

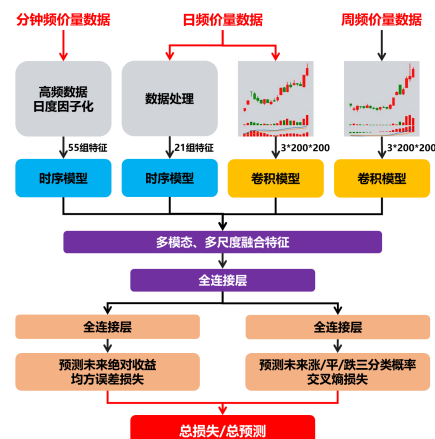
股价预测之多模态多尺度

深度学习研究报告

报告摘要:

- **AI看图:** 在去年发布的《基于卷积神经网络的股价走势AI识别与分类》研究报告中,创新性地采用基于深度学习的图像识别技术,将价量数据图表与未来股价走势进行建模,以实现股票价格预测。
- **股价预测之多模态多尺度:** 本文以“多模态、多尺度”为题,基于AI看图初版模型进行了大幅改进,新的模型结构如右图1所示。本文模型在日度价量数据图表的基础上,加入了高频因子数据、日频时序数据、周度价量数据图表,采用4个独立的深度时序模型和深度卷积模型进行多模态、多尺度的特征提取,并同时采用回归损失和分类损失以端到端的方式进行模型训练,有效提升了模型对未来股价的预测能力,取得了更为显著的超额收益。
- **对比提升:** 以2020/01/01~2024/10/31作为样本外回测区间,每20个交易日进行换仓,本文模型预测结果在全市场、沪深300、中证500、中证800、中证1000、国证2000、创业板的RankIC均值分别为8.7%、7.9%、6.6%、6.9%、8.2%、8.7%、10.4%,RankIC胜率分别为86.7%、69.0%、73.5%、75.2%、84.8%、86.1%、89.2%。以全市场作为回测统计口径,对比AI看图初始版本,本文模型的RankIC均值提升了3.0%,RankIC胜率提升了7.8%。
- **超额表现:** 以2020/01/01~2024/10/31作为样本外回测区间,基于本文模型因子在全市场、沪深300、中证500、中证800、中证1000、国证2000、创业板分别构建10分档多头股票组合,双边千三计费后,多头股票组合分别相对对应板块指数取得了12.97%、9.17%、5.30%、8.38%、7.47%、7.47%、11.52%的超额年化收益率。
- **与Barra风格因子相关性:** 整体而言,本文模型因子与Barra风格因子的相关性较低,其中,相关性最高的三个Barra风格因子为流动性因子、波动率因子和市值因子,相关系数分别为-18%、-16%和-8%。
- **展望:** 本报告以研发为主要目的,对模型进行了严谨的训练、验证、以及回测。其中,训练样本为2008~2016年数据,验证样本为2017~2019年数据,回测样本为2020~2024年数据。从理论上来说,通过加入更新、更多的训练样本,结合滚动训练等方式,能够进一步提升模型对未来股价的预测能力,取得更为显著的超额收益。
- **风险提示:** (1) 本文所述模型用量化方法通过历史数据统计、建模和测算完成,所得结论与规律在市场政策、环境变化时存在失效风险;(2) 本文策略在市场结构及交易行为改变时有可能存在失效风险;(3) 因量化模型不同,本文提出的观点可能与其他量化模型结论存在差异。

图 1: 多模态多尺度股价预测模型



数据来源: 广发证券发展研究中心

分析师:

安宁宁



SAC 执证号: S0260512020003

SFC CE No. BNW179



0755-23948352



anningning@gf.com.cn

分析师:

陈原文



SAC 执证号: S0260517080003



0755-82797057



chenyuanwen@gf.com.cn

请注意, 陈原文并非香港证券及期货事务监察委员会的注册持牌人, 不可在香港从事受监管活动。

相关研究:

基于卷积神经网络的股价走势 AI 识别与分类 2023-04-06

基于深度学习的高频数据因子挖掘 2023-08-11

联系人:

林涛



0755 - 82528531



gflintao@gf.com.cn

目录索引

一、AI 看图	5
二、多模态多尺度股价预测模型	6
(一) 模型结构	6
(二) 改进思路	7
三、实证分析	8
(一) 数据说明	8
(二) 对比提升	8
(三) 本文模型因子与 BARRA 风格因子相关性	8
(三) 实证分析 (全市场)	9
(四) 实证分析 (沪深 300)	11
(五) 实证分析 (中证 500)	13
(六) 实证分析 (中证 800)	15
(七) 实证分析 (中证 1000)	17
(八) 实证分析 (国证 2000)	19
(九) 实证表现 (创业板)	21
四、卷积模型与时序模型	23
(一) 循环神经网络	23
(二) TRANSFORMER 模型	23
(三) 卷积神经网络	24
五、总结与展望	28
六、风险提示	29

图表索引

图 1: AI 看图——基于卷积神经网络的股价走势 AI 识别与分类	5
图 2: 多模态多尺度股价预测模型	6
图 3: 10 分档表现 (全市场)	9
图 4: RankIC 表现 (全市场)	9
图 5: 多头及超额表现 (全市场)	9
图 6: 10 分档表现 (沪深 300)	11
图 7: RankIC 表现 (沪深 300)	11
图 8: 多头及超额表现 (沪深 300)	11
图 9: 10 分档表现 (中证 500)	13
图 10: RankIC 表现 (中证 500)	13
图 11: 多头及超额表现 (中证 500)	13
图 12: 10 分档表现 (中证 800)	15
图 13: RankIC 表现 (中证 800)	15
图 14: 多头及超额表现 (中证 800)	15
图 15: 10 分档表现 (中证 1000)	17
图 16: RankIC 表现 (中证 1000)	17
图 17: 多头及超额表现 (中证 1000)	17
图 18: 10 分档表现 (国证 2000)	19
图 19: RankIC 表现 (国证 2000)	19
图 20: 多头及超额表现 (国证 2000)	19
图 21: 10 分档表现 (创业板)	21
图 22: RankIC 表现 (创业板)	21
图 23: 多头及超额表现 (创业板)	21
图 24: 循环神经网络结构	23
图 25: Transformer 模型结构	24
图 26: LeNet-5 网络结构图	25
图 27: VGG16 网络结构图	25
图 28: 卷积 (Convolution) 运算示意图	26
图 29: 填充 (Padding) 示意图	26
图 30: 步幅 (Stride) 为 2 的卷积运算示意图	27
图 31: 池化 (Pooling) 运算示意图	28
表 1: 本文模型因子和 Barra 风格因子之间的相关性	8
表 2: 分年度收益统计 (全市场)	10
表 3: 分年度收益统计 (沪深 300)	12
表 4: 分年度收益统计 (中证 500)	14
表 5: 分年度收益统计 (中证 800)	16
表 6: 分年度收益统计 (中证 1000)	18
表 7: 分年度收益统计 (国证 2000)	20

表 8: 分年度收益统计 (创业板) 22

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/585003031043012001>