判别器的输入的分类损失L_B, 二是模型使用原始数据以及生成数据作为输入构造的 MLP 分类器的损失L_c, 两部分损失均使用交叉熵损失函数, 由于具有非常强的概率分布表征能力, 交叉熵损失函数常用于分类任务.

$$L_B = L_c \tag{10}$$

$$L_c = -\frac{1}{N}\sum_N^i (y_i \log y_i + (1-y_i)\log(1-\hat{y}_i)) \tag{11}$$

3 实验分析

本文在公开数据集 Ctriteo、Avazu 和 Movielen-20M 上分别在点击率预测效果以及系统冷启动性能两个方面与现有模型进行了对比实验, 详细说明如表 1. 参与对照实验的 CTR 模型包括仅使用初始化特征的线性方法 LR、考虑二阶特征交互的分解机方法 (FM、FFM) 以及高阶交互的深度学习模型 (DCN、Wide & Deep、AFM、Autoint、xDeepFM、FibiNet).

表 1 数据集说明

数据集	样本数量	特征域数量	特征数量
Criteo	45 840 617	39	998 960
Avazu	40 428 967	23	1544 488
MovieLen-20M	20 000 263	2	—

3.1 实验设置

本文的实验基于 PyTorch 框架实现, 在 3 个数据集上, 模型的训练参数如表 2 所示, 其中超参数$\lambda_1 = 1$、$\lambda_2 = 0.1$、$\lambda_3 = 0.5$、$\lambda_4 = 1$. 所有对照模型与 ADVAE 方法均使用相同的 MLP 训练参数.

表 2 模型训练参数

数据集	Criteo	Avazu	MovieLen-20M
嵌入维度	48	32	8
Batchsize	4096	4096	2048
卷积核大小	16, 8, 4	16, 8, 4	2, 2, 2
MLP神经元个数	1024	1024	128
MLP层数	3	3	3
Dropout比率	0.5	0.5	0.5
学习率	0.001	0.001	0.001

本文选取 AUC 和 LogLoss 值作为模型性能评估指标, AUC 值通过计算 ROC 曲线下的面积得到, LogLoss 通过计算预测结果与标签的交叉熵得到.

3.2 点击率预测实验

表 3 反映了各个模型在 Ctriteo、Avazu 和 Movielen-20M 数据集上的性能表现. 可以看出, AVDAE 在各个数据集上的表现均优于现有模型. 在实验中, 我们发现使用多层感知器 (MLP) 往往会导致过度拟合, 如图 3 所示对于 Wide & Deep, xDeepFM 和 DNN 之类的几种模型, 通过使用多层感知器 (MLP) 训练约 4 个 epoch 时的准确率开始下降. 但是, 在使用 ADVAE 模块后时, 在 Criteo 数据集上训练 10 个 epoch 以及 Avazu 数据集上训练 8 个 epoch 之后, 准确率仍有所提高, 实验结果说明, 这种动态嵌入方式极大地减轻了 MLP 的过度拟合的问题.

表 3 点击率预测实验模型性能表现

模型	Ctriteo		Avazu		MovieLen-20M	
	AUC	LogLoss	AUC	LogLoss	AUC	LogLoss
LR	0.7453	0.4986	0.7634	0.3891	0.7652	0.5636
FM	0.7951	0.4536	0.7759	0.3824	0.7727	0.5589
FFM	0.8027	0.4474	0.7792	0.3786	0.7717	0.5600
DCN	0.7942	0.4516	0.7694	0.3854	0.8336	0.4742
Wide & Deep	0.8006	0.4485	0.7754	0.3816	0.8192	0.5008
AFM	0.7903	0.4592	0.7620	0.3899	0.7663	0.5642
AutoInt	0.8081	0.4397	0.7802	0.3799	0.8293	0.4843
xDeepFM	0.8089	0.4387	0.7803	0.3802	0.8312	0.4765
Fibi	0.8115	0.4356	0.7812	0.3785	0.8301	0.4831
AVDAE(ours)	**0.8163**	0.4388	**0.7906**	0.3792	**0.8345**	0.4805

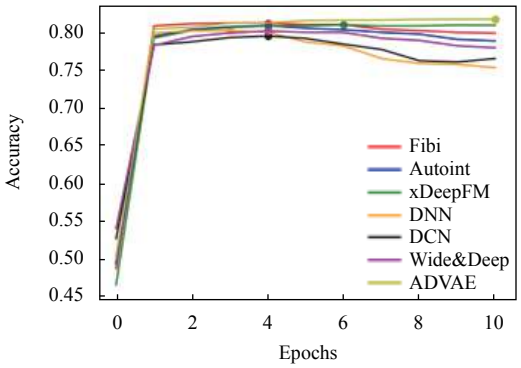
3.3 冷启动实验

冷启动在推荐系统中表示该系统积累数据量较少, 无法为新用户提供个性化推荐的问题, 是推荐系统的一个难题.

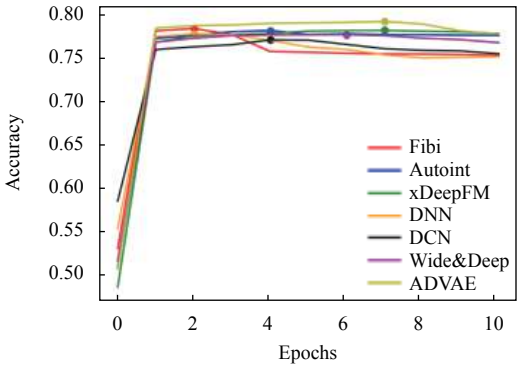
在冷启动实验中, 本文通过将输入特征的值置零来屏蔽一定数量的属性去模拟冷启动情况. 其中, K 值表示特定属性被屏蔽的概率. 如表 4 所示, K 值分别设置为 20%、40% 和 60%. 可以得到, 在 K 值等于 20%、40%、60% 的冷启动条件下, ADVAE 方法优于所有现有模型. 特别是在 K 值等于 20%、40% 的情况下, 与 AutoInt 和 xDeepFM 相比, ADVAE 性能有较明显的提升.

在 Criteo 和 Avazu 数据集这两个数据集中存在缺失值, xDeepFM、AutoInt 等算法更关注于数据嵌入后的特征交互, 但没有注意之前的嵌入操作是否合理. 而 ADVAE 模型通过引入噪声生成特征去动态修复原始嵌入, 使得模型在处理数据缺失问题时更加鲁棒.

结果表明, 在系统冷启动情况下, ADVAE 模型仍可以找出最合适的预测结果.



(a) Criteo 数据集性能变化



(b) Avazu 数据集性能变化

图 3 准确率变化曲线

4 结论与展望

本文提出了一种基于改进降噪编码器的点击率预测模型, 通过引入噪声数据生成新的嵌入特征来学习特征嵌入与特征交互的关系, 然后分别进行低阶和高阶的特征交互来预测用户点击行为. 本文和常见点击率预测模型, 如线性回归 (LR)、FFM、xDeepFM 等, 进行了比较. 实验结果表明, 本文提出的点击率预测算法在 AUC、LogLoss 等指标上显著优于现有模型, 同时, 在数据稀疏及系统冷启动条件下, 仍有较好的性能表现, 有效缓解过拟合现象.

本文提出的模型主要应用于当前互联网点击率预测任务, 其中 ADVAE 模块可以动态应用到各类点击率预测模型中, 具有很强的灵活性. 但是在实际应用中, 实际场景对算法的实时性要求较高. 所以, 如何在保证算法预测性能的同时降低模型复杂度是本文后续研究的重要工作之一.

表 4 冷启动实验模型性能表现

模型	K = 20%		K = 40%		K = 60%	
	Ctriteo	Avazu	Ctriteo	Avazu	Ctriteo	Avazu
LR	0.7860±0.0003	0.7501±0.0003	0.7743±0.0001	0.7455±0.0011	0.7561±0.0001	0.7369±0.0012
FM	0.7951±0.0001	0.7693±0.0003	0.7873±0.0006	0.7664±0.0008	0.7696±0.0004	0.7481±0.0008
FFM	0.7964±0.0005	0.7698±0.0007	0.7852±0.0004	0.7676±0.0008	0.7701±0.0007	0.7454±0.0004
DCN	0.8019±0.0009	0.7788±0.0005	0.7910±0.0003	0.7772±0.0004	0.7742±0.0007	0.7489±0.0005
Wide & Deep	0.7957±0.0008	0.7748±0.0003	0.7861±0.0007	0.7728±0.0004	0.7684±0.0008	0.7476±0.003
AFM	0.7956±0.0008	0.7541±0.0010	0.7764±0.0010	0.7510±0.0008	0.7596±0.0012	0.7428±0.0005
AutoInt	0.8022±0.0007	0.7782±0.0007	0.7921±0.0005	0.7480±0.0008	0.7712±0.0009	0.7482±0.0005
xDeepFM	0.8014±0.0001	0.7718±0.0004	0.7909±0.0008	0.7702±0.0009	**0.7746±0.0005**	0.7515±0.0006
Fibi	0.7987±0.0004	0.7845±0.0008	0.7896±0.0004	0.7812±0.0002	0.7703±0.0005	0.7507±0.0005
AVDAE(本文)	**0.8077±0.0004**	**0.7866±0.0003**	**0.7964±0.0006**	**0.7856±0.0008**	0.7703±0.0002	**0.7786±0.0005**

参考文献

1 Marz N, Warren J. Big data: Principles and Best Practices of Scalable Realtime Data Aystems. Greenwich, USA: Manning Publications Co, 2015.

2 Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6): 734 –749. [doi: 10.1109/TKDE.2005.99]

3 McMahan HB, Holt G, Sculley D, *et al*. Ad click prediction: A view from the trenches. Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Chicago, IL, USA. 2013. 1222–1230.

4 He XR, Pan JF, Jin O, *et al*. Practical lessons from predicting clicks on ads at facebook. Proceedings of the 8th International Workshop on Data Mining for Online Advertising. New York, NY, USA. 2014. 1–9.

5 Graepel T, Candela JQ, Borchert T, *et al*. Web-scale Bayesian click-through rate prediction for sponsored search advertising in Microsoft's bing search engine. Proceedings of the 27th nternational Conference on Machine Learning. Haifa, Israel. 2010. 13–20.

6 Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. Computer, 2009, 42(8): 30 –37. [doi: 10.1109/MC.2009.263]

7 Juan YC, Zhuang Y, Chin WS, *et al*. Field-aware factorization machines for CTR prediction. Proceedings of the 10th ACM Conference on Recommender Systems. Boston, MA, USA. 2016. 43–50.

8 Juan Y, Lefortier D, Chapelle O. Field-aware factorization machines in a real-world online advertising system. Proceedings of the 26th International Conference on World Wide Web Companion. Perth, WA, Australia. 2017. 680–688.

9 He KM, Zhang XY, Ren SQ, *et al*. Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770–778.

10 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Siem Reap, Combadia. 2012. 1097–1105.

11 Cho K, Van Merriënboer B, Gulcehre C, *et al*. Learning phrase representations using RNN encoder-decoder for statistical machine translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar. 2014. 1724–1734.

12 Mikolov T, Karafiát M, Burget L, *et al*. Recurrent neural network based language model. Proceedings of the 11th Annual Conference of the International Speech Communication Association. Makuhari, Japan. 2010. 1045–1048.

13 Zhang WN, Du TM, Wang J. Deep learning over multi-field categorical data. Proceedings of the 38th European Conference on Information Retrieval. Padua, Italy. 2016. 45–57.

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 Main Track. Melbourne, Australia. 2017. 3119–3125.

15 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, MA, USA. 2016. 7–10.

16 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, Australia. 2017. 1725–1731.

17 Bengio Y, Lamblin P, Popovici D, *et al*. Greedy layer-wise training of deep networks. Proceedings of the Advances in Neural Information Processing Systems. Vancouver, BC, Canada. 2007. 153–160.

18 Verbert K, Manouselis N, Ochoa X, *et al*. Context-aware recommender systems for learning: A survey and future challenges. IEEE Transactions on Learning Technologies, 2012, 5(4): 318–335. [doi: 10.1109/TLT.2012.11]

19 Mooney RJ, Roy L. Content-based book recommending using learning for text categorization. Proceedings of the 5th ACM Conference on Digital Libraries. San Antonio, TX, USA. 2000. 195–204.

20 Breese JS, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence. Madison, WI, USA. 1998. 43–52.

21 Balabanović M, Shoham Y. Fab: Content-based, collaborative recommendation. Communications of the ACM, 1997, 40(3): 66–72. [doi: 10.1145/245108.245124]

22 Wang X, He X, Nie L, *et al*. Connecting social regularization. Proceeding of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Tokyo, Japan. 2017. 185–194.

23 Liu Q, Wu S, Wang DY, *et al*. Context-aware sequential recommendation. Proceedings of the 2016 IEEE 16th International Conference on Data Mining. Barcelona, Spain. 2016. 1053–1058.

24 He XN, DuXY, Wang X, *et al*. Outer product-based neural collaborative filtering. arXiv preprint arXiv:1808.03912, 2018.

25 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, Australia. 2015. 1743–1746.

26 Liu B, Tang RM, Chen YZ, *et al*. Feature generation by convolutional neural network for click-through rate prediction. The World Wide Web Conference. San Francisco, CA, USA. 2019. 1119–1129.

27 Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature, 1986, 323(6088): 533–536. [doi: 10.1038/323533a0]