

短视频用户行为数据的深度学习框架分析

段玉凤¹
DUAN Yufeng

摘要

针对短视频用户行为数据规模庞大、时序依赖复杂及特征提取困难等问题，文章构建了一套数据预处理、特征融合与多任务建模的深度学习分析框架。该框架采用卷积神经网络（CNN）与长短期记忆网络（LSTM）相结合的融合模型，分别应用于行为序列预测、用户偏好识别、行为动机分类三大核心任务，并通过双向长短期记忆网络（Bi-LSTM）与注意力机制进一步增强时序建模与特征聚焦能力。实验结果表明，所提融合模型在三大任务中的整体预测精度分别为 89%、92% 和 90%，综合性能显著优于单一 CNN、LSTM 等主流模型，有效解决了单一模型难以兼顾局部特征提取与长时序依赖捕捉的瓶颈。

关键词

短视频用户行为；深度学习；卷积神经网络；长短期记忆网络

doi: 10.3969/j.issn.1672-9528.2026.05.011

0 引言

用户在短视频平台上的行为数据，如观看时间、点赞、评论和分享等，蕴含着丰富的信息，能够揭示用户偏好和行为习惯^[1]，然而现有研究在分析这些数据时仍面临一些挑战，例如数据规模庞大、实时性要求高以及用户行为复杂多变等

问题。本研究通过运用深度学习技术，构建一个集成数据输入、数据预处理、特征工程、深度学习模型与评估的框架，以系统地分析短视频用户行为数据^[2]。本文主要采用卷积神经网络（convolutional neural network, CNN）和长短期记忆网络（long short-term memory, LSTM）来进行特征提取和时序分析，CNN 能够有效提取用户行为数据中的局部特征，而 LSTM 则擅长处理时间序列数据中的长程依赖关系，这种结

1. 运城开放大学开放教育部 山西运城 044000

[6]Shao Rui , Wu Tianxing , Wu Jianlong ,et al.Detecting and grounding multi-modal media manipulation and beyond[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(8):5556-5574.

[7]Xu Junfeng, Liu Xintao, Lin Weiguo, et al. Localization and detection of deepfake videos based on self-blending method[J]. Scientific Reports, 15(1): 3927-3927.

[8] Abdelnabi S, Hasan R, Fritz M. Open-domain, content-based, multi-modal fact-checking of out-of-context images via online resources[PP/OL].(2022-03-20)[2026-04-27]. <https://doi.org/10.48550/arXiv.2112.00061>.

[9] Aneja S, Bregler C, Nießner M.COSMOS: catching out-of-context misinformation with self-supervised learning[PP/OL](2021-04-21)[2026-04-27].<https://doi.org/10.48550/arXiv.2101.06278>.

[10]Jin Zhiwei, Cao Juan, Guo Han, et al. Multimodal fusion with recurrent neural networks for rumor detection on microblogs[C]//MM '17: Proceedings of the 25th ACM international

conference on Multimedia. NewYork: ACM, 2017: 795 -816.

[11] Khattar D, Goud J S, Gupta M,et al.MVAE: multimodal variational autoencoder for fake news detection[C]//Proceedings of the 2019 World Wide Web Conference. NewYork: ACM, 2019: 2915-2921.

[12]Liu Huan, Tan Zichang , Chen Qiang,et al.Unified frequency-assisted transformer framework for detecting and grounding multi-modal manipulation[PP/OL].(2023-09-18)[2026-04-27]. <https://doi.org/10.48550/arXiv.2309.09667>.

[13] Kobayashi H, Hou Yufang, Ng V.Constrained multi-task learning for bridging resolution[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Brussels: ACL, 2022: 759-770.

【作者简介】

张洛笙（2004—），男，北京人，本科，研究方向：信息安全。

（收稿日期：2025-10-10 修回日期：2026-05-09）

合能够更加准确地识别用户观看习惯和交互模式，帮助理解用户的偏好和行为趋势，为短视频平台优化用户体验提供了重要参考。

1 深度学习分析框架

为实现短视频用户行为数据的精准分析与价值挖掘，本研究构建了一套深度学习分析框架，包括数据输入、数据预处理、特征工程、深度学习模型、模型评估五个核心层级^[3]，如图1所示。



图1 短视频用户行为数据的深度学习分析框架

(1) 数据输入层，实现短视频用户相关数据的全面采集，涉及四类核心数据：用户基础数据、用户行为数据、内容特征数据、场景时序数据，多源数据的全面接入，确保了后续分析的全面性与多维性，避免了单一数据维度导致的分析偏差。

(2) 数据预处理层，实现对原始数据的净化、转换与规范化处理：包括数据清洗、数据标准化、类别数据编码、时序数据处理、数据降维、数据集划分等环节。其中数据降维环节，重点针对高维数据采用了t-分布随机邻域嵌入(t-distributed stochastic neighbor embedding, t-SNE)降维算法，以精准捕捉数据中的非线性结构与局部特征关联，适配短视频用户行为数据中多源异构特征；数据集划分按照7:2:1的比例将处理后的数据划分为训练集、验证集与测试集，分别用于模型训练、参数调优与性能验证。

(3) 特征工程层，主要实现三类特征的提取与融合：一是行为序列特征，基于用户的历史行为数据，提取行为发生的顺序、频次、时间间隔等序列特征，反映用户行为的动态变化规律；二是用户偏好特征，通过统计用户对不同类型视频的互动频率、停留时长等数据，构建用户偏好向量，精准刻画用户的内容偏好；三是时空关联特征，结合用户行为的时间与地域信息，提取时段行为特征、地域行为特征等，挖掘时空因素对用户行为的影响。

(4) 深度学习模型层，针对短视频用户行为分析的三

大核心任务，构建三类专属的深度学习模型：一是行为序列预测模型，以LSTM为核心，融合CNN提取的局部特征，预测用户未来行为趋势；二是用户偏好识别模型，结合CNN局部偏好特征提取与LSTM时序规律捕捉^[4]，辅以注意力机制，精准归类用户观看习惯与内容偏好；三是行为动机分类模型，采用“CNN+双向长短期记忆网络(bidirectional long short-term memory, Bi-LSTM)+全连接层”架构，实现行为动机的精准分类^[5]。

(5) 模型评估层，该层对构建的复杂深度学习模型进行全面的性能验证与迭代优化，主要涉及环节：评估指标设计，分别对分类任务与回归任务进行指标设计^[6]；实验设计，将构建的深度学习模型与基础深度学习模型LSTM、CNN进行性能对比，验证该模型的优越性；模型优化，针对模型存在的过拟合、收敛速度慢等问题，采用正则化、学习率衰减、早停机制等策略进行优化。

2 深度学习模型层的关键模型实现

深度学习任务模型各自承担着不同的功能，每个模型都通过深度学习技术对复杂的数据进行处理，以提取有价值的信息。

2.1 行为序列预测模型

行为序列预测模型首先由CNN从用户行为数据中提取重要的局部特征，然后将特征输入LSTM中，构建时序预测网络。在时序预测网络中，LSTM通过门控机制(遗忘门、输入门、输出门)管理信息流^[7]，其中遗忘门决定了前一个时刻的记忆单元状态中哪些信息需要丢弃，计算公式为：

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

式中： σ 表示sigmoid激活函数； W_f 表示权重矩阵； b_f 表示偏置输出； $[h_{t-1}, x_t]$ 表示连接向量； h_{t-1} 表示前一个时间步的隐含状态(输出)； x_t 表示当前时间步的输入向量； f_t 表示数值范围在0~1之间的向量，乘以上一时刻的单元状态，从而决定信息保留的比例。

输入门决定当前时刻输入信息的多少被存入单元状态，计算过程包括：信息更新计算，如式(2)所示；创建新的候选值，如式(3)所示。

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

式中： W_i 表示输入门的权重矩阵； b_i 表示输入门的偏置； W_c 表示候选单元状态的权重矩阵； b_c 表示候选单元状态偏置； $\tanh$ 表示双曲正切激活函数，用于将线性组合的结果压缩到-1~1之间； i_t 表示输入门的激活值，是一个范围为(0,1)的向量，该值决定了当前输入信息被更新到单元状态中的比例，将 i_t 与式(3)的计算结果候选值 $\tilde{C}_t$ 相乘，再与遗忘门处理后的 C_{t-1} 相加，即可完成单元状态的更新，同时也保留了

重要信息。

输出门用于决定当前时刻的输出，其计算公式为：

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (4)$$

式中： W_o 表示输出门的权重矩阵； b_o 表示输出门的偏置； o_t 表示输出门的激活值，用来决定当前时刻的单元状态对输出的影响程度，捕捉用户行为的时序依赖关系，实现对用户未来行为趋势的有效预判。

2.2 用户偏好识别模型

用户偏好识别模型主要采用了“CNN+LSTM”的串联结构，过程如下：

(1) CNN提取用户-内容交互的局部偏好特征，通过卷积核滑动窗口操作发现交互数据中的局部模式并生成高层次的特征表示^[8]。用公式表示为：

$$y_{i,j,k} = \sum_{m=0}^{m-1} \sum_{n=0}^{n-1} x_{i+m,j+n,c} \cdot W_{m,n,c,k} + b_k \quad (5)$$

式中： $y_{i,j,k}$ 表示输出特征图在位置 (i,j) 的第 k 个通道的值； $x_{i+m,j+n,c}$ 表示输入数据在 $(i+m,j+n)$ 位置的第 c 个通道的值； $W_{m,n,c,k}$ 表示卷积核在 (m,n) 的第 c 个输入通道和第 k 个输出通道的权重； b_k 表示第 k 个输出通道的偏置。

针对特征图采用最大池化进行特征降维，采用非线性激活函数进一步增强模型的表达能力。

(2) LSTM挖掘时序偏好变化^[9]，将CNN输出的特征向量作为LSTM的输入序列，每个时间步表示用户在该时刻的特征，LSTM的输出则表示了用户偏好的时间序列变化，为后续步骤提供时序上下文信息，同时避免了传统RNN中梯度消散或梯度爆炸的问题。

(3) 引入注意力机制强化核心特征权重^[10]，在LSTM的输出上应用注意力机制，计算每个时间步的上下文重要性，将LSTM的输出进行加权求和，得到一个上下文向量，以代表用户偏好的综合特征。

2.3 行为动机分类模型

构建“CNN+Bi-LSTM+全连接层”架构，其中CNN提取行为触发的局部特征，CNN模块设置为“Conv2d(3×3, 32)→LeakyReLU→MaxPool2d(2×2)→Conv2d(3×3, 64)→LeakyReLU→MaxPool2d(2×2)”，选择3×3卷积核是因为预实验显示，其对用户行为“时段-交互”局部模式的提取能力优于1×1（特征太浅）和5×5（冗余特征多）；Bi-LSTM双向捕捉时序行为关联，该模块设置为“256神经元+双向+Dropout=0.3”，初始选择128神经元时模型欠拟合（验证集精度仅82%），提升至256后精度达89%，继续提升至512则出现过拟合；全连接层实现行为动机类别的映射，将CNN输出的4×4×64特征展平为1024维向量，与Bi-LSTM的256×2=512维双向输出拼接为1536维特征，该拼接方式保留了局部特征与时序特征的独立性，最终精度提

升2%。

3 实验验证

为验证“CNN+Bi-LSTM”融合架构的技术合理性，本实验聚焦消融实验与超参数敏感性分析两类核心验证逻辑，通过量化模块贡献、优化训练细节，明确融合模型在短视频用户行为分析任务中的技术优势，同时规避单一模型的性能瓶颈。

3.1 评估指标

针对短视频用户行为分析的三大核心任务特性，结合回归任务与分类任务的评价标准，选取精确率（Precision, P ）、召回率（Recall, R ）、 F_1 值、整体预测精度（Accuracy, ACC）、平均绝对误差（MAE）作为核心评估指标，全面量化模型在常规场景与高峰/突发性行为场景下的性能表现。各指标计算公式为：

$$S_P = [N_{TP} / (N_{TP} + N_{FP})] \times 100\% \quad (6)$$

$$S_R = [N_{TP} / (N_{TP} + N_{FN})] \times 100\% \quad (7)$$

$$F_1 = 2 \times \frac{S_P \times S_R}{S_P + S_R} \quad (8)$$

$$ACC = [(N_{TP} + N_{TN}) / (N_{TP} + N_{TN} + N_{FP} + N_{FN})] \times 100\% \quad (9)$$

$$MAR = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (10)$$

式中： N_{TP} 为真正样本且预测为正的数量； N_{FP} 为真实负样本但预测为正的数量； N_{FN} 为真正样本但预测为负的数量； N_{TN} 为真实负样本且预测为负的数量； $\hat{y}_i$ 为模型预测值； y_i 为真实值； N 为样本总数。

3.2 消融实验

为验证深度学习分析框架中CNN局部特征提取模块、LSTM(Bi-LSTM)时序建模模块、注意力机制特征聚焦模块各改进点的独立有效性及融合协同增益，设计4组消融实验，以单一LSTM模型为基础基线（实验1），依次加入各核心模块进行对比，所有实验保持超参数（学习率1e-4、Adam优化器、交叉熵损失函数、训练轮数100）一致，仅改变特征提取与建模模块，实验结果如表1所示。

由实验结果可知，实验2在基线模型中加入CNN局部特征提取模块后，行为序列预测MAE降至0.24，用户偏好识别精度提升6.3%，行为动机分类 F_1 值提升0.05，三大任务平均精度提升5.5%，验证了CNN模块对用户行为局部关联特征的有效提取能力；实验3在实验2基础上引入Bi-LSTM双向时序建模模块，相比实验2平均精度进一步提升1.6%，说明Bi-LSTM模块能更全面地挖掘行为时序规律；实验4为本文最终构建的融合模型，加入注意力机制后，模

表 1 融合模型消融实验结果

实验组	模型配置	行为序列预测 MAE	用户偏好识别 ACC /%	行为动机分类 F_1 值	三大任务平均精度 /%
1	单一 LSTM 模型（基线）	0.35	82.6	0.81	83.2
2	LSTM+CNN 模块	0.24	88.9	0.86	88.7
3	LSTM+CNN+Bi-LSTM 模块	0.20	90.5	0.88	90.3
4	LSTM+CNN+Bi-LSTM+ 注意力机制（本文模型）	0.18	92.0	0.90	91.6

型对核心行为特征的聚焦能力显著增强，行为序列预测 MAE 降至 0.18，用户偏好识别精度达 92.0%，行为动机分类 F_1 值达 0.90，三大任务平均精度提升至 91.6%，相比基线模型提升 8.4%，充分验证了各模块融合的合理性与协同性。

3.3 对比实验

为验证本文融合模型的优越性，选取短视频用户行为分析领域的主流算法与单一基础模型作为对比对象，设置单一 CNN 模型、单一 LSTM 模型、CNN+GRU 融合模型、Transformer 时序模型 4 类对比模型，与本文 CNN+Bi-LSTM+ 注意力机制融合模型在相同实验环境、数据集与评估指标下进行性能对比，重点对比各模型在常规行为场景与高峰 / 突发性行为场景下的表现，实验结果如表 2 所示。

由实验结果可知，单一 CNN 模型与单一 LSTM 模型性能最差，其中单一 CNN 模型因缺乏时序建模能力，在高峰 / 突发场景下精度仅 75.2%，MAE 达 0.42；单一 LSTM 模型虽能捕捉时序特征，但对局部特征提取不足，整体性能受限；CNN+GRU 融合模型相比单一模型性能有明显提升，但 GRU 对长时序依赖的捕捉能力弱于 LSTM，且无双向建模能力，高峰 / 突发场景下精度仅 82.5%，低于本文模型；Transformer 时序模型虽在时序特征捕捉上表现较好，常规场景精度达 88.7%，但因模型复杂度高，推理速度仅 68.9 ms/ 样本，且对局部特征提取能力不足，高峰 / 突发场景精度仍低于本文模型；本文模型在常规场景与高峰 / 突发场景下的平均精度分别达 91.6% 与 89.5%，均为所有对比模型中最高，行为序

表 2 不同模型性能对比实验结果

模型类型	常规场景平均精度 /%	高峰 / 突发场景平均精度 /%	行为序列预测 MAE	模型推理速度 / (ms·样本 ⁻¹)
单一 CNN 模型	78.5	75.2	0.42	12.5
单一 LSTM 模型	83.2	79.0	0.35	28.6
CNN+GRU 融合模型	86.9	82.5	0.26	25.3
Transformer 时序模型	88.7	85.8	0.22	68.9
本文 CNN+Bi-LSTM + 注意力机制模型	91.6	89.5	0.18	32.1

列预测 MAE 降至 0.18，同时保持 32.1 ms/ 样本的推理速度，在精度与效率之间实现了良好平衡。

4 结语

本研究针对用户在短视频平台上的行为数据动态分析需求，采用 CNN+LSTM（Bi-LSTM）融合算法构建了深度学习模型，解决了单一 CNN 模型难以有效捕捉用户行

为长时序依赖、单一 LSTM 模型对局部突发性行为特征提取精度有限的问题，实现了对短视频用户行为序列预测、用户偏好识别、行为动机分类三大核心任务的高效统一建模。未来研究将引入注意力机制以强化模型对关键行为特征的聚焦能力，从而提升模型的泛化性能。

参考文献：

[1] 刘贺, 刘春生. 大数据背景下基于用户偏好的短视频推荐算法研究 [J]. 广播电视网络, 2024, 31(6): 104-106.

[2] 刘云浩. 基于 CNN-LSTM 的航空发动机剩余使用寿命预测 [D]. 哈尔滨: 哈尔滨工业大学, 2025.

[3] 史扬. 电商平台销售数据的统计分析与用户行为模式挖掘 [J]. 中国电子商情, 2024(22): 42-44.

[4] 谢明权. 基于多维时序数据的学生成绩预测及分析 [D]. 桂林: 桂林电子科技大学, 2024.

[5] 李慧玲. 基于数据挖掘的金融时序数据分析 [J]. 兵工自动化, 2024, 43(12): 35-37.

[6] 谢晓宇, 范深, 宋金玉, 等. 基于 MOOC 学习平台的学习行为数据分析与预测 [J]. 软件导刊, 2025, 24(2): 198-203.

[7] 徐林菊. 基于 CNN-LSTM 的高速列车转向架故障诊断 [J]. 自动化应用, 2025, 66(18): 47-51.

[8] 乔建民, 李睿, 李宁, 等. 基于改进双向长短期记忆网络的综合能源系统负荷预测 [J]. 电工技术, 2025(19): 40-46.

[9] 方则宇. 面向工业过程数据的时序表征学习与软测量方法研究 [D]. 兰州: 西北师范大学, 2024.

[10] 毛颖颖, 袁浩. 基于自注意力机制的轻量化网络流量检测模型研究 [J]. 信息与电脑, 2026, 38(2): 56-58.

【作者简介】

段玉凤（1982—），女，山西山阴人，硕士，讲师，研究方向：计算机技术。
（收稿日期：2026-03-04 修回日期：2026-04-17）