A Portfolio Optimization Algorithm Using Fuzzy Granularity Based Clustering

S. M. Aqil Burney
Institute of Business Management
Korangi Creek, Karachi, Karachi City, Sindh 75190, Pakistan
Phone: +92 21 111 002 004
aqil.burney@iobm.edu.pk

Tahseen Jilani
University of Nottingham
Nottingham NG7 2RD, UK
Phone: +44 115 951 5151
University of Karachi
Main University Rd, Karachi, Karachi City, Sindh 75270, Pakistan
Phone: +92 21 99261300
tahseen.jilani@notingham.ac.uk

Humera Tariq
University of Karachi
Main University Rd, Karachi, Karachi City, Sindh 75270, Pakistan
Phone: +92 21 99261300
humera@uok.edu.pk

Zeeshan Asim
Virtual University, Pakistan
M. A. Jinnah Campus, Defence Road, Off Raiwand Rd, Lda Avenue Phase 1 Lda Avenue, Lahore, Punjab, Pakistan
Phone: +92 42 111 880 880
ms080400010@vu.edu.pk

Usman Amjad
University of Karachi
Main University Rd, Karachi, Karachi City, Sindh 75270, Pakistan
Phone: +92 21 99261300
usmanamjad87@gmail.com

Syed Shah Mohammad
University of Karachi
Main University Rd, Karachi, Karachi City, Sindh 75270, Pakistan
Phone: +92 21 99261300
syed@vu.edu.pk

Abstract

Clustering algorithms are applied to numerous problems in multiple domains including historic data analysis, financial markets analysis for portfolio optimization and image processing. Recent years have witnessed a surge in use of nature inspired computing (NIC) techniques for data clustering to solve various real world optimization problems. Granular Computing (GC) is an emerging technique to handle pieces of information, known as information granules. In this paper, an ensemble of fuzzy clustering using Particle Swarm Optimization and Granular computing for stock market portfolio optimization. The model is then tested on stocks listed in Hong Kong Stock Exchange. Experimental results suggested that clusters formed through Fuzzy Particle Swarm Optimization (FPSO) with Granular computing are well suited and efficient for portfolio optimization. For comparison, we have used a benchmark index of Hong Kong Stock Exchange called as Hang Sang Composite Index (HSCI). Results proved that results of proposed approach are better in comparison to benchmark results of HSCI.

Keywords: Hybrid Approach for Portfolio Selection; Fuzzy C-mean Clustering (FCM); Fuzzy Particle Swarm Optimisation (FPSO); Granular Computing; Hong Kong Composite Index.

1. Introduction

Data Clustering is the mathematical method designed to identify relevant data within a collection of data (Nerurkar et. al., 2018). It can be described as a methodology for assignment of data into groups in a manner that the data points in same group or cluster are analogous to each other and unrelated with objects of other clusters or groups (Hammouda and Karray, 2000). It is being used efficiently in various domains for identification of natural groups present in large datasets. Data Clustering can be used by businesses for identification of potential customers of a product by analyzing or collecting buying patterns of customers, so as to design marketing strategies based on those behaviors (Ravi, Pradeepkumar and Deb, 2017). Finding out clusters in a large dataset is challenging task and usually require some data mining tool. Clustering tools usually assign data elements to a clusters based on their similarities to the group.

Clustering remained an area of interest for researchers in last few decades, thus various clustering techniques were developed. Clustering techniques can be generally divided into two types (Suganya and Shanthi, 2011). Based on classical set theory there is a type of clustering called hard clustering algorithms in which data items can only be assigned to only one group at a point of time. A widely used hard or crisp clustering algorithm is k-means. But for real datasets where there are no definite boundaries, this technique is not useful. (Izakian and Abraham, 2009).

Soon after the introduction of the fuzzy theory, the researchers applied fuzzy set theory on clustering algorithms (Izakian and Abraham, 2009). There is no sharp boundary in real world data, so Fuzzy clustering algorithms remained fruitful in those applications. It can handle real world uncertainties efficiently by assigning membership degree to items. Membership degree in such clusters relies on the proximity of values to the cluster centers. Widely used and famous fuzzy clustering algorithm is Fuzzy C-Means (FCM) introduced by Bezdek in 1974 and is being applied at large (Bezdek, 1984).

Swarm intelligence (SI) is an area of computational intelligence which comprise of algorithms getting inspiration from population based natural phenomenon working on the basis of decentralized control and self-organization (Shandilya et. al., 2017). It can be said that SI is "collective behavior of decentralized and self-organized systems" (Zhang et. al., 2013). On the other hand Granular Computing (GC) is a computation theory for efficiently using granules such as clusters, groups and subsets to build a computational model for complicated applications that contains huge amounts of data and information. A granule can be described as one of the various small data points or particles combining to form a larger unit.

In this paper we have used Fuzzy Particle Swarm Optimization (FPSO) using the concept of granular computing to divide the information granules into different clusters to build a portfolio that can optimize the weekly investor's returns. The experimental results using the Hong Kong Stock Exchange data indicate that our proposed method provides better returns than the benchmark index for the Hong Kong Stock exchange.

2. Literature Review

2.1. Fuzzy Data Clustering

Fuzzy logic concepts are based on degree of membership so imprecision concepts are dealt with fuzzy logic in better way. Fuzzy logic can be used in data clustering so as to deal with partial membership of data points. Fuzzy logic based data clustering algorithms assign data object partly to more than one cluster. FCM proposed by Bezdek (Bezdek, 1984), divides the collection of n data objects denoted as o = {o1,o2 ,...,on} in R dimensional space into c fuzzy clusters, where (1 < c < n) with centroids or cluster centers Z = {z1, z2,..., zc}. Fuzzy clustering can be represented by using fuzzy matrix μ with dimensions n x c. Here n is count of data objects whereas c is count of data clusters. Data item present at ith row and jth column is represented by μij. Degree of membership of ith and jth object is represented by μ. Degree of membership μ has following properties:

$$\mu_{ij} \in [0,1] \qquad \forall i = 1, 2, \cdots, n; \qquad \forall j = 1, 2, \cdots, c \quad (1)$$

$$\sum_{j=1}^{c} \mu_{ij} = 1 \qquad \forall i = 1, 2, \cdots, n \tag{2}$$

$$0 < \sum_{i=1}^{n} \mu_{ij} < n \qquad \forall j = 1, 2, \cdots, c \tag{3}$$

Fuzzy C-Means has the objective function to minimize the following equation:

$$J_m = \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij}^{m} d_{ij} \tag{4}$$

s.t,

$$d_{ij} = |o_i - z_j| \tag{5}$$

where m; (m > 1) is a scalar constant value called as "weighting exponent", which manages the fuzziness of clusters whereas dij is Euclidian distance between object oi and the cluster center zj. Where zj indicates cluster center of jth cluster and it is obtained using equation (6)

$$Z_j = \frac{\sum_{i=1}^{n} \mu_{ij}^{m} o_i}{\sum_{i=1}^{n} \mu_{ij}^{m}} \tag{6}$$

FCM is an iterative algorithm and described in below steps:

i) Select the weighting component m where (m>1) and initiate μij, the membership function values where , i = 1,2,..., n; j = 1,2,..., c .

ii) Using above mentioned Eq. 6 find out the cluster centers zj , where j = 1,2,..., c.

iii) Calculate the Euclidian distance dij, where i = 1,2,..., n; j = 1,2,..., c

iv) Using below mentioned Eq. 7 update μij, the membership function, where i = 1,2,..., n; j = 1,2,..., c.

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{(m-1)}}} \tag{7}$$

v) If not converged, go to step 2.

There are many conditions that can be used to stop the execution of this loop. One of them is to stop iterations of the algorithm when the change in the cluster center values becomes negligible or the objective function as specified in the equation (4), cannot be minimized more. One problem of FCM algorithm is that it is very much dependent on initial values and likely to fall in local optima problem.

2.2. Fuzzy Particle Swarm Optimization

(Peng et al., 2004) suggested a variant of PSO based on fuzzy logic, for Travelling Salesman Problem (TSP), known as "Fuzzy particle swarm optimization (FPSO)". In this algorithm the velocity and position of individuals are re-defined to characterize the fuzzy relation within variables. In Fuzzy PSO algorithm, X represents fuzzy relationship between collection of data objects, o = {o1,o2,..., on} , and the collection of cluster centers, Z = {z1, z2,..., zc} . Fuzzy relationship X is represented as:

$$X = \begin{pmatrix} \mu_{11} & \cdots & \mu_{1c} \\ \vdots & \ddots & \vdots \\ \mu_{n1} & \cdots & \mu_{nc} \end{pmatrix} \quad (8)$$

In the above mentioned matrix X, μij denotes the membership value of object i to cluster j with constraints specified in the eq. 9 and 10

$$\mu_{ij} \in [0,1] \qquad \forall i = 1, 2, \cdots, n; \qquad \forall j = 1, 2, \cdots, c \quad (9)$$

$$\sum_{j=1}^{c} \mu_{ij} = 1 \qquad \forall i = 1, 2, \cdots, n \quad (10)$$

The position matrix specified in the above mentioned equation of each individual is similar to fuzzy matrix μ specified in Fuzzy C-Means algorithm. Velocity of each individual is specified by a matrix of dimension [n, c] where n denotes number of rows and c is the number of columns. Elements of the matrix are within the range of [-1, 1]. Eq. 11 and 12 are used for changing the velocities and positions of every particle on the basis of matrix operations.

$$V(t+1) = w \otimes V(t) \oplus (c_1 r_1) \otimes (\text{pbest(t)} \Theta \text{X(t)}) \oplus (\text{c}_2 \text{r}_2) \otimes (\text{gbest(t)} \Theta \text{X(t)}) \quad (11)$$

$$X(t+1) = X(t) \oplus V(t+1) \quad (12)$$

Here $\oplus$ denotes the matrix addition and $\otimes$ represents the matrix multiplication. It is important to note here that constraints stated in eq. 9 and 10 may be violated after update of position matrix. Thus normalizing position matrix is necessary here. For normalization purpose, all the negative values in matrix are made zero. And if all matrix elements turn out to be zero then the matrix is evaluated again using random numbers within range of [0, 1] and then matrix is transformed without violating the conditions.

$$Xnormal = \begin{pmatrix} \mu_{11} / \sum_{j=1}^{c} \mu_{1j} & \cdots & \mu_{1c} / \sum_{j=1}^{c} \mu_{1j} \\ \vdots & \ddots & \vdots \\ \mu_{n1} / \sum_{j=1}^{c} \mu_{nj} & \cdots & \mu_{nc} / \sum_{j=1}^{c} \mu_{nj} \end{pmatrix} \quad (13)$$

Similar to other nature inspired algorithms Fuzzy PSO algorithm uses a fitness function for assessing the general solution. Following equation will be used for evaluation of the solutions.

$$f(X) = \frac{K}{J_m} \quad (14)$$

In equation 14, K is a constant while Jm is objective function for Fuzzy C-Means algorithm given in eq. 15.

$$Jm = \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij}^{m} d_{ij} \quad (15)$$

As the value of Jm is smaller, clustering results will be better fitness value f(X) being higher. Fuzzy PSO algorithm for fuzzy clustering is described as under:

1. Instantiate the following parameters: P the population size, w, c1, c2 and maximum number of iterations.
2. Initialize a swarm with P number of individuals. Here X, gbest, pbest and V are matrices of size n, c.
3. Instantiate X, V and pbest values for every individual and gbest for the whole population.
4. Determine the centers of cluster for every individual using eq. 16.

$$Z_j = \frac{\sum_{i=1}^{n} \mu_{ij}^{m} o_i}{\sum_{i=1}^{n} \mu_{ij}^{m}} \quad (16)$$

5. Evaluate the objective value of every individual using eq. 15.
6. Evaluate the value of pbest for every individual.
7. Evaluate the value of gbest for the whole population.
8. Change velocity matrix for every individual by using eq. 11 and 12.
9. Change position matrix for every individual by using eq. 13.
10. Go back to step 4 until stopping criteria is met.

Stopping criteria is either the predefined maximum number of iterations or no progress in global best fitness for a specified number of generations.

In this paper we have used Fuzzy PSO clustering algorithm on the granules created from the original dataset. To divide the dataset into granules on the basis of Market Capital Value of the company, the companies with similar market value are placed into same group. For the clustering of these granules Fuzzy PSO algorithm is used which gives the benefit of lower computational time and offer better results than Fuzzy C-Means clustering algorithm (Mehdizadeh, 2009).

3. Methodology Used

From the literature review it is revealed that lot of work was done for data clustering and portfolio management but not much work is done on clustering stock data for the portfolio optimization using Fuzzy PSO. The use of data clustering for stock data helps in segmenting different stocks in a way that all stocks having similar characteristics are grouped together (Cheng, Chen and Jian, 2015). A method for creating efficient portfolios with Markowitz model by using the clustering method to select stocks, called clustering based selection was designed by (Nanda, Mahanty and Tiwari, 2010). To classify stocks into clusters they used Fuzzy C-Mean data clustering algorithm. After classification of stocks, some stocks were selected from clusters for building an optimized portfolio to minimize the risk by diversifying the portfolio. According to them, the problem of efficient frontier can be solved more efficiently by clustering the stocks. Although fuzzy c-means (FCM) algorithm is considered one of the most popular and widely used fuzzy clustering techniques because of its efficiency, straightforwardness, and convenience of implementation. But problem is that fuzzy c-means is very sensitive to initialization and it can easily be trapped in local optima. On the other hand Particle swarm optimization (PSO) is a stochastic global optimization tool which is used for various optimization problems. (Li, Liu and Xu, 2007) proposed a fuzzy PSO based data clustering algorithm to overcome the shortcomings of FCM. Their suggested method uses the power of global search in PSO algorithm to overcome the shortcomings of Fuzzy C-Means.

In our methodology the focus is on how stocks can be divided into granules and how Fuzzy PSO based data clustering algorithm can be applied on these granules to further divide data into small clusters and how to design a diversified portfolio using stocks from different cluster to maximize portfolio returns. So Fuzzy PSO algorithm is applied to create clusters for each granule. The dataset used for the experiment contains the information about financial ratios of companies listed in Hong Kong Stock Exchange. This dataset is divided into six different sub groups known as

information granule. Information granules are collections of entities that are arranged together due to their similarity, functional or physical adjacency, coherence etc. A granulation criterion deals with the question of why two objects are put into the same granule. We divided the dataset into 6 different partitions or granules based on their market capitalization value.

This is followed by calculating the optimal number of clusters in each group. Then Fuzzy PSO data clustering algorithm is applied on each granule to divide it into optimal number of clusters as calculated in the previous step. After that 1 to 3 stocks are selected from every cluster according to the important fields as indicated in the Principal Component Analysis. Then average weekly return of each stock selected is calculated in the last step on the basis of their market value during January- 2012 to June-2012. The stocks having good positive average weekly returns are selected for the portfolio creation.

Finally, Variance – Covariance matrix for the selected stocks is calculated in the last step. MATLAB is used for the development of efficient portfolio against the efficient frontier. For this purpose, we have used the MATLAB financial tool box command frontcon. This command returns optimized portfolios as per the provided input parameters. We have taken 3 portfolios from the given set of portfolios and calculated the actual weekly portfolio returns for each portfolio on the basis of market value during July 2012 to December 2012. Then we calculated the Hang Seng Composite Index weekly performance from July 2012 to December 2012 from the Bloomberg website. Hang Seng Composite Index is benchmark index for Hong Kong Stock exchange. HSCI is a comprehensive benchmark index and covers about 95% of total listed companies on main board of stock exchange of Hong Kong ("SEHK"). HSCI is used as a basis for performance benchmarks.

In next step we compared these portfolio results against the HSCI and the comparison showed that these portfolio returns are better than the HSCI. Flow chart of proposed model is shown in Figure 1.

4. Data Description

Dataset of the Hong Kong Stock Market companies' data for the financial year 2011 was taken from the New York University Dataset page. In data preprocessing step, it was checked for missing values by removing the instances with missing data from dataset. There are 774 companies' data present in the dataset after the removal of missing values. This dataset contains companies' data from 77 different industry groups and represents almost all industry groups of the Hong Kong Stock Exchange. There are 42 fields for each company which includes many different types of financial ratios to represent the financial position of that company at the end year 2011. Some of the financial ratios include Market Capital (in US$), Total Debt (in US$), Firm Value (in US$), Cash, Enterprise Value (in US$), Cash Firm Value, Liquidity Ratio, Book Debt to Capital Ratio, Market Debt to Capital Ratio, Book Debt to Equity Ratio, Market Debt to Equity Ratio, Beta, Correlation with Market, PBV, PS, Return on Equity and Return on Capital etc.

In this dataset variety of data values are used. Some fields contain very large values like Market Value, Enterprise Value, Market Capitalization and some very small fields like Beta, Debit to Equity Ratio. Data is first pre-processed to deal with large values. Initially, the data was transformed to z- scores to get similar variability of the values.

Another problem is that there are 42 fields in the dataset which are difficult to handle for calculations while performing data clustering; therefore we used Principal Component Analysis (PCA) for this purpose. PCA uses a mathematical procedure to transforms a number of correlated variables into a smaller number of un-correlated variables called principal components. This process is also known as Dimension Reduction. The transformation in PCA is done in such a way that the 1st principal component has the largest possible variance, and each succeeding component in turn has the highest variance possible under the constraint that it will be uncorrelated with the preceding components. Before performing PCA the data must be standardized to remove the influence of different measurement scales and to give approximately equal weightage to all the values. We have used SPSS tool for this purpose.

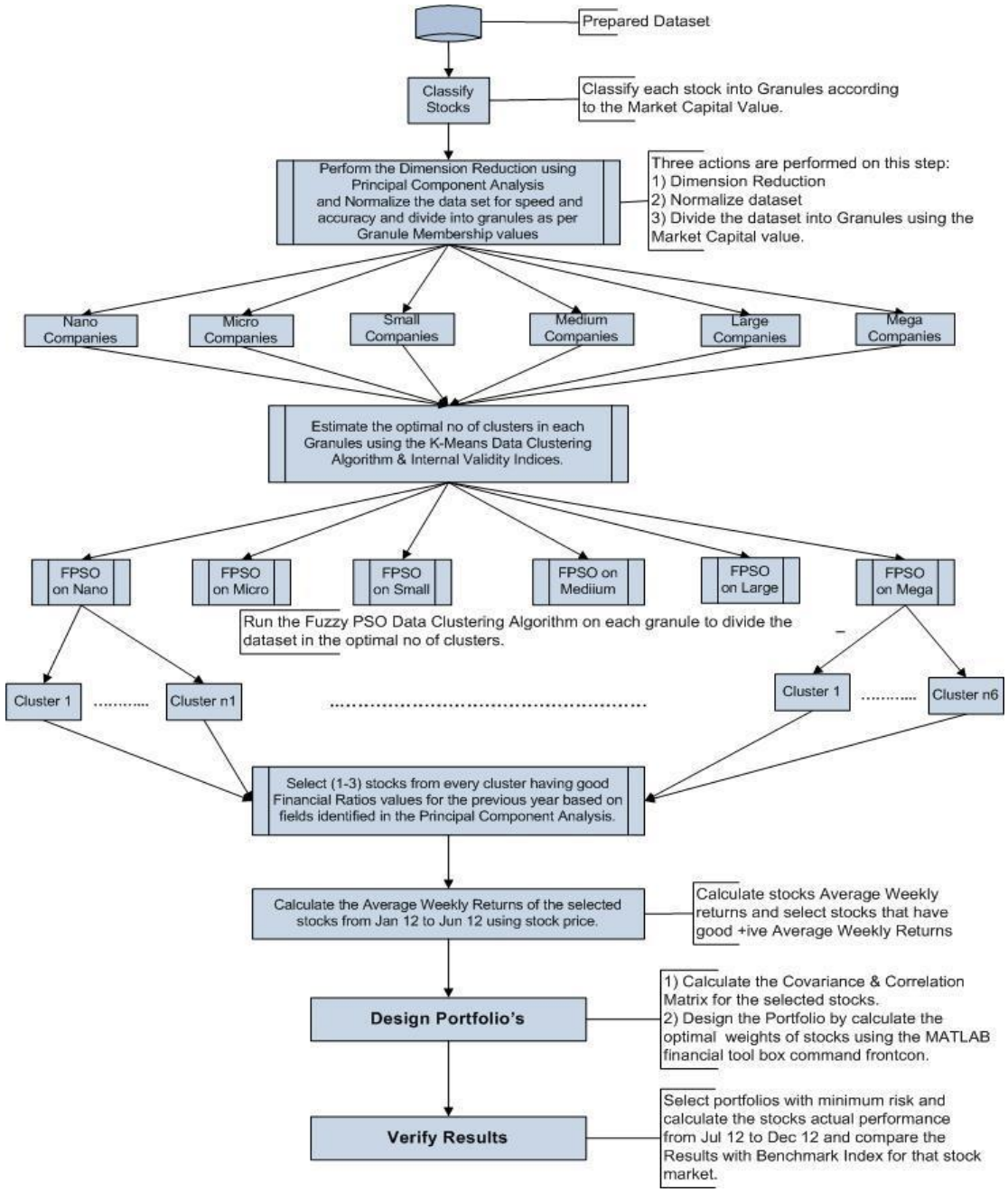

Figure 1. Flow chart of proposed model for portfolio optimization

After performing PCA on our data, 12 principle components cover 94.7% variation of the full dataset with 42 variables. The identified fields sorted by respective Eigen values are,

1. Firm Value (in US$),
2. Book Debt to Capital Ratio,
3. Price to Sale Ratio (PS),
4. Free Cash Flow to Firm (FCFF),
5. Beta (A measure of the volatility of a portfolio in comparison to the market as a whole),
6. Liquidity Ratio,
7. Correlation with Market,
8. Return on Capital,
9. Net Profit Margin,
10. Net Debt Issued,
11. Cash Firm Value,
12. EV Invested Capital.

First four fields represent around 62% of the dataset variation as per the PCA so we will use these four fields for selection of stocks from different clusters.

5. Experimental Results

5.1. Granules Formation

The dataset was divided into six different sub groups also known as information granules. Information granules are a group of objects that are organized together based on their similarity, coherence or physical adjacency. A granulation criterion describes the rules for dividing data objects into different granules. The categorization of companies into different partitions is made based on market capitalization. The companies were divided into groups namely Mega, Large, Mid, Small, Micro and Nano. There is no formal definition of the exact cutoff values. Therefore, following market capitalization values as granule criterion are used:

1. Mega Companies: Over $10000 Million	2. Large Companies: $5000 □ $10000 Million
3. Mid Companies: $1000 □ $5000 Million	4. Small Companies: $250 □ $1000 Million
5. Micro Companies: Below $250 Million	6. Nano Companies: Below $50 Million

After the granules formation the number of companies in each granule is shown in the table 1:

Table 1. Granules Frequency Table

Granule Name	Frequency	Percent	Cumulative Percent
Nano	182	23.5	23.5
Micro	312	40.3	63.8
Small	151	19.5	83.3
Medium	95	12.3	95.6
Large	17	2.2	97.8
Mega	17	2.2	100
Total	774	100	

5.2. Granules Formation Optimal Number of Clusters Estimation in Each Granule

Number of records of companies' data in each granule is different, with wide range of values so there would be different number of clusters in each granule. To divide each granule into clusters we have used K-means data clustering algorithm and to identify optimal number of clusters Internal Validity Indices. For the cluster estimation we have used available tool for this purpose designed by Mr. Kaijun Wang MATLAB code:
(http://www.mathworks.com/matlabcentral/fileexchange/13916). This tool is designed in MATLAB release 7.2 (R2006a). The tool is suitable for the performance comparison of different indices on the estimation of the number of clusters, algorithm design for applications by using or improving part codes, etc. This tool also provides internal validity indices which we have used to estimate the optimal number of clusters between 2 and 14. Following internal validity indices are used in this method for the cluster estimation.

- Silhouette (SIL)
- Davies-Bouldin (DB)
- Calinski-Harabasz (CH)
- Krzanowski-Lai (KL)

The results of optimal number of clusters estimation for each granule using above mentioned tool are described in table 2:

Table 2. Optimal number of Clusters or each Granule Table

Granule	Silhouette (Sil)	Davies-Bouldin (DB)	Calinski-Harabasz (CH)	Krzanowski-Lai (KL)	Optimal no. of Cluster
Nano	6	6	6	8	6
Micro	5	5	5	7	5
Small	2	2	2	2	2
Medium	3	3	3	3	3
Large	2	2	2	2	2
Mega	2	2	2	-	2

5.3. Clustering Granules using Fuzzy PSO

After the optimal numbers of cluster estimation for every granule we have performed data clustering using Fuzzy Particle Swarm Optimization algorithm. The code for this data clustering technique is written in MATLAB 2010 using the Fuzzy PSO algorithm. The code is run for 300 times for each granule and the best clustering result are used for further processing. To evaluate the best clustering result objective function value and Internal validity indices are used. Following internal validity indices are used for the evaluation of clustering results.

- Partition Coefficient (PC)
- Classification Entropy (CE)
- Partition Index (SC)
- Separation Index (S)
- Xie and Beni's index (XB)

5.4. Selection of Companies from Each Cluster

After applying the Fuzzy PSO (FPSO) data clustering algorithm on all granules we have developed clusters in each granules. In our next step we selected 1-3 companies from each cluster based on the performance of financial ratios during the year 2011. The ratios used for the selection of companies to build a portfolio are those fields which are identified in the principal component analysis. The fields & ratios used for selection of companies are Firm Value in US$, Book Debt to Capital Ratio, PS (Price Sale), FCFF (Free Cash Flow to Firm), Beta, Liquidity Ratio, Correlation with Market, Return on Capital , Net Profit Margin, Net Debt Issued/Repaid, Cash Firm Value and EV Invested Capital. As a result of this process 33 companies were selected for further processing. The detail of number of companies selected from each granule is given in Table 3:

Table 3. No of companies selected from each granule

Granule Name	No of Companies Selected	Total no of Companies	No of Clusters
Nano	4	182	6
Micro	6	312	5
Small	5	151	2
Medium	8	95	3
Large	5	17	2
Mega	5	17	2
Total	33	774	20

The number of companies selected for hybrid optimal portfolio from each cluster of every granule is given in Table 4:

Table 4. No of companies selected from each cluster

Cluster No. **Granule Name**	**1**	**2**	**3**	**4**	**5**	**6**
Nano	1	1	0	0	1	1
Micro	1	1	1	1	2	--
Small	3	2	--	--	--	--
Medium	1	4	3	--	--	--
Large	3	2	--	--	--	--
Mega	3	2	--	--	--	--

5.5. Selection of Companies for Portfolio Management

For the selection of companies for the portfolio management we calculated the average weekly earnings of above mentioned companies for the period of January 2012 to June 2012 (26 weeks). For the calculation of the average weekly earnings we have downloaded the company historical share prices from yahoo finance. For further processing we selected only those companies that have weekly earnings of 0.08%. There were 15 companies out of 33 that have weekly earning greater than or equal to 0.08%. The names of those companies are as under:

Table 5. List of companies selected for portfolio creation

Sr #	Company Name	Industry Group	Granule	Cluster
1	VS International Group Ltd. (SEHK:1002)	Machinery	Nano	2
2	Chun Wo Development Holdings Ltd. (SEHK:711)	Engineering	Nano	5
3	Huafeng Group Holdings Limited (SEHK:364)	Apparel	Nano	6
4	National Electronics Holdings Ltd. (SEHK:213)	Apparel	Micro	2
5	Hon Kwok Land Investment Co. Ltd. (SEHK:160)	Real Estate (Development)	Micro	3
6	Convenience Retail Asia Ltd. (SEHK:831)	Retail (Grocery and Food)	Small	2
7	Shimao Property Holdings Ltd. (SEHK:813)	Real Estate (Development)	Medium	2
8	Kerry Properties Ltd. (SEHK:683)	Real Estate	Medium	2
9	Cafe de Coral Holdings Ltd. (SEHK:341)	Restaurant	Medium	2
10	Franshion Properties (China) Ltd. (SEHK:817)	Real Estate	Medium	3
11	Haier Electronics Group Co., Ltd. (SEHK:1169)	Furn./Home Furnishings	Medium	3
12	Galaxy Entertainment Group Limited (SEHK:27)	Hotel/Gaming	Large	1
13	China Resources Land Ltd. (SEHK:1109)	Real Estate (Development)	Large	2
14	China Overseas Land & Investment Ltd. (SEHK:688)	Real Estate (Development)	Mega	1
15	BOC Hong Kong Holdings Ltd. (SEHK:2388)	Bank	Mega	2

5.6. Design Portfolios using the FrontCon

To design different portfolios, we have used frontcon function from MATLAB's financial toolbox. This function returns the mean-variance efficient frontier with user specified covariance and returns. FrontCon is a MATLAB 2010 function that helps us to design portfolios of asset investment weights which minimize the risk for given values of the expected return. The portfolio risk is minimized subject to constraints on the asset weights or on groups of asset weights. To use the frontcon we need a Variance-Co Variance matrix of given companies, expected return and number of portfolios required to be designed. We have calculated the Var-Co Variance matrix using the average weekly earnings calculated before, for the expected returns we have used the average weekly earnings of these companies between January 2012 and June 2012. To calculate the average weekly earnings we downloaded historical share price of these companies from the yahoo finance. The frontcon function returns the portfolios with specified number of shares of each company. The frontcon also returns the estimated portfolio returns and the risk associated with that portfolio. The syntax of frontcon is as under:

[PortRisk, PortReturn, PortWts] = frontcon (ExpReturn, ExpCovariance, NumPorts, PortReturn, AssetBounds, Groups, GroupBounds, varargin)

We have used this function for 20 portfolios. The portfolio 1 gives a portfolio weekly return of 1.1127% at the risk of 1.3694% and the portfolio comprised of 8 companies. This portfolio has the lowest risk and the risk is lowered by diversifying investment in 8 companies. The portfolio 20 gives a portfolio return at the risk of 9.0925% and the portfolio comprised of only one company. This portfolio gives highest return but also contains the highest risk. The efficient frontier for our given values is as under. The associated return and risk of each portfolio is described in Table 6:

Table 6. Risk associated with each portfolio

Portfolio No	Portfolio Return	Portfolio Risk
1	1.1127%	1.3694%
2	1.1903%	1.4578%
3	1.2678%	1.6697%
4	1.3454%	1.9324%
5	1.4230%	2.2264%
6	1.5005%	2.5407%
7	1.5781%	2.8711%
8	1.6556%	3.2147%
9	1.7332%	3.5679%
10	1.8108%	3.9389%
11	1.8883%	4.3340%
12	1.9659%	4.7513%
13	2.0434%	5.1856%
14	2.1210%	5.6427%
15	2.1986%	6.1287%
16	2.2761%	6.6375%
17	2.3537%	7.1710%
18	2.4312%	7.7630%
19	2.5088%	8.4072%
20	2.5864%	9.0925%

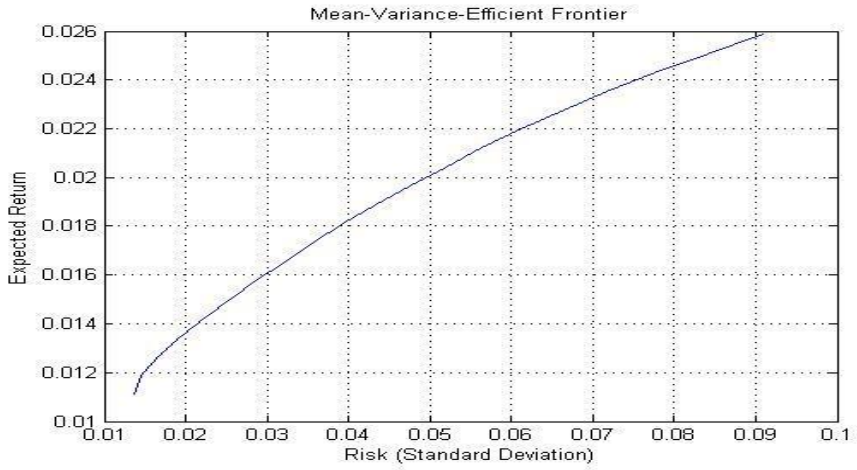

Figure 2. Risk and expected return of portfolio

A portfolio that offers maximum expected return for a given level of risk, or conversely the lowest level of risk for a given expected return is known as optimal portfolio. Efficient frontier is a set of optimal portfolios that suggests highest expected return for a defined level of risk or in other words lowest risk for a specified level of expected return. So we can say that the set of all efficient portfolios is called the efficient frontier, shown in graph presented in Figure 2.

5.7. Portfolio Performance

To assess the efficacy of our portfolios we measured the actual weekly performance of these stocks from July 2012 to Dec 2012 and compared it with the standard index of the Hong Kong Stock exchange for the same duration. For this we have used the Hang Seng Composite Index (HSCI) which is one of benchmark index for the Hong Kong Stock Exchange. The Hang Seng Composite Index (HSCI) offers a comprehensive Hong Kong market benchmark that covers about 95% of the total market capitalization of companies listed on the main board of the stock exchange of Hong Kong (SEHK). HSCI uses free float adjusted market capitalization methodology, and can be used as a basis for performance benchmarks. So to compare the portfolio performance weekly performance of HSCI is calculated and compared with the portfolio performance. Top three portfolios having least risk for the invested capital are used. The portfolio details of 3 portfolios formed are given in Tables 7, 8 and 9:

Table 7. Portfolio number 1 composition

Sr #	Company Name	Weight	Granule	Cluster Membership
1	Cr Asia - (831)	0.22186	Small	2
2	Franshion Ppt - (817)	0.00418	Medium	3
3	Haier Elec - (1169)	0.15436	Medium	3
4	Galaxy Ent - (27)	0.05250	Large	1
5	China Res Land - (1109)	0.02305	Large	2
6	National Elec H - (213)	0.16128	Micro	2
7	Hon Kwok Land - (160)	0.38278	Micro	3

Table 8. Portfolio number 2 composition

Sr #	Company Name	Weight	Granule	Cluster Membership
1	V.S. Intl - (1002)	0.01817	Nano	2
2	Cr Asia - (831)	0.24439	Small	2
3	Franshion Ppt - (817)	0.06214	Medium	3
4	Haier Elec - (1169)	0.15718	Medium	3
5	Galaxy Ent - (27)	0.03945	Large	1
6	National Elec H - (213)	0.14326	Micro	2
7	Hon Kwok Land - (160)	0.33540	Micro	3

Table 9. Portfolio number 3 composition

Sr #	Company Name	Weight	Granule	Cluster Membership
1	V.S. Intl - (1002)	0.040	Nano	2
2	Cr Asia - (831)	0.263	Small	2
3	Franshion Ppt - (817)	0.114	Medium	3
4	Haier Elec - (1169)	0.156	Medium	3
5	Galaxy Ent - (27)	0.024	Large	1
6	National Elec H - (213)	0.115	Micro	2
7	Hon Kwok Land - (160)	0.288	Micro	3

The actual weekly performance of our portfolios and the benchmark index for the July 2012 to December 2012 is given in Table 10:

Table 10. Comparison between Benchmark Index & Our Portfolios

Week No	HSCI	Portfolio # 1	Portfolio # 2	Portfolio # 3
Week # 1 Return	0.83%	0.26%	0.25%	0.27%
Week # 2 Return	-2.73%	-1.08%	-1.24%	-1.48%
Week # 3 Return	1.09%	0.01%	0.17%	0.26%
Week # 4 Return	-0.25%	0.23%	0.49%	0.82%
Week # 5 Return	0.85%	0.62%	0.48%	0.22%
Week # 6 Return	0.79%	3.06%	2.46%	1.86%
Week # 7 Return	-0.60%	-1.84%	-2.02%	-2.22%
Week # 8 Return	-0.65%	2.65%	2.63%	2.58%
Week # 9 Return	-2.58%	-1.42%	-1.36%	-1.40%
Week # 10 Return	1.90%	2.77%	2.62%	2.61%
Week # 11 Return	3.94%	3.15%	2.79%	2.59%
Week # 12 Return	0.52%	3.33%	2.66%	1.84%
Week # 13 Return	0.75%	1.06%	1.01%	1.05%
Week # 14 Return	0.60%	0.08%	0.30%	0.58%
Week # 15 Return	0.79%	-0.84%	-0.58%	-0.33%
Week # 16 Return	2.26%	1.41%	1.28%	1.17%
Week # 17 Return	0.12%	1.25%	1.18%	1.08%
Week # 18 Return	3.54%	3.73%	3.38%	3.17%
Week # 19 Return	-2.62%	-0.95%	-0.53%	-0.11%
Week # 20 Return	-1.15%	-0.31%	-0.41%	-0.44%
Week # 21 Return	3.15%	2.80%	2.79%	2.91%
Week # 22 Return	0.39%	3.21%	4.06%	4.94%
Week # 23 Return	0.91%	1.79%	2.03%	2.20%

The performance of our three portfolios and the HSCI benchmark index from July 2012 to December 2012 is shown above. The graphical view of weekly performance of our portfolios and benchmark index are shown in Figure 3:

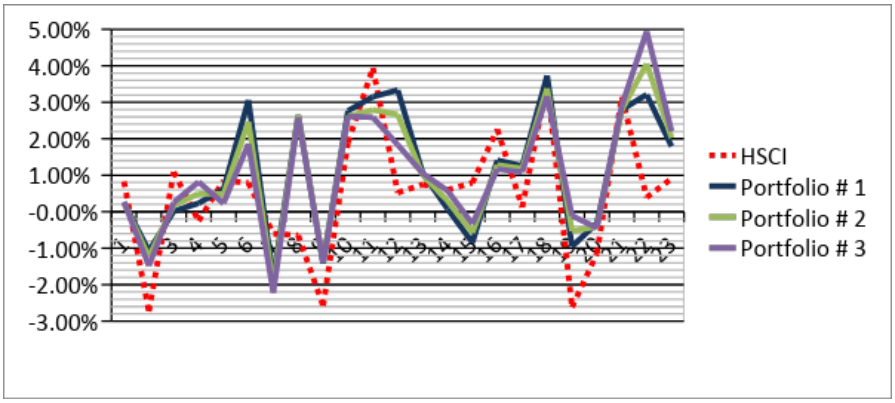

Figure 3. Performance comparison of three portfolios

The graph of Figure 3 clearly suggests that the performance of our three portfolios is greater than the Hong Kong Stock Exchange benchmark index HSCI. Another important point to note here is that in all the three portfolios, stocks belongs to different set of Granule and clusters. As a first step we divided the stocks into 5 different granules based on their market capitalization value and in the next step the granules are further sub divided into clusters. As we know that cluster helps in grouping similar records so we can say that clustering will divide stocks into similar groups. Therefore, selecting stocks from different groups will diversify our portfolio and as a result it will also reduce our risk, as we know that diversification reduces the risk. In the Portfolio 1 there are two stocks from Micro granule. First stock belongs to cluster 2 and second belong to cluster 3. There is one stock from the Small granule that belongs to cluster 2. There are two stocks from the Medium granule and both of them belong to cluster 3 of that granule. There are two stocks from the large granule that belong to cluster 1 and 2 of that granule. In the same way the stocks are also diversified from different granules and clusters in the portfolio 2 and 3 which helps in reduction of portfolio risk.

6. Conclusion

In this research a method for portfolio management by using granule based fuzzy data clustering is proposed. Granular Computing is integrated with Fuzzy based Particle Swarm Optimization technique to design portfolios that can offer returns matching the benchmark index. This method is also useful in selecting the stocks for the investment as initially there were 774 companies' data from 77 different industry groups of the Hong Kong Stock Exchange. There are 42 fields for every company present in the dataset which includes many different types of financial ratios to represent the financial position of these companies at the year-end 2011. This method reduces time for the selection of stocks from different categories and also helps in convenient grouping of stocks into a cluster and thus best performing stocks from those groups can be selected. The selection of stocks from different granules and then from different clusters will diversify the portfolio and as a result our portfolio risk will be reduced. The aim was to maximize return by investing in different groups of stocks that would individually react in a different ways for the same event. An important point to note here is that although diversification does not guarantee against loss, but diversification can be used as an important factor for maximizing returns while minimizing risk. We can reduce risk associated with a stock, but general market risks influence almost every stock. That's why for building an optimal portfolio we selected stocks from different granules and clusters so that each stock has its own characteristics and would react differently for the same event.

To design optimal portfolios, MATLAB function frontcon from financial toolbox is used. This helps us to design portfolios that maximize the return for the given value of risk. The results of designed portfolios are better than the benchmark index of the Hong Kong Stock Exchange which further validates the view. To summarize this work, it can be said that we have demonstrated

a granule based FPSO data clustering approach for the selection of stocks, portfolio management and designing portfolios on the efficient frontier.

References

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences, 10(2-3), 191-203. doi:10.1016/0098-3004(84)90020-7.

Cheng, S., Chen, S., & Jian, W. (2015). A Novel Fuzzy Time Series Forecasting Method Based on Fuzzy Logical Relationships and Similarity Measures. IEEE International Conference on Systems, Man, and Cybernetics, doi:10.1109/smc.2015.393.

Hammouda, K., & Karray, F. (2000). A Comparative Study of Data Clustering Algorithms. Retrieved from http://www.pami.uwaterloo.ca/pub/hammouda/sde625-paper.pdf.

Izakian, H., Abraham, A., & Snasel, V. (2009). Fuzzy clustering using hybrid fuzzy c-means and fuzzy particle swarm optimization. World Congress on Nature & Biologically Inspired Computing (NaBIC). doi:10.1109/nabic.2009.5393618.

Li, L., Liu, X., & Xu, M. (2007). A Novel Fuzzy Clustering Based on Particle Swarm Optimization. IEEE International Symposium on Information Technologies and Applications in Education. doi:10.1109/isitae.2007.4409243.

Li, X. Y., Sun, J. X., Gao, G. H., & Fu, J. H. (2011). Research of Hierarchical Clustering Based on Dynamic Granular Computing. JOURNAL OF COMPUTERS, 6(12), 2526-2533.

Maciel, L., Gomide, F., & Ballini, R. (2013). Forecasting Exchange Rates with Fuzzy Granular Evolving Modeling for Trading Strategies. Proceedings of the 8th conference of the European Society for Fuzzy Logic and Technology. doi:10.2991/eusflat.2013.40

Nanda, S., Mahanty, B., & Tiwari, M. (2010). Clustering Indian stock market data for portfolio management. Expert Systems with Applications, 37(12), 8793-8798. doi:10.1016/j.eswa.2010.06.026.

Nerurkar, P., Shirke, A., Chandane, M., & Bhirud, S. (2018). Empirical Analysis of Data Clustering Algorithms. Procedia Computer Science,125, 770-779. doi:10.1016/j.procs.2017.12.099

Östermark, R. (1996). A fuzzy control model (FCM) for dynamic portfolio management. Fuzzy Sets and Systems. doi:10.1016/0165-0114(96)84605-7.

Rajagopal, S. (2011). Customer Data Clustering Using Data Mining Technique. International Journal of Database Management Systems, AIRCC Publishing Corporation, 3(4).

Ravi, V., Pradeepkumar, D., & Deb, K. (2017). Financial time series prediction using hybrids of chaos theory, multi-layer perceptron and multi-objective evolutionary algorithms. Swarm and Evolutionary Computation, 36, 136-149. doi:10.1016/j.swevo.2017.05.003.

Shandilya, S. K., Shandilya, S., Deep, K., & Nagar, A. K. (2017). Handbook of research on soft computing and nature-inspired algorithms. Hershey, PA: Information Science Reference.

Shi, Y., & Eberhart, R. (1998). A modified particle swarm optimizer. IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360). doi:10.1109/icec.1998.699146.

Suganya, R., & Shanthi, R. (2012). Fuzzy C- Means Algorithm- A Review. International Journal of Scientific and Research Publications, IJSRP Inc, 2(11).

Zhang, Y., Agarwal, P., Bhatnagar, V., Balochian, S., & Yan, J. (2013). Swarm Intelligence and Its Applications. The Scientific World Journal, 2013, 3.

Zhu, Q., & Azar, A. (Eds.). (2015). Complex System Modelling and Control Through Intelligent Soft Computations (Vol. 319). Springer, Cham. doi: https://doi.org/10.1007/978-3-319-12883-2.

Zhu, H., Wang, Y., Wang, K., & Chen, Y. (2011). Particle Swarm Optimization (PSO) for the constrained portfolio optimization problem. Expert Systems with Applications, 38(8), 10161-10169. doi:10.1016/j.eswa.2011.02.075.

Professor Dr. **S. M. Aqil Burney** is the Head of Actuarial Sciences, Risk Management & Mathematics at Institute of Business Management (IoBM) Karachi. He holds M.Sc.(Statistics), M.Phil(Risk Theory and Insurance -Statistics) from University of Karachi (UoK) and Ph.D.(Mathematics) from Strathclyde University, Glasgow-UK along with many courses in Population Studies of UN, Computing. He has taught for more than 40 years at UoK and extensively delivered lectures at other institutions and universities of Pakistan and abroad. He also holds extensive experience of academic management and organization and Provost, Registrar, Project Director Development of Dept. of Computer Science and a Institute of Information technology and founding director of Main Communication Network of University of Karachi. Dr. Aqil Burney was Meritorious Professor at Dept of Computer Science University of Karachi prior to joining at IoBM. He has published more than 135 research papers and 7 books nationally and internationally in ICT, Mathematics, Statistics and Computer Science. He has supervised more than 10 PhD and 5 MS/M.Phil in Mathematics/Computer Science/Statistics and approved HEC Supervisor .Dr. Aqil Burney is Chairman (elect) National ICT Committee for Standard PSQCA- Ministry of Science & technology Govt. of Pakistan and member National Computing Education Accreditation Council (NCEAC), Member IEEE(USA), Member ACM(USA) was Fellow Royal Statistical Society UK) for 30 years or so. His fields of interests are algorithmic analysis & design of Multivariate Time series, Stochastic Simulation and Modeling, Software engineering, computer science, soft computing, risk theory and insurance e-health management and Data Sciences and Fuzzy and other logical systems.

Tahseen Jilani received the B.Sc. degree in Computer Science from Government Science Degree College, in 1998, and the M.Sc. (Statistics) and Ph.D. (Computer Science) from University of Karachi, Pakistan, in 2001 and 2007, respectively.
He is working as Associate Professor since 2014, in the Department of Computer Science, University of Karachi. Since January, 2016, he is engaged with the School of Computer Science and School of Medicine as post doc data scientist, University of Nottingham- UK. His current research interests include data sciences, machine learning in medical sciences, Statistical techniques for big data analytics, imprecise and uncertainty data modelling. Dr. Jilani is a member of Rough set society (RSS) and Association for Professional Health Analysts (APHA).
He is serving as member of technical committee and active reviewers for many national and international research activities. He was the recipient of the HEC Indigenous 5000 scholarship in 2003, the National Science Foundation grant 2010, the Nottingham University fellowship and honorary postdoc at University of Sterling.

Usman Amjad received BS. Degree in Computer Science from University of Karachi, in 2008. He recently completed his PhD in Computer Science from University of Karachi. His research interests include soft computing, machine learning, artificial intelligence and programming languages. He was the recipient of the HEC Indigenous 5000 scholarship in 2013. Currently, he is working as AI solution architect at Datics.ai Solutions.

Humera Tariq received B.E (Electrical) from NED University of Engineering and Technology in 1999. She joined MS leading to PhD program at University of Karachi in 2009 and completed her PhD in 2015. Currently she is working as Assistant Professor at Department of Computer Science, University of Karachi. Her research interest includes image processing, biomedical imaging, Modeling, Simulation and Machine Learning.

Mr. **Syed Shah Muhammad** is working as a Lecturer in the department of computer science in Virtual University of Pakistan. He obtained the degree of MS in computer sciences (MSCS) with the specialization of Computer Networks from the University of Agriculture, Faisalabad in 2005. Presently, he is Ph.D. Computer Science scholar at University of Engineering and Technology (UET) Lahore.