Improved Large-Scale Multi-objective Optimization Algorithm for Portfolio Management

Shengtao Zhang(✉), Xuyang Li, Jie Zhang, and Wanqing Li

School of Information Engineering, Huzhou University, Huzhou 313000, China
zjhu_zst@126.com

Abstract. For securities investors, the return and risk of investment are the two main aspects of their concern. However, in real life, the vast majority of investors are not professionally trained, which makes them confused about portfolio selection in the face of tens of thousands of investment targets in the financial market. The multi-objective optimization problem of investment return and risk is solved by using an improved large-scale multi-objective optimization algorithm. From experimental results, it can be seen that the improved algorithm can get better results than previous algorithms on the large-scale multi-objective problems. The portfolio with the highest Sharpe ratio produced by the improved algorithm outperforms the CSI 300 index over the same period in terms of return and maximum retracement. It shows that the improved algorithm can achieve the selection of the investors' ideal portfolio from a larger number of stocks.

Keywords: large-scale multi-objective optimization · portfolio management · immune cloning

1 Introduction

With the improvement of per capita disposable income, people's awareness of financial investment is significantly enhanced. According to data, 19.6336 million new investors were added in 2021, of which 19.5814 million were natural person investors. The number of investors at the end of 2021 is 197.4085 million. However, by the end of 2021, the number of registered employees in the securities industry was 359, 800 1, accounting for only 1.8 per thousand of all market investors. In addition, according to Wind Financial Terminal data, as of October 27, 2022, there were 4960 stocks in Chinese A-share market. Facing to such a huge stock market, many investors often feel confused while choosing assets and do not know how to make portfolio to get the maximum returns and take the minimum risk at the same time. Obviously, these two objectives are conflicting, which is a typical multi-objective optimization problem. To solve this multi-objective optimization problem, Markowitz proposed a mean-variance model 2, which assumes that investors who try to maximize their profits and abhor risk are rational.

In order to solve this multi-objective optimization problem, many scholars have proposed that intelligent algorithms such as genetic algorithm can solve the problem.

J. Yen et al. (Eds.): ICBIS 2023, AHCS 14, pp. 827–836, 2023.
https://doi.org/10.2991/978-94-6463-198-2_85

Kyong Joo Oh et al. used genetic algorithm to optimize an index fund management scheme for portfolio, and shows outstanding advantages in comparison with S & P 500 index and Korea KOSPI200 index 3. Zhang et al. also used genetic algorithm to prove the effectiveness of genetic algorithm in solving portfolio optimization problems under different risk measures 4. Bian used ant algorithm and simulated annealing algorithm to solve this problem 5. Xiong et al. used multi-objective evolutionary algorithm to solve the problem of long-term portfolio management of bank wealth management products 6. However, although these multi-objective optimization algorithms can get a large number of investment portfolios under different investment risks, these algorithms only select no more than 100 stocks for portfolio. It may make investors miss the stocks with higher returns in the market. In addition, investors still need to choose the most cost-effective portfolio from a large number of portfolios obtained by the algorithm.

In order to improve the return of investors in the A-share market and facilitate their work in portfolio selection, a large-scale multi-objective optimization algorithm based on adaptive immune cloning is proposed in this paper. Sharpe ratio is introduced into the optimization process of the algorithm to express the return-risk ratio of a certain investment portfolio. In this paper, 744 stocks are selected from the A-share market. The reason for choosing these stocks is that there are relatively long-term and complete historical data, so that the serendipity of risk, return and Sharp ratio will be less. The experimental results show that the improved algorithm can produce higher returns and withdraw smaller portfolios than the CSI 300 index, thus saving investors' time in portfolio selection.

2 Related Background

2.1 Markowitz Model

The Markowitz model, or mean-variance model, assumes that the future return and risk of a particular portfolio can be extrapolated from past experience and data. The mean-variance model considers two main factors: the rate of return and the risk taken by the investor. The overall return of a portfolio is defined as the weighted average combination of the expected rates of return of the assets contained within it, which can be formulated as (1).

$$E\left(\vec{R}|\vec{W}\right) = \sum_{i=1}^{N} w_i E(R_i) \tag{1}$$

where $\vec{W} = \{w_1, w_2, \ldots, w_i, \ldots, w_N\}, 1 \leq i \leq N,\ w_i$ is the weight of asset i in the portfolio. N is the total number of optional assets in the portfolio, and $E(R_i)$ is the expected return of asset i.

Similarly, the overall risk of each portfolio is defined as the weighted combination of the covariances of the returns of the assets included in it, i.e.

$$\sigma^2\left(\vec{R}|\vec{W}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij}\left(\vec{R}\right) \tag{2}$$

where $\sigma_{ij}\left(\vec{R}\right)$ is the covariance between the asset i and asset j. For example, $E\left(\vec{R}_i\right)$ and $\overline{E}_i$ are respectively the past return and average past return of asset i. $E\left(\vec{R}_j\right)$ and $\overline{E}_j$ are respectively the past return and average past return of asset j. The value of $\sigma_{ij}\left(\vec{R}\right)$ is the product of the elements in the corresponding positions in the $E\left(\vec{R}_i\right)$ and $E\left(\vec{R}_j\right)$ minus the product of $\overline{E}_i$ and $\overline{E}_j$. The risk $\sigma^2\left(\vec{R}|\vec{W}\right)$ is the sum of elements of vector $\vec{W}$ multiplied by the covariance multiplied by the transposed vector $\vec{W}$.

It can be seen from above formulas that finding the best investment portfolio is to find a reasonable asset weight combination $\vec{W}$, so that the return of $E\left(\vec{R}|\vec{W}\right)$ is as much as possible while the risk $\sigma^2\left(\vec{R}|\vec{W}\right)$ is as less as possible, and at the same time $\vec{W}$ needs to meet the restrictive condition $\sum_{i=1}^{N} w_i = 1,\ w_i \geq 0$ that is, short selling is not allowed.

2.2 Sharpe Ratio

In order to evaluate the return-risk ratio of the portfolio obtained by the algorithm, we need a measure that can consider not only the profitability of the relevant portfolio, but also its risk, and this measure can also compare different portfolios. For this reason, we choose Sharp ratio 7, which represents the return-risk ratio of a portfolio. The Sharp ratio represents how much excess profit can be reaped for every unit of risk borne by the investor who chooses the portfolio. The calculation formula is as follows:

$$Sharpe\left(\vec{R}|\vec{W}\right) = \frac{E\left(\vec{R}|\vec{W}\right) - R_0}{\sigma\left(\vec{R}|\vec{W}\right)} \tag{3}$$

where $Sharpe\left(\vec{R}|\vec{W}\right)$ represents the Sharpe ratio of the portfolio, R_0 is the risk-free rate of return, and $\sigma\left(\vec{R}|\vec{W}\right)$ is the standard deviation of the portfolio. Here, the risk-free return R_0 is expressed by the average closing price of the yield of Chinese 10-year treasury bonds during the period from June 30, 2004 to June 30, 2022.

2.3 Large-Scale Multi-objective Optimization Framework

In order to accelerate the convergence speed of existing multi-objective algorithms, Cheng He et al. proposed a large-scale multi-objective optimization framework (LSMOF) based on problem reconstruction 8, which transforms a large-scale multi-objective problem into a low-dimensional single-objective optimization problem. First of all, some reference solutions are selected in the decision space to construct multiple bidirectional reference vectors. Then the distance from the ideal or the lowest point in the decision space to the frontier surface is defined as the decision variable of the restated problem. Finally, the target space is restated by performance metrics (for example, hypervolume (HV) metrics 9), where the values obtained through the performance metrics are used as the fitness of the restated problem.

3 Proposed Framework

In order to solve problems that genetic algorithm random and undirected iterative search, immune algorithm (IA) was proposed by imitating the immune system of human body. The immune cloning mode in the immune algorithm is the key to determine the effectiveness of the algorithm, so this paper improves LSMOF by the immune cloning mechanism of the immune algorithm, including antigen recognition, antibody cloning and mutation, and antibody adsorption.

3.1 Antigen Identification

In human immune system, antibodies must accurately identify the correct antigen, which is the basic premise for antibodies to destroy antigens. In order to enable the antibody to quickly find the antigen that needs to be eliminated, we directly select several evenly distributed points on the real front in the target space as our antigens, which are the targets that other antibodies should pursue. In the application of investment portfolio, this can ensure that the portfolio generated by the algorithm can obtain a wider range of returns and risks as far as possible, so that later investors can choose their preferred investment methods according to their own preferences.

3.2 Antibody Cloning and Mutation

After accurately identifying the antigen, we need to clone and mutate the antibody in order to solve the problem better. The cloning operation provides enough excellent individuals for subsequent mutation operation and optimization. Under the stimulation of antigen, the algorithm quickly clones multiple identical individuals, and then mutates each decision variable of each cloned individual.

In general, the mutation operator takes the real constant in the search process of the algorithm, but it is difficult to determine the size of the mutation operator in the implementation process. If the mutation rate is too large, the search efficiency of the algorithm and the accuracy of the global optimal solution are not good enough; if the mutation rate is too small, the population diversity will be reduced, which is prone to precocious phenomenon. Therefore, we design and use an adaptive mutation operator, which can adjust itself according to the search progress of the algorithm. The adaptive mutation operator is as follows:

$$\lambda = e^{1-\frac{G_{max}}{G_{max}-G+1}} \tag{4}$$

$$F = F_0 \times 2^{\lambda} \tag{5}$$

where F_0 represents the initial mutation operator, F represents the mutation probability, G_{max} represents the maximum generation of iterations, and G represents the current generation of iterations. When the algorithm starts, the adaptive mutation probability is $2F_0$. With the operation of the algorithm, the current number of iterations G gradually increases, λ gradually approaches zero, and the late mutation rate approaches F_0.

The adaptive method can not only maintain individual diversity and avoid precocity in the initial stage of the algorithm, but also retain the excellent individuals and avoid the destruction of the optimal solution, increasing the probability of finding the global optimal solution.

3.3 Antibody Adsorption

After cloning and mutation operations, we need to select and retain the best clones and give up other clones. Our method is to screen the best clones by calculating the Minkowski distance from the clone to the antigen in the target space, and these excellent antibodies have the shortest distance to the antigen. The Minkowski distance formula is as follows:

$$D(x, y) = \left(\sum_{d=1}^{D} |x_d - y_d|^p\right)^{\frac{1}{p}} \tag{6}$$

where D is the upper limit of the number of decision variables, d is the lower limit of the number of decision variables, and d starts from 1. x represents the clonal solution set, and y represents the imaginary antigen. The method repeats N (population size) times in each iteration and selects each clonal antibody.

In addition, this part is modified according to the application of the portfolio that needs to be realized after the algorithm, that is, the Sharpe ratio method is used to select and retain the best clones. The method of calculating the Sharp ratio has been described above.

4 Experimental Results

4.1 Performance Indicator

In order to verify the performance of the improved algorithm, this paper compares improved algorithm with other representative large-scale multi-objective optimization algorithms MOEA/DVA 10, WOF 11 and previous algorithm LSMOF on nine problems of test problem set LSMOP 12. Each algorithm runs 30 times independently on each test problem, and uses Wilcoxon rank sum test 13 to compare the improved algorithm with the classical algorithms at a significant level of 0.05. The symbols ‘ +’, ‘-’ and ‘ = ’ indicate that the comparison algorithm is significantly better than, significantly worse than, and statistically equal to improved algorithm, which named by LSMOF-IC.

In this experiment, the performance evaluation index IGD 14 is used to evaluate and compare the performance of the four algorithms. IGD is a comprehensive performance evaluation index, which mainly evaluates the convergence performance and distribution performance of the algorithm by calculating the minimum distance between each solution set on the real frontier and the individual set obtained by the algorithm. The smaller the IGD value is, the better the comprehensive performance of the algorithm is, including convergence performance and distribution performance. IGD is defined as follows:

$$IGD(P, Q) = \frac{\sum_{v \in P} d(v, Q)}{|P|} \tag{7}$$

where P is the point set evenly distributed on the real front surface, and $|P|$ is the number of point sets distributed on the real front. Q is the solution set obtained by the algorithm. $d(v, Q)$ is the minimum Euclidean distance from individual v in population P to population Q. IGD evaluates the comprehensive performance of the algorithm by calculating the average of the minimum distance from the point on the real Pareto frontier to the obtained population. Through the above formula, we can see that the better the convergence performance of the algorithm is, the smaller the $d(v, Q)$ is. Not only that, but also if most of the individuals in the population are concentrated in a narrow region, the worse the distribution performance of the algorithm is, the greater the individual $d(v, Q)$ is. In this way, we can evaluate the convergence and distribution performance of the algorithm.

4.2 Experimental Settings

In order to be fair, this paper uses the parameter settings recommended by each algorithm in the theory to ensure that each algorithm can achieve the best performance reported in the literature. All the comparison algorithms are implemented on PlatEMO 15. The parameters of the algorithm in the experiment are set as follows:

(1) population size N. For test cases with two goals, the overall size is set to 100, and for test cases with three goals, the overall size is set to 105.
(2) the specific parameter setting of each algorithm. In WOF, the number of evaluations used to optimize each original question T1 is 500, and for refactoring questions, T2 is 250, parameter Q is 3, the number of packets is 4, and the grouping method uses ordered grouping. At the same time, NSGA-II is embedded into WOF and LSMOF, and LSMOP is compared. In MOEA/DVA, the number of sampling solutions used to identify the control attributes of decision variables is set to 20, and the maximum number of trajectories used to judge the interaction between two variables is set to 6; in LSMOF, the reference solution r is set to 10, the population size of single-objective optimization is set to 30, and the mutation factor Fm in evolutionary algorithm is set to 0.8. In LSMOF-IC, the number of clones NCL is set to 50, and the initial value of adaptive operator F0 is 0.45. The disturbance range of decision variables depends on the value range of decision variables of specific test problems. The tr for both LSMOF and LSMOF-IC is set to 0.5.
(3) termination conditions. The termination condition of all test instances is 50000 times of evaluation.

4.3 General Performance

The statistics of the final IGD experimental results of these algorithms are shown in Table 1. The most effective examples have been represented with a gray background.

By analyzing Table 1, we can see that among the 18 test cases shown, compared with the other three algorithms, the best number of IGD indicators of LSMOF-IC has reached 16. Among them, for the three-objective LSMOP1 problem, LSMOF is the best of the four algorithms. For the LSMOP9 problem, the WOF is the best except when the number of decision variables is 1000. In most other examples of problems, LSMOF-IC works best among the four algorithms. It can be seen that the immune cloning strategy can make the LSMOF get a better performance and produce a better optimization effect.

Table 1. Comparison between LSMOF-IC and other three classical large-scale multi-objective optimization algorithms on LSMOP

Problem	M	D	MOEA/DVA	WOF-NSGAII	LSMOF	LSMOF-IC
LSMOP1	2	1000	1.05e + 1 (2.68e-1) -	6.79e-1 (1.20e-1) -	6.37e-1 (1.97e-2) -	4.18e-1 (8.03e-2)
	3	1000	1.07e + 1 (2.36e-1) -	8.01e-1 (1.30e-1) -	7.25e-1 (1.24e-2) +	7.51e-1 (1.93e-2)
LSMOP2	2	1000	4.05e-2 (2.84e-4) -	1.92e-2 (1.54e-3) -	1.81e-2 (5.41e-4)-	1.12e-2 (3.36e-4)
	3	1000	6.37e-2 (3.82e-3) -	7.00e-2 (4.76e-4) -	7.05e-2 (3.08e-3) -	6.00e-2 (1.66e-3)
LSMOP3	2	1000	1.07e + 3 (9.29e + 2) -	1.58e + 0 (1.25e-1)-	1.57e + 0 (1.08e-3) =	1.57e + 0 (1.03e-3)
	3	1000	4.47e + 2 (4.02e + 2) -	8.61e-1 (3.66e-4) -	8.60e-1 (3.00e-5) -	8.54e-1 (2.12e-2)
LSMOP4	2	1000	7.66e-2 (3.00e-4) -	5.99e-2 (2.38e-3) -	3.20e-2 (9.51e-4) -	2.51e-2 (4.32e-4)
	3	1000	1.30e-1 (3.19e-3) -	1.39e-1 (4.90e-3) -	1.14e-1 (2.83e-3) =	1.13e-1 (2.99e-3)
LSMOP5	2	1000	2.26e + 1 (5.20e-1) -	7.42e-1 (1.13e-16) =	7.42e-1 (1.13e-16) =	7.42e-1 (1.13e-16)
	3	1000	1.92e + 1 (5.90e-1) -	5.41e-1 (7.27e-5) =	7.20e-1 (1.98e-1) -	5.39e-1 (3.44e-3)
LSMOP6	2	1000	3.32e + 3 (2.78e + 3) -	6.82e-1 (1.39e-1) -	3.17e-1 (4.88e-3) =	3.13e-1 (2.53e-2)
	3	1000	3.92e + 4 (6.78e + 3) -	1.18e + 1 (3.25e + 1)-	9.39e-1 (2.02e-1) -	8.14e-1 (1.12e-1)
LSMOP7	2	1000	7.97e + 4 (3.37e + 3) -	1.51e + 0 (4.03e-2) =	1.51e + 0 (4.08e-4) =	1.51e + 0 (3.56e-3)
	3	1000	2.05e + 3 (1.83e + 3) -	9.23e-1 (3.49e-2) -	9.10e-1 (5.90e-2) =	8.58e-1 (8.26e-2)
LSMOP8	2	1000	1.90e + 1 (4.00e-1) -	7.42e-1(1.13e-16) =	7.42e-1 (1.13e-16) =	7.42e-1 (1.13e-16)
	3	1000	7.43e-1 (1.16e-1) -	3.56e-1 (6.34e-2) -	3.60e-1 (2.90e-2) -	1.31e-1 (5.38e-2)
LSMOP9	2	1000	5.51e + 1 (1.47e + 0) -	8.09e-1 (2.13e-3) -	8.07e-1 (2.13e-3) -	7.85–1 (2.17e-2)
	3	1000	1.36e + 2 (4.84e + 0) -	1.08e + 0 (1.14e-1) +	1.34e + 0 (1.99e-1)-	1.17e + 0 (9.94e-2)

4.4 Application of Portfolio Management

In order to verify the effectiveness of the algorithm proposed in this paper in solving the portfolio optimization problem of Markowitz mean-variance model, this paper selects 744 stocks from A-share market, calculates the historical data of these stocks from June 1, 2004 to May 30, 2022, and simulates the trend of the resulting portfolio from June 1, 2022 to September 30, 2022. The experimental results are shown in Fig. 1.

According to the analysis of Fig. 1, the Abscissa represents the time and the ordinate represents the cumulative rate of return of the portfolio. Here we choose the CSI 300 index as the control group for comparison. It can be seen that the combination with the highest Sharp ratio obtained by the algorithm has a cumulative return of 4.71% after 86 trading days, while the cumulative return of the CSI 300 index in the same period is −6.82%.

In addition, the portfolio obtained by the algorithm reached the highest return of 14.36% on the 21st trading day, while the highest return of the CSI 300 index reached only 10.11% in the same period. The excess return of the investment portfolio obtained by the algorithm is 4.25% higher than that of the CSI 300 index.

Finally, the portfolio obtained by the algorithm is also better than the CSI 300 index in controlling the maximum pullback. The maximum withdrawal of the portfolio from the highest point of the 21st trading day to the lowest point of the 80th trading day is only −12.86%, which is far better than the 16.93% of the CSI 300 index. In other words, the portfolio obtained by the algorithm allows investors to avoid at least 4.07% of capital losses.

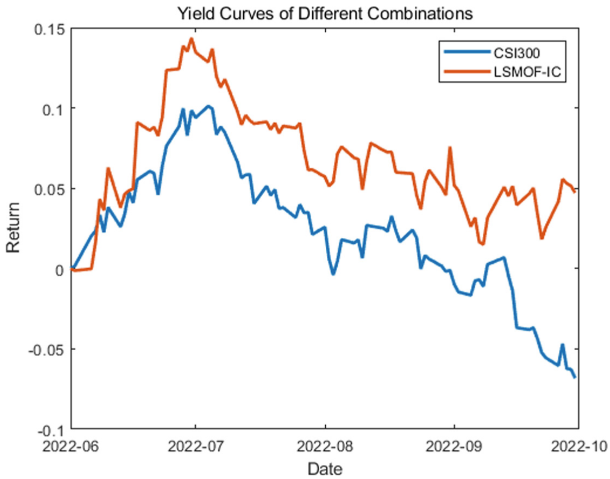

Fig. 1. Comparison between the Investment portfolio obtained by LSMOF-IC and the CSI 300 Index.

5 Conclusion

In this paper, the immune cloning mechanism is added to the first stage of LSMOF. The improved algorithm's performance in the LSMOP test problem set is compared with three classical large-scale multi-objective optimization algorithms. Experimental results prove that the immune cloning mechanism can make a certain optimization effect on LSMOF and enhance the comprehensive ability of the original algorithm. After that, this paper uses the improved LSMOF-IC to calculate the best portfolio of A-share's 744 stocks, and selects the portfolio with the highest Sharpe ratio to compare with the trend of CSI 300 index in the same period. The experimental results prove the feasibility and practicability of the improved immune algorithm in obtaining excellent investment portfolio. However, in fact, there are still some problems in this application, such as the mathematical model and algorithm do not take into account the impact of corporate fundamentals, market trading sentiment and other aspects on the stock trend. Therefore, how to take the basic financial situation of assets into account in the choice of investment portfolio and provide investors with a portfolio with a higher safety factor is one of the research directions in the future.

References

1. Yang Yi. "China Securities Industry Development report (2022) was published," Financial Times, (2022).
2. Harry M. Markowitz, "Portfolio selection," Journal of Finance, 7, 77–91, (1952).
3. Oh K J, Kim T Y, Min S, "Using genetic algorithm to support portfolio optimization for index fund management." Expert Systems with applications, 28(2), 371-379, (2005).
4. Chang T J, Yang S C, Chang K J, "Portfolio optimization problems in different risk measures using genetic algorithm," Expert Systems with applications, 36(7), 1052910537, (2009).
5. Bian Beili, "Research on the Application of Ant Colony algorithm in Multi-objective Optimization portfolio," South China University of Technology, (2012).
6. Xiong J, Zhang C, Kou G, et al, "Optimizing long-term bank financial products portfolio problems with a multiobjective evolutionary approach," Complexity, (2020).
7. W.F. Sharpe, "Mutual fund performance," Journal of Business, 39, 119–138, (1966).
8. C. He, L. Li, Y. Tian, X. Zhang, R. Cheng, Y. Jin, and X. Yao, "Accelerating large-scale multiobjective optimization via problem reformulation," IEEE Transactions on Evolutionary Computation, 23(6), 949–961, (2019).
9. L. While, P. Hingston, L. Barone and S. Huband, "A faster algorithm for calculating hypervolume," in IEEE Transactions on Evolutionary Computation, 10(1), 29-38, (2006).
10. X. Ma et al, "A Multiobjective Evolutionary Algorithm Based on Decision Variable Analyses for Multiobjective Optimization Problems with Large-Scale Variables," in IEEE Transactions on Evolutionary Computation, 20(2), 275-298, (2016).
11. H. Zille, H. Ishibuchi, S. Mostaghim and Y. Nojima, "A Framework for Large-Scale Multiobjective Optimization Based on Problem Transformation," in IEEE Transactions on Evolutionary Computation, 22(2), 260-275, (2018).
12. R. Cheng, Y. Jin, M. Olhofer, and B. Sendhoff, "Test problems for large-scale multiobjective and many-objective optimization," IEEE Transactions on Cybernetics, 47, 4108 – 4121, (2017).
13. W. Haynes, "Wilcoxon rank sum test," in Encyclopedia of Systems Biology. Springer, 2354–2355, (2013).

14. A. Zhou, Y. Jin, Q. Zhang, B. Sendhoff, and E. Tsang, "Combining model-based and genetics-based offspring generation for multi-objective optimization using a convergence criterion," IEEE Congress on Evolutionary Computation, 892–899, (2006).
15. Y. Tian, R. Cheng, X. Zhang, and Y. Jin, "PlatEMO: A MATLAB platform for evolutionary multi-objective optimization [Educational Forum]," IEEE Computational Intelligence Magazine, 12(4), 73–87, (2017).