



# 智能系统学报

CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

## 基于多路交叉的用户金融行为预测

程鹏超, 杜军平, 薛哲

引用本文:

程鹏超, 杜军平, 薛哲. 基于多路交叉的用户金融行为预测[J]. 智能系统学报, 2021, 16(2): 378–384.

CHENG Pengchao, DU Junping, XUE Zhe. Prediction of user financial behavior based on multi-way crossing[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(2): 378–384.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202006054>

## 您可能感兴趣的其他文章

### 深度学习的双人交互行为识别与预测算法研究

Human interaction recognition and prediction algorithm based on deep learning

智能系统学报. 2020, 15(3): 484–490 <https://dx.doi.org/10.11992/tis.201812029>

### 基于生成式对抗网络的道路交通模糊图像增强

Enhancement of blurred road-traffic images based on generative adversarial network

智能系统学报. 2020, 15(3): 491–498 <https://dx.doi.org/10.11992/tis.201903041>

### 多特征融合的兴趣点推荐算法

A point of interest recommendation algorithm based on multi-feature fusion

智能系统学报. 2019, 14(4): 779–786 <https://dx.doi.org/10.11992/tis.201801048>

### 基于改进卷积神经网络的多标记分类算法

A multi-label classification algorithm based on an improved convolutional neural network

智能系统学报. 2019, 14(3): 566–574 <https://dx.doi.org/10.11992/tis.201804056>

### 多特征融合的lncRNA识别与其功能预测

lncRNA recognition by fusing multiple features and its function prediction

智能系统学报. 2018, 13(6): 928–934 <https://dx.doi.org/10.11992/tis.201806008>

### PG-RNN:一种基于递归神经网络的密码猜测模型

PG-RNN: a password-guessing model based on recurrent neural networks

智能系统学报. 2018, 13(6): 889–896 <https://dx.doi.org/10.11992/tis.201712006>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.202006054

## 基于多路交叉的用户金融行为预测

程鹏超, 杜军平, 薛哲

(北京邮电大学 智能通信软件与多媒体北京市重点实验室, 北京 100876)

**摘 要:** 针对通过挖掘用户的金融行为来改善金融领域的服务模式和服务质量的问题, 本文提出了一种基于多路交叉特征的用户金融行为预测算法。根据数据包含的属性构建训练的特征, 基于因子分解机模型 (FM) 利用下游行为预测任务对金融数据的特征进行预训练, 获取数据特征的隐含向量。引入特征交叉层对金融数据的高阶特征进行提取, 解决 FM 线性模型只能提取低阶特征的缺点。利用残差网络对金融数据的高阶特征进行提取, 解决深度神经网络在提取金融数据高阶特征时由于网络层数过深而导致的梯度消失的问题。最后, 将 FM、特征交叉网络和残差网络整合为统一的多塔模型进行用户金融行为预测, 并融合低阶特征与高阶特征进行用户金融行为预测。在多个数据集上对算法的有效性进行了实验验证, 实验结果表明, 所提出的算法能够取得较好的用户金融行为预测的准确率。

**关键词:** 行为预测; 金融; 多路交叉; 残差; 多塔模型; 预训练; 挖掘; 联合训练

**中图分类号:** TP391    **文献标志码:** A    **文章编号:** 1673-4785(2021)02-0378-07

中文引用格式: 程鹏超, 杜军平, 薛哲. 基于多路交叉的用户金融行为预测 [J]. 智能系统学报, 2021, 16(2): 378-384.

英文引用格式: CHENG Pengchao, DU Junping, XUE Zhe. Prediction of user financial behavior based on multi-way crossing[J]. CAAI transactions on intelligent systems, 2021, 16(2): 378-384.

## Prediction of user financial behavior based on multi-way crossing

CHENG Pengchao, DU Junping, XUE Zhe

(Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China)

**Abstract:** To improve the service mode and service quality in the financial field by mining the financial behaviors of users, a user financial behavior prediction algorithm based on multi-way crossing (MCUP) is proposed in this paper. First, the training features are constructed based on the attributes contained in the data. Based on the FM model, the downstream behavior prediction tasks are used to pre-train the features of the financial data, and the hidden vectors of the features are obtained. Second, the feature cross-layer is introduced to extract high-order features of financial data, overcoming the disadvantage that the FM linear model can only extract low-order features. Then, the residual network structure is used to extract high-order features of financial data, solving the gradient disappearance problem caused by the too deep network layer. Finally, a unified multi-tower model integrated by the FM, feature cross network, and residual network is used to predict user financial behavior, blending low-order and high-order features. Experimental results show that the proposed algorithm can achieve a better accuracy rate in predicting user financial behavior.

**Keywords:** behavior prediction; financial; multi-way crossing; residual; multi-tower model; pre-training; mining; joint training

随着人们生活水平的提高, 人们更加关注对

收稿日期: 2020-06-30.

基金项目: 国家重点研发计划项目 (2018YFB1402600); 国家自然科学基金项目 (61772083, 61802028); 广西科技重大专项 (桂科 AA18118054).

通信作者: 杜军平. E-mail: junpingdu@126.com.

生活质量的追求, 在闲余的时间买卖基金、股票等以获取更多地利润。随着人们金融行为和意识的提升, 金融领域的用户行为挖掘可以探究用户的行为规律, 通过挖掘用户金融行为的规律与变化, 为用户推荐或提供更好的服务。金融大数据

中包括结构化文本数据、非结构化文本数据、数值数据等。挖掘金融大数据中用户的金融行为来改进金融领域服务模式,提高金融领域服务质量,已经成为了一个重要的研究方向<sup>[1]</sup>。

为了有效挖掘金融大数据中用户潜在的金融行为特征,提高金融领域的服务模式和服务质量,本文提出了一种基于多路交叉特征的用户金融行为预测算法(user financial behavior prediction algorithm based on multi-way crossing, MCUP)。首先,根据数据包含的属性构建训练的特征,利用因子分解机模型<sup>[2]</sup>(factorization machines, FM)模型借助下游行为预测任务对金融数据的特征进行预训练,可以获取到数据中特征的隐含向量。对于文本特征而言,利用预训练模型对文本特征进行表示以获得语义丰富的特征向量。然后,引入特征交叉层来对金融数据的较高阶特征进行提取,解决FM线性模型只能提取低阶特征的缺点。同时,利用残差网络结构对金融数据的高阶特征进行提取,解决深度神经网络在提取金融数据高阶特征时会因为网络层数过深而导致梯度消失的问题。最后,利用将FM、特征交叉网络和残差网络结合在一起的多塔模型进行用户金融行为的预测,融合低阶特征与高阶特征进行实验。

本文的主要贡献如下:

1) 提出了一种基于多路交叉的用户金融行为预测方法,结合数据低阶特征和高阶交叉特征对金融大数据用户行为进行预测;

2) 使用预训练FM模型对训练特征隐向量表示进行初始化,加速训练和模型收敛,同时FM可以提取数据特征的一阶和二阶交叉特征;

3) 引入交叉特征层和残差网络提取数据高阶特征,将稀疏的身份(identity, ID)特征表示成稠密向量作为输入,同时考虑到高阶特征对用户行为的预测效果影响。

## 1 用户行为预测的研究现状

目前现有的用户行为的预测的方法一般分为2种类型,传统的机器学习方法和深度学习方法。对于传统的机器学习方法而言,FM在理论上可以进行任意阶特征的交叉,由于计算复杂度的限制,目前只是用来进行二阶特征交叉。FM核心在于提出了特征隐向量的概念,将原始特征的出现都作为一个特征,每个特征利用稠密的向量来进行表示,解决了多项式特征交叉时交叉系数更新慢的缺点。FM的不足之处在于所有特征对于行为预测的结果影响系数是相同的,导致数

据各个特征之间没有区分性。为了解决FM无法区分特征对行为预测性能的影响,文献[3]提出了场敏感因子分解模型(field-aware factorization machines, FFM),FFM是在FM的基础上引入了域概念的行为预测模型。FFM将具有相同性质的特征归结于同一个域,每一维特征针对每一个域都会学习到一个隐向量,所以隐向量不仅仅与单个特征相关而且和域相关。FM和FFM仅仅考虑到了数据的低阶特征,而没有考虑数据高阶交叉特征,类似用户ID和物品ID特征需要深度神经网络才能提取高阶交叉特征<sup>[4]</sup>。随着深度学习技术的发展,深度学习在推荐、广告、计算机视觉、自然语言处理、语音识别等领域取得了重要进展,深度学习几乎无限的表达能力被广泛地研究<sup>[5]</sup>。文献[6]利用结构化和非结构化大数据对商品购买行为进行了预测。文献[7]基于用户行为序列数据和选择模型的方法,对用户金融行为选择出最优的因子模型并预测用户的行为。文献[8]利用支持向量机模型对用户的消费情况进行预测并获得了较好的预测结果。文献[9]结合支持向量机和逻辑回归模型进行了购买行为的预测,进一步提升了金融行为预测的准确性。文献[10]采用改进的决策树构建用户金融行为预测模型,实现了大数据环境下的模型构建和训练。文献[11]构建了Pareto/NBD模型,结合协变量进一步提升了用户购买行为预测的准确率。文献[12]通过改进了传统的预测模型,在产品购买行为数据集上验证了方法的有效性。文献[13]挖掘了用户购买行为规律,分析出了符合用户购买意愿的商品序列。文献[14]构建了线上消费购买率预测模型,预测了未来用户购买的行为规律。文献[15]结合遗传算法与传统算法进行最优模型组合,并通过实验验证了该方法的有效性。

深度学习技术同样也应用到了用户行为预测方面。文献[16]采用CNN-LSTM(convolutional neural network long short term memory)模型预测用户购买行为,利用卷积神经网络(convolutional neural network, CNN)进行特征抽取,通过长短期记忆网络(long short term memory, LSTM)<sup>[17]</sup>建立时间序列,实现对特征的自动抽取和行为预测。支持因子分解机的神经网络模型(factorization-machine supported neural networks, FNN)<sup>[18]</sup>就是来解决FM和FFM仅考虑一阶和二阶特征的问题,FNN利用FM作为预训练模型,将训练好的特征的隐向量作为后续深度神经网络的输入,来得到特征的高阶组合。它可以解决类似用户



ID 没有出现过的泛化问题,但是忽略了低阶特征的重要性,即在数据中频繁出现的低阶特征组合也能显示出用户行为。因此,低阶特征和高阶组合特征对用户行为的预测都很重要。深度因子分解模型 (deep factorization machines, DeepFM)<sup>[19]</sup> 提出了双塔模型分别对数据低阶特征和高阶交叉特征进行提取。利用 FM 模型自动地提取数据一阶与二阶特征,深度神经网络提取数据的高阶特征,通过融合数据低阶特征和高阶特征来进行用户行为的预测,利用 FM 天然可以自动组合二阶特征和一阶特征的优点来避免人工干预,并且在模型横向和纵向来共享 FM 的特征隐向量,真正实现了端到端的用户行为预测。DeepFM 模型的输入对于连续特征是不友好的,需要把文本特征作为用户行为预测的特征之一,文本特征大多数情况会用一个低维稠密向量来表示。深度交叉网络 (Deep&Cross network)<sup>[20]</sup> 共享输入,即 deep 部分和 cross 部分输入是相同的,在其输入的时候对于类似文本特征而言会将其与其他特征进行连接作为输入。cross 部分在理论上可以做到对高阶特征的组合,而且其参数随输入维度是线性增加的,FM 要想对高阶特征进行组合其复杂度是呈指数增长的。Deep&cross 虽然既有高阶特征又有低阶特征,但是对于输入的离散特征进行 one-hot 处理后,特征之间是同等重要的,没有 field 的概念。

## 2 基于多路交叉特征的用户行为预测的算法 (MCUP)

基于多路交叉特征的用户行为预测的方法的框架如图 1 所示。

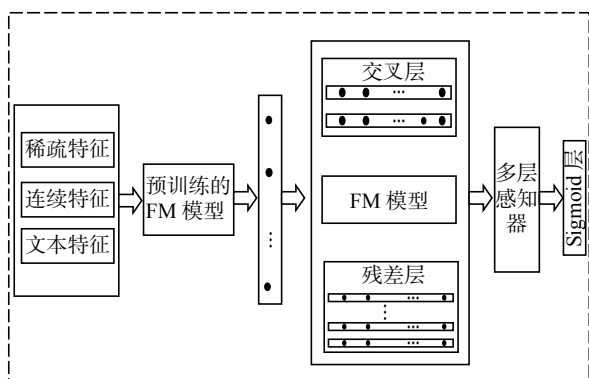


图 1 基于多路交叉特征的用户行为预测框架

Fig. 1 The framework of user behavior prediction based on multi-way

模型由 3 部分构成,如图 1 所示,左侧模块表示利用预训练 FM 模型获取离散特征和连续特征的稠密向量表示;中间模块利用交叉层对输入特征进行特征交叉获得较高阶特征向量表示;右侧

输入特征经过残差网络模块来提取高阶特征,得到特征的高阶组合。将 3 个模块的输出特征向量进行连接作为浅层全连接网络的输入,最后得到用户行为的预测结果。

### 2.1 基于多项式的交叉特征提取

FM 模型在理论上可以拟合任意高阶特征,由于 FM 模型在拟合高阶特征的同时,模型复杂度将会成指数倍增加,所以一般只使用 FM 到二阶特征交叉。为了在保证模型复杂度低的情况下,对高阶特征进行交叉组合提取,引入基于多项式的交叉特征提取方法。

基于多项式的交叉特征提取方法核心思想是利用乘法计算来进行特征交叉。同样,在理论上基于多项式的交叉特征层可以拟合任意高阶组合特征。主要利用  $\mathbf{x}_0 \mathbf{x}_l^T$  这种表达方式进行特征交叉,使用原始特征与基于多项式的交叉特征层进行乘法运算,使得特征交叉通过一个 rank-one 的数字来表示,大大减少了模型的参数量。具体为

$$\mathbf{e}_{l+1} = \mathbf{e}_0 \mathbf{e}_l^T \mathbf{w}_l + \mathbf{b}_l + \mathbf{e}_l \quad (1)$$

式中:  $\mathbf{e}_{l+1}$  表示  $l+1$  层特征向量;  $\mathbf{e}_0$  表示模型原始输入特征向量;  $\mathbf{w}_l$  表示第  $l$  层模型参数;  $\mathbf{b}_l$  表示第  $l$  层偏移量;  $\mathbf{e}_l$  表示第  $l$  层特征向量。

每一层的特征都由其上一层的特征进行交叉组合,并把上一层的原始特征重新加回来。这样既能特征组合,自动生成交叉组合特征,又能保留低阶原始特征,随着 cross 层的增加,可以生成任意高阶的交叉组合特征,且在此过程中没有引入更多的参数,有效控制了模型复杂度。

### 2.2 基于预训练的分解因子机

FM 的核心在于提出了特征隐向量的概念,将原始特征的出现都作为一个特征,每个特征利用稠密的向量来进行表示,解决了由于数据稀疏导致多项式特征交叉时交叉系数更新慢的缺点。FM 利用交叉的特征隐向量之间计算得分来代替多项式计算时的特征交叉系数,根据随机梯度下降的计算公式可以计算出参数的更新公式,由参数更新公式可以得出参数的更新只要交叉特征中的一个值不为 0 即可,从而解决了百万级的类别特征经过 one-hot 编码后数据特征稀疏的问题。本文提出利用下游任务去预训练 FM 模型,使用预训练 FM 模型对 MCUP 算法中 FM 模块进行初始化,以达到加速模型训练和收敛。

FM 一般只用到二阶特征组合,为

$$y = \theta_0 + \sum_{i=1}^n \theta_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \theta_{ij} x_i x_j$$

式中:  $\theta_0$  表示偏置参数;  $\theta_i$  表示一阶特征的参数;

$\theta_{ij}$  表示二阶组合特征参数;  $x_i$  表示第  $i$  个特征;  $x_j$  表示第  $j$  个特征。

$$y = \theta_0 + \sum_{i=1}^n \theta_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \langle v_i, v_j \rangle x_i x_j$$

式中:  $v_i$ 、 $v_j$  分别表示第  $i$ 、 $j$  个特征对应的隐向量;  $\langle \cdot, \cdot \rangle$  表示含义同式 (1)。

在 MCUP 算法中利用预训练 FM 模型对数据中的特征进行预训练, 类似自然语言处理中词向量的操作方式, 可以获得数据特征的隐向量表示。利用预训练 FM 模型对 MCUP 算法进行初始化, 可以通过数据集对特征隐向量进行微调获得更好的用户行为预测的效果。

### 2.3 基于残差网络的高阶特征提取

线性回归 (linear regression, LR) 算法、FM 算法等仅仅考虑到了低阶特征的组合, 在下游任务中低阶特征可以记住用户历史行为, 从而对下游任务的目标起到正向作用。但是在收集到的数据中存在大量的 ID 类特征, 即 One-hot 操作后会得到离散稀疏的向量表示, 通常 ID 类特征在特征组合中起到非常重要的作用。高维离散特征通用的处理方式是使用 One-hot 进行表示后, 利用 LR 算法进行用户行为预测等下游任务。随着深度神经网络和嵌入方法的发展更新, 高维离散特征不仅使用 One-hot 进行表示, 后续还使用词嵌入的方法对 ID 类特征进行嵌入学习, 使用低维稠密特征对 ID 类特征进行表示。

基于残差网络的高阶特征提取方法将 ID 类特征和连续型特征放在一起作为模型输入, 利用深度神经网络中的残差概念来将网络做深。随着网络深度的增加, 特征不断进行交叉组合, 最后得到具有泛化能力的高阶特征。残差网络的计算为

$$e_l = e_{l-1} + f(e_{l-1}, w_l)$$

式中:  $e_l$  表示第  $l$  层的输出特征;  $e_{l-1}$  表示第  $l$  层的特征输入;  $w_l$  表示第  $l$  层的模型参数。

## 3 实验与结果

### 3.1 度量标准

评价用户行为预测结果好坏的指标一般可以使用精确率、召回率、 $F_1$ 。在实际场景中, 也用 AUC 和 MAP 来表示模型效果的好坏。

$$P = \frac{TP}{TP+FP}, R = \frac{TP}{TP+FN}$$

$$F_1 = \frac{2PR}{P+R}$$

式中: TP 表示标记为真正例; FP 表示标记为假正例; FN 表示标记为假反例; TN 表示标记为真

假例。

$$AP_i = \frac{1}{R} \sum_{k=1}^m \frac{R_k}{k} \text{isrel}_k$$

$$\text{MAP} = \frac{1}{q} \sum_{i=1}^q AP_i$$

式中:  $m$  表示与第  $i$  个查询相关的数据数量;  $R_k$  表示相关性排序前  $k$  个数据中与查询数据相关的数据数量;  $\text{isrel}_k$  表示第  $k$  个数据是否与查询相关, 如果相关值为 1, 反之为 0;  $q$  表示查询集合中数据数量;  $AP_i$  表示第  $i$  个查询的平均精确率。

### 3.2 数据集

实验数据为 2019 年天弘基金用户点击行为、2019 年长信基金用户点击行为和鹏华基金用户点击行为。点击基金的用户行为表示用户在界面点击基金对基金基本信息进行查看。基金数据集详细描述见表 1。

表 1 基金数据集  
Table 1 Fund dataset

数据集	训练集	验证集	测试集
天弘基金	13 872	3 000	1 000
长信基金	15 624	4 000	1 500
鹏华基金	14 476	3 500	1 000

### 3.3 实验和结果

为了验证基于多路交叉特征的用户行为预测 (MCUP) 算法的有效性, 在 3 个不同的基金数据集上对用户点击行为进行预测。对于每个基金数据集设置 3 组对比实验, MCUP 算法分别与 FM 算法、FNN 算法和 Deep&Cross 算法。利用精确率、召回率、 $F_1$  和 MAP 对 MCUP 算法的性能进行评测, 特别地在计算 MAP 指标的时候选取前 2、4、6、8 和 10 结果来进行统计, 从而验证 MCUP 算法在用户行为预测任务的有效性。

根据表 2 可以看出, MCUP 算法在 MAP@2、MAP@4、MAP@6、MAP@8 和 MAP@10 指标均优于传统的 FM 与基于深度学习的 FNN 算法和 Deep&Cross 算法。在天弘基金数据集上, 用户点击基金信息的行为预测任务中 MCUP 算法相较于 Deep&Cross 算法而言, 在 MAP@2、MAP@4、MAP@6、MAP@8 和 MAP@10 上分别高 0.7%、0.5%、0.8%、0.7%、1%。MCUP 算法使用预训练 FM 模型对特征隐向量进行初始化, 相对于 Deep&Cross 算法具有更好的收敛速度和预测效果。MCUP 算法比起传统的 FM 算法在 MAP@10 提升最大, FM 算法仅仅考虑到一阶和二阶特征对

用户点击行为预测结果的影响, MCUP 算法不仅考虑了低阶特征, 而且使用残差网络来提取高阶交叉特征。

表 2 MCUP 与对比算法在天弘基金数据集上点击 MAP  
Table 2 MAP of MCUP and comparison algorithm click behavior on Tianhong fund dataset

算法	MAP@2	MAP@4	MAP@6	MAP@8	MAP@10
FM	0.736	0.744	0.751	0.756	0.760
FNN	0.749	0.757	0.761	0.765	0.769
Deep&Cross	0.759	0.766	0.768	0.772	0.774
MCUP	0.766	0.771	0.776	0.779	0.784

根据图 2 可以看出, MCUP 算法在前 2、4、6、8 和 10 的数据中  $F_1$  指标均优于对比算法。在天弘基金数据集上, 用户点击基金信息的行为预测任务中, MCUP 算法相对于 FM 算法在 top@2 上提升最大, MCUP 算法同时考虑了低阶特征和高阶交叉特征对用户点击行为预测的影响, 利用残差网络隐式地提取数据高阶交叉特征和 FM 算法获取数据一阶和二阶特征。

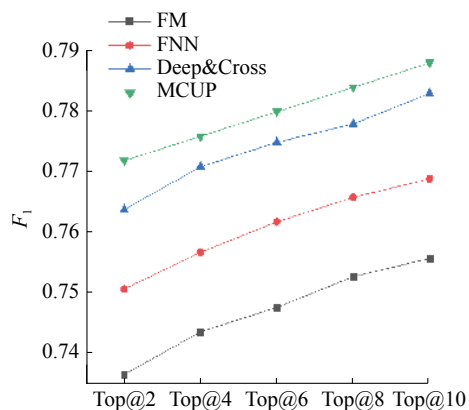


图 2 MCUP 与对比算法在天弘基金数据集上点击  $F_1$  曲线

Fig. 2  $F_1$  curves of MCUP and comparison algorithm click behavior on the Tianhong Fund dataset

根据表 3 可以看出, MCUP 算法在 MAP@2、MAP@4、MAP@6、MAP@8 和 MAP@10 指标均优于传统的 FM 与基于深度学习的 FNN 算法和 Deep&Cross 算法。在长信基金数据集上, 用户点击基金信息的行为预测任务中 MCUP 算法在 MAP@2、MAP@4、MAP@6、MAP@8 和 MAP@10 上比 FNN 分别高 1.8%、1.6%、1.9%、1.6%、1.7%。MCUP 算法不仅使用预训练 FM 算法对特征隐向量进行初始化, 利用 FM 提取低阶特征并且使用残差网络提取高阶交叉特征, 而 FNN 算法只利用预训练的 FM 算法进行高阶交叉特征的提取, 没

有考虑低阶特征对于用户点击行为预测的影响。MCUP 算法比具有相似结构的 Deep&Cross 算法在 MAP@10 提升最大, Deep&Cross 算法没有考虑到一阶和二阶特征对用户点击行为预测结果的影响, MCUP 算法利用 FM 来进行低阶特征的提取。

表 3 MCUP 与对比算法在长信基金数据集上点击 MAP  
Table 3 MAP of MCUP and comparison algorithm click behavior on Changxin fund dataset

算法	MAP@2	MAP@4	MAP@6	MAP@8	MAP@10
FM	0.743	0.747	0.752	0.755	0.762
FNN	0.757	0.763	0.766	0.772	0.776
Deep&Cross	0.769	0.773	0.777	0.781	0.784
MCUP	0.775	0.779	0.785	0.788	0.793

根据图 3 可以看出, MCUP 算法在前 2、4、6、8 和 10 的数据中  $F_1$  指标均优于对比算法。在长信基金数据集上, 用户点击基金信息的行为预测任务中, MCUP 算法相对于 Deep&Cross 算法在 top@6 以后提升的幅度逐渐增大, 相较于 Deep&Cross 算法的 Deep 神经网络, MCUP 算法使用残差网络隐式地提取数据高阶交叉特征, 具有更强的表达能力。

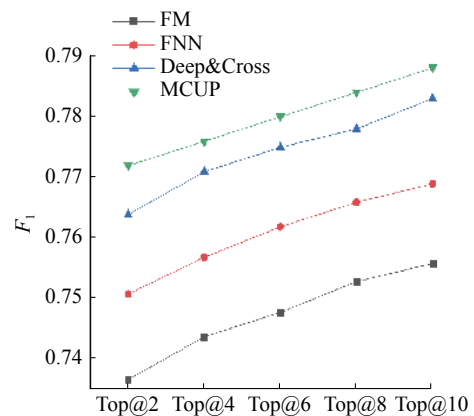


图 3 MCUP 与对比算法在长信基金数据集上点击  $F_1$  曲线  
Fig. 3  $F_1$  curves of MCUP and comparison algorithm click behavior on the Changxin fund dataset

根据表 4 可以看出, MCUP 算法在 MAP@2、MAP@4、MAP@6、MAP@8 和 MAP@10 指标均优于传统的 FM 算法、FNN 算法和 Deep&Cross 算法。在鹏华基金数据集上, 用户点击基金信息的行为预测任务中 MCUP 算法相较于 Deep&Cross 算法而言, 在 MAP@2、MAP@4、MAP@6、MAP@8 和 MAP@10 上分别高 0.3%、0.2%、0.3%、0.2%、0.4%。MCUP 在 MAP@10 上的性能最好, 说明 MCUP 算法能够尽可能地将用户点击过的基金数



据排到前面。MCUP算法融合低阶特征和高阶交叉特征对用户点击行为进行预测,极大地挖掘了用户潜在的兴趣意图。

表4 MCUP与对比算法在鹏华基金数据集上点击MAP  
Table 4 MAP of MCUP and comparison algorithm click behavior on Penghua fund dataset

算法	MAP@2	MAP@4	MAP@6	MAP@8	MAP@10
FM	0.743	0.747	0.752	0.755	0.762
FNN	0.757	0.763	0.766	0.772	0.776
Deep&Cross	0.769	0.773	0.777	0.781	0.784
MCUP	0.775	0.779	0.785	0.788	0.793

根据图4可以看出,MCUP算法在前2、4、6、8和10的数据中 $F_1$ 指标均优于对比算法。在鹏华基金数据集上,用户点击基金信息的行为预测任务中,MCUP算法相对于FNN算法提升更加平稳且 $F_1$ 性能更优,MCUP算法利用特征交叉层显示进行数据特征的有限阶交叉,而且使用残差网络隐式地进行数据高阶特征的提取,同时考虑了低阶特征和高阶交叉特征对用户点击行为预测的影响。

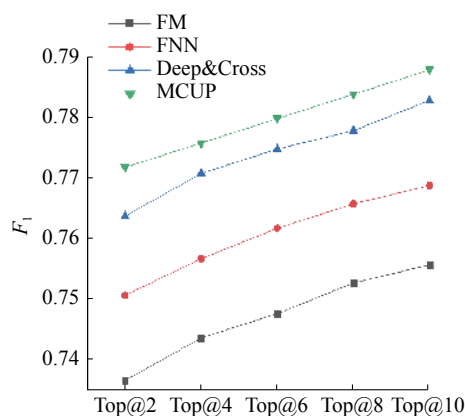


图4 MCUP与对比算法在鹏华基金数据集上点击 $F_1$ 曲线

Fig. 4  $F_1$  curves of MCUP and comparison algorithm click behavior on the Penghua Fund dataset

## 4 结束语

为了有效挖掘金融大数据中用户潜在的金融行为特征,本文提出了一种基于多路交叉特征的用户金融行为预测算法。首先,利用FM模型对金融数据的特征进行预训练,获取到数据中特征的隐含向量。然后,引入特征交叉层和残差网络结构来对金融数据的较高阶特征进行提取,解决了FM线性模型只能提取低阶特征、模型无法有效训练等问题。最后,将FM、特征交叉网络和残

差网络结合为统一的多塔模型进行用户金融行为的预测。在多个金融行为预测数据集上的实验结果表明,本方法能够有效融合金融大数据的低阶特征与高阶特征,并准确地预测了用户的金融行为。

## 参考文献:

- [1] MAIA M, HANDSCHUH S, FREITAS A, et al. WWW'18 open challenge: financial opinion mining and question answering[C]// Companion Proceedings of the Web Conference 2018. Lyon, France: International World Wide Web Conferences Steering Committee, 2018: 1941–1942.
- [2] SHI Hongyu, CHEN Ling, XU Zhenxing, et al. Personalized location recommendation using mobile phone usage information[J]. *Applied intelligence*, 2019, 49(10): 3694–3707.
- [3] CHEN Jiehao, LI Xueyi, ZHAO Ziqian, et al. A CTR prediction method based on feature engineering and online learning[C]// Proceedings of the 17th International Symposium on Communications and Information Technologies. Cairns, Australia, 2017: 1–6.
- [4] ZHAO Xibin, WANG Nan, ZHANG Yubo, et al. Beyond pairwise matching: person reidentification via high-order relevance learning[J]. *IEEE transactions on neural networks and learning systems*, 2018, 29(8): 3701–3714.
- [5] HAO Tong, WANG Qian, WU Dan, et al. Adaptive recommendation for photo pose via deep learning[J]. *Multimedia tools and applications*, 2018, 77(17): 22173–22184.
- [6] RYU G A, NASRIDINOV A, RAH H, et al. Forecasts of the amount purchase pork meat by using structured and unstructured big data[J]. *Agriculture*, 2020, 10(1): 1–14.
- [7] 曾宪宇, 刘淇, 赵洪科, 等. 用户在线购买预测: 一种基于用户操作序列和选择模型的方法[J]. *计算机研究与发展*, 2016, 53(8): 1673–1683.
- [8] ZENG Xianyu, LIU Qi, ZHAO Hongke, et al. Online consumptions prediction via modeling user behaviors and choices[J]. *Journal of computer research and development*, 2016, 53(8): 1673–1683.
- [9] LIU Xiaoman, LI Jing. Using support vector machine for online purchase predication[C]//Proceedings of 2016 International Conference on Logistics, Informatics and Service Sciences (LISS). Sydney, Australia, 2016: 1–6.
- [10] 祝歆, 刘潇蔓, 陈树广, 等. 基于机器学习融合算法的网络购买行为预测研究[J]. *统计与信息论坛*, 2017, 32(12): 94–100.
- [11] ZHU Xin, LIU Xiaoman, CHEN Shuguang, et al. Research on network purchase behavior prediction based on machine learning fusion algorithm[J]. *Statistics and information forum*, 2017, 32(12): 94–100.
- [12] 杜刚, 黄震宇. 大数据环境下客户购买行为预测[J]. *管*

- 理现代化, 2015, 35(1): 40–42.
- DU Gang, HUANG Zhenyu. Customer purchasing behavior prediction under the environment of big data[J]. *Modernization of management*, 2015, 35(1): 40–42.
- [11] 李美其, 齐佳音. 基于购买行为及评论行为的用户购买预测研究 [J]. 北京邮电大学学报(社会科学版), 2016, 18(4): 18–25.
- LI Meiqi, QI Jiayin. Customer purchase prediction based on buying behavior and comment behavior[J]. *Journal of Beijing University of Posts and Telecommunications (social sciences edition)*, 2016, 18(4): 18–25.
- [12] 张春莲. 客户购买行为的 BG/NBD 预测模型及其应用研究 [D]. 哈尔滨: 哈尔滨工业大学, 2006.
- ZHANG Chunlian. Research on BG/NBD prediction model and its application of the customer purchase behavior[D]. Harbin: Harbin Institute of Technology, 2006.
- [13] 刘柏麟. 基于客户网络购物行为分析及预测系统研究 [D]. 北京: 北京工业大学, 2016.
- LIU Bailin. The research on the analysis of customer's online shopping behavior and prediction system[D]. Beijing: Beijing University of Technology, 2016.
- [14] 陈洁, 谢文昕, 杨升荣. 在线渠道消费者动态品牌选择购买率预测 [J]. *工业工程与管理*, 2011, 16(3): 52–56.
- CHEN Jie, XIE Wenxin, YANG Shengrong. Forecasting the purchase rate of online consumer's dynamic brand choice[J]. *Industrial engineering and management*, 2011, 16(3): 52–56.
- [15] 舒方, 马少辉. 客户重复购买的组合预测方法 [J]. *计算机与现代化*, 2015(5): 67–70.
- SHU Fang, MA Shaohui. A composition forecasting approach of customer repeat purchasing[J]. *Computer and modernization*, 2015(5): 67–70.
- [16] 胡晓丽, 张会兵, 董俊超, 等. 基于 CNN-LSTM 的用户购买行为预测模型 [J]. *计算机应用与软件*, 2020, 37(6): 59–64.
- HU Xiaoli, ZHANG Huibing, DONG Junchao, et al. Prediction model of user buying behavior based on CNN-LSTM[J]. *Computer applications and software*, 2020, 37(6): 59–64.
- [17] 卢泓宇, 张敏, 刘奕群, 等. 卷积神经网络特征重要性分析及增强特征选择模型 [J]. *软件学报*, 2017, 28(11): 2879–2890.
- LU Hongyu, ZHANG Min, LIU Yiqun, et al. Convolution neural network feature importance analysis and feature selection enhanced model[J]. *Journal of software*, 2017, 28(11): 2879–2890.
- [18] ZHANG Weinan, DU Tianming, WANG Jun. Deep learning over multi-field categorical data[C]//*Proceedings of 38th European Conference on Information Retrieval*. Padua, Italy, 2016: 45–57.
- [19] GUO Huifeng, TANG Ruiming, YE Yunming, et al. DeepFM: a factorization-machine based neural network for CTR prediction[C]//*The International Joint Conference on Artificial Intelligence(IJCAI)*. Melbourne, Australia, 2017: 1725–1731.
- [20] WANG Ruoxi, FU Bin, FU Gang, et al. Deep and cross network for Ad click predictions[C]//*Proceedings of the ADKDD'17*. NS, Canada, 2017: 1–7.

#### 作者简介:



程鹏超, 硕士研究生, 主要研究方向为机器学习、广告推荐、信息检索。



杜军平, 教授, 博士生导师, 主要研究方向为人工智能、社交网络分析、数据挖掘、运动图像处理。主持国家重点研发计划、“863”、“973”计划项目、国家自然科学基金重点项目、国家自然科学基金重大国际合作项目、北京市自然科学基金重点项目等多项。发表学术论文 400 余篇, 出版学术专著 6 部。



薛哲, 副教授, 主要研究方向为机器学习、人工智能、数据挖掘、图像处理。主持国家自然科学基金青年基金项目、参与国家重点研发计划项目等多项。发表学术论文 30 余篇, 出版学术专著 1 部。