

基于细粒度特征交互选择网络的农产品推荐算法^①



白雪^{1,2}, 王霞光^{1,2}, 金继鑫¹, 宋春梅¹, 赵思彤³

¹(中国科学院 沈阳计算技术研究所, 沈阳 110168)

²(中国科学院大学, 北京 100049)

³(沈阳工业大学, 沈阳 110870)

通信作者: 王霞光, E-mail: 2310240647@qq.com

摘要: 在数字化的时代里, 越来越多人偏爱在电商平台购物, 随着农产品电商平台的发展, 消费者面对众多选择时难以找到适合自己的产品. 为了提高用户满意度和购买意愿, 农产品电商平台需要根据用户的兴趣偏好向其推荐合适的农产品. 考虑到季节、地域、用户兴趣和农产品属性等多种农业特征, 通过特征交互可以更好地捕捉用户需求. 传统的点击通过率 CTR (click through rate) 预测模型只关注用户评分, 以简单的方式计算特征交互, 而忽略了特征交互的重要性. 本文提出了一种名为细粒度特征交互选择网络 FgFisNet (fine-grained feature interaction selection networks) 的新模型. 该模型通过引入细粒度交互层和特征交互选择层, 组合内积和哈达玛积有效地学习特征交互, 然后在训练过程中自动识别重要的特征交互, 并删除冗余的特征交互, 最后将重要的特征交互和一阶特征输入到深度神经网络, 得到最终的 CTR 预测值. 在农产品电商真实数据集上进行广泛的实验, FgFisNet 方法取得了显著的经济效益.

关键词: 农产品推荐; 点击率预测; 特征交互; 特征选择; 深度神经网络

引用格式: 白雪, 王霞光, 金继鑫, 宋春梅, 赵思彤. 基于细粒度特征交互选择网络的农产品推荐算法. 计算机系统应用, 2024, 33(5): 271-279. <http://www.c-s-a.org.cn/1003-3254/9519.html>

Agricultural Product Recommendation Algorithm Based on Fine-grained Feature Interactive Selection Network

BAI Xue^{1,2}, WANG Xia-Guang^{1,2}, JIN Ji-Xin¹, SONG Chun-Mei¹, ZHAO Si-Tong³

¹(Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)

²(University of Chinese Academy of Sciences, Beijing 100049, China)

³(Shenyang University of Technology, Shenyang 110870, China)

Abstract: In the digital era, an increasing number of people prefer shopping on e-commerce platforms. With the development of agricultural product e-commerce platforms, consumers find it challenging to discover suitable products among numerous choices. To enhance user satisfaction and purchase intent, agricultural product e-commerce platforms need to recommend appropriate products based on user preferences. Considering various agricultural features such as season, region, user interests, and product attributes, feature interactions can better capture user demands. This study introduces a new model, fine-grained feature interaction selection networks (FgFisNet). The model effectively learns feature interactions using both the inner product and Hadamard product by introducing fine-grained interaction layers and feature interaction selection layers. During the training process, it automatically identifies important feature interactions, eliminates redundant ones, and feeds the significant feature interactions and first-order features into a deep neural network to obtain the final click through rate (CTR) prediction. Extensive experiments on a real dataset from agricultural e-

① 基金项目: 辽宁省应用基础研究计划 (2022JH2/101300126)

收稿时间: 2023-12-04; 修改时间: 2024-01-09; 采用时间: 2024-01-23; csa 在线出版时间: 2024-04-07

CNKI 网络首发时间: 2024-04-10

commerce demonstrate significant economic benefits achieved by the proposed FgFisNet method.

Key words: agricultural product recommendation; click through rate (CTR) prediction; feature interaction; feature selection; deep neural network (DNN)

近年来,我国着重强调乡村振兴,构建新的农业发展格局,提出以“互联网+农业”的新型组合为抓手,科技创新为动力,结合传统农业的发展方向,推动农产品电商的发展。随后,《数字乡村发展战略纲要》等文件陆续出台,强调加强农产品电子商务平台的建设,将电商服务纳入数字乡村振兴的整体战略。

随着科技的发展,农业电子商务网站的崛起成为未来带动农产品消费的新风口。数字化转型和创新是农产品电商发展的新热点^[1]。农产品电商作为农业现代化推动和提升农产品附加值的主要力量,正经历着规模扩大、多元化发展、创新趋势的时期。在政府和行业的支持下,它不仅在规模上不断增长,还呈现出规范、标准、绿色、数智、创新等多方面的发展特点。农产品电商不再受制于传统电商,而是逐渐引入直播电商、预制菜电商等新兴模式,数字化转型成为各企业的普遍目标,阿里、京东等行业领军企业在这一过程中发挥着引领作用。农产品电商的发展趋势也包括网上网下融合和跨境上行,这些趋势正塑造着未来农产品电商的新格局。

为了给用户提供更好的购物体验,电商购物平台利用推荐技术为用户提供个性化推荐,减轻商品过载的同时还为用户提供多样化的选择。农产品电商平台和通用电商平台在多个方面存在显著差异,主要表现在它们所服务的行业、商品属性以及用户需求等方面。农产品推荐通过深入分析季节性、地域性、用户兴趣和农产品属性等农业特征,以实现农业环境的精准适应,从而提供个性化、符合用户实际需求的推荐服务。相比之下,通用电商平台推荐更注重广泛的商品品类、全球化市场、用户评价和促销优惠等通用特征,致力于为用户提供全面而多样的购物选择。

在农产品推荐中,特征交互的复杂性体现在对众多农业特征的全面分析。这些特征涵盖了季节性、地域差异、用户兴趣和农产品属性等因素。通过深入的特征交互,系统能够在多个维度上进行全面而精细的分析。首先,系统利用季节性因素捕捉用户需求的变化,实现了在不同季节为用户提供最合适和受欢迎的农产

品推荐。这不仅是根据时间调整推荐内容,更是在深刻理解季节性需求的基础上进行个性化推荐,提升了系统的实用性。其次,地域差异是农产品推荐中另一个重要的考虑因素。由于不同地域的农业特点和气候条件存在差异,通过特征交互,系统更好地辨别和理解地域性特点,为用户提供更加个性化、符合实际需求的推荐服务。除此之外,用户的历史行为、兴趣和偏好等信息也是推荐效果的决定性因素。特征交互通过深度挖掘和分析这些信息,为用户提供更符合个体差异的推荐建议。通过了解用户过去的交互行为,系统能够更好地预测未来的需求,提高推荐的准确性和个性化水平。总体而言,特征交互在农产品推荐中的应用不仅是简单的特征组合,更是对众多农业特征进行深入挖掘和分析的手段。这种全面性的特征交互有助于系统更好地理解用户和农产品之间的关系,提高推荐系统在复杂农业环境下的适应性和性能。

CTR (click through rate) 预测在农产品电商推荐系统中至关重要,其任务是预测用户点击推荐物品的概率,然后根据预测的点击率做出许多推荐决策。传统的CTR推荐算法的核心是提取重要的低阶和高阶特征交互。然而,所有的这些模型中,特征交互要么简单枚举所有的特征,要么手动设计识别重要的特征。枚举所有的特征交互会带来大量的内存和计算成本,无用的交互可能会引入噪音并使训练复杂化。同时,手动识别重要的交互,人工成本很高,并且有可能遗漏一些违反直觉的交互。并且许多CTR预测工作以简单的方式计算特征交互,例如内积和哈达玛积,无法有效地对稀疏数据集中的特征交互进行建模。

本文提出了一种名为细粒度特征交互选择网络 FgFisNet (fine-grained feature interaction selection network) 的新模型,它是细粒度特征交互选择网络的缩写,用于细粒度特征交互和选择重要的特征交互。我们凭经验验证,模型训练预先识别有用的特征交互,模型可以专注学习重要的特征交互,而不必处理无用的特征交互。本文提出了一种特征交互选择算法来自动选择模型中重要的低阶和高阶特征交互。为每个特征

交互引入一组架构参数,通过梯度下降获得每个特征交互的相对重要性,然后重新训练模型,删除不重要的交互后,保留架构参数作为注意单元以帮助模型的学习。本文提出一种新的细粒度特征交互方法来计算与双线性函数的特征交互。组合内积和哈达玛内积,引入3种类型的双线性交互层来以细粒度的方式学习特征交互。在农产品电商真实数据集上进行广泛的实验, FgFisNet 在 CTR 方面改进了深度因子分解机 DeepFM (deep factorization machine) 模型约 20%, 验证了改进算法在农产品数据集上的有效性和可行性, 为农产品电商平台提供了新的思路和方法。

1 相关工作

1.1 农产品推荐技术研究现状

美国作为最早开展农产品电商推荐的国家,其涉及农产品的网站数量和比例在全球排名第1,远超其他国家,有26%的农产品通过互联网进行销售,欧洲是继美国之后农产品电商发展最快的地区, Farmigo 和 Amazon 作为欧洲和美国最具代表性的两个农产品电商企业, Farmigo 采用社区团购模式,根据社区订单利用当地或周边的农场进行生产配送,确保了农产品新鲜度,同时减少仓储和物流的开销。Amazon 率先在农产品电商平台中引入个性化推荐吸引了大量的忠实消费者,为公司带来了不菲的收益。

与国外相比,我国农产品电子商务起步较晚,但在“精准扶贫”和“互联网+农业”的推动下迅速崛起,传统的线下超市、电商企业和互联网投资者纷纷投入农产品电商领域的探索和投资。涌现出众多农产品平台,如美团买菜、京东生鲜、多多买菜等,这些平台降低交易成本,使农民能够获得更高利润,同时推动农村经济的发展和农业产业结构的优化。

农产品电商推荐主要是在传统推荐技术的基础上,结合农产品本身的特点来进行推荐技术的选择与优化。王明珠采用谱聚类技术解决了农产品电商推荐中协同过滤算法中导致的矩阵稀疏问题,通过对用户进行聚类构建用户-评分矩阵,成功降低了稀疏性,提高了推荐性能^[2]。Chen 等人提出基于 SVM 的农产品商品评论情感分析方法,利用情感评分代替好评率,综合考虑农产品评论数量、评论长度以及评论的情感因素,有效解决了高好评率排名中的问题,提供更可靠的排序结果^[3]。Xie 等人以知识图谱空间内方式建

模高阶连接性,通过递归传播邻居节点嵌入,以注意因子分解器作为邻居信息聚合,提取农产品潜在特征向量。通过多层感知器 (multilayer perceptron, MLP) 整合农产品和用户特征向量,并通过向量内积操作获得用户点击率预测^[4]。金沛璇构建面向农产品对接推荐场景下的农产品知识图谱,提出基于翻译转换和图卷积的知识图谱嵌入模型,学习农产品图中的实体嵌入表示,利用农产品实体和地点实体的相似度计算得出销售地推荐^[5]。

1.2 CTR 预测模型研究现状

CTR 预测模型广泛应用于个性化推荐,用于学习和预测用户的反馈,根据预测的用户对物品的评分进行排序,生成下一时刻的推荐列表。在本节中将简要回顾一下 CTR 预测的因子分解机模型和推荐系统中的自动机器学习 AutoML (automated machine learning) 模型。

因子分解机 FM (factorization machine) 与传统的简单线性模型不同的是,因子分解机考虑了特征间的交叉,对所有嵌套变量交互进行建模^[6]。场因子分解机 FFM (field-aware factorization machine) 在 FM 的基础上引入了“场 (Field)”的概念而形成的新模型,每个特征将被映射为多个隐向量,每个隐向量对应一个场,使得每两组特征交叉的隐向量都是独立的,可以取得更好的组合效果^[7]。梯度提升决策树 GBDT (gradient-boosting decision tree)+逻辑回归 LR (logistic regression) 将原始特征向量转换为离散型特征向量,利用 GBDT 进行“自动化”的特征组合,特征工程模型化,使模型具有了更高阶特征组合的能力^[8]。

随着深度学习的发展,一些模型使用多层感知器 MLP 来改进 FM。广义线性模型和神经网络模型 Wide&Deep 由单层 Wide 和多层 Deep 组合形成混合模型,单层的 Wide 部分善于处理大量稀疏的 ID 类特征,Deep 部分利用神经网络表达能力强的特征,进行深层的特征交叉,挖掘藏在数据背后的数据模式^[9]。深度交叉网络 Deep crossing 实现端到端的训练,没有任何人工特征工程的参与,并且相对于 FM、FFM 模型只具备二阶特征交叉的能力,Deep crossing 模型可以通过调整神经网络的深度进行特征之间的“深度交叉”^[10]。基于乘积的神经网络 PNN (product-based neural network) 设计了 Product 层先对低阶特征进行充分的交叉组合之后再送入深度神经网络 DNN (deep neural

network) 中做特征交叉组合, 加强了特征交叉能力^[11]. 基于因子分解机的神经网络模型 FNN (factorization machine supported neural network) 利用 FM 参数来初始化深度神经网络的嵌入向量 embedding 参数, 加快整个网络的收敛速度, 但模型主结构比较简单, 没有针对性的特征交叉层^[12].

深度交叉网络 DCN (deep & cross network) 增加特征之间的交互力度, 使用多层交叉层 (crossing layer) 对输入向量进行自动的特征交叉, 避免更多的人工特征组合^[13]. 深度因子分解机 DeepFM 针对 Wide&Deep 模型中 Wide 部分不具备自动的特征组合能力, 用 FM 代替 Wide 部分进行特征组合^[14]. 神经因子分解机 NFM (neural factorization machine) 在 embedding 层和多层神经网络之间加入特征交叉池化层, 相比于 Wide&Deep 加强了特征交叉^[15]. 注意力因子分解机 AFM (attention factorization machine) 在 NFM 基础上通过加入区分不同特征交叉的重要程度的注意力 attention 网络, 消除不同特征组合之间的重要性来优化 FM, 提高模型的性能^[16]. XDeepFM (extreme deep factorization machine) 使用压缩交互网络 CIN (compressed interaction network) 代替 DCN 里面的 cross network, 用向量级别 vector-wise 的特征交叉代替元素级别 bit-wise 的特征交叉, 使网络具有更有效的高阶显性交叉特征, 更高的泛化能力^[17]. 自动特征交叉 AutoCross (automatic feature crossing) 自动实现表数据特征组合, 利用集束搜索策略构建有效的组合特征, 同时包含尚未被现有工作覆盖的高阶特征组合. 提出多粒度离散化和连续小批量梯度下降, 提高了特征组合的效率和有效性^[18].

2 细粒度特征交互选择网络算法

在本节中, 我们描述用于农产品电商 CTR 预测的细粒度特征交互选择网络算法 FgFisNet. 这是一种在因子分解机模型中细粒度特征交互和选择重要的特征交互算法. 模型的架构如图 1 所示. 我们提出的模型由以下部分组成, 输入层、嵌入层、细粒度交互层、交互特征选择层、组合层、多个隐藏层和输出层. 输入层、嵌入层与 DeepFM 相同, 对输入特征和采用稀疏表示, 将原始特征输入嵌入到稠密向量中. 细粒度交互层对二阶特征交互进行建模. 选择层识别出细粒度交互层重要的特征交互. 随后, 这些交叉特征与特征嵌入

层通过组合层连接起来. 最后, 我们将组合层特征输入深度神经网络, 网络输出预测分数.

2.1 输入层和嵌入层

输入层和嵌入层广泛应用于基于深度学习的 CTR 模型, 例如 DeepFM 和 AFM. 输入层对原始输入特征采用稀疏表示, 原始特征包括农产品属性、地理位置、用户行为数据等. 原始输入特征大多是离散的并且包含多个字段, 通过 one-hot 编码将其转换为高维稀疏二值特征表示. 由于极端的稀疏性, 传统模型从数据中挖掘低阶特征组合的能力会受到限制, 以及深度神经网络特征空间巨大, 无法直接适用高维输入. 因此使用嵌入层将稀疏特征嵌入到低维、稠密实值向量中. Embedding 对每一个单独的离散特征进行转化, 将离散特征 c 的某一取值 c_i 经过映射转化为嵌入向量 e_i , 这个映射相当于一个单层线性神经网络. 嵌入层将各离散特征经过映射得到的嵌入向量组合在一起, 如式 (1) 所示:

$$E = [e_1, e_2, \dots, e_F] \quad (1)$$

其中, F 表示特征域的数量. e_i 表示第 i 个特征域的嵌入向量, $e_i \in R_k$, R_k 表示一个 k 维的实数向量空间, k 是嵌入层的维度.

2.2 细粒度特征交互和交互特征选择层

对于一个基于 CTR 预测的推荐场景, 最重要的是学习到用户点击行为背后隐含的特征组合. 由于农产品推荐任务的独特性, 我们需要深入分析包括季节性、地域性、用户兴趣和农产品属性等在内的农业特征. 对于季节性特征, 我们引入时间戳特征, 将时间划分为具体的月份或季度, 并通过细粒度特征交互层获取不同月份之间的特征关联, 以更精细地反映季节性需求; 同时, 引入地理位置信息, 如经纬度或地理编码, 首先通过细粒度特征交互层捕捉不同地域之间的相关性, 然后通过交互特征选择层选出与农产品推荐关联度高的地域性特征; 此外, 引入用户行为数据, 例如点击记录和购买历史, 通过细粒度特征交互层捕捉不同用户兴趣之间的关联, 以更全面地理解用户的个性化需求. 引入农产品属性信息, 考虑不同产品的生长周期、产地、品种等特性, 通过细粒度特征交互层, 例如引入“生长周期段”的特征和“产地为河南”的特征, 以更好地反映农产品属性对用户购买行为的影响.

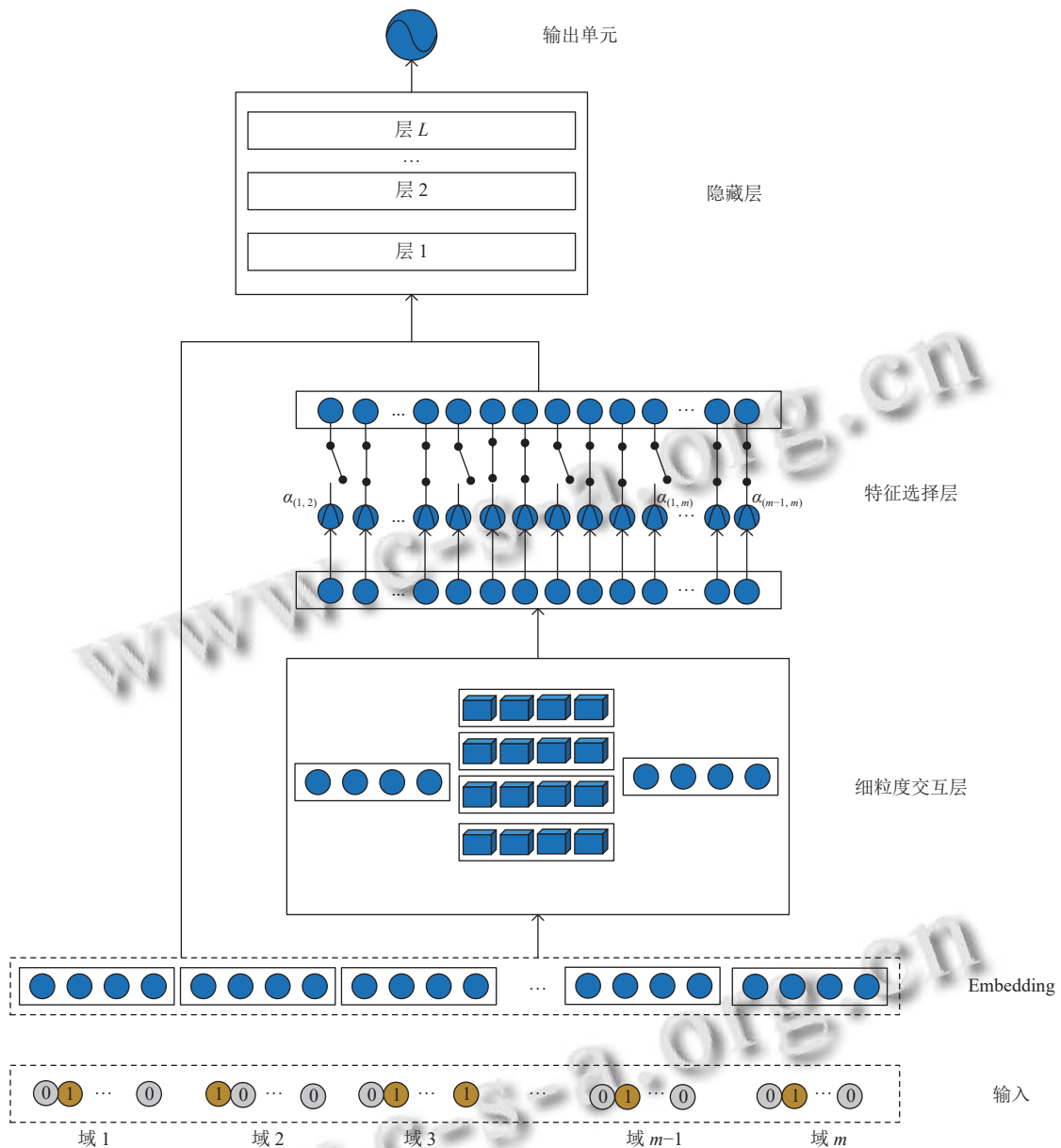


图 1 提出的 FgFisNet 模型架构

目前 CTR 模型离散型特征交互这一方面,交互层的特征交互要么是两两 embedding 内积,如 FM 和 FFM 等浅层模型.或者是两两 embedding 进行哈达玛积,比如, NFM 和 AFM 等深层模型.这两种交互方式过于简单,要让模型学习到更多的信息,因此我们提出组合内积和哈达玛积来更细粒度的特征交互.矩阵 w 和向量 v_i 之间使用内积,矩阵 w 和向量 v_j 之间使用哈达玛积.3 种特征交互方式如图 2 所示.

细粒度积共有 3 种类型的特征交互,分别是所有向量共用一个矩阵 w 、每一个域共用一个矩阵 w 和每

个交互共用一个矩阵 w .以第 i 个字段嵌入 v_i 和第 j 个字段嵌入 v_j 为例,特征交互的结果 $p_{i,j}$ 分别表示为:

$$p_{i,j} = v_i \cdot w \odot v_j \tag{2}$$

其中, $w \in R^{k \times k}$, $v_i, v_j \in R^k$, $1 \leq i \leq f$, $1 \leq j \leq f$,任一两域特征交叉 (v_i, v_j) 共用一个矩阵 w ,额外学习的参数量 $k \times k$.

$$p_{i,j} = v_i \cdot w_i \odot v_j \tag{3}$$

其中, $w_i \in R^{k \times k}$, $v_i, v_j \in R^k$, $1 \leq i \leq f$, $1 \leq j \leq f$,每一个域特征与其他域特征交叉时使用一个矩阵 w_i ,额外学习

的参数量 $f \times k \times k$.

$$p_{i,j} = v_i \cdot w_{ij} \odot v_j \quad (4)$$

其中, $w_{ij} \in R^{k \times k}$, $v_i, v_j \in R^k$, $1 \leq i \leq f, 1 \leq j \leq f$, 每两两域特征交叉时使用一个矩阵 w_{ij} , 额外学习的参数量 $f \times (f-1) \times k \times k$.

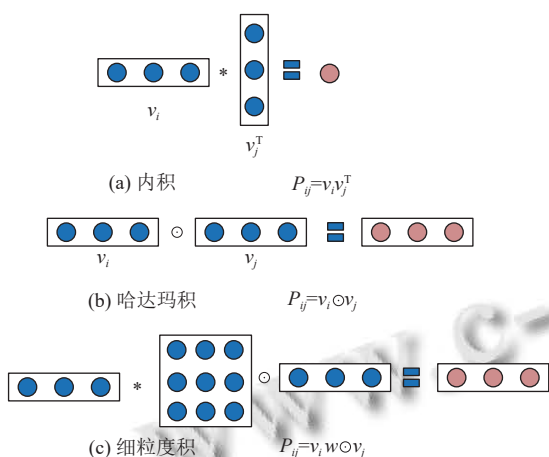


图2 计算特征交互的不同方法

随着推荐模型的不断变化, 特征交叉的方法也逐渐多样化. 从最开始的 LR 模型根据经验人工选择交叉特征, 到现在在神经网络自动计算所有特征对的交叉, 构造特征交叉往往需要先验知识和大量的实验测试, 同时引入很多没有意义的交叉特征, 可能形成噪声干扰模型. 因此我们设计门控机制, 在特征交叉层和 MLP 层之间为每个交叉特征分配一个全局的权重 α , 通过联合优化阶段和统一训练阶段, 选择保留非 0 权重 α 特征交互, 然后将权重 α 作为注意力单元进一步提高模型性能. 特征交互选择如图 3 所示.

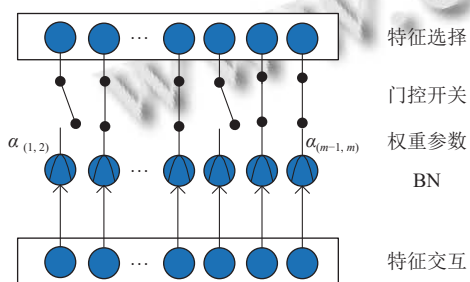


图3 特征交互选择

(1) 联合优化阶段. 通过梯度下降, 使用 GRDA^[19] 优化器训练出所有特征交叉项的权重 α , 同时使用 NAG 优化器训练模型参数. 两阶特征交叉 p 的表达式如式 (5) 所示:

$$p = \sum_{i=1}^m \sum_{j>i}^m \alpha_{(i,j)} p_{ij} \quad (5)$$

由于 $\alpha_{(i,j)}$ 与 p_{ij} 的值是共同学习, 它们尺度的耦合会导致 $\alpha_{(i,j)}$ 的估计不稳定, 使得 $\alpha_{(i,j)}$ 不再能够表示 p_{ij} 的相对重要性. 因此我们使用批量归一化 BN (batch normalization) 解决 $\alpha_{(i,j)}$ 与 p_{ij} 耦合问题, 同时实现快速收敛. 每组特征交互 p_{ij} 的 BN 操作计算表达式如式 (6) 所示:

$$BN(p_{ij}) = \frac{p_{ij} - \mu_B p_{ij}}{\sqrt{\sigma_B^2 p_{ij} + \epsilon}} \quad (6)$$

其中, μ_B 和 σ_B 是小批量数据中 p_{ij} 的平均值和标准差, ϵ 为数值稳定性常数, 同时将训练尺度和平移参数分别设置为 1 和 0, 使 $\alpha_{(i,j)}$ 得到稳定估计. 权重 α 与某个样本的特征交叉值无关, 仅与其对应的两个交叉特征域有关.

使用广义正则化对偶平均 GRDA (generalized regularized dual averaging) 优化器对权重 α 进行训练, 使其结果具有一定稀疏性, 权重 α 的具体更新方式如式 (7) 所示:

$$\alpha_{t+1} = \arg \min \left\{ \alpha^T \left(-\alpha_0 + \gamma \sum_0^t \nabla L(\alpha_t; Z_{t+1}) \right) + g(t, r) \|\alpha\|_1 + \frac{1}{2} \|\alpha\|_2^2 \right\} \quad (7)$$

其中, t 为 global step, Z 为输入数据, r 为学习率, $g(t, \gamma) = c\gamma^{\frac{1}{2}}(t\gamma)^u$, c 和 u 为超参数, α_0 为初始参数向量, γ 为学习率, $\sum_0^t \nabla L$ 为损失函数关于参数向量 α_t 的梯度在当前数据样本 Z_{t+1} 上的累加和, α_t 为第 t 个迭代步骤的参数向量, Z_{t+1} 为当前的数据样本. 式 (7) 第 1 项是期望更新后的权重和梯度下降法得到的结果不要相差太远, 后面两项是正则项.

使用 NAG (Nesterov accelerated gradient) 优化器对其他所有网络权重 v 进行优化, 权重 v 和 α 利用梯度下降在联合优化阶段一起更新. 使用动量的随机梯度下降算法, 让每一次的参数更新方向不仅取决于当前位置的梯度, 还受到上一次参数更新方向的影响, 利用上一次迭代的更新量, 来平滑这一次迭代的梯度. 而 NAG 优化器就是在使用动量的随机梯度下降基础上加入一个本次梯度相对于上次梯度的变化量, 本质上

是对目标函数的二阶导近似,具有更快的收敛速度,同时抑制梯度下降上下震荡的情况.NAG优化器的参数 θ 更新公式如式(8)和式(9)所示:

$$\hat{d}_i = \beta \hat{d}_{i-1} + g(\hat{\theta}_{i-1}) + \beta [g(\hat{\theta}_{i-1}) - g(\hat{\theta}_{i-2})] \quad (8)$$

$$\hat{\theta}_i = \hat{\theta}_{i-1} - \alpha \hat{d}_i \quad (9)$$

其中, d_i 和 d_{i-1} 分别是这一次和上一次的梯度更新方向, $g(\theta)$ 表示目标函数在 θ 处的梯度,超参数 β 是上一次更新方向的衰减权重, α 为学习率.

(2) 统一训练阶段.通过联合优化阶段,会得到一些权重为0的特征,我们直接丢弃这些特征,选择权重 α 大于0的特征交互.两阶特征交叉 p 的表达式变为:

$$p = \sum_{i=1}^m \sum_{j>i}^m \alpha_{(i,j)} G_{i,j} p_{i,j} \quad (10)$$

其中, G 为门控开关,取值为0或1.此时权重 α 不再是特征选择的参数,而是作为一种Attention机制存于模型中,跟随模型参数一起采用NAG优化器再次进行训练.

2.3 组合层

组合层由embedding特征的一阶特征和二阶交叉特征拼接而成.为了进一步提高性能,拼接后的向量输入到FgFisNet的下一层深度神经网络中.组合层 $h^{(0)}$ 表示形式如式(11)所示:

$$h^{(0)} = F_{\text{concat}}[E, P] = [c_1, c_2, \dots, c_k] \quad (11)$$

其中, E 为一阶特征项, P 为二阶特征交叉项, c_i 为组合后的嵌入向量, k 为合并后嵌入向量的数量.

2.4 深度神经网络

DNN在分类和回归任务中表现出了强大的能力,包括计算机视觉、语音识别和自然语言处理.DNN可以自动学习更具表现力的特征表示并提供更好的预测性能,DNN通过多层神经网络能够学习到农产品电商数据中更复杂的和更高阶的交叉特征,这对于具有复杂关系的特征组合非常有用,例如用户在某个时段、某个地理位置、某个商品类别下的交互.深度神经网络由多个全连接层组成,隐式捕获高阶特征交互.深度神经网络输入 $h^{(0)} = [c_1, c_2, \dots, c_k]$ 是组合层的输出, $h^{(0)}$ 被输入到深度神经网络,前馈过程表达式如式(12)所示:

$$h^{(l)} = \sigma(w^{(l)} h^{(l-1)} + b^{(l)}) \quad (12)$$

其中, l 是深度, σ 是激活函数, $w^{(l)}$ 是第 l 层的模型权

重, $b^{(l)}$ 是第 l 层的偏差, h_l 是第 l 层的输出.之后,将生成的密集实值特征向量输入到Sigmoid函数中进行农产品CTR预测,所述预测表达式如式(13)所示:

$$\hat{y} = \sigma(w^{|L|+1} h^{|L|} + b^{|L|+1}) \quad (13)$$

其中, $|L|$ 是DNN的深度.

2.5 输出层

我们提出的FgFisNet模型输出 $\hat{y} \in (0, 1)$,是农产品CTR预测值.我们模型的参数 $\theta = \{w^{(i)}\}_{i=1}^{|L|}, \{b^{(i)}\}_{i=1}^{|L|}$,FgFisNet模型通过训练,目标函数旨在最小化交叉熵损失,所述目标函数 $loss$ 表达式如式(14)所示:

$$loss = -\frac{1}{N} \sum_{i=1}^N (y_i \log(\hat{y}) + (1 - y_i) \times \log(1 - \hat{y})) \quad (14)$$

其中, y_i 是第 i 个实例的真实标签, \hat{y} 是点击率预测值, N 是样本的总长度.

3 实验分析

在本节中将详细介绍实验验证分析,首先介绍实验所需的数据集和实验设置,然后对FgFisNet模型进行实验验证和性能评估.在农产品电商真实的数据集上,FgFisNet模型在农产品CTR预估任务中优于主要的最先进模型.

3.1 数据集和实验设置

实验是在真实的农产品电商数据集上进行的,该数据集包含10万用户6个月内的行为记录,行为记录包含商品属性信息(商品类别、商品价格、是否打折或促销等)、用户属性信息(ID、年龄、性别等)、日期、地理位置信息以及用户操作类型(点击、收藏、加入购物车、购买).首先进行数据清洗,对存在缺失值的行为数据采用众数进行填充;对异常值使用中位数进行替换;对不一致数据根据相同特征在其他字段中记录进行数据调整,使其保持一致.对清洗后数据进行划分训练集和测试集,以2022年10月1日-2023年3月1日用户行为数据作为训练集,2023年3月1日-2023年4月1日用户行为数据作为测试集.

农产品电商数据集的小批量为1000.嵌入层的维度设置为15,学习率设置为0.001.所有深度神经网络模型的层数设置为3,所有激活函数均为ReLU,每层神经元数量为1000,dropout率设置为0.5.所有模型的最后一层都使用Sigmoid函数.

3.2 性能评估和实验结果

我们实验选择CTR预测常用的评估指标受试者工作特征曲线ROC (receiver operating characteristic curve) 曲线下与坐标轴围成的面积AUC (area under the curve) 和交叉熵 (log loss) 作为评估指标: AUC即ROC曲线下面积, 是评估分类问题中广泛使用的指标, AUC对分类阈值不敏感, 上限为1, 越大越好. log loss即对数损失, 是二元分类中广泛使用的指标, 测量两个分布之间的距离, log loss的下限为0, 表示两个分布完美匹配, 值越小表示性能越好.

表1显示在真实的农产品电商数据集上, FgFisNet中细粒度交互层使用不同类型特征交互性能. 细粒度交互层共有3种类型特征交互, 分别为所有向量共用一个矩阵 w (All)、每一个域共用一个矩阵 w (Each) 和每个交互共用一个矩阵 w (Interaction), 实验结果表明, 对于当前农产品推荐不同类型的特征交互方式, 每个交互共用一个矩阵 w 性能最好. 因此本实验FgFisNet采用每个交互共用一个矩阵 w .

表1 不同类型的特征交互性能

类型	log loss	AUC
All	0.3763	0.7858
Each	0.3763	0.7859
Interaction	0.3764	0.7862

我们在实验与5个模型进行比较, 这些模型在CTR预估和个性化推荐任务中有许多成功的应用, 都使用内积或哈达玛积进行特征交互, 但忽略了特征交互的重要性, 使用TensorFlow 1.5实现, 并使用NAG优化器进行训练, 参数设置与FgFisNet参数保持一致. FM捕捉特征交互, 限制在低阶无法处理复杂的特征交互. FNN使用FM的隐向量作为embedding初始化, 能够捕获多字段分类数据的高阶潜在模式. GBDT+LR能够捕获非线性和高阶特征交互, 适用于处理复杂模式, 但需要更多的特征工程, 计算成本高. AFM引入自注意力机制, 能够对特征交互进行权重分配, 提升了模型对特征之间关系的建模能力. DeepFM结合因子分解机和深度神经网络, 能够捕获低阶和高阶特征交互, 适用于复杂特征模式. 为了防止过拟合, 在训练GBDT+LR和FM时, 在损失函数中加入L2正则化. 同时采用dropout作为正则化方法, 以防止训练神经网络时过度拟合.

表2显示在真实的农产品电商数据集上, FgFisNet与这些主流的推荐算法整体性能. 我们发现FgFisNet模型始终优于其他模型. FgFisNet模型优于DeepFM

模型, 在AUC方面, FgFisNet相对于DeepFM的性能提升为0.39%, 相对于AFM性能提升为0.57%. 结果表明, 将细粒度特征交互和重要特征交互选择对农产品电商数据集进行预测是真实有效的, 并且优于其他没有进行更细粒度的特征交互和重要特征交互选择的模型的. 一方面原因, 引入细粒度特征交互, 有效地对系数数据集中的特征交互进行建模. 进行重要特征交互选择, 模型可以专注学习重要的特征交互, 减少降低了模型训练的复杂性. 另一方面原因, 将细粒度特征交互和重要交互特征和深度神经网络结合成一个统一的模型, 模型的性能获取进一步的提升, 隐式的高阶特征交互有助于模型获得更强的表达能力.

表2 不同方法的性能

方法	log loss	AUC
FM	0.3843	0.7788
FNN	0.3785	0.7809
GBDT+LR	0.3842	0.7728
AFM	0.3804	0.7817
DeepFM	0.3775	0.7831
FgFisNet	0.3764	0.7862

3.3 超参数研究

我们将在模型中进行一些超参数研究, 我们重点关注嵌入的维度和DNN的深度. 表3为不同嵌入维度在农产品电商数据集上的性能. 我们将嵌入维度大小从10更改为30, 并将实验结果总结在表3中. 我们发现随着维度从10扩展到30, 增大嵌入维度意味着增加嵌入层和DNN部分的参数数量, 我们的模型随着参数数量增加在农产品电商数据集上获得了实质性改进.

表3 不同嵌入维度的性能

嵌入维度	log loss	AUC
10	0.3793	0.7737
15	0.3787	0.7839
20	0.3786	0.7841
25	0.3767	0.7853
30	0.3764	0.7862

在深度神经网络层, 增加层数会增加模型的复杂性. 从表4可以看出, 增加层数一开始会提高模型性能, 然而, 如果层数不断增加, 性能会下降, 这是因为过于复杂的模型容易过拟合. 对于农产品电商数据集, 我们发现将隐藏层数设置为3是一个合适的选择.

4 结论

本文提出FgFisNet来进行细粒度特征交互选择,

旨在进行细粒度的特征交互和选择重要的特征交互。我们引入细粒度交互层和交互特征选择层,组合内积和哈达玛积,每个特征交互共用一个权重矩阵 w ,而不是简单地使用内积和外积。然后利用门控机制,在我们的模型细粒度交互层和深度神经网络层之间为每个交叉特征分配一个全局的权重 α ,通过在训练过程中自动识别重要的特征交互,并删除冗余的特征交互。然后将细粒度交互层和特征交互选择层和深度神经网络相结合,隐式的高阶特征交互有助于模型获得更强的表达能力,模型的性能获取进一步的提升。所提出的方法已经用到农产品电商推荐平台上,取得了显著的经济效益。

表4 不同DNN深度的性能

DNN深度	log loss	AUC
1	0.3773	0.7733
2	0.3767	0.78341
3	0.3764	0.7862
4	0.3762	0.7854
5	0.3768	0.7842
6	0.3763	0.7857

参考文献

- 农产品电商: 转型与创新成发展新热点. 农村新技术, 2022, (6): 49–50.
- 王明珠. 结合谱聚类和季节函数的农产品协同过滤推荐算法研究及应用 [硕士学位论文]. 南昌: 南昌大学, 2022. [doi: 10.27232/d.cnki.gnchu.2022.002004]
- Chen RL, Liu M. Agricultural commodity recommendation based on emotion analysis. Proceedings of the 2020 International Conference on Internet of Things and Intelligent Applications. Zhenjiang: IEEE, 2020. 1–4. [doi: 10.1109/ITIA50152.2020.9312324]
- 金沛璇. 基于网络表示学习的农产品供求对接推荐系统的设计与实现 [硕士学位论文]. 西安: 西安电子科技大学, 2022. [doi: 10.27389/d.cnki.gxadu.2022.001736]
- Xie HH, Yang J, Huang C, *et al.* Recommendation algorithm for agricultural products based on attention factor decomposer and knowledge graph. Proceedings of the 2022 Asia Conference on Algorithms, Computing and Machine Learning. Hangzhou: IEEE, 2022. 626–631. [doi: 10.1109/CACML55074.2022.00110]
- Rendle S. Factorization machines. Proceedings of the 2010 IEEE International Conference on Data Mining. Sydney: IEEE, 2010. 995–1000. [doi: 10.1109/ICDM.2010.127]
- 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.
- 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: ACM, 2014. 1–9.
- 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.
- Shan Y, Hoens TR, Jiao 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.
- Zhang WN, Du TM, Wang J. Deep learning over multi-field categorical data—A case study on user response prediction. Proceedings of the 38th European Conference on Information Retrieval. Padua: Springer, 2016. 45–57.
- Qu YR, Cai H, Ren K, *et al.* Product-based neural networks for user response prediction. Proceedings of the 2016 IEEE 16th International Conference on Data Mining. Barcelona: IEEE, 2016. 1149–1154. [doi: 10.1109/ICDM.2016.0151]
- Wang RX, Fu B, Fu G, *et al.* Deep & cross network for ad click predictions. Proceedings of the ADKDD 2017. Halifax: ACM, 2017. 12.
- 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: IJCAI.org, 2017. 1725–1731.
- He XN, 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.
- 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.org, 2017. 3119–3125.
- 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.
- Luo YF, Wang MS, Zhou H, *et al.* AutoCross: Automatic feature crossing for tabular data in real-world applications. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage: ACM, 2019. 1936–1945.
- Chao SK, Cheng G. A generalization of regularized dual averaging and its dynamics. arXiv:1909.10072, 2019.

(校对责编: 张重毅)