

# 一种基于全连接神经网络的 A 股投资策略的研究与实现

## 1. 策略概述

在 2013 年至 2019 年期间，我们基于全 A 股票，采用滚动训练全连接神经网络模型的方式，即按照一定的频率更新模型，来对接下来一段时间的股票表现进行实时预测，进而动态调整持仓。在因子筛选阶段，我们综合 IC 分数与主成分分析（PCA）筛选有效因子，达到冗余因子去除与降维的目的，以提升模型预测表现。模型表现与回测结果表明，经过因子筛选与全连接神经网络模型预测，策略能够持续跑赢大盘，经过对冲后能够为投资者持续创造利润。

## 2. 策略流程

在 2013 年至 2019 年期间，每隔 `update_years` 年，我们运行以下步骤：

1. 【计算因子 IC 序列】提取全 A 股票所有因子历史 `train_years` 年的数据，计算每个因子的 IC 序列。第  $t$  期因子  $i$  的 IC 定义为

$$IC_{i,t} = spearman\_corr(factor_{i,t}, ret(reallocate\_days)_t)$$

其中， $factor_{i,t}$  表示第  $t$  期所有股票在因子  $i$  的暴露， $ret(reallocate\_days)_t$  表示第  $t$  期所有股票未来 `reallocate_days` 个交易日相对基准的超额收益率，`spearman_corr` 表示 spearman 相关系数。将每个因子每一期的 IC 计算后得到每个因子的 IC 序列。

2. 【计算因子 IC 分数】因子  $i$  的 IC 分数定义为

$$ICScore_i = w * rank(abs(mean(IC_i))) + (1 - w) * rev\_rank(std(IC_i))$$

其中， $IC_i$  表示因子  $i$  的 IC 序列，`mean` 表示取均值，`std` 表示取标准差，`abs` 表示取绝对值，`rank` 表示升序排名，`rev_rank` 表示降序排名， $w$  表示 IC 均值所占权重。该公式表示：当 IC 均值绝对值较大，IC 标准差较小时，因子有效性较高，IC 分数也较高。IC 分数综合考虑了 IC 均值（预测能力）与 IC 标准差（稳定性），具有较好的参考意义。

3. 【构造数据集】筛选 IC 分数最高的前 `factor_num` 个因子作为机器学习模型特征，未

来 reallocate\_days 个交易日相对基准的超额收益排名前 30% 的样本作为正例 1，排名后 30% 的样本作为负例 0。将历史 train\_years 年的数据作为训练集，未来 update\_years 年的数据作为测试集。

4. 【数据预处理】利用去空值、中位数去极值、标准化、主成分分析对数据进行预处理。

5. 【模型搭建】全连接神经网络结构如下：

层	神经元个数	参数个数	激活函数
输入层	50	0	linear
隐藏层_1	40	2040	relu
隐藏层_2	20	820	relu
输出层	2	42	softmax

我们选择 adam 与二分类交叉熵作为网络的优化器与损失函数。

6. 【模型训练】设定训练代数为 5，并在测试集上验证模型表现。

7. 【实时预测】每隔 reallocate\_days 个交易日，利用训练好的模型预测所有股票的上涨概率，筛选出上涨概率最高的 max\_hold\_num 支股票，将上涨概率占比作为持仓权重持有 reallocate\_days 个交易日。

### 3. 参数设定

策略参数设定如下：

参数名称	参数值	参数说明
train_years	2	训练数据长度（年）
update_years	1	模型更新频率（年）
reallocate_days	10	调仓频率（天）
max_hold_num	20	最大持仓个数
w	0.7	IC 均值所占权重
factor_num	50	筛选出的因子个数
min_non_nan	0.3	去除空值时最小非空值比例
min_sample	0.5	去除空值后最小样本比例

## 4. 回测结果



## 5. 策略创新点

策略创新点如下：

1. 提出 IC 分数概念。在因子筛选阶段，我们综合考虑 IC 序列均值与标准差，筛选出 IC 均值绝对值大（预测能力强）与 IC 标准差小（稳定性高）的因子。
2. 滚动训练模型。由于金融数据具有时间序列特征，因此需要经常更新模型，以捕捉市场变化。我们按照固定频率更新模型，而非使用一个模型，解决了这一问题。

## 6. 参考文献

- [1]. 李聪. 基于 BP 神经网络的股票指数期货价格预测[D]. 青岛大学.
- [2]. 吴微, 陈维强, 刘波. 用 BP 神经网络预测股票市场涨跌[J]. 大连理工大学学报, 2001, 041(001):9-15.
- [3]. 王莎. BP 神经网络在股票预测中的应用研究[D]. 中南大学.
- [4]. 杨一文, 刘贵忠. 基于神经网络的多变量时间序列预测及其在股市中的应用[J]. 信息与控制, 2001, 30(5):413-417.
- [5]. 禹建丽, 孙增圻, Kroumov V, et al. 基于 BP 神经网络的股市建模与决策[J]. 系统工程理论与实践, 2003, 23(5).
- [6]. Kaastra, I, Boyd, M. Designing a neural network for forecasting financial and economic time series[J]. Neurocomputing, 1996.
- [7]. Yun Z, Quan Z, Caixin S, et al. RBF Neural Network and ANFIS-Based Short-Term Load Forecasting Approach in Real-Time Price Environment[J]. IEEE Transactions on Power

Systems, 2008, 23(3):853-858.

[8]. V. Kodogiannis, A. Lolis. Forecasting Financial Time Series using Neural Network and Fuzzy System-based Techniques[J]. Neural Computing & Applications, 11(2):90-102.