

融合AEMD与趋势交叉注意力的多尺度Transformer油价预测框架

李滕佳, 马春爱

引用本文

李滕佳, 马春爱. [融合AEMD与趋势交叉注意力的多尺度Transformer油价预测框架](#)[J]. 计算机科学, 2026, 53(5): 157-163.

LI Tengjia, MA Chun'ai. [Multi-scale Transformer Oil Price Prediction Framework with AEMD and Trend Cross-attention](#) [J]. Computer Science, 2026, 53(5): 157-163.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于多尺度Transformer融合多域信息的伪造人脸检测](#)

Forgery Face Detection Based on Multi-scale Transformer Fusing Multi-domain Information

计算机科学, 2023, 50(10): 112-118. <https://doi.org/10.11896/jsjcx.220900048>

融合 AEMD 与趋势交叉注意力的多尺度 Transformer 油价预测框架

李滕佳 马春爱

中国石油大学(北京)经济管理学院 北京 102249

摘要 在全球低碳转型和能源结构调整的背景下,原油价格预测不仅是能源市场分析的重要课题,也是政策制定与投资决策的核心参考。然而,原油价格序列往往表现出高度非线性和显著的非平稳性。现有方法在特征提取与时序建模方面仍存在不足:一方面,多时间尺度特征的挖掘不充分,导致对短期剧烈波动与长期趋势演化的刻画存在偏差;另一方面,短期与长期信息在融合过程中常出现耦合不当,难以兼顾预测精度与趋势稳定性。针对上述问题,提出了一种自适应频率解耦的双分支 Transformer 模型(Multi-frequency Decoupled Dual-Branch Variant Transformer, MFD-DBV-Transformer),用于布伦特原油价格预测。首先,引入自适应经验模态分解(Adaptive Empirical Mode Decomposition, AEMD)方法,将原油价格序列自适应分解为多个本征模态函数(Intrinsic Mode Functions, IMF),并区分高频短期成分与低频长期趋势成分。在此基础上,设计自适应频率解耦模块(Adaptive Frequency Decoupling Module, AFDM),构建双分支特征表示网络,以分别捕获短期波动模式和长期趋势特征。为了突破现有方法中短期与长期特征融合不合理的局限性,进一步提出趋势融合模块,利用交叉注意力机制实现长期趋势信息对短期预测的自适应调节。为防止模型在长期趋势建模中过度拟合,引入带温度调节的自适应掩码机制,以提升模型在复杂市场环境下的泛化能力。实验结果表明, MFD-DBV-Transformer 在时频特征捕获与预测性能方面均表现优异,显著优于传统 LSTM 及部分主流深度学习模型,不仅在预测精度上取得提升,也在趋势跟踪能力上展现出更强的稳定性与适应性。该成果为政策制定者和能源投资者在应对原油市场剧烈波动时提供了一种高效、可靠的预测与决策支持工具,同时也为复杂非平稳时间序列的建模研究提供了新的思路与方法。

关键词: 多尺度 Transformer; 自适应频率解耦; 原油价格预测; 趋势融合; 低碳经济

中图分类号 F831.5

Multi-scale Transformer Oil Price Prediction Framework with AEMD and Trend Cross-attention

LI Tengjia and MA Chun'ai

School of Economics and Management, China University of Petroleum(Beijing), Beijing 102249, China

Abstract Against the backdrop of global low-carbon transition and energy restructuring, crude oil price forecasting has become not only a key topic in energy market analysis but also an essential reference for policy-making and investment decisions. However, crude oil price series often exhibit strong nonlinearity and pronounced non-stationarity. Existing methods still face limitations in feature extraction and temporal modeling: on the one hand, the mining of multi-scale features is often insufficient, leading to biased characterization of short-term fluctuations and long-term trend evolution; on the other hand, the integration of short- and long-term information is frequently mishandled, making it difficult to balance predictive accuracy with trend stability. To address these challenges, this paper proposes a multi-frequency decoupled dual-branch Transformer model(MFD-DBV-Transformer) for Brent crude oil price forecasting. The method firstly employs adaptive empirical mode decomposition(AEMD) to decompose the crude oil price series into multiple intrinsic mode functions(IMFs). This distinguishes high-frequency short-term components from low-frequency long-term trends. An adaptive frequency decoupling module(AFDM) is then designed to construct dual-branch feature representations, separately capturing short-term volatility patterns and long-term trend features. A trend fusion module is further introduced, where cross-attention is used to achieve adaptive modulation of short-term predictions with long-term trend information. In addition, a temperature-regulated adaptive masking mechanism is incorporated to prevent overfitting in long-term trend modeling and to enhance the model's generalization ability in volatile market environments. Experimental results demonstrate that the proposed MFD-DBV-Transformer achieves superior performance in capturing complex time-frequency char-

到稿日期:2025-09-12 返修日期:2025-11-18

基金项目:国家自然科学基金(71202118);北京市社会科学基金一般项目(22JB016)

This work was supported by the National Natural Science Foundation of China(71202118) and General Project of Beijing Social Science Foundation(22JB016).

通信作者:李滕佳(litj62069831@163.com)

acteristics of crude oil prices, significantly outperforming traditional LSTM and several mainstream deep learning models. The model not only improves forecasting accuracy but also demonstrates stronger stability and adaptability in trend tracking. The proposed approach provides policymakers and energy investors with an efficient and reliable forecasting and decision-support tool for coping with crude oil market volatility, while offering new insights and methodologies for modeling complex non-stationary time series.

Keywords Multi-scale transformer, Adaptive frequency decoupling, Crude oil price forecasting, Trend fusion, Low-carbon economy

1 引言

原油作为全球最重要的能源商品和战略物资,其价格波动对世界经济具有深远影响。准确预测原油价格不仅关系到能源企业的投资决策和风险管理,更直接影响着各国的经济政策的制定。然而,原油价格受到供需关系、地缘政治、宏观经济等多重因素影响,呈现出高度的非线性和多时间尺度特征,使得准确预测极具挑战。因此,原油价格预测^[1-3]一直是经济学、金融学和能源市场研究的热点方向。

当前,众多学者已经从不同角度对原油价格预测问题进行了深入探索。传统的预测方法如 ARIMA 和 VAR^[4-5]等均基于线性假设,通过数学模型描述原油价格序列的统计规律。这类方法在处理具有明显增长趋势和季节性的时间序列时表现良好,但难以捕获价格序列中的复杂非线性模式和突发性变化,特别是在面对市场剧烈波动时预测精度显著下降。

近年来,深度学习方法在时间序列预测领域展现出强大潜力。LSTM 等循环神经网络通过门控机制能够有效捕获序列的长期依赖^[6-11],在原油价格预测中得到了广泛应用,验证了深度学习在处理时间序列长期依赖关系方面的有效性。然而,RNN 系列模型存在梯度消失、训练效率低等固有缺陷,其在长序列建模中的表现受到限制。随着 Transformer 架构在时间序列预测领域的兴起^[12-14],研究者们开始探索其在原油价格预测中的应用潜力。Transformer 通过自注意力机制并行处理序列信息,在建模长距离依赖关系方面表现尤为出色。Zhou 等提出的 Autoformer 使用自注意力与自回归解码器,大幅降低了长序列建模的时间与内存复杂度,同时保持对远距离依赖的刻画能力,适合更长周期的预测^[12]。Nie 等将时间序列切分为 patch 做通道独立建模,用 Transformer 在 patch 级别学习局部到全局的时序模式,模型兼具长程依赖捕获与训练效率^[15]。

为了结合不同方法的优势,研究者们探索了将信号分解技术与深度学习相结合的混合方法^[3,16-22]。这类方法通常首先采用 EMD 和 CEEMDAN 等分解技术^[5,14]将复杂的原油价格序列分解为多个相对简单的子序列,然后分别建模预测,最后重构得到最终结果。Bi 等^[23]提出了一种基于变分模态分解(VMD)与 BiTCN-BiGRU-Attention 深度学习技术的混合预测模型,通过 VMD 将原油价格序列分解为不同频率的模态分量,再利用双向时间卷积网络和门控循环单元的组合架构捕获时间序列的长短期依赖特征。虽然混合方法能够有效提升预测性能,但现有方法大多采用固定的分解策略和简单的信息融合机制,难以充分挖掘时间序列不同时间尺度特征

之间的复杂关系。

现有方法在原油价格预测方面仍面临两个关键挑战:首先,原油价格序列包含多个时间尺度的信息,传统方法往往采用单一架构处理所有频率成分,导致高频短期波动与低频长期趋势特征相互干扰,难以有效分离和建模不同时间尺度的价格变化模式;其次,现有多尺度融合方法多采用简单的特征拼接或加权平均,缺乏对短期波动与长期趋势之间复杂耦合关系的深入建模,在融合过程中容易出现信息冗余或关键特征丢失,影响预测的准确性和稳定性。EMD 能够有效将原油价格序列分离为不同频率成分:短期高频成分(IMF1—IMF3)主要反映市场的价格调整和随机扰动,呈现出明显的高频波动特征;而长期低频成分(IMF4—IMF7 和残余项)则体现原油价格的基本面驱动因素和长期演化规律,展现出相对平缓的变化趋势。受此启发,本文提出了一种自适应频率解耦的双分支 Transformer 模型。该方法使用改进的 AEMD^[14]技术将价格序列分解为不同频率成分,实现对非平稳时间序列中长短趋势特征的高效分解,然后利用双分支架构分别建模短期和长期信息,并设计趋势融合模块实现多尺度信息的自适应整合,从而抽取长期趋势以及短期趋势进行准确预测。

本研究的主要贡献包括:(1)提出 AEMD 与双分支 Transformer 相结合的预测框架,有效缓解了处理多时间尺度信息时的特征混叠问题;(2)设计基于交叉注意力的趋势融合模块和温度调节机制,实现短期波动与长期趋势的自动耦合,减少了信息冗余,同时提高了数据利用率;(3)通过自适应掩码机制防止模型过度依赖单一时间尺度信息,改善了在市场剧烈波动时的预测性能。

2 自适应频率解耦的双分支 Transformer 模型

2.1 总体框架

原油价格预测任务旨在基于历史价格序列预测未来价格,核心挑战在于有效处理价格序列中的多时间尺度特征和复杂的非线性动态。为此,本研究提出了自适应频率解耦的双分支 Transformer。如图 1 所示,模型采用四阶段处理流程:首先,利用自适应经验分解模块将原始价格序列自适应分解为短期高频成分和长期低频成分;随后,通过双分支 Transformer 编码器分别提取两类成分的特征表示;接着,设计趋势融合模块,采用交叉注意力机制实现短期和长期信息的自适应整合;最终,通过前馈神经网络输出预测结果。整体架构采用端到端的训练方式,通过频域分解、双分支编码和趋势融合的有机结合捕获原油价格的多尺度时频特征,并实现精确预测。

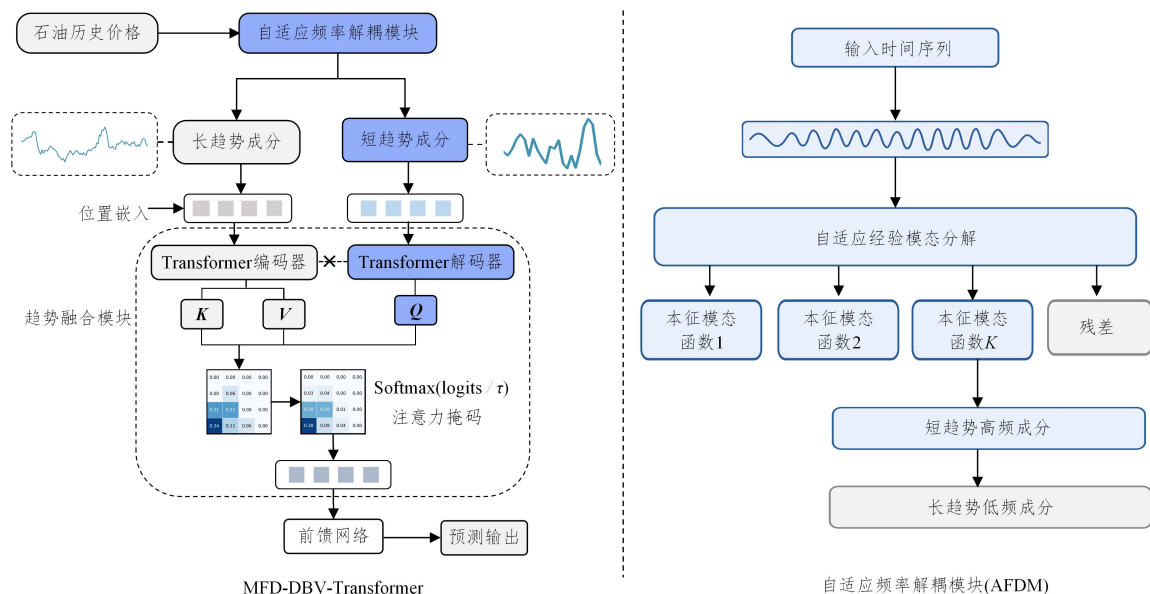


图1 自适应频率解耦的双分支 Transformer 模型

Fig. 1 Multi-frequency decomposition dual-branch transformer

2.2 自适应频域分解模块

由于原油价格时间序列具有复杂性和多尺度特征,直接学习经验模态分解(Empirical Mode Decomposition, EMD)技术分解后的全部分量,会导致计算量大幅增加且更易受随机噪声的影响。因此,自适应频域分解模块将会在 EMD 的基础上根据信号本身的特征对所有的分量进行二次的再分组合并,旨在针对原油价格预测的实际意义,将原油价格序列自动分离为长期趋势和短期波动两个部分,以降低计算复杂度并提高预测效率,使其更适合处理非平稳的时间序列。

具体而言,设原油价格时间序列为 $X = \{x_1, x_2, \dots, x_t\}$, 通过 EMD 分解得到多个本征模态函数(IMFs)和残差项:

$$X(t) = \sum_{i=1}^n IMF_i(t) + r(t) \quad (1)$$

其中, $X(t)$ 表示 t 时刻的原油价格, $IMF_i(t)$ 表示第 i 个本征模态函数, $r(t)$ 为残差项^[1]。

如图 2 所示, EMD 将 2002—2018 年的布伦特原油价格时间序列分解为 7 个 IMF 分量和 1 个残差分量。从分解的可视化结果可以观察到: IMF1—IMF3 主要捕获了价格的高频波动和短期噪声信息; 而 IMF4—IMF7 及残差项的频率则显著降低, 说明主要包含长期趋势和低频成分。

基于频率特性分析和 IMF 分量的时频特征, 本研究将分解得到的 IMFs 重新组合为短期高频成分和长期低频成分:

$$X_{\text{short}} = \sum_{i=1}^k IMF_i(t) \quad (2)$$

$$X_{\text{long}} = \sum_{i=k+1}^n IMF_i(t) + r(t) \quad (3)$$

其中, k 为频率划分阈值, 根据 IMFs 的频率中位数的计算结果自动确定高低频分组。

自适应频率解耦策略(AEMD)能更有效地分离原油价格序列中的不同时间尺度信息, 为后续的双分支 Transformer 结构提供了结构化的输入特征, 有助于模型更好地理解 and 建模原油价格的复杂动态特性。

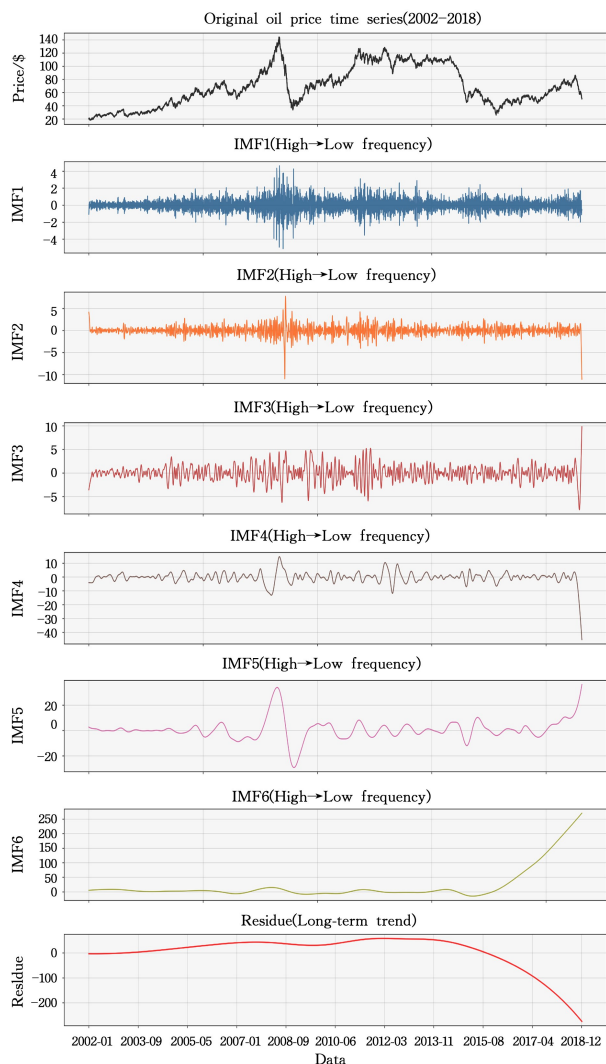


图2 采用 EMD 技术对原始价格序列进行分解后的信号可视化

Fig. 2 Signal visualization after decomposing the original price series using EMD

2.3 双分支 Transformer 编码器

为了有效处理不同时间尺度的信息,设计了专门的双分支 Transformer 结构来提取原油价格时间序列的多尺度特性:短期价格波动往往受到市场情绪、突发事件等因素的影响,表现出高频且不规律的特征;而长期价格趋势则主要由供需基本面、经济周期等宏观因素驱动,呈现相对稳定的演化规律。因此,需要采用不同的编码策略来分别处理这两类信息。

短期分支设计:针对短期高频成分 X_{short} ,采用标准的 Transformer 编码器结构,通过细粒度的位置嵌入和多头自注意力机制捕获短期价格波动的局部特征。

$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right) \quad (4)$$

$$PE(pos, 2i+1) = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right) \quad (5)$$

其中, pos 表示时间步位置, i 表示维度索引, d_{model} 为 Transformer 模型的隐藏层维度。

$$H_{\text{short}} = \text{Transformer}(PE(X_{\text{short}})) \quad (6)$$

$$H_{\text{long}} = \text{Transformer}(PE(X_{\text{long}})) \quad (7)$$

长期分支设计:对于长期低频成分 X_{long} ,同样采用 Transformer 编码器结构,但需通过调整位置编码的频率参数来适应更长的时间跨度。长期分支专注于建模供需平衡、地缘政治、宏观经济政策等基本面因素对价格的长期影响,生成稳定的趋势特征表示。

通过这种双分支设计,模型能够同时保持对短期市场动态的敏感性和对长期趋势的把握能力,为后续的趋势融合提供互补特征 H_{short} 和 H_{long} 。

2.4 趋势融合模块

2.4.1 交叉注意力机制

原油价格预测任务具有独特的多时间尺度耦合特性:短期波动反映市场情绪和投机行为,而长期趋势体现供需基本面和宏观经济因素。这种特性要求预测模型既要捕获短期价格的敏感变化,又要保持与长期趋势的一致性。基于上述原油价格预测任务的特点,在设计短期价格变化时往往需要以长期趋势为参考背景,但预测的主体仍应是短期动态。因此,本研究将短期分支特征作为 Query(Q),长期分支特征作为 Key(K)和 Value(V),这种设计使得短期特征能够主动“查询”相关的长期趋势信息,抽取长期趋势进行辅助预测。

$$\text{CrossAttn}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (8)$$

其中, $Q = H_{\text{short}}W_Q$, $K = H_{\text{long}}W_K$, $V = H_{\text{long}}W_V$ 。 W_Q, W_K, W_V 是交叉注意力机制中可学习的变换矩阵,通过反向传播优化。得益于该设计,短期特征作为 Query 确保了预测过程以短期动态为主导,长期趋势仅作为辅助参考。此外,通过注意力权重的动态分配,模型能够自适应地选择与当前短期状态最相关的长期趋势特征。

2.4.2 基于温度的自适应注意力裁剪

为了防止模型过度依赖长期趋势而导致对短期波动的敏

感性下降,设计了基于温度调节的自适应掩码机制。该机制的核心思想是通过温度参数动态调整注意力分布的锐度,实现对长期信息使用程度的精细控制。

$$\text{Mask}_{ij} = \text{sigmoid}\left(\frac{\alpha \cdot \text{sim}(q_i, k_j)}{\tau}\right) \quad (9)$$

$$\text{Attention}_{ij} = \text{Mask}_{ij} \cdot \text{softmax}\left(\frac{q_i k_j^T}{\sqrt{d_k}}\right) \quad (10)$$

相似度函数 $\text{sim}(q_i, k_j)$ 采用余弦相似度计算。其中, τ 为温度系数; q_i 表示第 i 个位置的查询向量,来自短期分支的隐藏状态通过线性变换得到;而 k_j 则表示第 j 个位置的键向量,来自长期分支的隐藏状态通过线性变换得到。

$$\text{sim}(q_i, k_j) = \frac{q_i \cdot k_j}{\|q_i\| \cdot \|k_j\|} \quad (11)$$

根据式(9)和式(10),当温度系数 τ 较小时,注意力分布更加集中,模型倾向于只关注与当前短期状态高度相关的长期信息,而长期趋势信息则会被弱化;当 τ 较大时,根据式(6)计算得到的注意力分布则会变得更加平滑,模型将会考虑更多的长期趋势信息。对 τ 的调控,有助于模型在不同市场状态下平衡短期敏感性和长期稳定性,从而引导模型同时从短期以及长期趋势中去探索有用信息来进行下一时间步的价格预测。

2.5 损失函数设计

考虑到原油价格预测任务的特点,设计了包含预测损失及平滑性约束项的复合损失函数来优化预测精度并提高模型稳定性。主损失函数采用均方误差(MSE),结合 L_2 正则化防止过拟合。

$$\hat{y} = FC(\text{FFN}(\text{CrossAttn}(H_{\text{short}}, H_{\text{long}}))) \quad (12)$$

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda \sum_j \theta_j^2 \quad (13)$$

其中, CrossAttn 表示 2.4.1 节中的交叉注意力机制; FFN 表示前馈神经网络层进行特征变换; FC 为全连接层,将高维特征映射为单一的原油价格预测值 \hat{y} ; θ_j^2 代表模型的第 j 个参数; λ 为正则化项的权重系数,用于控制惩罚模型权重的强度。

此外,为了确保预测序列的平滑性并避免异常波动,引入了梯度惩罚项。

$$\mathcal{L}_{\text{smooth}} = \frac{1}{N-1} \sum_{i=1}^{N-1} (\hat{y}_{i+1} - \hat{y}_i)^2 \quad (14)$$

最终的总损失函数可以表示为:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MSE}} + \gamma \mathcal{L}_{\text{smooth}} \quad (15)$$

其中, γ 为平衡不同损失项的权重系数,通过验证集网格搜索确定最优组合。

3 实验设计与结果分析

3.1 实验设置

3.1.1 数据集

本研究选择欧洲布伦特原油现货价格作为预测目标,数据来源于美国能源信息署(US Energy Information Administration, EIA)。布伦特原油价格被广泛认为是国际原油价格的基准,其价格数据不仅反映了全球供需动态,还包含了市场对未来价格的预期以及各种经济和政治因素的影响。该数据

集涵盖了 1987 年 1 月 1 日至 2025 年 8 月 4 日的布伦特原油日度现货价格,总计约 38 年的历史数据。

在数据集划分方面,按照 8:1:1 的比例将数据集划分为训练集、验证集和测试集。为了保持时间序列的连续性,采用时间顺序划分方式,即前 80% 的时间段作为训练集,中间 10% 作为验证集,最后 10% 作为测试集。

3.1.2 实验环境设置

本研究采用 Python 3.8 作为开发环境,使用 PyTorch 1.12.0 框架实现模型构建,在配备单张 NVIDIA RTX 3090 GPU 的服务器上进行实验。模型采用 6 层 Transformer 编码器,模型维度为 256,多头注意力头数为 8,AEMD 分解后将 IMF1-IMF3 归类为短期高频成分,IMF4 及残差归类为长期低频成分。在趋势融合模块中,温度参数 τ 设置为 0.5。训练时使用 AdamW 优化器,学习率设置为 1×10^{-4} ,批次大小为 32,训练 100 轮,损失函数权重设置为 MSE 损失权重 1.0、平滑性损失权重 $\gamma=0.1$ 。

3.2 基准模型对比实验

为了全面评估 MFD-DBV-Transformer 的有效性,选择了多种在时间序列预测领域具有代表性的基准方法进行对比实验:ARIMA^[4] 代表经典统计方法的线性模型;XGBoost^[24] 和 Random Forest^[25] 代表集成学习方法;LSTM^[6] 代表递归神经网络的序列模型;Transformer^[13] 代表基于注意力机制的深度学习方法;EMD-LSTM^[1],CEEMDAN-SVR^[23],VMD-BiTCN-BiGRU^[22] 以及 TN-LP-LSTM-SVM^[26] 则代表近年来提出的频域分解与机器学习结合的先进混合方法。为了公平比较,对于 Transformer 基准模型,同样采用 6 层 Transformer 编码器,多头注意力数量设置为 8,而其余模型则使用默认配置进行实验。

实验结果如表 1 所列。可以发现,MFD-DBV-Transformer 方法在所有评估指标上均取得最优性能;传统统计方法 ARIMA 受限于线性假设,MAE 达到 1.234;机器学习方法,如 XGBoost 和 Random Forest,通过集成学习提升了预测性能,MAE 分别为 0.892 和 0.934,但缺乏对时序依赖关系的专门建模。

表 1 与多种先进方法的比较结果

Table 1 Comparison results with various advanced methods

方法名称	MAE	RMSE	R^2	MAPE/%
ARIMA(2,1,2) ^[4]	1.234	1.672	0.789	4.87
XGBoost ^[24]	0.892	1.245	0.856	3.74
RandomForest ^[25]	0.934	1.298	0.841	3.96
LSTM ^[6]	0.823	1.156	0.878	3.42
Transformer ^[13]	0.792	1.123	0.883	3.29
EMD-LSTM ^[1]	0.768	1.089	0.889	3.15
CEEMDAN-SVR ^[23]	0.781	1.102	0.885	3.21
VMD-BiTCN-BiGRU ^[22]	0.763	1.068	0.889	3.02
TN-LP-LSTM-SVM ^[26]	0.759	1.032	0.894	2.93
MFD-DBV-Transformer	0.756	0.934	0.912	2.67

深度学习方法中,标准 Transformer 凭借自注意力机制表现较好,MAE 为 0.792, R^2 达到 0.883,但其统一的架构难以有效处理多时间尺度特征的差异化需求。混合分解方法表现出明显优势:EMD-LSTM 通过频域分解将 MAE 降至

0.768,但传统 EMD 的模式混叠问题限制了其性能的进一步提升;CEEMDAN-SVR 虽然改善了分解质量,但 MAE 为 0.781,略逊于 EMD-LSTM,主要是由 SVR 在长序列建模方面的局限性所致。值得注意的是,复杂混合方法 VMD-BiTCN-BiGRU-Attention 取得了较好的效果,MAE 为 0.743,但其 RMSE 为 1.068,说明在处理极值时仍不稳定。本文提出的方法 MAE 为 0.756,虽然略高于 VMD-BiTCN-BiGRU-Attention,但 RMSE 显著降低至 0.934,相比 VMD-BiTCN-BiGRU-Attention 提升了 12.5%, R^2 达到 0.912,MAPE 降至 2.67%。相比传统 Transformer,本文方法 MAE 提升 4.5%,RMSE 提升 16.8%,这充分体现了自适应频率解耦和双分支融合架构的优越性。

3.3 消融实验

3.3.1 各组件对性能的影响

为了验证所提出方法中各个关键组件的有效性,采用排除法设计消融实验。从完整的 MFD-DBV-Transformer 模型出发,逐一移除关键组件以评估其对整体性能的贡献,系统分析每个组件在完整系统中的真实作用。实验结果如表 2 所列。

表 2 自适应频率解耦的双分支 Transformer 中各组件对性能的影响

Table 2 Impact of each component on performance in the

MFD-DBV-Transformer

模型配置	MAE	RMSE	R^2	MAPE/%
完整模型	0.756	0.934	0.912	2.67
- AEMD 分解	0.834	1.089	0.896	3.12
- 长期趋势分支	0.812	1.045	0.899	3.05
- 趋势融合模块	0.897	1.134	0.885	3.58
- 自适应注意力掩码	0.782	0.967	0.908	2.89
Baseline(Transformer)	1.247	1.562	0.834	4.28

从表 2 可以看出,移除各个组件均导致了显著的性能下降。移除 AEMD 分解后,MAE 从 0.756 上升至 0.834,性能下降 10.3%,验证了频域分解对于处理复杂时间序列的有效性,通过将原始价格序列分解为不同频率成分,模型能够更好地捕获多尺度的价格变化模式。移除长期趋势分支后,MAE 上升至 0.812,性能下降 7.4%,说明长期趋势信息对于提高预测准确性具有重要作用,特别是在捕获价格的宏观变化方向方面。移除趋势融合模块的影响最为显著,MAE 上升至 0.897,性能下降 18.7%,这表明通过交叉注意力机制实现的短期和长期信息融合是提升预测性能的关键。移除自适应注意力掩码后,MAE 上升至 0.782,性能下降 3.4%,验证了基于温度调节的注意力裁剪机制在防止过度依赖长期趋势及保持短期预测敏感性方面的有效性。完整模型相比基础 Transformer 在 MAE 上提升了 39.4%,充分验证了所提出方法的有效性。

3.3.2 多种注意力机制的性能比较

为了进一步分析趋势融合模块中多层次注意力机制的作用,设计了详细的消融实验来分析不同注意力配置对模型性能的影响。实验采用逐步添加的方式,从基础配置开始,逐步引入各种注意力机制的改进组件,系统地评估每个设计选择的贡献。实验结果如表 3 所列。

表3 趋势融合模块中不同的注意力设置对性能的影响

Table 3 Impact of different attention settings in the trend fusion module on performance

注意力配置	MAE	RMSE	R^2	MAPE/%
无注意力机制	0.923	1.156	0.887	3.89
+共享权重自注意力	0.887	1.123	0.892	3.74
+分离权重自注意力	0.856	1.089	0.896	3.61
+交叉注意力机制	0.821	1.045	0.902	3.46
+ QKV 互换	0.798	1.012	0.906	3.37
+温度调节机制	0.773	0.976	0.909	3.26
+自适应注意力掩码(完整模型)	0.756	0.934	0.912	2.67

从表3可以看出,基线模型采用简单的特征拼接方式融合短期和长期信息,MAE为0.923。在引入共享权重自注意力后,模型开始具备捕获内部特征依赖关系的能力,MAE降至0.887,MAPE降至3.74%。由于权重共享限制了模型对不同分支特征的差异化处理,而权重分离的自注意力模块使模型能够为短期和长期分支分别学习专门的注意力模式,因此引入分离权重自注意力后,MAE进一步降至0.856,MAPE降至3.61%。交叉注意力机制的引入带来了显著的性能提升,MAE降至0.821,这验证了短期和长期信息之间交互建模的重要性,通过让短期分支关注长期趋势,长期分支感知短期波动,实现了更有效的信息融合。而QKV的互换策略通过让长期趋势作为Query、短期信息作为Key和Value,进一步优化了注意力计算,MAE降至0.798,MAPE降至3.37%,这种配置更符合以长期趋势为指导、从短期信息中选择相关特征的预测逻辑。此外,温度调节机制的引入使注意力分布更加集中和可控,MAE降至0.773,MAPE降至3.26%,通过调节注意力的锐度避免了过度平滑的问题。最终,自适应注意力掩码的引入实现了最优性能,MAE达到0.756, R^2 为0.912,该机制能够动态调整注意力范围,在保持长期趋势指导的同时确保对短期变化的敏感性。

3.4 可视化分析

为了直观展示MFD-DBV-Transformer方法在解决原油价格预测关键点方面的有效性,本研究进行了额外的可视化分析。图3展示了2018年1月至2025年7月期间的原油价格预测结果对比,其中黑色实线为真实价格,红色虚线为趋势平滑预测方法,青色实线为MFD-DBV-Transformer方法。结果表明,MFD-DBV-Transformer方法的预测曲线与真实价格走势高度吻合,在多时间尺度特征有效建模方面表现突出,尤其在2020年初原油价格急剧下跌和2021—2022年的显著上涨等关键时期表现出色。MFD-DBV-Transformer方法能够同时捕获短期剧烈波动和长期趋势演化,避免了传统单一架构方法中高频与低频特征相互干扰的问题。

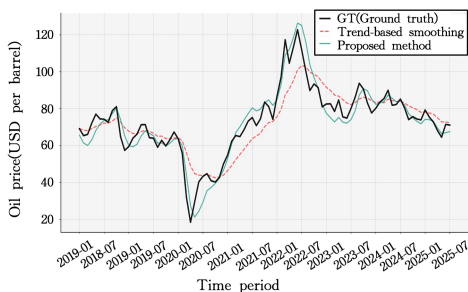
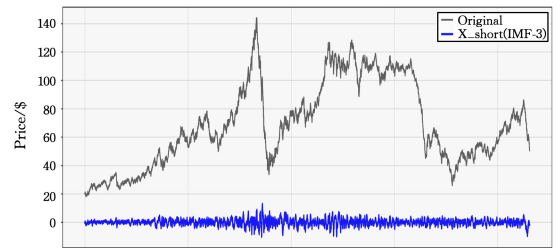


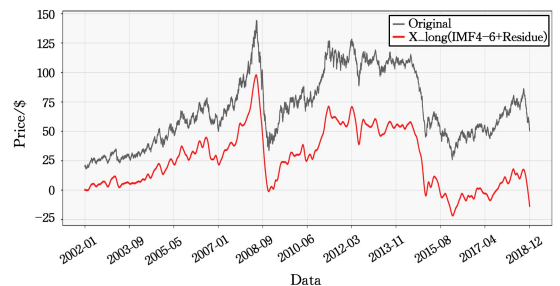
图3 所提出方法在2018—2025年的预测结果(电子版为彩图)

Fig. 3 Forecast results from 2018 to 2025 of the proposed method

图4展示了AEMD分解的效果,包括短期高频成分(IMF1—IMF3的重构)和长期低频成分(IMF4及残差的重构)。该分解结果直接验证了本文方法在解决多时间尺度特征混叠难点方面的有效性:短期成分捕获了价格的日常波动和短期噪声;而长期成分则准确提取了价格的基本面趋势,有效分离出2008年金融危机、2014—2016年油价下跌等重要宏观事件的长期影响,避免了这些不同时间尺度信息的相互干扰。更重要的是,两个重构信号的叠加还原了原始价格序列,既保留了对短期市场变化的敏感性,又确保了预测结果与长期基本面的一致性,从而进一步避免了简单拼接或加权融合导致的信息冗余和关键特征丢失,使得模型能够在长期与短期趋势这两个维度提供更精确的预测。



(a) Short-term Component: X_short(2002—2018)



(b) Long-term Component: X_long(2002—2018)

图4 长期与短期趋势的重建结果比较

Fig. 4 Comparison of the reconstruction results of long-term and short-term trends

虽然AEMD分解能够识别出2008年金融危机等重大事件的影响,但这种识别是基于历史价格模式的后验分析,存在一定的滞后性,缺乏对此类事件的前瞻性预警能力。因此,在面对极端的市场冲击时,纯粹基于价格历史的模型可能表现不佳。

结束语 本研究提出了一种基于AEMD分解的双分支Transformer原油价格预测模型。AEMD分解技术有效分离了原油价格序列中的不同频率成分,双分支Transformer架构通过短期分支处理高频波动、长期分支捕获宏观趋势,实现了多尺度信息的有效整合。消融实验验证了AEMD分解、趋势融合模块等关键组件的重要性。与多种基准方法的对比表明,所提出的方法在各项评估指标上均达到最优性能(MAE为0.756, R^2 达到0.912)。本研究通过AEMD与双分支Transformer的有机结合,以及多层级注意力融合机制的创新设计,为原油价格预测提供了新的技术框架,对推动能源金融领域的时间序列预测方法发展具有重要意义。

参考文献

[1] COLOMINAS M A, SCHLOTTHAUER G, TORRES M E. Im-

- proved complete ensemble EMD: A suitable tool for biomedical signal processing [J]. *Biomedical Signal Processing and Control*, 2014, 14:19-29.
- [2] NIU D, CUI X. Crude Oil Price Prediction Based on Hybrid Deep Learning [J]. *Journal of North China Electric Power University (Social Sciences Edition)*, 2023, 4(6):30-42.
- [3] ZHANG D, HUANG J, LING L, et al. Carbon Price Multi-decomposition Ensemble Prediction Integrating News Impact Decay [J]. *Journal of Henan University of Science & Technology (Natural Science)*, 2024, 45(1):51-61.
- [4] MORANA C. A semiparametric approach to short-term oil price forecasting [J]. *Energy Economics*, 2001, 23(3):325-338.
- [5] BAUMEISTER C, PEERSMAN G. The role of time-varying price elasticities in accounting for volatility changes in the crude oil market [J]. *Journal of Applied Econometrics*, 2013, 28(7):1087-1109.
- [6] LI T, HU Z, JIA Y, et al. Forecasting crude oil prices using ensemble empirical mode decomposition and sparse Bayesian learning [J]. *Energies*, 2018, 11(7):1882.
- [7] LIAO J. Research on artificial intelligence prediction of international crude oil price based on VMD-LSTM-ELMAN model [J]. *Journal of Chengdu University of Technology (Natural Science Edition)*, 2024, 51(1):164-180.
- [8] SU Z H, SHANG L, LIU Z Z, et al. A stock price prediction method based on improved empirical mode decomposition and A-LSTM hybrid neural network [J]. *Journal of Nanjing University (Natural Sciences)*, 2025, 61(4):613-623.
- [9] JIANG F, HU C, WANG H. Crude oil futures price prediction based on multi-source and multi-task automatic encoder [J]. *Systems Engineering-Theory & Practice*, 2025, 45(2):702-716.
- [10] PAN S, AARON L, WANG Y, et al. LSTM neural network crude oil price forecast [J]. *Computer Technology and Development*, 2021, 31(5):180-185.
- [11] YANG J, XUE X. International Natural Uranium Price Prediction Based on TF-CNN-BiLSTM Model [J]. *Atomic Energy Science and Technology*, 2025, 59(6):1352-1360.
- [12] ZHOU H, ZHANG S, PENG J, et al. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting [C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021:11106-11115.
- [13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]// *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017:6000-6010.
- [14] LIU Y, ZENG A, XU Q, et al. iTransformer: Inverted Transformations Are Effective for Time Series Forecasting [C]// *The Twelfth International Conference on Learning Representations*. Vienna, Austria, 2024.
- [15] YU W, FANG F. Research on China crude oil futures index prediction based on LSTM neural network [J]. *Natural science journal of Harbin normal university*, 2020, 36(2):32-37.
- [16] NIE Y, NGUYEN N H, SINHTHONG P, et al. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers [C]// *Proceedings of the 11th International Conference on Learning Representations*. Addis Ababa: ICLR, 2023.
- [17] TORRES M E, COLOMINAS M A, SCHLOTTHAUER G, et al. A complete ensemble empirical mode decomposition with adaptive noise [C]// *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Prague, Czech Republic: IEEE, 2011:4144-4147.
- [18] PAN S, YANG F, ZHAO C. Crude Oil Futures Price Prediction Based on CEEMDAN-SA-TCN [J]. *Natural Gas & Oil*, 2025, 43(3):147-154.
- [19] GAO H, HU Y, YU L. Oil price prediction based on decomposition and integration of LSTM neural network model [J]. *Computer Applications and Software*, 2021, 38(10):78-83.
- [20] LIN Y, YU Y, ZHANG X, et al. Research on crude oil futures price prediction based on error correction and deep reinforcement learning [J]. *Systems Engineering Theory and Practice*, 2023, 43(1):206-221.
- [21] ZHAI D, ZHANG T, LIANG G, et al. Research on crude oil futures price prediction methods: A perspective based on quantum deep learning [J]. *Energy*, 2025, 320:135080.
- [22] METWALLY D S, ALI M, ALGHAMDI S M, et al. A novel hybrid model to forecast the stock price based on CEEMDAN and support vector regression [J]. *Journal of Radiation Research and Applied Sciences*, 2025, 18(2):101385.
- [23] BI M, LIU Z, YANG X. A Hybrid Prediction Model for International Crude Oil Price Based on Variational Mode Decomposition with BiTCN-BiGRU-Attention Deep Learning Techniques [EB/OL] [2025-10-10]. <https://doi.org/10.21203/rs.3.rs-5284460/v1>.
- [24] CHEN T, GUESTRIN C. Xgboost: A scalable tree boosting system [C]// *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2016:785-794.
- [25] BREIMAN L. Random forests [J]. *Machine Learning*, 2001, 45(1):5-32.
- [26] ZHAO G Y, XUE M G. Prediction of the next-day price change direction and volatility of crude oil futures based on TN-LP-LSTM-SVM mixed model [J]. *Control and Decision*, 2022, 37(10):2627-2636.



LI Tengjia, born in 1987, Ph.D candidate. Her main research interests include technical economy, and energy economy and management.