Selecting Valuable Stock Using Genetic Algorithm

Ninh Nguyen Quang¹, Linh Dang Hong²

Abstract
This paper proposes the use of genetic algorithms Genetic Algorithm (Genetic Algorithm) for stock options to support quality investors in the stock market (stock market) value school of investing treatment. Value investing is the cornerstone of long-term growth, investors experience the ups and downs of the market the stock market is more likely to be more successful and have more investment strategies applied at each stage of development the stock market. Use these resources to learn more about value investing techniques can help investors increase the property value over time. With the basic financial information and trading stocks using genetic algorithms to determine market shares are undervalued than its true value. Experimental results show that genetic algorithm proposed portfolio selection provides a very flexible tool and useful to assist investors in selecting valuable stocks.

Keywords: genetic algorithms, algorithms and investment value.

1. Introduction
Stock market is growing, the number of shares listed and traded registration and more. When to invest in stocks that many investors do not have enough time and attention to understanding all the shares are all registered and listed transactions. A difficult task for investors when investment decisions are choosing to invest in stocks. There are many studies on the selection of stocks to invest in that value investing is a strategy to search for stock market valuation is lower than its true value, or seem to have lost gradually attractive to investors. Many studies have shown that, if patience to master the principles of this strategy, investors can earn higher returns than growth investing strategies. A distinctive feature of the investment value is not aimed at the preferred shares and does not make judgments adventure with the growth rate of a company in the future. Instead, seek out the good stocks according to their own standards of the market share is being "processed", however this is a very difficult task because it takes a lot of time and effort to find out the true value of the stock. With a focus on business criteria and apply artificial intelligence to select and optimize the investment portfolio is one way to meet the above challenges. A number of studies have presented to solve the problem analysis and asset selection. Levin (1995) used neural networks to stock options have value, Chu et al (1996) fuzzy multiple attributes analysis to select stocks for the portfolio, and similar Zargham Sayeh (1999) using fuzzy rules file system basis to evaluate listed shares and stock options.

However APPROACH often have a number of limitations in the selection of investment securities portfolio. For example, the fuzzy approach (Chu et al, 1996 and Zargham and Sayeh, 1999), difficulties in establishing the function of law and the basis for the fuzzy logic system. One popular solution is to use the optimal procedure, but this approach needs a lot of time to train the system before applying. There are many ways of dimming, types of content, multiple access legal representation basis, the techniques and methods of removal enter blurry ... should lead to too many fuzzy models can be built. The selection of a suitable model to need more testing, while the neural network approach often fall into the trap negative test does not find the global optimum. To overcome this weakness of applying genetic algorithms will help investors choose investment portfolio worth more efficient.

¹ Faculty of Information Technology, Vinh University
² Faculty of Information Technology, Vinh University
2. Stock Selection Process Based On Genetic Algorithms

An overview of genetic algorithms

Genetic algorithms [5], [6] provide an approach based on simulation evolution. The general theory is described by the bit string, the string out of this bit depends on the application, and the idea of the theory can also be described by the expression wildcards or even computer programs. Search theory proper begins with a population, or a selected set of initial hypotheses. The individuals of the current population originate populations by the next generation of hybrid operation and random mutations - were sampled after the process of biological evolution. At each step, the hypothesis of the current population is estimated quantities contact adapted to, with the most consistent hypothesis is chosen according to the probability that the seeds for the production of the next generation.

Today, genetic algorithms have been extended to boost research in many areas of science and social life. Genetic algorithm is used in multiple optimization problems such as the problem of cutting materials, the travel problems, network protocols, computer networks, traffic control, scheduling problem, the algorithm transportation ... genetics is often a popular choice for problems in the field of optimization and has been successfully applied to various learning tasks, and for other optimization problems. For example, they are used for learning robot control laws and to optimize the training parameters and topology to artificial neural network.

Genetic algorithms table form:

<table>
<thead>
<tr>
<th>GA (Fitness, Fitness threshold, p, r, m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input:</td>
</tr>
<tr>
<td>Fitness: evaluation function of hypotheses,</td>
</tr>
<tr>
<td>Fitness threshold: the threshold used as termination criterion,</td>
</tr>
<tr>
<td>p: size of the population,</td>
</tr>
<tr>
<td>r: fraction of population to be replaced by Crossover,</td>
</tr>
<tr>
<td>m: mutation rate</td>
</tr>
</tbody>
</table>

- Initialize: \( P \leftarrow \{ h_1, ..., h_p \} \), p random hypotheses
- Evaluate: for each \( h \) in \( P \), compute \( \text{Fitness} (h) \)
- While \( \max_{h \in P} \text{Fitness} (h) \) < Fitness threshold

Create a new generation \( P_S \):

1. Select: Probabilistically select \((1 - r)p\) members of \( P \) to add to \( P_S \), using

\[
\Pr(h_i) = \frac{\text{Fitness}(h_i)}{\sum_{j=1}^{p} \text{Fitness}(h_j)}
\]

2. Crossover: Probabilistically select \( \frac{r \times p}{2} \) pairs of hypotheses from \( P \). For each pair \( < h_1, h_2 > \), produce two offspring by applying the Crossover operator. Add all offspring to \( P_S \).

3. Mutate: Probabilistically select \( m*p \) members of \( P_S \) and inverta randomly selected bit.

4. Update: \( P \leftarrow P_S \)

5. Evaluate: for each \( h \) in \( P \), compute \( \text{Fitness}(h) \)

- Return the hypothesis from \( P \) that has the highest fitness.

Meanwhile, many well-known investors in the world have successfully pursued a strategy of value investing and has become one of the leading investment value to organizations and individual learning.
Strategic investors value stocks can proceed through the following steps: (i) Buy stocks with a P / E ratios; (ii) Search the list of company growth is not too hot; (iii) Interested dividend payout ratio; (iv) Consider carefully before buying or selling stocks; (v) Consistent with the proposed strategy, not governed by the crowd.

Warren Buffett grave investors value the world famous once said: "To be a successful investor, we read hundreds and hundreds of annual reports of companies". This is creating big difference between him and other investors. When you read the report, Warren Buffett paid special attention to the indicator, the following issues:

- The growth rate of the past and present. This is one of the basis for predicting the growth of the company in the future.
- Profit / capital investment. Warren Buffett paid special attention to this indicator, and he only invests in companies that generate profits on your investment.
- The ratio of debt / capital; profit / debt. Suppose not achieve profitability expectations, the company has the ability to repay the debt limit?
- Reinvestment. Warren Buffett concerned about reinvestment. According to him, if the company business efficiency, gain margin / high equity, the company should keep most of the profits for reinvestment, rather than take a dividend out. Meanwhile, the company will become a "print money" with the increasing speed of the effects of compound interest amazing.

Choice of investment portfolio between stocks on genetic algorithm

Genetic algorithms use stock charts to identify quality stocks, stocks with high rankings in the ranking is considered the best stocks for investment options. Genetic algorithm uses a number of financial indicators of listed companies to evaluate stocks. Department of financial indicators of listed companies are considered as the input variables and genetic algorithms will give scores to evaluate stocks. In this article uses four key financial indicators to evaluate stocks: return on capital employed (ROCE) ratio price / earnings (P / E), earnings per share (EPS) and liquidity ratios.

Said ROCE efficient use of capital, including equity capital and loans, are:

\[
ROCE = \text{profit before interest and tax} / \text{capital employed} \tag{1}
\]

Higher index (ROCE) is the P / E ratio that between market value and net income per share and is calculated:

\[
P / E = \text{(stock price)} / \text{(earnings per share)} \tag{2}
\]

EPS is the only income per share is calculated:

\[
EPS = (\text{Net profit} - \text{total preferred dividends}) / \text{total bonus shares} \tag{3}
\]

Solvency ratio of short-term debt measures the extent to which a company can quickly liquidate assets to cover short-term debts. It is calculated as follows:

\[
\text{Solvency ratio of short-term debt} = (\text{liquid assets and short-term investments}) / (\text{current liabilities}) \times 100\% \tag{4}
\]

Solvency of the business will be better if the liquid assets and short-term investments shifted upward trend and short-term debt shifted downward trend; or are moving in the same increasing trend, but the growth rate of liquid assets and short-term investments is greater than the growth rate of short-term debt; or are moving in the same reduction trend, but the rate of decline of liquid assets and short-term investment is less than the rate of decline of short-term debt. When the input variables are determined, we can use genetic algorithms to discriminate and identify the quality of each share.

First of all, the population consists of a certain number of chromosomes originally created by assigning random bits "1" and "0" to all genes. In case shares are ranked, a gene that contains only a single bit string for
the status of input variables. In this study, the initial population of the genetic algorithm is created by encrypting the four input variables. For the test case of ROCE index, the authors designed 8 states representing ROCE quality indicator corresponding to different levels, from 0 (very bad) to 7 (very good). The selection of this level concerning the use of binary bits in the encryption process. Examples of coding ROCE is shown in Table 1 different input variables are encoded by the same principle, which is the binary sequence of a genome consists of three single bits (Table 1).

Table 1: Examples of Coding ROCE

<table>
<thead>
<tr>
<th>ROCE value index</th>
<th>Encryption</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-∞, -60%)</td>
<td>0</td>
<td>000</td>
</tr>
<tr>
<td>(-60%, -40%)</td>
<td>1</td>
<td>001</td>
</tr>
<tr>
<td>(-40%, -20%)</td>
<td>2</td>
<td>010</td>
</tr>
<tr>
<td>(-20%, 0%)</td>
<td>3</td>
<td>011</td>
</tr>
<tr>
<td>(0%, 20%)</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>(20%, 40%)</td>
<td>5</td>
<td>101</td>
</tr>
<tr>
<td>(40%, 60%)</td>
<td>6</td>
<td>110</td>
</tr>
<tr>
<td>(60%, +∞)</td>
<td>7</td>
<td>111</td>
</tr>
</tbody>
</table>

The encryption process performed on 3 bits used to simplify the study. Of course, using 4 or more bits in the coding process is also performed. The next job is to evaluate the chromosomes generated by operations and the hybrid single point mutation previously called functional training (RM), the design of functional training is very important in the use genetic algorithms, it determines the optimization using genetic algorithms. When the list of designated shares put to the test, a list of shares will be determined prior to design functional training. Usually good stocks will achieve the best price in the market the most favorable growth will then go down, these stocks tend to meet or exceed the historical high price of the transaction the previous year at the time of favorable market growth for the year. Here the authors use margin due to increase / decrease the price (yield value) every year (APR - Annual Price Return) to rank the listed shares represented by the formula:

$$APR_n = \frac{ASP_n - ASP_{n-1}}{ASP_{n-1}} \quad (5)$$

Among them: APRn yield value for the nth year, ASPn the stock price of the nth year. Typically, the stocks with profit margins high annual rates are considered good stocks. With APR assessed value of N shares, they will be assigned rank r ranges from 1 and N, where 1 is the highest value of APR, N is the lowest value. To facilitate comparison, ranking r of shares will be mapped linearly into stock charts ranged 0-7 with the following equation:

$$R_{actual} = 7 \times \frac{N - r}{N - 1} \quad (6)$$

Thus, the function of RM training can be designed to minimize the difference between the index charts the previous financial year and the actual rankings of all current listed company for an infection specific chromosomes, represented by:

$$RM = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (R_{derived} - R_{actual})^2} \quad (7)$$

Among them is Rderived r of shares ranking prior year, m is the number of individuals of the parent population.

After developing the number of individuals in the population, the best chromosome with the highest value is selected and trained. Use technical crossover mutant chromosomes. A crossover mutant chromosome is called a work; choose a random crossover point in the chromosome. Then, two parent chromosomes are exchanged at this point to produce two new children. Then, the chromosome is mutated with a probability of
0.005 for each random genetic change from "0" to "1" and vice versa. The mutation prevents genetic algorithm converges quickly in a small area of the search space. The last generation will be assessed if there is, then optimize results. Otherwise, reproduction steps are repeated until the individual satisfies the fitness function (7). In the ideal case, all the chromosomes of the last generation have identical genes represent the optimal solution.

3. Test Analysis

The daily data used in this study is the closing price of the shares obtained from the Stock Exchange Hanoi Securities (HNX). The data sample period from 02/01/2013 until 02/01/2014. Data monthly and annually in this study were obtained by monitoring daily transaction data, 170 stocks selected from AAA to VND code. First of all, the financial information of the company are the input variables of the genetic algorithm to get top ranking as indicated by (5) and (6), for example, bank stocks in 2014 prices the highest average trading 40,300 contracts in 2013 and the average transaction price is 28,300 highest.

According to (5), the ACB is APR_{2013}. Completely done similar calculations for the 10 stocks APR typical obtained results (ACB: 0.42, PVX: 0.07, PGS: 0.55, PVC: 0.27 PVS 0.08, KLS: 0.34, PLC: 1.64, HOM: 0.34, BCC: 0.29, VGS: 0.25), after APR of 170 stocks, ranked done r from 1 to N based on the APR of the same stock charts ROCE (table 1). List stocks are ranked r is mapped into stock charts from 0-7 according to (6). In the process of genetic algorithm optimization, RM is used to assess the genetic algorithm process. Best chromosome obtained using n to rank top shares was selected for the portfolio. For testing purposes, the top 10 stocks (Figure 1) were selected for testing in accordance with the charts of the stock quality by using genetic algorithms. To evaluate the usefulness of the optimization genetic algorithm, the authors compare the list of stocks selected from a genetic algorithm with a benchmark portfolio, benchmark per share is determined at the highest average price of the first two years of data included in the test. List of selected stocks in the investment value will tend to accumulate higher return rates in the past at the time of market flourishes.

![Figure 1: Yields of Portfolio Value](image)

Source Deduction from the Results of Running the Algorithm On Data Transactions From 02/01/2013 Until 02/01/2014 HNX.

Figure 1 shows the results of the portfolio. We can find the price yields the same proportion of the portfolio formed by the stocks selected by genetic algorithm significantly outperformed the benchmark. Selection of good quality stock is prerequisite to get a good portfolio. Although the portfolio with a large number of shares may be difficult for investors to track, difficult to control the status of all shares, the related information can not be share timely updates dispersion ability to track stocks making purchasing decisions is
limited; a number of shares of poor quality can lead to portfolios, affect portfolio efficiency. Meanwhile, the results also demonstrate that the portfolio of large stocks is not necessarily better than smaller portfolios. So it would be wise for investors if selecting a limited number of quality stocks to build a portfolio.

4. Conclusion

This study uses genetic algorithms to perform stock selection for portfolios. Experimental results show that the approach using genetic algorithm was shown to be useful in helping investors pick stocks that have value to the portfolio. When the input parameter of the stock is set, from a portfolio of stocks with large amounts of genetic algorithms to quickly put out a list of stocks with financial indicators is ideal for value investing treatment helps investors quickly select portfolio that does not take too much time for finding portfolios.

References


Huỳnh Đức Thuận (2011). Ứng dụng kết hợp trong khai phá dữ liệu giúp nhà đầu tư ra quyết định đầu tư trong thị trường chứng khoán Việt Nam. Trường Đại học Đông Á.