An Evolutionary Weight Encoding Scheme and Crossover Methodology in Portfolio Assets Allocation

P. C. Lin[1], P. C. Ko[2], H. C. Wang[2], Y. J. Tien[2], J. A. You[1]
[1]Institute of Finance and Information
[2]Department of Information Management
[1,2]National Kaohsiung University of Applied Sciences

Abstract

Most of GA-based portfolio assets allocation uses normalization method to allocate investment asset's weight. However, the normalization process will cause unease converging and even diverging characteristics, because it changes the gene's relativity of address in chromosome. In this paper, we propose a weighed encoding scheme and crossover algorithm to allocate suitable assets in portfolio. Each gene encoded as a real number in a chromosome is denoted as the weighted number of assets in our approach. Due to no specific relationship assumed in our encoding scheme, the crossover process would not influence overall converging speed. In addition, in order to avoid losing optimal asset allocations through quicker converging, we also introduce the expanding rate to allow enlarging the possible range of asset allocation weight during the evolutionary process. Our experiments also show that higher expanding rate produces higher excess profit of investment.

Keywords: Portfolio, Asset allocation, Genetic Algorithms, Encoding, Crossover.

1. Introduction

Portfolio asset allocation problem is concerned with the percentage of the overall portfolio wealth allocated to each portfolio individual. The problem of Markowitz mean-variance approach [1] uses the covariance matrix derived to predict the investment risk of any allocation of assets. This approach only solves the efficient set problem for one portfolio rate of return at a time. Furthermore, the algorithm solves a set of linear simultaneous equations with efficient time complexity. To solve the asset allocation problem, the majority of research employs linear programming solvers. However, asset allocation optimization is complex and NP-hard problem. It is nonlinear with many local optima.

The genetic algorithm (GA), introduced by Holland in 1975, is a well-known efficient nonlinear search methodology in large problem spaces [2][7]. The efficient set selections within a portfolio based on the Markowitz model have been efficiently solved by using GA [5][6]. These papers optimize return and risk simultaneously, and give rise to more significant performance than traditional quadratic programming approach. Chang et al. [7] indicated the representation of a solution with two distinct parts: a set of assets and proportion of individual. The proportion of each asset will be weighted average to new proportion. A GA is designed to solve such a portfolio optimization corresponding to non-concave maximization problem with an order of the expected returns of stocks [10]. Xia, Wang and Deng [12] further considered transaction costs with V-shaped function based on variance covariance matrix of return to construct the portfolio compromise solution. Ehrgott, Klamroth and Schwehm [7] defined five different objectives related risk and return with consideration of individual investors' preferences based on GA to optimize the utility function of portfolio specified by decision maker. Orito and Yamazaki [10] used GA to select index fund portfolio to minimize risk in portfolio. Oh., Kim and Min [2] applied GA to optimize index fund portfolio management with fundamental variables. Their experimental results showed that the GA-based index fund portfolio provide better performance than S&P500 and FTSE100. Each asset weight is optimized by linear combination of minimum portfolio risk.

However, the asset allocation weight optimization is difficult for encoding scheme of Genetic algorithms in portfolio selection. Most of previous studies use normalization method to precede their asset weight of portfolio in GA mechanism. Unfortunately, the normalization process would cause unstable oscillation situation, because one gene's value changing would affect other genes' values. It will reduce the converging speed seriously. In this article, we will introduce

an evolutionary weighted encoding scheme and its corresponding crossover methodology to solve assets allocation in portfolio. It provides quicker converging speed without losing forecasting accuracy.

The rest of this paper is organized as follows. Section 2 describes our evolutionary weight encoding scheme and crossover algorithm. Section 3 shows our experimental results. Finally, Section 4 gives our conclusions.

2. An Evolutionary Weight Encoding Scheme and Crossover Algorithm

The traditional GA mechanism cannot produce reasonable asset allocations, except normalizing the final results. In this section, we will introduce a novel weight encoding scheme and its crossover algorithm.

3. Encoding Scheme

Each chromosome represents one possible assets allocation. It is noted that each gene is a real number not a bit string. Let X_i and X_j represent two chromosomes in parent population shown in Figure 1, where $w_{p,l}$ denotes the asset allocation weight for *l-th* asset in *p-th* chromosome. Assume the chromosome length is L and the population size is P, $1 \le p \le P$, $1 \le l \le L$ and $\sum_{l=1}^{L} w_{p,l} = 1$ for each chromosome p.

$X_i \Rightarrow$	$w_{i,1}$	$w_{i,2}$	…	$w_{i,L}$
$X_j \Rightarrow$	$w_{j,1}$	$w_{j,2}$	…	$w_{j,L}$

Figure 1 Two chromosomes representation

4. Crossover Algorithm

The crossover method would select two chromosomes from parent population. Assume X'_i and X'_j shown in Equation (1) are their generated offsprings, where c is a constant.

$$\begin{aligned} X'_i &= cX_i + (1-c)X_j \\ X'_j &= (1-c)X_i + cX_j \end{aligned} \qquad (1)$$

The sum of asset allocation weights in X'_i and X'_j maintains 100%, which will be shown as below. Let $\{w'_{ik}\}_{k=1}^{L}$ and $\{w'_{jk}\}_{k=1}^{L}$ denote the asset allocation weight sets in X'_i and X'_j, respectively.

$$\sum_{k=1}^{L} w'_{ik} = c\sum_{k=1}^{L} w_{ik} + (1-c)\sum_{k=1}^{L} w_{jk}$$

$$Because \ \sum_{k=1}^{L} w_{ik} = 1 \ and \ \sum_{k=1}^{L} w_{jk} = 1$$

$$\Rightarrow \sum_{k=1}^{L} w'_{ik} = 1, \ Similarly, \ \sum_{k=1}^{L} w'_{jk} = 1$$

Because each gene (w'_{ik}) of child chromosome (X'_i) represents the newly generated asset allocation weight in our portfolio, $0 \le w'_{ik} \le 1$. It means that

$$\begin{cases} 0 \le c \cdot w_{ik} + (1-c) \cdot w_{jk} \le 1 \\ 0 \le c \cdot w_{jk} + (1-c) \cdot w_{ik} \le 1 \end{cases}, for\ each\ k \qquad (2)$$

Let R_{ik} and R_{jk} denote the reasonable ranges of c in the *k-th* gene of X_i and X_j. We have

$$R_{ik} = \begin{cases} \left[\max\left(\frac{-w_{jk}}{w_{ik}-w_{jk}}, \frac{-(1-w_{ik})}{w_{ik}-w_{jk}}\right), \min\left(\frac{w_{ik}}{w_{ik}-w_{jk}}, \frac{1-w_{jk}}{w_{ik}-w_{jk}}\right)\right] & w_{ik} > w_j \\ \left[\max\left(\frac{-w_{ik}}{w_{jk}-w_{ik}}, \frac{-(1-w_{jk})}{w_{jk}-w_{ik}}\right), \min\left(\frac{w_{jk}}{w_{jk}-w_{ik}}, \frac{1-w_{ik}}{w_{jk}-w_{ik}}\right)\right] & w_{ik} < w_j \end{cases} \qquad (3)$$

Therefore, the reasonable range of c in the *i-th* chromosome denoted as R_{ic} would be shown as Equation (4).

$$R_{ic} = \bigcap_{k=1}^{n} R_{ik} \qquad (4)$$

The recombination process works by selecting a reasonable range of c denoted as R_c first, then randomly selecting one real number from R_c, and creating the two children by Equation (2). R_c is defined as Equation (5).

$$R_c = R_{ic} \bigcap R_{jc} \qquad (5)$$

The evolutionary process would be iterated until a candidate with sufficient fitness is reached. It means that higher number of offspring with higher viability will match the problem requirements more closely. Nevertheless, R_c will diverge because $w_{ik} \approx w_{jk}$, for each k.

Besides, the evolutionary process will converge gradually, if $R_c=[0, \alpha]$, $0<\alpha\le 1$, because $min(w_{ik}, w_{jk}) \le w'_{ik}, w'_{jk} \le max(w_{ik}, w_{jk})$. Regrettably, if the randomly generated c approaches 0.5, it will carry the danger of missing the optimal solution. Furthermore, the variety of individuals in population would decrease rapidly. For example, if $w_{ik}=0.2$, $w_{jk}=0.6$, $c=0.5$ and the optimal weight is 0.3, their offspring are $w'_{ik}=0.4$, $w'_{jk}=0.4$. It is possible not to achieve the optimal weight 0.3. To avoid such a problem, our crossover algorithm allows R_c a little expansion and not only to be limited in [0, 1]. Assume the allowable range of R_c is $[min_c, max_c]$ shown in Figure 2. Let η denote a

user-defined expansion rate. Therefore, the newly allowable range R'_c is shown in Equation (6).

Figure 2. The allowable range of R_c

$$R'_c = [\eta \times min_c, 1 + \eta \times (max_c - 1)] \quad (6)$$

Our proposed crossover algorithm is illustrated in Figure 3.

5.1 **Input**: S_P: Parent population
η: Expanding rate
5.2 **Output**: S_O: Offspring population
5.3 **DO**
5.4 Select two Chromosomes: X_i, X_j from S_P
5.5 **FOR** each X_i *and* X_j
5.6 Calculate $(R_{ik} \mid R_{jk})$ by Equation (3).
5.7 $R_c = R_c \bigcap (R_{ik} \mid R_{jk})$
5.8 **End FOR**
5.9 $R'_c = \left[-\frac{\eta}{min_c}, 1 + \frac{\eta}{max_c - 1}\right]$,
where $R_c = [min_c, max_c]$
5.10 Random generate c from R'_c
5.11 Generate offspring:
$X'_i = cX_i + (1-c)X_j$
$X'_j = (1-c)X_i + cX_j$
5.12 **End DO**

Figure 3. Crossover Algorithm

5. Fitness Function

In order to find effective portfolio and asset weight under maximum return and minimum risk, the fitness function evaluates the return of the investment of trading. The fitness $f(E)$ is derived from Equation (7), where ROI is return of investment from all trades and σ is the standard deviation from the net value of all of the trade days. Let NV_i be the net value of i-th trades, $\overline{NV}$ is the mean of NV, and N is the number of trades; σ is evaluated from Equation (8).

$$f(E) = \frac{ROI}{\sigma} \quad (7)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (NVi - \overline{NV})^2}{N}} \quad (8)$$

6. Experimental Results

The simulation environment, sample data and experimental results are described in this Section. Our weighed encoding scheme and crossover methodology of GA program is written in J2SE 5.0 and run in Microsoft Window XP environment.

7. Parameter Setting and Sample Data

The parameters used in our GA algorithm are described as follows. The population size is 100 and the chromosome length is n real positive values if considering n assets in portfolio. The number of generations is set to 700. The crossover rate is set to 0.9 and mutation rate is set to 0.01.

Five companies are arbitrarily selected to be our testing targets. The stocks of these companies have been listed and traded on the Taiwan Stock Exchange (TSE). The experimental period extends from January 1, 2005 to December 31, 2005. The market value and the total asset of those companies are given in Table 1. The mean market value is $202,024.92 million and the mean total asset is $360,624,462.20 thousands.

Table 1 The market value and total asset of data samples

Stock ID	Company Name	Market Value (Millions)	Total asset (Thousands)
2317	Foxconn Electronics Inc.	622,933.49	312,200,692
2301	Lite-On Technology	96,728.14	89,666,488
2354	FOXCONN Technology Co	78,136.77	19,613,566
2801	Chang Hwa Bank	81,280.13	1,288,444,615
2880	Hua Nan Financial Holdings	131,046.07	93,196,950
Mean		202,024.92	360,624,462.20

8. Analysis of Results

To avoid singularity arising from someone experiment, the average performances of eleven experiments under different expanding rate (0, 0.1…1) are used, respectively. In Figure 4, higher expanding rate would produce higher ROI generally. Especially, ROI will be improved if the expanding rate is larger than 1. Figure 5 shows the capital allocation of our portfolio, where the highest ratio is 79.37% (Lite-On Technology) and the next ratio is 18.44% (Foxconn Electronics Inc.).

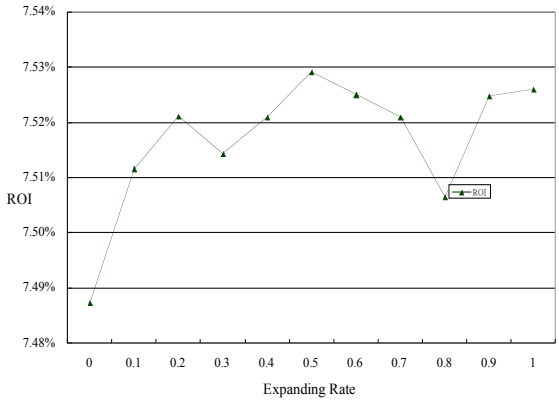

Figure 4. Investment ROI of testing phase

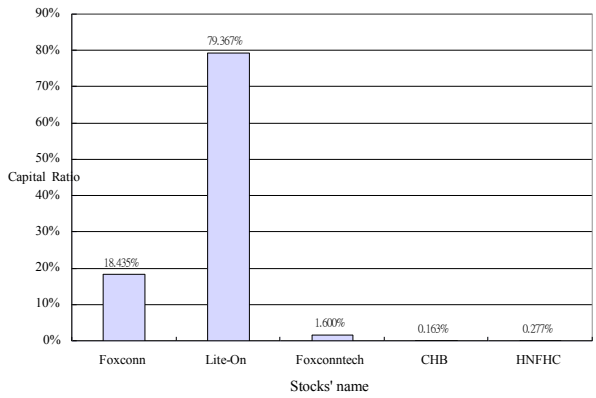

Figure 5. The capital allocation of portfolio

9. Conclusion

Asset allocation is the most important issue of portfolio management. The conventional Mean-Variance approach uses linear estimating a larger number of parameters such as co-variance metric to compute investment weights in portfolio. However, optimizing asset allocation is a complex and NP-hard problem. The traditional *Simple GA* finds efficient set with nonlinear capability in large problem spaces, but it can not produce reasonable asset allocation weights after crossover and mutation operations, except using normalization process. Unfortunately, the normalization process would cause serious unstable oscillation situation. It leads to converge slowly even diverge. In this article, we propose a novel weighted encoding scheme and crossover method to select an efficient asset allocation set in portfolio. From our experiments, higher expanding rate would produce higher ROI. It is interesting that ROI will be improved evidently, if the expanding rate is larger than 1. It means that restricting the expanding rate to 0 would lose optimality in the asset allocation problem. In addition, our experiments also demonstrate the stability and robustness of our algorithm.

References

[1] H. Markowitz, "Portfolio selection," Journal of Finance, Vol. 7, pp. 77-91, 1952.

[2] K. J. Oh, T. Y. Kim and S. Min, "Using genetic algorithm to support portfolio optimization for index fund management," Expert Systems with Application, Vol. 28, pp. 371-379, 2005.

[3] K. J. Oh, T. Y. Kim, S. Min and H. Y. Lee, Portfolio algorithm based portfolio beta using genetic algorithm," Expert Systems with Application, Vol. 30, pp.527-534, 2006.

[4] J. Holland: Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence. (University of Michigan Press, 1975).

[5] J. Shoaf and J. A. Foster, "A genetic algorithm solution to the efficient set problem: A technique for portfolio selection based on the Markowitz model," Tech. Report, Dept. of Computer Science, University of Idaho, Moscow, ID, 1995.

[6] J. Shoaf and J. A. Foster, "The efficient set GA for stock portfolios," In Proceedings of the IEEE International Conference on Computational Intelligence, pp. 354 – 359, 1998.

[7] M. Ehrgoot, K. klamroth and C. Schwehm, "An MCDM approach to portfolio optimization," European Journal of Operational Research, Vol. 155, pp. 752-770, 2004.

[8] M. Srinivas and M. P. Lalit: Genetic algorithms: A survey. IEEE Computer, 27 (1994) 18–20.

[9] T. J. Chang, N. Meade, J.E. Beasley and Y. M. Sharaiha, "Heuristics for cardinality constrained portfolio optimization," Computers & Operations Research, Vol. 27, pp. 1271-1302, 2000.

[10] Y. Orito and G. Yamazaki, "Index fund portfolio selection by using GA," International Conference Intelligence Economic and Finance, 2005.

[11] Y. Xia, B. Liu, S. Wang, and K. K. Lai, "A model for portfolio selection with order of expected returns," Computers & Operations Research, 27, pp. 409-422, 2000.

[12] Y. Xia, S. Wang and X. Deng, "A compromise solution to mutual funds portfolio selection with transaction costs," European Journal of Operational Research, Vol. 34, pp. 564-581, 2001.