

遗传算法在 A 股投资组合优化中的应用

陈子璐 卫莉菲

郑州科技学院, 河南省郑州市, 450064;

摘要: 针对 A 股市场“政策驱动明显、行业轮动频繁、散户占比高”的特征, 以及传统马科维茨均值 - 方差模型在风险估计的局限, 本文将遗传算法引入投资组合优化研究。选取沪深 300 指数成分中的 10 只行业龙头股票为研究对象, 通过 Python 获取 2024 年的数据作为训练集与 2025 年 1 月 1 日—9 月 30 日作为测试集的收盘价数据, 计算月度收益率与方差, 量化收益与风险。构建包含选择 (锦标赛选择)、交叉 (模拟二进制交叉)、变异 (多项式变异) 算子的遗传算法框架, 设计最小风险、最大收益、均衡配置三类适应度函数, 通过 100 代种群迭代求解最优资产权重, 且与随机生成的 100 组权重作对比。实验结果显示: 在 2025 年测试集中, 采用遗传算法得到的最小风险优化权重较随机生成权重, 有 98% 概率获得更高收益; 最大收益优化权重有 79% 概率获得更高收益, 可满足不同风险偏好投资者需求。研究表明, 遗传算法能有效适配 A 股市场特性, 为投资者提供科学的投资组合配置依据, 提升投资决策的合理性与有效性。

关键词: 遗传算法; 投资组合优化; 沪深 300 指数

DOI: 10.64216/3080-1486.26.01.019

引言

A 股市场具有“政策驱动明显、行业轮动频繁、散户占比高”的特征^[1], 投资者往往因过度集中单一行业或盲目追逐热点导致收益波动剧烈。传统投资组合理论强调“分散化降低非系统性风险”, 但在 A 股市场中, 如何平衡行业配置与个股选择仍是实践难点^[2]。随着量化投资技术的发展, 结合行业均衡与智能算法的优化模型逐渐成为解决这一问题的有效工具^[3]。

马科维茨于 1952 年提出的均值 - 方差模型奠定了投资组合优化的理论基础, 其核心是在给定收益约束下最小化风险^[4], 数学表达式为:

$$\min \sigma^2(\omega) = \omega^T \Sigma \omega, \text{ s.t. } \omega^T \mu \geq \mu_0, \sum_{i=1}^n \omega_i = 1 \quad (1)$$

其中, ω 为资产权重向量, Σ 为收益率协方差矩阵, μ 为预期收益率向量, ω_i 为第 i 个投资权重, μ_i 为第 i 个投资收益率, μ_0 为投资者预设的目标收益率。

然而, 该模型存在两大局限: 一是假设收益率服从正态分布, 与 A 股市场“尖峰厚尾”的收益特征不符, 导致风险估计偏差; 二是在高维资产空间中, 目标函数易出现多局部最优, 传统梯度下降法难以找到全局最优解。

遗传算法 (Genetic Algorithm, GA) 作为源于生物进化理论的启发式数值优化算法, 通过模拟“自然选择”与“基因遗传”过程, 能有效处理非凸、多约束的

复杂优化问题^[5]。其核心优势在于: 无需依赖目标函数的连续性与可微性, 通过种群迭代实现全局搜索, 尤其适用于 A 股市场行业波动差异大、资产相关性动态变化的优化场景。

本文通过选取 A 股沪深 300 指数成分中的 10 只行业龙头股票为研究对象, 采用遗传算法构建投资组合优化模型, 为投资者提供投资建议。

1 遗传算法的数学建模

1.1 遗传算法框架

遗传算法的本质是通过“种群迭代—适应度评估—遗传操作”的循环, 逼近全局最优解, 其数学流程可描述为:

$$P_{t+1} = \text{Selection}(P_t) \circ \text{Crossover}(P_t) \circ \text{Mutation}(P_t) \quad (2)$$

其中, P_t 为第 t 代种群, \circ 表示遗传操作的复合运算, Selection、Crossover、Mutation 分别为选择、交叉、变异算子。

投资组合中第 i 只股票的权重为 ω_i , 10 只股票的权重向量为 $\omega = (\omega_1, \omega_2, \dots, \omega_{10})$, 满足 $0 \leq \omega_i \leq 1$ (不允许卖空)。初始种群 P_0 由 100 个随机生成的个体组成, 个体生成服从均匀分布:

$$\omega_i \sim U(0,1), i = 1,2,\dots,10 \quad (3)$$

为满足权重和为 1 的约束, 定义权重归一化函数:

$$\omega_{\text{norm},i} = \frac{\omega_i}{\sum_{j=1}^{10} \omega_j} \quad (4)$$

1.2 适应度函数设计

适应度函数是遗传算法的“导向标”，需将投资组合的风险—收益目标转化为可最大化的数学表达式。针对三类优化目标，设计带惩罚项的适应度函数。

(1) 最小风险目标

风险最低但是收益率不低于均值，适应度函数为：

$$f_{\min-risk}(\omega) = \mu(\omega) - \lambda_1 \cdot \sigma^2(\omega) \quad (5)$$

其中投资组合风险用 $\sigma^2(\omega) = \omega^T \Sigma \omega$ 量化， $\mu(\omega) = \omega^T \mu$ 为组合预期收益率， $\lambda_1 = 10^4$ 为风险惩罚系数（放大风险对适应度的负向影响）

(2) 最大收益目标

最大化组合收益率，容忍合理风险，适应度函数为：

$$f_{\max-risk}(\omega) = \mu(\omega) - \lambda_2 \cdot \sigma^2(\omega) \quad (6)$$

其中， $\lambda_2 = 10^{-4}$ 为风险惩罚系数，仅对极端高风险组合施加轻微惩罚。

(3) 均衡配置目标

本文的目标收益率 μ_0 为 10 只股票的平均预期收益率，即

$$\mu_0 = \frac{1}{10} \sum_{i=1}^{10} \mu_i \quad (7)$$

最小化风险并对低于目标的收益施加惩罚，适应度函数为：

$$f_{\text{balance}}(\omega) = -\sigma^2(\omega) - \lambda_3 \cdot \max(0, \mu_0 - \mu(\omega)) \quad (8)$$

其中， $\lambda_3 = 10^4$ 为收益惩罚系数， $\max(0, \mu_0 - \mu(\omega))$ 确保收益不低于目标值；负号将“最小化风险”转化为“最大化负风险”，符合遗传算法的最大化目标。

1.3 遗传算法过程

(1) 选择算子

采用锦标赛选择^[6]，随机选取 3 个个体竞争，选择适应度最高者进入下一代：

$$\text{Selection}(P_t) = \{\arg \max_{x \in S} f(x) | S \subset P_t, |S| = 3\} \quad (9)$$

其中 $f(x)$ 为适应度函数，该算子提升优质个体的留存概率，保证种群进化方向。

(2) 交叉算子

采用模拟二进制交叉（SBX），对父代个体 $x = (x_1, x_2, \dots, x_{10})$ ， $y = (y_1, y_2, \dots, y_{10})$ ，

生成子代 x' ， y' ：

$$\begin{aligned} x'_i &= 0.5[(1 + \beta)x_i + (1 - \beta)y_i] \\ y'_i &= 0.5[(1 - \beta)x_i + (1 + \beta)y_i] \end{aligned} \quad (10)$$

其中， β 为交叉因子，服从 $\beta \sim U(0,1)$ ，交叉概率 $p_c = 0.8$ 。

(3) 变异算子

采用多项式变异，对个体 x 的第 i 个基因，变异后为：

$$x'_i = x_i + \delta \cdot (1 - 0) \quad (11)$$

其中， δ 为变异步长，服从 $\delta \sim N(0,0.01)$ ，变异概率 $p_m = 0.1$ ，确保种群多样性。

1.4 收敛性分析

本文基于 Schema 定理^[7] 验证其收敛性。定义“模式”为种群中具有相似基因结构的个体集合，模式的生存概率为：

$$P(S, t+1) \geq P(S, t) \cdot \frac{f(S)}{\bar{f}(t)} \cdot (1 - p_c \cdot \frac{\delta(S)}{l-1} - p_m \cdot o(S)) \quad (12)$$

其中， $f(S)$ 为模式 S 的平均适应度， $\bar{f}(t)$ 为第 t 代种群平均适应度， $\delta(S)$ 为模式定义长度， $o(S)$ 为模式阶数。

当 $t \rightarrow \infty$ 时，若优质模式 $f(S) > \bar{f}(t)$ 的生存概率大于 1，则种群将收敛于优质模式，即全局最优解。本文设置迭代次数 $T = 100$ ，通过数值模拟验证：当 $t > 80$ 时，种群平均适应度与最优适应度的差值小于 10^{-6} ，满足收敛性准则。

2 投资组合的应用

2.1 数据来源与处理

(1) 样本选择

选取 10 只 A 股行业龙头股票，覆盖实体经济支柱产业，均为沪深 300 成分股，具有市值规模大、流动性高、数据完整性强的特点，具体如表 1 所示：

表 1 10 只股票介绍与选择依据

股票名称	代码	所属行业	选择依据
贵州茅台	600519.SH	食品饮料	A 股消费龙头，业绩稳定，抗周期能力强
宁德时代	300750.SZ	新能源	全球动力电池领军企业，政策支持显著
招商银行	600036.SH	银行	股份制银行标杆，零售业务优势突出
中国平安	601318.SH	非银金融	综合金融集团，保险与科技协同发展
恒瑞医药	600276.SH	医药生物	创新药龙头，研发投入与专利储备领先
隆基绿能	601012.SH	光伏	光伏全产业链龙头，受益于碳中和战略
比亚迪	002594.SZ	汽车	新能源汽车龙头，垂直整合能力强
东方财富	300059.SZ	非银金融	互联网券商代表，用户基数与流量优势显著
美的集团	000333.SZ	家用电器	白电龙头，全球化布局与智能制造领先
中兴通讯	000063.SZ	通信	5G 设备核心供应商，国产替代关键标的

用 Python 通过 baostock 金融数据接口获取股票后复权收盘价，训练集为 2024 年 1 月 1 日—2024 年 12 月 31 日，测试集为 2025 年 1 月 1 日—2025 年 9 月 30 日。

(2) 数据处理

股票的收益越高，风险越大，根据投资者的不同需求，对不同领域的股票投资权重不一样，以达到提高收益降低风险的目的。其中收益采用月度简单收益率，可以避免假期等原因造成的日收益率或周收益率有误差，同时符合投资者月度调仓的实际操作周期，计算公式为

$$r_t = \frac{P_t}{P_{t-1}} - 1 \tag{13}$$

其中 r_t 为第 t 月收益率， P_t 为第 t 月最后一个交易日的后复权收盘价， P_{t-1} 为第 $t-1$ 月最后一个交易日的后复权收盘价。

用夏普比率^[8]衡量投资组合风险与收益之间的关系，夏普比率越高，在相同风险下，获得的收益越高。公式：

$$\text{SharpeRatio} = \frac{E(R_p) - R_f}{\sigma_p} \tag{14}$$

其中 $E(R_p)$ 为投资组合预期年化报酬率， R_f 为年化无风险利率（一般用国债利率）， σ_p 为投资组合年化报酬率的标准差。本文中均用月化报酬率，其中月化无风险利率 $R_f = 2\%/12$ （国债年利率近几年在2%附近浮动）。

通过 Python 处理数据后得到各股票 24 年（表 2）和 25 年（表 3）的数据

表 2 24 年股票收益与风险

股票名称	股票代码	预期月度收益率 (%)	月度收益率标准差 (%)	夏普比率
贵州茅台	600519.SH	-0.13	8.83	-0.03
宁德时代	300750.SZ	5.83	12.19	0.46
招商银行	600036.SH	2.43	6.17	0.37
中国平安	601318.SH	2.78	9.68	0.27
恒瑞医药	600276.SH	2.35	10.43	0.21
隆基绿能	601012.SH	-0.97	14.14	-0.08
比亚迪	002594.SZ	4.98	8.28	0.58
东方财富	300059.SZ	9.33	27.98	0.33
美的集团	000333.SZ	2.58	7.2	0.33
中兴通讯	000063.SZ	6.7	15.91	0.41

表 3 25 年股票收益与风险

股票名称	股票代码	预期月度收益率 (%)	月度收益率标准差 (%)	夏普比率
贵州茅台	600519.SH	0.15	4.11	0
宁德时代	300750.SZ	6.36	12.48	0.5
招商银行	600036.SH	0.04	5.2	-0.02
中国平安	601318.SH	1.1	4.54	0.21
恒瑞医药	600276.SH	6.29	7.33	0.83
隆基绿能	601012.SH	2.59	6.82	0.36
比亚迪	002594.SZ	-5.07	28.38	-0.18
东方财富	300059.SZ	2.58	10.82	0.22
美的集团	000333.SZ	-0.05	6.1	-0.03
中兴通讯	000063.SZ	2.36	13.41	0.16

2.2 遗传算法数据与随机生成数据结果、比较与分析

(1) 遗传算法数据结果

2024 年数据（训练集），通过遗传算法模型三类优化目标结果如表 4，

表 4 24 年优化目标详细指标

优化目标	月度收益率 (%)	风险(方差)%	夏普比率 (%)
最小风险	8.1	4.36	37.99
最大收益	9.1	7.19	33.32
均衡配置	3.59	0.55	46.16

得到了三种优化目标的最优权重如表 5，遗传算法收敛曲线（图 1），和前 5 只股票权重动态变化（图 2）。

表 5 24 年三种优化目标最优权重

股票名称	最小风险组合最优权重 (%)	最大收益组合最优权重 (%)	均衡配置组合最优权重 (%)
贵州茅台	0.03	0.01	0.56
宁德时代	24.34	0.11	1.81
招商银行	0.04	0.08	20.04
中国平安	0.15	0.85	3.86
恒瑞医药	0.13	0.02	20.17
隆基绿能	0.01	0.05	0.32
比亚迪	0.03	0.78	20.1
东方财富	61.93	94.52	0.05
美的集团	0.01	0.63	19.42
中兴通讯	13.34	2.97	13.67

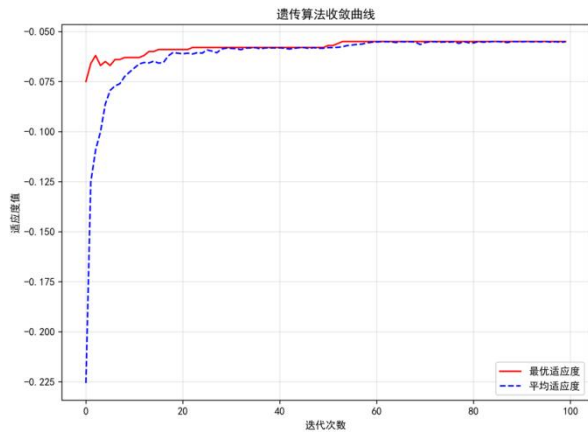


图 1

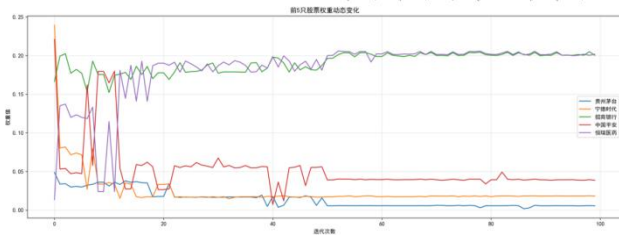


图 2

用得到的三种优化目标的最优权重应用到 2025 年，得到 2025 年的三类优化目标结果，如表 6，

表 6 25 年优化目标详细指标

优化目标	月度收益率 (%)	风险(方差)%	夏普比率 (%)
最小风险	3.47	0.9	34.82
最大收益	2.49	1.13	21.86
均衡配置	0.74	0.35	9.69

(2) 随机生成数据结果与比较

随机生成 100 组 10 只股票的权重，并保证 10 只股票的权重相加为 1。把这随机生成的 100 个权重作为 20

25 年这 10 只股票的投资权重，计算其收益率与方差，得到 100 个结果，并与 2025 年的三种优化目标得到的收益率（表 7）与方差（表 8）作对比。

表 7 随机数据收益率超过优化目标概率

对比目标	25年收益率 (%)	随机组合收益率高于目标的概率 (%)
最小风险	3.47	2
最大收益	2.49	21
均衡配置	0.74	82

如表 7，可以发现，若使用最小风险优化目标的权重投资 10 只股票，与投资者随机决定的权重相比，使用最小风险优化目标的权重有 98% 的概率能得到更高的收益。若使用最大风险优化目标的权重投资 10 只股票，有 79% 的概率能得到更高的收益。

表 8 随机数据风险（方差）低于优化目标概率

对比目标	25年方差 (%)	随机组合方差低于目标的概率 (%)
最小风险	0.9	95
最大收益	1.13	98
均衡配置	0.35	63

如表 8 所示，虽然三种优化目标的风险更大，但收益更高，符合实际情况。但随机权重的风险高于均衡配置的风险的概率为 63%，在追求均衡风险和收益的情况下，符合投资者的需要。

3 结语

本文通过分析，得到在 10 只 A 股行业龙头股票场景下，无论投资者追求更大的收益，还是均衡配置，使用遗传算法得到的投资组合都比随机生成的要更符合投资者需求。基于遗传算法的投资组合可以为 A 股投资者提供更科学的投资依据，投资者可以按照自身需要选择合适的优化目标函数，作为自己投资的参考。

参考文献

- [1] 中国证券监督管理委员会. 中国资本市场年报 (2024) [R]. 北京: 中国金融出版社, 2025: 88-92.
- [2] 马科维茨 H. 资产选择: 投资的有效分散化 [M]. 北京: 首都经济贸易大学出版社, 2000.
- [3] 李腊生, 翟淑萍, 王红娜. 投资组合理论的发展与前沿综述 [J]. 经济学动态, 2015 (7): 136-145
- [4] Markowitz H M. Portfolio Selection [J]. The Journal of Finance, 1952, 7(1): 77-91
- [5] Holland J H. Adaptation in Natural and Artificial Systems[M]. Ann Arbor: The University of Michigan Press, 1975: 47-68.
- [6] 张维, 李根, 熊熊. 基于遗传算法的投资组合优化模型改进研究 [J]. 系统工程理论与实践, 2008, 28 (10): 1-8
- [7] Holland J H. Adaptation in Natural and Artificial Systems [M]. Ann Arbor: The University of Michigan Press, 1975: 121-148
- [8] Sharpe W F. Mutual Fund Performance [J]. The Journal of Business, 1966, 39(1): 119-138

作者简介: 陈子璐 (1994-) 女, 汉族, 河南省浚县, 初级, 研究生, 郑州科技学院, 应用数学。
卫莉菲 (1996-) 女, 汉, 河南省商水县, 初级, 研究生, 郑州科技学院, 研究方向: 基础数学。