

含交叉注意力机制的趋势预测模型

——学界纵横系列之七

核心观点

本文解析了 Wood, K.等人关于时序趋势预测模型的最新研究成果，通过构造上下文集并引入交叉注意力机制来优化时序趋势预测模型，比传统趋势追踪策略和深度动量网络模型有更高的风险调整收益，并且在市场波动期间回撤修复的速度更快。

□ 要点一

预测模型如何快速适应市场条件变化一直是量化交易领域的难题。传统的深度学习模型对于表示学习通常依赖于大型静态数据集，而市场的非平稳性可能导致这些模型的性能下降。X-Trend 策略模型是利用交叉注意力机制优化的时序趋势预测模型。它利用小样本学习和变点检测技术以适应新的市场环境。回测结果表明，X-Trend 在 2020-2021 公共卫生事件期间出现回撤的恢复速度是同类型趋势策略的两倍左右。在 2018 年至 2023 年期间，相较于基线模型，夏普比率提高了 18.9%；与传统时序动量策略相比，夏普比率提高了约 10 倍。

□ 要点二

X-Trend 模型优势来源于交叉注意力机制，它可以将目标序列和从上下文集中的相似模式相关联，从而提高预测胜率。甚至在样本或计算资源缺乏的零次学习场景下，X-Trend 模型相较于深度动量网络基线模型和传统趋势追踪策略在风险收益比上均有优势。同时，X-Trend 模型在训练过程中可以使用灵活的目标函数，不论是极大似然估计还是分位数回归，均有超越基准的收益表现。

□ 要点三

使用深度神经网络构建系统性交易策略，考虑到实际样本数量限制、以及市场行情快速变化带来的样本差别较大，是典型的小样本学习场景。X-Trend 模型是金融领域小样本学习的成功案例。构造上下文集，并引入交叉注意力机制，能够将其他资产或历史市场行情中相似模式的信息，应用到不同资产或新的市场环境中，从而提高策略模型的胜率。经过训练，X-Trend 模型参数中包含了市场通用模式或规则（regime）的丰富信息。

□ 风险提示

本报告中包含公开发表的文献整理的模型结果，涉及的收益指标等结果的解释性请参考原始文献。回测结果是基于历史数据的统计归纳，模型力求跟踪市场规律和趋势，但仍存失效可能，不构成投资建议，须谨慎使用。

分析师：陈奥林

执业证书号：S1230523040002

chenaolin@stocke.com.cn

相关报告

1 《C 端落地 Sora 催化，关注 AI 投资价值》 2024.03.27

2 《全球宏观投资：宏观趋势与宏观套利》 2024.03.26

3 《财报季：重回基本面定价》 2024.03.24

正文目录

1 选题背景	4
1.1 金融市场通用模式的假设	4
1.2 时序动量模型的演变	4
1.3 小样本学习	5
2 核心结论	6
3 时序趋势预测模型 X-Trend	6
3.1 X-Trend 模型设计	7
3.1.1 序列表征与基线模型	7
3.1.2 由“上下文语境”衍生的环境构建	8
3.1.3 交叉注意力机制的实现	8
3.1.4 解码器与损失函数的设计	9
3.2 实战效果	9
3.2.1 Few-shot 实验	10
3.2.2 Zero-shot 实验	11
3.2.3 模型可解释性分析	12
4 总结	12
4.1 原文结论	12
4.2 我们的思考	13
5 风险提示	13
参考文献	13

图表目录

图 1: Momentum Transformer 结构及深度动量网络的回测表现.....	5
图 2: Encoder-decoder 结构及信息传导示意图	6
图 3: 目标序列中的信息如何影响上下文集	8
图 4: Encoder-decoder 结构及信息传导示意图	9
图 5: 小样本情景下各策略的累计收益对比 (2013/01-2023/01)	10
图 6: 小样本情景下策略的最大回撤对比 (2018/01-2023/01)	10
图 7: 零次学习情景下各策略收益对比 (1995/01-2023/01)	11
图 8: 零次学习情景下策略最大回撤对比 (2018/01-2023/01)	11
图 9: Encoder-decoder 结构及信息传导示意图	12
表 1: 小样本情景下策略模型收益表现对比	10
表 2: 零次学习情景下策略模型收益表现对比	11

1 选题背景

预测模型如何快速适应市场条件变化一直是量化交易领域的难题。传统的深度学习模型对于表示学习通常依赖于大型静态数据集，而市场的非平稳性可能导致这些模型的性能下降。

Wood, K.等人设计了一种含交叉注意力的时序趋势预测模型，通过将输入与市场环境变化的时间和幅度等指标相结合来判断市场所处的状态，同时利用交叉注意力来模拟不同资产和不同市场条件下的趋势，增强了预测模型的性能，提高了策略的风险收益比。本文将详细分析这一时序趋势预测模型，

1.1 金融市场通用模式的假设

通用模式（Universal Pattern）是一个跨学科的概念。Wood, K 认为，在金融时间序列分析和趋势预测的背景下，通用模式或市场的潜在规则，指在多个市场、不同时段、多种资产中普遍存在的、可识别的行为模式或趋势，其内容包含：

趋势跟随：价格在一段时间内上升或下降的趋势，在多种资产和市场中都可能出现。

均值回归：价格在偏离长期平均水平后往往会回归到平均水平的现象。

季节性模式：某些金融资产的价格或成交量会随季节变化而呈规律性波动，例如农产品价格受季节性因素影响。

市场情绪：投资者情绪可能导致市场出现过度反应或羊群行为。

波动率聚集：价格波动往往会在高波动期后跟随着高波动期，或在低波动期后跟随着低波动期，在多个市场中均有体现。

本文将要分析的时序趋势预测模型正是基于以上金融市场存在通用模式或潜在规则的假设。Wood, K. 将这种模式或规则称为“regime”。在假设模式存在的基础上，投资者可以通过模型来识别和利用这些通用模式或规则，从而在新的、未出现过的市场条件下进行有效的预测和交易决策，提高策略在不同市场环境下的泛化能力。

1.2 时序动量模型的演变

传统时序动量策略受限于交易拥挤表现下滑

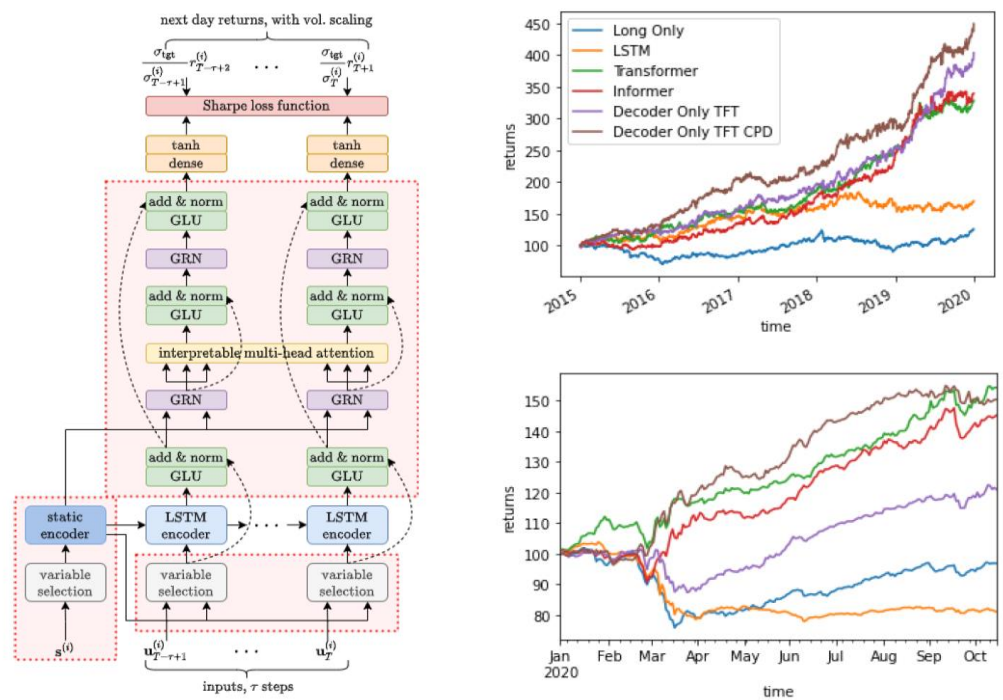
时序动量策略由来已久，又被称为趋势追踪策略。在管理期货策略（CTA）及投资组合管理中，趋势追踪是策略的主要构成部分。基础形式是根据过去一段时间的收益率来判断是否持有或卖出资产，在此基础上也有不少投资者使用调整后的技术指标代替收益率作为动量的判断依据，如 MACD 或 EWMA。尽管趋势追踪在历史上长期有效，但在市场环境发生明显变化时，动量策略会出现明显回撤，这种现象又被称为“动量崩溃”。

在对传统时间序列动量（TSMOM, Moskowitz, 2012）模型的研究中，Wood 发现其风险调整收益在 2018 年至 2023 年的时期较 1995 年至 2000 年的时期下降了 80%以上。这一现象与“因子拥挤”（Baltas, 2019）概念有关，即套利者在相似因子上交易相同的资产，导致市场效率下降和流动性驱动的尾部事件风险增加（Brown, 2022）。

深度动量网络的发展和演变

在应用动量模型时，我们经常会观察到不同的趋势和均值反转，可将其视为在多个时间尺度上同时出现的短期趋势集合。在此基础上，市场规则的变化更加大了多种交易信号融合的难度。深度动量网络（DMN）应运而生，引入深度神经网络来提升时序动量模型的风险调整收益。Wood, K 等人在 2020-2021 年先后发表了 3 篇文献，将深度神经网络用于动量与反转策略的优化，其所设计的深度动量网络如图 1 所示。在 Transformer 基础上，结合了变点检测（CPD）来调整策略，在期货市场超越了常规的 Transformer 及其衍生模型。

图1： Momentum Transformer 结构及深度动量网络的回测表现



资料来源：Wood, K. (2021)，浙商证券研究所

1.3 小样本学习

在深度学习领域中，样本是否被认为是一个小样本，与样本数量、特征维度、以及类别区分度有关。样本数量较少，或相较于特征维度 D 样本数量小于 $\log D$ ，或是某类别与其他类别之间区分度较大，均可视为小样本学习的场景。小样本和数据不平衡是共生问题并且普遍存在。

深度学习中的小样本学习为快速适应新数据提供了可能性，在计算机视觉领域已取得明显突破。在应用过程中，小样本学习通过训练神经网络来适应从最少的数据中学习。训练后的模型在未出现的图像类别上进行测试，或是解决不可见的强化学习环境，这对于系统性的交易策略研发有重要意义。将快速变化的市场规则作为系统性策略探索的外部环境，来学习资产价格变动的趋势。

2 核心结论

本文介绍了 Wood, K.等人提出的名为 **Cross Attentive Time-Series Trend Network**（简称 X-Trend）的新型时序预测模型，用于构造系统性交易策略。

该模型旨在解决传统预测模型难以快速适应金融市场条件变化的问题。X-Trend 通过在包含金融时间序列市场环境的上下文集上执行位置关注，从而在不同市场环境之间传输相似模式的趋势。具体而言，它能够快速适应新的金融市场环境，相对于传统的神经预测模型，夏普比率提高了 18.9%。

X-Trend 在 2018 年至 2023 年的市场波动期间表现出色，相较于传统的时间序列动量策略，其性能提高了 10 倍。尤其值得注意的是，X-Trend 在 2020 至 2021 年世界公共卫生事件期间，其回撤期恢复的速度是神经网络预测模型的两倍。

此外，X-Trend 还展现了在零次学习（zero-shot）情境下对新颖的、未包含在训练集中的金融资产进行仓位管理的能力。在相同时期，X-Trend 均优于基线模型，特别是在 2018 至 2023 年期间，X-Trend 的平均年化夏普比率相较于基准提高了 2 倍以上。

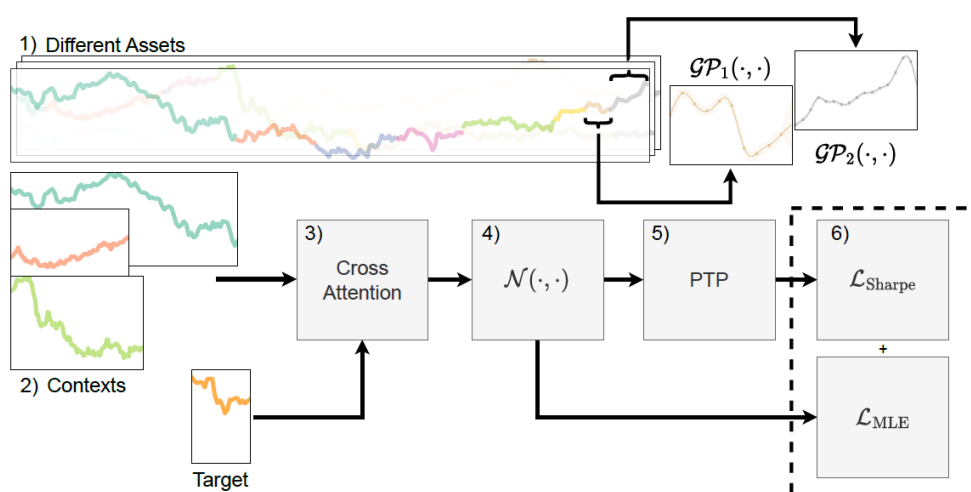
X-Trend 不仅可以预测隔夜价格，还能够生成交易信号。其交叉注意力机制使研究人员能够解释预测结果与上下文集中模式之间的关系。同时，通过辅助输出步骤输出预测，揭示了最优交易信号与预测收益率之间的关系。X-Trend 模型为交易策略的仓位管理提供了一种创新、可靠的方法。

总体而言，X-Trend 通过引入交叉注意力机制并利用上下文集中的信息，成功地提高了时序预测模型在金融市场中的适应性和风险收益性价比。

3 时序趋势预测模型 X-Trend

时序预测模型 X-Trend 的组成单元和工作原理如图 2 所示。

图2: Encoder-decoder 结构及信息传导示意图



资料来源: Wood, K. (2023), 浙商证券研究所

X-Trend 的工作原理可以概括为：任一资产的输入特征会经过相同的变点检测算法分为不同环境对应的区间，从每一种资产中进行随机抽样构造上下文集，并通过交叉注意力机制进行计算，结合未来的趋势变化优化对上下文集的信息提取，输出未来收益的概率分布，根据收益概率分布生成目标仓位。策略模型学习过程中，参数更新通过同时优化夏普比率损失和对数似然函数来实现。

3.1 X-Trend 模型设计

3.1.1 序列表征与基线模型

如何将金融资产的时序特征转化为高维数据特征，便于策略模型进行学习？X-Trend 模型借鉴了自然语言处理领域中的“文本概括”解决方案，即，将过去一段时间每一时刻/交易日的特征按时间顺序逐一“合并”，将时间维度包含的信息压缩在长数值向量中。

特征向量构造

特征向量包含两部分信息，价格趋势和资产类别。在任一时刻单个资产的特征是由收益率（1 天，1 个月，3 个月，6 个月，1 年）和 MACD 指标拼接而成的向量。其中，收益率需要根据回看时长进行调整：

$$\mathbf{x}_t^{(i)} = \text{Concat} \left(\left[\hat{r}_{t-t',t}^{(i)} \mid t' \in \{1, 21, 63, 126, 252\} \right], \left[\text{MACD} \left(p_{1:t}^{(i)}, S, L \right) \mid \forall (S, L) \right] \right)$$

$$\hat{r}_{t-t',t}^{(i)} = r_{t-t',t}^{(i)} / \sigma_t^{(i)} \sqrt{t'},$$

其中回看时长为 t' ， $\sigma_t^{(i)}$ 是资产 i 在 t 时刻的波动率， $r_{t-t',t}^{(i)}$ 是资产 i 在过去 t' 时间区间内的收益率。

此外还要将资产类别信息作为附带信息（side information）计入表征向量。不同大类资产的行情特征有明显的区别，例如原油期货和 5 年期国债，在同一时期的价格趋势大概率截然不同。因此，模型将这一类别信息通过映射学习转化为一个嵌入向量，和价格特征合并。

变量选择网络（VSN）

变量选择网络的作用是依据资产类别，权衡不同的滞后归一化回报和 MACD 特征。其工作原理如下：

$$\mathbf{w}_t = \text{Softmax} \circ \text{FFN}(\mathbf{x}_t, s)$$

$$\text{VSN}(\mathbf{x}_t) = \sum_{j=1}^{|\mathcal{X}|} w_{t,j} \text{FFN}_j(x_{t,j}),$$

编码解码构造

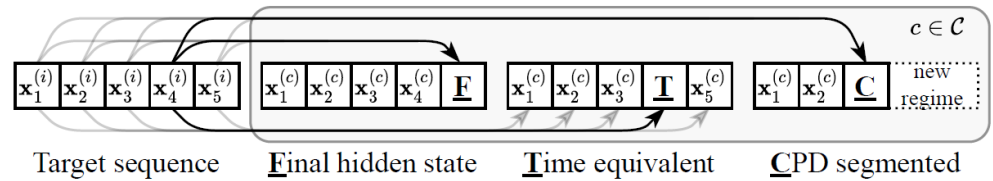
Wood 使用前期研究中的深度动量网络（Wood, K. 2021）作为基线模型，该模型只有解码器结构。而 X-Trend 模型则同时使用了编码器和解码器，并且在解码之前使用了交叉注意力机制。模型中对时间序列的表征计算过程如下：

$$\begin{aligned} \mathbf{x}'_t &= \text{LayerNorm} \circ \text{FFN}_1 \circ \text{Concat}(\text{VSN}(\mathbf{x}_t, s), \mathbf{y}_t), \\ (\mathbf{h}_t, \mathbf{c}_t) &= \text{LSTM}(\mathbf{x}'_t, \mathbf{h}_{t-1}, \mathbf{c}_{t-1}), \\ \mathbf{a}_t &= \text{LayerNorm}(\mathbf{h}_t + \mathbf{x}'_t) \\ \Xi_{\text{Dec}}(\mathbf{x}_{l:t}, s, \mathbf{y}_{l:t}) &= \text{LayerNorm}(\text{FFN}_2(\mathbf{a}_t, s) + \mathbf{a}_t). \end{aligned}$$

3.1.2 由“上下文语境”衍生的环境构建

根据历史表现判断当前市场环境处于何种状态，就是构建“上下文集”（context set）。判断出环境所处的状态，能更好地将特征和预测目标中的因果关系进行提取。在回测过程中，Wood 认为有 3 种不同方法来构建上下文集，其原理如图 3 所示。

图3：目标序列中的信息如何影响上下文集



资料来源：Wood, K. (2023)，浙商证券研究所

方法一：最终隐藏状态和固定长度的随机序列（Final hidden state）：在时间和资产上随机采样固定长度的随机序列，并在这些序列的最终隐藏状态为条件进行总结。

方法二：等时隐藏状态（Time equivalent）：随机采样与目标序列相同长度的随机上下文序列。对于目标序列中的每个时间步，根据等时隐藏状态进行条件，即第 k 个目标步骤与第 k 个上下文步骤相对应。这允许模型通过在目标序列的每个时间步上的不同表示进行调节来融合额外的相邻规则。

方法三：变点检测分段表示（CPD segmented）：使用高斯过程变点检测算法将上下文集合分割成不同的模式。随机采样变点分割序列，并以这些变点时间序列段的最终隐藏状态上为条件来构建上下文集环境。

3.1.3 交叉注意力机制的实现

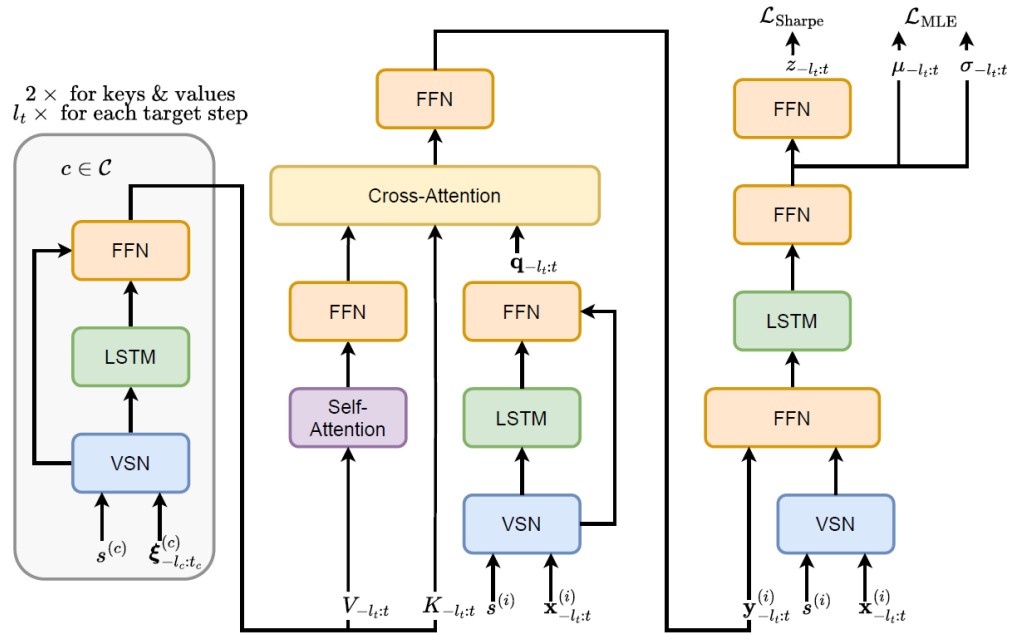
结合上下文集环境，X-Trend 使用了目标序列和上下文序列之间的交叉注意机制。这使得模型可以在不同时间和不同模式（regime）之间关注来自不同资产的不同序列，通过赋权实现。其基本原理是，与同一资产的近期历史相比，上下文集环境中包含了更广泛的信息（近期历史，长期历史，以及时列演变信息）。

在交叉注意力机制中，每个序列都被编码为一组查询（Query）、键（Key）和值（Value）。查询通常代表目标序列，而键和值来自上下文集/环境中的序列。具体计算流程可以参考图 4。

3.1.4 解码器与损失函数的设计

X-Trend 的解码器结构及交叉注意力的计算方法如图 4 所示。

图4: Encoder-decoder 结构及信息传导示意图



资料来源: Wood, K. (2023), 浙商证券研究所

上图中,除了常见的前馈神经网络(FFN)、长短期记忆神经网络(LSTM)、注意力机制等,还包含功能性网络,即变量选择网络(Variable Selection Network, VSN)。

目标函数的设计有 3 种选择:

(1) 夏普比率,与深度动量网络(Wood, K. 2021) 种 的设置一致:

$$\mathcal{L}_{\text{Sharpe}}^{\text{DMN}(\cdot)}(\theta) = -\sqrt{252} \frac{\text{mean}_{\Omega} \left[\frac{\sigma_{\text{tgt}}^{(i)}}{\sigma_t^{(i)}} r_{t+1}^{(i)} \text{DMN}(\mathbf{x}_{-l_t:t}^{(i)}) \right]}{\text{std}_{\Omega} \left[\frac{\sigma_{\text{tgt}}^{(i)}}{\sigma_t^{(i)}} r_{t+1}^{(i)} \text{DMN}(\mathbf{x}_{-l_t:t}^{(i)}) \right]},$$

(2) 极大似然估计:

$$\mathcal{L}_{\text{Joint}}^{\text{MLE}}(\theta) = \alpha \mathcal{L}_{\text{MLE}}(\theta) + \mathcal{L}_{\text{Sharpe}}^{\text{PTPQ}(\cdot)}(\theta),$$

(3) 分位数回归:

$$\mathcal{L}_{\text{Joint}}^{\text{QRE}}(\theta) = \alpha \mathcal{L}_{\text{QRE}}(\theta) + \mathcal{L}_{\text{Sharpe}}^{\text{PTPQ}(\cdot)}(\theta).$$

3.2 实战效果

在测试 X-Trend 的表现时, Wood, K. 等人使用了 Pinnacle Data Corp CLC 数据库中的 50 个流动性最好、最常见的连续期货合约作为投资组合,包含 1990 年至 2023 年初的历史数据。具体实验所使用的合约数会随着回测的需要进行调整。

训练和测试：采用了扩展窗口方法，初始训练集为 T 年至 T+5 年的数据，然后进行外样本测试，接着将训练窗口扩展至 T 年至 T+10 年进行再次测试，以此类推。样本外测试区间始终为紧跟训练集之后的 5 年。

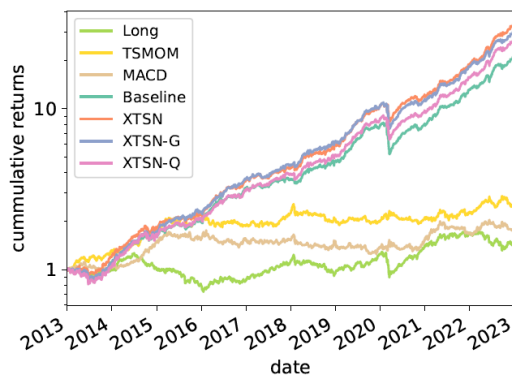
情景测试：针对 2020 年至 2022 年这段时间也进行了单独回测，因为这段时间的行情状态与训练集中的情况明显不同，其中 2020-2021 期间的公共卫生事件影响不可忽视。

回测实验包括零次学习实验（zero-shot setting）、少量学习实验（few-shot setting）、模块消融实验，以及可解释性分析。在进行模型比较时，使用夏普比率计算目标函数的模型是 X-Trend 的基础形式，使用高斯极大似然估计的模型记为 X-Trend-G，使用分位数回归作为目标函数的模型记为 X-Trend-Q。

3.2.1 Few-shot 实验

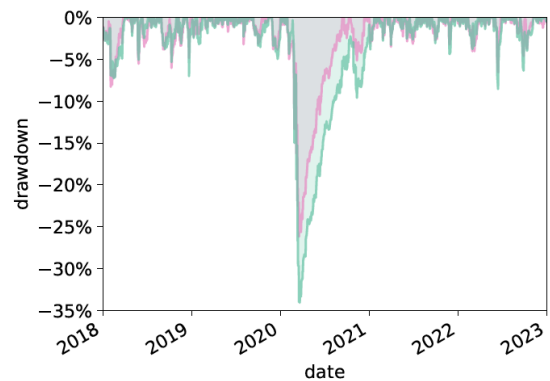
小样本学习场景中使用了样本池内全部 50 种连续期货合约进行回测。整体上，深度动量网络相较于使用传统趋势追踪策略有明显优势，并且 X-Trend 及其衍生模型相对于基准策略（Wood, K. 2021）的收益也有所提升。

图5：小样本情景下各策略的累计收益对比（2013/01-2023/01）



资料来源：Wood, K. (2023)，浙商证券研究所

图6：小样本情景下策略的最大回撤对比（2018/01-2023/01）



资料来源：Wood, K. (2023)，浙商证券研究所

表1：小样本情景下策略模型收益表现对比

策略模型	目标函数	区间年化夏普比率		
		2018-2023	2013-2023	1995-2023
买入持有	/	0.48	0.40	0.60
时序动量	/	0.23	0.71	1.01
MACD	/	0.27	0.45	0.71
基线模型 (Wook, K. 2021)	Sharpe	2.27	1.93	2.91
	J-Gauss	2.43	2.06	3.04
	J-QRE	2.26	1.96	2.89
X-Trend	Sharpe	2.65 (+16.9%)	2.17	3.17
	J-Gauss	2.52 (+11.0%)	2.10	3.05
	J-QRE	2.70 (+18.9%)	2.14	3.11

资料来源：Wood, K. 2023，浙商证券研究所（注：X-Trend 相关模型已进行超参数优化，取最好结果）

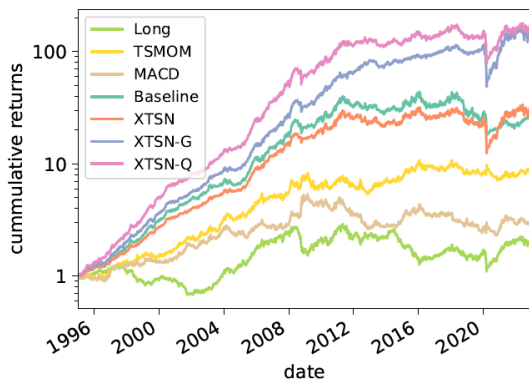
我们特别关注了 2018 年至 2023 年这段时间区间内的策略表现，因为它表现出显著的行情波动、前所未见的市场动态和许多制度性变化，包括 2018 至 2019 年的牛市，2020 至 2021 年的全球公共卫生事件，以及 2022 年的俄乌冲突。由图 6 可知，X-Trend 模型在这 5 年的期间的最大回撤明显小于基准，同时回撤修复的速度也更快。结合表 1，X-Trend 模型相较于基线策略的夏普比率提高了 16.9%，且使用分位数回归的 X-Trend-Q 夏普比率高于基准 18.9%。

交叉注意力机制是改善风险调整回报的主要推动因素。我们观察到 X-Trend-Q 和 X-Trend 胜过 X-Trend-G，这表明模型能够学习并受益于更复杂的回报分布，而不是一成不变的高斯回报分布。这可能是因为在返回分布的左（和右）尾部的分位数上执行分位数回归强制模型专注于大幅波动。

3.2.2 Zero-shot 实验

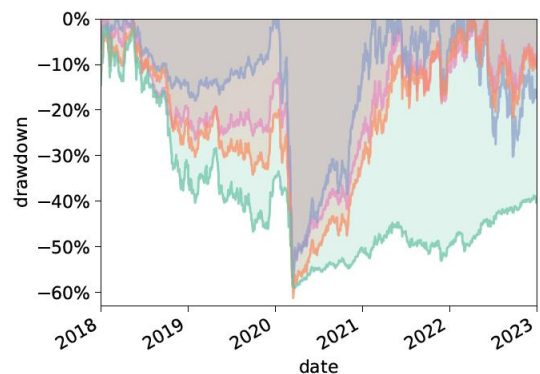
零样本学习情景下，测试所涉及的资产及其特征序列是策略模型训练时未曾学习过的。因此回测设置上述 50 种期货合约中的 30 种合约的历史数据作为训练集，另外 20 种期货合约进行样本外回测。与小样本学习情景不同，在该设置下，联合损失函数成为结果改进的主要驱动因素。

图7：零次学习情景下各策略收益对比（1995/01-2023/01）



资料来源：Wood, K. (2023)，浙商证券研究所

图8：零次学习情景下策略最大回撤对比（2018/01-2023/01）



资料来源：Wood, K. (2023)，浙商证券研究所

表2：零次学习情景下策略模型收益表现对比

策略模型	目标函数	区间年化夏普比率		
		2018-2023	2013-2023	1995-2023
买入持有	/	0.28	0.02	0.28
时序动量	/	-0.26	0.05	0.61
MACD	/	-0.14	0.11	0.32
基线模型 (Wook, K. 2021)	Sharpe	-0.11	0.02	1.00
	J-Gauss	0.14	0.19	1.25
	J-QRE	0.16	0.10	1.26
X-Trend	Sharpe	0.13	0.18	1.17
	J-Gauss	0.47	0.47	1.44
	J-QRE	0.12	0.18	1.27

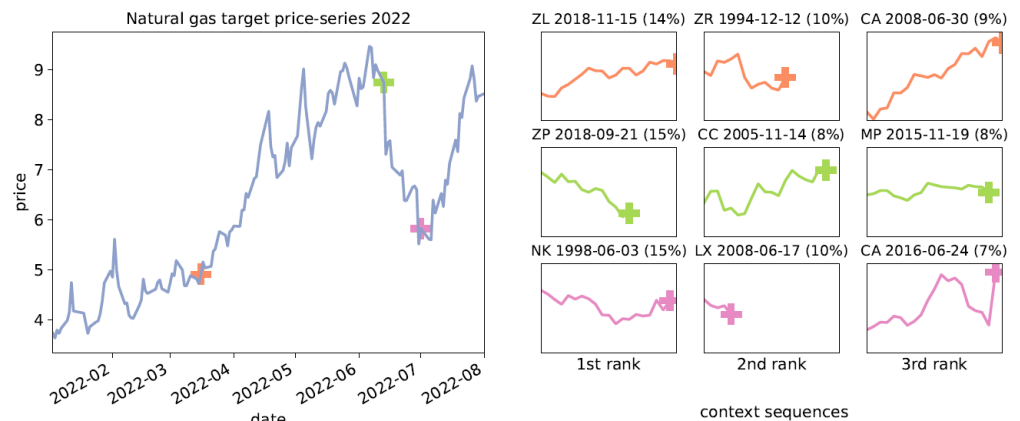
资料来源：Wood, K. 2023，浙商证券研究所（注：X-Trend 相关模型已进行超参数优化，取最好结果）

对比小样本学习情景，零次学习情景下，尽管模型优于传统趋势策略，但其收益表现出现明显下滑，尤其是在 2008 年之后，甚至部分策略模型在 2018 年之后持续处于亏损状态。这表明不同期货合约之间的趋势行情逐渐分化，表现出不同的周期和特征，因此零次学习得到的模型泛化效果较差。同时，也因为这种样本分布上的差别，使用高斯分布的极大似然估计作为目标函数得到的 X-Trend-G 模型表现最优。

3.2.3 模型可解释性分析

交叉注意力机制是否能让策略模型准确找到与环境状态（regime）关联的特征？为回答这一问题，Wood 对交叉注意力权重进行了可视化，选取了俄乌冲突期间 3 个不同的时间点（上行趋势开始，下行趋势开始，趋势反转）的天然气期货历史数据来进行说明，如图 9 所示。上下文集序列的筛选是通过变点检测（3.1.2 方法三）实现的。

图9：Encoder-decoder 结构及信息传导示意图



资料来源：Wood, K. (2023)，浙商证券研究所

交叉注意力的计算结果表明，策略在 2022 年 3 月天然气期货行情启动时，将标的资产与 2018 年 11 月的豆油期货（ZL）行情、1994 年 12 月的稻谷期货（ZR）行情、以及 2008 年的法国 CAC40 指数（CA）相关联从而做出决策，并且模型所搜索到的这些关联序列在历史上也符合价格上涨趋势开始这一特点。X-Trend 在天然气期货 2022 年 6 月（左图绿色标识）和 2022 年 7 月（左图紫色标识）时点也构建了对应的上下文集，并分别做出了下跌起始和趋势反转的判断。因此，实验结果表明，交叉注意力机制能够将来自上下文集中类似模式的信息，应用到不同资产或新的市场环境中。

4 总结

4.1 原文结论

X-Trend 策略模型是利用交叉注意力机制优化的时序趋势预测模型。它利用小样本学习和变点检测技术以适应新的市场环境。回测结果表明，X-Trend 在 2020-2021 全球公共卫生事件期间出现回撤的恢复速度是同类型趋势策略的两倍左右。在 2018 年至 2023 年期间，相较于基线模型，夏普比率提高了 18.9%；与传统时序动量策略相比，夏普比率提高了约 10 倍。

X-Trend 模型优势来源于交叉注意力机制，它可以将目标序列和从上下文集中的相似模式相关联，从而提高预测胜率。甚至在样本或计算资源缺乏的零次学习场景下，X-Trend 模型相较于深度动量网络基线模型和传统趋势追踪策略在风险收益比上均有优势。未来的研究方向是将这一框架应用于加密货币等新兴资产类别。

4.2 我们的思考

1. 使用深度神经网络构建系统性交易策略，考虑到实际样本数量限制、以及市场行情快速变化带来的样本差别较大，是典型的小样本学习场景。

2. 零次学习场景下，难以把握不同资产或不同时期的样本分布，因此使用高斯分布极大似然估计的效果反而更好。

3. 构造上下文集，并引入交叉注意力机制，是改善策略风险调整回报的主要推动因素，能够将来自上下文集中类似模式的信息，应用到不同资产或新的市场环境中，从而提高策略模型的胜率。

4. 交叉注意力机制对于策略输入端的启发：可以检测全市场不同类别资产，找到关联的历史信息或模式来优化策略；也可以在资产自身时序上使用自注意力机制来强化特征。

5 风险提示

本报告中包含公开发表的文献整理的模型结果，涉及的收益指标等结果的解释性请参考原始文献。回测结果是基于历史数据的统计归纳，模型力求跟踪市场规律和趋势，但仍存失效可能，不构成投资建议，须谨慎使用。

参考文献

- [1] Tobias J Moskowitz, Yao Hua Ooi, and Lasse Heje Pedersen. Time series momentum. Journal of financial economics, 104(2):228–250, 2012.
- [2] Nick Baltas. The impact of crowding in alternative risk premia investing. Financial Analysts Journal, 75(3):89–104, 2019.
- [3] Gregory W Brown, Philip Howard, and Christian T Lundblad. Crowded trades and tail risk. The Review of Financial Studies, 35(7):3231–3271, 2022.
- [4] Kieran Wood, Sven Giegerich, Stephen Roberts, and Stefan Zohren. Trading with the momentum transformer: An intelligent and interpretable architecture. arXiv:2112.08534, 2021.
- [5] Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. Deeptime: Deep time-index metalearning for non-stationary time-series forecasting. arXiv:2207.06046, 2022.
- [6] Tom Liu, Stephen Roberts, and Stefan Zohren. Deep inception networks: A general end-to-end framework for multi-asset quantitative strategies. arXiv:2307.05522, 2023.

股票投资评级说明

以报告日后的 6 个月内，证券相对于沪深 300 指数的涨跌幅为标准，定义如下：

1. 买入：相对于沪深 300 指数表现 + 20% 以上；
2. 增持：相对于沪深 300 指数表现 + 10% ~ + 20%；
3. 中性：相对于沪深 300 指数表现 - 10% ~ + 10% 之间波动；
4. 减持：相对于沪深 300 指数表现 - 10% 以下。

行业的投资评级：

以报告日后的 6 个月内，行业指数相对于沪深 300 指数的涨跌幅为标准，定义如下：

1. 看好：行业指数相对于沪深 300 指数表现 + 10% 以上；
2. 中性：行业指数相对于沪深 300 指数表现 - 10% ~ + 10% 以上；
3. 看淡：行业指数相对于沪深 300 指数表现 - 10% 以下。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重。

建议：投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者不应仅仅依靠投资评级来推断结论。

法律声明及风险提示

本报告由浙商证券股份有限公司（已具备中国证监会批复的证券投资咨询业务资格，经营许可证编号为：Z39833000）制作。本报告中的信息均来源于我们认为可靠的已公开资料，但浙商证券股份有限公司及其关联机构（以下统称“本公司”）对这些信息的真实性、准确性及完整性不作任何保证，也不保证所包含的信息和建议不发生任何变更。本公司没有将变更的信息和建议向报告所有接收者进行更新的义务。

本报告仅供本公司的客户作参考之用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本报告仅反映报告作者的出具日的观点和判断，在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议，投资者应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。

本公司的交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。本公司的资产管理公司、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

本报告版权均归本公司所有，未经本公司事先书面授权，任何机构或个人不得以任何形式复制、发布、传播本报告的全部或部分内容。经授权刊载、转发本报告或者摘要的，应当注明本报告发布人和发布日期，并提示使用本报告的风险。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

浙商证券研究所

上海总部地址：杨高南路 729 号陆家嘴世纪金融广场 1 号楼 25 层

北京地址：北京市东城区朝阳门北大街 8 号富华大厦 E 座 4 层

深圳地址：广东省深圳市福田区广电金融中心 33 层

上海总部邮政编码：200127

上海总部电话：(8621) 80108518

上海总部传真：(8621) 80106010

浙商证券研究所：<https://www.stocke.com.cn>