

MATLAB EXPO

从线性模型到深度学习

——和MATLAB共同见证股票收益预测的进化

冯佳睿, 海通证券股份有限公司



概要

线性模型：股票收益预测的起点

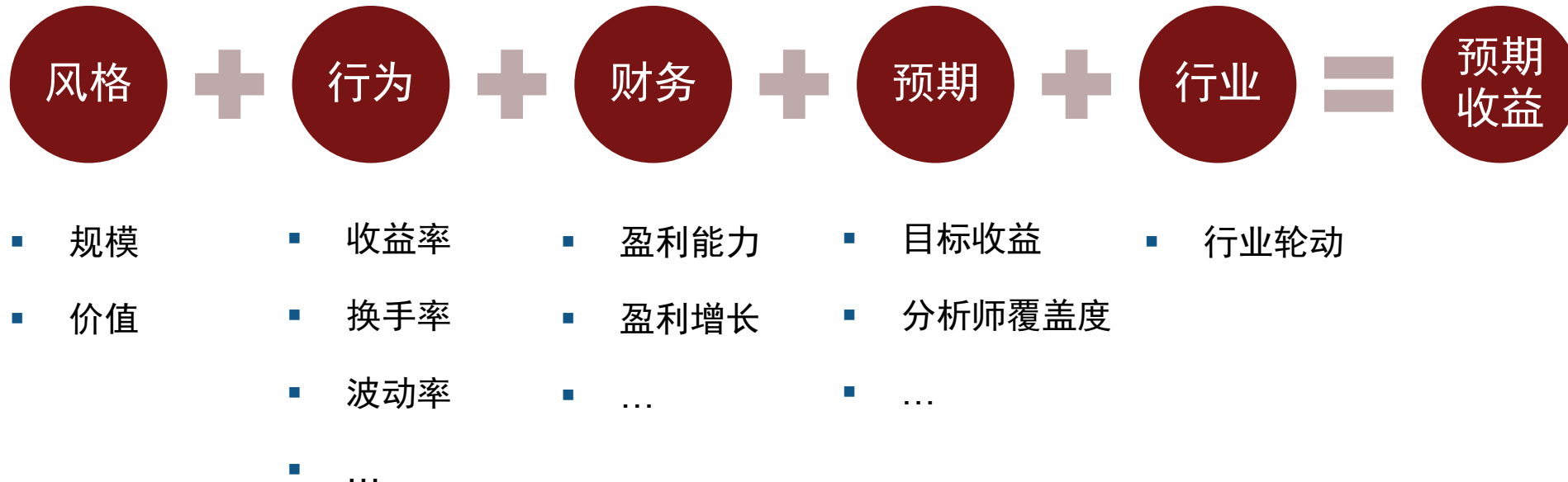
深度学习：股票收益预测的进化

线性模型+深度学习：股票收益预测的融合

线性模型：用一系列个股指标预测未来收益

r_i : 第*i*个股票的收益, F_i : 第*i*个股票的一系列指标

$$F_i \longrightarrow \mu_i = E(r_i)$$



线性模型： Fama-MacBeth横截面回归估计因子溢价

- 每个月进行一次横截面回归
- 回归因变量为所有个股t月的收益率，自变量为个股t-1月末的因子值，回归系数即为因子t月的因子溢价

$$r_{i,t} = c_t + \beta_t \cdot F_{t-1} + \varepsilon_{i,t}$$

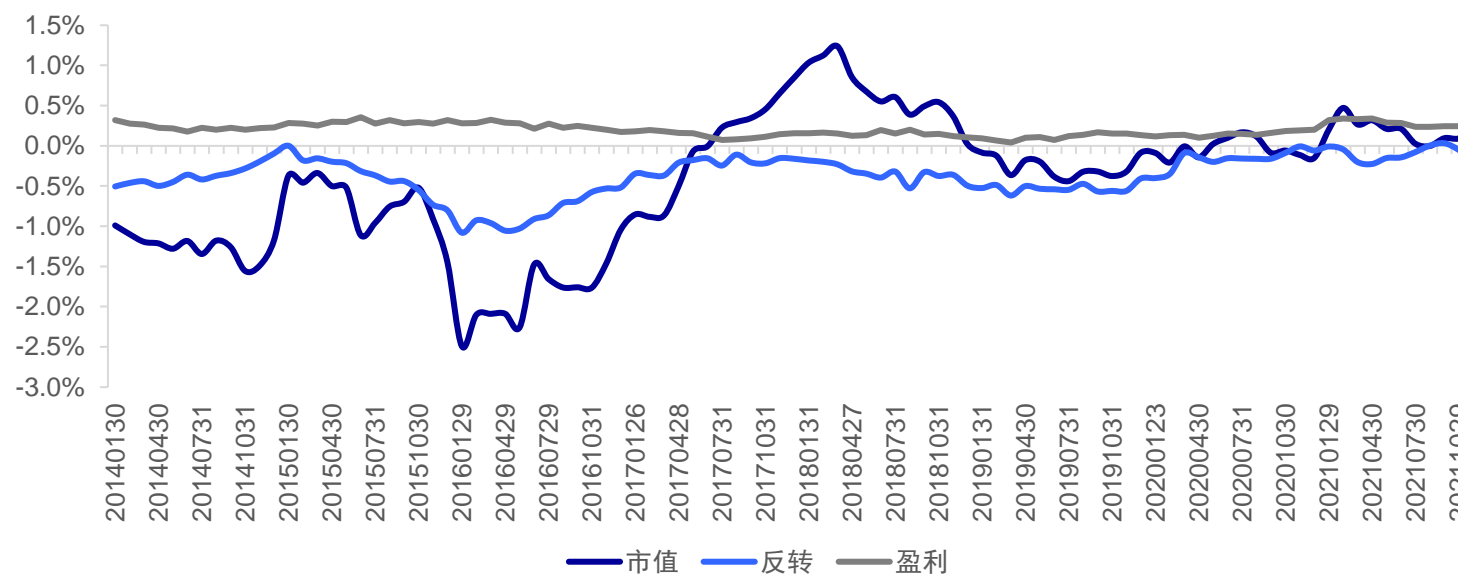
- 利用估计的因子溢价，预测下一个月的股票收益

$$\mu_i = F_{i,t} \cdot \widehat{\beta}_{t+1}$$

- **Statistics and Machine Learning Toolbox™** : regress

线性模型：因子溢价预测

- $t+1$ 月的因子溢价，通常采用 $(t-12+1)$ 至 t 月的因子溢价均值估计



线性模型：收益因子与风险因子的区分

- 能够在横截面上解释股票收益波动的因子，可以视作风险因子
- 在时间序列上溢价稳健且逻辑可靠的风险因子，可以视作收益来源
- **Statistics and Machine Learning Toolbox™** : ttest

| | 溢价>0比例 | 显著比例 | 月均溢价 | 月波动率 | 信息比率 |
|----|--------|------|--------|-------|-------|
| 盈利 | 0.68 | 0.69 | 0.44% | 0.79% | 1.94 |
| 估值 | 0.37 | 0.73 | -0.23% | 0.85% | -0.93 |

线性模型： A股市场上溢价显著异于0的常见收益因子

| | | 年溢价 | 月胜率 | 年波动率 | 信息比 | T值 |
|----|---------|--------|--------|-------|-------|-------|
| 风格 | 市值 | -6.33% | 59.09% | 7.37% | -0.86 | -2.31 |
| | 非线性市值 | 5.42% | 68.18% | 3.50% | 1.55 | 4.51 |
| | 预期EP | 0.69% | 45.45% | 3.06% | 0.23 | 0.74 |
| 行为 | 反转 | -5.05% | 70.45% | 2.61% | -1.94 | -5.54 |
| | 波动率 | -6.98% | 73.86% | 2.77% | -2.52 | -7.59 |
| | 换手率 | -4.14% | 61.36% | 5.15% | -0.80 | -2.36 |
| 财务 | ROE | 2.19% | 62.50% | 2.09% | 1.05 | 2.93 |
| | SUE | 3.48% | 84.09% | 1.06% | 3.28 | 8.99 |
| 预期 | 预期净利润调整 | 2.32% | 77.27% | 1.17% | 1.98 | 5.99 |
| | 分析师覆盖度 | 2.82% | 60.23% | 2.90% | 0.97 | 2.93 |
| | 分析师推荐 | 1.94% | 70.45% | 1.51% | 1.28 | 3.61 |

线性模型：模型好坏的评价

- IC (RankIC)：预测值和实际值的相关系数（秩相关系数）
- 模型预测IC、RankIC显著为正，月胜率为88.6%
- 预期收益最高的100只股票，相对市场等权组合月均超额2.35%
- **Statistics and Machine Learning Toolbox™** : corr

| | IC | RankIC | 多头收益 | 空头收益 | 多空收益 |
|-----|--------|--------|--------|--------|--------|
| 月均值 | 11.35% | 13.96% | 2.35% | 2.87% | 5.22% |
| 月胜率 | 88.64% | 88.64% | 80.68% | 87.50% | 88.64% |
| 信息比 | 3.40 | 4.07 | 2.34 | 2.50 | 2.85 |

概要

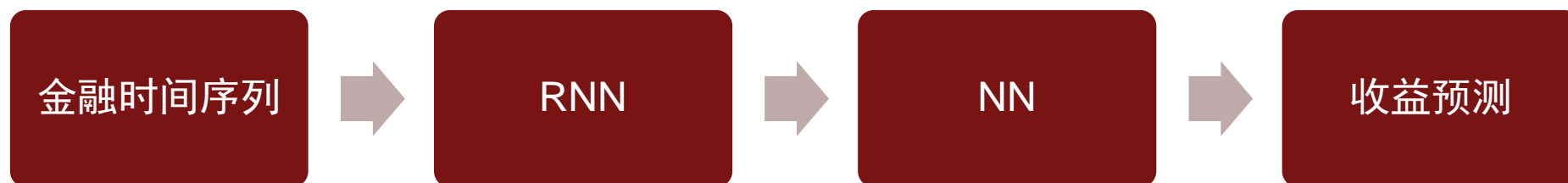
线性模型：股票收益预测的起点

深度学习：股票收益预测的进化

线性模型+深度学习：股票收益预测的融合

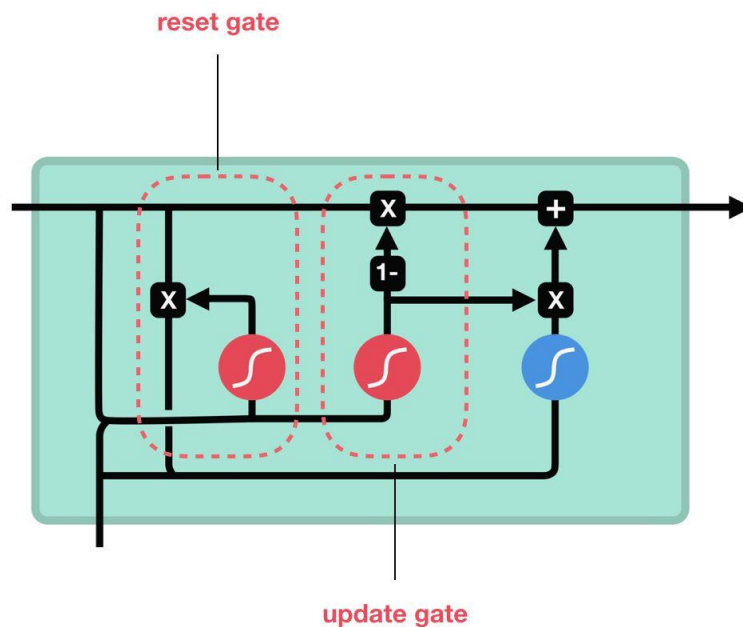
深度学习：股票收益预测流程

- 模型框架：循环神经网络（RNN）+全连接神经网络（NN）
- 金融数据的特征和RNN模型较为适配
- RNN提取金融时间序列中的信息，NN将RNN处理后的信息转化为收益预测



深度学习：GRU模型

- 结构简单，由重置门和更新门两部分组成
- 参数少，计算量小，训练速度快



深度学习：特征输入

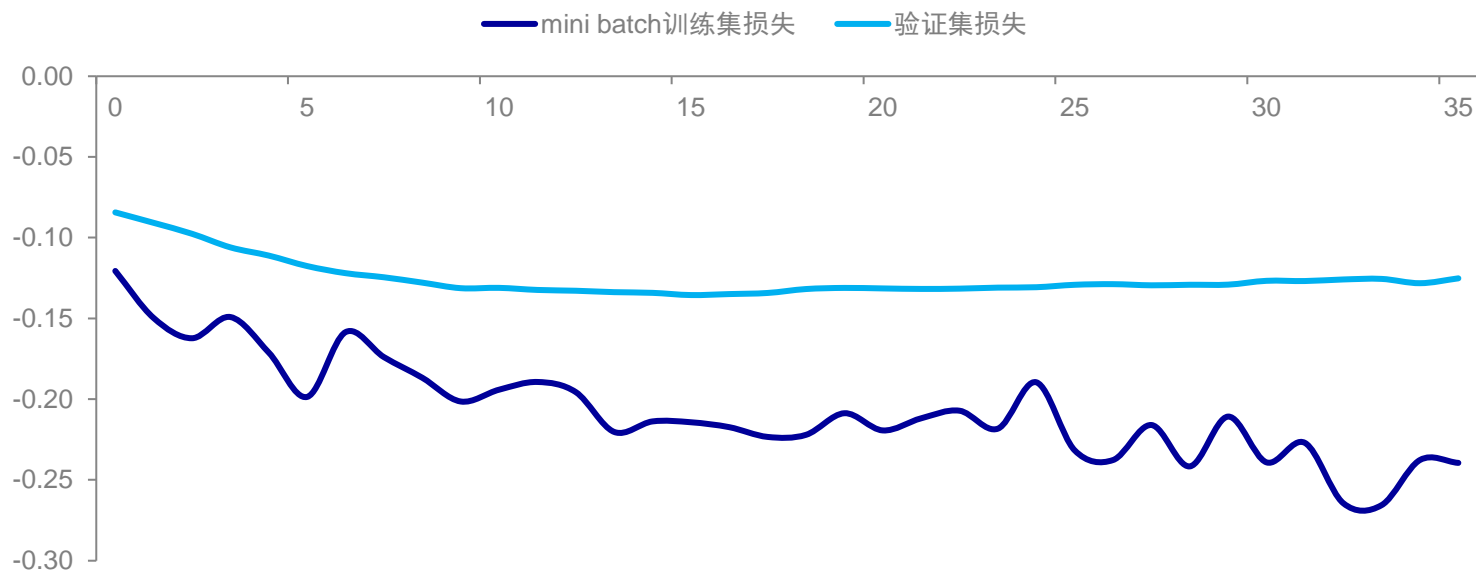
- 基于股票过去20个交易日的分钟K线、3秒TICK盘口、0.01秒逐笔成交数据，构建30分钟频率的指标序列
 - 分钟K线：收益、收益波动、收益偏度、成交额（占比）、成交笔数（占比）等
 - 3秒TICK盘口：委买增额、委卖增额、净委买增额等
 - 0.01秒逐笔成交：主动买入/卖出、净主买、买入意愿、大单买入/卖出、大单净买入等
- 共计164个特征（X），每次训练，特征输入为 $N*160*164$ 的tensor

深度学习：参数设定

- 预测目标（Y）：未来5个交易日的预期收益
- 使用过去120个交易日的数据训练模型，前100个为训练集，后20个为验证集
- 每5个交易日重新训练，每次训练5个模型，将5个模型的预测均值输出
- 随机失活：20%；激活函数：ReLU
- 损失函数： $-1 * IC$ ；优化器方法：ADAM；学习率：0.001
- 以mini-batch的方式训练，每个batch的量为1000
- RNN层数：[1, 2]；隐含状态大小：[10, 50]
- NN层数：[1]；神经元数量：[10]
- **Deep Learning Toolbox™** : grulayer

深度学习：早停是防止过拟合的重要技术

- 当验证集上的损失未随训练次数的增多而继续下降时，可考虑停止训练



深度学习：与常见因子的相关性在-0.2~0.2之内

| | GRU(10,1)+NN(10) | GRU(50,1)+NN(10) | GRU(10,2)+NN(10) | GRU(50,2)+NN(10) |
|------------|------------------|------------------|------------------|------------------|
| 市值 | -0.07 | -0.05 | -0.04 | -0.04 |
| 中盘 | -0.09 | -0.09 | -0.09 | -0.09 |
| 估值 | 0.04 | 0.04 | 0.03 | 0.05 |
| 换手率 | -0.18 | -0.15 | -0.17 | -0.15 |
| 改进反转 | -0.17 | -0.15 | -0.17 | -0.15 |
| 市场波动占比 | 0.07 | 0.08 | 0.07 | 0.07 |
| 盈利 | 0.00 | 0.00 | 0.01 | 0.00 |
| 盈利增长 | 0.00 | 0.01 | 0.02 | 0.02 |
| 分析师推荐 | 0.01 | 0.01 | 0.03 | 0.02 |
| 尾盘成交占比 | -0.02 | -0.03 | -0.03 | -0.03 |
| 开盘后买入意愿强度 | 0.10 | 0.11 | 0.12 | 0.11 |
| 开盘后大单净买入占比 | 0.08 | 0.09 | 0.10 | 0.10 |

深度学习：DL因子的选股有效性检验

- 周度多头收益0.4%左右，IC胜率超90%，但空头效应更明显

| | IC均值 | ICIR | 周度胜率 | 周均多空收益 | 周均多头超额 | 周均空头超额 |
|------------------|-------|------|------|--------|--------|--------|
| GRU(10,1)+NN(10) | 0.065 | 8.58 | 91% | 1.23% | 0.36% | -0.87% |
| GRU(50,1)+NN(10) | 0.067 | 9.17 | 90% | 1.32% | 0.42% | -0.89% |
| GRU(10,2)+NN(10) | 0.073 | 9.00 | 91% | 1.42% | 0.45% | -0.97% |
| GRU(50,2)+NN(10) | 0.072 | 9.38 | 92% | 1.41% | 0.46% | -0.96% |

深度学习：DL因子的选股有效性检验

| | GRU(10,1)+NN(10) | | GRU(50,1)+NN(10) | | GRU(10,2)+NN(10) | | GRU(50,2)+NN(10) | |
|------|------------------|--------|------------------|--------|------------------|--------|------------------|--------|
| | 多头超额 | 多空超额 | 多头超额 | 多空超额 | 多头超额 | 多空超额 | 多头超额 | 多空超额 |
| 2014 | 16.4% | 58.6% | 24.0% | 65.8% | 31.2% | 70.8% | 35.0% | 76.4% |
| 2015 | 68.5% | 144.3% | 85.3% | 166.5% | 76.4% | 162.2% | 102.2% | 185.4% |
| 2016 | 23.3% | 51.9% | 29.2% | 57.6% | 29.2% | 61.8% | 32.0% | 64.7% |
| 2017 | 15.3% | 41.9% | 13.9% | 39.6% | 16.2% | 46.1% | 18.3% | 47.2% |
| 2018 | 11.9% | 40.2% | 12.1% | 41.7% | 15.6% | 46.8% | 14.0% | 44.5% |
| 2019 | 17.9% | 66.5% | 21.8% | 74.4% | 27.8% | 82.6% | 28.1% | 81.7% |
| 2020 | 13.9% | 54.9% | 15.0% | 57.5% | 20.0% | 63.1% | 5.7% | 50.4% |
| 2021 | 16.3% | 46.5% | 22.0% | 50.0% | 13.4% | 46.0% | 14.2% | 45.2% |
| 全区间 | 20.9% | 59.7% | 24.9% | 64.6% | 26.3% | 68.7% | 26.9% | 68.9% |

概要

线性模型：股票收益预测的起点

深度学习：股票收益预测的进化

线性模型+深度学习：股票收益预测的融合

线性模型+深度学习：DL因子加入多因子模型

- 收益预测

$$\mu_i = F_{i,t} \cdot \widehat{\beta}_{t+1} + F_{DL,i,t} \cdot \widehat{\beta}_{DL,t+1}$$

- 目标函数

最大化组合预期收益 $\max_w w' \mu$ 。

- 约束条件

- 个股权重偏离、组合风格暴露
- 行业、市值中性、换手率上限

- 求解最优化问题，得到股票权重

- Optimization Toolbox™** : fmincon/linprog

$$\max_w w' \mu$$

$$\text{s.t. } w' e = 1, l \leq w \leq u$$

$$|(w - w_B)' F| \leq x$$

$$(w - w_B)' \mathbb{I}_{\{i \in D\}} = 0$$

$$\sum_{i=1}^N |w_{i,t} - w_{i,t-1}| \leq y$$

线性模型+深度学习：加入DL因子的中证500指数增强

| 周度单边 换手约束 | 2015-2021 | | | | |
|--------------|-----------|------------------|------------------|------------------|------------------|
| | 基础模型 | GRU(10,1)+NN(10) | GRU(50,1)+NN(10) | GRU(10,2)+NN(10) | GRU(50,2)+NN(10) |
| 30% | 26.6% | 28.6% | 29.3% | 32.2% | 29.4% |
| 40% | 26.8% | 29.5% | 28.0% | 32.6% | 33.9% |
| 50% | 27.0% | 28.8% | 27.7% | 31.2% | 32.8% |
| | 2021-2021 | | | | |
| 30% | 10.5% | 21.5% | 12.8% | 12.9% | 16.4% |
| 40% | 11.4% | 21.2% | 13.6% | 18.9% | 25.5% |
| 50% | 12.2% | 21.4% | 11.6% | 17.5% | 25.7% |

总结

- 股票收益预测技术的革新力撑，也是一部计算工具和计算能力的进化史，我们有幸和MATLAB共同成长

MATLAB EXPO

Thank you



© 2022 The MathWorks, Inc. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See [mathworks.com/trademarks](https://www.mathworks.com/trademarks) for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.