Genetic Algorithm for Designing an Index Fund

02991784 大阪大学経済学研究科 *韓 尚憲 HAN Sang-Heon
01300234 大阪大学経済学研究科 田畑吉雄 TABATA Yoshio

1 INTRODUCTION

It is well known that a stock index of a market portfolio, such as TOPIX, S & P 500 etc., is a measure of the performance of a hypothetical basket of stocks. In recent, it has become very popular in the investment industry to design portfolios or an index fund with a securities that closely track visible indices.

In this study, we consider the traditional asset allocation of Markowitz type and develop an efficient method based on GA(Genetic Algorithm) for designing an index fund. The GA is known one of the most powerful tools for solving hard constrained combinatorial problems such as Kanpsack Problem, Travelling Salesman Problem, Portfolio Selection Problem, etc. The problem of minimizing the tracking error under the given number of securities in the index fund is a combinatorial problem, which is formulated as quadratic programming problem with 0-1 variables[1]. Since it is very difficult to solve in practice, we propose an efiicient algorithm. Next, we discuss the efficiency of the algorithm by utilizing the Japanese index NIKKEI 225 and demonstrate several modifications of the basic genetic procedures including a new fitness and mutation rate.

2 MODEL

We are particulary interested in forming an index-like m security portfolio R_m of the n original securities($n > m$) in the mean-variance framework of Markowitz, satisfying the traditonal criterion that the tracking error $E(R_n - R_m)^2$ is minimized. It is varified that tracking error between portfolios R_n and R_m is

$$E\{R_n - R_m\}^2 = Var\{R_n - R_m\} + \{E(R_n) - E(R_m)\}^2.$$

Let $y_i (i = 1, 2, \ldots, n)$be a 0-1variable defined as

$$y_i = \begin{cases} 0, & \text{if the } i\text{th security is not included in } R_m \\ 1, & \text{if the } i\text{th security is included in } R_m. \end{cases}$$

Then the problem to minimize tracking error can be written as

$$\begin{aligned} \textbf{Minimize} \quad & E(R_n - R_m)^2 \\ & = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(w_i - x_i y_i)(w_j - x_j y_j) \\ \text{subject to} \quad & x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = 1 \\ & y_1 + y_2 + \cdots + y_n = m \\ & y_i = 0 \text{ or } 1. \end{aligned}$$

$R_n = \sum_{i=1}^{n} w_i P_i$ $(R_m = \sum_{i=1}^{n} x_i y_i P_i)$: random return of $n(m)$ portfolios

a_{ij}: variance-covariance matrix of security i and j.

w_i: fraction of investor's wealth invested in security i.

x_i: weight of security $i (i = 1, 2, \ldots, n)$

This problem is a qudratic programming problem with $0 - 1$ variables. However this optimization problem is very difficult to solve, since it belongs to the class of NP(non-polynomial) complete problems, even if the values of x_i are given. The GA approach will be described in the next section for solving the above problem.

3 GENETIC ALGORITHM

S0: **Cromosome representation**
binary representation of length=n
(1:selected seccurity, 0:non-selected seurity)

S1: **Generate an initial populaton**
Produce pop_{size} chromosomes $v_k (k = 1, 2, \ldots, pop_{size})$randomly.

S2: **Repeat**

S2-1: **Evaluate fitness of each individual**
fitness function is denoted as follows:

$$\alpha * \frac{1}{r+1} + (1 - \alpha) * \frac{1}{(c - select)^2 + 1}.$$

Where,

$$r += \quad a[i][j] * (w[i] - \frac{1}{select} chrom[i]) * (w[j] - \frac{1}{select} chrom[j]),$$

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

α:weight, c:numbers of '1'in the string.

S2-2: **Selection**(Roulette Selection method)
Let $\{I_1, I_2, \ldots, I_N\}$ be set of population. Select solution I_i as a parent string according to selection probability $P(I_i)$.

$$P(I_i) = \frac{f(I_i)}{\sum_{j=1}^{N} f(I_j)}$$

$f(\cdot)$ is fitness value of the slution I.

S2-3: **Crossover**
Two-point crossover

S2-4: **Mutation**
Invert each bit in the solution with a small probability.

4 A NUMERICAL EXAMPLE

We test numerical examples in cases of (n=14,m=7), (n=60,m=15) and real case(n=225,m=30) by using **NIKKEI 225**(1984–1988) for discussing the efficiency of proposed GA. In case of (n=14,m=7), GA result is compared with known optimal solution.

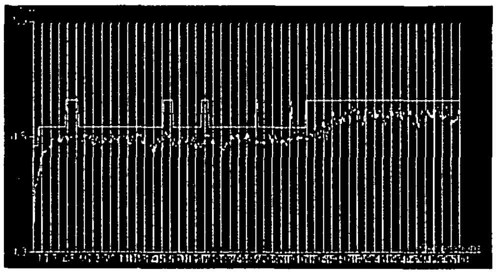

Fig 1. $n = 14, m = 7$ (Global optima is known)

best fitness = 0.663826

generation for best fitness = 38

chrom for maximum fitness is [00111001001101]

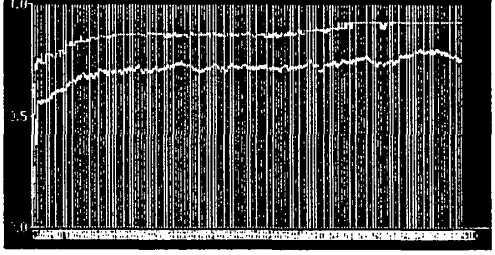

Fig 2. $n = 60, m = 15$ (Global optima is unknown)

best fitness = 0.919908

generation for best fitness = 971

chrom for maximum fitness is [0011000000100000
10000101100000010100100100100001000000100000l]

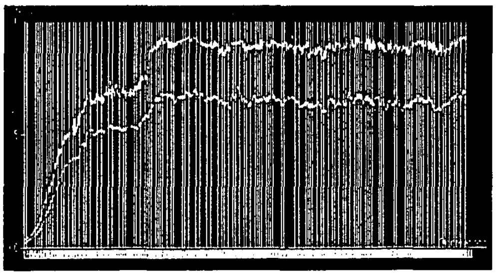

Fig 3. $n = 225, m = 30$ (Global optima is unknown)

best fitness = 0.938751

generation for best fitness = 1540

chrom for maximum fitness is [0100000010000100100010100000
00000000000000000011010000000000000100011000000000000000
1000000000000000000000000010100100000000000001000000000
0000000000000000010001010000010101011000000000000001
0100000010000000000011000000000000000000]

5 CONCLUDING REMARKS

In this research, we developed an efficient GA to design an index fund with a given number of securities that minimizes the tracking error between the benchmark portfolio and the index fund. The proposed method can find the solutions as optimal or at least sub-optimal. Many problems are left for future research. In particular, it is necessary more theoretical analysis, simplification of operator and setting of parameter for better GA convergency.

参考文献

[1] Y.Tabata and E.Takeda :"Bicriteria Optimization Problem of Designing an Index Fund", *Journal of Operational Ressearch Society*, Vol.46, pp.1023–1032 (1995)

[2] D.Goldberg :"Genetic Algorithm in serach, optimization, and Machine Learning", Addison-Wesley (1989)

[3] 北野宏明："遺伝アルゴリズム", 産業図書明 (1993)

[4] 石田良平："パソコンで学ぶ遺伝アルゴリズムの基礎と応用", 森北出版 (1997)

© 日本オペレーションズ・リサーチ学会. 無断複写・複製・転載を禁ず.