Expert Systems with Applications 37 (2010) 8793–8798

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Clustering Indian stock market data for portfolio management

S.R. Nanda, B. Mahanty, M.K. Tiwari *

Department of Industrial Engineering and Management, Indian Institute of Technology, Kharagpur, West Bengal, India

ARTICLE INFO

Keywords:
Portfolio management
Markowitz model
K-means clustering
Self organizing maps
Fuzzy *C*-means

ABSTRACT

In this paper a data mining approach for classification of stocks into clusters is presented. After classification, the stocks could be selected from these groups for building a portfolio. It meets the criterion of minimizing the risk by diversification of a portfolio. The clustering approach categorizes stocks on certain investment criteria. We have used stock returns at different times along with their valuation ratios from the stocks of Bombay Stock Exchange for the fiscal year 2007–2008. Results of our analysis show that *K*-means cluster analysis builds the most compact clusters as compared to SOM and Fuzzy *C*-means for stock classification data. We then select stocks from the clusters to build a portfolio, minimizing portfolio risk and compare the returns with that of the benchmark index, i.e. Sensex.

1. Introduction

One of the decision problems in the financial domain is portfolio management and asset selection. Under the extremely competitive business environment, in order to face the complex market competitions, financial institutions try their best to make an ultimate policy for portfolio selection to optimize the investor returns. A formal model for creating an efficient portfolio was developed by Markowitz (1952). In his model the return of an asset is its mean return and the risk of an asset is the standard deviation of the asset returns. Risk was quantified such that investors could analyze risk-return choices. Moreover, risk quantification enabled investors to measure risk reduction generated by diversification of investment. So diversification of investment is essential to create an efficient portfolio. The problem of selecting well diversified stocks can be tackled by clustering of stock data.

Clustering as defined by Mirkin (1996) is "a mathematical technique designed for revealing classification structures in the data collected in the real world phenomena". Clustering methods organize a data set into clusters such that data points belonging to one cluster are similar and data points belonging to different clusters are dissimilar. In this paper we demonstrate the implementation of stock data clustering using well known clustering techniques namely *K*-means, self organizing maps (SOM) and Fuzzy *C*-means. The stock market data is clustered by each of the above methods. The optimal number of clusters for the stock market data using each clustering technique is carried out. The stock data contains attributes as a series of its timely returns as well as the valuation ratios to present a clear position of their market value. These are the direct investment criteria that are being considered for stock selection. Thus the resulting clusters are a classification of high dimensional stock data into different groups in view of the difference between return series along with current market valuation of stocks. After clustering stock samples are selected from these clusters to create efficient portfolio. The process is simulated for certain iterations and average risk and return is found out. It is easy to get the portfolios with lowest risk for a given level of return, using certain optimization model which is demonstrated at the end of the paper.

In order to create efficient portfolios with Markowitz model, we use the clustering method to select stocks in the paper, called clustering-based selection in our paper. The remainder of the paper is organized as follows.

The remainder of this paper is organized as follows. Section 2 describes relevant literature review. Section 3 presents the clustering-based stock selection method. Section 4 shows problem description. Section 5 depicts experimental results. In Section 6, the conclusion is presented.

2. Literature review

In this section, portfolio management and clustering techniques are briefly reviewed.

2.1. Portfolio management

A model of creating efficient portfolio was developed by Markowitz (1952). In the Markowitz model, the return of a stock is the mean return and the risk of a stock is the standard deviation of the stock returns. The portfolio return is the weighted returns of stocks. The efficient frontier of portfolios is the set of portfolios that

* Corresponding author. Tel.: +91 3222 283746.
E-mail address: mkt09@hotmail.com (M.K. Tiwari).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.06.026

offer the greatest return for each level risk (or equivalently, portfolios with the lowest risk for a given level of return). Investors measure risk reduction by diversification of investment. A lot of work has been done on portfolio management henceforth. Topaloglou, Vladimirou, and Zenios (2008) worked on a dynamic stochastic programming model for international portfolio management, a solution that determines capital allocations to international markets, the selection of assets within each market, and appropriate currency hedging levels. Genetic algorithms have been used for portfolio optimization for index fund management by Oh, Kim, and Min (2005). Fernandez (2005) states a stochastic control model that includes ecological and economic uncertainty for jointly managing both types of natural resources. Fuzzy models (Östermark, 1996) for dynamic portfolio management have also been implemented.

From the literatures reviewed we could see that there are very few studies on clustering stock data but there have been a lot of work for portfolio optimization. The initial cluster indexing of stock data can be helpful for optimization models thus improving their efficiency. Therefore, in this study we would like to focus on the stock data clustering and develop a model for diversified stock selection.

2.2. Clustering techniques

Various clustering techniques have been used in problems from various research areas of math, multimedia, biology, finance and other application domains. There are various studies within the literature that used different clustering methods for a given classification problem and compared their results. For instance Chiu, Chen, Kuo, and He (2009) applied *K*-means for intelligent market segmentation. Many variants of the normal *K*-means algorithm have also been used in various fields. Kim and Ahn (2008) used a GA version of *K*-means clustering in building a recommender system in an online shopping market. Kuo, Wang, Hu, and Chou (2005) developed a variant of *K*-means which modifies it as locating the objects in a cluster with a probability, which is updated by the pheromone, while the rule of updating pheromone is according to total within cluster variance (TWCV). Fuzzy *C*-means was a development by Bezdek (1981). The Fuzzy *C*-means clustering algorithm is a variation of the *K*-means clustering algorithm, in which a degree of membership of clusters is incorporated for each data point. The centroids of the clusters are computed based on the degree of memberships as well as data points. Over the time Fuzzy *C*-means has found an increasing use in data clustering. Ozkan, Türkşen, and Canpolat (2008) published a paper on analyzing currency crisis using Fuzzy *C*-means. A Fuzzy system modelling with Fuzzy *C*-means (FCM) clustering to develop perception based decision matrix is employed here. Tari, Baral, and Kim (2009) proposes a variant GO Fuzzy *C*-means which is a semi supervised clustering algorithm and it utilizes the Gene Ontology annotations as prior knowledge to guide the process of grouping functionally related genes.

Other popular clustering techniques use artificial neural networks for data clustering and one of the most popular is self organizing maps (SOM). Some of the recent works include use of self organizing maps in detection and recognition of road signs (Prieto & Allen, 2009), for clustering of text documents (Isa, Kallimani, & Lee, 2009), for classification of sediment quality (Alvarez-Guerra, González-Piñuela, Andrés, Galán, & Viguri, 2008) and many more.

There are papers showing comparison of different clustering methods (Budayan, Dikmen, & Birgonul, 2009; Delibasis, Mouravliansky, Matsopoulos, Nikita, & Marsh, 1999; Mingoti & Lima, 2006) and also adapting different clustering methods for a particular problem. In case based reasoning (CBR) (Chang & Lai, 2005; Jo & Han, 1996; Kim & Ahn, 2008) the problem of cluster indexing the case base to build a hybrid CBR has adapted many clustering methods.

In this paper we consider the *K*-means, Fuzzy *C*-means and self organizing maps for clustering stock data. We will use validity indexes in each case to find the optimal number of clusters.

3. Methodology

Through our literature survey we found that the problem of efficient frontier can be solved more efficiently by clustering the stocks and then choosing to enhance the criteria of diversification. We propose clustering of high dimensional stock data by the popular clustering methods *K*-means, SOM and Fuzzy *C*-means and then selecting stocks to build an efficient portfolio.

All the clustering methods are used to cluster financial stock data from Bombay Stock Exchange that consists of returns for variable period lengths along with the valuation ratios. Through the step of clustering, the aim of least diversity within a group and most difference among groups is be reached. The optimal number of clusters for each method is to be found out using certain internal validity indexes. The framework of our problem is shown in Fig. 1. A brief explanation of Markowitz model along with the clustering techniques is given below.

3.1. Markowitz model

As stated before Markowitz's has enabled investors to measure risk reduction generated by diversification of investment. We can say the return of a portfolio is the weighted return of the underlying stocks. If σ_p is the portfolio risk and n be the number of underlying stocks then

$$\sigma_p^2 = \sum_{i=1}^{n} \sum_{i=1}^{n} w_i w_j \sigma_{ij} \tag{1}$$

where σ_{ij} is the covariance between the stock price of i and j. w_i and w_j are the weights assigned to stock i and j. If return is fixed then the problem of minimization of risk can be stated as:

$$\min \quad \sigma_p^2 = w^T S w \tag{2}$$

$$\text{subject to} \quad w^T I = 1 \tag{3}$$

$$w^T R = R_E \tag{4}$$

where w is the weight vector which is a value between 0 and 1. S is the variance covariance matrix of the stocks and R_E is the expected return and R is the mean return of each stock defined as $R_t = \log\left(\frac{S_t}{S_{t-1}}\right)$. S_t is the price of the stock at time 't'.

3.2. K-means

K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. *K*-means starts with a single cluster with its center as the mean of the data. This cluster is split to two and the means of the new clusters are trained iteratively. These clusters again split and the process continues until the specified number of clusters is obtained. If the specified number of clusters is not a power of two, then the nearest power of two above the number specified is chosen. Then the least important clusters are removed and the remaining clusters are again iteratively trained to get the final clusters. This is a non hierarchical method.

3.3. Fuzzy C-means

In Fuzzy clustering methods data points can be assigned to more than one cluster with different degree of membership. In

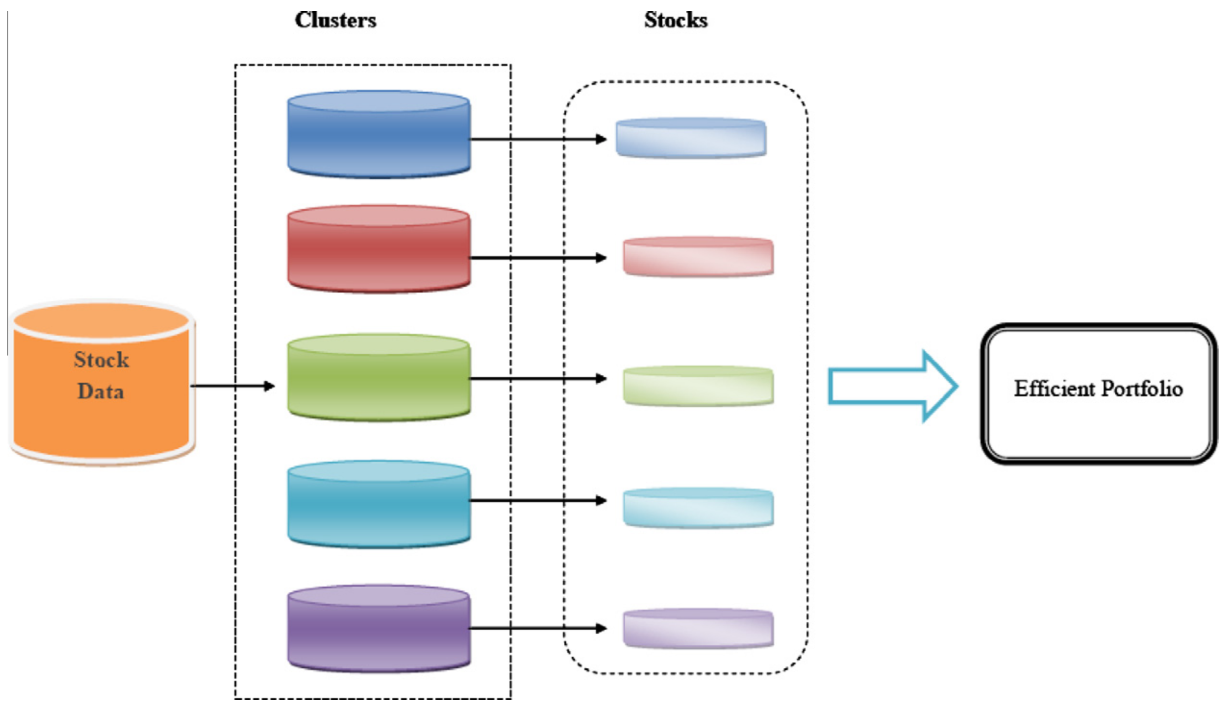

Fig. 1. Creation of efficient portfolios.

some cases it can perform well over hard clustering methods because of its ability to assign probability of a data point belonging to a particular cluster. It aims to minimize the Fuzzy version of classical "within groups' sum of squared error objective function" according to the fuzziness exponent by using Picard iterations.

The Bezdek's objective function is:

$$\min J_m(U,V) = \sum_{i=1}^{c}\sum_{k=1}^{n}(\mu_{ik})^m d^2(X_k, V) \quad (5)$$

where c is the number of the clusters, n is the number of data points and $m > 1$ is the fuzziness index. The expression of $d^2(X_k, V_i)$ shows the distance between observation X_k and the cluster centroid V_i. We have used Euclidian distance measure. The parameters affecting the results are number of clusters (c) and degree of fuzziness (m).

3.4. Self organizing maps (SOM)

SOM imitates the function of 'grouping by categories' operated by human brain and every output processing element would affect each other. It includes a set of neurons usually arranged in a two-dimensional structure, in such a way that there exist neighbourhood relations among the neurons, which dictates the topology, or structure. The neurons are well connected to each other from input to output layers but they are not connected to themselves.

It uses unsupervised learning which is known as self organization to visualize topologies and hierarchical structures of high dimensional input spaces. The algorithm of SOM is initialized by assigning the values of weight vectors of each output neuron linearly or randomly. Training process of SOM starts by representing a data point randomly in the network. The distances between these data points and the weight vectors of all neurons are computed by using distance measures such as Euclidean distance. The nearest neuron wins and is thus updated to move closer to the data point. SOM converges into a stable structure and represents information that is learned. SOM reflects an input value onto a point of plane, and the points around neighbouring area have similar functions.

In this paper, SOM is used to group similar kind of stocks based on the high dimensional stock data.

3.5. Validity

There are various indexes and functions to provide validity measures for each partition. The validation indexes also provide a clear picture on the optimal number of clusters. Some of the indexes used are discussed in brief.

Silhouette index: Better quality of a clustering is indicated by a larger Silhouette value (Chen et al., 2002).

Davies–Bouldin index: The lower the value the better the cluster structures (Kasturi, Acharya, & Ramanathan, 2003).

Calinski–Harabasz index: It evaluates the clustering solution by looking at how similar the objects are within each cluster and how dissimilar are different clusters. It is also called a pseudo F-statistic (Shu, Zeng, Chen, & Smith, 2003).

Krzanowski–Lai index[1]: Optimal clustering is indicated by maximum value (Krazanowski & Lai, 1988).

Dunn's index (DI): This index is proposed to use for the identification of "compact and well-separated clusters". Large values indicate the presence of compact and well-separated clusters (Bezdek & Pal, 2005).

Alternative Dunn index (ADI): The aim of modifying the original Dunn's index was that the calculation becomes more simple, when the dissimilarity function between two clusters.

In case of overlapped clusters the above index values are not very reliable because of repartitioning the results with the hard partition method. So the indexes below are very relevant for Fuzzy clustering (Bezdek & Pal, 2005).

Xie and Beni's index: It aims to quantify the ratio of the total variation within clusters and the separation of cluster. The optimal number of clusters should minimize the value of the index (Xie & Beni, 1991).

[1] These parameters have been scaled in our work.

Partition index: It is the ratio of the sum of compactness and separation of the clusters. It is a sum of individual cluster validity measures normalized through division by the Fuzzy cardinality of each cluster. A lower value of *Partition index* indicates a better partition (Bensaid et al., 1996).

Separation index[1]: On the contrary of *Partition index, the Separation index* uses a minimum-distance separation for partition validity (Bensaid et al., 1996).

In our work we give more importance to *Silhouette*, *Davies–Bouldin* and *Dunn's* indexes.

4. Data description

In this paper stock data for companies listed in Bombay Stock Exchange (BSE) was collected from Capitaline Databases Plus (a financial information database). Data for 106 stocks was collected for the fiscal year 2007–2008. We collected a fairly mixed data from various sectors and capital sizes. Some of the stocks which were a part of study are a part of various sectoral indices like BSE BANKEX, BSE Auto, BSE Pharma, BSE IT, BSE Midcaps, BSE Smallcaps, etc. Thus we ensured stocks being considered are covering almost all categories in the market and thus quite representative for the market during the fiscal year of 2007–2008

The returns of the stock for variable time intervals along with the validation ratios were chosen as a part of stock dimension. The dimensions used are noted in Table 1.

5. Experimental results

All the above three clustering algorithms were performed on data set. The results of the validity indexes of the clusters from all three of them are noted as below. We must mention a, that no validation index is reliable only by itself, that is why all the programmed indexes are shown, and the optimum can be only detected with the comparison of all the results.

Table 1
Stock classification factors.

Factor type	Factors	Importance
Returns	1 day	Short term
	1 week	
	30 days	
	3 months	Long term
	6 months	
	1 year	
Validation ratios	Price earning (P/E)	It shows how much investors are willing to pay per dollar of earnings
	Price to book value (P/BV)	A valuation ratio expressing the price compared to its book value
	Price/cash EPS (P/CEPS)	It shows investors valuation to cash flow generated on per share basis
	EV/EBIDTA	Measure of companies value to its earnings
	Market cap/sales	Indicator of performance with respect to market capital
EV = enterprise value		EBIDTA = earnings before interest, taxes, depreciation and amortization

5.1. K-means

The above stock data worked as input data in SOM MATLAB Toolbox for *K*-means clustering. The batch processing algorithm type was used. The number of clusters varied from 2 to 12 .The internal validity indices were calculated and tabulated in Table 2.

From Table 2 we can infer that number of clusters can be 5 or 6 for optimal clustering of the given data.

5.2. Self organizing maps (SOM)

SOM toolbox designed for MATLAB was used for SOM analysis. The variations in clustering outputs can occur due to changes in initial setting of parameters such as map size, shape of the output neurons, etc. The value of α was 0.5 and 0.05 for training and fine tune phases and a batch training algorithm was followed. The experimental results are shown in Table 3.

From above comparing individually each validity index shows a different cluster number to be optimal which was not seen as in the case of *K*-means. It proves that SOM is inconsistent as far as this stock data is concerned. However considering Davies–Bouldin and Dunn's index we can assume 7 to be an optimal cluster number.

5.3. Fuzzy C-means

The Fuzzy clustering toolbox of MATLAB was used for Fuzzy *C*-means clustering of stock data. The number of clusters was varied from 2 to 12. The fuzziness weighting exponent was equal to 2 and the maximum termination tolerance was 10^{-6}. Standard Euclidian distance norm was used to calculate cluster centers. The results of the validity indexes are tabulated as in Table 4. In case of Fuzzy clustering, Partition index, Separation index and Xie and Beni's index was also calculated.

The Partition index and the Xie and Beni's index shows that optimal number of clusters could be 11 for Fuzzy *C*-means clustering. Also Davies–Bouldin index is lower and Dunn's index is higher in case of $k = 11$. So it can be considered as optimal cluster number.

5.4. Performance evaluation using Intraclass inertia

Performance comparison of the clustering methods is done using Intraclass inertia (Michaud, 1997). Intraclass inertia $F(K)$ is a measure of how compact each cluster (class) is when the number of clusters is fixed. It can also be defined as the average squared Euclidean distance between each observation and its cluster mean. The variables are linearly scaled in (0 1) range. The results of Intraclass inertia are presented in Fig. 2.

The resulting compactness of clusters of the three clustering methods (*K*-means, SOM, Fuzzy *K*-means) shows that in our case, *K*-means method turns out to be the best. This is similar to the re-

Table 2
Validity indexes of *K*-means clustering.

K-means	No. of clusters										
Validity indexes	2	3	4	5	6	7	8	9	10	11	12
Silhoutte	0.469	0.555	0.073	0.555	0.555	0.555	0.555	0.555	0.555	0.555	0.555
Davies–Bouldin	1.303	0.992	1.429	0.799	0.912	0.985	0.972	0.97	0.824	0.973	0.946
Calinski–Harabasz	20	32.22	7.428	15.8	15.8	12.51	8.76	10.32	10.32	7.587	7.587
Krzanowski–Lai	1.399	1.105	1.553	1.212	1.212	1.253	1.32	1.289	1.289	1.349	1.349
Dunn's index	0.032	0.025	0.012	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025
Alternative Dunn's	0.528	0.019	0.245	0.019	0.112	0.112	0.112	0.153	0.03	0.019	0.086

Table 3
Validity indexes of SOM.

SOM	No. of Clusters										
Validity indexes	2	3	4	5	6	7	8	9	10	11	12
Silhoutte	0.331	-0.026	-0.141	-0.1498	-0.1498	-0.0365	-0.0365	0.0023	0.0023	-0.0006	-0.0006
Davies–Bouldin	1.4761	1.7683	1.6098	1.8038	1.8038	1.563	1.563	1.5542	1.5542	1.3156	1.3156
Calinski–Harabasz	17.483	10.932	10.489	8.9324	8.9324	12.298	12.298	6.8569	6.8569	8.2866	8.2866
Krzanowski–Lai	1.4281	1.4813	1.4461	1.4081	1.4081	1.1427	1.1427	1.3606	1.3606	1.1731	1.1731
Dunn's index	0.0137	0.0109	0.0159	0.0173	0.0173	0.0205	0.0205	0.0162	0.0162	0.0173	0.0173
Alternative Dunn's	0.8928	0.0948	0.0015	0.0029	0.0029	0.0009	0.0009	0.0006	0.0006	0.0054	0.0054

Table 4
Validity indexes of FCM.

Fuzzy *C*-means	No. of Clusters										
Validity indexes	2	3	4	5	6	7	8	9	10	11	12
Partition index	0.0187	0.0156	0.0139	0.0141	0.0144	0.0007	0.0113	0.0118	0.0112	0.001	0.0011
Separation index	0.1765	0.2498	0.2126	0.217	0.2217	0.0095	0.1795	0.1813	0.1686	0.0118	0.0128
Silhoutte	0.4621	0.184	0.1403	0.1089	0.143	0.1338	0.0543	0.136	0.1274	0.1601	0.1318
Davies–Bouldin	1.4192	1.6611	1.5119	1.5526	1.6415	1.3385	1.5714	1.7052	1.6657	1.2877	1.3658
Xie and Beni's	2.8528	1.8564	1.4633	1.259	1.146	1.0976	1.0193	0.8986	0.9303	0.8333	0.8577
Calinski–Harabasz	19.178	13.175	9.9768	8.0054	7.1705	21.118	5.7826	5.2101	4.7155	19.6284	19.0152
Krzanowski–Lai	1.5319	1.5554	1.5912	1.6274	1.6309	0.9994	1.6522	1.6683	1.6859	0.8068	0.7792
Dunn's index	0.0171	0.0121	0.0083	0.0115	0.0111	0.0183	0.0105	0.0074	0.0074	0.0193	0.0212
Alternative Dunn's	0.0075	0.0006	0.001	0.0007	0.0004	0	0	0.0001	0.0001	0	0.0001

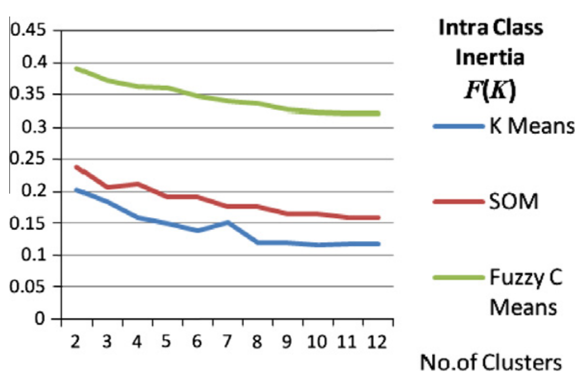

Fig. 2. The Intraclass inertia for the three clustering techniques.

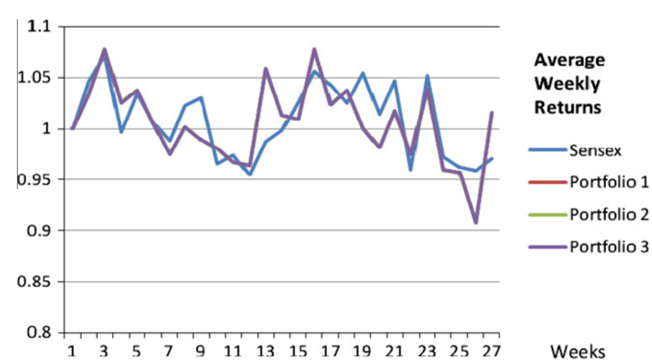

Fig. 3. Scaled plots showing returns of the portfolios with respect to Sensex.

sults obtained for customer segmentation by Shin and Sohn (2004). This is also consistent with the result drawn from *Dunn's index* which also measures the compactness of the clusters.

5.5. Creating efficient portfolio

After getting the clustering results stocks from different groups are selected and thus their corresponding weights are found as given in Section 3.1.This cluster based approach will highly reduce the possible times of creating an efficient portfolio and can help a lot of investors in practice. We took some sample stocks from the clusters that have been built by all the three clustering algorithms and build portfolios as stated in (1).The results are shown in Table 5.

A scaled plot showing the weekly returns of all the portfolios for 6 months from March 08 to September 08 is shown and compared to Sensex. The return of portfolios is higher than Sensex as shown in the plots in Fig. 3.

As seen from the plots all three portfolios give the same kind of returns as was expected due to the Markowitz model deciding the ratio of the stocks to give a fixed return. The return path followed by the portfolio was in around the same as that of the Sensex thus proving the effectiveness of Markowitz model. Thus due to diversification of portfolio the risk level was minimized and it also per-

Table 5
Weights of the stocks chosen for portfolio.

Portfolio 1 (*K*-means)		Portfolio 2 (SOM)		Portfolio 3 (FCM)	
Companies	Weights	Companies	Weights	Companies	Weights
Infosys	0.49	Eveready	0.21	Dr reddys	0.19
Hero Honda	0.32	Reliance Ind	0.17	NTPC	0.14
MRF	0.12	Infosys	0.22	Areva T & D	0.33
Lanco Infra	0.03	Pfizer	0.23	Gitanjali Gems	0.21
Financial Tech	0.04	Bharti	0.17	Tata Steel	0.13

formed fairly well as compared to the benchmark without much downward drift in returns.

6. Conclusions

This paper suggests how to integrate clustering techniques like *K*-means, SOM and Fuzzy *C*-means into portfolio management and build a hybrid system of getting efficient portfolios. It can reduce a lot of time in selection of stocks as stocks of similar categories can be easily grouped into a cluster and thus best performing stocks from those groups can be selected. Our work can find a lot of applications in software development for areas like asset management, algorithmic trading and Investor's technical information in financial markets. In our research we chose timely stock returns and valuation ratios, however from investment perspective other dimensions or factors that influence the performance of a stock can also be considered and thus that would help to refine the classification. Cases where certain temporary macroeconomic factors affect market performances for a short period of time could also be considered in our approach. We considered data for the fiscal year 2007 and since, it was the beginning world recession with markets worldwide acting bearish, and most of the stocks considered in the data have negative returns. But from clustering perspective we have compared the portfolio performances with the benchmark index, i.e. BSE Sensex. We summarize relevance of our work as:

1. In this study, various stocks from BSE were analyzed for their timely returns and market valuations. The data collected was actual rather than simulated data and the results can be considered more practical.
2. After analysis of the clustering methods, for our stock data *K*-means formed well compact clusters as compared to Fuzzy *C*-means and SOM neural network.
3. A clustering approach to portfolio management and selection of stocks to reach the efficient frontier was demonstrated.

References

Alvarez-Guerra, M., González-Piñuela, C., Andrés, A., Galán, B., & Viguri, J. R. (2008). Assessment of self-organizing map artificial neural networks for the classification of sediment quality. *Environment International, 34*, 782–790.

Bensaid, A. M., Hall, L. O., Bezdek, J. C., Clarke, L. P., Silbiger, M. L., Arrington, J. A., et al. (1996). Validity-guided (re)clustering with applications to image segmentation. *IEEE Transactions on Fuzzy Systems, 4*, 112–123.

Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press.

Bezdek, J. C., & Pal, N. (2005). Cluster validation with generalized Dunn's indices. In *Proceedings of the 2nd New Zealand two-stream international conference on artificial neural networks and expert systems* (p. 190).

Budayan, C., Dikmen, I., & Birgonul, M. T. (2009). Comparing the performance of traditional cluster analysis, self-organizing maps and fuzzy C-means method for strategic grouping. *Expert Systems with Applications, 36*(9), 11772–11781.

Chang, P.-C., & Lai, C.-Y. (2005). A hybrid system combining self-organizing maps with case-based reasoning in wholesaler's new-release book forecasting. *Expert Systems with Applications, 29*, 183–192.

Chen, G., Jaradat, S. A., Banerjee, N., Tanaka, T. S., Ko, M. S. H., & Zhang, M. Q. (2002). Evaluation and comparison of clustering algorithms in analyzing ES cell gene expression data. *Statistica Sinica, 12*, 241–262.

Chiu, C.-Y., Chen, Y.-F., Kuo, I. T., & He, C. K. (2009). An intelligent market segmentation system using *k*-means and particle swarm optimization. *Expert Systems with Applications, 36*, 4558–4565.

Delibasis, K. K., Mouravliansky, N., Matsopoulos, G. K., Nikita, K. S., & Marsh, A. (1999). MR functional cardiac imaging: Segmentation, measurement and WWW based visualisation of 4D data. *Future Generation Computer Systems, 15*(2), 185–193.

Fernandez, L. (2005). A diversified portfolio: Joint management of non-renewable and renewable resources offshore. *Resource and Energy Economics, 27*, 65–82.

Isa, D., Kallimani, V. P., & Lee, L. H. (2009). Using the self organizing map for clustering of text documents. *Expert Systems with Applications, 36*, 9584–9591.

Jo, H., & Han, I. (1996). Integration of case-based forecasting, neural network, and discriminant analysis for bankruptcy prediction. *Expert Systems with Application, 11*(4), 415–422.

Kasturi, J., Acharya, R., & Ramanathan, M. (2003). An information theoretic approach for analyzing temporal patterns of gene expression. *Bioinformatics, 19*, 449–458.

Kim, K.-j., & Ahn, H. (2008). A recommender system using GA *K*-means clustering in an online shopping market. *Expert Systems with Applications, 34*, 1200–1209.

Krazanowski, W. J., & Lai, Y. T. (1988). A criterion for determining the number of groups in a data set using sum of squares clustering. *Biometrics, 44*, 23–34.

Kuo, R. J., Wang, H. S., Hu, T.-L., & Chou, S. H. (2005). Application of ant *K*-means on clustering analysis. *Computers and Mathematics with Applications, 50*, 1709–1724.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance, 7*, 77–91.

Michaud, P. (1997). Clustering techniques. *Future Generation Computer System, 13*(2), 135–147.

Mingoti, S. A., & Lima, J. O. (2006). Comparing SOM neural network with fuzzy *c*-means, *K*-means and traditional hierarchical clustering algorithms. *European Journal of Operational Research, 174*, 1742–1759.

Mirkin, B. G. (1996). *Mathematical classification and clustering*. Dordrecht, The Netherlands: Kluwer Academic Publishing.

Oh, K. J., Kim, T. Y., & Min, S. (2005). Using genetic algorithm to support portfolio optimization for index fund management. *Expert Systems with Applications, 28*, 371–379.

Östermark, R. (1996). A fuzzy control model (FCM) for dynamic portfolio management. *Fuzzy Sets and Systems, 78*, 243–254.

Ozkan, I., Türkşen, I. B., & Canpolat, N. (2008). A currency crisis and its perception with fuzzy *C*-means. *Information Sciences, 178*, 1923–1934.

Prieto, M. S., & Allen, A. R. (2009). Using self-organising maps in the detection and recognition of road signs. *Image and Vision Computing, 27*, 673–683.

Shin, H. W., & Sohn, S. Y. (2004). Segmentation of stock trading customers according to potential value. *Expert Systems with Applications, 27*, 27–33.

Shu, G., Zeng, B., Chen, Y. P., & Smith, O. H. (2003). Performance assessment of kernel density clustering for gene expression profile data. *Comparative and Functional Genomics, 4*(3), 287–299.

Tari, L., Baral, C., & Kim, S. (2009). Fuzzy *c*-means clustering with prior biological knowledge. *Journal of Biomedical Informatics, 42*, 74–81.

Topaloglou, N., Vladimirou, H., & Zenios, S. A. (2008). A dynamic stochastic programming model for international portfolio management. *European Journal of Operational Research, 185*, 1501–1524.

Xie, X. L., & Beni, G. A. (1991). Validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 3*(8), 841–846.