INFORMATICA, 2001, Vol. 12, No. 4, 573–584

© 2001 *Institute of Mathematics and Informatics, Vilnius*

Forming of the Investment Portfolio Using the Self-organizing Maps (SOM)

Giedrius STANKEVIČIUS

Kaunas University of Technology
Studentų 48, 3028 Kaunas, Lithuania
e-mail: giedrius@earthling.net

Received: October 2001

Abstract. The problem of comparison of different companies is facing, when looking for possible candidates for the investment portfolio. Screening of the companies, using "well-known" trading strategy parameters, is one of the ways to solve this problem. Actually, using this procedure much more companies appear on the list, than the trader is willing to buy. To define the best companies or group of the best companies self-organizing (Kohonen's) map (SOM) could be used. Using fundamental financial parameters as inputs, the output of SOM forms the different groups of companies located into a number of disjoint clusters.

Then, by the special averaging technique, the 3D map of quality of investment could be formed. Investing portfolios also could be formed by simple technical analysis approach.

Non-linear ranging technique was applied as an alternative to self-organizing map procedure in this paper. The certain meanings of weights were given to the factors, which characterize the companies. Then, by estimation of all weights, companies were assigned to their place in the general listing.

Four different portfolios were formed as a result of these researches. The performance of these portfolios showed which of the researched techniques gave better result. The real data from USA stock markets was used for the realization of the whole idea.

Key words: self-organizing maps, ranging techniques, stock market.

1. Introduction

Due to the globalization of financial markets, expansion of the electronic trade and the growth of information about the market, the specialists of investment funds more frequently try to use artificial intelligence methods for the market analysis (Curry *et al.*, 1997). These methods are widely used in so-called intelligent process control and monitoring systems. Works related with the use of neural networks, obscure sets methods, fuzzy logic and expert systems for financial analysis and formation of trade decisions, receive the greatest recognition and interest in financial markets today.

The possibilities of SOM and ranging techniques in financial markets analysis are analyzed in this work.

2. Screening of the Companies

The number of individual investors increased dramatically in recent years with the growth of electronic trade services. The "right" choice of the "right" company is one of the biggest problems the individual investor is facing. The use of different screeners for selection of stocks is one of the ways to solve this problem. The choice of the factors and parameters, which most comprehensively characterize the analyzed companies, is the case there. The financial analysts more frequently try to combine the methods of technical and financial analysis during stock selection procedures lately. That's what the individual investors also should do.

It is accepted that price changing tendencies could be established with the use of prehistoric data of stock parameters and factors [2]. There are some of the parameters, which are used for the technical stock analysis:

1) shares outstanding;
2) opening price of share;
3) closing or last bid price of share;
4) highest price of share during the trading day;
5) lowest price of share during the trading day;
6) average daily trading volume;
7) the values of different stock indexes, etc.

The detailed analysis of the company activities and its financial reports is carrying out, when using fundamental stock analysis. Stock market analysts turn their attention at these parameters (Allrich, 1995):

1) earning per share growth;
2) return on assets and equity ratios;
3) debt to assets and equity;
4) net profit margin, etc.

To build the list of the companies for the possible investments certain meanings of mentioned parameters should be chosen. One of the possible ways to do so is to follow the recommendations of financial analysts again.

In our case the "Netscreen" screener was chosen for the screening of the companies. The meanings of the parameters for the screening were chosen according to the recommendations of William O'Neil (1995) and Ted Allrich (1995).

3. Application of SOM

To create portfolio from the selected stocks after the screening operation, self-organizing maps could be applied. The major function of self-organizing maps is to automatically classify input patterns into a number of disjoint clusters (Kohonen, 2001; Tsoukalas, 1997). The patterns located in the same cluster have similar features. The self-organizing map is formed in terms of unsupervised learning, i.e., learning without a teacher, for instance winner-take-all competitive learning. Here we introduce the algorithm of competitive learning self-organizing maps. In a self-organizing map, a vector quantizer can be performed by adjusting weights from N input nodes to $M \times L$ output nodes. When the input patterns have been presented sequentially to the map without specifying the desired output, the input patterns can be automatically classified into $M \times L$ output nodes or clusters. The structure of SOM and the geometrical explanation of competitive learning are schematically illustrated in Figs. 1 and 2.

The detailed learning algorithm can be summarized as follows:

Step 1. Randomly initialize small valued weights

$$W_{ijk}(0), \quad i = 1, N;\ j = 1, M;\ k = 1, L.$$

Step 2. Present input vector

$$xi, \quad i = 1, N.$$

Step 3. Calculate the distance between the input vector and the weight vector for all individual output nodes (Fig. 2b)

$$d_{jk} = \sum_{i=1}^{N} \big(x_i(t) - W_{ijk}(t)\big)^2, \tag{1}$$

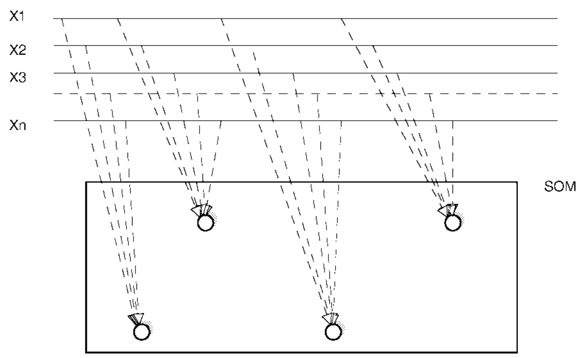

Fig. 1. Topology of Kohonen self-organizing map.

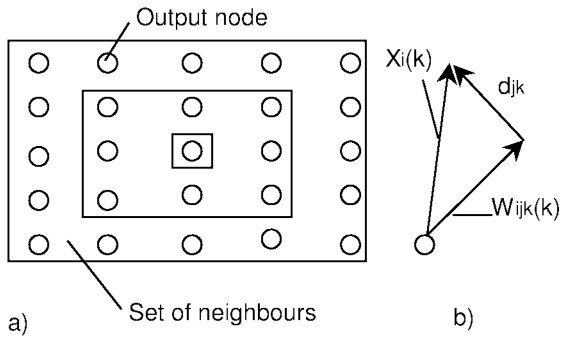

Fig. 2. Geometrical illustration of competitive learning.

where $x_i(t)$ is the input to the node i at time t, and $W_{ijk}(t)$ is the weight from input node i to output node jk at time t.

Step 4. Select the most active output node $(jk)^*$, or the so-called "winner" which has the least distance, i.e.,

$$d_{\min} = \min\{d_{jk}, \quad j = 1, M;\ k = 1, L, \}. \tag{2}$$

If $d_{jk} = d_{\min}$ then $jk = (jk)^*$ and $y^*_{jk} = 1$, otherwise $y_{jk} = 0$ (jk is not equal $(jk)^*$).

Step 5. Upgrade the weights for the "winner" node

$$W_{ijk}(t+1) = W_{ijk}(t) + g(t)y_{ik}\big(x_i - W_{ijk}(t)\big), \tag{3}$$

where $g(t)$ denotes learning rate and is defined as a time decreasing function within the range $(0, 1)$.

Step 6. Repeat by going to Step 2.

From (2), we can see that eventually the weights are upgraded only for the winner node j^*. However, in practice, the weights can also be upgraded only for the winner node j^*, but also for all nodes in the neighborhood of the winner. The size of the neighborhood $NE_j(k)$ can be predefined and can start large and slowly decrease in size with time. The weights upgrading may follow a modified version:

$$W_{ij}(k+1) = W_{ij}(k) + g(k)\big(x_i - W_{ij}(k)\big), \tag{4}$$

for all j which are located in the neighborhood $NE_j(k)$.

The use of SOM is based on the possibilities of SOM to find the similar companies using companies' fundamental parameters and factors during the SOM training procedure. The technical and financial data about the market's state and the company's financial state is used as the inputs for the SOM in this case. The merit there is that relations between different factors are not fixed a priori, but are identified during the training with the help of experimental data. Thus their outputs are safe from so-called "human factor", when the desirable result is obtained.

There are lots of different software packages for the realization of self-organizing maps. In our study, we used "Viscovery SOMine" software.

The list of the companies and their financial ratios was used as an input. Then the clustering operation was implemented. The map with some quite clearly formed groups of the companies appeared as an output (Fig. 3). Each cluster, which represents the set of nodes with the similar features, has different color. Intensity of the colors represents quantitative differences between neighboring clusters.

4. Ranking of the Companies

Alternative way to form the portfolio is the application of ranking procedure to the companies, which appeared on the list after the screening Ted Allrich (1995) in his book "The

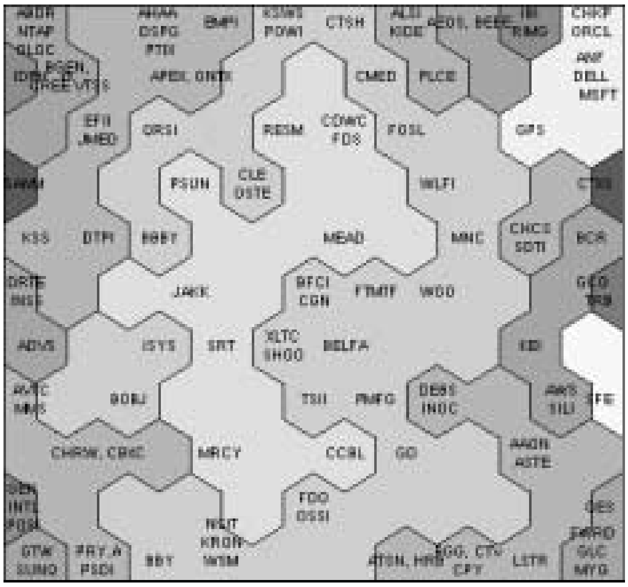

Fig. 3. Clustering result.

On-line Investor" proposed the method of ranking of the stocks of the companies. The basic concept for this procedure is to assign a value for each ratio of the stock, add up those valuations and compare that number with all others. Then it is possible to rank the stocks from the highest value to the lowest.

The most interesting there are the weights used for the ranging. T. Allrich picked 12 main financial parameters and stock ratios that have biggest influence on the stock price change (Table 1). Then he empirically formed non-linear functions of weights, attributed to these financial parameters and ratios. Fig. 4 and Fig. 5 illustrate a couple of non-linear functions used for attribution of weights.

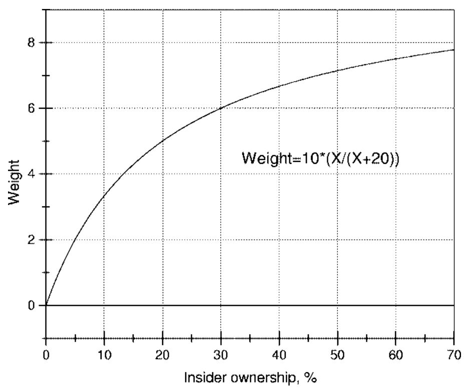

Fig. 4. Non-linear function of weights, attributed to insider ownership parameter.

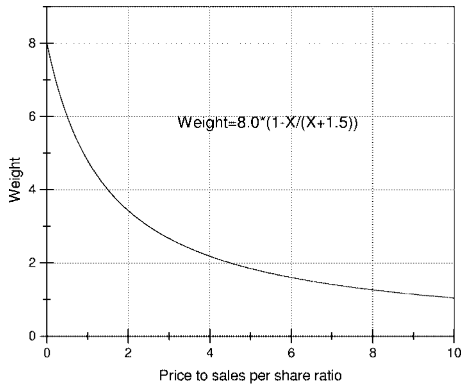

Fig. 5. Non-linear function of weights, attributed to price to sales ratio.

The maximal values of weights are shown on Table 1.

In this paper alternative values for the ranking procedure were proposed, because there were some doubts about the accuracy of T. Allrich proposed maximal values of weights. The alternative values are based on nonlinear regression analysis of the relationships between different stock ratios and the stock price change during last 52 weeks. The performance of about 500 companies in the stock market was analyzed (NYSE, NASDAQ and AMEX markets). The polynomial regression of the 2nd degree was used to describe the relationship between stock ratio and stock price change.

Such relationships were constructed for every stock ratio/parameter. The more complicated non-linear regression didn't improve significantly the root mean square error of estimated relationship. The values of the ratios were determined according to the value of

Table 1

Ranking values

Ratios and parameters	Maximal values of weights (Allrich's ranking)	Maximal values of weights (Alternative ranking)
Free cash flow per share	8	12
Price to sales ratio	8	11
Price earning ratio	10	10
Price to book ratio	7	9
Current ratio	5	8
Return on assets	6	7
Profit margin	10	6
Debt to equity ratio	5	5
Earning per share growth during 12 trailing months	9	4
Return to equity	8	3
Institutional ownership	7	2
Management ownership	8	1

the root square error value – the smaller value of the root square error, the bigger value of the ratio. It appeared that free cash flow per share ratio had the strongest relation with the stock price change. The parameter of management ownership had the weakest relation with the stock price change (Table 1).

Finally, earlier used non-linear functions were adjusted using alternative values of weights.

The investor should pay attention at the top companies on the list after the application of ranging procedure, analyze them closely (graphs, inside data and profile) and most probably buy the stocks of these companies.

5. Results

After the implementation of screening, clustering course and ranging procedures four portfolios of the shares were formed. The data set used for the screening was formed on 23 August 1999.

The formation of the portfolios according to ranging results was quite simple. The companies, which appeared on the top of the list, were chosen for the possible investments. One of the portfolios was named “Allrich portfolio” and another one – “Alternative portfolio”.

The forming of the investment portfolio according to the clustering results was complicated. Despite the fact, that some groups of the stocks appeared quite clearly (Fig. 3), it was still difficult to decide, which one should belong to the investment portfolio. There the simple technical analysis approach was used. Using the averaging technique a 3D map was drawn.

There is the basic idea of the averaging technique. The mean growth (diminution) of the price for each cluster of the map should be calculated. By choosing the “neighbors” of the each cluster in the map the certain area will be covered (Fig. 6). Then the price change for the past 12 trailing months should be calculated for every company, which belongs to the covered area. Finally, the mean of these price changes should be calculated. The covering area of the “neighbors” will vary depending on the chosen cluster.

A 3D map was drawn using the results of the averaging (Fig. 7). High percentage values of the peaks of the map reflect the bullish mood that prevailed in the markets during the analyzed period (1998–1999).

This 3D map was merged with the clustering map for the clarity reasons (Fig. 8). Two portfolios were formed according to this new map. First portfolio consisted of the companies, which were on the peaks of the map. The growth of the share prices of these companies was the biggest during the last 12 trailing months. It was supposed, that the share prices of these companies will grow for some time and then the correction of the prices should emerge. This portfolio was called a “Winners portfolio”.

Second portfolio consisted of the “losers”. The share prices of these companies declined during the last 12 trailing months. It was supposed, that the downtrend will continue for a very short time and the share prices should go upwards shortly.

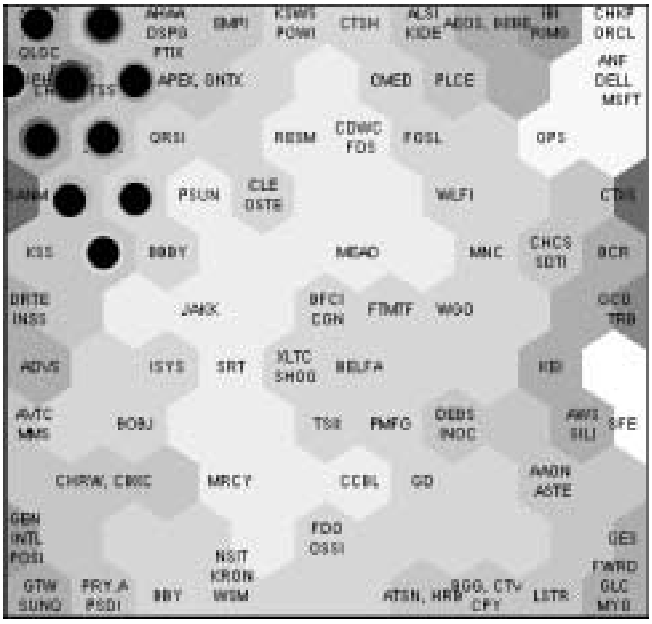

Fig. 6. Selecting the neighbors.

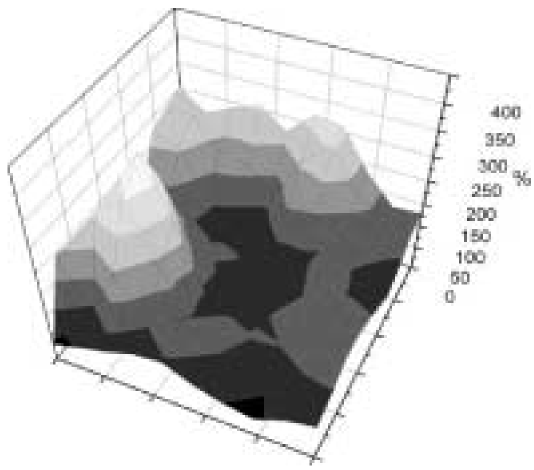

Fig. 7. 3D map.

It was supposed, that 100000 USD are available for the investments of the each portfolio. This imaginary capital was divided into equal parts and the stocks were "bought". The contents of all portfolios are shown on Table 2.

All of the portfolios were tracked from August 1999 to November 1999. The S&P 500 index was also tracked for the comparison of the portfolios performance with the whole situation in the market.

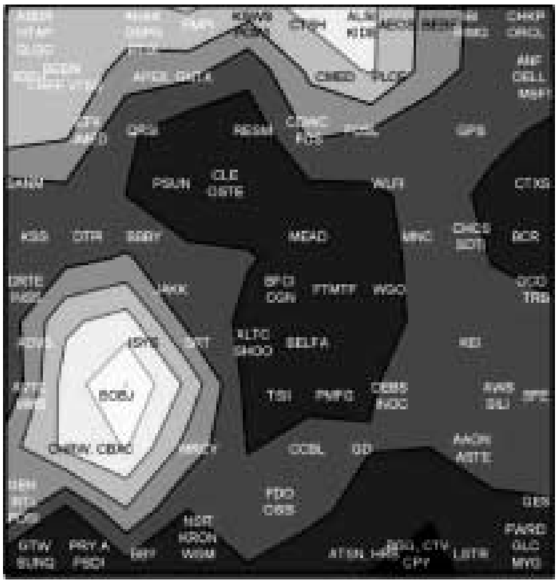

Fig. 8. Map of the clustering results merged with the 3Dmap.

Table 2

Contents of portfolios formed using different methods of selection

Winners portfolio	Losers portfolio	Allrich portfolio	Alternative portfolio
CTSH	GTW	BELFA	PMFG
ALSI	SUNQ	DEBS	KEI
KDE	ATSN	KEI	BELFA
AEOS	HRB	POWI	GCO
BEBE	BGG	KSWS	WGO
BOBJ	CTV	INOC	DEBS
ISYS	CPY	KDE	MNC
CHRW	PSDI	CHKP	INOC

Final research results (Fig. 9) showed, that earlier assumptions about the performance of the “Winners” and “Losers” were not quite correct.

The value of the “Winners” and “Losers” portfolio declined during the first tracked month. So as the value of S&P 500 index.

Value of the“ Winners” and “Losers” went upwards during the following two months “Winners” showed impressive gain during this period. The whole market was on the positive mood too.

“Allrich” portfolios showed stable gains during the whole period. “Alternative” portfolio results were not so good during the first two months, but during the final month it showed some gain, mostly according to the positive mood on the whole market.

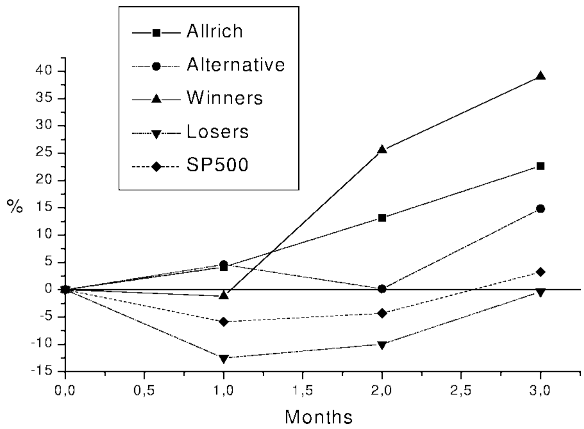

Fig. 9. Performance of the portfolios during the August–November 1999 period.

6. Conclusions

The results showed that ranging techniques and SOM could be applied for the financial market analysis.

Self-organizing maps are a bit complicated when investor should decide which group of companies could form the portfolio. The whole preprocessing of the data for the 3D map is a time consuming job, but the results are quite impressive.

Research also showed, that screening and clustering course should be done every month. By introducing the data of the new companies to the SOM they will appear in one of the earlier formed groups. By tracking the movement of the earlier chosen companies within the map, it is possible to decide, towards which direction new companies will move.

Ranging techniques are quite simple and easy to use. The portfolio formed using Allrich's ranging showed better results than the portfolio formed using alternative ranging parameters. Thus it is possible to come to the conclusion that pure statistical approach cannot reflect clearly the relationships between price change and different financial parameters. The stable gains of "Allrich" portfolio proved that ranging parameters were chosen correctly.

Finally, it should be noted, that further research is required for the improvement of the use of the SOM for the financial market analysis.

References

Allrich, T.(1995). *The On-line Investor. How to Find the Best Stocks Using Your Computer*. New York, St. Martin's Griffin.

Curry, B., P. Morgan, M. Beynon (1997). Neural networks and forecasting: "orthodox" methods and new research. In J.F. Marshall (Ed.), *Conference on Computational Intelligence for Financial Engineering (CIFER)*, New York Crown Plaza Manhattan.

Perk, R. (1990). *Professionelle Aktienanalyse fur Jedermann*. Munchen, Verlag C.H.Beck.

O'Neill, W.J. (1995). *How to Make Money in Stocks*. New York, McGraw-Hill Inc.

Kohonen, T. (2001). *Self-Organizing Maps*, 3rd ed. Springer Series in Information Sciences, **30**, Springer–Verlag.

Tsoukalas, L., R. Uhrig (1997). *Fuzzy and Neural Approaches in Engineering*. New York John Wiley & Sons.

G. Stankevičius is PhD student of Process Control Department, Kaunas University of Technology. The main research field is the use of artificial and computational intelligence in stock markets. G. Stankevičius has published 8 scientific publications and took part in 5 international scientific conferences dedicated to artificial intelligence and financial markets.

Investicinio portfelio formavimas taikant saviorganizuojančius neuroninius tinklus

Giedrius STANKEVIČIUS

Formuojant investicinį portfelį dažnai susiduriama su kompanijų palyginimo problema. Straipsnyje pateikiami du metodai, kuriais remiantis galima atlikti finansų rinkoje aktyviai dalyvaujančių kompanijų palyginimą. Vieno metodo esmė – saviorganizuojančių neuroninių tinklų panaudojimas. Atlikus klasterizavimo operaciją susidaro kelios kompanijų klasės. Klasėse sugrupuojamos atsiliekančios ir pirmaujančios kompanijos. Kitas kompanijų palyginimo metodas – netiesinis rangavimas. Straipsnyje nagrinėjami minėtų metodų privalumai, trūkumai, bei praktinis taikymas.