

# 基于增量式因果Transformer的高频金融时序预测

肖焕瑀, 郭躬德\*

(福建师范大学 计算机与网络空间安全学院, 福建 福州 350117)

**摘要:** 高频金融数据的非线性特征和剧烈波动, 使传统预测模型难以有效捕捉动态关系和跨市场泛化规律。针对这一问题, 提出一种融合元学习与动态因果推理的模型(Meta-causal Transformer, MCT)。设计一种基于滑动窗口和Granger检验的增量式因果发现算法, 动态更新市场因果图, 并以自适应衰减机制实时追踪因果关系演变。在此基础上构建因果约束的transformer架构, 通过注意力掩码显式抑制噪声导致的伪相关性。进一步结合元学习优化机制, 以快速适应不同市场情境下的动态因果模式。基于中国A股和美国股市Level2高频订单簿数据上的实验结果表明, 方向预测准确率相比其他模型提高了7%~15%且预测延迟降低1~3个时间步。为高频金融时序预测提供了可解释的因果推理框架, 并增强了实时交易决策的可靠性。

**关键词:** 元学习; 因果推理; 高频股价预测; Transformer; 订单簿分析

**中图分类号:** TP181 **文献标志码:** A **文章编号:** 0253-2395(2026)02-0199-10

## High-frequency Financial Time Series Prediction Based on Incremental Causal Transformer

XIAO Huanyu, GUO Gongde\*

(College of Computer and Cyber Security, Fujian Normal University, Fuzhou 350117, China)

**Abstract:** High-frequency financial data exhibit pronounced nonlinear characteristics and high volatility, significantly complicating accurate market forecasting. Traditional predictive models, reliant upon static correlations among features, fail to capture dynamic causal structures, thereby lacking robust generalization across varying market conditions. To address these challenges, this study proposes a novel forecasting model, the meta-causal transformer (MCT), which synergistically integrates meta-learning with dynamic causal inference. Specifically, an incremental causal discovery algorithm, leveraging a sliding-window approach and adaptive Granger causality testing, is introduced to dynamically reconstruct causal relationships among market variables. An adaptive decay mechanism further enhances this capability, allowing real-time tracking of rapidly evolving causal patterns. The causal-constrained Transformer architecture utilizes attention masks explicitly to eliminate spurious correlations driven by market noise, thus strengthening the model's interpretability and predictive robustness. Additionally, a meta-learning framework is employed to ensure rapid adaptation of model parameters to diverse causal scenarios during market regime shifts. Empirical evaluations on Level 2 high-frequency order book datasets from the Chinese A-share and U.S. stock markets demonstrate that the MCT model achieves substantial improvements, yielding a 7%-15% increase in directional prediction accuracy and a reduction in prediction latency by 1-3 time steps compared to state-of-the-art benchmarks. This research provides an interpretable and dynamically adaptive causal inference methodology, offering robust decision-making support for real-time high-frequency trading systems.

**Key words:** meta-learning; causal inference; high-frequency stock prediction; Transformer; order book analysis

收稿日期: 2025-07-01; 修回日期: 2025-09-23

基金项目: 国家自然科学基金(61976053; 62171131); 福建省自然科学基金(2023J01532)

作者简介: 肖焕瑀(1998-), 男, 福建永安市, 硕士研究生, 研究方向为人工智能。E-mail: 893600833@qq.com

\* 通信作者: 郭躬德(GUO Gongde), E-mail: ggd@fjnu.edu.cn

引文格式: 肖焕瑀, 郭躬德. 基于增量式因果Transformer的高频金融时序预测[J]. 山西大学学报(自然科学版), 2026, 49(2): 199-208. DOI: 10.13451/j.sxu.ns.2025109.

## 0 引言

随着算法交易与高频量化策略的发展,高频订单流数据成为金融市场微观结构研究的核心。国际清算银行统计,高频交易量已占全球股市成交量的60%以上,每秒生成数千笔订单,形成多尺度复杂系统。这些数据不仅反映价格形成机制,还蕴含市场博弈的因果关系。其中基于订单簿数据的预测模型成为决定交易策略盈亏的关键要素<sup>[1-3]</sup>。然而,高频交易预测面临的主要挑战包括:

1) 高频订单流数据中噪声占比高,传统模型易受虚假统计相关性误导,导致过拟合与泛化能力不足;

2) 市场微观结构中的因果关系随波动率、流动性状态非线性演化,现有方法难以实时捕捉因果结构突变;

3) 微观因子对价格的影响路径具有显著状态依赖性,传统黑箱模型缺乏因果机制建模能力,难以适应突发市场状态切换。

针对上述挑战,本文提出元因果 transformer (Meta-causal Transformer, MCT) 模型。主要贡献体现在三个层面:

1) 提出元学习驱动的增量式因果发现算法,结合滑动窗口 Granger 检验与自适应衰减机制,实时追踪市场状态驱动的因果结构演化;

2) 提出将动态因果图编码为 transformer 注意力掩码,显式抑制非因果交互路径,增强模型抗噪能力与可解释性;

3) 提出基于市场波动率聚类构建元任务分布,通过模型无关元学习 (Model-agnostic Meta-Learning, MAML) 框架实现模型参数快速迁移,提升对未知市场状态的预测鲁棒性。

## 1 相关工作

传统时序预测模型<sup>[4-6]</sup>多基于线性假设,难以刻画订单簿不平衡度,订单流加速度等微观特征的非线性演化。Tsantekidis 等<sup>[7]</sup>基于限价订单簿 (Limit Order Book, LOB) 的平稳特征,提出了卷积长短时记忆网络 (Convolutional Neural Network, CNN-LSTM) 混合模型,用于预测 LOB 的中间价格变动方向。Zhang 等<sup>[8]</sup>基于深度学习理论,根据中国顶尖企业的特点构建了股票价格预测指标体系,采用长短期记忆 (Long

Short-term Memory, LSTM) 循环神经网络预测。尽管基于时间序列预测的深度学习方法,通过递归或卷积结构提升了特征表达能力,这些深度学习方法受黑箱特性限制,难以捕捉市场因果结构,导致在概念漂移场景下泛化能力不足。近期, Hu 等<sup>[9]</sup>将静态元数据和动态时间序列数据融合于时序融合 transformer (Temporal Fusion Transformer, TFT) 模型中,验证了 TFT 在拥有足够训练数据时对股票价格的预测精度更高,且能够自动筛选重要特征。Choe 等<sup>[10]</sup>结合小波变换降噪与 TFT 模型,降低了金融数据噪声并提升了解释性,实现了对股票指数的有效预测。Shabani 等<sup>[11]</sup>引入注意力机制模型能识别关键时序模式,但仍依赖统计相关性,在市场极端波动时易受噪声误导,造成系统性误判。

因果推理技术为上述问题提供了新的解决思路。Granger 因果检验<sup>[12]</sup>方法可通过统计检验识别变量间的因果依赖,但其静态假设与高频场景的动态特性存在根本冲突。在订单流数据中,大额买单对价格变动的因果效应会随市场波动率状态发生显著变化。动态因果发现方法<sup>[13]</sup>通过滑动窗口更新因果图,但计算复杂度高且难应对突发性市场状态切换。与此同时,元学习<sup>[14]</sup>在少样本快速适应任务中展现出独特优势,但目前金融预测中的应用仍主要局限于跨资产泛化,未充分结合动态因果推理。

## 2 预备知识

### 2.1 元学习

元学习<sup>[15-16]</sup>通过“学会学习”机制,使模型从任务分布中提取可迁移的元知识,从而在少量样本下快速适应新任务。其核心范式可形式化为双层优化问题。在给定任务分布  $\mathcal{T} \sim p(\mathcal{T})$  下,学习一个最优参数初始化  $\theta$ , 使模型在新任务  $\mathcal{T}_i \in \mathcal{T}$ , 上通过少量梯度更新即可优化目标函数。

相较于传统迁移学习仅依赖源域与目标域的隐式相似性假设,元学习通过显式优化任务间共享的归纳偏置,在政策调整、“黑天鹅”事件等突发场景中展现出更强的鲁棒性。

### 2.2 因果推理

因果推理旨在区分因果关系与统计相关性。主流方法包括供应链管理 (Supply Chain Management, SCM)<sup>[17]</sup>, 潜在结果框架<sup>[18]</sup> 和

Granger 因果关系<sup>[19]</sup>。

Granger 因果检验基于时间序列数据,给定平稳时间序列  $X_t$  和  $Y_t$ , 构建自回归模型:

$$Y_t = \sum_{i=1}^p \alpha_i Y_{t-i} + \sum_{j=1}^q \beta_j X_{t-j} + \epsilon_t, \quad (1)$$

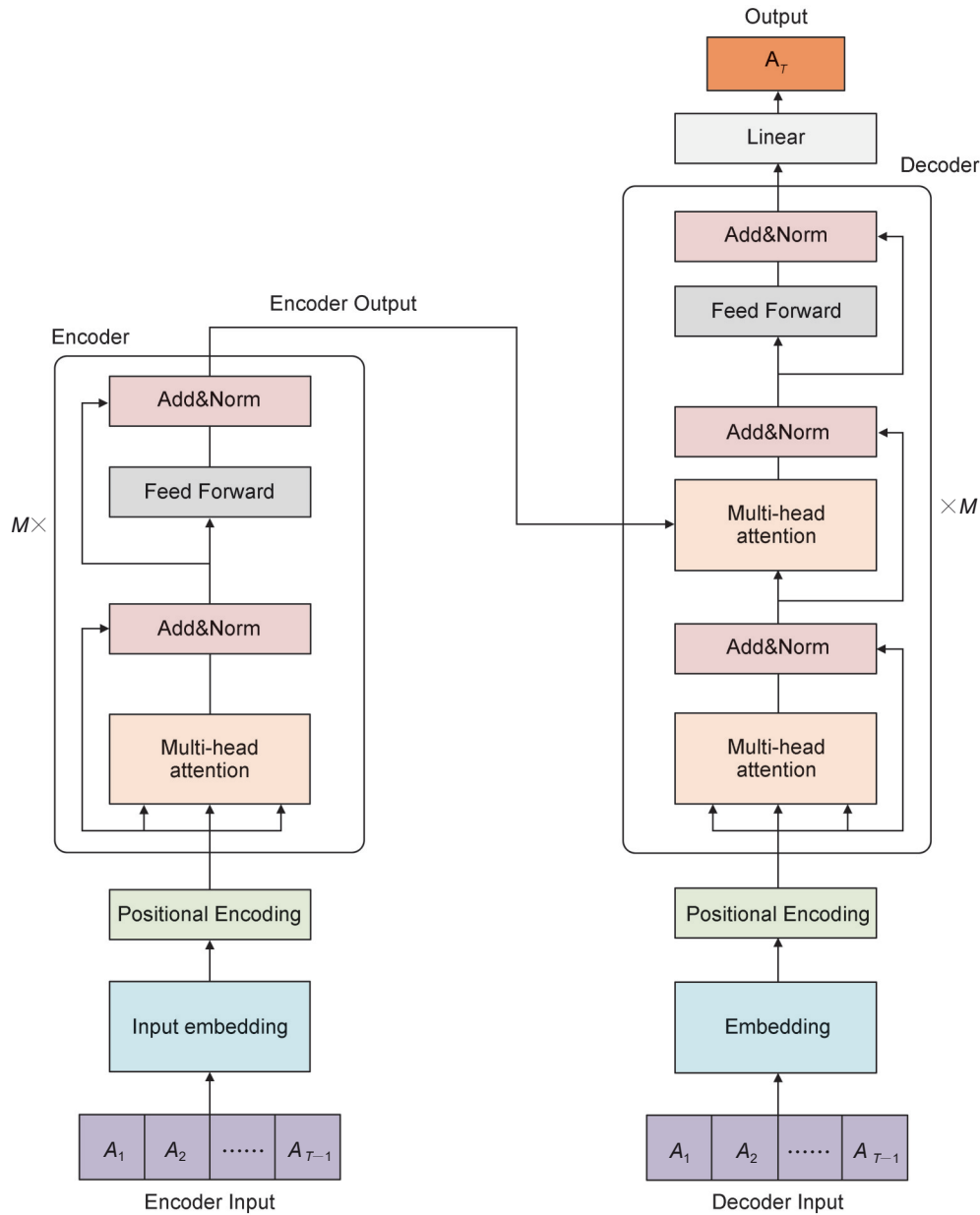
其中  $\alpha_i$  和  $\beta_j$  为对应滞后项的回归系数,  $\epsilon_t$  为白噪声误差项,若拒绝原假设  $\beta_j = 0$ , 则称  $X$  Granger 导致  $Y$ 。然而,传统 Granger 检验存在两点局限。一是静态性假设,二是线性依赖假设。为克服上述局限,动态因果发现通过滑动窗口机制实时更新因果图。定义窗口  $w$  内的因果图  $G_t$ :

$$G_t = \mathcal{F}(D_{t-w:t}, \Theta), \quad (2)$$

其中  $\mathcal{F}$  为因果发现算法,  $\Theta$  为超参数。在高频股价预测中,动态因果推理可识别两类关键因果路径。通过抑制非因果关联,可减少噪声干扰。

### 2.3 Transformer

Transformer<sup>[20]</sup> 基于自注意力机制,如图 1 所示。克服了循环神经网络(Recurrent Neural Network, RNN)的时间依赖性限制,提升了长时依赖建模能力,在高频金融时序预测任务中展现出显著优势。相较于传统 RNN 的递归计算模式更适配高频数据的实时处理需求。此外,注意力权



注:  $M$  为特征维度;  $T$  为时间步数,  $A_T$  表示第  $T$  步序列表示。

图 1 标准 Transformer 框架图

Fig. 1 Framework of standard Transformer

重矩阵的可视化特性为金融时序预测中的因果推理提供了直观支持,例如通过权重分布识别订单流不平衡对价格波动的驱动效应。

### 3 MCT 模型

MCT 模型是一种动态适应性金融预测框架,如图 2 所示。MCT 框架通过深度融合元学习与因果增强技术,构建了一个针对高频金融数据的自适应预测体系。数据预处理与特征工程模块从原始订单簿与成交流中提取多尺度、高维度的时序特征,如短周期滚动不平衡度,价差加速度,累积加权平均价(Volume-Weighted Average Price, VWAP)及动量指标。并借助平滑插值,离散化处理与主成分分析(Principal Component Analysis, PCA)方法,有效去噪与降维,为下游模型提供高质量输入。元学习优化器则基于市场流动性与波动率信息,以 MAML 双层迭代机制重构任务分布:内

循环在支持集上实现快速参数微调以捕捉局部市场特性,外循环通过查询集梯度反馈不断更新全局初始化参数,从而在多样化市场环境中展现出卓越的迁移与泛化能力。增量式因果发现模块结合滑动窗口内的 Granger 因果检验与统计显著性检验,动态生成稀疏布尔因果掩码,显式过滤伪相关信号并揭示特征间潜在因果结构。因果 Transformer 利用该掩码对多头注意力权重进行加权约束,使模型在编码阶段即融入因果推断信息。后续的 Sigmoid 激活与深度全连接网络进一步对因果特征进行精炼与映射,最终输出多步预测序列。

实时监控与反馈调整机制则持续监测预测误差、因果一致性及市场微观结构指标,当关键指标越过预设阈值时自动触发因果掩码更新频率与元任务划分策略的自适应调整,以保障模型在高频交易环境下的稳健性与鲁棒性。

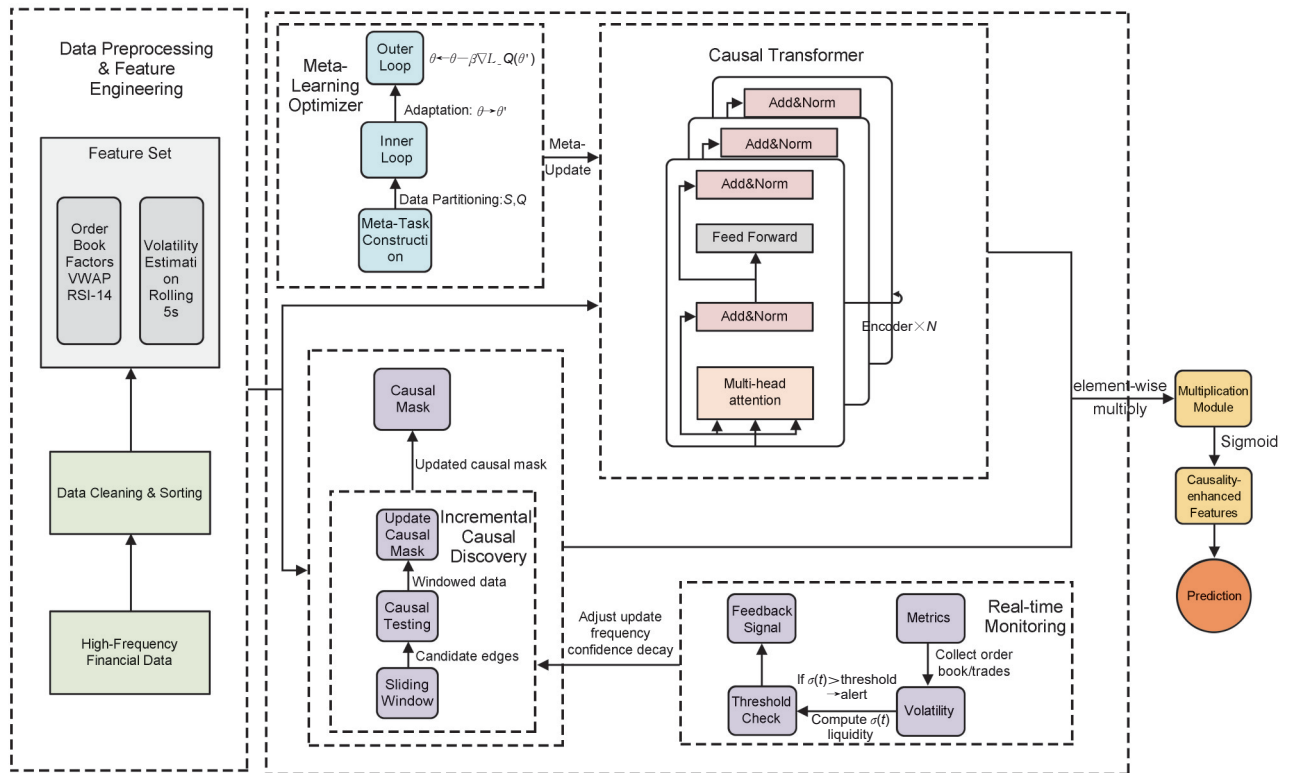


图2 MCT 模型总体框架图

Fig. 2 Overall framework of the MCT model

#### 3.1 元学习优化器

本文提出了一种结合市场波动率和流动性指标的分层任务划分策略,以增强元学习的适

应性与泛化能力。通过滑动窗口方式将时间序列数据划分为多个任务,每个任务包含支持集  $D_{\text{support}}$  和查询集  $D_{\text{query}}$  两部分数据。其中支持集

用于内循环快速适应局部市场特征,查询集用于外循环优化模型的泛化能力。每个任务划分时间,时间序列的前80%用于支持集,后20%用于查询集,以最大化数据的时间依赖性。

在每个元任务上,模型先使用当前全局模型参数 $\theta$ 在 $D_{\text{support}}$ 上进行梯度下降更新,得到任务特定的临时参数 $\theta_i^*$ 。更新步长由元学习率 $\eta$ 控制:

$$\theta_i^* = \theta - \eta \nabla_{\theta} \mathcal{L}(f_{\theta}(D_{\text{support}})). \quad (3)$$

模型在 $D_{\text{query}}$ 上评估临时参数 $\theta_i^*$ 的损失,并利用该梯度反向优化全局参数 $\theta$ ,形成元更新:

$$\theta \leftarrow \theta - \beta \sum_{i=1}^K \nabla_{\theta} \mathcal{L}(f_{\theta_i^*}(D_{\text{query}})), \quad (4)$$

其中 $\beta$ 为元梯度学习率, $K$ 表示元任务数。通过这种双层更新机制,模型在每个小任务上快速适应,同时在更广泛的市场环境中积累知识,从而在面对新环境或市场状态突变时仍能保持较高的预测精度。

### 3.2 增量式因果发现算法

MCT采用增量式因果发现算法以动态捕捉市场因果结构,并通过自适应衰减机制增强模型对市场状态变化的响应能力。首先,对高频订单流数据进行ADF单位根检验,识别非平稳序列并进行差分处理,以确保因果分析的统计稳健性。随后,在滑动窗口 $w$ 内,对各变量进行Granger因果检验,筛选显著因果关系并构建市场因果图 $G_t$ 。

在因果权重更新过程中,采用指数衰减方式平衡历史信息与当前市场状态的影响,从而增强对近期因果关系变化的敏感度。权重更新公式定义为:

$$w_{i,j}^{(t)} = \alpha \cdot w_{i,j}^{(t-1)} + (1 - \alpha) \cdot I_{\text{Granger}}(X_i \rightarrow X_j), \quad (5)$$

其中 $\alpha$ 为衰减因子, $I_{\text{Granger}}$ 为Granger因果检验的显著性判别函数。 $\alpha$ 值较小时,模型更依赖近期市场信息,以适应市场结构的快速变化。 $\alpha$ 取值较大时,因果图更新较为平稳,可避免过度响应短期噪声波动。

### 3.3 因果约束Transformer

本文设计了一种因果约束Transformer以增强MCT对市场因果关系的建模能力,如图3所示。通过注意力机制<sup>[21]</sup>显式引入因果关系约束,以提升模型的可解释性和预测稳定性。在标准Transformer结构的基础上,引入因果注意力掩码 $M \in \{0, 1\}^{d \times d}$ ,其中 $M_{i,j} = 1$ 表示 $i$ 对 $j$ 有

因果影响,否则 $M_{i,j} = 0$ 。

在计算注意力权重时,因果掩码用于约束注意力矩阵,使模型仅关注因果相关的变量,抑制统计相关性驱动的非因果交互:

$$A_{i,j} = \frac{\exp\left(\frac{Q_i K_j^T}{\sqrt{d_k}}\right) \cdot M_{i,j}}{\sum_{k=1}^d \exp\left(\frac{Q_i K_k^T}{\sqrt{d_k}}\right) \cdot M_{i,k}}, \quad (6)$$

其中 $Q$ 和 $K$ 分别表示查询矩阵和键矩阵, $d_k$ 为注意力权重的缩放因子。当 $M_{i,j} = 0$ 时,对应的注意力得分将被屏蔽,以确保信息流仅沿着因果路径传播。

此外,为进一步增强因果结构的显式建模能力,本文引入因果投影模块(causal projection module),利用掩码引导的特征重构机制,将注意力输出转化为因果导向的特征表示,并通过双层全连接网络与ReLU激活函数实现最终的非线性映射。

## 4 实验与结果分析

为验证MCT模型的有效性,本文基于沪深300和标普500股票的Level-2高频订单簿数据进行实验,并与传统时序预测方法<sup>[22]</sup>,深度学习模型<sup>[23]</sup>及前沿模型TFT<sup>[7]</sup>,随机欠采样提升算法(Random Under-sampling Boosting, RUS-BOOST)<sup>[24]</sup>,自适应学习策略引擎(Adaptive Learning Policy Engine, ALPE)<sup>[25]</sup>进行对比。实验采用方向预测准确率Acc,ROC曲线下面积(Area Under the Curve of the Receiver Operating Characteristic, AUC),预测延迟及因果结构稳定性(jaccard指数 $T_{\text{Latency}}$ )作为评估指标。其计算公式分别为:

$$\text{Acc} = \frac{1}{N} \sum_{t=1}^N \mathbf{1}(\text{sign}(\hat{y}_t) = \text{sign}(y_t)), \quad (7)$$

$$\text{AUC} = \frac{1}{N_+ N_-} \sum_{i: y_i=1} \sum_{j: y_j=0} \mathbf{1}(\hat{p}_i > \hat{p}_j), \quad (8)$$

$$T_{\text{Latency}} = \frac{1}{M} \sum_{i=1}^M (t'_i - t_i). \quad (9)$$

### 4.1 实验设置

实验采用网格搜索<sup>[26]</sup>与贝叶斯优化<sup>[27]</sup>相结合的方法优化MCT的关键参数。首先,在20%的训练数据上进行粗粒度网格搜索,筛选

出内循环学习率,滑动窗口长度,注意力头数等候选参数。随后,基于贝叶斯优化在完整训练集上迭代 50 轮,确定  $\gamma=0.05$  和  $\alpha=0.9$  的最优组合,其中  $\gamma$  表示因果掩码生成时所用的显著性水平阈值,  $\alpha$  为历史因果权重的指数衰减系数。调参过程中引入早停机制,避免过拟合。最终,最优参数组合设置自适应学习率  $\eta$  通过贝叶斯优化动态调整为 0.009 8,外循环学习率  $\beta$  经网格搜索确定为 0.003 2, dropout 率,学习率以及 L2 正则化系数确定为 0.25, 0.001, 0.000 1。

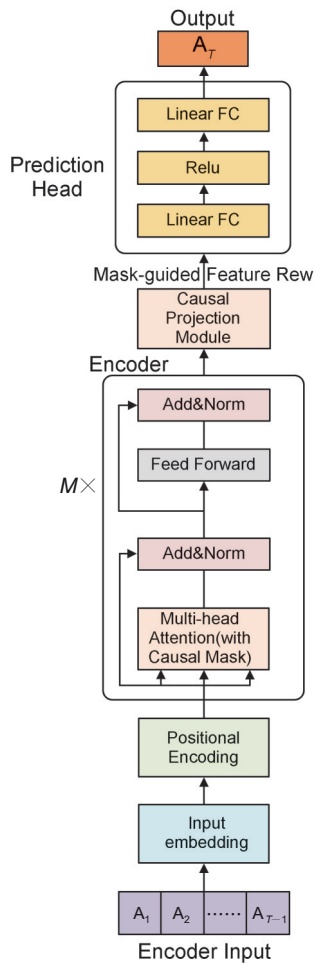


图3 因果Transformer框架图  
Fig. 3 Framework of causal Transformer

为评估模型在短期市场动态中的泛化能力,实验采用 5 日滚动时间窗进行时序交叉验证<sup>[28]</sup>,如表 1 所示。

4.2 评价指标

MCT 在沪深 300 和标普 500 市场的实验结果如表 2 和表 3 所示。结果表明,MCT 在方向预测准确率上较对比模型提高 4%~15%。

表1 短线交叉验证结果

Table 1 Short-term cross-validation results

验证轮次	训练时段	验证时段	Acc/%
1	2025.01.01—01.05	2025.01.06—01.10	70.1
2	01.06—01.10	01.11—01.15	69.8
3	01.11—01.15	01.16—01.20	71.3
4	01.16—01.20	01.21—01.25	73.4
5	01.21—01.25	01.26—01.30	70.6

而通过 AUC 指标的跨模型对比,验证了因果约束机制对 MCT 性能提升的核心作用。在沪深 300 市场中,MCT 的 AUC 值达 0.710 6,较无因果约束版本,TFT,RUS-BOOST 和 ALPE 分别提升 11.0%, 9.7%, 5.7% 与 2.5%,且显著超越 LSTM 与 XGBoost。标普 500 市场的实验结果进一步强化了这一结论。表明因果建模能力具有跨市场普适性。

表2 各模型在沪深 300 运行结果

Table 2 Performance of different models on CSI 300

模型	准确率	AUC-ROC	延迟(时间步)	Jaccard
LSTM	0.583	0.580 7	5.1	—
Prophet	0.521	0.482 5	4.4	—
XGBoost	0.552	0.537 0	3.5	—
TFT	0.675	0.648 0	2.8	—
RUS-BOOST	0.695	0.672 0	2.6	—
ALPE	0.705	0.693 0	2.3	—
MCT(无因果约束)	0.667	0.640 2	3.2	0.33
<b>MCT</b>	<b>0.737</b>	<b>0.710 6</b>	<b>1.95</b>	<b>0.63</b>

注:—表示模型不存在 Jaccard 指数;加粗表示本文提出模型及结果。

表3 各模型在标普 500 运行结果

Table 3 Performance of different models on S&P 500

模型	准确率	AUC-ROC	延迟(时间步)	Jaccard
TFT	0.642	0.615 3	3.0	—
RUSQ-BOOST	0.670	0.667 0	3.3	—
ALPE	0.690	0.687 0	3.1	—
MCT(无因果约束)	0.583	0.657 0	3.6	0.31
<b>MCT</b>	<b>0.716</b>	<b>0.702 0</b>	<b>2.9</b>	<b>0.58</b>

注:—表示模型不存在 Jaccard 指数;加粗表示本文提出模型及结果。

MCT 在沪深 300 市场的平均预测延迟仅 1.95 时间步,比 TFT,RUS-BOOST 与 ALPE 分别降低 30.4%, 25%, 15.2%。显示因果约束有效抑制噪声干扰,显著提高了模型的市场响应速度。

在因果结构稳定性方面,采用 Jaccard 指数

进行衡量。移除因果掩码后,因果结构稳定性显著下降,说明未显式建模因果关系时,模型易受市场噪声影响。而MCT模型则在动态因果图的约束下,Jaccard指数分别达到沪深300市场的0.63和标普500市场的0.58,显示出稳健的因果关系捕捉能力和更好的预测表现。

#### 4.3 实验结果分析

方向预测混淆矩阵如图4—图5所示。MCT在两类市场中均能精准区分价格涨跌趋势,如沪深300下跌趋势的误报率仅为12.8%。

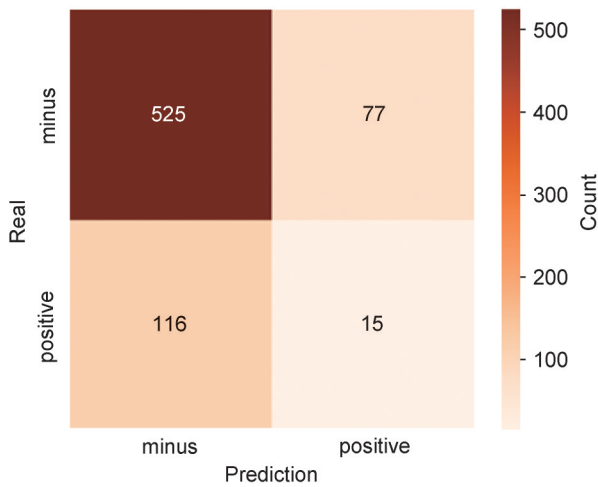


图4 沪深300市场方向预测混淆矩阵热力图

Fig. 4 Confusion matrix heatmap of direction prediction for CSI 300 market

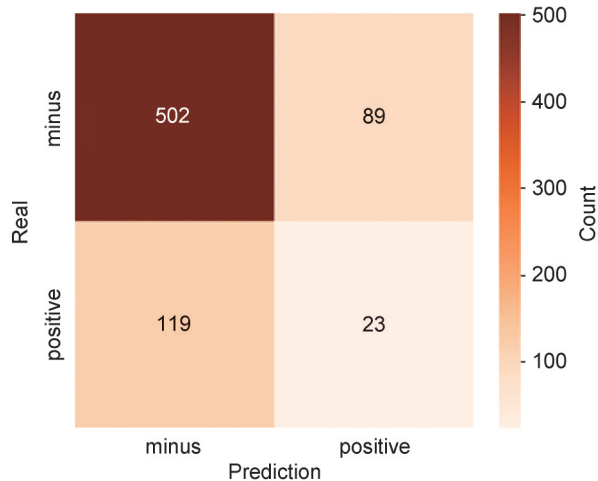


图5 标普500市场方向预测混淆矩阵热力图

Fig. 5 Confusion matrix heatmap of direction prediction for S&P 500 market

通过主成分分析对沪深300市场状态进行深入分析,如图6所示。将市场微观特征划分为流动性充裕、平衡态与流动性枯竭三类<sup>[29]</sup>。

从图中可以看出,PC1与PC2分别解释了60%和25%的方差,三种状态在主成分空间中表现出明显差异。进一步分析发现,流动性充裕状态下市场波动率较低且订单簿不平衡度较小。平衡态的波动率和订单簿不平衡度处于中等水平。流动性枯竭状态下市场波动率显著升高且订单簿不平衡度明显加剧。这种明确的市场状态区分为MCT模型提供了动态调整因果关系分析策略的基础,使其能够实时精准捕捉不同状态下的关键市场特征。

进一步对比不同模型的注意力权重分布,如图7所示。标准Transformer过度关注非因果交互路径,如某些特征对VWAP和流动性特征的注意力权重异常升高,与金融领域中量价联动的基本原理不符。此外,兄弟特征之间的冗余注意力进一步表明标准Transformer过度依赖历史值,忽视了实际因果驱动因素。

而MCT通过因果掩码机制有效减少非因果特征间的冗余交互,使注意力更集中于真实的因果路径。显著强化了经济学中符合量价联动原理的因果路径如price→volume的权重提高至0.85,同时削弱了flow→flow的冗余自回归依赖,减少了对自身历史信息的过度依赖。此外,MCT还突出学习到了一些隐藏的因果特征独立性,如内部特征间注意力权重提升至0.99,暗示模型捕捉到潜在的独立订单流信号,这是订单簿不平衡的经济学体现。

为定量验证因果掩码机制的有效性,引入IoU指标<sup>[30]</sup>(Intersection over Union)进行严格的量化评估:

$$I_{ou} = \frac{|M_{pred} \cap M_{true}|}{|M_{pred} \cup M_{true}|}, \quad (10)$$

其中 $M$ 为二值化后的因果掩码矩阵。实验表明,MCT模型的IoU值达到0.68,显著高于标准Transformer的0.32,进一步验证了因果建模的优势。

## 5 结论

本研究提出了融合元学习与动态因果推理的MCT模型,通过增量式因果发现算法实时更新因果图,引入自适应衰减机制抑制伪因果关系,实验揭示高频订单流数据中的状态依赖性

与因果稳定性,并验证了传统统计相关性方法在极端市场波动下易导致系统性误判的缺陷。因果约束机制显著提升了方向预测准确率与稳定性,完善了高频金融预测理论。然而,本研究尚未深入探索多市场、多资产联动的复杂因

果网络,且模型在处理更长期的非平稳市场状态和罕见事件如政策突发调整时的泛化性能仍需进一步验证。未来研究方向可考虑引入跨市场联动机制和长期趋势预测能力,并进一步整合跨市场宏观因子以增强模型泛化能力。

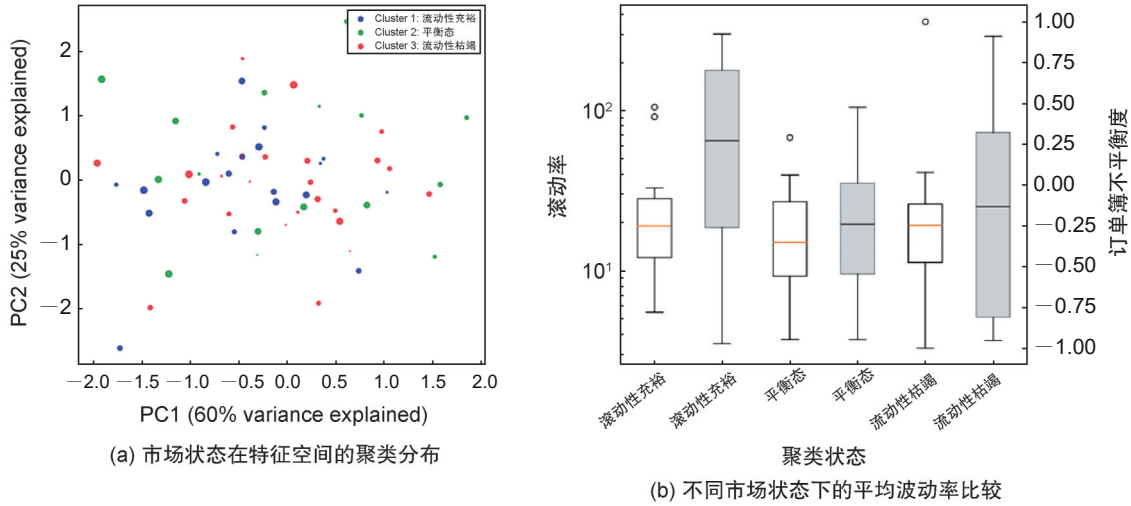


图6 市场状态聚类与波动特征图

Fig. 6 Market state clustering and volatility feature analysis

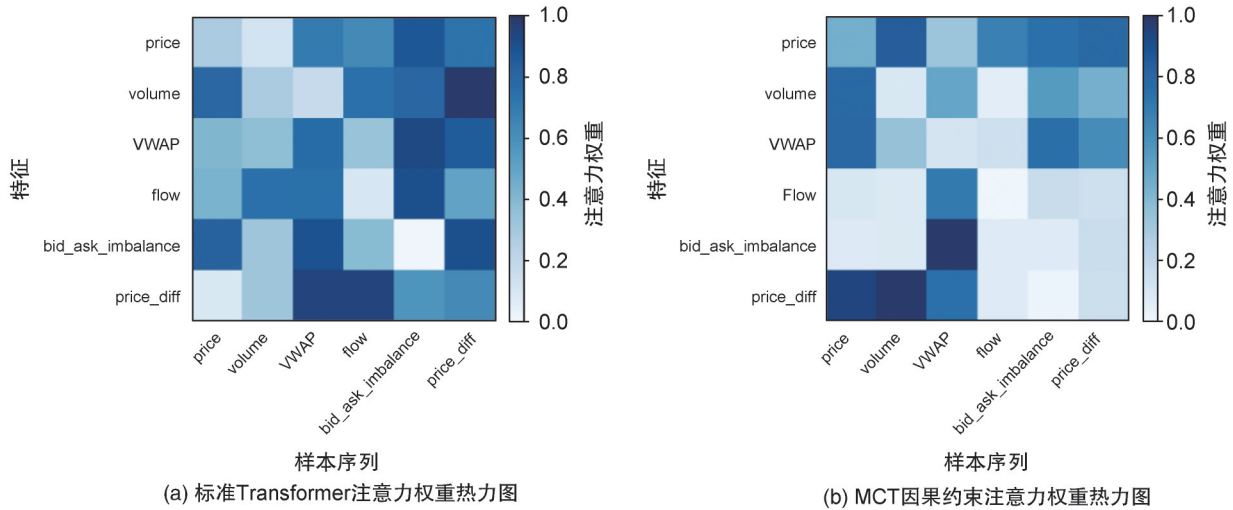


图7 模型特征注意力权重热力对比图

Fig. 7 Comparison of model feature attention weight heatmaps

参考文献:

[1] GOODELL J W, KUMAR S, LIM W M, *et al.* Artificial Intelligence and Machine Learning in Finance: Identifying Foundations, Themes, and Research Clusters from Bibliometric Analysis[J]. *J Behav Exp Finance*, 2021, **32**: 100577. DOI: 10.1016/j.jbef.2021.100577.

[2] LIU Z H, SHAM C W, MA L Y, *et al.* ViT-LOB: Efficient Vision Transformer for StockPrice Trend Prediction Using Limit Order Books[C]//2024 10th International Conference

on Applied System Innovation (ICASI). New York: IEEE, 2024: 436–438. DOI: 10.1109/ICASI60819.2024.10547868.

[3] LEON D, SANDOVAL J, CRUZ A, *et al.* Deep Heterogeneous AutoML Trend Prediction Model for Algorithmic Trading in the USD/COP Colombian FX Market through Limit Order Book (LOB) [J]. *SN Comput Sci*, 2024, **5**(5): 631. DOI: 10.1007/s42979-024-02930-1.

[4] LI Z W, LIU F, YANG W J, *et al.* A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects[J].

- IEEE Trans Neural Netw Learn Syst*, 2022, **33**(12): 6999–7019. DOI: 10.1109/TNNLS.2021.3084827.
- [5] YAO Y M. Data Analysis on the Computer Intelligent Stock Prediction Model Based on LSTM RNN and Algorithm Optimization[C]//2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA). New York: IEEE, 2022: 480–485. DOI: 10.1109/EEBDA53927.2022.9744859.
- [6] SIRISHA U M, BELAVAGI M C, ATTIGERI G. Profit Prediction Using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison[J]. *IEEE Access*, 2022, **10**: 124715–124727.
- [7] TSANTEKIDIS A, PASSALIS N, TEFAS A, *et al.* Using Deep Learning for Price Prediction by Exploiting Stationary Limit Order Book Features[J]. *Appl Soft Comput*, 2020, **93**: 106401. DOI: 10.1016/j.asoc.2020.106401.
- [8] ZHANG Y J, TUMIBAY G M. Stock Market Prediction Model Based on LSTM Deep Learning: The Case of Top Corporate Company in China[C]//2022 4th International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM). New York: IEEE, 2022: 451–455. DOI: 10.1109/AIAM57466.2022.00092.
- [9] HU X K. Stock Price Prediction Based on Temporal Fusion Transformer[C]//2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI). New York: IEEE, 2021: 60–66. DOI: 10.1109/mlbdbi54094.2021.00019.
- [10] CHOE U, SON H, RYU B, *et al.* Predicting Stock Price Index Using the Wavelet Transform and Temporal Fusion Transformer Model[C]//Proceedings of Annual Conference of KIPS. Seoul, Korea: Korean Information Processing Society, 2024: 401–404. DOI: 10.3745/PKIPS.Y2024M10A.401.
- [11] SHABANI M, TRAN D T, MAGRIS M, *et al.* Multi-head Temporal Attention-augmented Bilinear Network for Financial Time Series Prediction[C]//2022 30th European Signal Processing Conference (EUSIPCO). New York: IEEE, 2022: 1487–1491.
- [12] 任仙玲, 孙文岳. 中国股指期货市场间的非对称风险溢出效应: 基于分位数 Granger 因果检验[J]. *运筹与管理*, 2021, **30**(8): 190–197. DOI:10.12005/orms.2021.0265.
- REN X L, SUN W Y. Asymmetric Risk Spillovers between the Chinese Stock Market and Index Futures Market Based on Granger Non-causality Test in Quantiles[J]. *Oper Res Manag Sci*, 2021, **30**(8): 190–197. DOI:10.12005/orms.2021.0265.
- [13] 陈济斌, 蔡瑞初. 面向非稳态霍克斯过程的格兰杰因果发现算法研究[J]. *计算机科学与应用*, 2021, **11**(4): 821–831. DOI: 10.12677/CSA.2021.114084.
- CHEN J B, CAI R C. Research on Algorithms of Granger Causality Discovery for Non-stationary Hawkes Process[J]. *Comput Sci Appl*, 2021, **11**(4): 821–831. DOI: 10.12677/CSA.2021.114084.
- [14] 崔树银, 汪昕杰. 基于特征工程的集成学习短期光伏功率预测[J]. *科学技术与工程*, 2022, **22**(2): 532–539. DOI: 10.3969/j.issn.1671-1815.2022.02.013.
- CUI S Y, WANG X J. Short-term Photovoltaic Power Prediction by Ensemble Learning Based on Feature Engineering[J]. *Sci Technol Eng*, 2022, **22**(2): 532–539. DOI: 10.3969/j.issn.1671-1815.2022.02.013.
- [15] SHI J T, YAO H X, WU X, *et al.* Relation-aware Meta-learning for E-commerce Market Segment Demand Prediction with Limited Records[C]//Proceedings of the 14th ACM International Conference on Web Search and Data Mining. New York: ACM, 2021: 220–228. DOI: 10.1145/3437963.3441750.
- [16] CHANG S H, HSU C W, LI H Y, *et al.* Short-term Stock Price-trend Prediction Using Meta-learning[C]//2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC). New York: IEEE, 2021: 2900–2905. DOI: 10.1109/smc52423.2021.9658607.
- [17] FALATOURI T, DARBANIAN F, BRANDTNER P, *et al.* Predictive Analytics for Demand Forecasting: A Comparison of SARIMA and LSTM in Retail SCM[J]. *Procedia Comput Sci*, 2022, **200**: 993–1003. DOI: 10.1016/j.procs.2022.01.298.
- [18] 马忠贵, 徐晓晗, 刘雪儿. 因果推断三种分析框架及其应用综述[J]. *工程科学学报*, 2022, **44**(7): 1231–1243. DOI: 10.13374/j.issn2095-9389.2021.07.04.002.
- MA Z G, XU X H, LIU X E. Three Analytical Frameworks of Causal Inference and Their Applications[J]. *Chin J Eng*, 2022, **44**(7): 1231–1243. DOI: 10.13374/j.issn2095-9389.2021.07.04.002.
- [19] 李林泰, 崔巍. 铁矿石价格波动的因果推断: 影响因素与效应分析[J]. *技术经济*, 2024, **43**(8): 36–45. DOI: 10.12404/j.issn.1002-980X.J24051802.
- LI L T, CUI W. Causal Inference of Iron Ore Price Fluctuations: Analysis of Influencing Factors and Effects[J]. *J Technol Econ*, 2024, **43**(8): 36–45. DOI: 10.12404/j.issn.1002-980X.J24051802.
- [20] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is All You Need[C]//Proc of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Red Hook, NY: Curran Associates Inc, 2017: 6000–6010. DOI:10.48550/arXiv.1706.03762.
- [21] KISIEL D, GORSE D. Axial-LOB: High-frequency

- Trading with Axial Attention[C]//2022 IEEE Symposium Series on Computational Intelligence (SSCI). New York: IEEE, 2022: 1327–1333. DOI: 10.1109/SSCI51031.2022.10022284.
- [22] SEKHAR P C, PADMAJA M, SARANGI B, *et al.* Prediction of Cryptocurrency Using LSTM and XGBoost[C]//2022 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS). New York: IEEE, 2022: 1–5. DOI: 10.1109/ICBDS53701.2022.9935871.
- [23] SONG Y P, CAI C C, MA D X, *et al.* Modelling and Forecasting High-frequency Data with Jumps Based on a Hybrid Nonparametric Regression and LSTM Model [J]. *Expert Syst Appl*, 2024, **237**: 121527. DOI: 10.1016/j.eswa.2023.121527.
- [24] GOUDARZI M, BAZZANA F. Identification of High-frequency Trading: a Machine Learning Approach[J]. *Res Int Bus Finance*, 2023, **66**: 102078. DOI: 10.1016/j.ribaf.2023.102078.
- [25] NTAKARIS A, IBIKUNLE G. Minimal Batch Adaptive Learning Policy Engine for Real-Time Mid-Price Forecasting in High-Frequency Trading [EB/OL]. (2024–12–26)[2025–05–21]. <https://arXiv.org/abs/2412.19372>.
- [26] SHAMS M Y, ELSHEWEY A M, EL-KENAWY E M, *et al.* Water Quality Prediction Using Machine Learning Models Based on Grid Search Method[J]. *Multimed Tools Appl*, 2024, **83**(12): 35307–35334. DOI: 10.1007/s11042-023-16737-4.
- [27] RIMAL Y, SHARMA N, ALSADOON A. The Accuracy of Machine Learning Models Relies on Hyperparameter Tuning: Student Result Classification Using Random Forest, Randomized Search, Grid Search, Bayesian, Genetic, and Optuna Algorithms[J]. *Multimed Tools Appl*, 2024, **83**(30): 74349–74364. DOI: 10.1007/s11042-024-18426-2.
- [28] HIDAYAT T, MANONGGA D, HENDRY, *et al.* Performance Prediction Using Cross Validation (Grid-SearchCV) for Stunting Prevalence[C]//2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS). New York: IEEE, 2024: 1–6. DOI: 10.1109/AIMS61812.2024.10512657.
- [29] BOGOUSSLAWSKY V, COLLIN-DUFRESNE P. Liquidity, Volume, and Order Imbalance Volatility[J]. *J Finance*, 2023, **78**(4): 2189–2232. DOI: 10.1111/jofi.13248.
- [30] SU K K, CAO L H, ZHAO B T, *et al.* N-IoU: Better IoU-based Bounding Box Regression Loss for Object Detection[J]. *Neural Comput Appl*, 2024, **36**(6): 3049–3063. DOI: 10.1007/s00521-023-09133-4.