Journal of Management and Sustainability; Vol. 6, No. 1; 2016
ISSN 1925-4725 E-ISSN 1925-4733
Published by Canadian Center of Science and Education

Stock Fundamentals Model Based on Genetic Algorithm-Rough Set

Weizhong Jiang[1] & Xi Xie[1]

[1] Electronic information science and technology of Electrical and Information School, Jinan University, Zhuhai, Guangdong Province, China

Correspondence:Weizhong Jiang, Electronic information science and technology of Electrical and Information School, Jinan University, Qianshan road 206#, Zhuhai City, Guangdong Province, 519070, China. E-mail: 768242601@qq.com

Received: January 26, 2016 Accepted: February 18, 2016 Online Published: February 25, 2016

doi:10.5539/jms.v6n1p206 URL: http://dx.doi.org/10.5539/jms.v6n1p206

Abstract

Security investment problems are mostly highly-nonlinear and have huge operations, hence quantitative investment is applied to make decisions frequently. Considering the example of Medicine plate in Chinese stocks, fundamental indicators and technical indicators are combined, and then investment decisions are made and optimized stepwise based on rough set model, where generical gorithm is applied to solve the model, aiming at searching for a portfolio with high value and growth inside the whole medicine plate. In addition, such strategy elements as trend and goodness are considered in the prediction, evaluation and correction of the model, resulting in lower uncertainty of index selection. In the empirical example, with the use of the improved model, stock rankings inside the plate achieve an accuracy of 60.7%, which proves the model makes sense to some extent in the security investment.

Keywords: quantitative investment, fundamental analysis, genetic algorithm, adaptive optimization, rough set

1. Introduction

1.1 Background

In Efficient Markets Hypothesis, technical analysis alone could not help us to obtain the excess earnings in weak-form efficient market. In long term, numbers of investors, investment institutions and equity funds mesh fundamentals and technical analysis to make investment strategy. The fact proves that it is useful in the past period. Chinese stock market transfers fast to semi-strong efficient market, which lead to the voidance of this approach. What's more, the characteristic of information in equity market is numerous and mercurial, in this situation, we must consider the quantitative investment approach before making investment decision.

1.2 Correlational Studies

Black & Scholes (1973) derived B-S model applying in financial derivatives market, and Ross (1976) derived APT model applying in equity markets. Based on equilibrium security pricing theory structured by Sharpe (1964), Perold (1984), Ma (2004), Li (2008) applied program trading to the study of equity investment strategy, but these studies only focus on theoretical research, leaving lots of defect in empirical study.

For indexes, prior stock models' evaluation only focus on return-variance, ignore corporate financial situation and operating performance, which are the real reason of stock appreciation. This arise a number of researchers and investment institutions to analysis corporate financial situation, including some financial factors which could affect the stock intrinsic value, such asprofit ability, solvency, cash-flow ability and so on, and found corporate financial situation makebig difference in the fluctuation of stock price. Asness (1997) analysed stock price in terms ofcorporate fundamentals. Lee (1998) believed investors should consider corporate fundamentals firstly, and next, consider stock price. However, these researches only have text description until the 21 century. Piotroski (2000) selected 9 financial indexes from fundamentals, scored every stock, and selected some good stocks as a portfolio. On these basis, Mohanram (2005) commended BP as a standard, make the top ranked stocks as a sample, scored every stock and make the investment portfolio in terms of profitability, growth stability and financial conservatism, and found this portfolio obtain high excess return. However, we could not identify highly nonlinear law selected in tons of data, merely by traditional statistic method and investors' parts of experience. Therefore, lots of researchers such as Jiang (2013) & Zhang (2014) do some empirical studies on quantitative investment based on multi-influencing-factor model, trying to select a stock portfolio whichcontents higher value and growth. Inspired by these studies, we combine investment with roughset.

As an emerging data mining tool, rough set is characterized by attribute reduction function, attribute significance principle, and objective & quantitative data mining function.so rough setcould combine with comprehensive assessment theory, and then solve questions in corporate comprehensive assessment in fundamentals. Bao (2010) researched on equity investment strategy using multiple attribute decision making based on rough set. However, it shows useless in stock analysis with numerous data. This study should reduce the index system with 32 indexes, so we should traverse hundreds million times (2^{32}-1=4 294 967 295), it is impossible to finish it. But genetic algorithm is a kind of efficient global searching algorithm, which is put forward by Wroblewski (1995), shows good performance in solving NP-hard questions. Many researchers such as Lu (2008) improved this algorithm in a meliorating its complexity and adding to heuristic information. Li (2010) & Guo (2007) improved rough set according to genetic algorithm, accelerate its operation. But both of these are the research on algorithm itself, didn't apply it on practice, especially the application inequity investment strategy.

1.3 The Basic Train of Thought

This study considers the truth of equity investment portfolio strategy, reduces redundant indexes in initial index, and then determinates the weight of core attributes using approximation quality of rough set, next, applies it to Chinese equity market data to do empirical research inpast times and present times. What's more, this study synthetically analysis the result of index system in several times, improve the index system based on period moving. Further, this study evaluates portfolio strategy goodness in past times, reset the goodness using adaptive optimization, adjusts the investment strategy constantly, adapts to the volatile market.

2. Method

2.1 Selection of Data and Index Determination

2.1.1 Selection of Data

Determination of industry: this study selects listing corporations in the pharmaceutical industry, because the seasonal cycle and the economic cycle would have a significant impact on some industries operating conditions, and then affect its stock market, we should try to reduce these external factors' interference on our model, In consideration of the data we selected should have consistent seasonal characteristics, we need to think about select corporations in the same industry. What's more, we should make sure corporate financial situation has a lower sensitivity to the economic cycle. Considering these comprehensive considerations, we choose the medicine industry as our research sample for its more stable development.

Selection of period: in view of the effect of seasonal factors, we select the data from January 1, 2012 to March 31, 2014 as our sample. Because of the effect of seasonal factors, we divide it into 9 parts according to the quarterly time interval.

2.1.2 Determination of Condition Attributes

The selection of condition attributes should consider the availability of data and the comprehensiveness of evaluation of stock. Starting from the performance corporation, we should consider corporate capability in 5 terms, including profitability, operating capacity, solvency, cash ability and growth ability; for this study, the corporations we analysis are all listing corporations, which arises the importance of the evaluation of stock appreciation ability improve. Therefore, we do this study from 6 terms, and select some appropriate condition attributes, and do quantitative research on evaluation of stock comprehensively.

In terms of profitability: profitability represents the level of enterprises profit in a certain period of time. In order to avoid the differences among companies caused by base's difference. Firstly, we should consider corporate comprehensive profitability, such as profit rate of asset; secondly, we should consider the profitability in aspect of corporate business activities, such as OPE and gross margin; thirdly, we should consider the profitability investors could attain from investing these corporations, such as return on equity.

In terms of operating capacity: operating capacity is mainly used for representing the enterprise asset management efficiency. There is no doubt that corporate asset management efficiency should be considered in evaluating stocks. What's more, in the real asset operation, companies could only operate part of assets to adjust its asset management efficiency, this part of asset is called liquid assets. Combining both 2 terms of consideration, we use turnover of total assets to measure corporate comprehensive asset management efficiency, combined with liquid assets turnover in days, the inventory turnover in days and the receivable turnover in days, we consider corporate liquid assets management efficiency in every term.

In terms of solvency: solvency measures a company's degree of assurance of repayment of debts, reflects corporate financial situation and operating conditions, is the key to achieve sustainable management. This capability includes corporate short-term debt-paying ability and long-term debt paying ability: for the short-term debt paying ability, we use current ratio, quick ratio and cash ratio to measure the enterprise short-term cash payment ability, we use the interest coverage ratio to measure the short-term enterprise ability to bear the interest on the debt; then we use equity ratio and asset liability ratio to measure the enterprise's long-term liquidity cash payment ability.

In terms of cash ability: cash capacity is the ability of company use cash to pay for the other expenses. The better the Cash capacity of company， the more cash company has to pay for the other expenses. The most important is the cash income of enterprise business activities among cash capability, it involves company can give investors probability of benefits after debt service in short term. Accordingly, we select net operating cash flow to sales ratio, net operating cash flow to debt ratio and net operating cash flow per share as condition attributes.

In terms of growth ability: growth ability refers to the company's future development trend and development speed. We mainly consider the capacity for growth of the corporate assets and profits. We first introduce the following financial indicators to measure the increase in their capital: prime operating revenue growth rate, net profit growth rate, NAV, total assets growth rate. At the same time, we select 4 rates of change indicators to measure the growth rate of corporate profits, including ROC of PE, ROC of ROE, ROC of ROA and ROC of gross profit rate.

In terms of stock appreciation ability: stock appreciation is the main source of investment income, the essence of which is to investigate the stock present returns and future potential income. For its existing values, we shall use net assets value per share (adjusted), operating cash flow per share, undistributed earnings per share, share capital reserve, the circulation market value and book-to-market to measure; its future potential value can be measured by PE and P/B. In conclusion, select specific condition as shown in Table 1.

Table 1. Candidate condition attributes

Profitability	Ability to operate	Debt service ability	Cash ability	Stock appreciation ability
Total return on assets	Accounts receivable turnover in days	Interest coverage ratio	NET operating cash flow to sales ratio	Operating cash flow per share
prime operating revenue growth rate	Current asset turnover in days	Debt to asset ratio	NET operating cash flow to debt ratio	Share Capital reserve
Total asset's growth rate	Total assets turnover in days	Equity ratios	Per-share net operating cash flow	Undistributed earnings per share
Net asset's growth rate	Inventory turnover in days	Current ratio		Net assets per share
ROC of gross margin		Quick ratio		The circulation market value
ROE gross margin		Cash ratio		ROC of PE Book-to-market
OPE				P/E
ROC of ROA				P/B
ROC of ROE				

Notes. The unit of absolute index such as net assets per share is yuan, net assets per share is adjusted.

2.1.3 Determination of Decision Attributes

The direct standard whether investors buy a stock is if the stock can bring the income, the income mainly includes issuing dividends and capital gains, where the capital gains ,which is also called the up and down of stock price, is mainly factor that investors concern about. So we start from technical analysis, analyze the stock price fluctuation, make sure the comprehensiveness of the evaluation of stock returns.

1) Abnormal return $RP_{(m,t)}$, it is used to measure the level of company i could excess average return in medicine industry.

$$RP_{(m,t)} = R_{(m,t)} - R_{(f,t)} \tag{1}$$

Where $RP_{(m,t)}$ represents the m -th company's abnormal return in the t -th quarter, $m=1,2,...,M, m=101$, $R_{(m,t)}$ represents the m -th company's return in the t -th quarter, $R_{(f,t)}$ represents the industry's average return

2) Risk-adjusted yield $RR_{(m,t)}$, it is used to measure the stock's return per risk which investors purchase.

$$RR_{(m,t)} = \frac{R_{(m,t)}}{\sigma_{(m,t)}} \tag{2}$$

Where $RR_{(m,t)}$ represents the m -th company's return per risk in the t –th quarter, $R_{(m,t)}$ represents the m -th company's return in the t -th quarter, $\sigma_{(m,t)}$ represents the m -th company's risk in the t -th quarter, it is also called deviation.

3) Probability of beating the market $P_{(m,t)}$, it is used to measure the probability that the corporate investors' return is higher than average return in the equity.

$$P_{(m,t)} = \frac{\sum_{k=1}^{N} T_{(\mathrm{m},k,t)}}{N_t} \tag{3}$$

Where $P_{(m,t)}$ represents the m -th company'sprobability of beating the market in the t -th quarter, $T_{(\mathrm{m},k,t)}$ is 0-1 variables, we suppose 1 means the m -th company in the k-th day beat the market, 0 means the m -th company in the k-th day lose the market. N_t means he number of days in the t -th quarter.

4) The slope of the trend line $slope_{(m,t)}$, it is used to measure corporate stock price changing trend. We could use the m -th company's daily stock price in the t -th quarter to build a regression, and then draw a trend line, this trend is $slope_{(m,t)}$.

2.2 Attributes Discretization

Before we apply financial indicators for each business in the pharmaceutical industry to rough set, we require each discretization of condition attributes values because of the characteristic of rough set. And the values of discrete need appropriate intervals as a precondition, because there is inevitably missing some data or having some abnormal data during data collection, which will affect the efficiency of discretization. So, we need to determine the proper interval for each attribute before scattering both condition attributes and decision attributes.

In addition, when we select 32 condition attributes and 4 decision attributes which we need to analyze, we find there are still having a significant seasonal variation. This suggests that when we scatter the data, we should consider seasonal factor of data. Therefore, when we scatter the data, not only should we meet conditions of using of rough sets, but also preserve data characteristics and normal development level, the steps are as follows:

Step1**:** The determination of the time interval.

We select the data from January 1, 2012 to March 31, 2014, taking into account seasonal variations in data, divide data in accordance with the quarterly into 9 stages, naming a quarter as a period (the first period is from January 1, 2012 until March 31, 2012).

Step2**:** The determination of the appropriate interval.

In order to make data truly reflecting the actual level of economic development and avoid interval dislocating because of the interference of outliers. So we use the data to draw boxplots, which could help us to achieve the elimination of abnormal value which are not in the normal range, then we commend the upper and lower boundaries of the normal data as each attribute's rang. It is noteworthy that there may be significant differences may be different quarter intervals for the same attribute, due to seasonal factor.

Step3: The discretization of data.

Considering that the fundamentals of pharmaceutical companies were investigated in divided periods in the study, therefore all attributes of each period should be divided into equal n intervals, and then assign them according to the ascending order. At the same time, the outliers eliminated from step2 should be involved again. Set the value of outliers under the proper interval as 0, and set the value of outliers above the proper interval as n+1. Finally get the discrete value attributed to different seasons. And the distribution range are $0,1,2,\cdots,n+1$.

2.3 Attribute Reduction and Weight Determination

2.3.1 The Attributes Reduction Model Based on Index System of the Approximate Quality

Rough sets' the index system of information system is $S=(U,A,V,f)$, of which the evaluation object set U is a non-empty and finite set of evaluation objects. Set A is a non-empty and finite index set, including condition attributes set C and decision attributes set D, of which condition attributes set C is made up of fundamental indexes. That is to say, the formula $A=C\bigcup D$ and $C\bigcap D=\varnothing$ is established. Set V is the value set of condition attributes and decision attributes set, f is an information function.

When $P \subseteq A, X \subseteq U, x \in U$, the lower approximation of set X on I is $apr_{-p}(X) = \bigcup\{x \in U : I(x) \subseteq X\}$, it represents the set that we assure it belongs to evaluation object of X according to the judgment of index subset P. $POS_P(X) = apr_{-P}(X)$ referred to as P positive region of X.

If the number of equivalence classgenerated by the index set A and the number of equivalence class generated by the index set $A - a_i$ is the same, the index a_i is unnecessary condition attribute for index set A. Otherwise the index a_i is necessary condition attributefor index set A. If $\forall a_i \in A$ is necessaryfor index set A, the index set A is called as independent index system.

When the evaluation results $D = \{d\}$, the approximation quality γ_C of condition attributes set C made up of fundamental indexes is:

$$\gamma_C(d) = \frac{|POS_C(d)|}{|U|} \tag{4}$$

Index dependencies: If the approximation quality γ_C=1, it means the knowledge of evaluation object depend entirely on the fundamental index system C. If $0 < \gamma_C < 1$, it means some of the knowledge of evaluation object is dependent on the fundamental index system. If γ_C=0, it means knowledge of evaluation object is completely independent on the fundamental index system (Li, 2009). In the screening of the index system, therefore, should try to ensurethe approximatequality of the index system close to 1. Only by this way, the knowledge of the object will better preserved in the index system after the reduction. In addition, considering the mad of index system, we introducing the reduction degree index γ_n to ensure the adequate reduction of the index system, its formula is:

$$\gamma_n = \frac{l - l_r}{l} \tag{5}$$

Where l represents the number of the indexes, l_r represents the number of the indexes after the reduction.

Based onthe approximate quality ofindex system, the process of screening index is as follows:

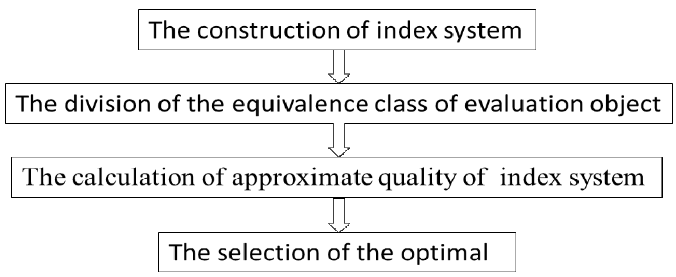

Figure 1. Process of screening index

2.3.2 The attributes Reduction Model Based on Genetic Algorithm

Attribute reduction can be considered as nonlinear optimization problems. Combining with this study, its essence is a reduction of the index system including 32 indexes in order to get a simple index system. This needs traverse 4 294 967 295 times on different index system, to obtain the excellent index system who has the high approximate quality and the contracted degree. However, the efficiency of this reduction method of traverse type is very low. Therefore, the study use 32 virtual 0-1 variables to represent if the index is retained and combines with the genetic optimization algorithm to remove redundant indexes. By this way, in the premise of the quality of approximation of reduced index system, we increase the degree of reduction and efficiency. As a result, we get the reduced index system (Liu & Chen, 2010).

Combined with the genetic algorithm, based on the approximate quality of index system, index screening procedure is as follows:

Step1: Input the complete decision table and eliminate duplicate rows, keeping only one.

Step2: Produce the initial population of which the size is $M = 30$ of. A chromosome carrying l =32 genes is used to represent each individual. Each of these genes corresponds to a condition attribute and value of 0 or 1. For the index system $(c_1, c_2, ..., c_l)$, for example, if an individual contains the j-th $(j = 1, 2, ..., l)$ attribute, thevalue of j-th gene willequal 1, otherwise 0. In other words, the reduced index system is

$\text{reduct}(C)=\text{reduct}(C)\cup\{c_j\}$, where the value of j-th gene is 1.

Step3: Construct the objective function and calculate the target population. Try to ensure the approximate quality of the reduced index system close to 1. At the same time, as far as possible to make the index system adequate reduction.Building the objective function is as follows:

$$\min f(c_1,c_2,...,c_l)=-w_1\times\frac{l-l_r}{l}-w_2\times\gamma_C(d) \tag{6}$$

Where l_r represents the number of genes, whose value equal to 1, in chromosome. $\gamma_C(d)$ represents the approximate quality of the index system.

Step4: Use the way of a single point of crossover to inherit.Design as follows, for the selected two parent individuals $f_1=c_1c_2...c_l, f_2=c_1'c_2'...c_l'$, with crossover probability = 0.8, randomly select the t genes as the point of crossover. As a result, we can get their offspring code s_1 and s_2 to ensure that off spring can inherit the excellent characteristics of their parent.

Step5: Use uniform mutation operator toimplement mutation. For the selected individuals, with mutation probability = 0.35, randomly produce 3 integers, which meet the conditions $1<u<v<w<l$. Then, the gene segments between u and v (including u and v) is put behind w . As a result, we can generate a new individuals, so as to realize the diversity of population, also guarantee the global optimization.

Step6: According to the principle of survival of the fittest, using a deterministic selection strategy that is to choose individuals of which the size is $M=30$ with the smallest value of objective function to give birth to the next generation. So as to ensure the good qualities of parent being preserved, rather than being eliminated.

Step7: Determine the maximum number of iterations of the stopping rule. When the number of iterations achieve the maximum number of iterations (G = 100), stop operation and select individual with the biggest value of objective function. Taking out the position where genes are equal to 1, namely which indexes are necessary to retain. As a result, get the final reduced index system.Otherwise, go to step4.

2.3.3 The Solution of Attributes Reduction Model

For of the reduced index system can reserve enough information and the index system is adequate reduction, double objective optimization model is established and by using the genetic algorithm to solve it.The objective function is as follows:

$$\min f(c_1,c_2,...,c_l)=-w_1\times\frac{l-l_r}{l}-w_2\times\gamma_C(d)$$
$$\gamma_C(d)=\frac{|POS_C(d)|}{|U|} \tag{7}$$

Where l_r represents the number of genes, whose value equal to 1, in chromosome. $\gamma_C(d)$ represents the approximate quality of the index system.

Using genetic algorithm function Ga in the software Matlab , to reduce the index system for the first quarter of 2012 the decision attribute of excess yield. The result is shown in Figure1.

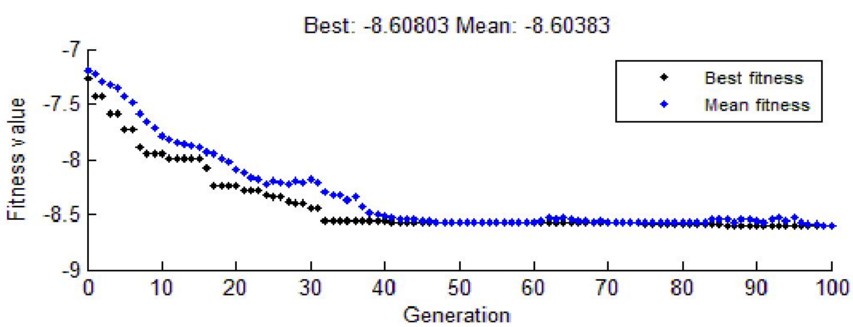

Figure2. Iteration of target value f

From the Figure 2, we can conclude that the group of optimal target value f and the average target value $\overline{f}$ gradually decreases with the increase of the number of iterationsat the start of the iterative process. When the number of iterations reaches 40 times, the optimal target value tends to be stable and converges to the optimal value f= -8.608 . After deleting redundant indexes, the index system after the reduction is $\{c2, c9, c10, c12, c15, c24, c30\}$. The number of index system from 32 reduce to 7 and the approximationquality γ_C is equal to 0.952. The result shows that through the approximation quality of attributes and combined with the genetic optimization algorithm to eliminate the redundant indexes, we can get the adequate reduced index system and reserve the most original knowledge.

2.3.4 The Weight Model Based on the Attribute Importance

After the attributes reduction, the important of each index in the reduced index system should be measured. This needs to dig out the influence degree in decision attribute by each index. Therefore, the study is based onattribute importance to objectively determine the weight of each index. By this way, to truly reflect the contribution of each index to decision attribute in the reduced index system.

By using the thought sensitivity analysis, the study analyzes the degree of change of classification when an attribute is removed from the reduced index system. And the degree of change represents the importance of the attribute. The classification of variations is larger after removed a certain attribute, the attribute significance of this attribute is higher. Otherwise, the attribute significance of this attribute is lower. The attribute significance $\mathrm{Sig}(c)$ of index c_j for the index system C defined as:

$$\mathrm{Sig}(c_j) = 1 - \frac{\left|C \cup \{c_j\}\right|}{|C|} \tag{8}$$

Where C represents the reduced index system removed the index c_j . $|C|$ represents $|IND(C)|$.

When $\dfrac{U}{|IND(C)|} = \dfrac{U}{C} = \{C_1, C_2, ..., C_k\}$,

namely according to the theory of equivalence relation, the evaluation object set U is divided into k classes. And $|C| = |IND(C)| = \sum_{j=1}^{k} |C_k|^2$, $|C_k|$ represents the number of individuals in the k-th class.

If the index c_j is not important to the index system C , the $\mathrm{Sig}(c_j)$ will equal to 0 that means $IND(C \cup \{c_j\}) = IND(C)$ (Li, 2009).

2.3.5 Synthesis of Index Weight Based on Multiple Decision Attributes

To choose high quality stocks which better than the industry level, we need to consider four decision attribute, including abnormal return, risk-adjusted yield, probability of beating the market and the slope of the trend line. However, the contribution degree of the condition attribute for four decision attributes is different. Therefore, the study respectively uses the discrete numerical of the four decision attributes to filtered to all condition attributes and objectively determine the weights of the reduction conditions attributes by the above weight model. Corresponding to the four decision attributes, each condition attributes get the corresponding weights of the four groups. At last, four groups weights numerical of each condition attribute is average processing, as the initial weight of the condition attribute of this season. Its formula is as follows:

$$w^0_{(t,j)} = \frac{\sum_{z=1}^{4} w^0_{(t,j,z)}}{4} : \tag{9}$$

$$W^0_t : \{w^o_{(t,j)} \mid j = 1, 2, ..., l\}, (t = 1, 2, ..., 7);$$

Where $w^0_{(t,j,z)}, (\mathrm{z} = 1, 2, 3, 4)$ refers to the weight of the j-th condition attribute forthe z-th decision attribute. $w^0_{(t,j)}$ refers to the comprehensive weight of the j-th condition attribute in the t-th period, a total of l condition attributes.

2.4 Adaptive Evaluation Model Based on Fundamentals

In considering the seasonal factors, we drive period of the data (from January 1, 2012 to March 31, 2014) to 9 period, we define the first quarter of 2012 (January 1st to March 31st) as the 0th period, the second quarter of

2012 (April 1st to June 30th) as the first period, and can deduce the rest from this, the first quarter of 2014 is defined the eighth period. And t represents the No. of period, $t = 0,1,2,\ldots,8$.

2.4.1 Adaptive Weight System

According to the fourth part model, after reduce condition attributes and achieve its weight determination, we could use decision attributes in the present period (t) and condition attributes in the last period (t-1) to obtain a set of weighting system. We define it W_t^0, then:

$$W_t^0 : \left\{ w_{(t,j)}^0 \mid j = 1,2,\ldots l \right\}, (t = 1,2,\ldots 7) \tag{10}$$

Where $w_{(t,j)}^0$ represents the j-th condition attributes in the t–th period weight system, this system includes l condition attributes.

2.4.2 Period-Modification Weight-Determination Model

The weighting system we mentioned only use two periods data to build weights to evaluate companies comprehensively and to do further selection, so it is impossible to avoid chance. Therefore, in order to determinate the weighting system more accurately, we need to use more data as our sample to buid weight modification based on period (combined with the data in past period)

We hypothesis the t-th period-modification weighting system as $W_t^1 : \left\{ w_{(t,j)}^1 \mid j = 1,2,\ldots l \right\}$, so the -th weight in this set of weighting system is:

$$\begin{aligned} w_{(t,j)}^1 &= f_0 \cdot w_{(t,j)}^0 + (1 - f_0) \cdot w_{(t-1,j)}^1 \\ &= \sum_{i=1}^{t} f_0 \cdot (1 - f_0)^{i-1} \cdot w_{(t,j)}^0 \quad , \quad (t \ge 2) \end{aligned} \tag{11}$$

Where f_0 is period-modification coefficient, we assume it as a constant, and define $f_0 = 0.6$, it shows apparently that if the past period of data is closer to the modification period, the data in this past time would be defined a bigger weight.

2.4.3 Adaptive-Modification Weight-Determination Model

Using index weight and the weighted distance square method to determine the enterprise of -th period whose decision attribute is closest to the enterprise of $t+1$ t-th period. Then, the decision attribute of the closest the enterprise of t-th period is as expected decision attribute of the enterprise of $t+1$-th period. Compared the prediction of decision attribute and the real of decision attribute, we can evaluate the goodness of the index weight. Combined with the factor of goodness, higher goodness of the past index weight, higher influence degree on the latest index weight so as to realize the adaptive. Combined with the time factors, we make the importance of index weight decreases with time. As a result, we get the index weight according to the adaptive of goodness and time.

Firstly, in order to evaluate past-period weight system, we commend the next-period return as our standard. We scatterreturn rate, if the high next-period return has a high forecasted decision attributes, we could say this weight system has a high goodness. The following is our model:

We scatter corporate return in the $t+1$-th period, get the real decision attributes set:

$$Var_c = \min_{n_c}\{Var_{(c,c')}\} = \min_{n_c}\{\sum_{j=1}^{l} w_{(t,j)} \left| C_{(t+1,c,j)} - C_{(t,c',j)} \right|^2\} \ (c' = 1,2,\ldots,n_c) \tag{12}$$

Where $r_{(t+1,c)}^R$ represents the c-th corporate discrete return in the $t+1$-th period, n_c represents the number of companies.

For the corporate condition attributes set $C_{(t+1,c)}$ in the $t+1$-th period, and the corporate condition attributes set $C_{(t,c)}$ in the t-th period, we calculate every company's least sum of squares of deviations in the $t+1$-th period:

$$Var_c = \min_{n_c}\{Var_{(c,c')}\} = \min_{n_c}\{\sum_{j=1}^{l} w_{(t,j)} \left| C_{(t+1,c,j)} - C_{(t,c',j)} \right|^2\} \ (c' = 1,2,\ldots,n_c) \tag{13}$$

Where $C_{(t+1,c,j)}$ represents the j-th condition attributes of the c–th company in the $t+1$-th period, $C_{(t,c',j)}$ represents the j-th condition attributes the c'-th company in the t-th period, n_c represents the number of companies, this system includes l condition attributes.

If the c-th company in the $t+1$-th period and the c''-th company in the t–th period assess the least sum of squares of deviations, the we could define the forecasted decision attributes of the c-th company in the $t+1$-th period the same as the decision attributes of the c''-th company in the t-th period, that is to say:

$$s_{(\mathrm{t+1,c})}^{s1} = s_{(\mathrm{t,c''})}^{s1}\ ,\ \ c = 1,2,\ldots n_c \tag{14}$$

Where $s_{(\mathrm{t,c''})}^{s1}$ the decision attributes of the c''-th company in the t–th period. We apply this approach to all companies in the $t+1$-th period, forecasting every corporate decision attributes in the $t+1$-th period, then we attain every corporate forecasted decision attributes in the $t+1$-th period:

$$S_{t+1}^{S1} = \{s_{(\mathrm{t+1,c})}, c = 1,2,...,n_c\} \tag{15}$$

Afterward, we calculate correlation coefficient between forecasted decision attributes sequence S_{t+1}^{S1} in the $t+1$-th period and real return sequence S_{t+1}^{R}, the bigger the correlation coefficient is, the higher the goodness of the t-th weight system W_t^1 is:

$$Corr_t = Corr(S_t^{S1}, S_{t+1}^{R}), (t = 1,2,\ldots,7) \tag{16}$$

Where $Corr_t$ is the t-th correlation coefficient sequence, correlation coefficient sequence is:

$$CORR_t : \left\{ Corr_i \middle| i = 1,2,\ldots t \right\}, (t = 1,2,\ldots,7) \tag{17}$$

Because of $Corr_t \in [-1,1]$, if we use $Corr_t$ to modify $w_{(t,j)}^{1}$, the uncertainty of comprehensive score index $CS_{(t,c)}^{1}$ would go up, so we need to use a new coefficient—goodness-modification coefficient, we assume the goodness-modification coefficient in the t–th weight system W_t^1 is g_t:

$$g_t = \frac{Corr_t}{Average(CORR_t)}, (t = 1,2,\ldots,7) \tag{18}$$

This expression shows that if the correlation coefficient in a period is higher average correlation coefficient among all past periods, then $g_t > 1$, or $0 < g_t \le 1$. So we could obtain the weight of condition in the adaptive-modification weight-determination model:

$$\begin{aligned} w_{(t,j)}^{2} &= \sum_{i=1}^{t} g_t \cdot w_{(t,j)}^{1} \\ &= \sum_{i=1}^{t} g_t \cdot f_0 \cdot (1-f_0)^{i-1} \cdot w_{(t,j)}^{0} \end{aligned} \tag{19}$$

$$W_t^2 : \left\{ w_{(t,j)}^{2} \mid j = 1,2,\ldots l \right\}$$

Up to now, there is still a question we need to solve: the weight of adaptive modification is in past period, so when we solve the weight system in the t-th period, we define:

$$g_t = g_{t0} = 1 \tag{20}$$

Where g_{t0} is a constant, as fictitious value of goodness-modification coefficient g_t in the present period.

2.4.4 The Screening – Evaluation Model of Stocks

First of all, according to decision attribute prediction model based on adaptive-modification weight-determination, we can predict the decision attribute set $S_{t+1}^{s} = \{s_{(\mathrm{t+1,c})}, c = 1,2,...,n_c\}$ in a period and by this way to evaluate the enterprise stocks. Then we select the enterprise stocks $S_{t+1}^{s'} = \{s_{(\mathrm{t+1,c})}, c = 1,2,...,N\}$, whose decision attribute is greater than or equal to 4. The number of these enterprise stocks is N.

Then we attain real decision attribute set $S_{t+1}^{R} = \{s_{(\mathrm{t+1,c})}^{R}, c = 1,2,...,n_c''\}$, according to the discretization of the real yield, and we select t the enterprise stocks $S_{t+1}^{R'} = \{s^{R}_{(\mathrm{t+1,c})}, c = 1,2,...,\mathrm{N}''\}$ whose real decision attribute is greater than or equal to 4.

Finally, we calculate the proportion of the companies which really have good performance in terms of return rate, the denominator value is the number of companies we predict according to its prior fundamentals, molecular is the number of companies which really have high return rate in companies (the denominator) we predict have good performance.

$$p = \frac{count\{S_{t+1}^{s^*} \cap S_{t+1}^{R^*}\}}{N} \quad (21)$$

The $p = 1$ shows that all real decision attribute values of those chosen stocks, by the model of stock screening, are greater than or equal to 4. That is to say, the stock screening model has very high credibility.

3. Results

3.1 Index selection & Initial Weights Determination

First of all, using the approximation quality of the index system, combined with genetic algorithm to delete redundant condition attributes, get different initial reduced index system during different periods. The reduced condition attribute system as shown in Table 2.

Table 2. Result of the initial reduced attributes

Period(t)	The code set of reduced condition attributes
1	2,3,7,8,9,10,12,15,19,21,24,27,29,30,32
2	1,2,5,6,8,9,11,15,17,20,22,23,24,28,30,32
3	1,2,4,8,9,10,11,12,14,18,20,21,22,24,30,31
4	1,3,4,5,6,7,8,11,12,13,20,21,30,32
5	1,2,3,5,6,7,9,10,19,24,26,27,29,30,31,32
6	1,3,5,6,7,8,9,11,12,16,17,19,21,24,28,29,30
7	1,4,7,8,11,12,13,15,16,21,24,27,29,30,32

For Table 2, reduced index system, through the basic model mentioned above, based onthe important degree of the attributes, we can get the initial weights of condition attribute in each period. According to the weight size screen the former nine condition attributesin every period. Table 3 shows codesof condition attributes and their corresponding weights under the basic model:

Table 3. The condition attributes and the initial weights under the basic model

Period(t)	1	2	3	4	5	6	7
Codes of condition attributes (The initial weights)	30 (0.071)	30 (0.122)	8 (0.063)	11 (0.047)	29 (0.045)	9 (0.083)	11 (0.033)
	21 (0.053)	5 (0.049)	21 (0.059)	5 (0.045)	2 (0.036)	30 (0.045)	8 (0.013)
	32 (0.04)	22 (0.041)	30 (0.047)	4 (0.032)	3 (0.036)	7 (0.039)	30 (0.021)
	8 (0.024)	20 (0.039)	9 (0.04)	32 (0.027)	32 (0.034)	1 (0.035)	24 (0.027)
	12 (0.021)	2 (0.034)	12 (0.04)	12 (0.027)	1 (0.031)	11 (0.03)	21 (0.02)
	24 (0.02)	8 (0.029)	1 (0.025)	30 (0.02)	27 (0.026)	8 (0.029)	15 (0.021)
	7 (0.016)	11 (0.021)	2 (0.014)	1 (0.019)	7 (0.025)	24 (0.026)	13 (0.02)
	10 (0.016)	23 (0.02)	20 (0.013)	8 (0.019)	26 (0.021)	6 (0.026)	12 (0.014)
	3 (0.015)	28 (0.016)	22 (0.013)	21 (0.013)	6 (0.021)	3 (0.021)	29 (0.02)

3.2 Solution of Adaptive-Modification Model

As stated earlier, according to Table 3 we consider periodfactors and goodness factors. Through modified the above conditionsattributes based on periodmodification and goodness modification, we can get the revised condition attributes and the corresponding weights and screen the former nine condition attribute condition in every period according to the weight size. By the adaptive-modification weight-determination model, we can get condition attributes and the corresponding weights. The specific situation are shown in Table 4.

From Table 4, we could conclude that the part of relatively good condition attributes stand outwith the moving of the period. Statistics the frequency of the 7-th condition attributes in the previous 0-th~6-th period, we can get the result in Table 5.

FromTable 5, it is not hard to see that high frequency ofthe condition attribute in the reducedcondition attribute of every period. The condition attribute is eventually preserved, which largely reduce the accident of the selection of indexes.

Table 4. Condition attributes and its initial weights in adaptive-modification model

Period(t)	1	2	3	4	5	6	7
Codes of condition attributes (The modified weights)	30 (0.122)	30 (0.148)	8 (0.107)	11 (0.11)	29 (0.12)	9 (0.015)	32 (0.014)
	5 (0.049)	5 (0.06)	21 (0.098)	5 (0.107)	2 (0.097)	30 (0.01)	11 (0.008)
	22 (0.041)	22 (0.049)	30 (0.086)	4 (0.075)	3 (0.097)	1 (0.009)	30 (0.008)
	20 (0.039)	20 (0.047)	9 (0.067)	32 (0.065)	32 (0.096)	7 (0.009)	24 (0.007)
	2 (0.034)	2 (0.041)	12 (0.066)	12 (0.063)	1 (0.086)	3 (0.007)	1 (0.006)
	8 (0.029)	8 (0.035)	1 (0.043)	30 (0.053)	27 (0.07)	6 (0.006)	9 (0.006)
	11 (0.021)	11 (0.025)	2 (0.026)	8 (0.047)	7 (0.067)	11 (0.006)	29 (0.006)
	23 (0.02)	23 (0.025)	20 (0.025)	1 (0.046)	6 (0.058)	8 (0.005)	7 (0.005)
	28 (0.016)	28 (0.019)	22 (0.025)	21 (0.031)	26 (0.057)	24 (0.005)	8 (0.005)

Table 5. Frequency of condition attributes in 7-th period

Name of index	Code of index	weight	Times	Frequency
ROC of gross margin (%)	32	0.014	5	71.4%
Total asset's growth rate (%)	11	0.008	5	71.4%
Book-to-market (%)	30	0.008	7	100.0%
Operating cash flow per share(yuan)	24	0.007	6	85.7%
Prime operating revenue growth rate (%)	9	0.006	5	71.4%
Share capital reserve(yuan)	1	0.006	6	85.7%
P/B (%)	29	0.006	4	57.1%
Gross margin (%)	7	0.005	5	71.4%
ROE (%)	8	0.005	6	85.7%

3.3 Solvation of Selection Model

Based on the stock screening-evaluation model, we select 28 companies whose forecasted decision attribute is above or equal to 4. And we attain these companies' real decision attributes in the 8th period, the data we mentioned is as follow, Table 6.

Table 6. Real decision attributes (≥4) of companies in the 8th period

No.(company)	real decision attributes	No.(company)	real decision attributes	No.(company)	Realdecision attributes
1	3	36	4	73	4
3	5	42	4	77	3
5	4	48	4	81	6
10	5	51	3	82	4
13	4	53	3	84	3
15	4	56	3	85	4
21	2	58	4	88	3
22	3	63	2	97	3
26	4	65	5	--	--
31	5	70	4	--	--

Table 6 suggests that there are 17 companies really have good performance in technical analysis (real decision attribute ≥4) in 28 companies which we forecast according to fundamentals (forecasted decision attribute ≥4) in last period. The accuracy rate is up to 60.7%.

4. Discussion

For the construction of measurable stock-selection, this study combines fundamental analysis and technical analysis, selects some fundamental index which make important difference in stocks' return based on the rough set model, to develop the financial position and operating performance of enterprise. And then we commit the technical index as our standard, to study the fundamental index' effects on its shares on the securities market performance, so that we could look for more valuable, more growth stock portfolio in China's stock market. In addition, in the face of enormous information, this study uses genetic algorithm to optimize a rough set model, greatly improve the arithmetic speed with promising the effect of optimization.

Based on a quarter as a cycle, we firstly use technical indexes in the second to eighth period and the fundamental indexes in the first to seventh period. So we could screen for seven cycle system's condition attributes and the corresponding weights, and thus to evaluate the current fundamentals of each company, in order to sort, filter for companies to realize the purpose of the stock selection. Among them, the article gives full consideration to the trend factors and goodness factors. For the trend factors, we select condition attributes from the first period to the seventh period, and the later the period is, the bigger weight the condition attributes have in this period; for the goodness factors, we use the method of the adaptive, so that the more accurate the forecast results the condition attributes have, the bigger weight the condition attributes have. In this process, we achieve condition attributes' selection from one period to another and the determination of weights, and corrects the condition attribute system based on the period and the adaptive optimization factors in different periods, so the generally good condition attributes in every period can stand out, and it reduces the accident of the selection of indicators in a large extent. Finally, on the sorting of the 8th period issue of the shares, decision-making achieved 60.7% accuracy, and it can be used as the basis of fundamental analysis in the practice of the securities investment. Not only this set of model could be used to another industry's stock selection strategy, but also could be used to different industries' stock selection strategy.

Acknowledgments

Thanks Haijun Huang, Qing Su and Fanfeng Zeng for their personal assistance in manuscript preparation. We are also particularlygrateful to Jinan University.

References

Asness, C. (1997). The interaction of value and momentum strategies. *Financial Analysts Journal,* (3/4), 29-36. http://dx.doi.org/10.2469/faj.v53.n2.2069

Bao, X. Z., & Xiao, M. (2010). Portfolio investment decision-making based on rough set theory. *Systems Engineering-Theory Methodology Application, 19*(5), 526-533.

Frankel, R., & Lee, C. M. C. (1998). Accounting valuation, market expectation, and cross sectional stock return. *Journal of Accounting and Economics, 21*(7), 283-319.

Guo, C. G. (2007). *Rough Set Attribute Reduction Algorithm Based on Genetic Algorithm.* HeFei University of Technology.

Jiang, F. M. (2013). *A-share stock portfolio analysis base on multiple-factor quantitative models.* Southwest Jiaotong University. http://dx.doi.org/10.1016/S0165-4101(98)00026-3

Li, Q., & Zhang, J. etc. (2008). Effectiveness test of investment ways MCSD based on investment scale of listed companies of china's a-share market. *Systems Engineering, 26*(8), 56-64. http://dx.doi.org/10.3969/j.issn.1001-4098.2008.08.010

Li, W. S., & Yi, Z. (2010). Rough set attribute reduction algorithm based on genetic algorithm. *Microelectronics, 3*, 71-74.

Li, Y. Y. (2009). *Research on the construction of indicator system and comprehensive evaluation method based on rough set.* Wuhan University of Technology, pp. 48-68.

Liu, B., & Chen, D. J. (2010). Analysis of road traffic accidents based on rough sets and genetic algorithms, *Journal of Lanzhou Jiaotong University, 29*(1), 69-71. http://dx.doi.org/10.3969/j.issn.1001-4373.2010.01.017

Lu, Y. J., Liu, N. X., & Chen, L. (2008). A rough set attribute reduction algorithm based on parallel genetic algorithm. *Computer Science, 35*(3), 219-221. http://dx.doi.org/10.3969/j.issn.1002-137X.2008.03.068

Ma, Y. K., & Tang, X. M. (2004). Study on multi-factor model for portfolio investment decision based on the market benchmark. *Systems Engineering-theory & Practice, 24*(7), 30-37. http://dx.doi.org/10.3321/j.issn:1000-6788.2004.07.004

Mohanram, P. S. (2005). Separating winners form losers among low book-to-market stocks using financial statement analysis, Review of Accounting studies, 10, 133-170. http://dx.doi.org/10.1007/s11142-005-1526-4

Perold, A. (1984). Large-scale portfolio optimization. *Man. Science, 30*, 1143-1160. http://dx.doi.org/10.1287/mnsc.30.10.1143

Piotroski, J. D. (2000). Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research, 38*, 1-41. http://dx.doi.org/10.2307/2672906

Sharpe, W. (1964). Capital Asset Prices: a theory of market equilibrium under condition of risk. *Journal of Finance, 19*, 425-442. http://dx.doi.org/10.1111/j.1540-6261.1964.tb02865.x

Zhang, L. P. (2014). *Stock selection base on multiple-factor quantitative models.* Hebei University of Economics and Business.

Copyrights