

面向电商营销决策的多模态服装销量预测方法研究

代丽, 杨浩男, 徐靖捷, 王世雄*

浙江理工大学经济管理学院, 浙江 杭州

收稿日期: 2026年3月10日; 录用日期: 2026年3月25日; 发布日期: 2026年5月29日

摘要

服装行业商品上新频繁、需求波动大, 传统仅依赖历史销量序列的预测方法难以准确刻画多因素共同作用下的销量变化规律。为提升预测准确性, 本文提出一种基于多模态信息融合的服装销量预测方法, 将商品图像、文本描述、用户评价与历史销量等多源数据统一建模, 以全面刻画影响销量的关键因素。模型由多模态编码、门控融合、跨模态对齐和GRU解码器四个模块组成。各模态特征分别编码并标准化至统一表示空间; 通过门控机制与跨模态注意力进行动态融合; 对比学习增强图文时序特征的一致性; 最终通过GRU完成未来多步销量预测。实验表明, 该方法在MAE、RMSE、MAPE与R²等指标上均显著优于基线模型; 多步预测结果显示, 模型在短期补货及中期库存规划任务中具备更高稳定性。消融实验进一步验证了门控融合与对比学习模块对提升性能的重要性。研究结果得出, 多模态融合能够有效弥补结构化数据表达不足, 为服装行业的补货策略与库存管理提供更可靠的预测支持。

关键词

多模态融合, 销量预测, 电商服装, 对比学习, 时序建模

Research on Multimodal Clothing Sales Forecasting Methods for E-Commerce Marketing Decision-Making

Li Dai, Haonan Yang, Jingjie Xu, Shixiong Wang*

School of Economics and Management, Zhejiang Sci-Tech University, Hangzhou Zhejiang

Received: March 10, 2026; accepted: March 25, 2026; published: May 29, 2026

*通讯作者。

文章引用: 代丽, 杨浩男, 徐靖捷, 王世雄. 面向电商营销决策的多模态服装销量预测方法研究[J]. 电子商务评论, 2026, 15(5): 805-813. DOI: 10.12677/ecl.2026.155580

Abstract

The clothing industry is characterized by frequent product updates and significant demand fluctuations, and traditional forecasting methods relying solely on historical sales sequences struggle to accurately capture sales variation patterns under the combined influence of multiple factors. To improve prediction accuracy, this paper proposes a multimodal clothing sales forecasting method based on multimodal information fusion, which jointly models multiple data sources, including product images, textual descriptions, user reviews, and historical sales data, to comprehensively characterize the key factors affecting sales. The proposed model consists of four modules: multimodal encoding, gated fusion, cross-modal alignment, and a GRU decoder. Features from different modalities are separately encoded and normalized into a unified representation space; dynamic fusion is achieved through a gating mechanism and cross-modal attention; contrastive learning is employed to enhance the consistency of image, text, and temporal features; and finally, a GRU decoder is utilized to perform future multi-step sales forecasting. Experimental results show that the proposed method significantly outperforms baseline models in terms of MAE, RMSE, MAPE, and R^2 . Multi-step forecasting results further demonstrate that the model exhibits higher stability in short-term replenishment and medium-term inventory planning tasks. Ablation studies further verify the importance of the gated fusion and contrastive learning modules in improving predictive performance. The findings indicate that multimodal fusion can effectively compensate for the limitations of structured data representation and provide more reliable forecasting support for replenishment strategies and inventory management in the clothing industry.

Keywords

Multi-Modal Fusion, Sales Forecasting, Apparel E-Commerce, Contrastive Learning, Time-Series Modeling

Copyright © 2026 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

随着电商平台与快时尚品牌的广泛发展, 服装行业逐渐呈现出商品生命周期短、上新频率高以及市场需求波动剧烈等特征, 企业在库存管理与供应链协同过程中面临更高的不确定性与决策压力[1][2]。在此背景下, 准确刻画服装商品的销量变化规律, 已成为提升运营效率与降低成本的重要基础。然而, 服装产品普遍具有季节性强、流行周期短、款式差异显著等特点, 其销量形成机制复杂多变, 单纯依赖历史销量序列的预测方法难以充分刻画多因素共同作用下的动态变化过程[3]。既有研究表明, 电商场景下的商品销量不仅受到历史销售趋势的影响, 还与商品视觉风格、文本描述、消费者评价以及搜索行为等多源信息密切相关, 多模态信息在反映消费者偏好与市场变化方面具有重要价值[4]-[6]。因此, 如何在充分利用多模态信息的基础上, 提高服装销量预测的准确性与稳定性, 成为当前服装电商运营实践中亟待关注的重要问题。

从方法层面看, 传统统计与机器学习模型主要依赖历史销量数据建模, 对非结构化信息利用有限[7]-[10]。多模态建模虽已成为趋势, 但现有方法在模态间动态融合与语义对齐方面仍存在不足[11][12]。

近年来, 循环神经网络与 Transformer 在时序建模中展现出较强能力[13][14], 视觉语言预训练模型推动了异构模态的深度融合[15]-[17], 但将上述技术有效整合应用于服装销量预测仍需针对性设计。

基于此,本文围绕服装电商场景下的销量预测任务,提出一种基于多模态信息融合的预测方法,综合利用历史销量序列、商品图像与文本信息,对影响销量变化的多源因素进行统一建模。该方法通过多模态特征编码与门控融合机制实现不同模态信息的动态协同,引入跨模态注意力与对比学习策略增强多源特征的一致性表达,并在此基础上构建多步销量预测模型,实现对未来销售趋势的精细化刻画。

2. 文献综述

(一) 传统时间序列方法

经典预测方法如 ARIMA、SARIMA、Prophet 等基于线性或加性结构,在长期被应用于流量预测与需求规划场景[18]。其优势在于结构简单、解释性强,但对于服装行业中常见的突发波动、非线性趋势以及促销驱动的销量跳变,拟合能力有限。此外,传统方法通常无法整合多模态特征,只能依赖单一序列输入,严重限制其在实际业务中的表现。

具体而言,ARIMA 模型对平稳序列具有良好拟合效果,Prophet 通过加性分解在业务预测中得到广泛应用,Hyndman 等发展的 ETS 模型在短期预测中表现出较高可解释性[19]。部分学者尝试将传统方法与机器学习组合建模,但仍局限于结构化特征。

(二) 基于机器学习的预测方法

Ke 等利用 LightGBM 结合促销标记验证了特征工程的关键作用[20],Thomassey 等通过聚类与分类识别服装销售模式[21],韩亚娟等提出组合模型提升电商预测稳定性,Bandara 等研究了基于聚类的多序列预测方法[22]。

以随机森林、XGBoost、LightGBM 为代表的机器学习模型在结构化数据预测任务中取得了良好效果,能够捕捉复杂的特征交互关系。相关研究常将价格、折扣、类目、周次等特征加入预测模型中,以提升预测稳定性。然而,这类方法仍存在两个显著不足:难以处理高维且连续的时序特征;无法直接使用图像与文本等非结构化模态,这一不足在国内相关研究中亦被强调[23][24]。因此,它们对服装类目这种“强视觉属性”的商品依然存在天然限制。

TFT 模型通过多头注意力实现多变量时序建模[25],Informer 通过稀疏注意力降低长序列计算复杂度[26]。石闻达等设计了层次化多模态注意力循环网络用于服装新品预测[27],Wu 等提出 Dual-Forecaster 框架,Zhang 等利用对比学习增强多模态时序特征判别性。

(三) 深度学习在销量预测中的应用

CLIP 通过大规模图文对比学习实现视觉与语言的统一表示,UNITER 通过联合嵌入实现细粒度图文融合。Hao 等提出多模态 Transformer 新品销量预测方法,Khan 等的 MoAT 模型将多模态增强引入时序预测[28],易文龙等引入分层动态邻域的多模态电商分析[29],Yang 等利用搜索指数与图像特征进行时尚预测。上述研究为本文提供了理论参考,但在门控融合与对比对齐方面仍有优化空间。

近年来,随着 RNN、GRU、LSTM 以及 Transformer 结构的成熟,基于深度学习的时间序列预测方法得到广泛关注。其优势在于能够自动学习复杂的非线性动态模式,相比传统方法具有更强的表达能力。一部分研究将卷积神经网络(CNN)或 Transformer 用于时间序列特征提取,一部分研究将外生变量与销量序列共同建模,但整体仍未有效整合文本、图像等模态所包含的产品属性信息。

(四) 多模态学习在电商领域的研究进展

多模态学习在商品分类、图文匹配、属性识别、推荐系统等任务上发展迅速,代表性模型包括跨模态注意力、图文对齐网络、视觉语言预训练模型等。对于服装行业,商品图像能够反映款式风格、色彩、材质等视觉特征;文本描述与用户评论则包含语义信息与消费者偏好。如何将这些模态有效融合,用于销量预测,是近年来逐渐受到关注的新方向。部分研究尝试将图像、文本与销量序列同时输入模型,但

多采用简单拼接或加权融合方式，缺乏对模态重要性的动态建模，也缺少对跨模态语义空间一致性的约束。此外，图像与文本的模态差异较大，若缺乏对齐机制，融合往往不能显著提升预测效果。

3. 基于多模态信息融合的服装销量预测方法

(一) 整体模型架构

服装商品销量受多种因素共同影响，本文构建了包含数据预处理、多模态特征编码、特征融合与对齐以及销量预测解码四个模块的统一框架(见图 1)。模型同时接收历史销量序列、商品图像与文本等多模态输入，经编码、融合后通过时序解码网络完成多步销量预测。

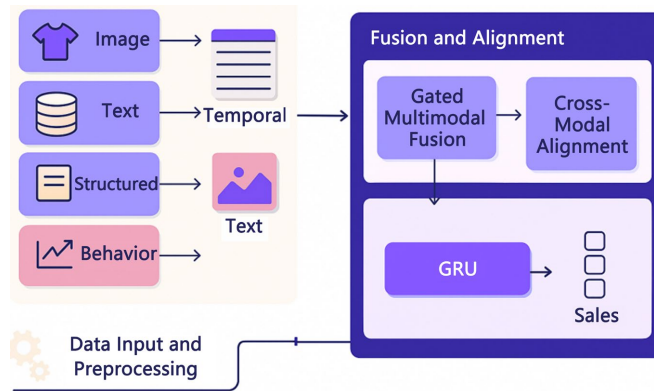


Figure 1. Overall model architecture
图 1. 模型架构图

(二) 多模态特征编码网络

本文分别采用 GRU 对历史销量序列建模，预训练语言模型结合双向 GRU 提取文本语义特征，预训练 CNN 提取视觉特征，并通过标准化与维度对齐处理统一各模态表示空间。

(三) 基于门控机制的多模态融合

在获得各模态编码后，本文设计了门控融合与跨模态注意力结构。

1) 门控权重计算原理

设不同模态的输入特征为 F_v, F_t, F_s ，分别表示视觉、文本和时序模态。门控单元计算每个模态的重要性权重：

$$g_i = \sigma(W_g F_i + b_g) \tag{1}$$

其中 σ 为 Sigmoid 函数。权重越大表示该模态在当前样本中的贡献越高。最终融合特征表示为：

$$F_{fusion} = \sum_i g_i \cdot F_i \tag{2}$$

2) 跨模态注意力机制

在门控加权的基础上，进一步引入跨模态注意力机制(Cross-modal Attention)，以捕捉模态间的关联依赖。例如文本与图像的互注意可帮助模型理解“红色长裙”等词汇与对应视觉特征的语义对应关系。注意力得分通过点积计算并经 softmax 归一化，实现跨模态特征的动态关联。

3) 特征融合策略

融合策略采用并行交互与残差连接结合的形式：

$$F_{out} = LayerNorm(F_{fusion} + CrossAttn(F_{fusion})) \tag{3}$$

这种设计既保留了单模态特征信息，又引入了模态间互补信息，从而增强融合特征的表达能力。

(四) 基于对比学习的多模态对齐

本文引入跨模态对比学习模块，使图像、文本与时序特征在同一语义空间中对齐。

1) 跨模态对比损失函数

采用改进的 Infonce 损失作为对齐目标。对于一组匹配的模态对(如图像 - 文本)，其相似度记为 $s_{i,j} = \cos(F_i, F_j)$ 。对比损失为：

$$L_{cl} = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp(s_{i,i}/\tau)}{\sum_{j=1}^N \exp(s_{i,j}/\tau)} \quad (4)$$

其中 τ 为温度参数，用于调节分布平滑程度。

通过对比学习将同一商品的多模态特征拉近、不同商品特征推远，形成语义一致的聚簇结构。同时设计可学习温度参数模块，通过反向传播自动优化相似度对齐。

(五) 多步销量预测解码器

融合特征输入至 GRU 解码器，采用自回归策略建模时序依赖，并支持滑动窗口式多步预测。

$$\hat{y}_t : t + K = W_o h_t + b_o \quad (5)$$

4. 实验与分析

(一) 实验设置

1) 数据集

数据来源于某大型服装电商平台，涵盖商品图像(统一缩放至 224×224 并归一化)、文本信息(标题、描述、评论，经分词与 BERT Tokenizer 处理)以及时序数据(日销量、价格、促销标记，经 Min-Max 标准化与缺失值填补)。

2) 对比方法

为验证模型有效性，本文选取 ARIMA、Prophet、XGBoost 等传统与机器学习模型，以及 TS-Only (GRU)、Text-Only (BERT)和 Image-Only (ResNet)等单模态深度学习模型作为基线，并构建 Time + Image 与 Time + Text 两种双模态模型进行对比。同时设计 MM-noGate、MM-noContra 和 MM-noAttn 三种消融模型，用于分析门控机制、对比学习与跨模态注意力对模型性能的影响，最终与本文提出的完整多模态融合模型(Full Model)进行综合评估。

3) 评价指标

本研究选取 MAE、RMSE、MAPE 和 R^2 四项指标评估模型表现，并设置 $K = 1, 2, 4, 8, 12$ 周的多步预测实验，分别对应短期补货、中期库存规划与长期采购决策场景。

(二) 结果分析

本节基于前述实验设置，对各类模型进行全面对比与分析，包括整体性能比较、多步预测表现以及关键模块的消融实验，以验证本文所提出多模态融合模型的有效性与优势。

1) 整体性能对比

各模型在测试集上的预测性能见表 1。

(1) 传统模型 vs 深度学习模型

传统模型(ARIMA、Prophet)受限于线性假设，整体表现最差。XGBoost 与 TS-Only 等模型 MAE 降低 10%~15%，但仍难以充分利用非结构化信息。

(2) 单模态模型性能有限

单模态模型(Text-Only、Image-Only)表现均未超过 TS-Only，证明单一模态难以实现全面表征。双模态模型引入后各项指标均显著改善，其中 Time + Text 优于 Time + Image，文本模态对销量波动具有更强解释力。

Table 1. Performance comparison of different sales forecasting models
表 1. 各模型销量预测性能对比

模型名称	MAE↓	RMSE↓	MAPE (%) ↓	R ² ↑
ARIMA	0.3542	0.4618	28.7	0.672
Prophet	0.3419	0.4462	27.5	0.688
XGBoost	0.3167	0.4125	25.4	0.721
TS-Only (GRU)	0.3012	0.3928	24.1	0.739
Text-Only (BERT)	0.3274	0.4173	25.8	0.716
Image-Only (ResNet)	0.3391	0.4297	27.1	0.701
Time + Image	0.2789	0.3531	22.4	0.762
Time + Text	0.2614	0.3291	21.2	0.781
MM-noAttn	0.2568	0.3226	20.7	0.789
MM-noContra	0.2509	0.3158	20.3	0.794
MM-noGate	0.2467	0.3097	20.0	0.799
Full Model	0.2347	0.2993	19.5	0.807

2) 消融实验分析

(1) 去除门控机制(MM-noGate)

去除门控机制后模型无法动态调节模态权重，性能略有下降但仍优于双模态模型。去除对比学习后各模态语义一致性降低，性能下降更明显。去除注意力机制后模态间深度交互受限，细粒度特征提取不足。三项消融实验验证了各模块的互补作用。

3) 多步预测性能分析

Table 2. Performance variation across prediction horizons K

表 2. 同预测步长 K 的性能变化

预测步长 K (周)	MAE (MM-Full)	RMSE (MM-Full)	MAE (TS-Only)	RMSE (TS-Only)
1	0.172	0.218	0.198	0.251
2	0.195	0.241	0.223	0.278
4	0.241	0.288	0.269	0.325
8	0.312	0.362	0.338	0.392
12	0.387	0.435	0.412	0.463

本研究设置了不同预测步长的多步预测实验(见表 2)。总体来看，随着 K 增大，两类模型误差均累积上升，但 MM-Full 始终优于 TS-Only。

(1) 短期预测(1~2周): 适合补货与快速响应场景

由表 2 可知, 短期预测(K = 1~2 周)中 MM-Full 表现出高精度与稳定性, 适用于实时补货决策。

(2) 中长期预测(4~8周): 误差增长更平缓、规划价值突出

中期预测(K = 4~8 周)中 MM-Full 误差增幅显著小于 TS-Only, 文本与视觉模态提供了稳定的结构性补充信息, 适用于库存规划与品类优化。

(3) 长期预测(12周): 仍保持优势, 但受季节与促销影响较大

长期预测(K = 12 周)中 MM-Full 的 MAE 仍低于 TS-Only 约 6%, 但受促销与季节性因素影响, 误差累积现象仍然存在, 未来可通过事件建模进行针对性优化。

4) 多步预测的细粒度分析

为揭示模型在不同商品子集上的表现差异, 本节从商品类别、生命周期阶段和价格区间三个维度对 K = 4 周的预测误差进行分组分析。

Table 3. Prediction performance by product category (K = 4 weeks)

表 3. 不同商品类别的预测性能(K = 4 周)

商品类别	MAE	RMSE	MAPE (%)	R ²
上装	0.218	0.267	17.8	0.823
下装	0.232	0.281	18.6	0.812
连衣裙	0.269	0.318	21.3	0.786
外套	0.257	0.303	20.1	0.795

由表 3 可知, 上装与下装预测精度最高, 连衣裙受季节与流行趋势影响最大(MAPE 达 21.3%), 外套介于两者之间。

Table 4. Prediction performance by product lifecycle stage (K = 4 weeks)

表 4. 不同生命周期阶段的预测性能(K = 4 周)

生命周期阶段	MAE	RMSE	MAPE (%)	R ²
新品期(<4 周)	0.286	0.339	23.1	0.761
成长期(4~12 周)	0.229	0.274	18.4	0.818
成熟期(>12 周)	0.212	0.258	17.1	0.831

由表 4 可知, 成熟期商品表现最优(R² = 0.831), 新品期预测难度最大(MAPE = 23.1%), 但多模态模型仍显著优于纯时序基线(TS-Only 新品期 MAE = 0.321), 表明图文信息可在历史数据匮乏时提供有效先验。

Table 5. Prediction performance by price range (K = 4 weeks)

表 5. 不同价格区间的预测性能(K = 4 周)

价格区间	MAE	RMSE	MAPE (%)	R ²
低价(<100 元)	0.253	0.305	20.4	0.798
中价(100~300 元)	0.224	0.270	18.0	0.821
高价(>300 元)	0.247	0.296	19.8	0.804

由表 5 可知, 中价位商品预测精度最高($R^2 = 0.821$), 低价位因促销波动大、高价位因序列稀疏而误差偏高。总体而言, 多模态模型在各子集上均保持较好预测能力, 但在高波动性与数据稀疏场景下仍有优化空间。

5. 结语

本文围绕服装电商场景中销量波动大、影响因素复杂、传统预测方法适应性不足等问题, 构建了一种基于多模态信息融合的服装销量预测模型。研究从商品图像、文本描述及历史销量等多源异构数据出发, 设计了包含多模态编码、门控融合、跨模态对齐与 GRU 解码的统一预测框架。实验结果表明, 该模型能够有效整合不同模态所蕴含的信息, 在整体预测精度、多步预测稳定性等方面均优于传统统计模型、机器学习模型以及单模态或双模态深度学习模型, 验证了所提出多模态融合机制在服装销量预测任务中的有效性。

从理论角度看, 本文为电商多模态数据在时序预测中的协同建模提供了可行思路; 从实践角度看, 所提方法能够为服装企业在补货决策、库存规划等业务中提供更可靠的预测支持。

尽管如此, 模型对促销等强扰动因素的刻画仍有改进空间, 预测与库存策略的耦合有待探索。未来可在引入业务特征、构建预测与决策一体化模型等方面展开研究。

基金项目

国家自然科学基金面上项目(24BGL114)。

参考文献

- [1] Fisher, M. and Raman, A. (1996) Reducing the Cost of Demand Uncertainty through Accurate Response to Early Sales. *Operations Research*, **44**, 87-99. <https://doi.org/10.1287/opre.44.1.87>
- [2] Cachon, G.P. and Swinney, R. (2011) The Value of Fast Fashion: Quick Response, Enhanced Design, and Strategic Consumer Behavior. *Management Science*, **57**, 778-795. <https://doi.org/10.1287/mnsc.1100.1303>
- [3] Skenderi, E., Joppi, R., Fiedler, J., et al. (2021) Well Googled Is Half Done: Fashion Product Image Search Trends and Sales Forecasting. arXiv Preprint.
- [4] Yang, T. (2021) Multimodal Fashion Sales Forecasting Using Google Trends and Image Features. *Journal of Retailing and Consumer Services*, **63**, Article 102685.
- [5] Li, X., Shen, J., Wang, D., Lu, W. and Chen, Y. (2024) Multi-Modal Transform-Based Fusion Model for New Product Sales Forecasting. *Engineering Applications of Artificial Intelligence*, **133**, 108606. <https://doi.org/10.1016/j.engappai.2024.108606>
- [6] 徐琪. 基于 ARIMA 模型与随机森林组合的零售服装动态销售预测[J]. 中国管理信息化, 2022, 25(6): 100-104.
- [7] Box, G. and Jenkins, G. (1976) *Time Series Analysis: Forecasting and Control*. Holden-Day.
- [8] Taylor, S.J. and Letham, B. (2018) Forecasting at Scale. *The American Statistician*.
- [9] Chen, T. and Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 785-794. <https://doi.org/10.1145/2939672.2939785>
- [10] 韩亚娟, 高欣. 基于机器学习组合模型的电商商品销量预测[J]. 计算机系统应用, 2022, 31(1): 315-321.
- [11] Radford, A., Kim, J.W., Hallacy, C., et al. (2021) Learning Transferable Visual Models from Natural Language Supervision.
- [12] Chen, Y., Li, L., Yu, L., El Kholy, A., Ahmed, F., Gan, Z., et al. (2020) UNITER: Universal Image-Text Representation Learning. In: *Lecture Notes in Computer Science*, Springer International Publishing, 104-120. https://doi.org/10.1007/978-3-030-58577-8_7
- [13] 潘志松, 韩笑, 黎维. 基于深度学习的时间序列预测方法综述[J]. 南京航空航天大学学报(自然科学版), 2025, 57(5): 799-821.
- [14] 陈嘉俊, 刘波, 林伟伟, 等. 基于 Transformer 的时间序列预测方法综述[J]. 计算机科学, 2025, 52(6): 96-105.

-
- [15] 马翌硕, 张光南, 刘亚婷, 等. 视觉-语言模型研究综述[J]. 计算机技术与发展, 2026, 36(3): 1-10.
- [16] 张浩宇, 王天保, 李孟择, 等. 视觉语言多模态预训练综述[J]. 中国图象图形学报, 2022, 27(9): 2652-2682.
- [17] 张虎成, 李雷孝, 刘东江. 多模态数据融合研究综述[J]. 计算机科学与探索, 2024, 18(10): 2501-2520.
- [18] 毛远宏, 孙琛琛, 徐鲁豫, 等. 基于深度学习的时间序列预测方法综述[J]. 微电子学与计算机, 2023, 40(4): 8-17.
- [19] Hyndman, R.J. and Athanasopoulos, G. (2021) *Forecasting: Principles and Practice*. 3rd Edition, OTexts.
- [20] Ke, G., Meng, Q., Finley, T., *et al.* (2017) LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, 4-9 December 2017, 3149-3157.
- [21] Thomassey, S. and Fiordaliso, A. (2006) A Hybrid Sales Forecasting System Based on Clustering and Decision Trees. *Decision Support Systems*, **42**, 408-421. <https://doi.org/10.1016/j.dss.2005.01.008>
- [22] Bandara, K., Bergmeir, C. and Smyl, S. (2020) Forecasting across Time Series Databases Using Recurrent Neural Networks on Groups of Similar Series: A Clustering Approach. *Expert Systems with Applications*, **140**, Article 112896. <https://doi.org/10.1016/j.eswa.2019.112896>
- [23] Wu, W., Li, Y., Wang, Z., *et al.* (2025) Dual-Forecaster: Learning to Forecast Multimodal Time Series. arXiv.
- [24] Zhang, Y., Li, T., Yin, Z., *et al.* (2024) Contraformer: Contrastive Multimodal Transformer. arXiv.
- [25] Lim, B., Arık, S.Ö., Loeff, N. and Pfister, T. (2021) Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting. *International Journal of Forecasting*, **37**, 1748-1764. <https://doi.org/10.1016/j.ijforecast.2021.03.012>
- [26] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., *et al.* (2021) Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 11106-11115. <https://doi.org/10.1609/aaai.v35i12.17325>
- [27] 石闻达, 杜劲松, 李笛出乘. 基于层次化多模态注意力机制循环神经网络的服装新品销售预测(英文) [J]. 东华大学学报(英文版), 2024, 41(1): 21-27.
- [28] Lee, G., Yu, W., Cheng, W., *et al.* (2024) MoAT: Multi-Modal Augmented Time Series Forecasting. *Proceedings of International Conference on Learning Representations*, OpenReview.
- [29] 易文龙, 黄暄, 刘木华, 等. 基于分层动态邻域的多模态电商特色水果评价情感分析方法[J]. 农业工程学报, 2025, 41(19): 206-217.