

Research on the Application of Artificial Intelligence Technology in Quantitative Stock Selection Models

Junru Wang

Jiangsu Branch of CITIC Securities Co., Ltd., Nanjing, Jiangsu, 210000, China

Abstract

As the financial market becomes increasingly complex, artificial intelligence technology brings new impetus to quantitative stock selection models. This paper analyzes the application framework of artificial intelligence in quantitative stock selection, examines the limitations of traditional quantitative stock selection methods, and explores the necessity of adopting artificial intelligence technology. The research mainly elaborates on the specific applications of machine learning, deep learning, and other technologies in factor discovery and model construction, and constructs a complete system design and backtesting evaluation system. Regarding the backtesting of the model, it details the main links such as system construction, performance evaluation, and overfitting detection, demonstrating that artificial intelligence technology can significantly improve the predictive accuracy and market adaptability of quantitative stock selection models. However, this technology also faces challenges such as data quality, model overfitting, and interpretability. It is necessary to strike a balance between technological innovation and risk control to promote the development of intelligent quantitative investment.

Keywords

Artificial Intelligence; Quantitative Stock Selection; Model Backtesting; Overfitting; Interpretability

人工智能技术在量化选股模型中的应用研究

王君铷

中信证券股份有限公司江苏分公司, 中国·江苏·南京 210000

摘要

金融市场不断复杂化的时候, 人工智能技术给量化选股模型带来新的动力。本文剖析了人工智能在量化选股中的应用框架, 分析了传统的量化选股方式的局限, 并且对采用人工智能技术的必要性展开了研究, 研究主要讲述了机器学习、深度学习等技术因子发掘, 模型构建等方面的具体应用, 构建了完整的系统设计和回测评价体系, 对于模型的回测情况, 详细叙述了系统搭建, 绩效评价, 过拟合检测这些主要环节, 显示人工智能技术可以大大改善量化选股模型的预测准确性和市场适应性, 不过该技术也存在数据质量, 模型过拟合, 可解释性等困难, 要在技术创新和风险控制之间找到平衡点, 以促进智能量化投资的发展。

关键词

人工智能; 量化选股; 模型回测; 过拟合; 可解释性

1 引言

量化选股属于现代投资决策的关键方法, 依靠数学模型和计算机技术达成投资的系统化与纪律化。近年来, 金融市场呈现出高维非线性特点, 传统量化选股方法应对复杂市场环境时渐渐显出不足, 人工智能技术凭借其强大的模式识别与预测能力, 给量化选股赋予了新的技术途径。本文希望系统考察人工智能技术在量化选股中的应用, 从系统构建, 模型回测到风险评估形成一个完整的体系, 通过剖析关键技术模块, 回测评估办法以及实际应用的困难, 给相关研究提

供理论参照和实践指引。研究显示, 人工智能技术不但可提升信息处理效率, 而且可加强模型对市场动态的适应能力, 不过其应用依旧存在数据质量, 模型稳健性和可解释性等问题。

2 量化选股模型概述

2.1 量化选股的基本概念

量化选股属于利用数学模型与计算机技术展开投资决策的一种方法, 其本质就是借助系统性的分析架构, 把投资理念转化为可量化的规则体系。这种方法侧重于投资的纪律性和系统性, 努力排除主观情绪等人为因素对投资决策的影响。量化选股模型一般建立在现代金融理论的基础之上, 包括资产定价理论、投资组合理论等等。模型经由对历史数

【作者简介】王君铷(1993-), 女, 中国江苏南京人, 硕士, 从事投资分析研究。

据做回溯测试，来验证策略是否有效，然后在实际交易里持续优化，一套完备的量化选股系统包含很多部分，比如获得数据、建立因子、改进组合和把控风险这些。从方法论角度来看，量化选股强调科学的实证精神。每一个投资决策背后都要有具体的数量化依据，这些依据要经受得住严格统计检验。这样科学的投资方法可以提升投资过程的可复制性和可验证性，给投资成绩持续稳定赋予制度保障。

2.2 传统量化选股方法

传统的量化选股方法主要是依靠线性模型和统计分析的方法。多因子模型是最具代表性的一种框架，其主要目的是寻找出与股票收益率有着稳定的相关性的特征变量从而形成投资组合。这些因子通常包括价值、成长、动量、质量等不同的维度。传统的处理市场数据的方法，通常基于一些简化假设来完成工作，例如收益率服从正态分布以及因子之间是线性关系。这样的方法虽然简单、容易理解，但却在把握市场复杂特性上存在着很大的欠缺，市场情况发生改变的时候，传统的模型就会显得应对不起来。另外，传统的量化选股方法对于因子的挖掘和选择主要依靠研究人员的先验知识以及经验判断。通过人工驱动的因子发现效率很低，无法充分利用市场上海量的信息。随着市场的有效性提高，通过传统的办法获取超额收益也越来越困难。

2.3 人工智能引入的必要性

金融市场本质上就是一个复杂的适应系统，它的运行方式存在非线性、高维以及动态变化等特性。这些特点让传统量化方法遭遇越来越多的考验，人工智能技术被引进来之后就给了这些问题新的应对办法。第一，人工智能技术能够很好地处理非线性问题。深度学习等算法可以对数据中复杂的模式进行自动学习，而不需要事先确定具体的函数形式。这种灵活性可以更好的适应市场的实际情况。第二，人工智能技术具有很强的特征抽取能力。采用神经网络等模型可以从原始数据中自动提取出有预测能力的特征，提高了信息利用的能力。这样的能力对现代金融市场中海量的数据的处理就更为重要。

3 人工智能量化选股的系统构建

3.1 系统架构设计

人工智能量化选股系统的架构设计是整个模型构建的基础，一般采用分层模块化的思路。系统自下而上可以分为数据层、因子层、模型层、组合层、交易层。数据层负责多源异构数据的采集与存储，行情数据、基本面数据、另类数据等，进行数据清洗、对齐等预处理操作。因子层基于原始数据构建特征因子，传统量化因子与 AI 自动挖掘的因子在此融合。模型层是系统的核心，用机器学习、深度学习等算法训练预测模型，输出股票收益率的预测结果，组合层依照预测结果做投资权重的分配，控制风险暴露，交易层负责执行具体的交易指令，还要考虑滑点、手续费等市场摩擦因素，

整个系统靠模块化设计，功能解耦，方便以后迭代和维护，而且系统要能处理高并发、低延迟，满足实时交易的需求。

3.2 关键技术模块

人工智能量化选股系统的关键技术模块包含特征工程、模型训练与优化算法，特征工程模块旨在从原始数据中提取具备预测能力的有效特征，除了传统的量价，基本面因子之外，更加重视借助深度学习（比如自动编码器）从高维数据中自动提取抽象特征，或者利用自然语言处理技术（如情感分析，主题建模）从新闻，研报等文本数据中生成另类因子，模型训练模块是技术的核心，根据不同的场景选择合适的算法，比如，截面预测时，可以选择梯度提升决策树（如 XGBoost, LightGBM），时序预测时，可以选择循环神经网络（RNN, LSTM）或者时序卷积网络（TCN）。优化算法模块负责模型超参数的调整（网格搜索、贝叶斯优化）和投资组合权重的优化（均值-方差模型、风险平价模型），提升模型的性能和控制下行风险。

3.3 人工智能系统建立的特点

建立人工智能量化选股系统存在几个核心特性，即数据推动，自动学习以及模型自适应。该系统依靠海量历史数据创建起来，要凭借高质量的数据预处理和特征工程给予模型学习有效的输入。其主要优点就是通过机器学习算法自行挖掘因子并找出复杂模式，削减对人工先验知识的倚重。而且，模型具有动态顺应市场变动的能力，能够按照新数据不断改良调整，进而改善预测的准确性以及策略的稳定性。不过，系统创建碰上诸多难题，譬如模型复杂度大，可解释性差等，所以要兼顾自动化和风险控制。

4 模型回测与评估

4.1 回测框架的搭建

回测框架是检验量化选股策略历史表现的关键工具，核心目的就是尽可能真实地模拟策略在历史中的执行状况，一个严谨的回测框架需包含几个重要组成部分：历史数据池、策略逻辑模块、交易模拟器和绩效分析器，搭建时务必注意避免前视偏差，保证在任何时刻 t ，模型只能利用 t 时刻及之前的信息实施预测和决策，交易模拟器得精确考虑交易成本（佣金、印花税等）、市场冲击成本和滑点，以便接近真实交易环境，回测框架还要支持不同频率的回测（日频、分钟频），并且可以灵活设定初始资金、仓位限制、再平衡周期等参数，稳固的回测框架是策略迈向实盘前的第一道关键检验，其设计的科学性和严谨性直接影响评估结果的可信度。

4.2 绩效评估指标

回测完成后，要用一套全面的绩效评估指标来衡量策略的优劣，这些指标要从收益、风险、风险调整后收益等方面去考察，常用的收益指标有年化收益率、累计收益率；风险指标有年化波动率、最大回撤（反映极端风险）、下行波

动率；风险调整后收益指标最经典的是夏普比率（单位总风险带来的超额收益）、索提诺比率（单位下行风险带来的超额收益）和卡玛比率（年化收益与最大回撤之比）。信息比率反映相对于基准（如市场指数）的超额收益的稳定性，对于 Alpha 策略，还需关注 Alpha 收益、Beta 值和跟踪误差，除了这些定量指标，也要分析策略收益的分布（偏度、峰度）以及在不同市场环境（牛市、熊市）下的表现，来判断其适应性和稳定性。

4.3 过拟合与稳健性检验

金融数据信噪比较低且具有非平稳性，量化模型很容易发生过拟合，即在历史数据上表现很好但在实盘（样本外）表现很差。防范过拟合与稳健性检验很重要。防范过拟合的第一步是样本外测试，即将数据严格划分为训练集、验证集、测试集，策略的最终表现以在测试集（即“未知”数据）上的表现为准。交叉验证在时序数据中需要使用时序交叉验证等特殊方法。稳健性检验则包括：参数敏感性分析，观察策略的绩效在关键参数略作变动时是否稳定；子样本分析，将整个回测期细分为多个时期，如不同年份，从而判定策略在不同时段是否保持相仿的绩效表现，也可采用添加噪声数据或者拔靴法（Bootstrap）抽取样本的方法，以此测试策略的抗干扰能力。通过这些检验，可以增强对模型未来表现的信心。

5 挑战与未来展望

5.1 当前面临的主要挑战

人工智能赋能量化选股的确拥有大大的可能性，不过在投入实际使用的时候，依然碰上了相当多的严峻挑战，首当其冲的是数据质量挑战，数据的准确与否，完备情况，一致性差异，又以及另类数据来源的可靠性等方面存有困难，模型容易过拟合，金融市场的演变没有止境，当年的走势可能不会重现，使模型在外部样本的预测水准大幅降低。模型过于复杂并不易解释，并且二者存在矛盾之处，深度学习之类“黑箱”模型的确预测能力很强，不过其决策逻辑却难以明白，这种情形增加了风险归因以及监管合规的困难，其次，市场流动性，交易成本这些实际存在的约束往往无法彻底达成，在模型当中也许造成收益的损失。最后，策略同质化风险也值得警惕，类似算法和数据的广泛使用可能导致因子失效或加剧市场波动。

5.2 可解释性与风险控制

增强模型的可解释性是提升其实用性和可信度的关键。为此，可解释人工智能（XAI）技术被引入，比如借助 SHAP、LIME 之类的工具来解释复杂模型的预测结果，找出对决策有关键影响的因子，这有益于研究人员认识模型逻

辑并执行风险归因，在风险控制方面，AI 量化模型要创建多层次的风控体系，模型层面，可以采用集成学习，加入正则化等方法来改善泛化能力，交易层面，设置严格的仓位控制，止损止盈规则，组合层面，利用风险模型（如 Barra）来监控并控制对各类风险因子的暴露，还要持续监控模型的性能衰减，创建模型失效的预警和切换机制，将 AI 模型的动态适应能力与严格的风控纪律相结合。

5.3 未来发展趋势

展望未来，人工智能量化选股研究将出现如下趋势，一方面，技术融合会加深，图神经网络（GNN）被用来分析上市公司关联、供应链等复杂关系网络，强化学习在动态资产配置和交易执行策略优化方面的应用会更加成熟，Transformer 等大模型架构处理多模态金融数据的能力会得到更多探索。另一方面，对模型可解释性、稳健性和因果关系的追求会更加强烈，因果推断技术的引入可能会使模型超越相关关系，探寻更稳定的驱动因子。而且，面对越来越复杂的模型，高效、自动化的机器学习（AutoML）平台会成为投研流程的标准配置，以提升研发效率，技术创新和风险管理的有效平衡，将是推动 AI 量化投资持续健康发展的关键。

6 结语

本文系统探讨了人工智能技术在量化选股中的应用，搭建起从系统设计，模型回测到风险评估的完整研究框架。研究表明，人工智能技术依靠机器学习，深度学习等方法，大大改善了量化选股模型在因子挖掘，股价预测，风险控制等方面的表现。不过数据质量，过拟合风险，模型可解释性等仍属于实际应用的主要难题。未来的研究重点应放在可解释人工智能，因果推断等前沿技术同量化投资的深入融合上，还要在模型复杂性同实用性之间达成平衡。随着算法创新与风险管理体系的不断完善，人工智能技术有望推动量化选股向更智能化、稳健化的方向发展，但需始终重视技术创新与风险控制的协同并进。

参考文献

- [1] 熊俊.基于投资者情绪的机器学习选股策略研究[D].中南财经政法大学,2024.
- [2] 安仲文,吴琦.公募基金投研差异化势在必行人工智能开辟新路径[N].证券时报,2024-06-03(A06).
- [3] 李杨.基于图神经网络的量化选股模型研究[D].浙江大学,2024.
- [4] 柯嘉晨.基于信息论方法的智能量化策略研究[D].东南大学,2023.
- [5] 左梦瑶.基于机器学习的因子选股与量化择时策略研究[D].广州大学,2023.