

学校代码：10240

学 号：Y1710203003

# 学 位 论 文

## 基于多目标进化算法的投资组合优化研究

黄文澜

指导教师姓名：	张雨 副教授	哈尔滨商业大学
申请学位级别：	硕 士	学 科 专 业 ： 计算机应用技术
论文提交日期：	2020-06-08	论文答辩日期： 2020-06-10
授予学位单位：	哈尔滨商业大学	授予学位日期：

哈爾濱商業大學

University Code: 10240  
Register Code : Y1710203003

Dissertation for the Degree of Master

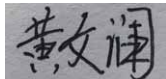
# Research on Portfolio Optimization Based on Multi-Objective Evolutionary Algorithm

<b>Candidate:</b>	Huang Wen Lan
<b>Supervisor:</b>	A/Prof. Zhang Yu
<b>Academic Degree Applied for:</b>	Master of Engineering
<b>Speciality:</b>	Computer Application Technology
<b>Date of Oral Examination:</b>	
<b>University:</b>	Harbin University of Commerce

# 哈尔滨商业大学硕士学位论文原创性声明

本人郑重声明：本论文的所有工作，是在导师的指导下，由作者本人独立完成的。有关观点、方法、数据和文献等的引用已在文中指出，并与参考文献相对应。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经公开发表的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本论文文责自负。

学位论文作者签名：



签字日期： 2020 年 6 月 15 日

# 哈尔滨商业大学硕士学位论文版权使用授权书

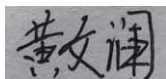
本论文系作者本人在哈尔滨商业大学攻读硕士学位期间，在导师指导下完成的硕士学位论文。本论文的研究成果归哈尔滨商业大学所有，本论文的研究内容不得以其他单位的名义发表。本人完全了解哈尔滨商业大学关于保留、使用学位论文的规定，同意学校保留、送交本论文的复印件，允许本论文被查阅和借阅；学校可以公布本论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

保密☐， 在 年解密后适用本授权书。

本学位论文属于

不保密☐。

学位论文作者签名：



签字日期： 2020 年 6 月 15 日

导师签名：

签字日期： 年 月 日

## 摘 要

近年来,得益于经济与科技的高速发展,我国人民的生活水平在不断提高,金融市场在经济活动中也逐渐占据重要的地位。然而,普通投资者具有较强的盲目性和从众心理,抗风险能力较低。因此,研究如何将资产进行有效的组合来寻找到更合理的投资策略,具有十分重要的意义。

多目标进化算法是解决多目标优化问题的主要方法,可得到使各个目标尽可能达到最优的一组均衡解。在股票投资过程中,投资者希望在提高收益的同时,也能降低风险,这是一种典型的多目标投资组合优化问题。本文的研究内容主要集中在以下两点:一方面,针对均值-方差模型的局限性,利用启发式函数链神经网络预测得到的收益率替代历史数据,采用基数约束对组合规模进行限制来降低资产管理难度、提高求解效率,引入偏度刻画实际收益率分布的非对称性,进而构建更具有实际意义的基于预测的带基数约束的均值-方差-偏度投资组合优化模型。另一方面,针对实际问题模型,在 NSGA-III 算法的基础上,环境选择阶段引入权衡收敛性与多样性的距离函数替代垂直距离,在配对选择阶段引入基于隶属度的锦标赛选择替代原来的随机选择方法,采用带资产信息的个体编码方案以及两阶段交叉、变异操作更新组合中的资产种类,利用可行解修复操作满足约束,降低模型求解复杂度,提出定制化的改进算法 c-NSGA-III。

本文以上证 A 股中的实际股票数据为基础,进行了大量的对比实验和结果分析,充分验证了改进模型和改进算法的有效性,且实验结果表明,定制化的 c-NSGA-III 算法能够优化模型的求解过程,相对于 NSGA-II、MOEA/D 效果更好,更具有有一定的实际应用价值,为高维目标投资组合优化问题提供了有效方法。

**关键词** 多目标;进化算法;组合投资;优化

# Abstract

In recent years, thanks to the rapid development of economy and technology, the living standards of our people have continued to improve, and financial markets have gradually occupied an important position in economic activities. However, with strong blindness and herd mentality, ordinary investors have a lower ability to resist risks. It is of great significance to study how to effectively combine assets to find a more reasonable investment strategy.

Multi-objective evolutionary algorithm (MOEA) is the main method to solve multi-objective optimization problem (MOP), It can obtain a set of equilibrium solutions that make each objective as optimal as possible. In the process of stock investment, investors hope to reduce risks while increasing returns, this is a typical multi-objective portfolio optimization problem. The content of this paper is mainly focused on the following two aspects: On the one hand, for the limitations of mean-variance model, the historical data is replaced by the return predicted by a heuristic functional link neural network. To reduce asset management difficulty and improve the computational efficiency, cardinality constraint is used to limit the size of portfolio, skewness is introduced to describe the asymmetry of the real return distribution, and then build a prediction based mean-variance-skewness cardinality constrained portfolio optimization model. On the other hand, for the practical problem, based on the NSGA-III algorithm, in the environment selection phase, a distance function that comprehensively considers convergence and diversity is introduced to replace the vertical distance, in the mating selection phase, tournament selection based on membership is introduced instead of random selection method, a coding scheme with asset information and two phase crossover and mutation operators are introduced to update asset type in a portfolio, feasible solution repairment operator is used to satisfy constraints and reduce the complexity of solving problem, and then propose a customized improved algorithm c-NSGA-III.

Based on the practical stock data of the SSE A Share, a large number of comparative experiments and results analysis are performed, which fully verified the effectiveness of the improved model and improved algorithm. The experimental results show that the c-NSGA-III algorithm can optimize the problem solving process, which is better than NSGA-II and MOEA/D and has practical application value, it also provides an effective method for many objectives portfolio optimization problems.

**Keywords** Multi-objective; evolutionary algorithm; portfolio; optimization

# 目录

摘 要.....	I
Abstract .....	II
1 绪论.....	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	2
1.2.1 投资组合模型研究现状.....	2
1.2.2 多目标进化算法研究现状.....	3
1.3 论文的主要工作和章节安排.....	4
2 Markowitz 投资组合理论 .....	6
2.1 基本概念.....	6
2.1.1 单一证券的收益与风险.....	6
2.1.2 多种证券组合的收益与风险.....	7
2.2 马科维茨均值-方差模型 .....	8
2.2.1 基本假设.....	8
2.2.2 均值-方差模型 .....	8
2.3 马科维茨投资理论的局限性.....	9
2.4 本章小结.....	9
3 多目标进化算法.....	10
3.1 多目标优化问题.....	10
3.2 多目标进化算法基本流程.....	11
3.3 多目标进化算法的分类.....	12
3.4 第三代非支配排序遗传算法 NSGA-III.....	14
3.4.1 定义参考点.....	14
3.4.2 种群成员的自适应标准化.....	16
3.4.3 关联操作.....	16
3.4.4 个体保留操作.....	17
3.5 本章小结.....	17
4 投资组合模型改进与求解.....	18
4.1 股票投资组合模型.....	18
4.1.1 基于 PSO-FLNN 的股票收益预测.....	18
4.1.2 基于预测的 M-V 投资组合模型.....	20
4.1.3 基于预测的三目标投资组合模型.....	20
4.2 利用多目标进化算法求解投资组合模型.....	21

4.2.1 相关算法 .....	21
4.2.2 性能评价指标 .....	22
4.3 实证分析 .....	23
4.3.1 实验说明 .....	23
4.3.2 经典模型实验 .....	24
4.3.3 改进模型实验 .....	27
4.4 实验结论 .....	29
4.5 本章小结 .....	29
5 定制化的多目标进化算法及应用 .....	31
5.1 定制化的多目标进化算法 .....	31
5.2 选择操作 .....	32
5.2.1 环境选择改进 .....	32
5.2.2 配对选择改进 .....	33
5.2.3 仿真实验 .....	33
5.3 个体更新 .....	34
5.4 实证分析 .....	37
5.4.1 实验说明 .....	37
5.4.2 实验结果与分析 .....	38
5.5 本章小结 .....	42
结论 .....	43
参考文献 .....	44
攻读学位期间发表的学术论文 .....	50
致谢 .....	51

# English Catalog

Chinese Abstract.....	I
Abstract .....	II
1 Introduction .....	1
1.1 Background and meaning of task .....	1
1.2 Research status .....	2
1.2.1 Research status of portfolio model.....	2
1.2.2 Research status of Multi-objective evolutionary algorithm .....	3
1.3 Main content and structure.....	4
2 Markowitz portfolio theory .....	6
2.1 Basic concepts.....	6
2.1.1 Return and risk of a security .....	6
2.1.2 Return and risk of a portfolio .....	7
2.2 Markowitz mean-variance model.....	8
2.2.1 Basic assumptions .....	8
2.2.2 Mean-variance model.....	8
2.3 Limitations of Markowitz portfolio theory .....	9
2.4 Chapter summary .....	9
3 Multi-objective evolutionary algorithm .....	10
3.1 Multi-objective optimization problem .....	10
3.2 Multi-objective evolutionary algorithm basic flow.....	11
3.3 Classification of multi-objective evolutionary algorithms.....	12
3.4 The third Generation Non-Dominated Sorting Genetic Algorithm NSGA-III .....	14
3.4.1 Define a reference point .....	14
3.4.2 Adaptive normalization of population members .....	16
3.4.3 Association operation .....	16
3.4.4 Niche-preservation operation .....	17
3.5 Chapter summary .....	17
4 Improvement of model and solving .....	18
4.1 Stock portfolio model.....	18
4.1.1 Stock return prediction based on PSO-FLNN.....	18
4.1.2 Prediction based M-V model.....	20
4.1.3 Prediction based three objectives portfolio model .....	20
4.2 Solving portfolio problem using multi-objective evolutionary algorithms.....	21



4.2.1 Relative algorithms .....	21
4.2.2 Performance indicators.....	22
4.3 Empirical analysis .....	23
4.3.1 Experiment description .....	23
4.3.2 Classic model experiment .....	24
4.3.3 Improved model experiment .....	27
4.4 Experimental results.....	29
4.5 Chapter summary .....	29
5 Customized multi-objective evolutionary algorithm and application.....	31
5.1 Customized Multi-Objective Evolutionary Algorithm.....	31
5.2 Selection operators .....	32
5.2.1 Environment selection improvement .....	32
5.2.2 Mating selection improvement.....	33
5.2.3 Simulation experiment .....	33
5.3 Individual update.....	34
5.4 Empirical analysis .....	37
5.4.1 Experiment description .....	37
5.4.2 Experimental results and analysis .....	38
5.5 Chapter summary .....	42
Conclusion .....	43
References .....	44
Relative academic papers .....	50
Acknowledgement.....	51

# 1 绪论

## 1.1 研究背景与意义

随着金融市场的迅速发展,投资理财不再局限于单一的储蓄方式,进行证券投资的人日益增多。然而各式危机一直在金融市场中游弋,2008 年金融海啸导致我国证券市场低迷不振,2015 年经历了行情上涨和股灾,随后又两次触发熔断,短期波动巨大,2018 年中美贸易战的展开,也为证券投资埋下了许多风险。中国股民总体呈现低龄、低学历的趋势,跟风炒股的错误心态使得其总体盈利状况较差。那么如何利用更科学、理性化的投资策略有效地实现资金保值或增值,已成为投资者必须考虑和关注的问题。总体而言,投资理财都是以寻求收益最大化为目的,但收益与风险往往同时存在。组合投资可以分散风险,其关键在于如何对资产进行配置以及如何处理风险与收益的关系。

1952 年,马科维兹提出均值-方差模型,量化了收益和风险<sup>[1]</sup>,为现代投资组合的发展打下了坚实的基础,在经济学的各个领域得到广泛的应用。该模型最大化投资组合的预期收益,同时最小化其相关的风险,从而在有限的风险资产集合中找到对有限资本的最优配置。但是该模型也受到众多基本假设的约束,比如使用历史股票收益的均值作为预期的未来收益,证券收益完全服从正态分布等等,这些假设会受到现实世界数据的威胁。多年来,专家、学者们不断对该模型进行研究和发展的,使其更好地适用于实际市场。实际市场往往涉及到更多的实际约束,它们反映了真实的投资组合决策过程,从计算的角度来看,如基数约束和交易费的引入使得问题转化为离散多目标非线性优化问题,属于 NP 难问题的类别<sup>[2]</sup>,难以用精确方法求解。与此同时,随着股票池中证券种类的增多,投资组合的管理难度也随之增加,虽然传统的手工计算,例如枚举、贪心算法等,可以取得全局最优解,但随着问题规模的增长,解的规模也呈指数级速度增长。

启发式算法可在多项式时间内获得令人满意的解,如进化算法,模拟退火算法,人工神经网络,量子算法等。由于投资组合问题的内在多目标性质,很多学者通常会用到多目标进化算法对其进行求解,相对于传统的求解多目标优化问题的方法,如加权法、约束法、目标规划法和极大极小法等,多目标进化算法既不需要知道所求问题的导数信息,也不需要聚合具有不同性质的各种优化目标,其能够在独立于问题领域的情况下,处理大规模的搜索空间,通过循环迭代的并行搜索对问题进行求解,逐代提高种群的平均适应度从而逼近全局最优解。

不论证券市场行情好坏,合理的证券投资组合都是投资者所追求的,多目标进化算法在投资组合问题中具有广泛的应用前景。基于以上考虑,本文一方面,在经典模型的基础上,对投资组合模型进一步完善,考虑均值-方差模型的固有缺陷和基本假设的局限性,在其基础上进行改进,从而更好地适用于实际的金融证券市场;另一方面,在

NSGA-III 算法的基础上, 针对问题模型, 对其进行定制化改进, 既拓展了多目标进化算法的相关研究, 也提高了投资者的决策效率。从长远看来, 本文的研究工作具有一定的理论价值和实际意义。

## 1.2 国内外研究现状

### 1.2.1 投资组合模型研究现状

投资组合是一种分散投资方式, 投资者将资金按特定比例分别投资于不同的资产, 形成一系列资产的组合, 在获得一定的收益的同时, 避免因为市场波动造成风险损失。现代投资组合理论是基于证券市场建立的, 其标志是 1952 年 Harry Markowitz 在其论文《Portfolio Selection》中提出的均值-方差模型<sup>[1]</sup>。马科维茨认为, 理想情况下, 投资者搜索最优的投资组合, 在最小化风险的同时, 也追求最大化收益。虽然该模型在一定程度上对风险进行了分散, 但其过分简单化, 且投资者需要面对现实世界的各种约束条件, 这些均导致该模型需要被不断改进, 增加了问题求解的复杂性。

以马科维茨的经典模型为基础, 研究者们不断对其进行改进。1963 年, William Sharpe 提出了单指数模型<sup>[3]</sup>, 随后在其基础上提出了多因素模型。Sharpe<sup>[4]</sup>, Lintner<sup>[5]</sup>以及 Mossin<sup>[6]</sup>各自独立提出资本资产定价模型 (Capital-Asset Pricing Model, CAPM), 该模型主要探求证券市场中资产收益与风险之间的数量关系, 可应用于解决资金成本预算、投资组合绩效等问题。但 CAPM 的假设前提过于苛刻且脱离实际, 贝塔参数也难测量, 很多学者认为经验式的检验与预言经济是不可能的, 对该模型的实用性表示质疑。Ross<sup>[7]</sup>基于无套利原则对 CAPM 进行了拓展, 提出了套利定价理论 (Arbitrage Pricing Theory, APT), 该理论以收益率形成过程的多因子模型为基础, 认为证券收益率与一组因子线性相关。Black 和 Scholes 提出了期权定价公式 (Option Pricing Model, OPM), 该模型以及它的一些拓展变形已经被期权交易商, 投资银行, 金融管理者等广泛使用, 但 OPM 中证券价格符合对数正态分布的前提假设与实际市场情况并不相符。虽然均值-方差模型各项基本假设与实际市场投资行为并不是完全一致, 但该模型可较精确地计算收益和风险的大小, 研究者们基于该模型在多个方面进行改进, 使其更符合实际情况。Yan<sup>[8]</sup>等把半方差 (Semi-variance, S-V) 作为投资组合收益的风险。Bradshaw 等采用风险价值 (Value-at-Risk, VaR) 作为风险度量的标准, VaR 衡量了在一定时间间隔上, 带有一定置信度的投资组合可能遭受的最大损失<sup>[9]</sup>。Hitaj<sup>[10]</sup>等研究了使用改进的赫芬达尔约束进行投资组合优化。

多年来, 国内学者也在不断深入投资组合方面的研究。李华<sup>[11]</sup>等在经典模型的基础上, 提出用熵和差熵作为风险的度量方法。张保帅<sup>[12]</sup>等通过对马科维茨的效率前沿的改进, 在系统性风险考量中引入了个别标的资产收益率变动的因素。考虑到交易成本, 陈国富<sup>[13]</sup>等以风险价值和夏普比率作为风险与效益的评价指标, 为达到全局与局部搜索的平衡, 在引力搜索和粒子群算法的基础上构建混合算法, 进而快速收敛到模型的最优

解。黄金波<sup>[14]</sup>等采用非参数核估计方法对条件风险价值（Conditional Value-at-Risk, CVaR）进行估计，替代理论上的 CVaR 来构建模型，实现风险估计与投资组合优化同时进行。

### 1.2.2 多目标进化算法研究现状

现实世界中，为解决一个问题通常需要同时优化多个相互冲突的目标，这就是多目标优化问题（Multi-objective Optimization Problem, MOP）。过去由于缺少合适的算法，多目标优化问题通常与单目标优化问题一样计算和求解<sup>[15]</sup>，然而单目标问题与多目标问题之间存在根本性的区别，多目标优化改善了某个目标必然会导致剩余其他目标性能变差，其需在各个目标之间进行协调权衡与折衷处理，进而得到一组均衡的非劣解。传统的多目标优化方法通过构造聚合函数，把多目标问题转化为一个单目标问题，每次只获得一个 Pareto 最优解，若想要获得近似解集，必须运行多次才可能有希望找到不同的 Pareto 最优解，而且一些传统方法对 Pareto 前沿（Pareto Front, PF）形状敏感，不能很好处理前沿的凹部，求解问题所需的启发式信息也很难获得。通常情况下，得到 PF 的近似解集往往更容易，也可以满足实际的需要，进化算法（Evolutionary Algorithm, EA）非常适合解决这种问题，在某种程度上能提供比较准确的均衡解集。

进化算法是一类启发式搜索算法，已成功应用于多目标优化领域，并逐渐发展为一个相对较热的研究方向<sup>[16]</sup>。群体搜索策略和群体中个体的信息交换是进化算法的两大特点<sup>[17]</sup>。其能够在独立于问题领域的情况下，处理大规模的搜索空间，通过循环迭代的并行搜索对问题进行求解，广受研究人员的欢迎，成为解决复杂优化问题的有效工具<sup>[18]</sup>。将进化算法运用在多目标优化问题上就是多目标进化算法（Multi Objective Evolutionary Algorithm, MOEA），因一次运行，便可得到近似 PF 的非支配解集，其多目标优化领域获得了广泛的认同和应用<sup>[19]</sup>。

近几十年，多目标进化算法的发展浪潮大致分为四个阶段<sup>[18]</sup>。初始阶段为 1967 年，Rosenberg<sup>[20]</sup>提出采用遗传搜索的方式来求解多目标优化问题，但没有具体实现。1985 年，Schaffer<sup>[21]</sup>提出了向量评估遗传算法（Vector-Evaluated Genetic Algorithms, VEGA），开创了用进化算法处理 MOP 问题的先河。1989 年，Goldberg<sup>[9]</sup>提出将 Pareto 理论与进化算法结合。第一次多目标进化算法浪潮为 20 世纪 90 年代初，引入了 Pareto 支配概念<sup>[22]</sup>，运用非支配排序进行等级分配，采用适应度共享机制保持种群多样性。典型算法有 Fonseca、Fleming 等提出的多目标遗传算法<sup>[23]</sup>（Multiple Objective Genetic Algorithm, MOGA），Srinivas、Deb 等提出的非支配排序遗传算法<sup>[24]</sup>（Nondominated Sorting Genetic Algorithm, NSGA），Horn、Nafpliotis 等提出的小生境 Pareto 遗传算法<sup>[25]</sup>（Niche Pareto Genetic Algorithm, NPGA）。该期间亟待改进的是如何降低计算复杂度以及寻找更好的多样性保持策略。第二次多目标进化算法浪潮为 1999 年至 2002 年，主要特点是采取了精英保留机制，该期间提出了一些更好的多样性保持策略。典型算法有

Zitzler、Thiele 等提出的强度 Pareto 进化算法<sup>[26]</sup> (Strength Pareto Evolutionary Algorithm, SPEA), Knowles、Corne 等提出的 Pareto 存档进化策略<sup>[27]</sup> (Pareto Archived Evolution Strategy, PAES) 和 Pareto 包络选择算法<sup>[28]</sup> (Pareto Envelope-Based Selection Algorithm, PESA) 以及 PESA-II<sup>[29]</sup>, Erichson、Mayer 等提出了 NPGAII<sup>[30]</sup>, Coello、Pulido 等提出的微遗传算法<sup>[31]</sup> (Micro-Genetic Algorithm, Mircro-GA)。最具有代表性的是 SPEAII<sup>[32]</sup>, 以及 Deb 等提出的非支配排序遗传算法 II<sup>[33]</sup> (Nondominated Sorting Genetic Algorithm-II, NSGA-II)。该期间的众多算法主导了进化多目标优化领域多年, 但仍无法很好地处理高维目标优化问题。第三次多目标进化算法浪潮, 主要发展时期为 2003 年至今, 呈现多方向、多样化的特点: (1) 高维多目标优化的研究, 基于指标和基于分解的 MOEA<sup>[34]</sup>得到了广泛的应用, Deb 等人提出的 NSGA-III<sup>[35]</sup> (the reference-point based many-objective NSGA-II) 也受到了广泛关注。(2) 引进一些新的进化范例, 如 Coello<sup>[36]</sup>等提出的多目标粒子群优化算法 (Multi-objective Particle Swarm Optimization, MOPSO)。(3) 在环境选择算子方面的研究, 如采用新型支配关系的算法<sup>[37]</sup>。(4) 众多学者相继设计更多针对不同类型问题的高效的标准测试问题集, 如 Deb 等人提出的 ZDT 测试集<sup>[38]</sup>, DTLZ 测试集<sup>[39]</sup>。(5) 结合机器学习技术的多目标进化算法的研究也在大力开展, 如 Zhang<sup>[40]</sup>等人利用 Pareto 解集在连续多目标问题上的结构特征提出了基于规则模型的多目标分布估计算法 (Regularity Model Based Multi-objective Estimation of Distribution Algorithm, RM-MEDA), 还有如基于反向学习的多目标进化算法<sup>[41, 42]</sup>等。最近的一些研究开始在以下两个方向展开: (1) 在保证解的收敛性的基础上建立一个更可靠的终止条件<sup>[43]</sup>。(2) 交互多目标优化, 将偏好交互地融入到进化算法中, 不需要在算法结束后再根据偏好进行选择, 而是直接产生决策者偏好的 Pareto 解集<sup>[44]</sup>。这两个领域的研究也可称为第四次多目标进化算法浪潮<sup>[43]</sup>。

投资组合问题是多目标进化算法应用的重要领域。Deb<sup>[45]</sup>等利用定制化的 NSGA-II 算法与局部搜索结合来求解投资组合问题; Subbu<sup>[46]</sup>等将进化计算和线性规划结合, 利用这种混合多目标优化方法来得到满足各项约束的有效投资组合; 孙雪莲<sup>[47]</sup>等在经典模型的基础上引入降低组合资产数目的目标, 构建三目标投资组合模型, 并利用 NSGA-II 算法对其进行多目标优化; 严丽娜<sup>[48]</sup>等提出使用基于种群密度的多目标协同进化算法来求解含预算约束的投资组合模型; 任雪婷<sup>[49]</sup>等对基本的多目标差分进化算法进行改进, 求解考虑交易费用的投资组合模型; 李锋刚<sup>[50]</sup>等对粒子群算法进行改进, 增强其离散型问题的搜索能力, 引入自适应粒子迁移保证多样性, 并对带有实际约束的均值-CVaR 投资组合模型进行了求解。

### 1.3 论文的主要工作和章节安排

本文以经典的均值-方差模型为基础, 考虑股票投资的实际情况以及经典模型的局限性, 构建基于预测的带有基数约束的三目标投资组合模型, 为投资者提供可行的投资

建议。针对实际问题模型，对 NSGA-III 进行定制化改进，环境选择阶段引入综合考虑收敛性与多样性的距离函数，配对选择阶段采用基于隶属度的锦标赛选择方法，采用有效的个体编码方案以及两阶段交叉、变异算子来提高模型求解的效率和效果。实证分析部分，为验证模型和算法的实践意义，本文在实际股票数据上，进行了大量的对比实验和结果分析。

本文各章节的内容安排为：

第一章主要介绍了本文的研究背景与意义、该领域研究现状以及本文的主要工作。

第二章具体介绍了马科维茨投资组合理论的基本概念，数学描述以及面对实际市场的局限性。

第三章对多目标优化问题进行描述，给出多目标进化算法框架，并按类别对其进行阐述，介绍了一些经典算法，重点分析了 NSGA-III 算法。

第四章针对经典模型的缺陷，提出改进的投资组合模型，进行多目标进化算法求解投资组合模型的基本实验与结果分析。

第五章针对实际问题模型，对算法进行定制化改进，并进行实验与分析。

## 2 Markowitz 投资组合理论

### 2.1 基本概念

在低利率时代，投资者将巨额财富存于银行不再是最佳选择。随着科技的高速发展以及金融体制的不断完善，我国金融市场积极与国外接轨，金融产品愈加多样化，债券、股票、期货、外汇、以及各种互联网金融衍生品逐渐走向普通民众的投资范围中。投资就是在某一段时间内，将资金转化为某一领域的资产或资本，从而获取经济回报或增值收益的行为，当然无论进行何种投资，目的都是为了获得较高的收益。但是收益与风险往往并存，每一个投资者都想获得尽可能多的收益，同时尽可能的承担最低的风险，如何通过投资理财，实现资产的保值增值是一个很大的挑战。

不将所有鸡蛋放在一个篮子里，这是投资组合规避风险的基本思想。投资人在综合考虑自己的资金实力以及风险承担能力的基础上，将有限的资本合理分配到不同的投资产品，形成一个投资集合。投资组合的风险随着资产多样性的增加而降低，从而达到在获取预期收益的前提下分散风险的目的。在过去很长一段时期里，投资组合的研究中，如何定量衡量产品的风险是一大难点，有关收益和风险的定性研究较多，却没有成功的定量分析方法。1952 年，马科维茨首次将数量关系应用到投资组合理论的研究中，对收益和风险进行了量化，提出最大化收益和最小化风险这两个目标，为投资者们解决有关金融产品选择以及资金比例配置的问题<sup>[1]</sup>。下面以股票投资问题为基础，对相关概念进行介绍，股票投资组合是指在 N 只股票中选择特定的多种股票，每种股票投资一定的金额，在可控的风险范围内，获得最大投资收益，收益与风险是其关注的两个综合指标。

#### 2.1.1 单一证券的收益与风险

大多数的证券产品的收益难以预测，例如投资一只股票，突发事件、市场行情的好坏等各种因素对股价的变动都具有不同程度的影响。在实际投资活动中，通常使用历史实际收益率均值替代预期收益率来衡量投资组合的优劣。假设证券池有 M 只证券，对于某只证券的实际收益率计算如式（2-1）所示：

$$r_i = \frac{S_i(t) - S_i(0)}{S_i(0)} \quad (2-1)$$

其中， $r_i$  是某证券的实际收益率， $S_i(t)$  是期末该证券收盘价以及持有期间的红利， $S_i(0)$  是期初该证券收盘价。

马科维茨模型认为公司的运营和业绩通常具有一定的稳定性，可使用历史实际收益率的均值作为该证券预期收益率的估计。设向量  $R_i = (r_i^1, r_i^2, \dots, r_i^T)^T$  表示第 i 种资产的多期历史收益率组成的向量，那么预期收益率如式（2-2）所示：

$$\overline{R}_i = E(R_i) = \sum_{t=1}^T r_i^t p_i^t \quad (2-2)$$

其中,  $\overline{R}_i$  是某证券的预期收益率,  $T$  是历史实际收益率的期数,  $r_i^t$  是第  $t$  期该证券的实际收益率,  $p_i^t$  是实际收益率为  $r_i^t$  时的概率。

在投资活动中, 使用历史收益率对预期收益率进行估计具有不确定性, 因为真实的收益可能高于预期, 也可能低于预期, 这是众多投资者所面临的风险。目前在风险的衡量上没有一致认同的模型, 马科维茨认为, 可以用预期收益率的方差来对风险进行量化, 对于单一证券, 可用式 (2-3) 计算其风险:

$$\sigma^2 = D(R_i) = \sum_{t=1}^T (r_i^t - \overline{R}_i)^2 p_i^t \quad (2-3)$$

上式利用证券收益的波动性, 即某证券预期收益率的方差  $\sigma^2$  对风险进行量化, 方差越大, 表示实际收益率与预期收益率偏离越大, 说明该证券收益具有较大的不确定性, 投资风险较强。

## 2.1.2 多种证券组合的收益与风险

马科维茨投资组合理论中, 若某投资组合是从证券池中选择  $N$  种资产并按一定的投资比例组合在一起的, 那么在某投资周期内, 该组合的收益是通过各资产收益率的加权平均和来进行量化的, 可用式 (2-4) 计算:

$$r_p = \sum_{i=1}^N x_i \overline{R}_i \quad (2-4)$$

其中,  $r_p$  是投资组合的期望收益率,  $x_i$  为第  $i$  种资产在组合中的所占的资金比例, 满足预算约束  $\sum_{i=1}^N x_i = 1$ , 且  $x_i > 0$ ,  $\overline{R}_i$  为该资产的预期收益率。

在实际市场中, 各种证券之间不会完全相互独立, 往往都存在一定的联系, 它们的收益率可能朝着相同的方向, 也可能朝着相反的方向变化。证券之间的这种关联性通常用协方差来表示, 它是衡量两个变量之间总体误差的期望。在实际的证券组合中, 通常希望选择一些尽量不相关资产构成组合来尽量分散风险。设  $R = (R_1, R_2, \dots, R_N)^T$  为投资组合中各个证券的实际收益率,  $i, j$  两种证券之间的协方差如式 (2-5) 计算:

$$\sigma_{ij} = \text{cov}(R_i, R_j) = \frac{1}{T} \sum_{t=1}^T (r_i^t - \overline{R}_i)(r_j^t - \overline{R}_j) \quad (2-5)$$

在马科维茨投资组合理论中, 使用方差  $\sigma_p^2$  作为投资组合的风险度量, 除了要考虑单个证券的特征, 还要考虑它们之间的关系, 组合风险表示为式 (2-6):

$$\sigma_p^2 = D(R) = \sum_{i=1}^N \sum_{j=1}^N x_i x_j \sigma_{ij} \quad (2-6)$$



## 2.2 马科维茨均值-方差模型

马科维茨在《Portfolio Selection》<sup>[1]</sup>中，详细地论述了投资组合的基本原理，同时利用数量化方法提出了确定最佳资产组合的基本模型，首次提出使用资产收益率的均值来衡量投资者获得的收益大小，利用收益率的方差来表示需要承担的风险，其中均值表示一组数据的集中趋势，方差表示一组数据偏离其集中趋势的大小程度。下面对马科维茨投资组合理论的基本假设进行介绍。

### 2.2.1 基本假设

马科维茨在建立均值-方差模型时，设定了一些基本假设，但在实际的证券市场中往往很难完全满足这些条件。具体包括：

- (1) 证券市场是完全有效的；
- (2) 投资者都是理性的，厌恶风险；
- (3) 历史收益率的均值和方差分别作为预期收益以及风险的度量，协方差是衡量每对证券共同风险的指标；
- (4) 证券的收益率服从正态分布；
- (5) 单期投资，期初投资，期末获取回报；
- (6) 交易中没有最小交易单位的限制；
- (7) 交易过程中忽略税收和交易成本。

### 2.2.2 均值-方差模型

马科维茨的均值方差模型，实质上是一个非线性的双目标的最优化问题，为的是达到最大化收益和最小化风险双目标的均衡状态。模型的数学描述如式（2-7）所示：

$$\begin{aligned} \min V(x) &= X^T V X = \sum_{i=1}^N \sum_{j=1}^N x_i x_j \sigma_{ij} \\ \max R(x) &= X^T \bar{R} = \sum_{i=1}^N x_i \bar{R}_i \\ \text{s.t. } \sum_{i=1}^N x_i &= 1 \quad (x_i > 0) \end{aligned} \quad (2-7)$$

其中， $X$  是投资到各个证券上的资产比例组成的权重向量， $X = (x_1, x_2, \dots, x_N)^T$ ， $N$  项资产的期望收益率向量为  $\bar{R} = (\bar{R}_1, \bar{R}_2, \dots, \bar{R}_N)^T$ ， $V = (\sigma_{ij})_{N \times N}$  为协方差矩阵。

如式（2-8）所示，M-V 模型可以看作一个典型的带约束的二次规划问题，结合背包问题可得出二次规划是 NP 难问题<sup>[51]</sup>。投资组合问题是一类非常复杂的决策问题，亦被证实是组合优化问题中带约束的 NP 难问题，难以使用常规的方法求解。

$$\begin{aligned} \min f(x) &= X^T V X - X^T \bar{R} \\ \text{s.t. } \sum_{i=1}^N x_i &= 1 \quad (x_i > 0) \end{aligned} \quad (2-8)$$

## 2.3 马科维茨投资理论的局限性

虽然马科维茨投资组合理论为现代投资组合研究发展奠定了基础，但在面对实际问题时，仍然存在着一定的局限性。

该理论假设只进行单期投资，即资金分配只在期初进行一次，之后不进行改变，这不符合实际的投资行为，不适用于多阶段问题。经典的均值-方差模型经常受到现实数据的威胁，在实际金融市场中，一系列收益的分布经常偏离正态分布，并且表现出峰度和偏度，这使得收益的方差不合适风险度量。使用历史收益均值作为股票未来收益的估计，对股票市场的动态行为产生了低通滤波效应。另外，随着投资组合中证券种类数量的增多，数据获取以及期望、协方差的计算都需要很大的工作量。因此，开发一个没有这些缺点的模型仍然是一个具有挑战性的研究领域。

## 2.4 本章小结

本章主要介绍投资组合的基本知识，同时重点阐述了马科维茨基本投资组合模型的数学表达，并分析了模型的局限性。

## 3 多目标进化算法

### 3.1 多目标优化问题

现实世界的许多情况下，为了解决一个问题通常需要同时优化多个相互冲突的目标，这就是所谓的多目标优化问题。投资组合的均值-方差模型就是典型的多目标优化问题，需要在最大化组合收益的同时，最小化组合风险。多目标优化通常没有对于所有目标都是最优的一个解。这就意味着，改善了某个目标必然会导致剩余其他目标性能变差，为了使各个目标尽可能最优，需要在它们之间进行协调权衡处理，得到一组均衡的非劣解，即 Pareto 最优解集合。

以所有目标最小化问题为例（最大化问题可对结果取反）。多目标优化问题由  $n$  个决策变量， $m$  个目标变量和若干个约束组成，数学描述<sup>[15]</sup>如式（3-1）所示：

$$\begin{aligned} \text{minimize } y = F(x) &= (f_1(x), f_2(x), \dots, f_m(x))^T \\ \text{subject to: } x &= (x_1, x_2, \dots, x_n)^T \in \Omega \end{aligned} \quad (3-1)$$

其中： $x = (x_1, x_2, \dots, x_n)^T$  是由  $n$  个决策变量组成的一个  $n$  维决策向量， $\Omega \subseteq \mathbb{R}^n$ ，是决策空间。 $x_i(L) \leq x_i \leq x_i(U)$ ， $i=1, 2, \dots, n$ ， $x_i(L)$  和  $x_i(U)$  分别是  $x_i$  取值的上下边界。 $y = (y_1, y_2, \dots, y_m)^T$  是  $m$  维目标向量， $y \in \mathbb{R}^m$ ， $\mathbb{R}^m$  是目标空间。 $F(x): \Omega \rightarrow \mathbb{R}^m$ ，是  $m$  个目标函数组成的，由决策空间到目标空间的映射。

（1）可行解： $\forall x \in \Omega$ ，若  $x$  满足（3-1）中的约束，那么  $x$  是可行解。

（2）可行解集合：记作  $X^f$ ，它是所有可行解组成的集合，也称为可行域或搜索空间。

（3）Pareto 支配：假设  $x, x' \in X^f$  是多目标优化问题的两个可行解向量，如果  $\forall i=1, 2, \dots, m, f_i(x) \leq f_i(x') \cap \exists j=1, 2, \dots, m, f_j(x) < f_j(x')$ ，记作  $x \succ x'$ ，称  $x$  支配  $x'$ 。支配的实质是“更优于”。

（4）Pareto 最优解（Pareto Optimal Solution）：不存在其他的可行解能够支配它，这是最优解的条件。即  $x$  是 Pareto 最优的，当且仅当  $\neg \exists x': x \succ x'$ 。值得注意的是，Pareto 最优解的信息可以辅助分析目标函数之间的冲突。

（5）Pareto 解集（Pareto Set, PS）：是所有 pareto 最优解的集合，实质上就是一组折衷解或均衡解的集合， $PS = \{x \mid \neg \exists x': x \succ x'\}$ ，一般多目标进化算法的目的是求得最优解的集合，然后由决策者从中进行筛选。

（6）Pareto 前沿（PF）：是 Pareto 最优解集中所有最优解对应的目标向量在目标空间所形成的曲面， $PF = \{F(x) = (f_1(x), f_2(x), \dots, f_m(x))^T \mid x \in PS\}$ 。

## 3.2 多目标进化算法基本流程

进化算法随机初始化产生一个种群，种群由一定数目的个体组成，在多目标优化问题中对应的是潜在的可行解集和解集中的可行解。个体在进化过程中相当于遗传物质的主要载体染色体，其内部由多个基因组成，称为基因型，个体性状的外部表现称为表现型。遗传操作要在基因层次进行，所以开始时需将个体的表现型编码映射成基因型，编码方式有实数编码，二进制编码和符号编码等。初始化后，种群模拟自然进化，使用适当的方法评估个体对环境的适应度，通过选择、交叉和变异操作逐代进行改良，并行地产生一组逼近最优解的较好的近似解。

通常对于某个特定的多目标优化问题，设计一个高效的进化算法需要考虑以下事项：（1）如何以某个编码步骤来设计一种能够表示个体并且满足问题约束的种群初始化方法。（2）如何通过带有解码步骤的适应度函数来评估个体。（3）怎样通过遗传操作改良种群。

典型的进化算法的基本流程：

- ① 随机产生初始种群  $P(t)$ ，当代  $t=0$ ；
- ② 评价群体  $P(t)$  中每个个体  $x$  的适应度值；
- ③ 选择个体进行交叉、变异： $P'(t) = \text{Crossover}[P(t)]$ ， $Q(t) = \text{Mutation}[P'(t)]$ ；
- ④ 评价  $Q(t)$  中个体的适应度值；
- ⑤ 进行环境选择，生成下一代种群  $P(t+1) = \text{Reproduction}[P(t) \cup Q(t)]$ ；
- ⑥ 终止条件判断：若不满足终止条件，则  $t = t + 1$ ，转至第③步，继续执行进化操作；若满足终止条件，则输出当前最优种群，算法结束。

选择操作是一种适者生存的过程，基于个体的适应度进行比较，按照一定的概率从种群中挑选出若干个体，主要用来确定能够进入配对池进行交叉和变异的个体以及通过环境选择产生下一代种群。不同的选择策略将会导致不同的选择压力。主要的选择方法有轮盘赌方法，精英个体保留策略，锦标赛选择方法，排序选择方法，共享函数法等。

交叉是结合父代用于交配的个体的信息来产生新个体的遗传操作，是把父代中配对的个体的部分基因加以替换重组进而生成新个体的操作。进行交叉时，要求不能过多破坏个体基因串中的优良模式，同时又能产生较好的新个体模式。根据个体编码方式的不同，可以有以下方法：（1）实值交叉：如离散交叉，线性交叉等。（2）二进制交叉：如单点交叉，均匀交叉等。

变异，实际上是个体基因按小概率扰动产生的变化，其目的主要有两个：（1）增强局部搜索能力，（2）保持种群的多样性，防止早熟现象出现。依据个体编码表示方法的不同分为实值变异和二进制变异两类，主要的方法有：均匀变异，非均匀变异，多项式变异等。

根据模式定理和积木块假设，那些具有低阶，短定义距，以及平均适应度高于种群平均适应度的模式，在遗传操作算子选择、交叉和变异的作用下，这类模式在子代中呈

指数增长, 逐代产生适应度更高的个体, 找到更好的解。

### 3.3 多目标进化算法的分类

研究人员在多目标优化领域中进行研究与探索, 相继提出了多种多目标进化算法, 根据当前算法的特点, 将其分为以下四类:

#### (1) 基于支配的多目标进化算法

这类算法的基本思想是, 为了达到良好的收敛性, 采用基于 Pareto 支配的等级划分对个体进行适应度分配, 为了保持解的多样性, 配备了明确的多样性保持方案, 在各个等级对个体进行密度估计。综合考虑个体的等级及其密度估计进行排序, 选择阶段: 若等级不同, 则选择优秀等级的个体, 若等级相同, 则选择密度较小的个体进入子代。典型的算法有 SPEAII、NSGA-II。SPEAII 是 Zitzler 和 Thiele 在 2001 年提出的对 SPEA 的改进版本, 在适应度赋值中增加了个体的密度值, 该值是个体到其第  $k$  个最近邻的距离所决定的密度估计, 引入基于最近邻域的环境选择来简化 SPEA 中基于聚类的外部种群更新方法。NSGA-II 是 2002 年 Deb 等人对其算法 NSGA 的改进, NSGA-II 采用基于 Pareto 支配的快速非支配排序对个体进行分级, 有效降低了计算复杂度  $O(MN^2)$ ; 采用拥挤比较操作算子替代 NSGA 的适应度共享方法, 不再指定共享参数; 引入  $(\mu+\lambda)$ <sup>[52]</sup> 的精英保留策略, 父代和父代经遗传操作产生的后代一同竞争来产生下一代种群。尽管基于 Pareto 支配的多目标进化算法很流行, 但是仍然存在一些缺点, 该类算法收敛至 PF 的速度较慢, 解决带有复杂 PF 的 MOPs 的性能较差, 且每个非支配解相对于 Pareto 最优解集的真实收敛性是未知的。随着目标数量的增加, 基于 Pareto 支配产生的非支配解在可行解中的比例呈指数形式增加, 算法的选择压力急剧下降, 没有能力使解集近似到 PF, 阻碍了进化, 多样性和收敛性都受到了破坏。

#### (2) 基于评估指标的多目标进化算法

基于指标的 MOEA, 其基本思想是将评价指标整合到 MOEA 中。即将性能指标作为环境选择的尺度来引导搜索, 不断优化整个种群的期望属性, 从而得到 MOP 的一组良好的近似最优解。第一个著名的基于指标的进化算法是基于指标的进化算法<sup>[53]</sup> (indicator-based evolutionary algorithm, IBEA), 是 2004 年由 Zitzler 和 Kunzli 提出的, 其将决策者的偏好定义为二元性能指标, 环境选择是基于这个提前定义的性能指标设计的。另一个早期有代表性的算法是由 Beume<sup>[54]</sup> 等人提出的 SMS-EMOA (Smetric selection evolutionary multi-objective algorithm), 使用超体积指标 (hypervolume, HV) 作为选择标准, 避免了 Pareto 支配比较引起的大量计算开销。随着目标数目增多, HV 计算开销呈指数增长, 所以很难解决高维目标优化问题 (many-objective optimization problems, MaOPs)。Bader 和 Zitzler<sup>[55]</sup> 提出一个基于快速 HV 的进化算法。目前基于超体积指标的多目标进化算法最为常见, 其不需要知道 PF, 且包含了近似, 多样性和相关性信息, 可综合评估近似解集的质量。还有以世代距离 (inverted generational

distance, IGD) 指标为基础作为选择标准的研究, 如 Ye Tian<sup>[56]</sup>等人提出用增强的世代距离指标 (IGD with noncontributing solution detection, IGD-NS) 来区分对 IGD 指标没有贡献的解。

### (3) 基于分解技术的多目标进化算法

大多数 MOEA 只把 MOP 看作一个整体, 在搜索期间主要依赖支配的方法来度量解的质量, 当求解高维目标优化问题时, 选择压力急剧下降, 妨碍了进化过程。可能不会很好地产生均匀分布于 PF 的解。基于标量函数的适应度估值相对于基于 Pareto 支配的方法有几个优点, 比如对问题目标数量的可伸缩性以及较高的搜索能力。同时, 标量函数改进了适应度估值的计算效率, 尤其是高维目标问题, 它的计算复杂度不随目标增加呈指数增长。因此, 很多基于分解的具有标量适应度函数的多目标进化算法被提出来了, 其中最具有代表性的是 2007 年 Zhang 和 Li<sup>[34]</sup>提出的基于分解的多目标进化算法 (Multi-objective Evolutionary Algorithm Based on Decomposition, MOEA/D), 分解策略这才广泛应用到多目标进化优化中。该算法的基本思想是根据权重向量, 利用分解策略将多目标优化问题分解为一组标量优化子问题, 每个子问题的解的优化通过其邻居子问题之间的进化操作来完成, 进化算法有机地组织这些子问题进行求解<sup>[57]</sup>。MOEA/D 不再将传统数学规划法与进化算法相互孤立, 而是结合起来成为一种求解多目标优化问题的有效的方法, 由于多目标被分解为一组单目标和近邻关系, 所以各种传统的单目标优化方法和局部搜索方法均可以应用到 MOEA/D 的算法框架中。当然分解不仅存在于目标空间, 也可以在搜索空间进行分解, 如 Darshan<sup>[58]</sup>等人提出的基于搜索空间的多目标计划算法 (search space-based multi-objective evolutionary algorithm, SSMOEA)。

虽然 MOEA/D 有很多优点, 但仍存在很多限制, 比如 Das 和 Dennis 提出的权重向量产生方法<sup>[59]</sup>, 其缺点是: 种群规模随着目标数量的增加而呈非线性增长, 种群规模不能随意设定, 而且对于三个或更多目标权重向量分布不是很均匀。MOEA/D 对于不同的问题很难选择合适的分解方法, 而且对于高维表现得不太好。针对 MOEA/D 算法中权重向量与个体分配不合理而导致种群多样性降低的问题, 张磊<sup>[60]</sup>等提出基于重新匹配策略的  $\epsilon$  约束多目标分解优化算法。

### (4) 混合算法

在设计算法时, 为了提高其求解能力, 很自然会想到在其中混合不同算法的特征, 运用它们的优点解决复杂的多目标优化问题, 由此提出了一系列的混合多目标进化算法。如 Hadka 和 Reed<sup>[61]</sup>提出一个自适应高维进化计算框架, 它把多个操作算子在一个标准的优化框架中重新组合起来。

① 混合不同的搜索方法, 可以结合全局搜索和局部搜索方法, 也可以结合不同算法的搜索操作算子。如带有自适应选择操作的混合非支配排序遗传算法<sup>[62]</sup> (Hybrid Nondominated Sorting Genetic Algorithm, HNSGA)。还有比较经典的是 Elhossini<sup>[63]</sup>等人把 PSO 和 EA 混合, 每一代由 PSO 操作算子产生个体后再施加 EA 的操作算子。

② 在搜索过程分段采用不同的搜索策略, 如 Yang<sup>[64]</sup>等, 把搜索过程分为不同的时期, 在不同的时期使用不同的搜索策略。

③ 将混合算法与新型进化模型的结合, 如谢承旺<sup>[65]</sup>等提出的一种增强型多目标烟花爆炸优化算法。邱飞岳<sup>[66]</sup>等提出将分组分解融合到多目标粒子群算法中以提高邻域局部搜索的速度。

目前大多数研究集中在如何进行混合搜索来产生子代。其他方面的研究工作不多。适应度分配对多目标进化算法的性能也有很大的影响, Gu<sup>[67]</sup>等提出了带有自适应多适应度分配的混合进化多目标优化算法, 它在一个标准的优化框架中混合不同的适应度分配策略, 并且为每个子种群自适应选择最合适的适应度分配方法。

### 3.4 第三代非支配排序遗传算法 NSGA-III

NSGA-III 算法框架与 NSGA-II 基本相同, 同样利用快速非支配排序将合并种群  $R_t$  的个体分层进入不同的非支配前沿, 在基于分层的非支配排序中: 定义了两个向量  $\{n_p\}$  和  $\{s_p\}$ ,  $p$  为种群中的个体, 支配  $p$  的个体数为  $n_p$ ,  $p$  支配的个体的集合为  $s_p$ , 计算每个个体的  $n_p$  和  $s_p$ 。因为第一层为种群的非支配个体的集合, 所以第一层个体的集合  $F_1 = \{q | n_q = 0, q \in R_t\}$ , 第二层为种群中去掉第一层后的非支配个体的集合, 第三层为去掉第一层和第二层个体后的非支配个体的集合, 依次按照该方法得到各层个体的集合:  $F_s = \{\text{所有个体 } q | n_q - s + 1 = 0\}$ 。

环境选择阶段: 非支配前沿 1 层到  $L$  层的个体被放入集合  $S_t$ , 如果  $|S_t| = N$ , 那么下一代的父代  $P_{t+1} = S_t$ 。如果  $|S_t| > N$ , 那么先将 1 层到  $L-1$  层的个体放入  $P_{t+1}$ , 此时还差  $k = N - |S_t|$  个个体, 需从临界层  $L$  中选择。NSGA-II 利用拥挤比较操作进行选择, 用以保持多样性。NSGA-III 最大的变化就是利用良好分布的参考点来保持种群的多样性, 当参考点均匀分布在标准超平面上时, 那么得到的解可能会接近或均匀分布在 Pareto 最优前沿。NSGA-III 是一种处理高维目标优化问题的有效方法, 但由于基于 Pareto 支配关系的低选择压力, 其收敛性仍然可以进行改进。NSGA-III 算法的基本流程如图 3-1 所示。

#### 3.4.1 定义参考点

NSGA-III 中, 参考点既可以用结构化的方法预定义也可以由使用者提前提供。Das 和 Dennis<sup>[59]</sup>提出的系统化边界交叉构造权重方法, 在目标维度较大的情况下, 参考点数量则会急剧增加, 为了避免这种情况, 本文采用 Deb<sup>[35]</sup>提出的两层参考点产生方法, 增加了内层, 这样既能保证中间点的产生也不会使参考点数目过多。

一个  $M-1$  维的标准单纯形, 它对所有的目标轴都有相同的倾斜度。同样, 在两层参考点产生法中, 内层和外层的参考点数目均可这样计算:  $H = \binom{M+P-1}{P}$ , 其中  $M$  为目标数目,  $P$  为沿着每个目标方向的分割数。比如在一个三目标  $M=3$  的问题中, 参考点在一个顶点为  $(1,0,0)$ ,  $(0,1,0)$  和  $(0,0,1)$  的三角形上产生。当外层  $P=2$ , 内层  $P=1$ , 那么

参考点的个数为 9，如图 3-2 所示。

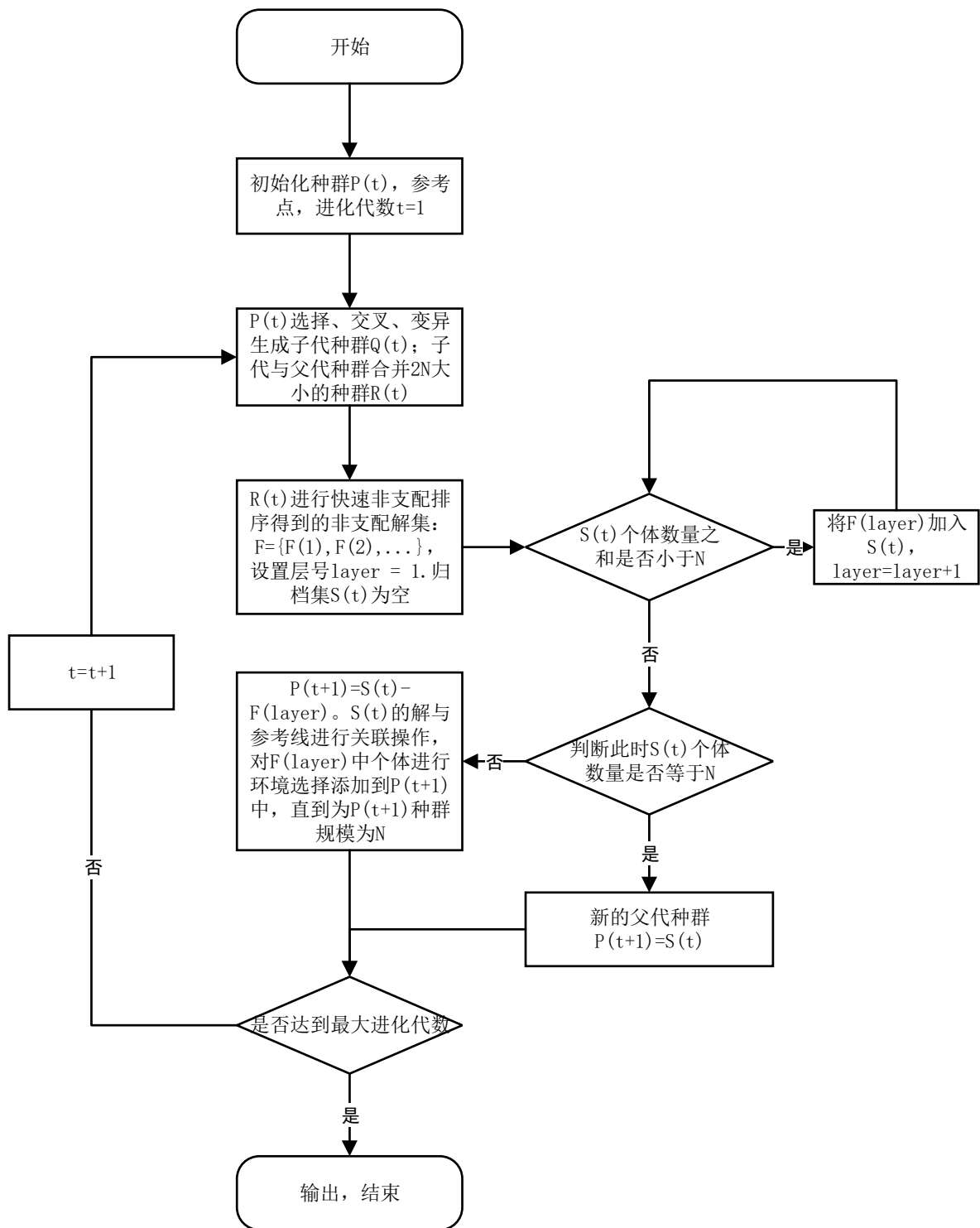


图 3-1 NSGA-III 算法流程



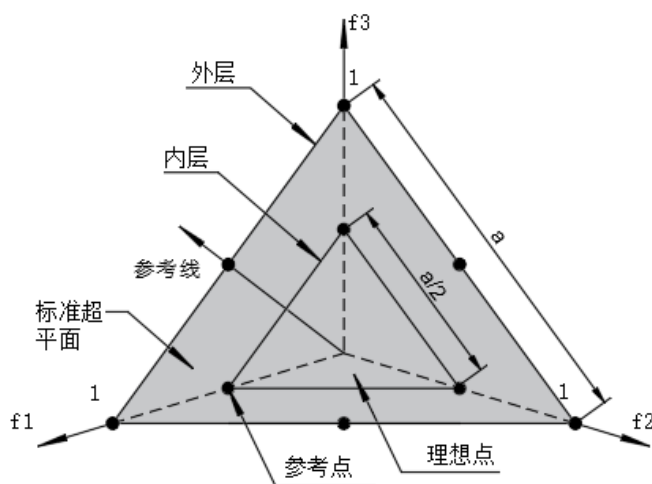


图 3-2 参考点分布

### 3.4.2 种群成员的自适应标准化

NSGA-III 每一代都使用极值点标准化，这样既可以自适应地保持多样性，也可以很好地解决具有不同尺度目标值带来的困难。

首先，利用种群  $S_t$  的所有个体在每一维目标上的最小值  $z_i^{\min}$  构成当前种群的理想点  $(z_1^{\min}, \dots, z_M^{\min})$ 。然后以理想点作为参考对所有个体的目标值进行转换，理想点变为原点，个体  $x$  的目标值转换，目标函数转换为： $f'_i(x) = f_i(x) - z_i^{\min}, i=1, 2, \dots, M$ 。

然后，计算每一维目标轴上的极值点。固定一个目标方向，则该方向向量  $w_i=1$ ，其他方向的权重设为  $10^{-6}$ （无限接近 0）， $w=(1, 10^{-6}, \dots)$ ，将令式（3-2）值最小的解作为  $i$  目标轴方向上的极值点。

$$\text{ASF}(x, w) = \max_{i=1}^M f'_i(x) / w_i, \text{ for } x \in S_t \quad (3-2)$$

最后，利用得到  $M$  个极值点，它们组成了线性超平面，可计算出各目标方向上的截距  $a_i$ ，结合  $f'_i(x)$  可计算出标准化后的超平面。式（3-3）利用截距、理想点进行目标函数标准化：

$$f_i^n(x) = \frac{f'_i(x)}{a_i - z_i^{\min}} = \frac{f'_i(x) - z_i^{\min}}{a_i - z_i^{\min}}, \text{ for } i=1, 2, \dots, M \quad (3-3)$$

标准超平面在每一维上的截距都是 1，并且该超平面使得  $\sum_{i=1}^M f'_i = 1$ 。NSGA-III 中结构化的参考点已经在该超平面上，若采用带有偏好的参考点，可使用式（3-3）对各参考点进行映射。

### 3.4.3 关联操作

参考点设置完成后，对参考线进行定义，它是在目标空间中从原点出发经过参考点的射线。关联操作就是将种群中的个体分别关联到相应的参考线上。具体方法是：计算

归档集  $S_t$  的每个个体到参考线的垂直距离，然后将个体与和它最近的参考线关联起来，如图 3-3 所示。

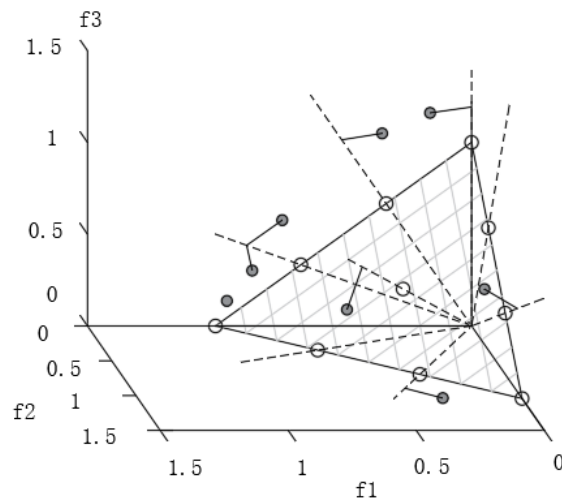


图 3-3 关联操作

### 3.4.4 个体保留操作

从临界层  $F_L$  中选择  $k$  个个体加入到下一代种群  $P_{t+1}$ 。在关联操作后，可能会出现以下情况：（1）参考点关联了一个或多个个体；（2）没有一个个体与之关联。算法记录了各个参考点在集合  $P_{t+1}=S_t/F_L$  中所关联的个体数目，将这个小生境计数记为  $\rho_j$ ，它是第  $j$  个参考点关联的个体数目。

首先选择  $\rho_j$  最小的参考点，如果有多个这样的点，随机选择一个即可。如果  $\rho_j=0$ ，表示集合  $P_{t+1}=S_t/F_L$  中没有与之关联的点，那么在临界层  $F_L$  中会出现以下两种情况：（1）在  $F_L$  中存在一个或多个个体与该参考点关联，则将距离最近的个体与之关联，并将该个体加入到  $P_{t+1}$  中；（2）若  $F_L$  中不存在个体与它关联，之后不再考虑该参考点。如果  $\rho_j \geq 1$ ，表示  $P_{t+1}=S_t/F_L$  中已经有一个个体与该参考点关联，如果  $F_L$  中有个体与之关联，则随机选择一个，将该个体加入到  $P_{t+1}$  中。重复以上操作直到下一代种群  $P_{t+1}$  的大小等于  $N$ 。

## 3.5 本章小结

本章首先介绍了多目标优化问题的数学描述以及多目标进化算法的基本流程，根据分类对多目标进化算法进行了阐述，介绍了一些比较经典的算法，如：NSGA-II 和 MOEA/D，分析了各个算法的优缺点，最后对 NSGA-III 算法进行了详细的介绍。

## 4 投资组合模型改进与求解

实际投资市场中，均值-方差模型的各项基本假设具有局限性，是不能完全满足的。目前对经典模型的扩展主要集中在三个方面：（1）简化输入数据的类型和数量；（2）引入其他新的目标（大部分是风险度量指标）；（3）融入额外的实际约束。最广泛使用的是预算约束，其次是上下限约束以及基数约束。基数约束对组合中资产数量加以限制，上下限约束限制了组合中每种资产的投资比例<sup>[68]</sup>。这些约束反映了真实的投资组合决策过程。作为风险度量，方差主要刻画了投资组合收益的波动性，但忽略了收益率分布的非对称性。随着越来越多金融衍生产品的出现，人们普遍认为，实际股票数据并不遵循多变量正态分布<sup>[69]</sup>。投资者偏好投资组合存在正偏度，并且在一个时期内为正偏度的资产可能会在下一个时期仍然保持正偏度<sup>[70]</sup>。目前除均值和方差之外，一些学者将刻画收益非对称性的偏度作为选择有效投资组合的标准融入到投资组合模型中。

为使均值-方差模型更好地适用于实际投资情况，本文尝试在马科维茨均值-方差模型基础上，引入基数约束，增加偏度作为第三个目标，利用函数链神经网络 FLNN 对股票收益率进行预测，替代历史收益率，构建一个基于预测的带基数约束的均值-方差-偏度投资组合优化模型 PBMVSCCPO（prediction based mean-variable-skewness cardinality constrained portfolio optimization）。

### 4.1 股票投资组合模型

#### 4.1.1 基于 PSO-FLNN 的股票收益预测

利用历史收益率数据，使用低复杂度的函数链神经网络（functional link neural network, FLNN）直接预测期望收益，利用粒子群优化（PSO）的方法训练 FLNN，更新网络权重。

##### （1）函数链神经网络 FLNN

FLNN 是 Pao<sup>[71]</sup>提出的一种没有隐层的单层神经网络，即输入变量以三角、指数函数，幂级数或切比雪夫型等非线性方式进行扩展。与多层感知器相比，减少了网络结构所需的层数，计算复杂度更低，可以形成任意复杂的决策区域，较好地解决高度非线性问题。基于 Chebyshev 扩展的 FLNN 预测结构如图 4-1 所示。

切比雪夫多项式可由式（4-1）递归生成：

$$\begin{aligned} T_{n+1} &= 2xT_n(x) - T_{n-1}(x) \\ T_0(x) &= 1, \quad T_1(x) = x, \quad T_2(x) = 2x^2 - 1, \quad T_3(x) = 4x^3 - 3x, \\ T_4(x) &= 8x^4 - 8x^2 + 1, \quad T_5(x) = 16x^5 - 20x^3 + 5x \end{aligned} \quad (4-1)$$

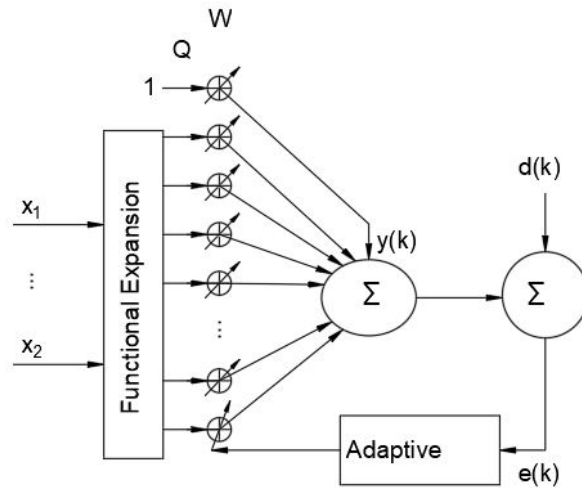


图 4-1 FLNN 网络结构

输入样本向量  $U$  中有  $e$  个元素，经过  $Z$  次切比雪夫非线性展开形成  $Q$  点向量。包括偏置后，拓展后的元素总数  $Q = Z \times e + 1$ 。

第  $k$  次迭代，设该向量的第  $q$  个元素表示为  $s_q$ ，对应的权重向量表示为  $w_q(k)$ ， $1 \leq q \leq Q$ ，瞬时输出  $y(k) = \sum_{q=1}^Q s_q w_q(k)$ ，误差项  $e(k) = d(k) - y(k)$ ，其中  $d(k)$  是从历史数据中获取的期望值，通过迭代更新权重来学习 FLNN 模型进而减少误差。

## (2) 利用粒子群算法训练 FLNN

FLNN 的学习算法大致分为两类：(1) 基于梯度的算法；(2) 无梯度的算法。本文采用无梯度的 PSO 算法。优化权重时，粒子的位置向量为 FLNN 的权重向量，初始值为 0 到 1 之间的随机数，误差函数为均方误差 MSE，粒子的速度和位置利用更新方程进行修改。基于 PSO 的 FLNN 模型预测股票周收益的基本步骤为：

① 收集并标准化特定股票的每周收益数据。FLNN 结构的输入为该股票特定周的收益的标准化值对应的  $D$  个统计特征，例如前 10 周收益的众数，中位数和移动平均值，以上所有特征被视为一项。利用该方式从该股票的周收益向量中生成  $M$  项， $N$  个用于训练， $M-N$  个用于测试模型。第  $n$  项表示为  $X_n = [x_{n1}, x_{n2}, \dots, x_{nD}]^T$ 。

② 输入项的特征数为  $D$ 。在  $D$  个变量上使用切比雪夫非线性扩展 5 次， $X_n = [x_{n1}, x_{n2}, \dots, x_{n5D}]^T$ 。

③ 粒子  $i$  的位置是由  $5D$  个随机权重组成，表示为  $W_i = [w_{i1}, w_{i2}, \dots, w_{id}, \dots, w_{i5D}]^T$ ，对应的速度矢量表示为  $V_i = [v_{i1}, v_{i2}, \dots, v_{id}, \dots, v_{iD}]^T$ 。最初，第  $i$  个粒子的初始最佳权重向量与  $W_i$  的初始值相同，表示为  $W_{il} = [w_{il1}, w_{il2}, \dots, w_{ild}, \dots, w_{il5D}]^T$ ，种群粒子初始化后，在搜索空间移动以寻找最佳位置。

④ 计算 FLNN 结构的输出  $y_i(n) = X_n^T W_i$ ，将该输出与对应的期望收益进行比较以产生误差  $e_i(n) = d(n) - y_i(n)$ 。

⑤ 得到  $N$  个样本的误差后，计算所有粒子的适应度函数，即均方误差 MSE，第  $i$

个粒子的均方误差表示为  $MSE(i) = \sum_{n=1}^N \frac{e_i(n)}{N}$ 。

⑥ 取 MSE 最小的粒子作为最佳粒子，其对应的权重向量为全局最佳，定义为  $W_g = [w_{g1}, w_{g2}, \dots, w_{gd}, \dots, w_{g5D}]^T$ 。

⑦ 下一次搜索的第  $i$  个粒子的第  $k$  个权重的速度和位置更新如式 (4-2)：

$$\begin{aligned} V_{id}(k+1) &= \psi v_{id}(k) + C_1 r_1 (w_{ild}(k) - w_{id}(k)) + C_2 r_2 (w_{gd}(k) - w_{id}(k)) \\ W_{id}(k+1) &= W_{id}(k) + V_{id}(k+1) \end{aligned} \quad (4-2)$$

其中  $\psi$  为惯性因子， $C_1$  和  $C_2$  为学习因子， $r_1$  和  $r_2$  为 [0,1] 间均匀分布的随机数。

⑧ 重复步骤④~⑧，达到最大迭代次数时截止。

#### 4.1.2 基于预测的 M-V 投资组合模型

使用预测收益替代历史数据，基于预测的均值-方差投资组合优化 PBMVPO (prediction based mean-variable portfolio optimization) 模型，如式 (4-3) 所示：

$$\begin{aligned} \max \quad R(\mathbf{x}) &= \mathbf{X}^T \bar{\mathbf{R}} = \sum_i^n x_i \bar{R}_i \\ \min \quad V(\mathbf{x}) &= \mathbf{X}^T \mathbf{V} \mathbf{X} = \sum_{i=1}^n \sum_{j=1}^n x_i x_j \sigma_{ij} \\ \text{s.t.} \quad & \sum_{i=1}^n x_i = 1 \\ & l_i \leq x_i \leq u_i, i = 1, 2, 3, \dots, n \end{aligned} \quad (4-3)$$

$n$  是可用资产数目； $X$  是各个目标达到最佳权衡时投入到各种资产上的投资比例构成的权重向量； $x_i$  是第  $i$  个资产上的投资比例； $\bar{R}_i$  是第  $i$  个资产的期望收益率； $\sigma_{ij}$  是资产  $i$  和资产  $j$  之间的收益率协方差； $R(x)$  是投资组合  $x$  的期望收益率； $V(x)$  是投资组合  $x$  的收益率方差； $l_i$  和  $u_i$  分别是资产  $i$  的投资占投资总额的最低和最高比例。模型的目标为最大化投资组合的期望收益与最小化投资组合的风险，约束为预算约束和上下限约束，前者满足每个资产上的投资比例之和为 1，后者保证了每个资产的最低和最高投资比例。

该模型为两目标优化问题，目的是找到所有可以被称为有效投资组合的非支配解。假设有两个投资组合  $\mathbf{x}^A$  和  $\mathbf{x}^B$ ，当  $R_A \geq R_B$ ， $V_A \leq V_B$ ，且至少有一个是不等式，这时可以说  $\mathbf{x}^A$  支配  $\mathbf{x}^B$ （表示为  $\mathbf{x}^A \succ \mathbf{x}^B$ ）。

#### 4.1.3 基于预测的三目标投资组合模型

随着偏度作为一个新的目标引入，增添基数约束和上下限约束，以及使用预测收益替代历史数据，M-V 模型变为基于预测的带基数约束的均值-方差-偏度投资组合优化 PBMVSCCPO 模型，如式 (4-4) 所示：

$$\begin{aligned}
\max \quad R(\mathbf{x}) &= \mathbf{X}^T \bar{\mathbf{R}} = \sum_i^n x_i \bar{R}_i \\
\min \quad V(\mathbf{x}) &= \mathbf{X}^T \mathbf{V} \mathbf{X} = \sum_{i=1}^n \sum_{j=1}^n x_i x_j \sigma_{ij} \\
\max \quad S(\mathbf{x}) &= E\left(\mathbf{X}^T (\mathbf{R} - \bar{\mathbf{R}})\right)^2 = \sum_{i=1}^n x_i^3 s_i^3 + 3 \sum_{i=1}^n \left( \sum_{j=1}^n x_i^2 x_j s_{ij} + \sum_{j=1}^n x_i x_j^2 s_{ij} \right) \quad (i \neq j) \\
\text{s.t} \quad & \\
& \sum_{i=1}^n x_i = 1 \\
& \sum_{i=1}^n K_{\min} \leq \theta_i \leq K_{\max} \\
& l_i \theta_i \leq x_i \leq u_i \theta_i, i = 1, 2, 3, \dots, n \\
& \theta_i \in \{0, 1\}, i = 1, 2, \dots, n
\end{aligned} \tag{4-4}$$

$n$  是可用资产数目； $\mathbf{X}$  是在各个目标达到最佳权衡时投入到各种资产上的投资比例构成的权重向量； $x_i$  是第  $i$  个资产上的投资比例； $\bar{R}_i$  是第  $i$  个资产的期望收益率； $\sigma_{ij}$  是资产  $i$  和资产  $j$  之间的收益率协方差； $s_i^3$ ， $s_{ii}$  和  $s_{ij}$  是偏度和斜偏度； $R(\mathbf{x})$  是投资组合  $\mathbf{x}$  的期望收益率； $V(\mathbf{x})$  是投资组合  $\mathbf{x}$  的收益率方差； $S(\mathbf{x})$  是投资组合  $\mathbf{x}$  的偏度。 $\theta_i$  指定某资产  $i$  是否存在于投资组合中。 $K_{\min}$  和  $K_{\max}$  是在投资组合中允许存在的最低和最高资产数目； $l_i$  和  $u_i$  分别是资产  $i$  的投资占投资总额的最低和最高比例。模型的优化目标为最小化投资组合的风险，最大化投资组合的期望收益，最大化投资组合偏度。约束为预算约束，上下限约束和基数约束，基数约束保证了投资组合中资产个数在一定范围内，大量经验研究表明拥有 6 至 15 种股票即可实现实质性的多元化收益，Tang<sup>[72]</sup>表示将  $K_{\max}$  设置为 10 即可。

该模型为三目标优化问题，目的是找到所有可以被称为有效投资组合的非支配的解，假设有两个投资组合  $\mathbf{x}^A$  和  $\mathbf{x}^B$ ，当  $R_A \geq R_B$ ， $V_A \leq V_B$ ， $S_A \geq S_B$ ，且至少有一个是不等式，这时可以说  $\mathbf{x}^A$  支配  $\mathbf{x}^B$ ，表示为  $\mathbf{x}^A \succ \mathbf{x}^B$ 。

## 4.2 利用多目标进化算法求解投资组合模型

多目标进化算法可用于求解多目标优化问题，在第 3 章分别对 NSGA-III、NSGA-II 和 MOEA/D 进行了介绍，虽然它们各自具有不同的特点，但是算法步骤大体相同，都具有选择，交叉和变异等进化操作。将 M-V、PBMVPO 和 PBMVSCCPO 模型简称为经典模型 1、改进模型 2 和改进模型 3，利用以上三种算法进行求解。

### 4.2.1 相关算法

NSGA-II 利用快速非支配排序对种群进行分层定级，采用基于 Pareto 的锦标赛选择机制选择个体进入配对池，交叉采用模拟二进制交叉（Simulated Binary Crossover，

SBX)，变异采用多项式变异（Polynomial Mutation，PM），环境选择采用拥挤比较操作。算法主要参数有：种群大小；交叉率和交叉分布指数；变异率和变异分布指数。

NSGA-III 同样采用快速非支配排序，采用随机选择的方法选择个体进入配对池，交叉和变异操作同 NSGA-II，环境选择利用提前定义的参考线进行。算法主要参数有：种群大小；交叉率和交叉分布指数；变异率和变异分布指数。

MOEA/D 将多目标问题分解为单目标进行求解，采用差分进化（Differential Evolution，DE）进行交叉，采用 PM 进行变异，主要参数有：种群大小；DE 的交叉率和步长；PM 的变异率和分布指数；邻域大小；交配可能性  $\delta$ ；种群中个体更新最大数量  $\eta$ 。

本文实验平台在开源 MOEA 框架基础上进行开发形成的，它是一款基于 Python 开发的多目标进化算法框架，方便进行二次开发。与当前的 PyGMO，DEAP 等框架相比，既提供了在大量广泛使用的多目标进化算法，如 NSGA-II，NSGA-III，MOEA/D，IBEA，SPEA2 等，也具有针对多目标优化性能进行评估的分析工具。框架主页见：<https://github.com/Project-Platypus/Platypus>。下面对该 MOEA 框架中的核心类进行说明，如表 4-1 所示。

表 4-1 多目标进化算法框架核心类

核心类	说明
Algorithm	算法的抽象类
Problem	定义多目标优化问题的类
Generator	进行种群初始化的抽象类
Selector	配对选择抽象类
Variator	进行交叉和变异操作的抽象类
TerminationCondition	定义截止条件的抽象类
Constraint	定义约束的类
Solution	表示问题解的类
Dominance	比较两个解之间支配关系的抽象类
Archive	保留非支配解的类
FitnessEvaluator	评估适应度的抽象类
Indicator	性能指标抽象类

### 4.2.2 性能评价指标

本文使用超体积 Hypervolume（HV）<sup>[73]</sup>指标和 epsilon<sup>[38]</sup>指标以及 set coverage<sup>[34]</sup>（C-metric）指标。

Hypervolume（HV）指标符合集合偏好关系，能够衡量算法在识别沿着完整 Pareto 前沿解的表现。解集 P 的超体积可以表示为 HV(P)，式（4-5）计算了产生的解所支配的目标空间的体积，空间边界为参考点  $\boldsymbol{r} = (r_1, \dots, r_m)^T$ ，该点被解集中的所有解所支配。多个算法进行对比时，参考点是可由所有算法运行中得到的全部解在每个目标方向上最差的值组成。超体积 HV 是一种综合度量指标，其值越高解集质量越好。

$$HV(P) = \text{Volume}\left(\bigcup_{F \in P} [f_1, r_1] \times \dots [f_M, r_m]\right) \tag{4-5}$$

epsilon 指标给出了近似解集 A 比另一个近似解集 B 更差的因子  $I_{\epsilon}(A, B)$ ，本文使用 additive 版本 epsilon 指标  $I_{\epsilon+}$ ，该指标不仅可以度量近似程度，也可以很好地度量多样性，参考集通常是真实或者是已知最好的有效前沿。 $I_{\epsilon+}$  指标越小越好，如果近似解集正好就是参考集，则该指标为 0。大部分实际问题，由于不知真实的非支配前沿，可使用针对特定问题的各个算法得到非支配解的集合作为参考集合，如式（4-6）所示：

$$I_{\epsilon+}(A, B) = \inf_{\epsilon \in R} \left\{ \forall z^2 \in B \exists z^1 \in A : z^1 \succ_{\epsilon+} z^2 \right\} \quad (4-6)$$

$$z^1 \succ_{\epsilon+} z^2 \text{ if and only if } \forall 1 \leq i \leq n : z^1_i \leq \epsilon + z^2_i$$

C-metric 指标，对于两个近似解集 A 和 B， $C(A, B)$  是 B 被 A 中至少一个解支配的解占 B 中解个数的百分比。 $C(A, B)=100$  表示 B 中所有的解都被 A 中至少一个解支配。 $C(A, B)=0$  表示 B 中没有解被 A 中的解支配，如式（4-7）所示：

$$C(A, B) = \frac{|\{u \in B \mid \exists v \in A : v \text{ dominates } u\}|}{|B|} \quad (4-7)$$

Wilcoxon 符号秩检验<sup>[74]</sup>是检测两种算法之间差异的常用工具，本文利用它对定制化的 NSGA-III 算法和其他算法进行统计分析。实验中，p-value<sup>[74]</sup>由 Scipy 软件包计算，显著性水平  $\alpha$  设定为 0.05， $\alpha < 0.05$  表示两种算法之间存在显著差异。

## 4.3 实证分析

### 4.3.1 实验说明

#### （1）实验数据集

本文利用 Tushare 金融数据接口采集国内市场的真实股票周线行情数据。由于部分数据丢失，经过预处理，最终选取了 100 种股票的真实数据。处理后的数据内容主要包括：股票代码，以及 2015 年 10 月 9 日至 2019 年 9 月 30 日各股票周收盘价数据，图 4-2 为数据导入 excel 中的截图，其由股票代码及其各周收盘价组成。

	A	B	C	D	E	F	G	H	I	J	K
1		0	1	2	3	4	5	6	7	8	9
2	0	600033.SH	3.07	3.08	3.13	3.16	3.12	3.08	3.1	3.06	3.05
3	1	600036.SH	34.75	35.2	34.51	35.29	35.89	34.28	36.27	35.16	34.99
4	2	600048.SH	14.3	14.24	14.58	14.64	13.77	13.45	14.47	13.85	13.51
5	3	600056.SH	12.93	13.03	13.4	13.72	13.6	13.14	13.85	13.35	12.66
6	4	600066.SH	13.9	13.94	14.49	15.45	15.64	14.53	13.37	13.35	12.65
7	5	600089.SH	6.5	6.52	6.68	6.83	6.8	6.58	6.7	6.58	6.58
8	6	600111.SH	10.87	10.75	11.68	11.84	11.67	11.62	11.57	11.45	11.78
9	7	600121.SH	2.35	2.4	2.47	2.57	2.49	2.41	2.42	2.36	2.29
10	8	600123.SH	6.51	6.54	6.71	6.88	6.78	6.45	6.58	6.43	6.43

图 4-2 部分数据

#### （2）实验环境

硬件环境：Win10 家庭版 64 位操作系统、CPU 处理器（Intel(R)Core(TM)i5-6200U）、8GB 内存。



软件环境：Python，Pycharm-2018.2.1，用于进行数据采集与处理，算法编程与实验；Matlab-2014a，主要用来进行数据绘图。

### （3）实验总体思路

实验分为经典模型和改进模型实验两部分。前者利用 NSGA-III，NSGA-II 和 MOEA/D 三种算法求解经典模型 1，对比分析各算法得到的 Pareto 解集的质量，确定最优参数组合，验证多目标进化算法求解投资组合问题的有效性。后者利用三种算法求解改进模型 2 和改进模型 3，对比分析改进模型 2 和经典模型 1 的实验结果，验证改进模型 2 的可行性，对于改进模型 3，因其涉及到更实际的问题，其不与前两个模型得到的结果进行对比。

## 4.3.2 经典模型实验

### （1）最优前沿

经典模型 1 在 2.2.2 小节已经进行了描述，针对该模型，利用 Cplex 求解最优解。目标为最小化组合风险，约束随着收益均匀变化，求出其对应的组合风险值，最优前沿如图 4-3 所示。

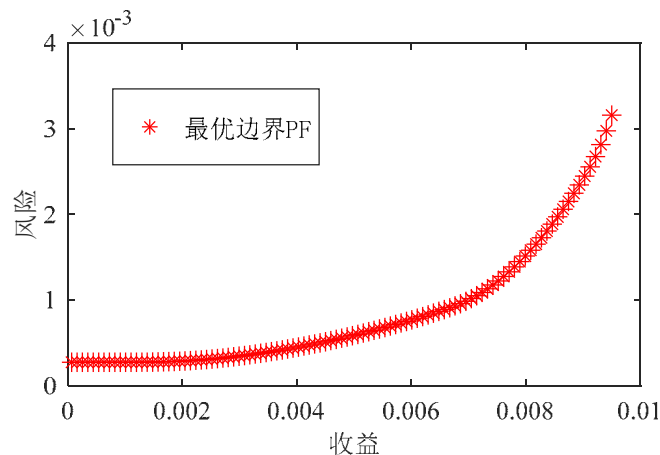


图 4-3 最优前沿

### （2）NSGA-III 算法实验

NSGA-III 算法实验参数说明：种群采用随机初始化的方法，初始种群大小默认为 100，种群大小以及最大迭代次数根据实验需要调整。模拟二进制交叉 SBX 的交叉概率为 0.9，分布度指数为 20；多项式变异 PM 的变异概率 0.1，分布度指数为 20。设置最大进化代数为 Gen = 100 和 500；种群大小 Pop 依次为 100，300，500。对于经典模型 1 中的预算约束，在初始化种群以及对个体进行交叉变异的过程中，对其进行权重归一化操作来满足约束，防止陷入局部最优。

由图 4-4 可知，进化代数为 100 的情况下，NSGA-III 在最优边界中间部分有一定的收敛性，但没有取得最优前沿前端和末端的解。随着代数提升至 500 次，解集质量提升明显，收敛性更好，多样性也得到了很大的提升，有近似最优解分布在最优前沿末端，但收敛性还有提升空间。在进化代数相同的情况下，种群大小为 100 时，解集质量更

好。

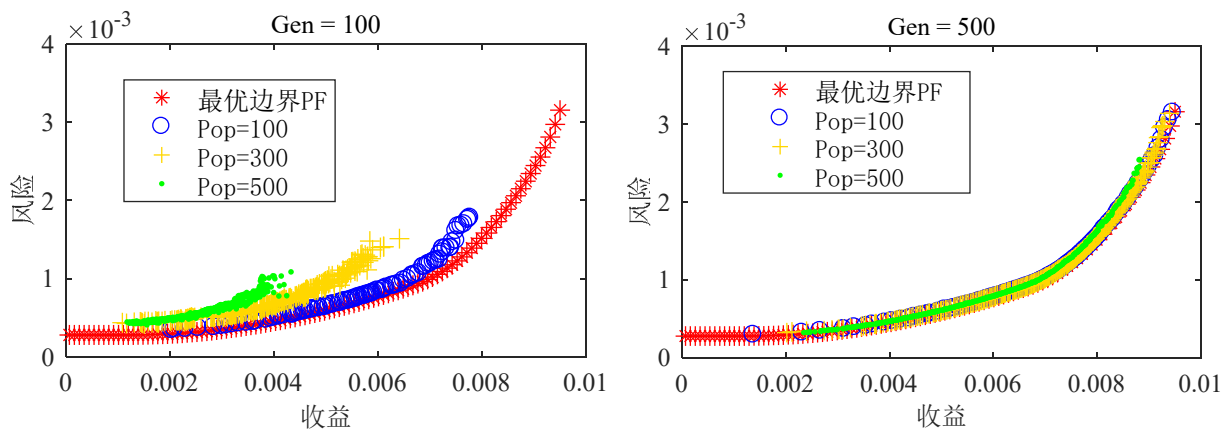


图 4-4 NSGA-III 算法经典模型 1 实验结果

(3) NSGA-II 算法实验

NSGA-II 算法参数说明：初始种群采用随机初始化方法，初始种群大小默认为 100，种群大小以及最大迭代次数根据实验需要调整。模拟二进制交叉以及多项式变异的参数同 NSGA-III。设置最大进化代数 100 和 500，种群大小依次为 100，300，500，采用权重归一化操作满足约束。

由图 4-5 可知，对比迭代 100 次，迭代 500 次得到的近似最优解集虽然多样性和收敛性均增强了，但也未完全覆盖最优前沿末端，且末端收敛性相对于中部较差。当迭代次数相同时，种群数目为 100 效果更好。与 NSGA-III 相比，其收敛性和多样性相对弱一些。

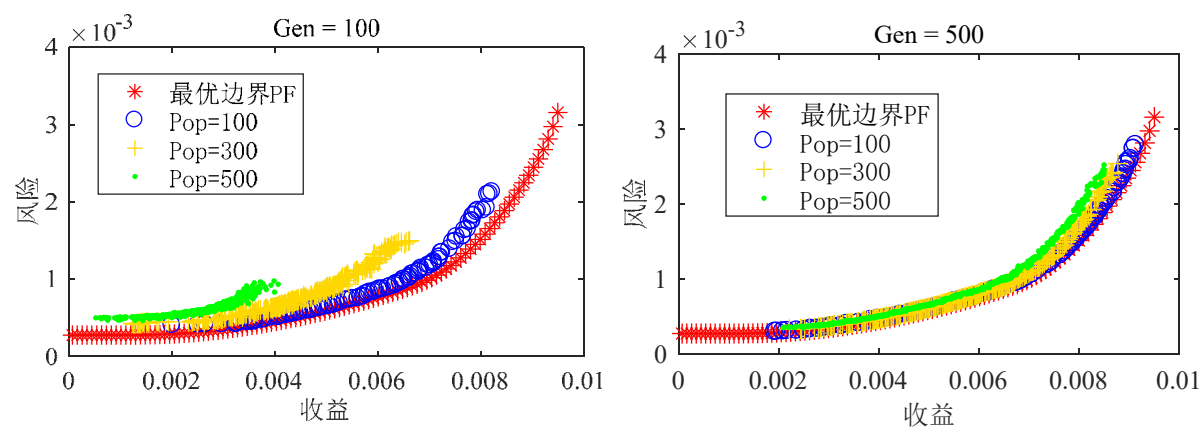


图 4-5 NSGA-II 经典模型 1 实验结果

(4) MOEA/D 算法实验

MOEA/D 算法参数说明：初始种群采用随机初始化方法，设种群规模为 N，N 初始值默认为 100，种群大小以及最大迭代次数根据实验需要调整。差分进化 DE 的交叉概率 0.9，步长大小 0.5；多项式变异率为 1/N，分布指数为 20；邻域规模为 0.1\*N，最小为 20；delta = 0.8；eta = 0.01\*N，最小为 2；设置最大进化代数 100 和 500，种群大小

依次为 100, 300, 500。采用权重归一化满足预算约束。

由图 4-6 可知，随着迭代次数的增加，解集质量明显改善，但其收敛性以及末端分布情况仍较差，迭代次数相同的情况下，种群规模为 100 时最好。与 NSGA-III 相比，无论是收敛性还是分布性都较差。

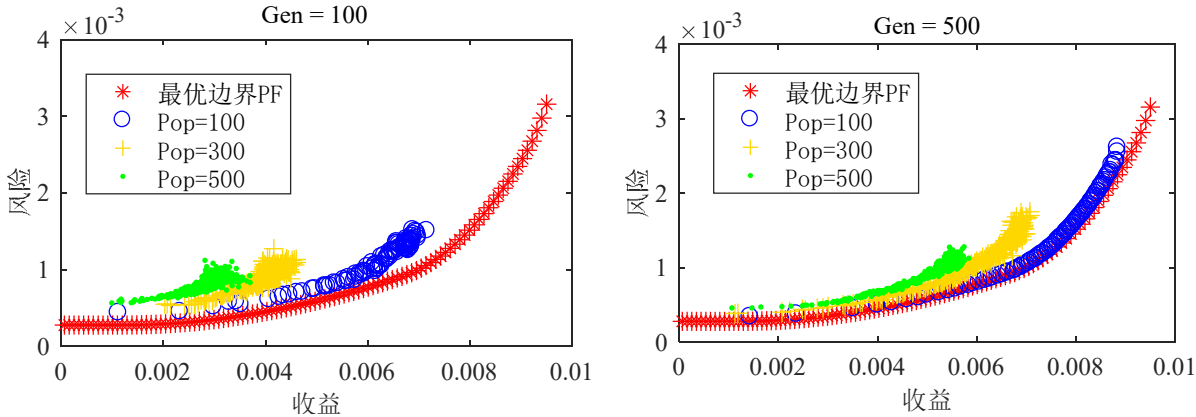


图 4-6 MOEA/D 算法经典模型 1 实验结果

以上利用实际数据集在经典模型 1 上对三种算法进行了测试，直观地了解到各个算法的应用效果和不同参数设置对解集质量的影响，在种群规模  $Pop=100$ ，最大进化代数为  $Gen=500$  时，各算法的性能相对较好。下面设定种群规模为 100，最大进化代数为 500，利用性能指标对三种算法的性能进行量化比较，给出实验结果对比图 4-7。

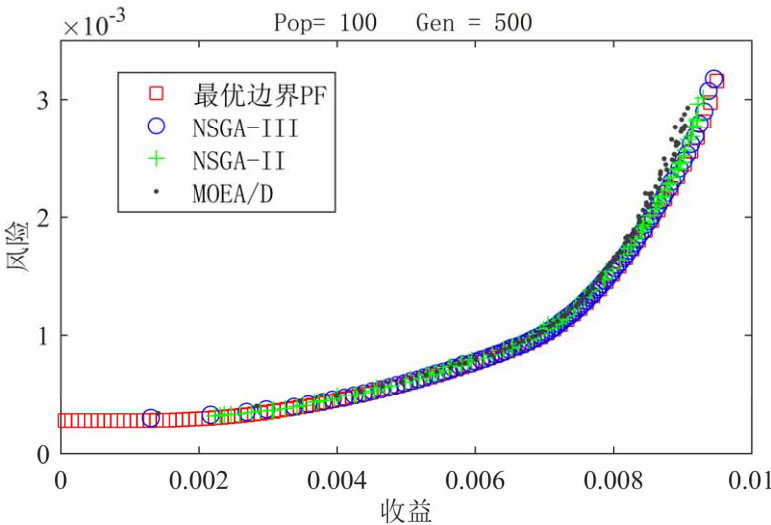


图 4-7 三算法经典模型 1 实验结果对比

由图 4-7 可知，除在组合收益和风险都非常小的最优前沿前一段，NSGA-III 在整个最优前沿 PF 上收敛性最好，最优前沿覆盖程度最高，NSGA-II 和 MOEA/D 在最优前沿高收益和高风险的一部分上没有分布。可见，NSGA-III 无论在收敛性还是分布性上，均优于其他两种算法。但三种算法对于边界左端（低风险，低收益的解）没有求出。

利用超体积 HV 指标和 Epsilon 指标综合衡量解集收敛性和多样性。各算法分别独立运行 20 次，由表 4-2 可知，相对于其他两种算法，NSGA-III 得到的超体积值更大，Epsilon 值更小，前者表示其得到的非支配解集与参考点围成的目标空间的体积最大，

后者表示其得到的非支配解集质量更接近于参考集，可见 NSGA-III 得到的解集质量更好。以上实验结果也有效地验证了多目标进化算法求解投资组合优化问题的有效性。

表 4-2 三算法 HV 指标与 Epsilon 指标

算法	HV. Mean	HV. Std.	Epsilon. Mean	Epsilon. Std.
NSGA-III	6.4298E-1	2.7124E-2	1.785E-5	5.3055E-6
NSGA-II	6.3611E-1	2.7671E-2	3.406E-5	1.2201E-5
MOEA/D	6.2576E-1	2.6108E-2	7.1754E-5	2.0475E-5

4.3.3 改进模型实验

改进模型 2 在 4.1.2 小节进行了描述，其与经典模型 1 的区别是使用基于预测的股票收益率，而不是历史数据来构建投资组合模型。利用三种算法对改进模型 2 进行求解，与经典模型 1 的求解结果进行对比分析，为解决更实际的问题提供参考。

(1) 股票周收益率预测

利用 4.1 节的 PSO-FLNN 对股票周收益率进行预测，并与真实收益率进行比较，经多次校验，惯性因子 $\psi$ 为 0.729，学习因子 $C_1$ 、 $C_2$ 均为 2，性能指标为均方误差 MSE，迭代 10000 次。以 600089.SH 和 600449.SH 两只股票为例，给出部分预测值与真实值对比数据，进行可视化比较，计算各自的均方误差值 MSE。600089.SH 的均方误差为 8.092e-06，600449.SH 的均方误差为 2.096e-06。图 4-8 是部分数值的对比结果，其由股票代码及其对应的各周收益预测值和真实值组成。两只股票周收益率预测值以及真实值的对比折线图如图 4-9、4-10 所示，可见利用 PSO-FLNN 对股票周收益率进行预测的结果与实际值拟合程度较好，虽有些许偏差，但不会造成收益率参照影响。

	A	B	C	D	E	F	G	H	I
1	600089.SH	周收益率预测值	-0.02288	-0.10991	-0.11404	0.01439	-0.11404	0.01376	0.02565
2		周收益率真实值	-0.02243	-0.09856	-0.11404	0.01596	-0.10576	0.01405	0.02656
3	600449.SH	周收益率预测值	-0.02819	-0.12484	-0.12875	0.01146	0.00391	-0.01165	0.06098
4		周收益率真实值	-0.02693	-0.12321	-0.12831	0.01051	0.00347	-0.01152	0.05944

图 4-8 600089.SH 和 600449.SH 部分预测值与真实值对比

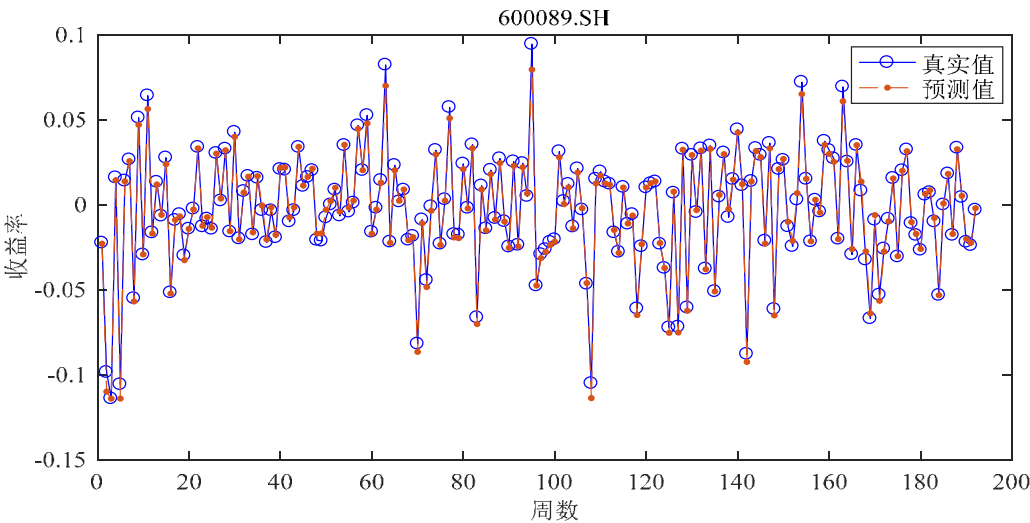


图 4-9 600089.SH 预测值与真实值对比折线图

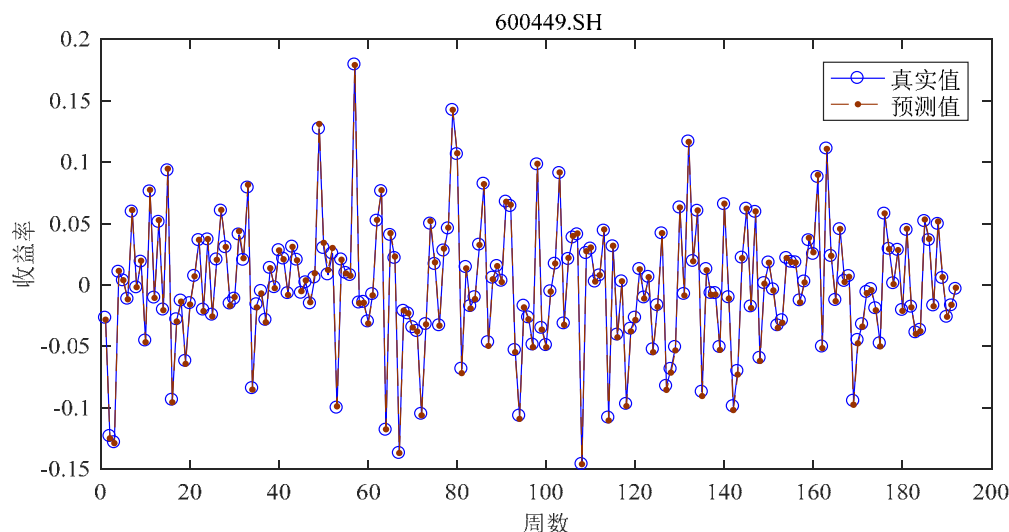


图 4-10 600449.SH 预测值与真实值对比折线图

### (2) 改进模型 2 与经典模型 1 的对比

由图 4-11 可知，使用股票的预测收益率替代历史收益率，各算法得到的最优前沿前半部分与经典模型 1 的最优边界基本重合，后半部分略微偏离，这符合实际情况，因为预测值往往和真实值存在误差。相对于使用历史数据对期望收益等指标进行量化，使用预测的未来收益估计期望收益等指标更具有实际投资意义。改进模型 2 得到的最优边界仍符合经典模型 1 最优边界的分布和趋势，且在收益相同的情况下，未低估风险。同样，NSGA-III 在改进模型 2 中的性能比其他两种算法更好。

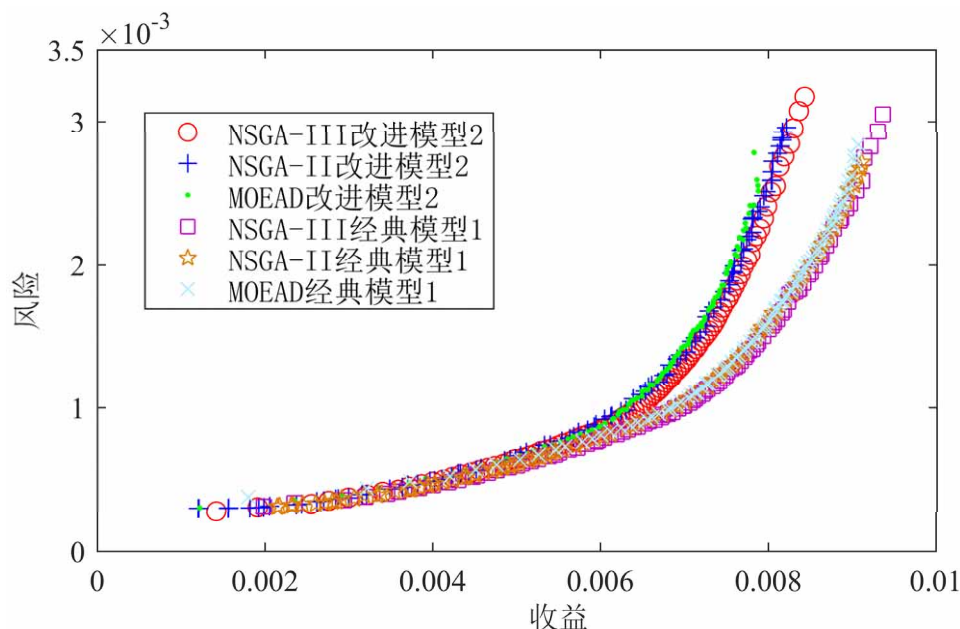


图 4-11 三算法改进模型 2 与经典模型 1 实验结果对比

### (3) 改进模型 3 基本实验

改进模型 3 涉及到更实际，更高维度的问题，其求解较为困难。利用 NSGA-III、NSGA-II 和 MOEA/D 算法对其进行基本求解。实验结果如图 4-12 所示，三种算法得到



的解集质量均一般，NSGA-III 相对于其它两种算法多样性更好一些，MOEA/D 算法得到结果最差。对结果分析发现：虽然风险水平较低，但 NSGA-III 最终解集的收益率均值为 0.21%，其收益水平仍有提高的空间。引入基数约束后，投资组合中的资产数目受到了限制，各算法在对模型的基本求解过程中，没有针对其设计有效的初始化方案和种群个体更新策略，各投资组合的资产种类初始化后便固定不变，进而对解集质量产生影响。

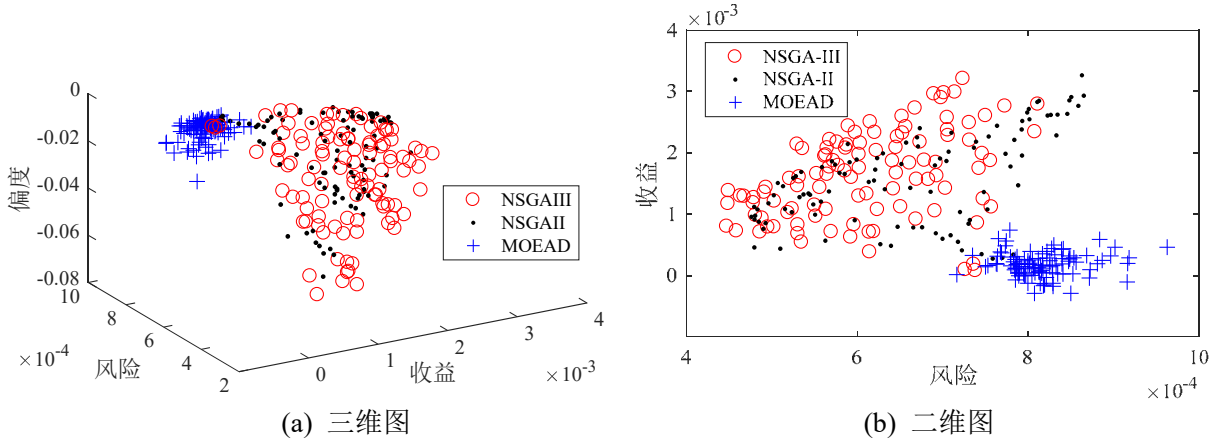


图 4-12 三种算法求解改进模型 3 实验结果

#### 4.4 实验结论

利用 NSGA-III、NSGA-II 和 MOEA/D 在实际数据集上进行了经典模型的实验，对种群大小、迭代次数等重要参数进行了调整，对比了各个算法得到的解集的质量，实验结果表明：相对于其他两种算法，NSGA-III 在整个最优前沿 PF 上分布性最好，更接近最优前沿，但三种算法对于边界左端（低风险，低收益的解）没有求出，且在收敛性和分布性仍有提高的空间。

改进模型 2 上的实验表明，使用预测收益率替代历史收益率构建投资组合模型是有效可行的，得到的最优边界仍符合经典模型 1 最优边界的分布和趋势，且相对于经典模型更具有实际投资意义。改进模型 3 考虑了更多的实际情况，影响了算法的性能，得到解集质量一般。

综上所述，多目标进化算法求解投资组合优化问题是有效的，可以找到最优边界，但在对于更实际的问题，仍然有需要改进的地方。面对解集质量较差的情况，需在算法各个进化过程中考虑收敛性因素，研究增强其搜索能力，进一步提高解集质量；基数约束的引入，限制了组合内资产的种类数，一定程度上降低了解集的质量，需研究更有效的组合内的资产更新能力，进一步提升解集的多样性和收敛性。下一章，围绕以上几个方面，对 NSGA-III 算法进行定制化改进。

#### 4.5 本章小结

本章首先阐述了基于 PSO-FLNN 的股票周收益率预测的原理及步骤，在此基础

上，对改进模型 2 和改进模型 3 的构建进行了详细的数学描述。其次介绍了多目标进化算法对模型求解的实现方法以及对算法进行定量比较的各项性能评价指标。然后进行实证分析，在实际数据集的基础上，利用 NSGA-III，NSGA-II，MOEA/D 三种算法分别对各个模型进行求解，了解了各个算法的性能特征和最优参数设置，验证了多目标进化算法求解投资组合优化问题的有效性，最后给出了实验结论和需要改进的地方。

## 5 定制化的多目标进化算法及应用

### 5.1 定制化的多目标进化算法

NSGA-III 可以有效求解高维目标优化问题，结合了基于支配和基于分解的多目标进化算法的思想，利用 Pareto 支配关系进行快速非支配排序，预先定义参考点形成参考线，在环境选择阶段，考虑距离参考线较近的个体，其多样性得到了良好的保持，收敛性仍有改进的余地。由第 4 章的实证分析可知，相对于 NSGA-II 和 MOEA/D 算法，NSGA-III 算法综合性能较好，算法框架简单，且在经典模型 1 和改进模型 2 的求解中，也获得了更好的效果。因此本文主要基于 NSGA-III 算法框架，针对更加实际的 PBMVSCCPO 模型进行定制化的改进，提出定制化的 c-NSGA-III 算法（customized NSGA-III），来提高股票投资组合的求解效率与质量。

本文针对改进模型 3 的定制化改进主要有三个方面：（1）环境选择阶段，由于 NSGA-III 使用 Pareto 支配关系促进种群朝 PF 进化，并且在选择阶段只考虑距离参考线近的个体，为了更好权衡收敛性和多样性，使用一种基于惩罚的边界交点（penalty-based boundary intersection, PBI）形式的距离函数替换垂直距离；（2）在选择父辈个体进行配对的过程中，利用模糊 C 均值聚类<sup>[75]</sup>求出各个个体的隶属度，使用一种基于隶属度的锦标赛选择替代原算法中的随机选择方法，进一步提高子代解的质量。（3）在投资组合个体表示中，引入资产信息，使用一种两阶段的交叉和变异操作替换原算法中的进化操作，加强个体更新，以弥补基数约束容易造成投资组合中资产种类无法进行更新的缺点。

c-NSGA-III 算法流程为：

输入：预先设定的参考点，基数的上下限  $K_{\min}$ 、 $K_{\max}$ ，权重的上下限  $l_i$ 、 $u_i$ ，以及种群规模  $N$ ，最大进化代数  $Gen$  等，模糊聚类簇数  $C$ ，惩罚参数  $\lambda$ ，各项进化参数等。

输出：最优种群  $P_{t+1}$ 。

- ① 种群初始化  $N$  个个体加入到  $P_t$  中，当个体违反约束，对其进行修复操作。
- ② 对  $P_t$  进行模糊聚类操作，获取隶属度矩阵，对  $P_t$  执行基于隶属度的锦标赛选择操作，以及两阶段交叉、变异操作，获得子代种群  $Q_t$ ，对违反约束的子代个体执行修复操作。
- ③ 合并  $P_t$  和  $Q_t$ ，生成合并种群  $R_t$ 。
- ④ 对  $R_t$  执行基于 PBI 距离的环境选择操作，得到下一父代  $P_{t+1}$ 。
- ⑤ 若达到终止条件，停止循环，输出结果  $P_{t+1}$ ，否则转至②。



## 5.2 选择操作

### 5.2.1 环境选择改进

环境选择阶段，使用 PBI 距离代替垂直距离。PBI 距离的形式化如式 (5-1) 所示：

$$\begin{aligned} d(x) &= d_{j,1}(x) + \lambda d_{j,2}(x), j \in \{1, 2, \dots, H\} \\ d_{j,1}(x) &= \left\| (F(x))^T w_j \right\| / \|w_j\| \\ d_{j,2}(x) &= \left\| F(x) - d_{j,1}(x) (w_j / \|w_j\|) \right\| \end{aligned} \quad (5-1)$$

其中  $d(x)$  使用 Yuan<sup>[76]</sup>提出的聚合函数形式。 $\lambda$  是预先定义的惩罚参数。第一个距离度量  $d_{j,1}$  是个体的目标点到参考线上垂足与原点的欧式距离， $d_{j,2}$  是该个体的目标点与参考线之间的垂直距离。两种度量在图 5-1 中有清晰的说明。可以看出， $d_1$  值越小，解的收敛性越好， $d_2$  值越小可保证良好的多样性，相比原始算法，PBI 距离函数引入了收敛性信息，进而平衡收敛性与多样性。

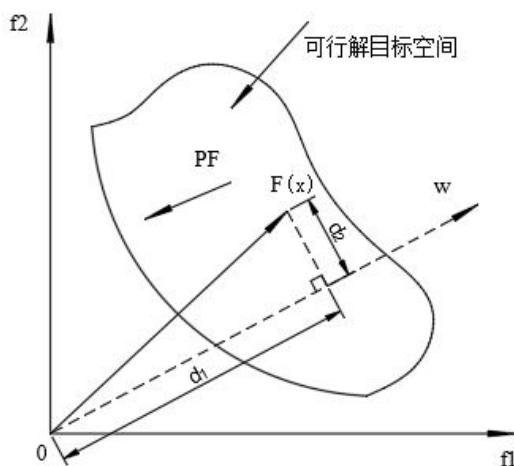


图 5-1 两种距离度量

基于 PBI 距离关系的环境选择流程为：

- ① 输入：合并种群  $R_t$ ，种群规模  $N$ ，权重向量  $W$  等
- ② 对  $R_t$  进行非支配排序得到各层：( $F_1, F_2, \dots$ )
- ③ 设归档集  $|S_i| = \emptyset, i = 1$
- ④ 进入循环：  $S_i = S_i \cup F_i$ ，  $i = i + 1$ ，直到  $|S_i| \geq N$  为止
- ⑤ 设定最后加入  $S_i$  的  $F_i$  为临界层  $F_l$
- ⑥ 如果  $|S_i| = N$
- ⑦ 那么下一代种群  $P_{t+1} = S_i$ ，输出  $P_{t+1}$
- ⑧ 否则
- ⑨ 令  $P_{t+1} = \bigcup_{j=1}^{l-1} F_j$ ，需要从临界层  $F_l$  选择出的个体数目： $k = N - |P_{t+1}|$
- ⑩  $S_i$  中个体与参考线进行关联操作
- ⑪ 计算  $S_i$  中所有个体相对于权重向量的 PBI 距离  $d(x)$

- ⑫ 在临界层  $F_l$  进行环境选择，选择  $k$  个个体进入  $P_{t+1}$ ，输出  $P_{t+1}$

### 5.2.2 配对选择改进

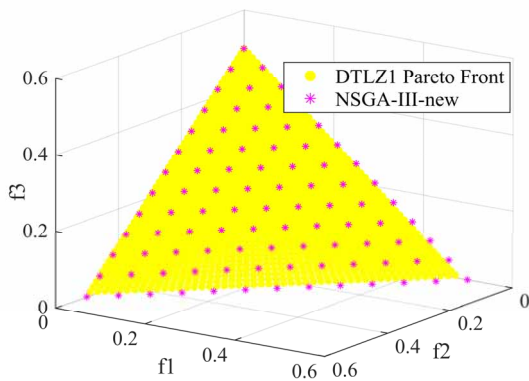
定制化的 c-NSGA-III 算法使用基于隶属度的锦标赛选择机制替代原算法中的随机选择方法，即使用种群的聚类信息（隶属度矩阵）替代原标准锦标赛选择中的 Pareto 支配关系进行父代个体的选择。同类中随机选择个体进行重组，在增强局部搜索能力的同时，一定程度上也保证了新个体的良好的多样性。该方法利用模糊 C 均值聚类对种群中  $N$  个个体聚为  $C$  类，对数据集进行模糊划分，从而获取聚类信息—隶属度矩阵。该矩阵可表示为  $U=\{u_{ij}\}$ ， $1 < i < N$ ， $1 \leq j \leq C$ ，其中  $u_{ij}$  是个体  $i$  隶属于第  $j$  类的隶属度。

c-NSGA-III 采用基于隶属度的锦标赛选择算子选择父代个体，流程为：

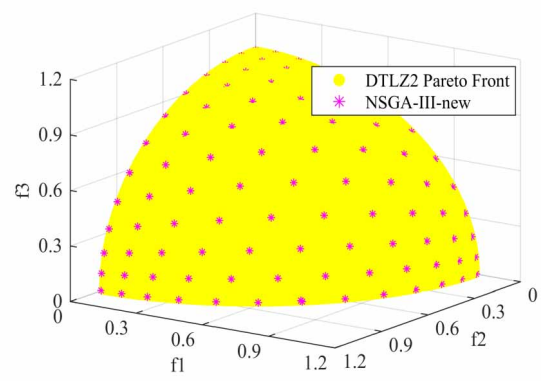
- ① 输入：隶属度矩阵  $U=\{u_{ij}\}$ ， $1 < i < N$ ， $1 \leq j \leq C$ ，父代归档集  $Q_P = \emptyset$
- ② 令  $k = 0$ ， $n = 2$ ，随机选择的个体  $x^i$ ，其归属的类为  $m$
- ③ 当  $k < 2$  时进入循环
- ④  $k = k + 1$
- ⑤ 随机从  $P_t$  选择  $n$  个个体
- ⑥ 选择父代个体  $\mathbf{x}^k = \{x^j \mid \{j = \operatorname{argmax}\{u_{1m}, \dots, u_{nm}\}\}, j = 1, \dots, n\}$
- ⑦ 更新父代归档集  $Q_P = Q_P \cup \{\mathbf{x}^k\}$
- ⑧ 结束循环，返回已选择的父代个体  $Q_P$

### 5.2.3 仿真实验

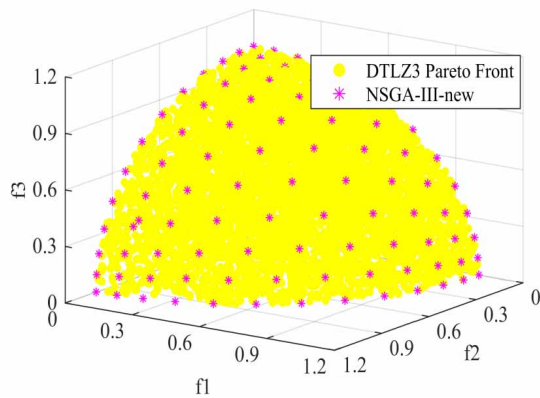
为了验证引入 PBI 距离函数以及基于隶属度的锦标赛选择的 NSGA-III-new 算法的有效性，本文将其与 NSGA-II 和 MOEA/D 以及原始 NSGA-III 算法在四个典型的多目标测试函数 DTLZ1~4<sup>[39]</sup>中进行对比实验。性能指标采用综合性指标 HV 和 Epsilon，所有算法在四个测试函数上均独立进行 20 次实验。模糊聚类中的类别数  $C$  为 10，最大迭代次数  $\maxiter=1000$ ，收敛精度  $e=0.005$ ，模糊指数  $m=2$ ，PBI 函数中惩罚参数  $\lambda$  设为 5，其余参数如第四章实验中设置。种群规模为 100，最大进化代数为 500，四个测试问题目标数均取 3，实验结果如图 5-2 所示。



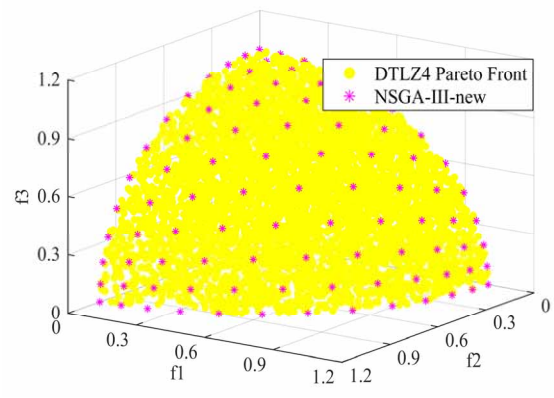
(a) DTLZ1 实验结果



(b) DTLZ2 实验结果



(c) DTLZ3 实验结果



(d) DTLZ4 实验结果

图 5-2 实验结果

可见, NSGA-III-new 算法得到的解多样性良好, 均匀地分布在 Pareto 前沿上, 说明其具有良好的优化性能。为了进一步定量比较 NSGA-III-new 与其他四种算法之间的性能差异, 在表 5-1 中统计了四种算法求解测试问题的 HV 指标和 Epsilon 指标, 20 次独立运行后的均值和标准差, HV 和 Epsilon 指标的均值, 反映了算法的收敛性和多样性, 标准差反映了算法的鲁棒性。

由表 5-1 可知, 在求解测试问题时, NSGA-III-new 算法在 HV 指标上取得 4 个最优, 在 Epsilon 指标上取得了 4 个最优值, 解集质量相对于 NSGA-III 算法有所改善, 且在所有测试问题, NSGA-III-new 性能均优于 NSGA-II 和 MOEA/D 算法。

表 5-1 HV 指标和 Epsilon 指标的均值 (标准差)

测试函数	算法	HV 指标	Epsilon 指标
DTLZ1	NSGA-III-new	7.82E-01 (1.69 E-03)	6.17E-02 (7.31E-03)
	NSGA-III	7.75E-01 (1.36E-02)	7.59E-02 (2.61E-02)
	NSGA-II	7.52E-01 (7.99E-03)	1.11E-01 (1.54E-02)
	MOEA/D	7.53E-01 (5.24E-03)	9.93E-02 (6.65E-03)
DTLZ2	NSGA-III-new	4.17E-01 (2.58E-03)	7.42E-02 (5.65E-03)
	NSGA-III	4.14E-01 (6.60E-04)	7.97E-02 (1.18E-03)
	NSGA-II	3.74E-01 (4.60E-03)	1.30E-01 (9.82E-03)
	MOEA/D	3.80E-01 (3.83E-03)	1.30E-01 (1.12E-02)
DTLZ3	NSGA-III-new	3.68E-01 (1.94E-02)	5.21E-02 (8.59E-03)
	NSGA-III	3.51E-01 (1.02E-02)	5.36E-02 (1.99E-03)
	NSGA-II	3.18E-01 (2.13E-02)	6.10E-02 (1.07E-02)
	MOEA/D	3.23E-01 (4.09E-02)	6.69E-02 (1.98E-02)
DTLZ4	NSGA-III-new	4.11E-01 (2.34E-03)	8.42E-02 (4.07E-03)
	NSGA-III	4.05E-01 (7.20E-03)	1.24E-01 (1.51E-02)
	NSGA-II	3.54E-01 (5.09E-02)	2.47E-01 (2.83E-02)
	MOEA/D	3.09E-01 (8.64E-02)	3.47E-01 (2.79E-02)

### 5.3 个体更新

在 NSGA-III-new 的基础上, 针对改进模型 3, 对个体的编码方式, 以及交叉、变异算子进行改进, 加强种群个体的更新, 进一步形成定制化的改进算法 c-NSGA-III。

本文采用一种混合表示方法, 利用实值向量  $x$  和二进制向量  $\theta$  共同定义投资组合个

体，结合基数约束，引入资产信息，从而更好地利用计算资源，式（5-2）中使用两个向量定义一个投资组合个体：

$$\begin{aligned}\theta &= \{\theta_1, \theta_2, \dots, \theta_n\} \quad \theta_i \in \{0, 1\} \quad i = 1, 2, \dots, K_{\max} \\ \mathbf{x} &= \{x_1, x_2, \dots, x_n\} \quad 0 \leq x_i \leq 1, \quad i = 1, 2, \dots, K_{\max}\end{aligned}\quad (5-2)$$

$\theta$  为二元向量，其中  $\theta_i$  决定了资产  $i$  是否在投资组合中， $\theta_i=0$  表示该资产不在组合中， $\theta_i=1$  表示该资产在投资组合中。 $\mathbf{x}$  为实值向量，其中  $x_i$  是投资在资产  $i$  上的满足上下限约束的资金比例， $x_i$  也包含了对应的资产种类信息  $a_i$ 。 $K_{\max}$  为基数约束上限，其作为投资组合个体决策变量数，本文设置  $K_{\max}=10$ ， $K_{\min}=2$ 。初始化阶段，随机从资产池选择  $K_{\max}$  个资产进入一个投资组合，然后随机产生  $K$  个满足上下限约束的权重并随机分配给组合内的各个资产，其中  $K \in [K_{\min}, K_{\max}]$ 。

对于不满足预算约束以及上下限约束的解用以下方法进行修复，如式（5-3）所示，决策变量个数  $n=K_{\max}$ 。

$$x'_i = l_i \theta_i + \frac{x_i \theta_i}{\sum_{i=1}^{K_{\max}} x_i \theta_i} \left( u_i \theta_i - \sum_{i=1}^{K_{\max}} l_i \theta_i \right) \quad (5-3)$$

本文采用一种两阶段交叉、变异算子，以 Deb 和 Tiwari<sup>[77]</sup>提出的模拟二进制交叉 SBX 和多项式变异 PM 为基础，考虑到资产池的规模大多远大于  $K_{\max}$ ，如果各个投资组合中的资产种类在初始化阶段便固定，那么可能有许多资产并未进入投资组合中，有些资产可能比较优质，这样既降低了投资组合的质量，也损害了解的多样性。两阶段交叉、变异算子允许实值解向量外某些资产替代实值解向量中某些资产，尽可能有效地开采搜索空间，对个体进行更新。本文在第二阶段的资产更新中，首先在总资产中排除投资组合内存在的资产，然后不重复地选择进入投资组合，达到资产种类更新的目的，既增加了资产种类多样性，也简化了进化操作。

多项式变异和模拟二进制交叉操作如下：

#### （1）多项式变异 PM

变异的分布度指数为  $\eta_m$ ，对于一个要进行变异的个体，其需要进行变异的实值决策变量为  $x_p$ ，资产权重的上下限为  $L$  和  $U$ ，令  $\delta_1 = \frac{x_p - L}{U - L}$   $\delta_2 = \frac{U - x_p}{U - L}$ 。

取  $[0, 1]$  之间的随机数  $r$ ，若  $r \leq 0.5$ ，那么： $\delta_q = \left[ 2r + (1 - 2r)(1 - \delta_1)^{\eta_m + 1} \right]^{\frac{1}{\eta_m + 1}}$ ，否则： $\delta_q = 1 - \left[ 2(1 - r) + 2(r - 0.5)(1 - \delta_2)^{\eta_m + 1} \right]^{\frac{1}{\eta_m + 1}}$ 。

更新： $x_c = x_p + \delta_q (U - L)$ ，二进制变量则采用位翻转进行变异。

#### （2）模拟二进制交叉 SBX

交叉的分布度指数为  $\eta_c$ ，两个进行交叉的个体，设需要进行交叉的实值决策变量为  $x_p^{(1)}$  和  $x_p^{(2)}$ ，各资产权重的上下限为  $L$  和  $U$ 。令  $\beta = 1 + \frac{2}{x_p^{(2)} - x_p^{(1)}} \min[(x_p^{(1)} - L), (U - x_p^{(2)})]$ ，

$\alpha=2-\beta^{-(\eta_c+1)}$ 。

取 $[0,1]$ 之间的随机数 $u$ ，若 $u \leq \frac{1}{a}$ ，那么 $\beta_a = (au)^{1/(\eta_c+1)}$ ，否则 $\beta_a = (\frac{1}{2-au})^{1/(\eta_c+1)}$ 。

更新： $x_c^{(1)} = 0.5[(x_p^{(1)} + x_p^{(2)}) - \beta_a |x_p^{(2)} - x_p^{(1)}|]$ ， $x_c^{(2)} = 0.5[(x_p^{(1)} + x_p^{(2)}) + \beta_a |x_p^{(2)} - x_p^{(1)}|]$ 。二进制变量进行单点交叉。

以 PM 和 SBX 为基础，两阶段变异算子和两阶段交叉算子流程分别如图 5-3 和图 5-4 所示。

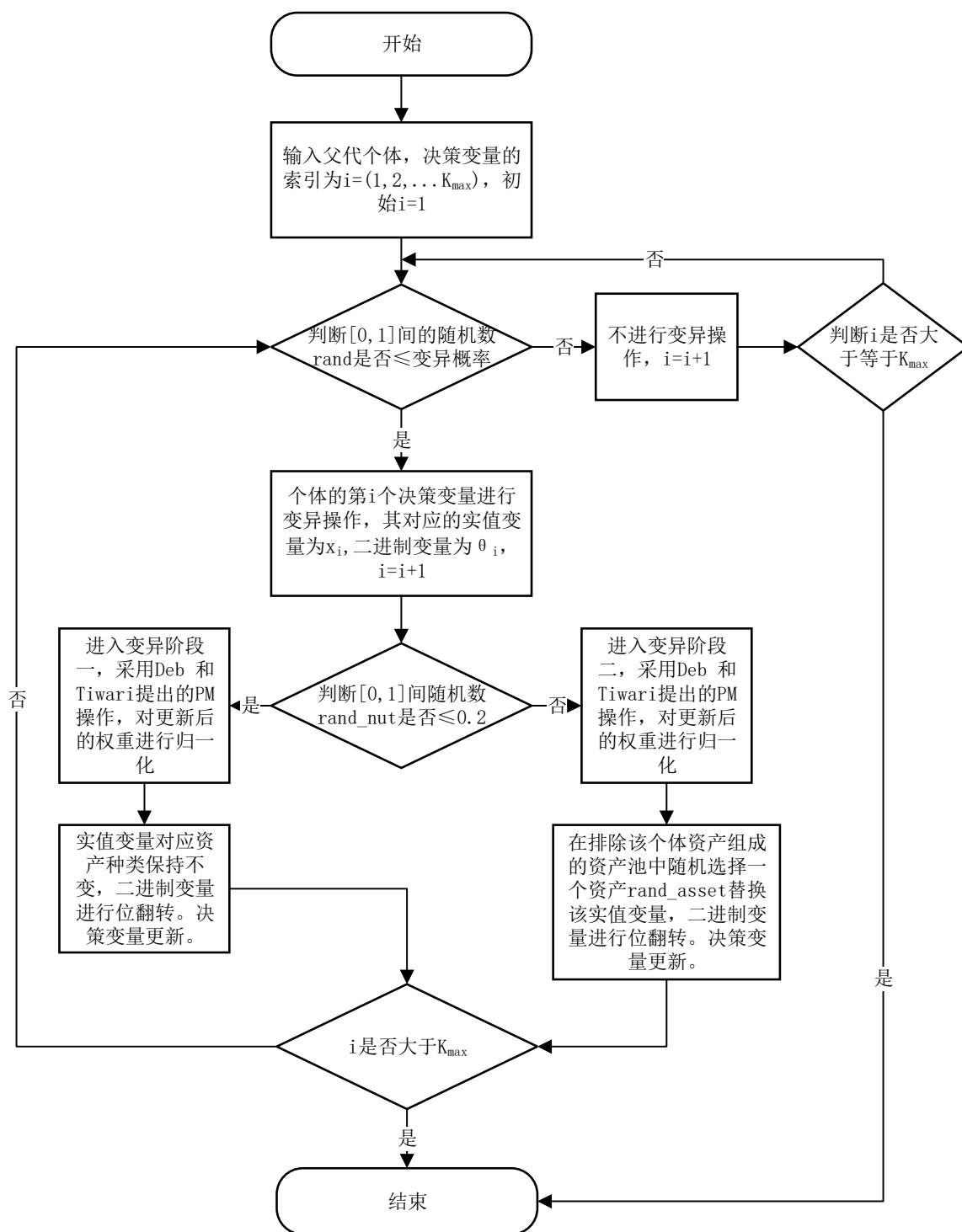


图 5-3 两阶段变异算子流程图

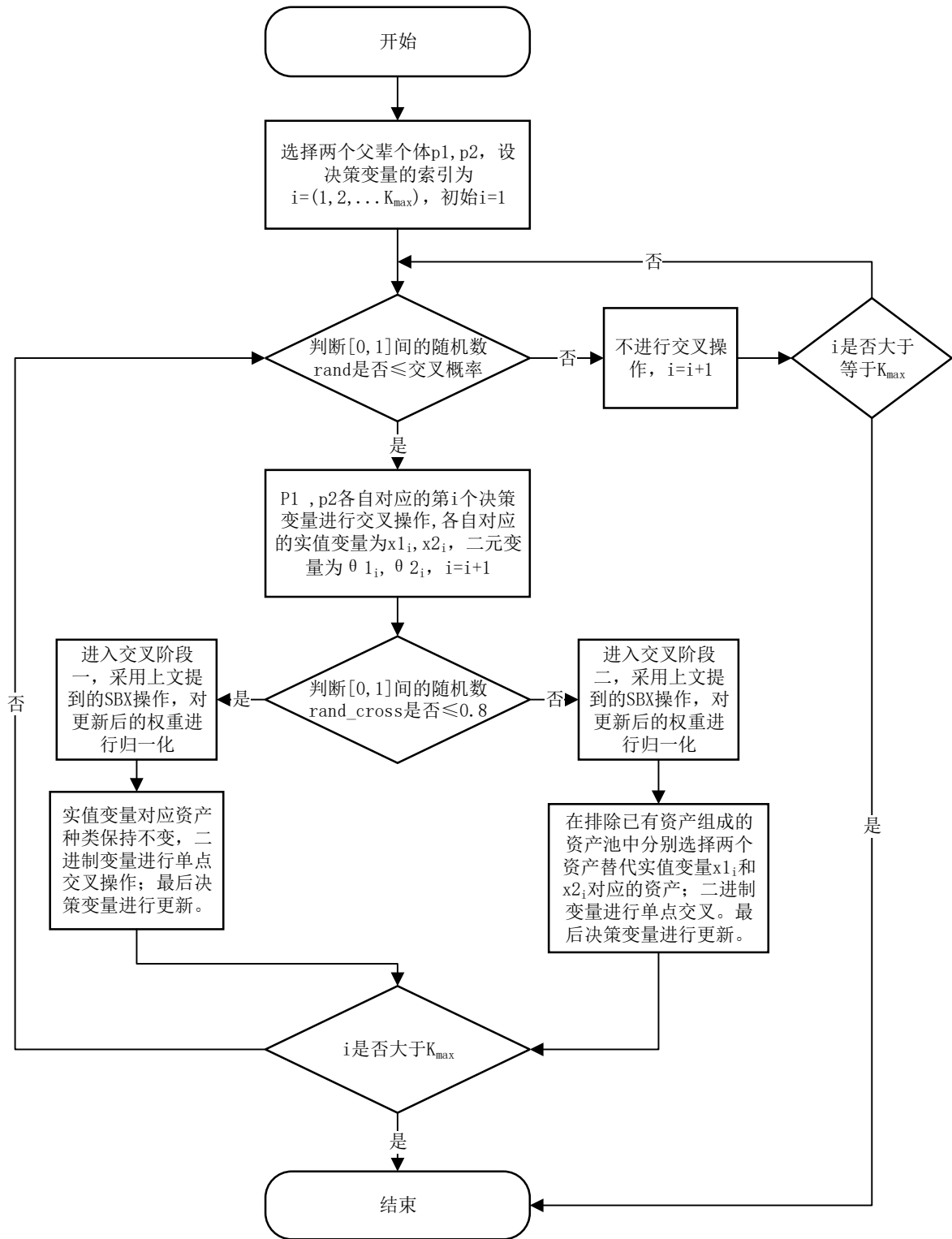


图 5-4 两阶段交叉算子流程图

## 5.4 实证分析

### 5.4.1 实验说明

实验数据及环境同第 4 章。改进模型 3，即 PBMVSCCPO 模型，见式 (4-4)，在

4.1.3 小节进行了描述。利用 c-NSGA-III 算法对改进模型 3 进行求解。c-NSGA-III 算法在配对选择阶段采用基于隶属度的锦标赛选择，经多次校验，设置类别数 C 为 10，最大迭代次数 maxiter=1000，收敛精度 e=0.005，模糊指数 m=2，环境选择阶段采用 PBI 函数，惩罚参数  $\lambda$  设为 5，同时采用了定制化的个体编码方法以及两阶段交叉、变异操作。基数约束上下限  $K_{\min}$  和  $K_{\max}$  分别为 2 和 10。资产  $i$  的权重的上下限为  $l_i=0.00$  和  $u_i=0.99$ ， $i=1,\dots,N$ 。除非另有明确说明，群体大小 N 设置为 100，最大进化代数 Gen = 500。

实验对比分析六种算法在求解改进模型 3 的性能。它们分别是定制化的 c-NSGA-III 算法、NSGA-III、NSGA-II、MOEA/D 算法、以及同样加入定制化的个体编码方法、两阶段交叉和变异操作的新算法 NSGA-II-new 和 MOEA/D-new。各算法在测试问题上分别独立运行 20 次。

### 5.4.2 实验结果与分析

首先，表 5-2 和表 5-3 给出 c-NSGA-III、NSGA-II-new、MOEA/D-new 以及各原始算法经过独立运行 20 次后得到的 HV 指标和 Epsilon 指标的统计结果。可见采用定制化操作获得的非支配投资组合解集得到了更高的 HV 值以及更低的 Epsilon 值（加粗加斜显示），说明使用带有定制化操作的算法求解 PBMVSCCPO 模型具有更好的收敛性和多样性，也揭示了使用多目标进化算法解决 PBMVSCCPO 模型的有效性。

表 5-2 HV 指标统计结果		
算法	均值	标准差
c-NSGA-III	<b><i>7.635E-01</i></b>	5.123E-03
NSGA-III	4.3E-01	2.729E-02
NSGA-II-new	<b><i>7.486E-01</i></b>	8.561E-03
NSGA-II	4.36E-01	3.26E-02
MOEA/D-new	<b><i>6.791E-01</i></b>	1.112E-02
MOEA/D	1.87E-01	2.207E-02

表 5-3 Epsilon 指标统计结果		
算法	均值	标准差
c-NSGA-III	<b><i>4.173E-02</i></b>	5.715E-03
NSGA-III	5.207E-01	3.257E-02
NSGA-II-new	<b><i>6.558E-02</i></b>	1.202E-02
NSGA-II	5.093E-01	4.473E-02
MOEA/D-new	<b><i>1.23E-02</i></b>	1.238E-02
MOEA/D	7.756E-01	2.618E-02

同时可知，c-NSGA-III 取得了最优的 HV 值和 Epsilon 值，说明该算法比同样采用定制化操作的 NSGA-II-new 和 MOEA/D-new 算法性能更好。利用统计分析中的箱线图单独展示每个算法 20 次运行的 HV 和 Epsilon 性能指标，如图 5-5 所示，进一步证实了前面的结论。

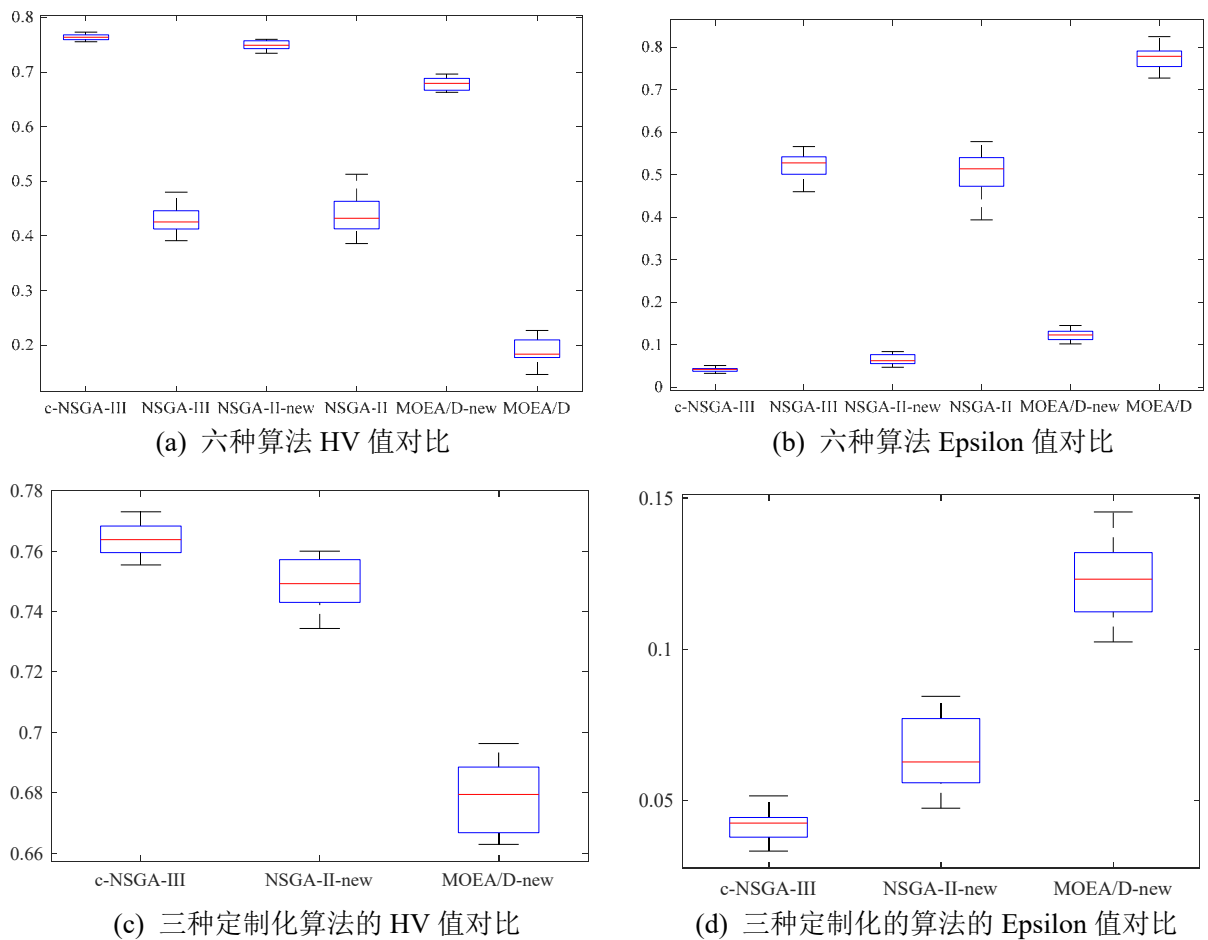


图 5-5 算法的 HV 值、Epsilon 值对比图

表 5-4 比较了各个算法使用定制化操作和无定制化操作分别运行 20 次后得到的 C-metric 均值。对于 NSGA-III 而言，26.5%的解被 c-NSGA-III 中至少一个解所支配，对于 NSGA-II 而言，16.2%的解被 NSGA-II-new 中至少一个解所支配，对于 MOEA/D 生成的所有的解被 MOEA/D-new 中至少一个解所支配。该结果表明采用定制化操作的算法获得了明显更好的结果。

表 5-4 各算法的 C-metric 指标 C(A,B)

A B	NSGA-III	c-NSGA-III	NSGA-II	NSGA-II-new	MOEA/D	MOEA/D-new
NSGA-III	—	26.5	—	—	—	—
c-NSGA-III	9	—	—	—	—	—
NSGA-II	—	—	—	16.2	—	—
NSGA-II-new	—	—	3.1	—	—	—
MOEA/D	—	—	—	—	—	100
MOEA/D-new	—	—	—	—	0	—

表 5-5 比较了采用定制化操作的三种算法的 C-metric 值，可见 c-NSGA-III 相对于其他两种算法的覆盖度更高，得到了更好的 Pareto 前沿。

利用 Wilcoxon 符号秩检验对实验结果进行统计比较，结果如表 5-6 所示。很明显，所有的 p-values 都比  $\alpha=0.05$  要小，说明 c-NSGA-III 与 NSGA-II-new、MOEA/D-new 的



性能差异具有统计学意义，置信度为 95%，进一步证实了 c-NSGA-III 表现更好。

表 5-5 三种定制化算法的 C-metric 指标

A B	c-NSGA-III	NSGA-II-new	MOEA/D-new
c-NSGA-III	—	8.14	6.35
NSGA-II-new	22.8	—	6.67
MOEA/D-new	14.1	4.05	—

表 5-6 Wilcoxon 符号秩检验

c-NSGA-III vs.	p-value
NSGA-II-new	0.000351
MOEAD-new	0.000293

图 5-6 是三种定制化算法求解 PBMVSCCPO 模型得到的最终解集三维图。风险由方差的算术平方根，也就是标准差进行度量。由图可知，各个算法得到的结果收敛性较好，收益相比原始算法有较大的提升，c-NSGA-III 均匀分布程度比其他两种算法更好。在分析投资组合问题时，风险和收益是两个重要的指标，本文对这两个指标也进行了二维图展示，由图 5-7 可知，虽然 c-NSGA-III 在高风险区域也有解的分布，但其更偏向中低风险区域，对比其它两种算法，相同风险程度下，获得的收益是偏大的，而且良好的多样性为投资者根据偏好进行投资组合选择提供了保障。

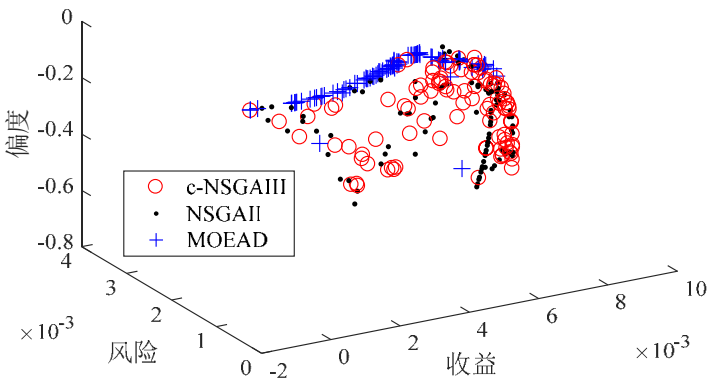


图 5-6 收益-风险-偏度三维图

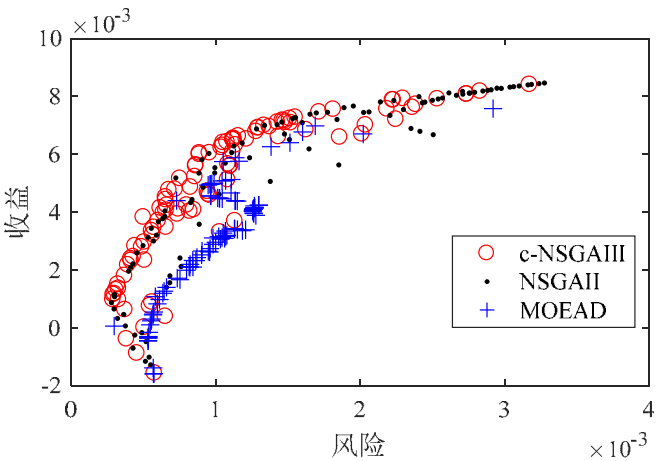


图 5-7 风险-收益二维图

在算法复杂度相同，软硬件配置也相同的情况下，相对于原始算法，c-NSGA-III、NSGA-II-new 和 MOEA/D-new 在求解改进模型 3 上的运行时间均有所提升。NSGA-III 从 423 秒缩短到 320 秒，NSGA-II 从 355 秒缩短到 289 秒，MOEA/D 从 499 秒缩短到 374 秒。该结果说明改进后的各个算法不仅得到的解集质量更好，在时间性能上也有所提升。

表 5-7 列出了由 c-NSGA-III 求解 PBMVSCCPO 模型得到的部分最优解集，由投资组合的具体情况可以看出，得到的解集满足实际约束，具有应用价值。

表 5-7 部分最优解

股票代码、名称、投资比例			收益 (%)	风险 (%)
600056.SH 中国医药 0.108	600125.SH 铁龙物流 0.882		0.804	0.273
600033.SH 福建高速 0.229	600125.SH 铁龙物流 0.761		0.789	0.225
600066.SH 宇通客车 0.292	600125.SH 铁龙物流 0.698		0.772	0.202
600033.SH 福建高速 0.258	600056.SH 中国医药 0.086	600125.SH 铁龙物流 0.646	0.745	0.181
600066.SH 宇通客车 0.410	600121.SH 郑州煤电 0.089	600125.SH 铁龙物流 0.491	0.711	0.144
600033.SH 福建高速 0.194	600036.SH 招商银行 0.437	600125.SH 铁龙物流 0.359	0.653	0.128
600036.SH 招商银行 0.050	600066.SH 宇通客车 0.284	600125.SH 铁龙物流 0.376	0.642	0.116
600111.SH 北方稀土 0.280				
600066.SH 宇通客车 0.205	600089.SH 特变电工 0.182	600125.SH 铁龙物流 0.358	0.627	0.102
600111.SH 北方稀土 0.245				
600036.SH 招商银行 0.205	600048.SH 保利地产 0.266	600111.SH 北方稀土 0.258	0.573	0.097
600066.SH 宇通客车 0.261				
600033.SH 福建高速 0.307	600036.SH 招商银行 0.231	600066.SH 宇通客车 0.238	0.565	0.095
600048.SH 保利地产 0.215				
600033.SH 福建高速 0.306	600036.SH 招商银行 0.193	600048.SH 保利地产 0.170	0.539	0.09
600066.SH 宇通客车 0.204	600121.SH 郑州煤电 0.116			
600033.SH 福建高速 0.268	600036.SH 招商银行 0.230	600048.SH 保利地产 0.188	0.525	0.085
600056.SH 中国医药 0.086	600066.SH 宇通客车 0.218			
600033.SH 福建高速 0.176	600036.SH 招商银行 0.163	600066.SH 宇通客车 0.181	0.455	0.079
600111.SH 北方稀土 0.171	600121.SH 郑州煤电 0.158	600125.SH 铁龙物流 0.141		
600033.SH 福建高速 0.292	600036.SH 招商银行 0.142	600048.SH 保利地产 0.119	0.453	0.06
600056.SH 中国医药 0.146	600066.SH 宇通客车 0.146	600111.SH 北方稀土 0.145		
600033.SH 福建高速 0.131	600048.SH 保利地产 0.229	600056.SH 中国医药 0.162	0.379	0.055
600066.SH 宇通客车 0.165	600089.SH 特变电工 0.042	600125.SH 铁龙物流 0.261		
600033.SH 福建高速 0.185	600048.SH 保利地产 0.082	600056.SH 中国医药 0.209	0.269	0.044
600066.SH 宇通客车 0.128	600089.SH 特变电工 0.107	600123.SH 兰花科创 0.018		
600125.SH 铁龙物流 0.261				

与不进行基数约束限制，只单纯满足预算约束的普通投资组合进行比较，对资产池中的 100 个资产随机赋予权重，形成 100 个投资组合，再与通过 c-NSGA-III 求解 PBMVSCCPO 模型得到的 100 个投资组合进行对比。比较各自组合的收益率均值及标准差，以及风险均值及标准差，由表 5-8 可知，c-NSGA-III 算法得到的近似最优解集相对于普通的投资组合而言，更合理，更有优越性。且与第 4 章中 NSGA-III 得到的结果相比，其风险水平相差不多的情况下，收益水平有了较大的提升。

表 5-8 近似解集与普通投资组合对比

求解方法	收益均值 (%)	收益标准差 (%)	风险均值 (%)	风险标准差 (%)
c-NSGA-III	0.5014	0.012	0.1085	0.0044
普通组合	0.00010649	0.000127	0.02935	0.00157

## 5.5 本章小结

本章针对 PBMVSCCPO 问题模型，对算法进行了定制化改进，在上海证券交易所的 100 支股票上与同样采用定制进化操作的 MOEA/D 和 NSGA-II 进行对比实验分析，同时也与各自对应的原始算法进行比较。利用 HV 指标，Epsilon 指标和 C-metric 指标对结果进行评估，并采用 Wilcoxon 符号秩检验进行统计性检验。结果显示 c-NSGA-III 具有更好的性能，采用定制化操作的算法相比原始算法在求解实际问题时性能更好，此外还给出了部分最优投资组合，供投资者实际使用。

## 结论

本文针对马科维茨经典模型的局限性，考虑实际股票投资情况，以 M-V 模型为基础，对其进行拓展。针对实际问题模型，对 NSGA-III 算法进行定制化改进，并与另外两种算法 NSGA-II 和 MOEA/D 对比，进行相关的实证分析。

论文主要完成如下工作：

(1) 建立新模型 PBMVSCCPO。考虑到 M-V 模型的局限性，引入基数约束降低投资组合管理难度，引入偏度刻画收益非对称性，利用启发式函数链神经网络预测的收益替代历史数据对证券期望收益、风险、偏度等指标进行量化，建立了基于预测的带有基数约束的三目标投资组合优化新模型，相对于经典模型更具有实际意义。

(2) 投资组合问题模型比较研究。改进模型 2 采用预测收益率对收益和风险进行计算，改进模型 3 引入基数约束和偏度。以上证 A 股中的实际股票数据为基础，利用 NSGA-III、NSGA-II、MOEA/D 对各模型求解，确定算法参数以及应用效果，验证了使用预测收益率替代历史收益率构建投资组合模型的可行性，以及多目标进化算法求解投资组合优化问题的有效性，并给出针对改进模型 3，算法需要进行的改进。

(3) 定制化改进算法 c-NSGA-III。改进算法以 NSGA-III 算法框架为基础，为引入收敛性信息，在环境选择阶段采用 PBI 距离函数替代垂直距离，在配对选择阶段采用基于隶属度的锦标赛选择机制，针对实际投资组合问题模型，采用带有资产信息的混合个体编码机制以及两阶段交叉和变异操作，对于违反约束的解，利用可行解修复操作对其进行修复。进一步拓展了多目标进化算法在投资组合优化领域的应用。

(4) 求解三目标新模型，对算法性能进行比较。利用定制化的 c-NSGA-III 算法，以及同样采用定制化进化操作的 NSGA-II 和 MOEA/D 求解 PBMVSCCPO 模型，与原始算法的求解结果进行对比。相对于原始算法求解效率低、解集收敛性的差的情况，采用定制化操作的算法的求解效果有了很大提升，且三个算法中 c-NSGA-III 的效果最好，与普通投资组合进行比较，其得到的组合也更具有实际投资价值，为高维目标投资组合优化问题提供了有效方法，具有一定的理论和实际意义。

后期研究展望：

本文在利用多目标进化算法解决投资组合问题上进行了一系列的研究与实验，但由于时间和精力限制，仍存在一些不足之处：(1) 投资组合模型未考虑市场上的其他实际因素，如市场行情，交易成本等，没有考虑更多的优化目标，后期可进一步研究 c-NSGA-III 求解更高维更实际的投资组合问题的有效性。(2) 在算法参数设置和调优上，后期需采用合理的自适应方法，进一步减少工作量。(3) 未考虑投资者的决策偏好，未来可将偏好交互融入到多目标进化算法中，直接产生决策者偏好的 Pareto 解集。

## 参考文献

- [1] H Markowitz. Portfolio Selection [J]. Journal of Finance, 1952, 7(1): 77-91
- [2] D X Shaw, S Liu, L Kopman. Lagrangian relaxation procedure for cardinality-constrained portfolio optimization [J]. Optimization Methods and Software, 2008, 23(3): 411-420
- [3] W F Sharpe. A Simplified Model for Portfolio Analysis [J]. Management Science, 1963, 9(2): 277-293
- [4] W F Sharpe. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk [J]. Journal of Finance, 1964. 19(3): 425-442
- [5] J Lintner. The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Assets [J]. Stochastic Optimization Models in Finance, 1969, 51(2): 220-221
- [6] J Mossin. Equilibrium in a Capital Asset Market [J]. Econometrica, 1966, 34(4): 768-783
- [7] S A Ross. The Arbitrage Theory of Capital Asset Pricing [J]. Journal of Economic Theory, 1976, 13(3): 341-360
- [8] W Yan, R Miao, S Li. Multi-period semi-variance portfolio selection: Model and numerical solution [J]. Applied Mathematics and Computation, 2007, 194(1): 128-134
- [9] N A Bradshaw, C Walshaw, C Ierotheou, et al. A multi-objective evolutionary algorithm for portfolio optimization [C], Proceedings of the Adaptive and Emergent Behaviour and Complex Systems Convention. 2009: 27-32
- [10] A Hitaj, G Zambruno. Portfolio optimization using modified herfindahl constraint[M], Handbook of Recent Advances in Commodity and Financial Modeling. Springer, Cham, 2018: 211-239.
- [11] 李华, 李兴斯. 证券投资组合中的熵优化模型研究 [J]. 大连理工大学学报, 2005(1): 157-160
- [12] 张保帅, 姜婷, 周孝华, 段俊. 投资组合优化的新方法: Mean-CoVaR 模型 [J]. 统计与决策, 2019, 35(11): 67-70
- [13] 陈国福, 陈小山, 张瑞. 基于引力搜索和粒子群混合优化算法的证券投资组合问题研究 [J]. 运筹与管理, 2018, 27(9): 170-175
- [14] 黄金波, 李仲飞, 姚海祥. 基于 CVaR 两步核估计量的投资组合管理 [J]. 管理科学学报, 2016, 19(5): 114-126
- [15] K Deb. Multi-Objective Optimization Using Evolutionary Algorithms [M]. John Wiley & Sons, 2001
- [16] 公茂果, 焦李成, 杨咚咚, 马文萍. 进化多目标优化算法研究 [J]. 软件学报, 2009,

20(2): 271-289

- [17]崔逊学. 多目标进化算法及其应用 [M]. 国防工业出版社, 2006
- [18]X Yu. Introduction to evolutionary algorithms [C], International Conference on Computers & Industrial Engineering. IEEE, 2010
- [19]A Zhou, B.Y Qu, H Li, et al. Multiobjective Evolutionary Algorithms: A Survey of the State of the Art [J]. Swarm and Evolutionary Computation, 2011, 1(1): 32-49
- [20]R Rosenberg. Simulation of Genetic Populations with Biochemical Properties [D]. Ann Harbor, Michigan, USA: University of Michigan, 1967
- [21]J D Schaffer. Multiple Objective Optimization with Vector Evaluated Genetic Algorithms [C]. Proceedings of the 1st International Conference on Genetic Algorithms. Hillsdale, New Jersey, USA: L. Erlbaum Associates Inc., 1985: 93-100
- [22]D E Goldberg. Genetic Algorithms in Search, Optimization and Machine Learning [J]. Menlo Park, California, USA: Addison-Wesley Publishing Company, 1989
- [23]C M Fonseca, P J Fleming. Genetic Algorithms for Multiobjective Optimization: Formulation Discussion and Generalization [C]. Proceedings of the 5th International Conference on Genetic Algorithms. San Francisco, California, USA: Morgan Kaufmann Publishers, 1993: 416-423
- [24]N Srinivas, K Deb. Multiobjective Optimization Using Non-dominated Sorting in Genetic Algorithms [J]. Evolutionary Computation, 1994, 2(3): 221-248
- [25]J Horn, N Nafpliotis, D E Goldberg. A niched Pareto genetic algorithm for multi objective optimization [C]. Proceeding of the 1st IEEE Conference on Evolutionary Computation. New York, USA: IEEE, 1994: 82-87
- [26]E Zitzler, L Thiele. Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach [J]. IEEE Transactions on Evolutionary Computation, 1999, 3(4): 257-271
- [27]J D Knowles, D W Corne. Approximating the Non-dominated Front Using the Pareto Archived Evolution Strategy [J]. Evolutionary Computation, 2000, 8(2): 149-172
- [28]D W Corne, J D Knowles, M J Oates. The Pareto Envelope-Based Selection Algorithm for 1 Multiobjective Optimization [C]. Proceedings of the 6th International Conference on Parallel Problem Solving from Nature. Heidelberg, Berlin, Germany: Springer-Verlag, 2000: 839-848
- [29]D W Corne, N R Jerram, J D Knowles, et al. PESA-II: Region-Based Selection in Evolutionary Multiobjective Optimization [C]. Proceedings of the 2nd Annual Genetic and Evolutionary Computation Conference. San Francisco, California, USA: Morgan Kaufmann Publishers, 2001: 283-290

- [30]M Erickson, A Mayer, J Horn. The Niched Pareto Genetic Algorithm 2 Applied to the Design of Groundwater Remediation Systems [C]. Proceedings of the 1st International Conference on Evolutionary Multi-Criterion Optimization. Heidelberg, Berlin, Germany: Springer-Verlag, 2001: 681-695
- [31]C A C Coello, G T Pulido. A Micro-Genetic Algorithm for Multiobjective Optimization [C]. Proceedings of the 1st International Conference on Evolutionary Multi-Criterion Optimization. Heidelberg, Berlin, Germany: Springer-Verlag, 2001: 126-140
- [32]E Zitzler, M Laumanns, L Thiele. SPEA2: Improving the Strength Pareto Evolutionary Algorithm [R]. Zurich, Switzerland: Swiss Federal Institute Technology, 2001
- [33]K Deb, A Pratap, S Agarwal, et al. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II [J]. IEEE Transactions on Evolutionary Computation, 2002, 6(2): 182-197
- [34]Q Zhang, H Li. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition [J]. IEEE Transactions on Evolutionary Computation, 2007, 11(6): 712-731
- [35]K Deb, H Jain. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems With Box Constraints [J]. Evolutionary Computation, IEEE Transactions on, 2014, 18(4):577-601
- [36]M R Sierra, C A C Coello. Improving PSO-Based Multi-Objective Optimization Using Crowding, Mutation and-Dominance [C]. Proceedings of the 3rd International Conference on Evolutionary Multi-Criterion Optimization. Heidelberg, Berlin, Germany: Springer-Verlag, 2005: 505-519
- [37]M Laumanns, L Thiele, K Deb, et al. Combining Convergence and Diversity in Evolutionary Multiobjective Optimization [J]. Evolutionary Computation, 2002, 10(3): 263-282
- [38]E Zitzler, K Deb, L Thiele. Comparison of Multiobjective Evolutionary Algorithms: Empirical Results [J]. Evolutionary Computation, 2000.8(2):173-195
- [39]K Deb, L Thiele, M Laumanns, E Zitzler. Scalable Test Problems for Evolutionary Multiobjective Optimization [C]. Proceedings of the 2002 IEEE Congress on Evolutionary Computation. New York, USA: IEEE, 2002: 825-830
- [40]Q Zhang, A Zhou, Y Jin. RM-MEDA: A Regularity Model Based Multiobjective Estimation of Distribution Algorithm [J]. IEEE Transactions on Evolutionary Computation, 2008, 12(1): 41-63
- [41]谢承旺, 王志杰, 夏学文. 应用档案精英学习和反向学习的多目标进化算法 [J]. 计算机学报, 2017, 40(3): 757-772
- [42]魏文红, 周建龙, 陶铭等. 一种基于反向学习的约束差分进化算法[J]. 电子学报, 2016,

- [43]M Abouhawwash, H Seada, K Deb. Towards faster convergence of evolutionary multi-criterion optimization algorithms using Karush Kuhn Tucker optimality based local search [J]. Computers & Operations Research, 2017, 79: 331-346
- [44]B Xin, L Chen, J Chen, et al. Interactive Multiobjective Optimization: A Review of the State-of-the-Art [J]. IEEE Access, 2018, 6: 41256-41279
- [45]K Deb, R E Steuer, R Tewari, et al. Bi-objective Portfolio Optimization Using a Customized Hybrid NSGA-II Procedure [C] International Conference on Evolutionary Multi-Criterion Optimization. Springer, Berlin, Heidelberg, 2011: 358-373
- [46]R Subbu, P P Bonissone, N H W Eklund, et al. Multiobjective financial portfolio design: a hybrid evolutionary approach [C] IEEE Congress on Evolutionary Computation. IEEE, 2005,2: 1722-1729
- [47]孙雪莲. 基于多目标优化的投资组合分析 [J]. 计算机工程与应用, 2013,49(9): 267-270
- [48]严丽娜, 吴军. 动态种群规模协同进化算法求解多目标投资组合优化问题 [J]. 数学的实践与认识, 2019,49(19):9-17
- [49]任雪婷, 卞京红, 张迷. 多目标差分进化算法在投资组合中的应用 [J]. 哈尔滨商业大学学报 (自然科学版), 2017,33(3): 343-346
- [50]李锋刚, 骆林, 陈亚波等. 求解均值-CVaR 投资组合模型的改进粒子群算法 [J]. 计算机工程与科学, 2016, 38(9):1870-1877
- [51]雍龙泉. 二次规划中 K-T 点的复杂性[J]. 喀什师范学院学报, 2006, 027(003):8-9
- [52]H P Schwefel, H P Schwefel, H P Schwefel, et al. Evolution and optimum seeking [M]. New York: Wiley, 1995
- [53]E Zitzler, S Künzli. Indicator-based selection in multiobjective search [C]. International Conference on Parallel Problem Solving from Nature. Springer, Berlin, Heidelberg, 2004: 832-842
- [54]N Beume, B Naujoks, M Emmerich. SMS-EMOA: Multiobjective Selection based on Dominated Hypervolume [J]. European Journal of Operational Research, 2007, 181(3): 1653-1669
- [55]J Bader, E Zitzler. HypE: an algorithm for fast hypervolume based many objective optimizations [J]. Evolutionary Computation, 2014, 19(1): 45-76
- [56]Y Tian, R Cheng, X Zhang, et al. An Indicator-Based Multiobjective Evolutionary Algorithm with Reference Point Adaptation for Better Versatility [J]. IEEE Transactions on Evolutionary Computation, 2018, 22(4): 609-622
- [57]A Trivedi, D Srinivasan, K Sanyal, et al. A Survey of Multiobjective Evolutionary



- Algorithm Based on Decomposition [J]. IEEE Transactions on Evolutionary Computation, 2017, 21(3): 440-462
- [58]D V Medhane, A K Sangaiah. Search space-based multi-objective optimization evolutionary algorithm [J]. Computers & Electrical Engineering, 2017, 58: 126-143
- [59]I Das, J E Dennis. Normal-Boundary Intersection: A New Method for Generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems [J]. Siam Journal on Optimization, 1998, 8(3): 631-657
- [60]张磊, 毕晓君, 王艳娇. 基于重新匹配策略的  $\varepsilon$  约束多目标分解优化算法 [J]. 电子学报, 2018, 46(5): 1032-1040
- [61]D Hadka, P Reed. Borg: An Auto-Adaptive Many-Objective Evolutionary Computing Framework [J]. Evolutionary Computation, 2013, 21(2): 231-259
- [62]W K Mashwani, A Salhi, O Yeniay, et al. Hybrid non-dominated sorting genetic algorithm with adaptive operators selection [J]. Applied Soft Computing, 2017, 56:1-18
- [63]A Elhossini, S Areibi, R Dony. Strength Pareto Particle Swarm Optimization and Hybrid EA-PSO for Multi-Objective Optimization [J]. Evolutionary Computation, 2014, 18(1): 127-156
- [64]D Yang, L Jiao, M Gong. Adaptive multi-objective optimization based on non-dominated solutions [J]. Computational Intelligence, 2009, 25 (2): 84-108
- [65]谢承旺, 许雷, 汪慎文. 一种增强型多目标烟花爆炸优化算法 [J]. 电子学报, 2017, 45(10): 2323-2331
- [66]邱飞岳, 胡烜, 王丽. 分组分解的多目标粒子群进化算法研究 [J]. 小型微型计算机系统, 2017, 38(8):1824-1828
- [67]F Gu, H L Liu, K C Tan. A hybrid evolutionary multiobjective optimization algorithm with adaptive multi-fitness assignment [J]. Soft Computing, 2015, 19(11): 3249-3259
- [68]T J Chang, N Meade, J E Beasley, et al. Heuristics for cardinality constrained portfolio optimization [J]. Computers & Operations Research, 2000, 27(13):1271-1302
- [69]H Konno, R Yamamoto. A mean-variance-skewness model: algorithm and applications [J]. International Journal of Theoretical & Applied Finance, 2005, 8(4):409-423
- [70]N Maghrebi. On skewness preference and persistence hypotheses[J]. Japan Financial Review, 1992, 15: 17-35
- [71]Y H Pao, S.M Phillips, D J Sobajic. Neural net computing and intelligent control systems [J]. Int. J. Control, 1992, 56 (2): 263-268
- [72]G Y N Tang. How efficient is naive portfolio diversification? an educational note [J]. Omega, 2004, 32(2): 155-160
- [73]E Zitzler, L Thiele, M Laumanns, et al. Performance assessment of multiobjective

optimizers: an analysis and review [J]. IEEE Transactions on Evolutionary Computation, 2003, 7(2):117-132

- [74]G Salvador, D Molina, M Lozano, et al. A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 Special Session on Real Parameter Optimization [J]. Journal of Heuristics, 2009, 15(6): 617-644
- [75]J C Bezdek, R Ehrlich, W Full. FCM: The fuzzy c-means clustering algorithm [J]. Computers & Geosciences, 1984, 10(2-3): 191-20
- [76]Y Yuan, H Xu, B Wang. An improved NSGA-III procedure for evolutionary many-objective optimization [C]. Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation. 2014: 661-668
- [77]K Deb, S Tiwari. Omni-optimizer: A generic evolutionary algorithm for single and multi-objective optimization [J]. European Journal of Operational Research, 2008, 185(3): 1062-1087

## 攻读学位期间发表的学术论文

- [1] Wenlan Huang, Yu Zhang, Lan Li. Survey on multi-objective evolutionary algorithms. J. Phys. Conf. Ser. , 2019, 1288
- [2] Lan Li, Wenlan Huang, Yu Zhang. The mean-variance-skewness cardinality constrained portfolio optimization problem using a customized multi-objective evolutionary algorithm. IEEE Int. Conf. Electron. Inf. Technol. Comput. Eng., EITCE, 2019, 1938~1944

## 致谢

寒来暑往，斗转星移，在哈尔滨商业大学三年的求学生活即将结束。值此毕业论文完成之际，谨向所有关心、帮助、激励我的人表示我最真诚的感谢。

本论文是在导师张雨副教授的悉心指导下完成的，几易更稿，倾注了导师大量的心血。三年来，我的点滴进步无不凝结着导师的关怀和教导，导师严谨的治学态度，敏捷的思维，渊博的知识，兢兢业业的奉献精神对我影响深远。在此向我的导师张雨副教授表示最诚挚的敬意和感谢。

本论文的完成同样离不开各位老师、同学、和师兄师妹们的关心和帮助。在此要感谢徐耀群教授、胡文教授、秦相林教授和辛海涛教授在论文开题、实验等过程中给予的支持和帮助，还要感谢李岚教授在科研过程中给我的鼓励和帮助。

还要感谢我的父亲、母亲和朋友们在求学路上对我无微不至的关怀、支持与鼓励，使我能够顺利完成学业。