Application of Self-Organizing Map to Assessing Performance Stabilities of Top Ranking Companies

Manimannan G[1*], R. Lakshmi Priya[2]

[1]Assistant Professor, Department of Mathematics, TMG College of Arts and Sciences, Chennai.

[2]Assistant Professor, Department of Statistics, Dr. Ambedkar Government Arts College, Chennai

*Corresponding Author: Manimannan G, Assistant Professor, Department of Mathematics, TMG College of Arts and Sciences, Chennai.

Abstract: *Recently, the use of Self-Organizing Map (SOM) has provided a benchmark in analyzing and visualizing high dimensional data as a rapid growing field in data mining, especially in the analysis of financial statements. The dataset relates to 445 companies of five major industrial sectors from Indian corporate database published by Capital market, India. The dataset comprises of important financial ratios of 78 companies from cement, 115 companies from steel, 102 companies from plastic, 66 companies from leather and 84 companies from hardware and software industries. The time frame of the data pertaining to the present study is 2006-2010. Out of numerous ratios, eleven financial ratios are carefully chosen that had different notions of the objectives and significant meaning in the literature. The analysis is a two-stage procedure. First, SOM is used on the data relating to all the 445 companies to produce prototype vectors which are then, partitioned applying k-means clustering algorithm that yielded only 3 meaningful clusters with optimum Davies-Boudl in index for each of the study period. The labels of the prototype vectors are then assigned to the data on the basis of the best mapping unit. This procedure is used to grade the companies as Grade A, Grade B and Grade C based on their financial ratios. Grade A companies are those with higher profile than B and C. Similarly, Grade B companies had their profile of performances superior to Grade C but inferior to Grade A.*

The grades of the 445 companies are compared for the successive years from the year 2006 to 2010, taking 2006 as the base reference year. The movements of every company are studied for each the said period. A company could move from Grade C to Grade B, Grade B to Grade A, or Grade C to Grade A from one year to the next. On the other hand, it is also possible for a company to have lower profile in the successive years by moving in the opposite direction; that is, from Grade A to Grade B, Grade B to Grade C, or Grade A to Grade C. Those companies, which are stable, maintain their grades over the study period. It is found that certain companies had moved upwards during the period 2002-2010 (11%), 2007-2008 (12%) and 2008-2009 (54%). But there was a reverse trend in the year 2009-2010, where many companies moved in the opposite direction, which accounted for about 82%. About 69% of the companies retained their grades in the years 2002-2004 while it was 78% in 2004-2006, indicating more stability in the performances of the companies. However, only 45% of the companies retained their grades from 2006 to 2008. There was a nose dive in 2008 -20010 accounting for only 18% of the companies retaining the same grades. This pattern suggests that the reason for the companies to have reverse effects may be due to the instability at the centre, and also the sales and price mix policy of the companies.

Keywords: *Financial Ratios, Data Mining, Self-Organizing Map, k-Means Clustering, Grading.*

1. INTRODUCTION

Bearing in mind the present banking system and the globalization of the economy, it is apparent that financial performance of a firm is more essential within the financial assessment of a company. Assessment facilitates managers to place their company's performances in perspective. Hitherto, numerous statistical classification models have been constructed for financial aspects and related problems of manufacturing companies in research contexts. It is also well known that the statistical information extracted from financial ratios is being extensively used in identification or classification of manufacturing unit. The focal point of issue in financial analyses is due to its severe effect on the operation of a firm, its environment and even on the whole economy of a country. In this context, much of statistical techniques in multivariate analysis are in use to explore the objective(s) of identification

or classification of manufacturing unit into any of the prior group information, which are known in advance, to assess the financial performance of a company over the years. Usually, only two groups are considered with the operational objective; to link a set of *independent* variables to a *dependent* variable that can take discrete values such as bankruptcy or non-bankruptcy, failure or non-failure etc. In this aspect, several researchers are inspired by the pioneering work of Altman (1968) on the application of multivariate discriminant analysis (MDA), for more reliable financial distress prediction models using certain set of financial ratios, computed from the fiscal statement of any company.

On the other hand, tremendous advances in computer technology have led researchers to the development of Decision Support Systems using Artificial Intelligence. This has attracted much of researchers in the recent periods, especially in the application of Self-Organizing Maps (SOM), Neural Network models etc., to meet various objectives of problem-solving as Knowledge Discovery in Data bases (KDD). Artificial Neural Networks (ANN) is being used by many researchers to predict/classify the manufacturing units as disaster units prior to failure in different spans (Altman and Marco, 1994; Martin-del-brio and Serrano-Cinca, 1995). As an alternate approach, the Self-Organizing Map (SOM) is one of the most popularly used neural network algorithm, based on unsupervised learning developed by **Kohonen** in early1980's to assess the financial performances. It has also proven to be a valuable tool in data mining and KDD with applications in financial data analysis.

The application of SOM in finance includes evaluation of financial performances status of manufacturing units (Back, Sere and Vanharanta, 1996; Juha Vesanto, 1997; Deboeck, 1998; Raquel 2001). In the present context, the problem of grading companies has been studied without making any aprior assumptions with regard to the number of groups or any other structural patterns that reflect the performance of companies, based on certain financial ratios (Lakshmi Priya R, *et.al.* 2014). The main objectives of this paper are

- To explore the financial performance patterns of 445 top ranked companies from 2006 to 2010.
- To compare the stability of their performance over the years from 2006 to 2010.

2. METHODOLOGY

In this section, a discussion on the principles of Self-Organizing Map, the database and the choice of financial ratios are presented.

2.1. Principles of Self-Organizing Map

The Self-Organizing Map (SOM), developed by academician Kohonen, is one of the most popular neural network models. The SOM algorithm is based on unsupervised competitive learning process which is widely used in the emerging areas of research especially as a Data Mining technique, to extract previously unknown and potentially useful information from the data. Basically, SOM network have only two layers, a standard input layer and an output layer known as the Competitive (Kohonen) layer as in Figure 1. Each input neuron is connected to each and every neuron on the competitive layer which are organized as a two dimensional grid. Each company is associated with exactly one neuron whereas each neuron may have one or more companies associated with it. This grid map projects certain statistical regularities and also visibly presents the tendency of input data to represent different classes.
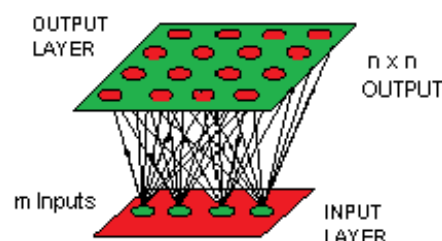

Figure1. *SOM Network*

2.2. Database

The financial data published by Capital Market, India is considered for this study; the list covers public and private sector companies in India, which are rated as the best, based on their net sales, from the year 2006 to 2010. From the top most 445 companies that are common in all the years were considered for the present analyses.

2.3. Choice of Financial Ratios

Ratios are simple financial indicators and easy to calculate from a typical set of financial statements. Various ratios are possible depending on financial parameters reported by a company. Owing to the financial parameters available for 445 companies, the choices of financial ratios are based on the notions of objective, their meaning and significance in the literature (Prasanna Chandra, 1997; Kaveri, 1980). Hence, altogether eleven ratios are carefully chosen to explore the objectives of the present study. The list of different ratios used in the present context is given in *Appendix*.

3. TRAINING THE NETWORK

In this section, a brief description of the procedure followed in the construction of self-organizing map, using the SOM toolbox for Matlab version 2.0, is discussed. The number of output neurons is chosen in such a way that it avoids excessive data overlapping. The map that has the minimum average distance between the data vectors and the trained map (minimum average quantization error) is selected. A map size of 12 by 12 neurons is found to be appropriate to exhibit the patterns present in the data. Initially, a SOM is trained separately, with the sequential procedure algorithm, where the map is linearly initialized in the subspaces spanned by the eigen vectors corresponding to the two largest eigen values of the data. The sequential algorithm involves training the map in two phases: (i) the rough learning phase with large initial neighbourhood width and learning rate of 0.5 that decreases to zero and (ii) the fine-tuning phase with small initial neighbourhood width and learning rate of 0.05 that decreases to zero. The Gaussian neighbourhood function and a map with hexagonal topology are used in the training process.

4. ALGORITHM

For the present data set, which includes 445 companies, the following algorithm is proposed to grade the companies and compare their stabilities during the study period (2006-2010).

Step1. Train the input data set and obtain a map of weight vectors with 12 x 12 neurons having hexagonal topology for neighbourhood.

Step2. Perform **k**-means analysis to partition the weight vectors obtained from **Step 1** and select the groups with optimum Davies-Bould in index.

Davies-Bould in validity index is given by $\dfrac{1}{C}\sum\limits_{k=1}^{C}\left[\max\limits_{\substack{l=1\\l\neq k}}^{C}\left\{\dfrac{S_c(Q_k)+S_c(Q_l)}{d_{ce}(Q_k,Q_l)}\right\}\right],$

Where C is the number of clusters, S_c is the within-cluster distance, d_{ce} is the between-clusters distance and Q_j is a partition with j clusters.

Step3. Assign labels to the companies in the data set using the clusters obtained in **Step 2** on the basis of Best Mapping Unit (BMU).

Step4. Obtain the mean of each cluster of companies formed in **Step 3** to assess the grades (Grade **A**, Grade **B** and Grade **C**) for the companies based upon the magnitude of financial ratios.

Step5. Repeat **Step 1** to **Step 4** for each of the study period from 2006 to 2010.

Step6. Compare the grades for each of the companies in the successive years obtained in **Step 4** with 2006 as the base year.

5. RESULTS AND DISCUSSION

The most important feature of the prototype vectors of SOM is that they preserve the topological relationship and statistical regularities of the input data set. Thus, the SOM is able to cluster the input information and their relationships on a lower dimensional map. This property of SOM paves the way to match the prototype vectors of the map of each year to the companies of that year and to assign the labels for the companies based on their best mapping unit (BMU). Corresponding to each vector in the data set, BMU vector in the prototype vectors of the map is found. The labels assigned to the prototype vectors in Step 2 are then assigned to the companies in the data set in Step 3.

Since the companies do not have predefined labels describing their financial status, the algorithm proposed in Section 4 is used to assign appropriate grades to the companies depending on their overall performances. In the construction process of SOM maps, several maps are initialized and trained separately for each of the study period. Among the prototype vector maps obtained, the best ones in respect of average quantization error are carefully chosen to explore the data and extract the hidden patterns. The prototype vectors extracted, having 144 neurons, is then subjected to MacQueens' k-means algorithm to partition into k groups. Although, the maximum allowable number of partition k, typically ranges from two to $\sqrt{N}$, where N is the number of samples in the data set, as many as 6 to 7 partitions are explored for the present study. Among the partitioned group the best one in respect to Davies-Bouldin validity index defined in Step 2 of the algorithm revealed only 3 clusters to be meaningful with optimal index that ranging from 0.8931 to 1.0185 for the study period from 2006 to 2010. Hence, the number of groups is 3, which also coincides with that of our earlier research work sited in Section 1.

Having decided upon the number of clusters as 3, an attempt is made to rate the clusters on the basis of the mean vectors. The grades are assigned based on the magnitude of mean vectors of clusters for each year taking into account certain financial ratios (Table 1). These three groups pave way to rate the members in the clusters as Grade A, Grade B and Grade C. Companies belonging to Grade A category are the ones that performs better than those of Grade B and Grade C. Similarly the companies belonging to Grade B category are superior to those of Grade C, indicating the members in the category Grade C are at a low profile in terms of the ratios considered in the present analysis.

Table1. *Cluster Sizes and Centroids*

Ratios	Cluster Centers with Grades														
	2006			2007			2008			2009			2010		
	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
OPM	0.1800	0.1472	0.0993	0.2204	0.1105	0.0845	0.1802	0.1071	0.0898	0.1454	0.1193	0.0985	0.1717	0.1201	0.0941
NPM	0.1013	0.0758	0.0488	0.1215	0.0439	0.0406	0.0789	0.0430	0.0424	0.0695	0.0498	-0.0246	0.1196	0.0517	0.0232
CF_NS	0.1161	0.0865	0.0521	0.1370	0.0553	0.0433	0.1037	0.0551	0.0515	0.079	0.0775	0.0357	0.1071	0.0597	0.055
ROTA	0.1235	0.1218	0.0887	0.1198	0.1146	0.1064	0.1235	0.1014	0.0940	0.1343	0.0837	0.0382	0.1361	0.1264	0.0694
EP	0.1696	0.1413	0.0935	0.1643	0.1325	0.1282	0.1632	0.1172	0.1050	0.1610	0.0897	0.0399	0.1873	0.1541	0.0752
RP_TA	0.0549	0.0544	0.0365	0.0634	0.0563	0.0411	0.0549	0.0329	0.0317	0.0664	0.0170	-0.0081	0.0876	0.0525	0.0073
NS_TA	1.8202	1.0285	0.5720	1.9558	1.1050	0.6410	1.8422	1.0816	0.6104	1.5303	0.7014	0.4002	1.7685	1.1125	0.6983
CF_TA	0.0893	0.0822	0.0587	0.0843	0.0828	0.0663	0.0913	0.0599	0.0597	0.1075	0.0459	0.0163	0.1122	0.0921	0.0381
RONW	0.2099	0.2024	0.1200	0.1901	0.1683	0.1295	0.1827	0.1078	0.0969	0.1787	0.0450	-0.0169	0.1969	0.1447	-0.0236
PBIT_NW	0.4252	0.3808	0.2312	0.4189	0.3359	0.3086	0.4070	0.2880	0.2655	0.3616	0.2193	0.1059	0.4932	0.2811	0.1678
LTIER	1.1704	1.0261	1.0005	1.1353	1.0628	1.0057	1.1226	1.0147	0.9845	1.1722	0.9468	0.9069	2.1149	1.0871	0.9388

No. of Comp.	55	184	206	78	212	155	65	120	260	198	205	42	55	140	250

The movements of the companies are investigated in addition to grading the companies. This involves identifying the number of companies that had a movement from one profile to another in the successive year. That is, a company could move from Grade **C** to Grade **B**, Grade **B** to Grade **A,** or Grade **C** to Grade **A** from one year to the next. This would indicate that the movements are in the positive direction and the companies do better in the successive years. On the other hand, it is also possible for a company to have lower profile in the successive years by moving in the opposite direction; that is, from Grade **A** to Grade **B**, Grade **B** to Grade **C**, or Grade **A** to Grade **C**. Those companies, which are stable, maintain their grades over the years. *Table 2* summarizes the movements of companies from one profile to another.

Table2. *Number of companies with changing profile in successive years*

Trend	Change over of Grades	Number of Companies			
		2006 – 2007	2007 – 2008	2008– 2009	2009 - 2010
Upward Trend	C → B	19	12	73	-
	B → A	6	15	30	-
	C → A	-	-	19	-
	Percentage	11%	12%	54%	-
Downward Trend	A → B	11	5	1	53
	B → C	33	17	2	100
	A → C	1	-	-	34
	Percentage	20%	10%	1%	82%
Constant	Companies	158	179	103	41
	Percentage	69%	78%	45%	18%

From *Table 2*, it is found that certain companies have moved upwards during the period 2006-2007 (11%), 2007-2008 (12%) and 2008-2009 (54%). But there was a reverse trend in the year 2009-2010, where many companies moved in the opposite direction, which accounted for about 82%. About 69% of the companies retained their grades in the years 2006-2007 while it was 78% in 2007-2008, indicating more stability in the performances of the companies. However, only 45% of the companies retained their grades from 2008 to 2009. There was a nose dive in 2009-2010 accounting for only 18% of the companies retaining the same grades. Further, from Table 1 one can find that the numbers of companies in each grade are more or less stable from 2006 through 2010 but not very stable for the remaining periods. The possible reasons that kept most of the companies in varying profile in the year 2009 and 2010 could be due to the union budget, political uncertainty in New Delhi and also the elections results.

6. CONCLUSION

The purpose of this paper is to explore the possibility of identifying meaningful groups of companies and their stability using SOM and unsupervised classification techniques. An attempt is made to analyze the 445companies covering both the public and private sector over a period of five years from 2006 to 2010, based on the financial data. The present analysis is carried out in two phases. Initially, a SOM is trained separately for each year choosing a map of size 12 x 12 neurons with different training parameters, adopting sequential algorithm. The final prototype vector with the least quantization error is obtained using SOM Toolbox version 2. In the second phase, the extracted prototype vectors containing 144 neurons are clustered using the conventional MacQueens' k- means algorithm, and the best partitions with optimal Davies-Bouldin index is chosen to explore the formation of clusters in the data set. It is found that only 3 clusters are meaningful during each of the study period. This facilitated to group the companies into 3 clusters and also to grade them as Grade A, Grade B and Grade C based on the group means of certain financial parameters. Grade A companies are those with higher profile than B and C. Similarly, Grade B companies have their profile of performances superior to Grade C but inferior to Grade A

The stability of performances of 445 companies during the period 2006 to 2010 is also assessed. It is found from the analyses that both upward and downward trends prevail with respect to their grades. About 69% of the companies retained their grades in the years 2006-2007 while it was 78% in 2007-2008, indicating more stability in the performances of the companies. However, only 45% of the companies retained their grades from 2008 to 2009. There was a nose dive during 2009-2010,

accounting for only 18% of the companies retaining the same grades. This pattern suggests that the reason for the companies to have reverse effects may be due to the political uncertainty in New Delhi, Union Budget and elections. Financial Analyst can make use of these techniques of rating, and the industries can project the performance on the basis of financial ratios that has been considered in this study. A generalization of the results is under investigation to obtain a set of 3 groups of industries for any given year.

REFERENCES

[1] Altman E I (1968), Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, The Journal of Finance, vol.23, pp. 589-609.

[2] Barbro Back *et.al* (1996), Data Mining Accounting Numbers Using Self Organizing Maps, Finnish Artificial Intelligence Society, pp. 35-47.

[3] Breaver W H (1966), Financial Ratios as Predictors of Failure, Journal of Accounting Research, pp. 179-192.

[4] R. Lakshmi Priya, *et.al.* (2014), Grading Companies on the Basis of Financial Ratios: A Data Mining Approach, *Presented at the National Conference on Applied Statistics and Mathematics,* Nagpur University, Nagpur.

[5] Doumpos M and Zopounidis (1999), A Multicriteria Discrimination Method for the Prediction of Financail Distress: The Case of Greece, Multinational Finance Journal, Vol.3, no.2, pp.71-101.

[6] JuhaVesanto (1999), SOM – Based Visualization Methods, http://www.cis.hut.f i/projecs/ide /publications/

[7] Juha Vesanto (2000), Using SOM in Data Mining. http://www.cis.hut.fi/projecs/ide/publications/.

[8] Juha Vesanto *et.al.*(2000), Clustering of the Self-Organizing Map, IEEE Transaction on Neural Networks, 11(3), pp. 586-600.

[9] Kaveri. V.S. (1980), Financial Ratios As Predictors of Borrowers Health, Sultan Chand & Sons, New Delhi

[10] Markus Siponen *et.al.* (2001), an Approach to Automated Interpretation of SOM, Information and Classification, Springer-Verlag.

[11] Pieter Adriaans and Dolf Zantinge (1996), Data Mining, Addison-Wesley Longman Limited, England.

[12] Prasanna Chandra (1997), Financial Management Theory and Practice, 4/ed, Tata McGraw-Hill Publication Company Limited, New Delhi.

APPENDIX

S.No.	Ratios	Abbreviations
1.	Profit Before Interest and Tax / Net Sales	OPM
2.	Net Profit / Net Sales	NPM
3.	Cash Flow/ Net Sales	CF/NS
4.	Net Profit /Total Assets	ROTA
5.	Profit Before Interest and Tax / Total Assets	EP
6.	Retained Profit / Total Assets	RP_TA
7.	Net Sales/Total Assets	NS_TA
8.	Cash Flow/Total Assets	CF_TA
9.	Net Profit / Net Worth	RONW
10.	Profit Before Interest and Tax/Net Worth	PBIT_NW
11.	Log((PBIT/Interest) + 5)	TIER

OPM   - Operating Profit Margin

NPM   - Net Profit Margin

ROTA - Return on Total Assets

EP   - Earning Power

RONW - Return on Net Worth

TIER   - Time Interest Earning Ratio