

基于Transformer-LSTM算法的电子商务平台用户行为预测研究

周文钦

浙江建设职业技术学院建筑工程学院, 浙江 杭州

收稿日期: 2025年6月4日; 录用日期: 2025年6月23日; 发布日期: 2025年7月23日

摘 要

本文探讨了在电子商务领域, 利用Transformer-LSTM混合神经网络算法对用户行为进行预测的方法。随着网络购物的普及, 用户浏览、收藏、加入购物车等在电商平台产生的大量行为数据成为预测其购买行为的关键信息。在处理此类序列性和依附关系复杂的数据时, 传统的机器学习方法有一定的局限性。因此, 本文提出了一种结合Transformer和LSTM的混合神经网络模型, 以提高预测的准确性和效率。Transformer模型擅长在序列数据中捕捉长时间的依赖关系和并行计算, 而LSTM则可以针对梯度消失和爆炸等问题, 对长序列数据进行有效处理。本文的模型首先通过Transformer对用户行为数据进行编码, 提取关键特征, 然后利用LSTM对提取的特征序列进行建模, 最终实现对用户购买行为的预测。实验结果显示, 该模型比传统的LSTM模型在精确度、精确率、召回率、F1-score等指标上都要好, 在电商用户行为预测上显示出潜力和优势。电商平台通过这种改良的模式, 能够将商品更精准地推荐给用户, 提升用户的购物体验, 同时也提高了平台的经济效率。

关键词

Transformer-LSTM, LSTM, 用户行为预测

Research on User Behavior Prediction of E-Commerce Platform Based on Transformer-LSTM Algorithm

Wenqin Zhou

School of Architectural Engineering, Zhejiang College of Construction, Hangzhou Zhejiang

Received: Jun. 4th, 2025; accepted: Jun. 23rd, 2025; published: Jul. 23rd, 2025

文章引用: 周文钦. 基于 Transformer-LSTM 算法的电子商务平台用户行为预测研究[J]. 电子商务评论, 2025, 14(7): 1902-1910. DOI: 10.12677/ec.2025.1472385

Abstract

This paper discusses the method of predicting user behavior by using the Transformer-LSTM hybrid neural network algorithm in the field of e-commerce. With the popularity of online shopping, the large amount of behavioral data generated by users on e-commerce platforms (such as browsing, collecting, adding to shopping carts, etc.) has become the key information for predicting their purchasing behaviors. Traditional machine learning methods have certain limitations when dealing with such data that have complex sequentiality and dependencies. Therefore, this paper proposes a hybrid neural network model combining Transformer and LSTM to improve the accuracy and efficiency of prediction. The Transformer model is good at capturing long-term dependencies and parallel computing in sequence data, while LSTM can effectively handle long sequence data and solve the problems of vanishing and exploding gradients. The model in this paper first encodes the user behavior data through Transformer to extract key features, and then uses LSTM to model the extracted feature sequence, ultimately achieving the prediction of user purchasing behavior. The experimental results show that this model outperforms the traditional LSTM model in terms of indicators such as accuracy rate, precision rate, recall rate and F1-score, demonstrating its potential and advantages in the prediction of e-commerce user behaviors. Through this improved model, e-commerce platforms can recommend products to users more accurately, enhancing the shopping experience of users and the economic benefits of the platform.

Keywords

Transformer-LSTM, LSTM, User Behavior Prediction

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,随着互联网的普及,在电子商务网站进行在线购物已成为一种流行的购物方式。在在线购物过程中,用户在决定购买商品之前会在电子商务平台上产生大量的行为数据,如浏览、收藏、加入购物车等。这些数据蕴含着丰富的用户行为模式和购物习惯信息,通过数据挖掘和分析这些数据,能够更好地理解用户的购物倾向,为预测用户的购买行为提供可能[1]。

准确预测用户的购物行为对电子商务平台具有重要意义。通过预测结果,平台可以个性化地向用户推荐商品,提高用户的购物效率和满意度,促成更多交易,增加营业收入。因此,众多国内外大型电子商务企业纷纷运用商品推荐算法,学者们也不断探索新的方法来提高预测的准确度。

传统的机器学习方法,如随机森林、逻辑回归和支持向量机(SVM)等,已经在用户行为预测方面取得了一定的成果。例如,雷名龙[2]采用这些算法对阿里巴巴电子商务平台4个月的购物数据进行分析,对用户未来是否会购买某种商品做出预测,取得了超过5%的准确率。张春生等[3]考察了品牌可信度、价格、付款人数等多种评价指标对于用户购买行为的相关性。此外,深度学习方法也在该领域崭露头角,如Vieira等[4]采用深度置信网络和自编码器等策略进行建模,发现其效果优于传统的决策树和随机森林等算法。

然而,用户行为数据往往具有复杂的序列性和依赖关系,传统的机器学习和深度学习方法在处理这类数据时存在一定局限性[5]。Transformer模型凭借其强大的并行计算能力和对序列数据中长期依赖关系

的有效捕捉，在自然语言处理等领域取得了显著成果。长短期记忆网络(LSTM)作为一种特殊类型的循环神经网络(RNN)，能够有效解决传统 RNN 在处理长序列数据时的梯度消失和梯度爆炸问题，适用于处理具有时间序列特性的用户行为数据[6]。

本文创新性地将 Transformer 与 LSTM 相结合，提出一种 Transformer-LSTM 混合神经网络算法用于电子商务用户行为预测。该算法首先利用 Transformer 对用户行为数据进行编码，提取关键特征，然后通过 LSTM 对提取的特征序列进行建模，以预测用户的购买行为。这种结合充分发挥了 Transformer 在特征提取和 LSTM 在序列建模方面的优势，有望进一步提高用户行为预测的准确度，为电子商务平台提供更精准的商品推荐服务。

2. 数据描述

本文所使用的数据来自阿里巴巴天池大数据竞赛公开数据集，包含 20,000 名用户某年 11 月 18 日至 12 月 18 日一个月的完整行为数据。每条购物行为包含 4 个字段，分别为用户 ID、品牌 ID、用户对商品的交互行为和行为时间。用户与商品的交互行为分为“浏览”“收藏”“加入购物车”和“购买”。

2.1. 异常值剔除

异常值的存在通常会严重影响建模和预测质量[7]，因此需要去除数据中存在的异常值。在获取的数据时间内包含了淘宝“双 12”购物节，用户当日总浏览、领取、添加购物车、购买次数分别是以往平均值的 1.8、1.4、2.4、4.5 倍，属于明显的异常数值，故在后续处理过程中去除了当日所有数据。另外，一个月内没有购买记录的用户，对于预测建模没有参考价值的这类用户，可能没有网购习惯，所以也剔除了这类数据[8]。

2.2. 特征筛选

在建模过程中，由于原始数据无法直接使用，需将其归纳为能充分描述商品、用户信息及用户 - 商品交互行为的统计特征。本文依据相关性、代表性和可解释性三个原则进行特征选择。所谓相关性，即特征需与用户购买行为强相关，能有效反映用户兴趣和购买可能；代表性意味着特征要能体现商品热度、用户购物习惯及交互模式；可解释性则要求特征具备明确业务含义，便于理解模型预测结果。基于这些原则，本文选定了交互特征、商品特征和用户特征三类共 31 个特征[9]。

针对特征相关性问题，本文计算了所有特征的皮尔逊相关系数，以此识别高度相关特征对，并设定相关系数绝对值大于 0.8 为阈值。对于高度相关的特征，依据其解释能力和对目标变量的影响，选择保留更具代表性和解释性的特征。同时，将部分高度相关特征进行合并或转换，如求和、求差、求比等，从而减少特征间冗余信息[10]。

通过以上严谨的特征筛选和处理流程，本文不仅确保了所选特征具有实际业务意义，还有效避免了多重共线性问题，为模型训练提供了高质量的输入数据，特征提取见表 1。

Table 1. Statistics for modeling
表 1. 用于建模的统计

特征描述	特征种类	编号
用户 1 日内对该商品的各类交互次数	交互特征	1~4
用户 1 日内对同类商品的各类交互次数	交互特征	5~8
用户 3 日内对该商品的各类交互次数	交互特征	9~12

续表

用户 3 日内对同类商品的各类交互次数	交互特征	13~16
商品总浏览(收藏、加入购物车、购买)量	商品特征	17~20
商品最近 3 日总浏览(收藏、加入购物车、购买)量	商品特征	21~24
用户总浏览(收藏、加入购物车、购买)量	用户特征	25~28
用户浏览(收藏、加入购物车)购买量比	用户特征	29~31

2.3. 样本划分

为了合理划分数据集，同时保证数据的随机性和代表性，本文在处理大量数据时采取了如下策略：

对于训练集的划分，本文选择了第 8、15、22 天的数据，这三天每隔一周选取一次，具有较好的时间分散性。从每天的数据中随机抽取 20,000 个样本，共 60,000 个样本构成训练集。训练集的每一个样本都是一个用户 - 商品配对，特征统计涵盖前三天的信息，这样的划分方式既能保证样本的时间顺序，又能较好地代表不同时期的用户行为特征。

此外，第 23 天的 60,000 个样本被选为测试集。测试集样本同样遵循时间顺序，且与训练集之间保持一定的间隔，以便真实地反映模型对新数据的预测能力。为保证数据的随机性，在抽取样本时，本文采用了随机抽样的方法，确保从每一天的数据中随机选取样本，避免出现因人为选择导致的偏差。为了应对样本不平衡性的问题，本文采用了过采样和欠采样相结合的方法。对于数量较少的阳性样本(购买行为)，适当增加其在训练集和测试集中的比例，以提高模型对阳性样本的识别能力；同时，对数量较多的阴性样本(非购买行为)进行适当的限制，避免其在数据集中占据绝对优势而影响模型的训练效果。在训练过程中，本文还从训练集中随机抽取 10%的样本作为验证集，用于实时监测模型的训练效果，防止过拟合，并辅助调整模型参数。

最终，本文详细统计了训练集和测试集中各类样本的分布情况，以确保数据集的合理性和多样性。在训练集的 60,000 个样本中，阳性样本为 100 个，而在测试集的 60,000 个样本中，阳性样本为 112 个。这种不平衡的分布反映了实际场景下用户行为的特点，即用户会浏览大量的商品，但实际购买的仅占少数。本文在模型训练过程中，通过调整样本权重等方法来适应这种不平衡性，以提高模型对少数阳性样本的预测性能。

3. 分类建模

3.1. Transformer 算法

如下图 1 所示，Transformer 编码器内部包含嵌入层和多层编码器结构，每层编码器内部含有多个子层：多头注意力层(Multi-head attention)、前馈神经网络层(Feed-forward neural network)、残差连接归一化层。这种结构使得模型能够对用户行为数据进行高效编码，提取关键特征。多头注意力机制允许模型在不同位置和层次上关注用户行为的不同方面，从而更全面地理解用户行为模式。Transformer 示意图见图 1。

3.2. LSTM 算法

长短期记忆网络(LSTM)是一种特殊类型的循环神经网络(Recurrent neural network, RNN)，通过引入了一个记忆神经元解决传统 RNN 模型在处理长序列数据时遇到的梯度消失和梯度爆炸问题。LSTM 网络图见图 2。

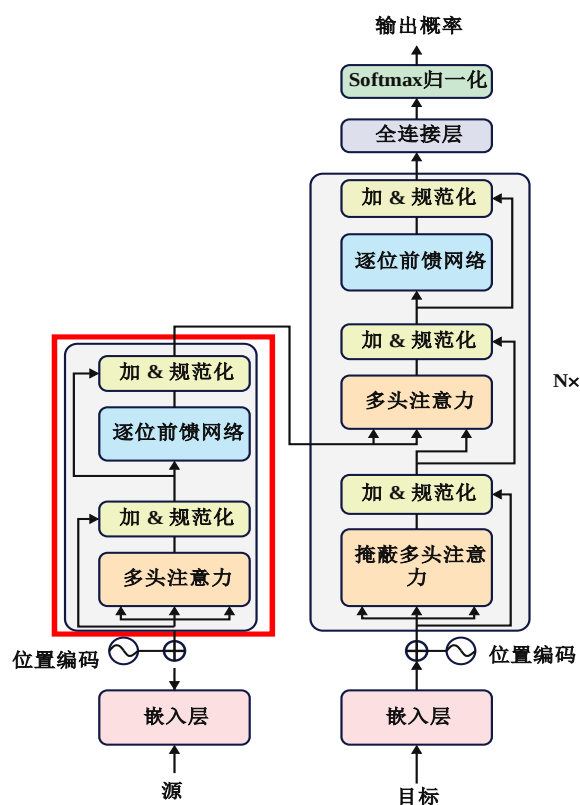


Figure 1. Schematic diagram of Transformer

图 1. Transformer 示意图

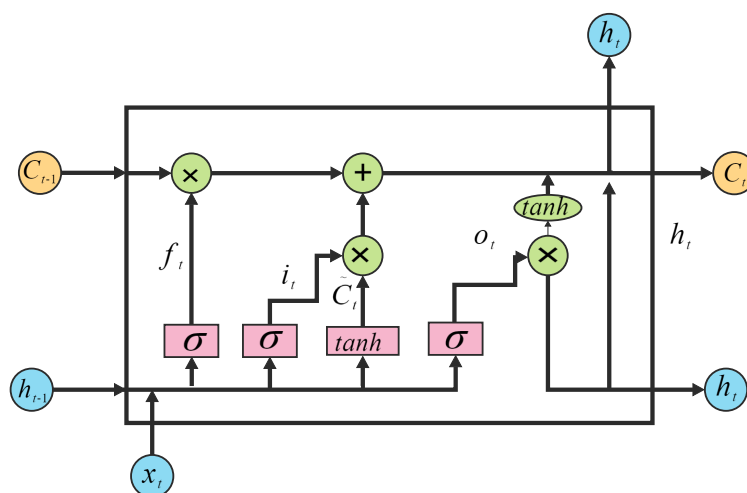


Figure 2. LSTM network diagram

图 2. LSTM 网络图

如图，LSTM 网络的核心是三个门的机制：遗忘门、输入门和输出门，这些门通过自适应的方式控制信息的流动，从而实现对长期以来信息的捕捉。LSTM 模型输入 Transformer 提取的用户行为特征序列，输出预测的用户行为。相应的计算过程如下：

其计算过程如下：

Step 1: 遗忘门:

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

Step 2: 输入门:

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

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

Step 3: 输出门:

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

Step 4: 长期记忆:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

Step 5: 短期记忆:

$$h_t = o_t \cdot \tanh(C_t)$$

其中, σ 表示激活函数, x_t 表示输入数据; h_{t-1} 表示前一时刻输出; b_f, b_i, b_c, b_o 分别为遗忘门, 输入门, 细胞状态, 输出门的偏置; W_f, W_i, W_c, W_o 分别表示遗忘门, 输入门, 细胞状态, 输出门的权重参数矩阵。价格相关特征序列数据输入后, 组成输入组合 $[h_{t-1}, x_t]$, $[h_{t-1}, x_t]$ 输入组合经过遗忘门得到决定保留 C_{t-1} 哪部分特征的概率值 f_t ; $[h_{t-1}, x_t]$ 经过输入门得到的表示现在学到的特征中哪些可以保留的概率值 i_t 与本次学习到的所有特征 $\tilde{C}_t$ 将 C_{t-1} 与 f_t 的乘积即上一时刻保留的特征和 $\tilde{C}_t$ 与 f_t 经的乘积即这一时刻选择的特征求和得到长期记忆 C_t , $[h_{t-1}, x_t]$ 经过输出门得到表示对长期记忆 C_t 即学习到的特征进行筛选后找到的解决当前问题的信息概率值 o_t , o_t 与 $\tanh(C_t)$ 的结果相乘为短期记忆 h_t 即需要的输出。

以下是 LSTM 的模型参数见表 2:

Table 2. Model parameter table of LSTM

表 2. LSTM 的模型参数表

模型名称	层内容	层参数
LSTM	LSTM 层 1	神经元数量 32
	Dropout 层 1	dropout 率为 0.1
	LSTM 层 2	神经元数量 32
	Dropout 层 2	dropout 率为 0.1
	全连接层	神经元数量 3

3.3. Transformer-LSTM 算法

将 Transformer 模型与 LSTM 模型相结合, 能够实现特征提取和序列建模的高效结合。Transformer 模型首先对用户行为数据进行编码, 提取出高维特征表示, 这些特征表示包含了用户行为序列中的长期依赖关系和关键信息。LSTM 模型则利用这些高维特征表示, 进一步捕捉序列中的时间依赖关系, 从而实现对用户购买行为的精准预测。

将 Transformer 与 LSTM 算法结合起来, 设定的模型参数见表 3:

Table 3. Parameter table of Transformer-LSTM hybrid neural network model
表 3. Transformer-LSTM 混合神经网络模型参数表

块名称	层名称	参数	输出尺寸
Transformer	输入层	/	(128, 10, 4)
	嵌入层	神经元数 64	(128, 10, 64)
	多头注意力层	注意力头数 8	(128, 10, 64)
	归一化层 1	输入形状 64	(128, 10, 64)
	Dropout 层 1	dropout 率 0.1	(128, 10, 64)
	前馈神经网络	神经元数 128	(128, 10, 128)
		神经元数 64	(128, 10, 64)
	归一化层 2	输入形状 64	(128, 10, 64)
	Dropout 层 2	dropout 率 0.1	(128, 10, 64)
链接块	全局平均池化层	池化窗口为 10	(128, 1, 64)
LSTM	LSTM 层 1	神经元数 32	(128, 1, 32)
	Dropout 层 3	dropout 率 0.1	(128, 1, 32)
	LSTM 层 2	神经元数 32	(128, 32)
	Dropout 层 4	dropout 率 0.1	(128, 32)
	全连接层(输出层)	神经元数 3	(128, 3)

4. 评价标准

对于商品推荐算法而言，我们比较关注的就是阳性样本的预测是否准确，所以我们采用阳性样本预测(Precision)的准确率(Accuracy)、召回率(Recall)和 F1 值作为评价指标，定义为：

准确率：

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

精确率：

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

召回率：

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

F1-score：

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

其中，P 为阳性样本总数，TP 为正确预测的阳性样本数量，NP 为错误预测的阳性样本数量。

4.1. 结果分析

通过 Transformer-LSTM 和 LSTM 的建模，可以判断每个特征变量对模型的贡献程度，从而判断用户购买行为受哪些特征变量影响比较显著，分析结果见图 3。

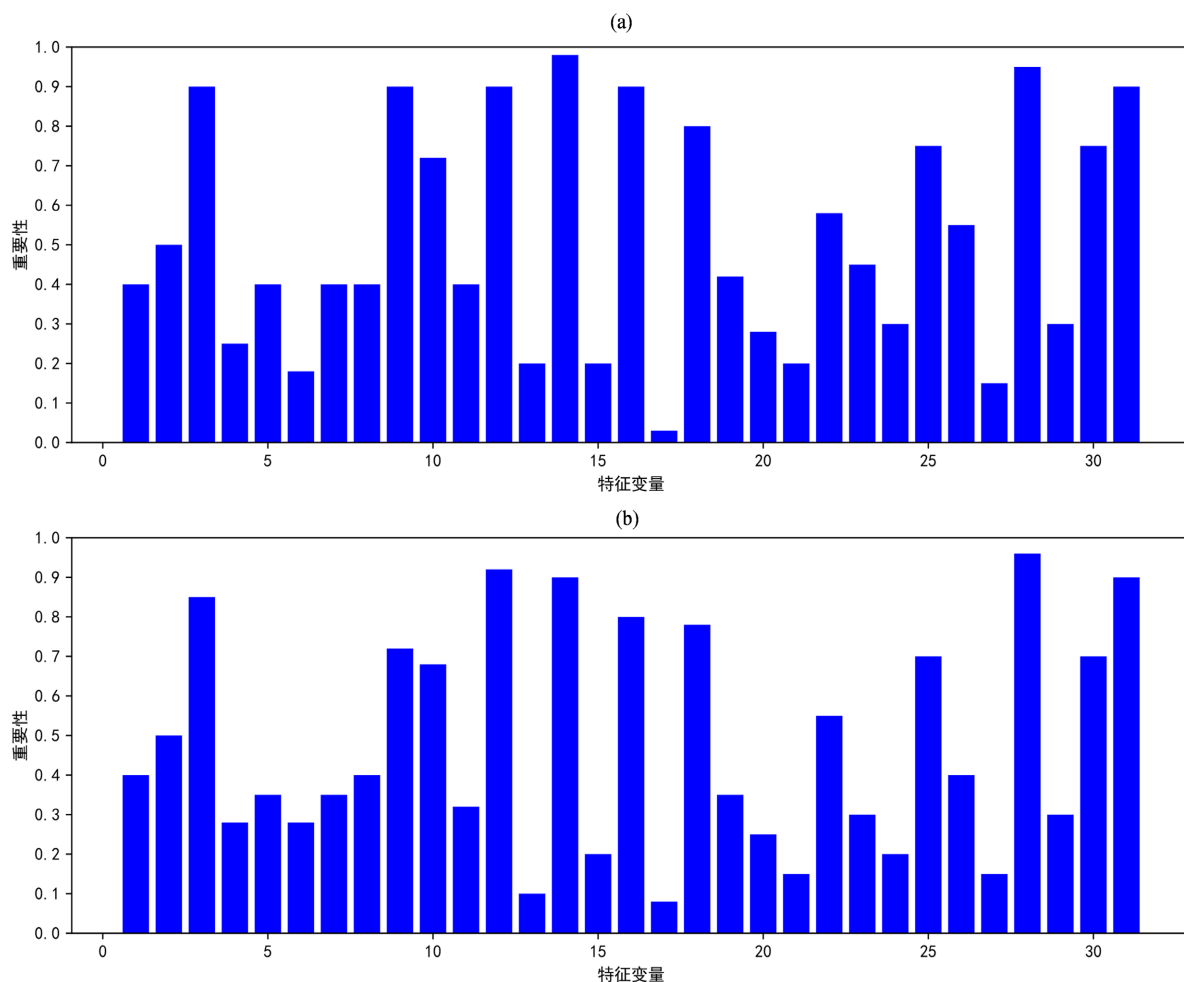


Figure 3. Feature importance map
图 3. 特征重要性图

上图展示了 Transformer-LSTM 和 LSTM 模型的特征重要性结果，A 和 B 分别对应两种模型。在两个子图中，特征重要性值的分布表明特征 17~20 (商品总浏览、收藏、加入购物车、购买量)和 21~24 (商品最近三日的总浏览、收藏、加入购物车、购买量)在两个模型中都属于重要特征，这说明商品相关的特征对预测用户行为具有重要意义。此外，在 Transformer-LSTM 模型中，特征 29~31 (用户浏览、收藏、加入购物车购买量比)也显示出较高的重要性，表明用户行为特征在 Transformer-LSTM 模型中被赋予了更多的权重。相比之下，在 LSTM 模型中，特征 5~8 (用户 1 日内对同类商品的各类交互次数)和 13~16 (用户 3 日内对同类商品的各类交互次数)的重要性有所下降。这表明，Transformer-LSTM 模型在捕捉用户行为特征方面可能具有优势。这些结果表明，Transformer-LSTM 模型在处理电子商务用户行为预测时，能够更有效地识别出关键特征，从而提升预测的准确性和效果。

4.2. 结果比较

通过交互检验分别优化两种算法的参数后，采用测试集样本对建模结果进行预测，最终，本文利用 Transformer-LSTM 与 LSTM 分别预测，并且进行上述准确率、精确率、召回率、F1-score 的检验，结果见表 4:

Table 4. Comparison table of LSTM and Transformer-LSTM effects
表 4. LSTM 与 Transformer-LSTM 效果对比表

模型	准确率	精确率	召回率	F1-score
LSTM	0.86	0.83	0.85	0.84
Transformer-LSTM	0.92	0.91	0.93	0.92

5. 结语

本文采用 Transformer-LSTM 分类算法，基于阿里巴巴的真实用户数据进行了特征提取和分类建模，并与 LSTM 算法进行了对比，得到了较准确的预测结果。通过分析变量的重要程度，对模型中贡献度较高的一个变量进行了鉴别。该研究有助于理解用户浏览信息与其购买行为的相关性，对个性化推荐系统性能的完善有重要的现实意义。

参考文献

- [1] Curme, C., Preis, T., Stanley, H.E., *et al.* (2014) Quantifying the Semantics of Search Behavior before Stock Market Moves. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 11600-11605. <https://doi.org/10.1073/pnas.1324054111>
- [2] 雷名龙. 基于阿里巴巴大数据的购物行为研究[J]. 物联网技术, 2016, 6(5): 57-60.
- [3] 张春生, 图雅, 翁慧, 等. 基于电子商务同类商品的推荐算法研究[J]. 计算机技术与发展, 2016, 26(5): 17-21.
- [4] Vieira, A. (2015) Predicting Online User Behaviour Using Deep Learning Algorithms. arXiv: 1511.06247.
- [5] 马月坤, 刘鹏飞. 基于知识库的客户网购意向预测系统[J]. 计算机工程与应用, 2016, 52(13): 101-109.
- [6] 杨守怀, 黄江流, 陈志华, 等. 基于 Transformer-LSTM 混合神经网络的迫弹外弹道及落点预测方法[J/OL]. 兵工学报: 1-10. <https://link.cnki.net/urlid/11.2176.TJ.20250520.1130.004>, 2025-05-30.
- [7] 王首晨, 王利民. LSTM 与 Transformer 融合模型在时间序列预测中的应用研究[J]. 网络安全与数据治理, 2025, 44(3): 27-31, 38.
- [8] 李沁远, 雷文平, 闫灏, 等. 基于特征选择与 Transformer-LSTM 的滚动轴承寿命预测[J]. 组合机床与自动化加工技术, 2025(2): 200-206, 211.
- [9] 王亿, 曹姗姗, 孙伟, 等. 融合 Transformer 和 LSTM 的蓝莓根区土壤含水量预测模型[J]. 湖北农业科学, 2024, 63(8): 78-84.
- [10] 张帆, 姚德臣, 姚圣卓, 等. 基于 Transformer-LSTM 网络的轴承寿命预测[J]. 振动与冲击, 2024, 43(6): 320-328.