

基于深度神经网络的点击率预测模型

刘弘历, 武 森, 魏桂英[✉], 李 新, 高晓楠

北京科技大学经济管理学院, 北京 100083

✉通信作者, E-mail: weigy@manage.ustb.edu.cn

摘 要 针对现有深度神经网络点击率预测模型在对用户偏好建模时, 难以有效且高效地处理用户行为序列的问题, 提出长短期兴趣网络(Long and short term interests network, LSTIN)模型, 充分利用用户历史记录上下文信息和顺序信息, 提升点击率预测精准性和训练效率. 使用基于注意力机制的 Transformer 和激活单元结构完成用户长、短期兴趣建模, 对用户短期兴趣进一步使用循环神经网络(Recurrent neural network, RNN)、卷积神经网络(Convolutional neural networks, CNN)进行处理, 最后使用全连接神经网络进行预测. 在亚马逊公开数据集上开展实验, 将提出的模型与基于分解机的神经网络(DeepFM)、深度兴趣网络(Deep interest network, DIN)等点击率预测模型对比, 结果表明提出的模型实现了考虑上下文信息和顺序信息的用户历史记录建模, 接受者操作特征曲线下面积(Area under curve, AUC)指标为 85.831%, 相比于基础模型(BaseModel)提升 1.154%, 相比于 DIN 提升 0.476%. 且因区分用户长、短期兴趣, 模型能够在提升预测精准性的同时保障训练效率.

关键词 点击率预测; 长短期兴趣网络; 深度神经网络; 注意力机制; 循环神经网络; 卷积神经网络

分类号 TP183

Click-through rate prediction model based on a deep neural network

LIU Hong-li, WU Sen, WEI Gui-ying[✉], LI Xin, GAO Xiao-nan

School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China

✉ Corresponding author, E-mail: weigy@manage.ustb.edu.cn

ABSTRACT The click-through rate (CTR) prediction task is to estimate the probability that a user will click an item according to the features of user, item, and contexts. At present, CTR prediction has become a common and indispensable task in the field of e-commerce. Higher accuracy of CTR prediction results conduces to present more accurate and personalized results for recommendation systems and search engines to increase users' actual CTR of items and bring more economic benefits. More researchers used a deep neural network (DNN) to solve the CTR prediction problem under the background of big data technology in recent years. However, there are a few models that can process time series data and fully consider the context information of users' history effectively and efficiently. CTR prediction models based on a DNN learn users' interests from their history; however, most of the existing models regard user interest, ignoring the differences between the long-term and short-term interests. This paper proposes a CTR prediction model named Long- and Short-Term Interest Network (LSTIN) to fully use the context information and order information of user history records. This use will help improve the accuracy and training efficiency of the CTR prediction model. Based on the attention mechanism, the transformer and activation unit structure are used to model long-term and short-term user interests. The latter is processed using the recurrent and convolutional neural networks further. Eventually, a fully-connected neural network is applied for prediction. Different from DeepFM and Deep Interest Network (DIN) in experiments on an Amazon public dataset, LSTIN achieves modeling with context and order information of user history. The AUC of LSTIN is 85.831%, which is 1.154% higher than that of BaseModel and 0.476% higher than that of DIN. Besides, LSTIN achieves distinguishing the long-term and short-term interests of users, which improves the

收稿日期: 2021-03-23

基金项目: 国家自然科学基金资助项目(71971025)

performance and maintains the training efficiency of the CTR prediction model.

KEY WORDS CTR prediction; long and short term interest network; deep neural network; attention mechanism; RNN; CNN

在信息时代,庞大的数据总量一方面加剧了人工筛选信息的难度,导致信息过载^[1],另一方面也为大数据技术的发展提供了基础。随着大数据技术逐步在各个领域展开应用并取得了极佳的效果^[2],用大数据技术解决信息过载问题成为学术界和工业界关注的问题之一。推荐系统、搜索引擎等基于信息检索的应用能够一定程度上解决信息过载问题,点击率(Click-through-rate, CTR)预测则在这类应用中提供支持与保障^[3]。CTR 预测^[4]根据用户、项目及上下文的特征,预测用户点击该项目的概率,帮助搜索引擎、推荐系统等向用户展示更加个性化、更加精准的内容。

解决 CTR 预测问题的模型可以分为传统机器学习模型和深度学习模型^[5]。传统机器学习模型中逻辑回归^[6]思路简单、易实现分布式计算,是该领域基础模型。但它作为线性模型,难以捕获高维非线性特征^[7]。Rendle 提出因子分解机(Factorization machine, FM)^[8]用以解决特征组合问题,FM 监督学习不同特征的隐因子向量,用内积完成特征交叉,但模型复杂度高,一般只用于 2 阶特征交叉。

深度神经网络(Deep neural network, DNN)可以很好地提取特征之间的非线性关系,使用大量数据完成模型训练,在图像识别、自然语言处理等领域已经得到广泛应用,将 DNN 应用于 CTR 预测的研究越来越受到关注。基于 DNN 的 CTR 预测模型大多基于“嵌入+多层感知机(Embedding+MLP)”范式。如 Zhang 等提出 FNN(Factorization-machine supported neural network)^[9],将 FM 的输出拼接后作为全连接层的输入,每个特征的嵌入向量预先采用 FM 模型训练得到,DNN 训练开销低,收敛速度快。但该模型难以克服 FM 的计算复杂度限制,同时没有对低阶特征建模。Qu 等提出的 PNN(Product-based neural networks)模型^[10]使用外积、内积实现特征交叉,在 DNN 的基础上使用特征提取层,完成高阶特征组合的同时融入低阶特征,无需进行两阶段训练。Cheng 等提出的 Wide&Deep 模型^[11]将传统的逻辑回归和 DNN 以并行方式组合,充分发挥了逻辑回归模型的拟合能力和 DNN 的泛化能力。但上述 DNN 模型依赖于人工特征工程。Guo 等提出的 DeepFM 模型^[12]不需要人工特征工程,它令 FM 与 DNN 在嵌入层

共享,同时学习高低阶特征组合。

以上 DNN 模型都为推动 CTR 预测的进一步发展做出了重要贡献,但均将用户兴趣视为整体。现实中用户的兴趣可能随着时间发生变化,如何通过用户历史点击行为探索其兴趣变化趋势,进而准确预测该用户对待预测项目的点击概率,成为一个需要解决的问题。自然语言处理领域广泛使用的“注意力机制”为解决该问题提供了有益启示。Zhou 等提出的 DIN(Deep interest network)模型^[13]使用注意力机制设计激活单元,针对不同待预测项目自适应调整赋予每条历史记录权重,提升了嵌入层的表达和感知能力。该模型将注意力机制应用于 CTR 领域的候选项目感知,但该模型在捕捉用户历史记录间的相对顺序时面临困难,难以关注某一记录的上下文信息。

自然语言处理领域处理词语和句子的方式对提取用户历史点击行为的上下文信息和顺序信息,建模用户长、短期兴趣两个问题具有重要参考意义。自然语言处理领域的 DNN 模型中,基于循环神经网络(Recurrent neural network, RNN)的模型和基于卷积神经网络(Convolutional neural networks, CNN)的模型是两个重要类别。传统 RNN 模型以序列数据作为输入,节点间按链式结构串联,但在实际应用中受限于梯度消失。长短期记忆网络(Long short-term memory, LSTM)^[14]能够通过门控在一定程度上解决梯度消失问题^[15],但其结构复杂,训练所需时间较长。门控循环单元(Gated recurrent unit, GRU)^[16]使用更简单的结构取得了与 LSTM 接近的效果,在自然语言处理等领域被广泛应用。CNN 使用卷积和池化操作,具有很强的语义特征捕捉能力^[17],结构合理的 CNN 在自然语言处理任务中已经表现出良好的性能^[18],但 CNN 偏向关注相邻信息,容易忽略局部与整体信息的关联性。谷歌研究团队的 Vaswani 等提出 Transformer 模型^[19]在机器翻译任务中使用自注意力和注意力机制,显著提升了训练速度的同时取得了更好的效果。

本文提出长短期兴趣网络(Long and short term interest network, LSTIN)点击率预测模型,解决 CTR 预测中基于用户历史记录建模时上下文信息和顺序信息难以高效提取和充分利用的问题,提

升点击率预测精准性,并通过对用户长、短期兴趣分别建模进一步提高训练效率.具体地,首先借鉴 Transformer 模型的自注意力机制,提取用户行为上下文信息,并使用基于注意力机制的激活单元结构建立用户行为与待预测项目之间的关系,构建兴趣提取层.接着,构建信息融合层,将用户历史行为序列划分为长期和短期,对短期序列中的项目使用 GRU 进一步处理.在亚马逊公开数据集上开展实验,结果表明 LSTIN 模型与对比模型相比,获得了更准确的点击率预测结果.此外,本文对比了在短期序列项目处理时使用 CNN 和 GRU 的效果,信息融合层使用 CNN 的 LSTIN 与使用 GRU 的 LSTIN 效果相近,训练效率更高.

本文所提出的 LSTIN 模型主要有以下创新点:

(1) 针对 CTR 预测问题设计结构,借鉴自然语言处理领域模型完成序列数据建模,使用激活单元结构建立用户历史行为与待预测项目的联系.

(2) 模型大量使用以注意力机制为基础的并行结构,训练效率较高.

(3) LSTIN 对同一用户长期和短期序列采用不同处理方式,使模型能够对用户长期兴趣和短期兴趣分别建模.

(4) LSTIN 在信息融合层可根据需要选用不同方法,具备灵活性和可拓展性.

1 相关定义和模型结构

1.1 长、短期序列定义

用户历史点击行为能够体现其特征和兴趣.将用户的历史点击序列定义如下,并在此基础上提出长、短期序列定义.

(1) 历史点击序列.

对于某一特定用户,按照时间顺序排列的历史点击项目组成该用户的历史点击序列.其中最近 N 个项目定义为历史点击序列的短期部分.

给定共有 n 条历史点击行为的用户 u , 记其历史点击序列为 $\mathbf{hs}_u$, t_i 表示第 i 条点击行为发生的时刻,该时刻被点击的项目记作 $\mathbf{c}_{t_i}$, 则用户历史点击序列可表示为式 (1). 选取短期部分长度为 l ($l \leq i$), 可以表示为式 (2). “项目”在数据集中即为商品.

$$\mathbf{hs}_u = \{\mathbf{c}_{t_1}, \mathbf{c}_{t_2}, \dots, \mathbf{c}_{t_n} | t_{k-1} < t_k, k \in [2, n]\} \quad (1)$$

$$\mathbf{rs}_u = \{\mathbf{c}_{t_{i-l+1}}, \mathbf{c}_{t_{i-l+2}}, \dots, \mathbf{c}_{t_i} | t_{k-1} < t_k, k \in [i-l+2, i]\} \quad (2)$$

(2) 长、短期序列.

长期序列: 由用户历史点击序列 $\mathbf{hs}_u$ 通过编码器和激活单元结构(即后文中的兴趣提取层)处理

后得到的向量序列 $\mathbf{hs}'_u$ 为长期序列.

短期序列: 短期部分 $\mathbf{rs}_u$ 通过编码器和激活单元结构(即后文中的兴趣提取层)处理后得到的向量序列 $\mathbf{rs}'_u$ 为短期序列.

1.2 模型结构

LSTIN 的模型结构如图 1 所示,该模型主要由五部分构成,从下至上依次为输入层(Input Layer)、嵌入层(Embedding Layer)、兴趣提取层(Interests Extraction Layer)、信息融合层(Information fusion layer)、评分层(Scoring layer). 输入数据自下而上经过输入层映射为独热编码(one-hot)向量,经过嵌入层映射为低维向量,通过兴趣提取层建模后按点击行为发生时间将序列划分为长期和短期两部分,分别在信息融合层合并为固定长度的向量用以表示用户的长、短期兴趣,将这两个向量与表示候选项目的嵌入向量拼接(Concat)后输入评分层,最终输出一个评分 $\text{score} \in \{0, 1\}$ 表示预测用户点击候选项目的概率. 兴趣提取层和信息融合层是 LSTIN 模型提取用户兴趣、组合用户长短期兴趣的核心.

(1) 输入层和嵌入层.

特征的选取对于模型的预测精度会产生显著影响^[20],为验证模型结构的有效性,减少特征工程的影响,选择项目 ID 和项目所属类别的 ID 作为输入,统一使用 one-hot 编码后通过嵌入层映射到低维、稠密的向量空间中并进行拼接.以项目 ID 为例,设该特征有 k 种取值,其第 i 项取值的 one-hot 向量 $\mathbf{o}_{i_id}^i$ 与权重矩阵 $\mathbf{W}_{i_id}^{\text{emb}}$ 进行点乘得到嵌入结果,以特征取第 3 个可能取值为例,其嵌入向量记为 $\mathbf{i_id}^i$,原理如式 (3) 所示, d 表示希望得到的低维向量维度,0 的不同下标只为说明其所在位置.

$$\mathbf{i_id}^i = \mathbf{o}_{i_id}^i \cdot \mathbf{W}_{i_id}^{\text{emb}} = [0_1, 0_2, 1, 0_4, 0_5, \dots, 0_k] \cdot \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1d} \\ w_{21} & w_{22} & \dots & w_{2d} \\ \dots & \dots & \dots & \dots \\ w_{k1} & w_{k2} & \dots & w_{kd} \end{bmatrix} = [w_{31}, w_{32}, \dots, w_{3d}] \quad (3)$$

用户历史点击序列 $\mathbf{hs}_u$ 表示若干个项目按点击行为发生时间排序后的序列,序列中每个向量均由表示该项目的 ID 和其类别 ID 两部分向量拼接而成.用户 u 的历史点击序列 $\mathbf{hs}_u$ 可表示为式 (4).

$$\mathbf{hs}_u = \{(\mathbf{i_id}_{t_1}, \mathbf{c_id}_{t_1}), \dots, (\mathbf{i_id}_{t_n}, \mathbf{c_id}_{t_n}) | t_{k-1} < t_k, k \in [2, n]\} \quad (4)$$

(2) 兴趣提取层.

兴趣提取层可以细分为一个编码器(Encoder)部分和一个激活单元(Activation unit)部分.

编码器部分的结构借鉴 Transformer, 其中自

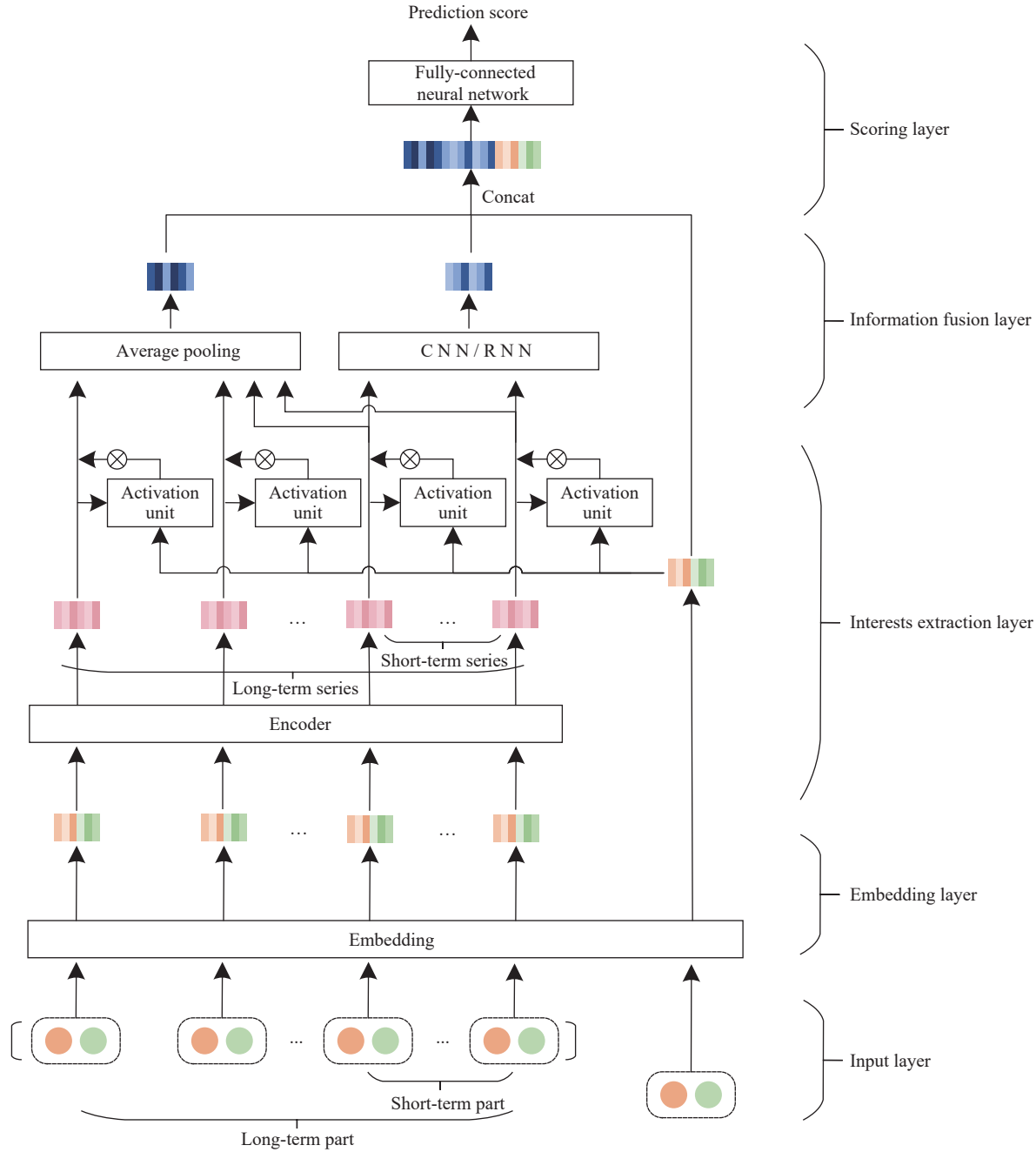


图1 LSTIN 模型结构
Fig.1 Structure of an LSTIN

注意力(Self-attention)机制能够关注上下文信息和顺序信息,并且高效地完成长序列建模,克服梯度消失的问题.编码器结构如图2所示.

通过输入层和嵌入层后,用户历史点击序列已经被映射为固定长度向量组成的序列,序列中每一个向量表示一次点击行为.编码器部分的作用是使得序列中的每个向量不仅包含本次点击行为的信息,也包含其在整个序列中的位置信息和上下文信息.

首先对得到的序列进行位置编码(Positional

encoding),并将位置编码得到的向量序列与原序列相加.采取正余弦位置编码^[19],如式(5)所示,将向量序列第 p 个位置映射为一个维度是 d_{pos} 的向量,该向量的第 k 个值记作 $\text{PE}_k(p)$.应用中,一般令 d_{pos} 和嵌入层得到的低维向量维度 d 相等.

$$\begin{cases} \text{PE}_{2i}(p) = \sin(p/10000^{2i/d_{\text{pos}}}) \\ \text{PE}_{2i+1}(p) = \cos(p/10000^{2i/d_{\text{pos}}}) \end{cases} \quad (5)$$

对上述结构输出的数据 $\mathbf{hs}_u^p$ 使用自注意力机制,使模型感知序列中某个向量的上下文信息,思路是基于每一个向量分别计算应该赋予序列中每

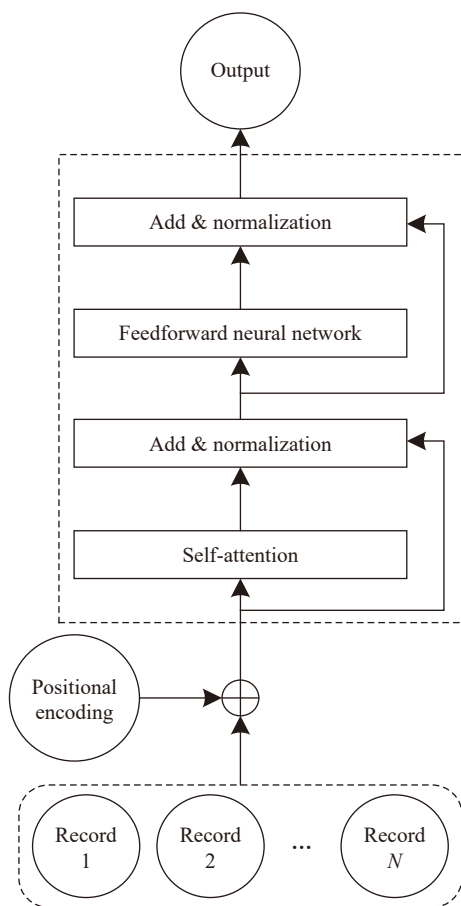


图2 编码器结构

Fig.2 Structure of an encoder

个向量的权重并进行加权求和,该权重可视为相似度,显然某特定向量与自身相似度最高,所以自身对应权重最大,其他向量与其相似度越高权重越大。对于处理后得到的向量序列,原序列某一向量所在位置的新向量包含原向量的信息,也包含其他位置向量的信息,即上下文信息。本部分使用基于缩放点积的自注意力机制,原理如式(6)所示,在本式中 d_k 是向量序列 $\mathbf{hs}_u^p$ 中的向量维度,为防止点乘结果过大,通过除以 $\sqrt{d_k}$ 进行缩放,softmax是归一化指数函数。

$$\text{Attention}(\mathbf{hs}_u^p, \mathbf{hs}_u^p) = \text{softmax}\left(\frac{\mathbf{hs}_u^p (\mathbf{hs}_u^p)^T}{\sqrt{d_k}}\right) \mathbf{hs}_u^p \quad (6)$$

自注意力机制的输出与其输入进行残差连接后进行归一化(Add & Normalization),为增加模型的非线性能力,将数据继续输入一个前馈神经网络(Feedforward neural network)中。残差连接的作用是缓解梯度消失的影响,避免模型退化。前馈神经网络的输出与输入仍然进行残差连接并归一化,结果即为编码器部分的输出 $\mathbf{E_out}$ 。该结果随后进入激活单元部分。

在编码器部分,可视为模型感知了每一次点击行为的上下文信息,在激活单元部分,模型进一步感知待预测候选项目信息。该部分仍使用注意力机制,思想与编码器部分注意力机制相似,但计算过程不同,将候选项目向量 $\mathbf{v}_c = (\mathbf{i_id}_{t_{n+1}}, \mathbf{c_id}_{t_{n+1}})$ 复制,并将其与编码器部分的输出 $\mathbf{E_out}$ 及二者的差三个矩阵拼接后输入三层全连接神经网络,如图3。该结构的目的是学习每个历史项目与候选项目的关系。

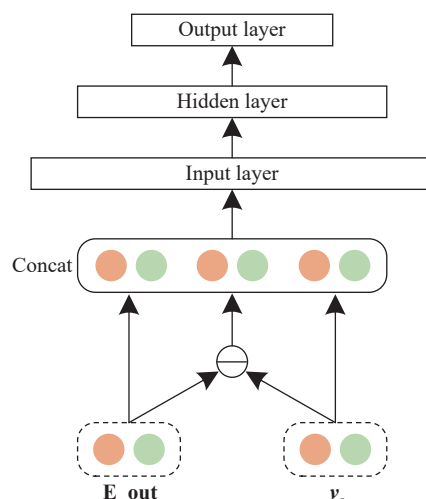


图3 激活单元注意力机制原理示意图

Fig.3 Attention mechanism in an activation unit

(3) 信息融合层。

在本层,此前结构输出的向量序列被分为长期和短期序列,对长期序列 $\mathbf{hs}_u'$ 建模关注用户长期以来相对稳定的兴趣,不同用户历史点击序列长度不同,为保证后续结构接收到的数据结构统一,采用平均池化的方式处理为定长向量,该结构输出的向量记作 $\mathbf{h}$ 。学习短期兴趣时,进一步使用GRU处理短期序列 $\mathbf{rs}_w'$,输出单个的定长向量 $\mathbf{r}$ 。

此外,针对短期序列建模,考虑将GRU替换为一维CNN。CNN最初为解决图像识别问题^[21],后被应用在多个领域,自然语言处理任务常用一维CNN^[22],它具有并行度高、提取局部信息能力强的特点。借鉴Kim^[23]及Zhang与Wallace^[24]将CNN用于自然语言处理的经验,设置LSTIN中的CNN网络第一层为卷积层,卷积核的长度为2,宽度为输入向量宽度,滤波器个数为100,输出矩阵的每一列包含一个滤波器的权值。第二层为平均池化层,第三层为与第一层结构相同的卷积层。最后,进行全局池化并抛弃一定数量的输出防止过拟合。

(4) 评分层和损失函数。

至此得到了表征用户长、短期兴趣的向量和

表征候选项目的向量, 将三个向量拼接后输入评分层. 评分层采取三层全连接神经网络, 含输入层、隐藏层和输出层. 全连接层输入层的处理过程可以表示为式 (7), 式 (7)~式 (9) 中的下标 1、2、3 分别表示变量位于全连接层神经网络的第几层. 其中, $[h, r, v_c]$ 由表征长期兴趣的向量、短期兴趣的向量和候选物品的向量拼接得到. 在式 (7) 中, W_1 为输入层的权重矩阵, b_1 是偏置向量, l_{f1} 是输入层神经元的数量. $o_1 \in \mathbf{R}^{1 \times l_{f1}}$ 是输入层的输出.

$$o_1 = \sigma([h, r, v_c]W_1 + b_1) \quad (7)$$

全连接层隐藏层原理与输入层相同, 输出层使用 softmax 函数, 如式 (8), 最终需要得到点击或不点击的二分类结果. softmax 函数进行归一化, o_3 第一维表示用户不会点击该候选项目的概率, 第二维则表示用户会点击该候选项目的概率. 随后通过 argmax 函数选择 o_3 中最大的值并输出其索引, 如式 (9) 和式 (10) 所示. 最终 $\hat{y} \in \{0, 1\}$ 表示模型预测用户是否会点击该候选项目, 0 表示不会点击, 1 表示会点击.

$$o_3 = \text{softmax}(o_2W_3 + b_3) \quad (8)$$

$$\hat{y} = \text{argmax}(o_3) \quad (9)$$

$$\text{argmax}_{f(x)} = \{x | \forall x: f(y) \leq f(x)\} \quad (10)$$

预测模型的目标函数为最大似然估计的对数损失 (Logloss), 如式 (11) 所示. N 是训练样本的样本总数, $y_i \in \{0, 1\}$ 代表用户的真实点击, 取值含义与 x_i 相同. p_i 为预测输入实例属于类别 1 的概率.

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \ln p_i + (1 - y_i) \ln (1 - p_i)] \quad (11)$$

2 实验设置

2.1 数据集、优化器和超参数设置

实验在亚马逊产品数据集 (Amazon dataset) 上开展, 该数据集包含来自用户对产品的评论数据和产品自身的属性数据, 是 CTR 预测领域常用的开源数据集, 如表 1 所示. 实验在其中一个电子类商品子集 (Electronics 5) 上进行, 该子集包含 19.24 万名用户、6.3 万项商品、801 个类别和 168.92 万条行为记录. 每个用户和商品都有 5 条以上的评分. DIN 模型曾使用本数据集验证效果, 在本数据集上进行实验有助于更加直观地对比模型效果. 数据集仅包含正样本, 随机抽取与正样本数量相同的负样本, 选取每个用户行为序列中最后一条样本作为测试集, 其他样本作为训练集.

表 1 数据集统计信息

Data set	Number of users	Number of commodities	Number of categories	Number of samples
Amazon (Electronics)	192403	63001	801	1689188

使用 Adam 优化器^[25]; 学习率初始为 0.001, 并且设置学习率随着迭代次数增加而衰减; 训练和测试的批尺寸 (Batch size) 设置为 128. 为公平起见, 本文模型和对比模型在嵌入维度上保持一致, 均为 32; 全连接层每层的神经网络单元数分别为 $4 \times 32, 1 \times 32, 1$. 短期序列长度设置为固定值 5.

2.2 指标和对比方法

AUC (Area under curve) 是 CTR 预测领域常用效果度量方法^[26], 将样本根据预测点击率由大到小排列, 随机取出一个正样本和一个负样本, 正样本排在负样本前的概率即为 AUC. 计算公式如式 (12) 所示, x_1 和 x_2 表示任取两个样本, X_+ 和 X_- 分别表示正、负样本的集合, N_+ 和 N_- 分别代表正负样本的数量, $\text{rank}(i)$ 指样本 i 的预测点击率的排名.

$$\begin{aligned} \text{AUC} &= P\{x_1, x_2 | \text{rank}(x_1) < \text{rank}(x_2) \cap x_1 \subseteq X_+ \cap x_2 \subseteq X_-\} \\ &= \frac{\sum_{i=1}^{N_+} \text{rank}(i) - \frac{(N_+ + 1)N_+}{2}}{N_+N_-} \end{aligned} \quad (12)$$

对比模型选用 BaseModel、DeepFM 和 DIN. BaseModel 是 CTR 预测领域的基础 DNN 模型, 是最简单的 Embedding+MLP 范式模型^[13], 遵循该范式的模型与之对比可以清晰地展示出结构改变后的效果提升. DeepFM 是 FM 的衍生算法之一, 提出后便成为 CTR 预测研究领域经典算法之一. DIN 成功将注意力机制应用到 CTR 预测中, 建立了候选项目与用户历史记录的联系, 与 DIN 对比可以得出 LSTIN 模型扩大模型感知范围并赋予模型序列数据建模能力后的效果提升.

为验证区分用户长短期兴趣对最终预测结果和训练效率的影响, 首先构建不区分用户长、短期兴趣的模型, 在兴趣提取层使用注意力和自注意力机制建立用户历史记录与候选项目和自身上下文之间的联系, 信息融合层使用平均池化的记作 LIN, 信息融合层使用 GRU 的记作 LIN_g, 使用 CNN 的记作 LIN_c. 接着, 对于分别考虑用户长、短期兴趣的 LSTIN 模型, 在信息融合层, 长期序列建模采用平均池化, 依据短期序列建模采用的方法, 将短期序列建模使用 GRU 和 CNN 的模型分别记作 LSTIN_g 和 LSTIN_c.

3 实验结果

对每个模型重复 5 次训练, 取每次训练最高 AUC, 5 次训练平均值记为该模型 AUC. Best AUC 是 5 次训练中 AUC 最高值. 记 $RP(\text{BaseModel})$ 为某一模型 AUC 对 BaseModel 的相对提升, 计算方式如式 (13) 所示, $AUC(\text{mesuared model})$ 表示测

试模型的 AUC 指标. 同理, 记 $RP(\text{DIN})$ 为某一模型 AUC 对 DIN 的相对提升, 结果如表 2 和图 4 所示.

$$RP(\text{BaseModel}) = \frac{AUC(\text{mesuared model}) - AUC(\text{BaseModel})}{AUC(\text{BaseModel})} \times 100\% \quad (13)$$

表 2 算法性能对比

Table 2 Algorithm performance

Category	Name	AUC/%	Best AUC/%	RP(BaseModel)/%	RP (DIN)/%
Existing models	BaseModel	84.852	84.946	0	-0.670
	DeepFM	85.012	85.095	0.189	-0.482
	DIN	85.424	85.468	0.674	0
Model of this paper	LIN	85.581	85.669	0.859	0.184
	LIN_c	84.803	84.911	-0.058	-0.727
	LIN_r	85.796	85.841	1.113	0.435
	$LSTIN_c$	85.781	85.834	1.095	0.418
	$LSTIN_r$	85.831	85.943	1.154	0.476

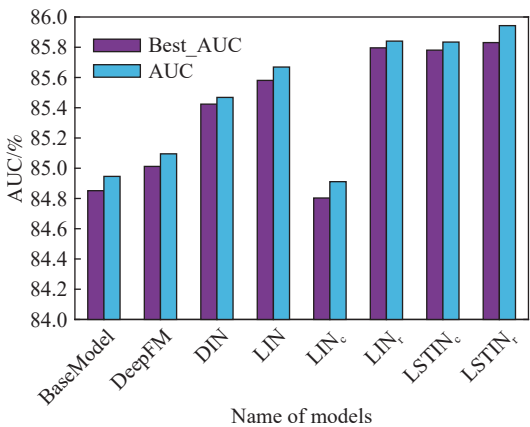


图 4 算法 AUC 对比

Fig.4 AUC of various algorithms

LIN 相比 DIN, 结构上的差异只是在信息融合层前加入编码器结构, 因此 LIN 的 AUC 指标高于 DIN 可以证明模型获得感知上下文信息的能力后能够取得更好的效果. LIN_r 效果进一步提高, 证明 GRU 结构相比平均池化能够更充分地提取和保留信息, 使模型获得更强的学习和表示能力, 但 LIN_c 效果不佳, 效果低于 BaseModel. 在对长期和短期兴趣采取不同的处理方式后, $LSTIN_r$ 和 $LSTIN_c$ 均取得了良好的效果. $LSTIN_c$ 相比 LIN_c 的效果提升, 可进一步验证 CNN 更加偏向关注局部信息, 适用于处理长度固定的短序列. 平均 AUC 和最高 AUC 的最大值均由 $LSTIN_r$ 取得.

图 5 对比了部分模型在达到最佳效果前 AUC 随

时间的变化. 从 AUC 的角度看, 性能最佳的三个模型为 LIN_r 、 $LSTIN_r$ 、 $LSTIN_c$, LIN_r 尽管性能较好, 但需耗费超过 60000 s 的训练时间取得最高 AUC, 此外 LIN_r 将用户全部历史数据通过 GRU 处理, 需要占用较多的内存资源. $LSTIN$ 通过区分用户长、短期兴趣, 对用户长、短期兴趣采取不同的处理方式, 实现训练效率提升的同时保障并提升模型效果. 在相同的实验环境下, $LSTIN_r$ 和 $LSTIN_c$ 均在 11000 s 以内达到最高 AUC, $LSTIN_c$ 略快于 $LSTIN_r$. 从模型效果及训练效率两个角度综合来看, 将用户历史点击行为分为长期和短期两部分都是十分有价值的.

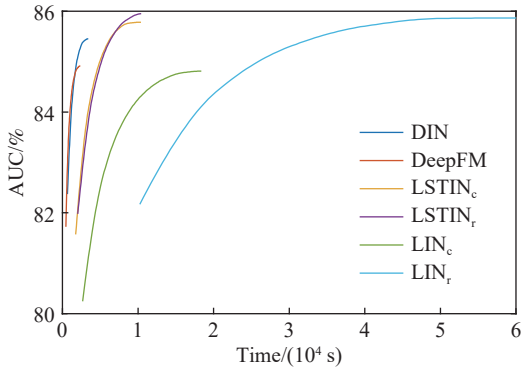


图 5 部分模型训练时间对比

Fig.5 Training time comparisons

4 结论

(1) 针对点击率预测问题中, 建模用户兴趣时

用户历史记录上下文信息和顺序信息难以有效提取的问题, 提出长短期兴趣网络模型, 使用基于注意力机制的结构解决了以上问题。

(2) 针对序列数据建模方法效率较低的问题, 将用户行为分为长期和短期两部分, 分别采用平均池化、RNN/CNN 进一步处理, 保障模型效果的同时提高了模型训练效率。

(3) 基于亚马逊公开数据集的对比实验表明, LSTIN 的预测精度有明显提升。

参 考 文 献

- [1] Shen X L, Li Z J, He C H. Hybrid recommendation algorithm based on rating filling and trust information. *J Comput Appl*, 2020, 40(10): 2789
(沈学利, 李子健, 赫辰皓. 基于评分填充与信任信息的混合推荐算法. 计算机应用, 2020, 40(10): 2789)
- [2] Song H Q, Du S Y, Zhou Y C, et al. Big data intelligent platform and application analysis for oil and gas resource development. *Chin J Eng*, 2021, 43(2): 179
(宋洪庆, 都书一, 周园春, 等. 油气资源开发的大数据智能平台及应用分析. 工程科学学报, 2021, 43(2): 179)
- [3] Tao Z L, Wang X, He X N, et al. HoAFM: A high-order attentive factorization machine for CTR prediction. *Inf Process Manag*, 2019, 57(6): 102076
- [4] Zhou A Y, Zhou M Q, Gong X Q. Computational advertising: A data-centric comprehensive web application. *Chin J Comput*, 2011, 34(10): 1805
(周傲英, 周敏奇, 宫学庆. 计算广告: 以数据为核心的Web综合应用. 计算机学报, 2011, 34(10): 1805)
- [5] Liu M J, Zeng G C, Yue W, et al. Review on click-through rate prediction models for display advertising. *Comput Sci*, 2019, 46(7): 38
(刘梦娟, 曾贵川, 岳威, 等. 面向展示广告的点击率预测模型综述. 计算机科学, 2019, 46(7): 38)
- [6] Richardson M, Dominowska E, Ragno R. Predicting clicks: Estimating the click through rate for new ADs // *Proceedings of the 16th International Conference on World Wide Web*. Alberta, 2007: 521
- [7] Chen J X, Sun B G, Li H, et al. Deep CTR prediction in display advertising // *Proceedings of the 24th ACM international conference on Multimedia*. Amsterdam, 2016: 811
- [8] Rendle S. Factorization Machines // *2010 IEEE International Conference on Data Mining*. Berlin, 2010: 995
- [9] Zhang W N, Du T M, Wang J. Deep learning over multi-field categorical data: A case study on user response prediction // *Proceedings of European Conference on Information Retrieval*. Padua, 2016: 45
- [10] Qu Y R, Cai H, Ren K, et al. Product-based neural networks for user response prediction // *2016 IEEE 16th International Conference on Data Mining*. Barcelona, 2016: 1149
- [11] Cheng H T, Koc L, Harmsen J, et al. Wide & Deep learning for recommender systems // *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*. Boston, 2016: 7
- [12] Guo H F, Tang R M, Ye Y M, et al. DeepFM: A factorization-machine based neural network for CTR prediction // *Twenty-Sixth International Joint Conference on Artificial Intelligence*. Melbourne, 2017: 1725
- [13] Zhou G R, Song C R, Zhu X Q, et al. Deep interest network for click-through rate prediction // *Proceedings of KDD International Conference on Knowledge Discovery & Data Mining*. London, 2018: 1059
- [14] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computat*, 1997, 9(8): 1735
- [15] Cheng Y, Yao L B, Zhang G H, et al. Text sentiment orientation analysis of multi-channels CNN and BiGRU based on attention mechanism. *J Comput Res Dev*, 2020, 57(12): 2583
(程艳, 尧磊波, 张光河, 等. 基于注意力机制的多通道CNN和BiGRU的文本情感倾向性分析. 计算机研究与发展, 2020, 57(12): 2583)
- [16] Cho K, Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation // *Proceedings of the Conference on Empirical Methods on Natural Language Processing*. Doha, 2014: 1724
- [17] Dai J H, Deng Y B. Extracting emotion-cause pairs based on emotional dilation gated CNN. *Data Anal Knowl Discov*, 2020, 4(8): 98
(代建华, 邓育彬. 基于情感膨胀门控CNN的情感-原因对提取. 数据分析与知识发现, 2020, 4(8): 98)
- [18] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences // *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Baltimore, 2014: 655
- [19] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need // *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, 2017: 6000
- [20] Jiang D, Xu R B, Xu X, et al. Multi-view feature transfer for click-through rate prediction. *Inf Sci*, 2021, 546: 961
- [21] Yu C M, Feng B L, An L. Sentiment analysis in cross-domain environment with deep representative learning. *Data Anal Knowl Discov*, 2017, 1(7): 73
(余传明, 冯博琳, 安璐. 基于深度表示学习的跨领域情感分析.

- 数据分析与知识发现, 2017, 1(7): 73)
- [22] Zhang Y F, Lu Z Q. Remaining useful life prediction based on an integrated neural network. *Chin J Eng*, 2020, 42(10): 1372
(张永峰, 陆志强. 基于集成神经网络的剩余寿命预测. 工程科学学报, 2020, 42(10): 1372)
- [23] Kim Y. Convolutional neural networks for sentence classification // *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. Doha, 2014: 1746
- [24] Zhang Y, Wallace B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *Comput Sci*, <https://arxiv.org/pdf/1510.03820.pdf>
- [25] Kingma D P, Ba J. Adam: A method for stochastic optimization // *International Conference on Learning Representations*. San Diego, 2015
- [26] Li Y J, Guo H X, Li Y N, et al. A boosting based ensemble learning algorithm in imbalanced data classification. *Syst Eng Theory Pract*, 2016, 36(1): 189
(李诒靖, 郭海湘, 李亚楠, 等. 一种基于Boosting的集成学习算法在不均衡数据中的分类. 系统工程理论与实践, 2016, 36(1): 189)