

从数据到均衡：AI 技术在大宗商品与金融市场中的实证验证路径

樊 滕

中共宜宾市南溪区委员会党校 四川宜宾 644601

摘 要：本文讲述了大宗商品以及金融市场数据的源头、类型、采集跟整合手段，存在数据清洗以及特征工程的技术。论述了机器学习、深度学习及自然语言处理于市场预测、时间序列剖析和市场情绪解析里的运用。规划出研究框架雏形，制订出研究假设，且搭建起针对样本选取的实验环境，开展对模型的训练、验证及调优。也对 AI 模型在大宗商品价格预测和金融市场波动掌握上的表现做了分析，还对 AI 在跨市场联动分析上的结果展开讨论。

关键词：人工智能；大宗商品市场；金融市场；数据预处理；实证验证

引言

伴随大数据与人工智能技术的迅猛发展，人工智能在金融市场的应用愈发广泛，尤以对大数据的处理分析与贵金属市场的量化交易，在其市场行情、价格预测、风险管理方面的问题越来越难以应对。而 AI 的机器学习、深度学习、自然语言理解等，由于其强大数据处理和精准预测优势，为解决问题，提出创新方法、策略提供了一条新路。

1. 数据获取与预处理

1.1 大宗商品市场数据来源与类型

大宗商品方面的数据来源主要有期货市场、官方统计机构、行业协会、卫星影像技术、第三方数据机构等。这些数据内容主要指的是现货、期货价格、实际库存量、进口 / 出口量、运输情况、气候影响数据等。以每天的价格变动、成交金额、持有量等为代表的结构化数据，其时间特征明显，用于建立 AI 模型。例如报纸的报道、行业报告、图片遥感等代表的非结构化数据，可以描述市场的情绪和基本面，提升对 AI 模型对市场的认知度。实际上我们需要根据各类产品（原油、农产品、工业品等）特性建立不同的数据输入规则，并考虑数据的时域、完整性、准确性，作为 AI 模型的支撑。

1.2 金融市场数据采集与整合

金融市场产生的不同类型金融数据一般来源于金融交易平台、金融数据分析数据库（如 Wind、Bloomberg）、中央银行公开数据、券商的分析数据库。除了价格和量价等交易数据之外，还包括公司的财务报表、政策公告、货币政策公告、宏观经济数据以及评级信息等。随着大数据时代的来

临，社交媒体、研究成果和分析员报告等非结构化数据也被加入分析数据之中。为了便于对多样产品、多样观点和不同时段资产管理，必须进行一系列数据清洗工作，如时间统一性、频率和格式规范化等。

1.3 数据清洗与特征工程方法

AI 模型是建立在对数据的清理和准备的基础上，这包括数据缺失问题的修补、异常数据的处理、数据重复及一致性、标准化处理等；而市场数据上所面临的一些典型难题比如停牌、暴跌、时滞等也需要重点关注。特征工程可以把原始数据转化为模型易读的高质量输入特征，包括使用滑动窗计算、使用滞后变量创造、取差分处理、波动率模型构建、技术指标（MACD、RSI）等等；而多元的时间序列数据构造变量交互特征，并使用 PCA 等降维方法。

2. AI 技术方法论

2.1 机器学习模型在市场预测中的应用

在金融预测领域中，因为机器学习的非线性建模方式，使得它们在不同场景中都有较高的使用度，对比参数化的传统方式，它可以更好地去把握和剖析复杂商品、资本市场中的大小关系。目前最为常见，且效果显著的预测工具有：随机森林、支持向量机（SVM）和梯度提升系列的算法（XGBoost、LightGBM）等。它们具备其在大规模特征空间、大类识别上的独特优势，可以用在市场趋势分析、价格变化的分析和应急反应等多个方面^[1]。其中最关键的是，使用特征重要性选择主要影响因素，同时交叉验证和提早停止法则更能增强模型的普适性。例如利用有宏观经济指数、存货数据、技

术分析参数在内的 XGBoost 模型,可以很好地对短期价格拐点进行预测。相对其他传统的预测方式,例如线性回归,机器学习对于市场的突变、不平稳序列更加敏感。但需注意,通过集成式学习策略(Bagging、Boosting)的应用,能更好地降低过拟合的可能,提升预测结果的稳定性。

2.2 深度学习模型与时间序列分析

除建模过程导致时序数据分析能力增强外,深度学习导致的特征获取更是一个革命性突破。例如,商业贸易和金融投资情景下 LSTM、GRU 和 Transformer 结构对于非平稳随机价格数据、多维规模效应及噪声扰动所造成之随机性均很善于应对,且超出一般时序分析方法。尤其是 LSTM 作为带记忆单元网络模型的代表,对于时长关系依赖有着优异的性能,有利于发掘连续和周期关系市场运动状态。不同于 LSTM 的阶层关系,代表多维信息传递的 Transformer 形式在并行处理和多维关系建模方面具有更广泛的应用前景,有潜力成为预测跨境资产联动性关系的有力工具。此外,借助图形神经网络可与卷积运算相结合,清晰地描述高频交易时的波动机理和市场各个维度之间的网络联系。

2.3 自然语言处理在市场情绪分析中的作用

市场定价是由期望推动的,期望驻留于语言与文本之中。利用自然语言处理(NLP)的方法,能够深入挖掘并量化情感,这对于 AI 模型的训练是非常有用的。在情绪词典法、情绪分类法和主题模型的帮助下,非结构化文本可以转化为结构性的特征变量。通过对与石油有关的新闻标题及推文进行分析可以提前预测市场对供给不足情绪的判断;从股票市场的交易层面来看,情绪指数和买卖量以及波动率之间的相关关系已经被验证了^[2]。除此之外,金融方面的特有 NLP 模型(如 FinBERT)可以解读行业术语、政策话术的变动,以提升市场行为预测清晰度;NLP 在扩展传统模型输入维度的同时,搭建起情绪、信息和价格三者间的连接桥梁,继而使市场模型在量和意两个层面上完成转换,这有助于市场模型实现“从数量到意义”的飞跃,是走向“从数量走向均衡”的重要一步。

3. 实证验证设计

3.1 研究框架与假设提出

实证研究的要义是设计周密的研究方案并提出可以证伪的研究假设。在大宗商品市场和金融市场上如何应用人工智能尚无普遍适用的理论,给出一个可能有用的研究框架,

即通常包含四个部分:数据采集和处理、模型的建立过程、预测试和检验评估。首先,要识别研究目标,如对某种商品未来价格的预测,对金融市场出现异常的表现的识别,对某种商品在两个及以上市场关联关系的建立。其次,在问题明确的情况下选择适当的应用 AI 进行建模的方法,并给出此技术适用条件下能够证明或否决的假设。例如假设 H1 为使用深度学习算法时模型的预测精度更高;假设 H2 为使用深度学习算法同时添加情感语料能够提高模型发现极端价格波动的能力。最后,考虑在模型的使用过程中因果推理、变量的滞后效应、异质市场结构等复杂因素,增加研究的说服力和实践价值。

3.2 样本选取与实验环境构建

就样本选择而言,需要选择具有及时性、完整性、实用性的样本。如果关注的是商品领域的话,往往关注交易量大、影响力强的商品(如原油、铁矿石、大豆、铜等),且需要长达五年的监测期间以捕捉周期轮回及其结构化特征;如果关注的是金融领域,可涵盖有广度的金融工具,比如股票指数、国家债券、货币对、交易型开放式商品基金等来进行不同的市场比较、联动性研究^[3]。就试验环境的建立而言,第一,使用真实的过往数据作为训练/检验的对象,让模型表现出贴近真实的市场行为,而不用给定一个(固定的)市场环境假设;第二,可以构建虚拟市场、回溯系统,让模型在不同的市场环境中接受检验。第三,为应对样本偏差和过度拟合,可以通过时间序列的交叉验证、滚动窗口学习、层次采样等方式,增强模型的通用性。第四,需要在数据处理能力强、模型计算速度快的分析平台上设计完备的实验体系,可借助 Python、R、Tensorflow、Pytorch 等进行人工智能试验环境平台搭建,形成规范、可重复、可拓展的应用环境。

3.3 模型训练、验证与调优策略

完成模型搭建的第二步,保质保量地进行模型训练和验证,也就是根据数据集选择合适的损失函数和优化策略进行参数拟合,其中对于时间序列的预测任务,损失函数多为均方误差(MSE)、平均绝对误差(MAE)等;对于类别型任务,如分类股票上涨还是跌落,可采用交叉熵损失函数等。不同的模型有其对应不同的调参方法,如 XGBoost、随机森林机器学习模型通过网格搜索(GridSearch)、贝叶斯优化(BayesianOptimization)方法寻优其超参数组合;基于深度学习提高模型鲁棒性的策略包括学习率衰减、批量归

一化、dropout 等。通过模型多维度指标如 R2、RMSE 平均绝对误差 (RootMeanSquareError)、MAPE 平均百分比误差 (AbsolutePercentError)、AUC 等以及预测走势图、残差图等来分析模型拟合情况。确定最终模型后,对最终模型进行稳健性检验和敏感性检验,检验出的模型是否可以适应不同市场环境。

4. 实证结果分析

4.1 AI 模型在大宗商品价格预测中的表现

从实证结果来看,人工智能技术方法对巨型产品价格的预测正确度和适应能力相比传统方法是大大提高的。对于原油和铁矿石来说,模型在不同周期的预测结果对比结果表明,AI 模型在对价格波动中非线性、迅速逆转和复杂季节等因素进行捕捉时更适应。对于存在巨大事件(如地缘冲突、OPEC 峰会)时出现的巨大的不确定环境来说,由于统计方法模型的误差会在不同的周期放大,因此 AI 模型利用模型的可自适特性对冲击具有一定的抵御能力^[4]。通过引入不同的信息源(如气候状况、港口交通状况、新闻媒体情感等)构造 AI 模型,可以有效对“结构性拐点”进行识别。AI 的模拟历史数据很好,并且能够应用于实时估计与趋势感知等应用方面。还可以通过模型的解释工具(SHAP 值等),研究人员发现驱动价格波动的基本因素,从而对政策分析、仓储管理、风险管理提供了解释和指南等。

4.2 AI 模型对金融市场波动的捕捉能力

金融市场的波动变化往往是因偶然的突发性情况、全局性危险对市场整体产生影响,因此,AI 模型对隐藏的市场风险有着更强的敏感性和准确度。实践比较发现,LSTM 和 Transformer 模型能在交易量、波动率、新闻情感等变化时提前预测市场的波动情况,其中,对于股票市场和债券市场而言,基于深度学习的方法都能提前感知到“恐慌”和“泡沫”信号,此外,站在风险管理的角度出发,AI 模型可以第一时间做出应激反应,在资产配置上发挥作用,例如,深度学习能够感知的关于系统未来发展的信号能够比实际波动提前 24 小时预警,以此来帮助投资者在市场波动前调整持仓结构以应对潜在的危机。此外,实验结果显示,AI 模型还可以通过建立宏观经济数据(CPI、利率决定)与市场波动间滞后响应的模型来验证这一结论。这都说明运用 AI 技术对波动进行分析的实用化运用,对于金融机构实现风险管理、提升监管前瞻性等方面都有着积极的价值体现。

4.3 跨市场联动的 AI 分析结果讨论

AI 模型另一优势在于其可以反映出各市场间的关联性与变化趋势,建立有关大宗商品、有关金融市场的结合机制的模型研究的结果显示:如油品价格变化会对能源类股、货币、通缩预测等多种变量有显著的领先性;农产品价格上升会因物价上涨而扰动债券市场的利率走向。但是,当需要考虑更多变量时,传统意义上的联结分析方法(例如协整检验、格兰杰因果检验)可能不适用,但是 AI 模型(如多输入多输出的 Transformer 模型构建方法)能够方便地进行各种市场、各种产品、各种时间段信息的处理,进行跨市场的价格、波动共同估计和测量^[5]。另外,对于跨市场的信息,如情绪信息与宏观事件的作用路径也已经能够通过模型建立并量化,例如,中国对于房地产政策变化的影响通过影响钢铁需求最终影响到全球铁矿石市场,再进而影响澳元汇率波动。人工智能带来的多市场研究能力上的突破和对未来趋势预测能力的提高,深入理解“全球经济一体化”真正内涵的能力,也将为全球资产配置及风险管理策略提供指引,并对政策决定产生重要影响。

结束语:综上所述,人工智能技术的应用可以对大量信息进行有效管理与精准解析,并借助其强大的机器学习与深度学习能力对市场价格变化作出精准判断,经过实践观测,其对大宗商品的价格预报和金融市场的价格震荡的控制皆能起到显著的作用,为投资人员与决策人员提供有力保障。随着 AI 技术的应用不断优化升级与大数据功能不断完善,人工智能将会在未来的大宗商品与金融市场研究中发挥越来越重要的作用。

参考文献:

- [1] 林化琛. 区块链与物联网技术在大宗商品管理中的运用探究[J]. 中国新通信, 2024, 26 (22): 53-55+112.
- [2] 于小林. AI 技术在金融市场的应用研究[J]. 网络安全和信息化, 2024, (08): 20-22.
- [3] 纪亮. 精准计量 匠心测试 构建大宗商品先进计量体系——访舟山市质量技术监督检测研究院院长李存军[J]. 中国计量, 2024, (04): 18-21.
- [4] 范玉坤, 吴尚睿, 高晓晓. 信息技术与大宗商品全生命周期检验的融合[J]. 质量与认证, 2023, (11): 57-59.
- [5] 徐鹏. 信息技术背景下大宗商品企业内部控制审计的创新[J]. 活力, 2023, (05): 97-99.