Journal of Optimization in Industrial Engineering 22 (2017) 49-59
DOI: 10.22094/joie.2017.276

A Hybrid Grey-Based Two-Step Clustering and Firefly Algorithm for Portfolio Selection

Farshad Faezy Razi[a,*], Naeimeh shadloo[b]

[a] *Assistant professor, Department of Industrial Management, Semnan Branch, Islamic Azad University, Semnan, Iran*
[b] *MSc, Department of Industrial Management, Semnan Branch, Islamic Azad University, Semnan, Iran*
Received 06 November 2014; Revised 02 July 2016; Accepted 15 September 2016

Abstract

Considering the concept of clustering, the main idea of the present study is based on the fact that all stocks for choosing and ranking will not be necessarily in one cluster. Taking the mentioned point into account, this study aims at offering a new methodology for making decisions concerning the formation of a portfolio of stocks in the stock market. To meet this end, Multiple-Criteria Decision-Making, Data Mining, and Multi-objective Optimization were employed. First, candidate stocks were clustered using two-step clustering method. Available stocks in each cluster were independently ranked using grey relational analysis. Firefly algorithm was employed for Pareto analysis of risk and ranking. The results of clustering in the stocks revealed that all candidate stocks were not placed in one cluster. The results of robustness analysis employed in ranking method verified the accuracy of calculations in the grey relational analysis through stock repetition of candidates in each cluster.

Keywords: Firefly algorithm, Grey relational analysis, Multiple-criteria decision-making, Portfolio optimization, Two-step clustering.

1. Introduction

Stock selection is considered as a challenging task in portfolio optimization. Selecting attractive stocks is regarded as significant short- and long-term decisions. Therefore, being equipped with right tools in the stock selection process plays a crucial role in supporting any decision-making process (Ince 2014). The main focus of the optimization is devoted to the amount of capital allocated for investment with the objective of enhancing the stock effectiveness. The mentioned topic was offered by Markowitz in a quantitative form. Markowitz formulated a question within a two-objective optimization problem. In this case, the optimization of expected returns is maximized, while the portfolio risk is minimized (Baykasoğlu, et al. 2015). Therefore, a decision-making problem has to be dealt with. Complexity, requirements, time, and information all vary in different decision-making processes. In each decision-making process, two factors, namely the value of each strategy for a decision-maker and the possible consequences, play key roles (Cabrera-Paniagua, et al. 2015). In any decision-making problems, a number of solutions can be offered. Any decision-makers' goal is to optimize the objectives. In most cases, multiple criteria, rather than one criterion, play significant roles in decision-making problems. Multiple-criteria decision-making techniques were introduced and developed based upon the mentioned assumption (Hamzaçebi and Pekkaya 2011). In Multiple-criteria decision-making techniques, indicators are evaluated on the basis of specific criteria and their weights (Xiao, et al. 2012). The present study addresses ranking, choosing, sorting, and describing a decision-making problem (Belton and Stewart 2002). In this article, ranking the stocks was designed based on clustering. Therefore, the main idea of this study was based on the fact that all stocks for choosing and ranking would not be necessarily present in one cluster. That is why ranking all stocks in one cluster cannot be effective concerning the accuracy. According to the methodology offered in this article, the stocks were divided into two using two-step clustering method. Available stocks in each cluster were independently ranked using grey relational analysis. The final portfolio was created with application of two-objective zero/one planning model and based upon firefly algorithm. Selected stocks were placed in the final portfolio with consideration of two objectives: risk minimization and ranking maximization. The second section of the article is dedicated to the literature review. The third, fourth and fifth sections address two-step clustering, grey relational analysis, and firefly algorithm, respectively. The sixth section offers a new methodology proposed for formation of portfolio. Seventh and eighth sections review the case study and sensitivity analysis, respectively. Conclusion is also offered in the ninth section.

* Corresponding author Email address: farshadfaezy@semnaniau.ac.ir

2. Review of the Related Literature

Zhou et al. (2013) studied the prediction mechanism of a portfolio of stocks with application of dynamic-dependent sparse factor models. They referred to net returns, risk-adjusted returns, and portfolio volatility as research variables (Zhou, Nakajima, & West, 2013). Yu et al. (2014) attempted to present a model for selecting a portfolio of stocks using principal component analysis and machine learning method. Variables, in this study, included the financial ratios of companies in Shanghai stock exchange (Yu et al, 2014). Zhang et al. (2014) studied the issue of stock selection by casual feature selection, principal component analysis, decision trees, and least absolute shrinkage and selection operator. Profitability, cash flow, and volatility factors were the main variables of this study (Zhang et al, 2014). Shen et al. (2014) used Multiple-attribute decision-making to present a compound decision-making method to study the problem of stock portfolio formation with application of an experimental approach. Financial ratios in five conductor industries of Taiwan stock exchange were the main variables of this study (Shen, Yan, & Tzeng, 2014). Bagheri et al. (2014) used ANFIS, QOSO, DTW, and WT models to present a mixed methodology based on artificial intelligence using machine learning and pattern recognition approaches. Beta risk coefficient was the main variable of this study (Bagheri et al, 2014). Ince (2014), in a study with experimental design, used CBR and MLP models, decision-tree algorithm, logistic regression, and Generalized Rule Induction (GRI) to study the short-term stock selection based on reasoning technique. The main variables of this study were average efficiency, Sharpe ratio, and ideal profit (Ince, 2014). Sonsino and Shavit (2014) made use of parametric tests, specialized randomization, and permutation tests to discover a number of patterns to predict purely technical efficiency from the provided experimental data. Rezaie, Dalfard, Hatami-Shirkouhi, and Nazari-Shirkouhi (2013) used fuzzy data envelopment analysis and goal programming to study investment portfolio formation. Financial ratios, sigma index, beta coefficient, net profits, expenses, incomes, and partners' return on assets were the main variables examined in the mentioned study. In another study, the mental patterns formed by shareholders gaining profit or loss from previous deals were studied. The variables of this study were financial ratios, beta coefficients per share, price instability, and past returns (Duxbury, Hudson, Keasey, Yang, & Yao, 2013). Chang, Yang, and Tsai (2014) made use of BAT algorithm to offer a methodology for stock portfolio formation. The variables studied in this study were inflation rate, risk-free interest rate, market price, gross profit margin, and stock price. Another pertinent study addressed the association rules, clustering, and decision-making tree to study the investment portfolio formation (Cheng, 2013). The mentioned study investigated the financial ratios, expenses, incomes, net growth rate, return on total assets, and profit per share in investment portfolio formation. In another study, the efficiency of firms under consideration was measured using Data Envelopment Analysis (DEA) model in order to form the portfolio. Therefore, in the first stage, initial stock portfolio was created. In the second stage, MCDM was used to determine the budget allocation for each stock (Huang, et al. 2015). Dynamism was studied in multiple periods of time in order to maximize the portfolio return and optimize the minimum and maximum risks of each stock. The results of stock selection were studied in one period. Dynamic simulation and planning results revealed that studying multiple periods outweigh that of one period (Sun, et al. 2016). A summary of studies conducted on investment portfolio formation is presented in Table 1. Although Change et al. (2013) focused on the portfolio creation using the clustering method, the final portfolio was not completed through ranking methods as well as meta-heuristic models. Other researchers did not cluster the studied shares, and assumed that all candidate shares were placed in one cluster. Alinezhad (2011) investigated the portfolio selection problem in fuzzy conditions. The tool adopted in the aforementioned study was goal programming. This study mainly aimed at analyzing the uncertain identity of portfolio. Naderi (2013) studied the portfolio problem within the framework of a hard problem. To handle this condition, three algorithms were developed, namely imperialist competitive, simulated annealing, and genetic algorithms. The respective problem was investigated as a mixed-integer programming. The results of running the algorithms revealed that imperialist competitive algorithm exhibited a better optimality in comparison with the other two algorithms (Naderi, 2013). Mehdizadeh and Moghaddam (2008) examined alternative clustering and adopted the practical swarm optimization technique as the clustering tool. The results indicated that the utilized algorithm improved fuzzy C-means clustering performance.

The present study is provided to fill the mentioned theoretical gap. Accordingly, candidate shares were first clustered and then ranked. The final portfolio was selected based on a meta-heuristic method.

Table 1
Studies conducted on investment portfolio formation

No,	Objective	Decision model	1						2											Researcher(s)
			A	b	c	d	e	f	A	b	c	d	e	f	g	h	k	m	n	
1	Improvement of the predictions and decision-making process about stocks	Dynamic dependent sparse factor models			*								*		*					(Zhou et al., 2013)
2	Creation of an effective model for selecting stocks	*SVM-PCA*			*									*						(Yu et al., 2014)
3	Investigation of CFS algorithm and feature selection for modeling stock prediction	*PCA-DT-LASSO*			*					*	*							*		(Zhang et al., 2014)
4	Presentation of MADM method of decision-making for investigation of glamour stock selection	MADM			*									*						(Shen et al., 2014)
5	Presentation of a hybrid artificial intelligence method as a business advisory system	*ANFIS-QPSO-DTW-WT*			*									*						(Bagheri et al., 2014)
6	Selection of a short-term stock based on reasoning technique	*CBR-MLP-DT-GRI-Logistic regression*			*				*			*							*	(Ince, 2014)
7	Discovering the patterns to predict purely technical efficiency from experimental data	parametric tests, specialized randomization test, permutation tests			*									*						(Sonsino & Shavit, 2014)
8	Comprehensive assessment and ranking of companies in stock exchange	Ranking			*				*		*		*					*	*	(Rezaie et al., 2013)
9	Investigation of the effects of previous gains and losses on preferences of investors and formation of basket of stocks	*FDEA*			*				*								*	*	*	(Duxbury et al., 2013)
10	Presentation of a method for offering an efficient basket of stocks for stock investment	*BAT*			*						*			*			*			(Chang et al., 2014)
11	Establishment of a relationship between financial data from public companies and return on investment using data mining technology	Association Rules, Clustering Analysis, Decision Tree			*						*			*		*	*			(Cheng, 2013)
12	Portfolio Selection	DEA-MODM			*									*						(Huang, et al. 2015)
13	Dynamic Stock Selection	Simulation			*									*						(Sun et al., 2016)

1. Approaches to Papers

a)Computational; b) machine learning; c) comparative; d) empirical; e) data mining; f) experimental

2. Research Variables

a)Financial and Sharpe ratios; b) risk; c) profit (ideal, gross, share, net); d) liquidity and cash flow; e) expenses and incomes; f) rates (inflation, interest, redemption yield, net growth, profit, exchange); g) series of random historical data; h) price; k) sigma index and beta coefficient; m) returns (net, risk, average, gross, regulatory risk, past); n) variability of stock market and factor

3. Two-Step Cluster Analysis

The Two-Step Cluster Analysis method was used for clustering the decision problems, in which input fields were of different measurement scales (Chiu et al, 2001). Like the K-means method, the mentioned method effectively deals with very large datasets (Sarstedt & Mooi, 2014). It clusters data records in two steps. The first step uses the K-means algorithm for clustering. The K-means algorithm is as follows:

Select a population and a distance between the individuals.
Choose k random individuals as initial centers.
Repeat.
For each individual,
Calculate its distance from all the centers, and Put it in the cluster with the nearest center.
Compute new cluster centers by taking the mean of each cluster till there are no more changes in the clusters.
Based on the results of the first step, the second step uses the Hierarchical Agglomerative Clustering Procedure algorithm. In agglomerative method, each item is considered as a cluster, and then these clusters merge in the clustering process to create a unique cluster. Common Agglomerative Hierarchical Clustering methods are presented in Figure 1 (Momeni, 2011).

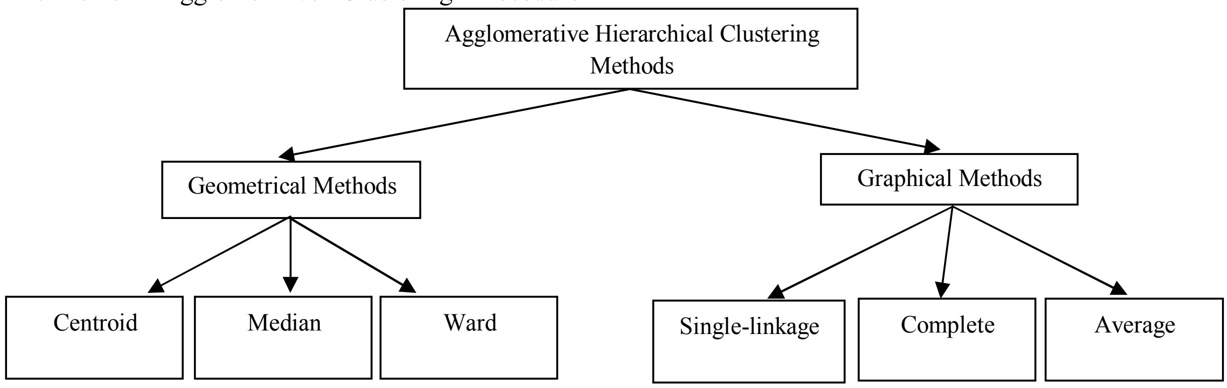

Fig. 1. Classification of hierarchical clustering methods

The advantage of this type of clustering in comparison with K-means is that the results are unbiased with initial parameters. The steps of hierarchical clustering algorithm are as follows (Merceron & Yacef, 2005):
Select a population and a distance between individuals.
Each individual forms an initial group.
Calculate distances between all groups to form a distance matrix.
While there is more than one group and the distance between two nearest groups falls below a given threshold,
Repeat.
Cluster these two nearest groups into one.
Recalculate the distances between all other groups and this newly formed group.
In this study, IBM SPSS Modeler 14.2 was used for clustering with two-step clustering. In this method, records were clustered with the K-means algorithm. In doing so, the advantage of linear run time was to make use of the K-means algorithm to manage the initial population. After implementing the K-means algorithm, K clusters were re-clustered through Agglomerative Hierarchical Clustering.

3. Grey Relational Analysis

The theory of Grey System was first presented by Deng in 1982 (Faezy Razi 2015). Grey relational analysis is an effective means for estimating the behavior of an uncertain system with insufficient or incomplete information (Tzeng & Huang, 2011). Grey relational analysis can efficiently analyze the correlation between processing agents and multi-responses (Khan, et al. 2014). The grey relational analysis method utilizes the theory of grey systems for analyzing the systems. The mentioned theory is based on qualitative and quantitative analyses (Sun, 2014).

3.1. Steps of Grey Relational Model:

Step one: Creating reference series based on equation 1.

$$x_0 = (x_0(1), x_0(2), \dots, x_0(j), epx_0(n)) \quad (1)$$

Therefore, this reference series will be defined as equation 2.

$$x_0 = \begin{bmatrix} x_1(1) & x_1(2) & finedx_1(n) \\ x_2(1) & x_2(2) &)inedx_2(n) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(1) & x_n(2) & & x_n(n) \end{bmatrix} \quad (2)$$

Step two: Normalization of data.
If the index is of the higher-the-better type, equation 3 will be used.

$$x_i^*(j) = \frac{x_i(j) - \text{Min } x_i(j)}{\text{Max } x_i(j) - \text{Min } x_i(j)} \quad (3)$$

If the index is of the lower-the-better type, equation 4 will be used.

$$x_i^*(j) = \frac{Max\, x_i(j) - x_i(j)}{\max x_i(j) - Min\, x_i(j)} \quad (4)$$

If the index is of the nearer-to-objective type, equation 5 will be used.

$$x_i^*(j) = \frac{x_0(j) - Min\, x_i(j)}{Max\, x_i(j) - Min\, x_i(j)} \quad (5)$$

Step three: Calculation of the distance from$\Delta_{0i}(j)$, which is in fact the absolute value of the distance between x_0^* andx_i^*.

$$\Delta_{0i}(j) = |\, x_0^*(j)\; x_i^*(j) \,| \quad (6)$$

Step four: Application of the grey confidence degree equation based on equation 7.

$$Y_{0i}(j) = \frac{\Delta Min + \varepsilon \Delta Max}{\Delta_{0i}(j) - \varepsilon \Delta Max} \quad (7)$$

$$\Delta Max = Max\; Max\; \Delta_{0i}(j) \quad (8)$$

$$\Delta Min = Min\; Min\; \Delta_{0i}(j) \quad (9)$$

Confidence degree $\varepsilon = [0,1]$

Step five: Calculation of grey confidence degree (Tzeng & Huang, 2011).

$$_{0i} = [\sum w_i(j) * Y_{0i}(j)] \quad (10)$$

4. Firefly Algorithm

Yang (2009) developed the Firefly algorithm (Poorbagheri 2015). The Firefly algorithm is inspired by the flashing behavior of fireflies (Upadhyay et al 2014). The algorithm inspired by the firefly assumes the following ideal rules:
All fireflies are unisexual.

- Attractiveness of the fireflies is proportional to their brightness (Kavousi-Fard et al 2014).
- Brightness of fireflies is defined by the value of the objective function.

Based on these three rules, the main basis of the firefly algorithm (FA) can be presented in a pseudocode format (Gao, 2015).

```
Objective function f(x)   x=(x1, x2, …,xd)^T
Generate initial population of fireflies xi(i=1,2, …n)
Light intensity Ii at xi is determined by f(xi).
Define light absorption coefficient γ
while (t <MaxGeneration )
for i =1:n all n fireflies
    for j =1:i all n fireflies
        if(Ii < Ij)
            Move firefly i towards j in d-dimension.
            Attractiveness varies with distance r via exp[-γr].
            Evaluate new solutions and update light intensity.
        end if
    end for j
  end for i
  Rank the fireflies and find the current best
  End while
  Post process the results and visualization.
```

5. A New Framework for Selecting Investment Portfolios

In this part of the paper, a comprehensive framework is presented for selecting investment portfolios. The main aim of this study is considering the current limitations on investment portfolio formation and designing a bi-objective zero-one mathematical programming model based on two-step clustering and grey relational algorithms generating the best optimal Pareto combination solved by firefly algorithm. As Figure 2 demonstrates, the data can be clustered using IBM SPSS Modeler 14.2 and later be ranked by grey relational analysis. Finally, the most optimal portfolio can be selected with application of firefly algorithm. The framework presented in the study is provided in Fig 2.

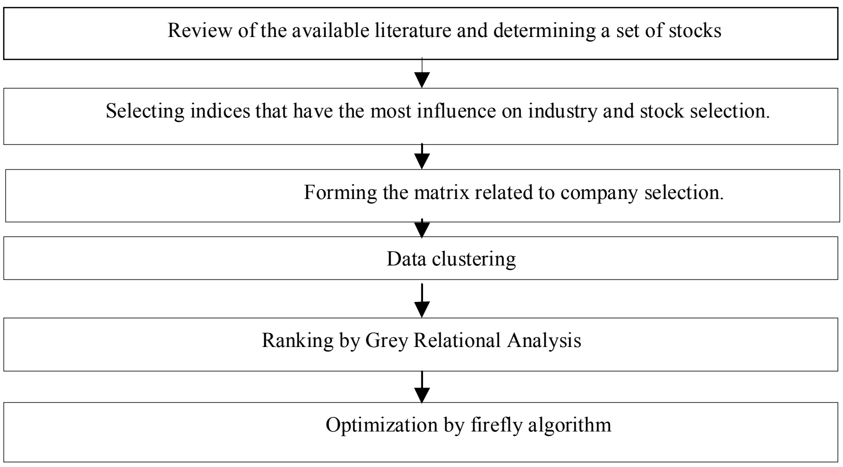

Fig. 2. A new framework to form investment portfolios

6. Case Study

In this section, a case study is presented, which describes the portfolio selection for investors in Tehran Stock Exchange using the hybrid algorithm of two-step clustering based on the gray relationships analysis and the firefly algorithm. This study used the historical data of firms. In this study, 4 variables were considered. The presented indicators were net income to sales (c1), return on assets (c2), return on investment (c3), and utility measurement (c4). Input data are presented in Table 2.

For clustering the data in Table 1, IBM SPSS Modeler 14.2 was used. In the second step, clustering with the two-step clustering method was based on the agglomerative clustering method. The centroids were determined based on the single-linkage method. The result of implementing the grey relationship analysis for complete ranking of shares per cluster is summarized in Table 3.

Table 2
Input data

A1	6.93	4.75	27.86	5.87	A16	14.01	16.08	37.43	**2.33**
A2	353.93	31.13	56.11	1.80	A17	5.82	27.68	49.93	**1.80**
A3	10.59	13.08	29.26	2.24	A18	13.55	10.15	50.25	**4.95**
A4	17.76	13.55	37.36	2.76	A19	10.66	1.24	9.95	**8**
A5	3.91	2.32	9.43	4.03	A20	31.02	21.11	35.44	**1.68**
A6	28.64	21.32	49.70	2.33	A21	24.09	22.45	48.79	**2.17**
A7	53	19.36	14.09	2.48	A22	14.99	13.12	16.03	**1.22**
A8	16.53	23.36	40.07	1.72	A23	19.73	21.90	69.05	**3.15**
A9	14.50	13.70	38.45	2.81	A24	1.73	0.86	3.10	**3.61**
A10	5.88	11.47	33.91	2.96	A25	12.43	12.94	27.92	**2.16**
A11	39.39	6.34	27.11	4.28	A26	11	14.16	41.73	**2.95**
A12	8.06	11.37	41.88	3.68	A27	10.85	16.88	40.89	**2.42**
A13	17.28	14.29	31.11	2.18	A28	13.23	13.67	56.76	**4.15**
A14	1.30	1.11	3.28	2.94	A29	6.76	4.34	13.91	**3.20**
A15	9.86	7.85	13.15	1.68	A30	11.06	12.74	23.09	**1.81**

Table 3
The results of running grey relational analysis for complete ranking of the data

(a)

cluster	2	1	1	1	2	1	1	1	1	**1**
GRA	1.9684	3.1248	1.1643	1.1935	1.3038	1.2128	1.1377	1.1359	1.1901	**1.1962**

(b)

cluster	2	1	1	2	2	1	1	1	2	**1**
GRA	3.0276	1.2315	1.1646	1.3394	1.4682	1.1608	1.1566	1.3153	1.3801	**1.1530**

(c)

cluster	1	1	1	2	1	1	1	1	2	**1**
GRA	1.1943	1.1531	1.4354	1.3348	1.1662	1.1921	1.1625	1.3005	1.3730	**1.1581**

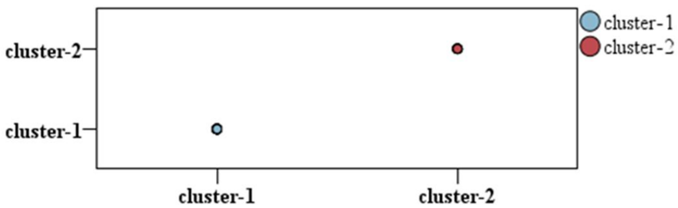

Fig. 3. The results of data clustering by two-step clustering method

To calculate the optimal number of clusters in the two-step clustering method, the Auto Cluster node in IBM SPSS Modeler 14.2 was used. The calculation summary for selecting the optimal number of clusters with four criteria is summarized in Table 4. As can be seen, the optimal number of clusters is 2.

Table 4
Implementation results of Auto Cluster node for determining the optimal number of clusters

Algorithm	Build Time(Min)	Silhouette	Number of Clusters
Two Steps	<1	0.586	2
K-means	<1	0.57	5
KOHONEN	<1	0.453	10

The bi-objective zero-one mathematical programming model of the research is presented as follows:

Max

$$Z_1 = 1.9684x_1 + 3.1248x_2 + 1.1643x_3 + 1.1935x_4 + 1.3038x_5 + 1.2128x_6 + 1.1377x_7 + 1.1359x_8 + 1.1901x_9 + 1.1962x_{10} + 3.0276x_{11} + 1.2315x_{12} + 1.1646x_{13} + 1.3394x_{14} + 1.4682x_{15} + 1.1608x_{16} + 1.1566x_{17} + 1.3153x_{18} + 1.3801x_{19} + 1.1530x_{20} + 1.1943x_{21} + 1.1531x_{22} + 1.4354x_{23} + 1.3348x_{24} + 1.1662x_{25} + 1.1921x_{26} + 1.1625x_{27} + 1.3005x_{28} + 1.3730x_{29} + 1.1581x_{30}$$

Min

$$Z_2 = 0.5x_1 + 0.8x_2 + 0.6x_3 + 0.3x_4 + 0.2x_5 + 0.1x_6 + 0.4x_7 + 0.2x_8 + 0.7x_9 + 0.8x_{10} + 0.9x_{11} + 0.1x_{12} + 0.2x_{13} + 0.3x_{14} + 0.2x_{15} + 0.4x_{16} + 0.6x_{17} + 0.5x_{18} + 0.4x_{19} + 0.7x_{20} + 0.8x_{21} + 0.5x_{22} + 0.4x_{23} + 0.3x_{24} + 0.2x_{25} + 0.4x_{26} + 0.2x_{27} + 0.3x_{28} + 0.2x_{29} + 0.1x_{30}$$

s.t:

$$0.1806x_1 - 0.0733x_2 + 0.1975x_3 + 0.2056x_4 + 0.2316x_5 + 0.2262x_6 + 0.5964x_7 + 0.2360x_8 + 0.3407x_9 + 0.1505x_{10} + 0.4995x_{11} + 0.2927x_{12} + 0.2455x_{13} + 0.0936x_{14} + 0.1031x_{15} + 0.1984x_{16} + 0.0790x_{17} + 0.2265x_{18} + 0.3853x_{19} + 0.5143x_{20} + 0.3749x_{21} + 0.1912x_{22} + 0.3545x_{23} + 0.1013x_{24} + 0.2267x_{25} + 0.1095x_{26} + 0.1304x_{27} + 0.25x_{28} + 0.3228x_{29} + 0.1590x_{30} \geq 23\%$$

$$0.10x_1 - 134.41x_2 - 0.46x_3 - 1.43x_4 + 0.12x_5 - 0.44x_6 - 1.28x_7 - 3.62x_8 - 1.40x_9 - 0.18x_{10} + 0.001x_{11} - 0.11x_{12} - 0.66x_{13} + 0.08x_{14} - 3.20x_{15} - 1.67x_{16} - 63.63x_{17} - 2.47x_{18} + 0.001x_{19} - 3.72x_{20} - 1.27x_{21} + 0.001x_{22} - 0.26x_{23} + 0.06x_{24} - 0.20x_{25} - 0.14x_{26} - 5.39x_{27} - 2.58x_{28} - 0.13x_{29} - 0.21x_{30} \leq 20\%$$

$$x_1 + x_2 \leq 1$$

$$x_4 + x_{19} \leq 0$$

$$X_j \in \{0,1\}, \qquad j = 1, \dots, 30$$

The results of solving Model 1 using the firefly algorithm are presented in Figure 4. For the firefly algorithm implementation, MATLAB R2010a was used. The main parameters of the algorithm are summarized in Table 5.

Table 5
Main parameters of firefly algorithm

Description	Value
Maximum Number of Iterations	100
Number of Fireflies (Swarm Size)	100
Light Absorption Coefficient	1
Attraction Coefficient Base Value	0.2
Mutation Coefficient	0.9
Mutation Coefficient Damping Ratio	0.99
Uniform Mutation Range	0.05*(VarMax-VarMin)
Decision Variables Lower Bound	VarMin=0
Decision Variables Upper Bound	VarMax=1

The time required to solve the firefly algorithm was 175.518 seconds.

7. Sensitivity Analysis

In this section, the number of research variables was increased, and the results were compared with the genetic algorithm to examine the validity of the model. In the first step, the number of criteria related to selection of shares was increased to 8. These measures included net income to sales (c1), return on assets (c2), return on investment (c3), utility measurement (c4), gross profit to sales (c5), operating profit (c6), margin to sales (c7), and profit to margin (c8). All steps described in the case

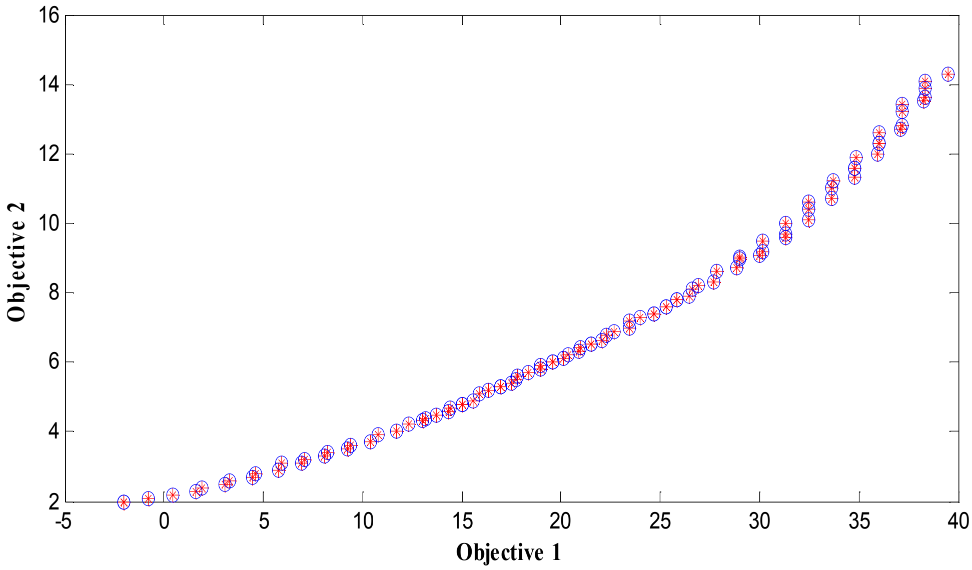

Fig. 4. The results of the model solved by the firefly algorithm

study were repeated once more with 8 criteria. The results of solving the model using the firefly algorithm are presented in Table 6. The time required to solve the firefly algorithm was 175.6561 seconds. When the number of variables was increased to 12 criteria, the variables were net income to sales (c1), return on assets (c2), return on investment (c3), utility measurement (c4), gross profit to sales (c5), operating profit (c6), margin to sales (c7), profit to margin (c8), return on fixed assets (c9), current ratio (c10), future ratio (c11), and liquidity ratio (c12). All steps described in the case study were repeated once more with 12 criteria. The results of solving the model using the firefly algorithm are presented in Table 6. The time required to solve the firefly algorithm was 176.4392 seconds. When the number of variables was increased to 16 criteria, the variables were net income to sales (c1), return on assets (c2), return on investment (c3), utility measurement (c4), gross profit to sales (c5), operating profit (c6), margin to sales (c7), profit to margin (c8), return on fixed assets (c9), current ratio (c10), future ratio (c11), liquidity ratio (c12), current assets ratio (c13), cash flow ratio (c14), inventory turnover (c15), and receivable collection period (c16). All the steps described in the case study were repeated once more with 16 criteria. The results of solving the model using the firefly algorithm are presented in Table 6. The time required to solve the firefly algorithm was 186.7923 seconds. When the number of variables was increased to 20 criteria, the variables were net income to sales (c1), return on assets (c2), return on investment (c3), utility measurement (c4), gross profit to sales (c5), operating profit (c6), margin to sales (c7), profit to margin (c8), return on fixed assets (c9), current ratio (c10), future ratio (c11), liquidity ratio (c12), current assets ratio (c13), cash flow ratio (c14), inventory turnover (c15), receivable collection period (c16), product to working capital ratio (c17), current capital turnover (c18), fixed asset turnover (c19), and total assets turnover (c20). The time required to solve the firefly algorithm was 181.513 seconds. Then, the number of variables was increased to 24, and the whole process was repeated. The variables included net income to sales (c1), return on assets (c2), return on investment (c3), utility measurement (c4), gross profit to sales (c5), operating profit (c6), margin to sales (c7), profit to margin (c8), return on fixed assets (c9), current ratio (c10), future ratio (c11), liquidity ratio (c12), current assets ratio (c13), cash flow ratio (c14), inventory turnover (c15), receivable collection period (c16), product to working capital ratio (c17), current capital turnover (c18), fixed asset turnover (c19), total assets turnover (c20), debt ratio (c21), debt to equity ratio (c22), special assets to fixed value ratio (c23), and long-term debt to equity ratio (c24). The time required to solve the firefly algorithm was 184.573 seconds. Finally, the number of variables was increased to 27, and the whole process was repeated. The variables were net income to sales (c1), return on assets (c2), return on investment (c3), utility measurement (c4), gross profit to sales (c5), operating profit (c6), margin to sales (c7), profit to margin (c8), return on fixed assets (c9), current ratio (c10), future ratio (c11), liquidity ratio (c12), current assets ratio (c13), cash flow ratio (c14), inventory turnover (c15), receivable collection period (c16), product to working capital ratio (c17), current capital turnover (c18), fixed asset turnover (c19), total assets turnover (c20), debt ratio (c21), debt to equity ratio (c22), special assets to fixed value ratio (c23), and long-term debt to equity ratio (c24). The time required to solve the firefly algorithm was 184.8406 seconds. The obtained were compared with the genetic algorithm whose parameters for execution in MATLAB 2010a were set as follows: Population Size 60, Crossover Rate 0.6, Probability of Mutation 0.4, and Maximum of Generation 200 iterations.

Table 6
Comparison of the results of genetic algorithm and firefly algorithms

4	175.518	41.393	23.06058	5.15425	22.2739	3.2
8	175.6561	41.284	32.55804	4.971088	29.7724	3.5
12	176.4392	41.369	22.98625	5.318859	24.0614	5.333333
16	184.7923	41.159	31.79223	4.745104	30.6651	2.333333
20	181.513	42.456	32.0983	5.260353	29.5765	4.333333
24	184.573	41.128	32.09536	5.146295	36.8889	5
27	184.8406	41.728	33.27823	5.129348	43.3968	6.8

The results presented in Table 6 reveal that when the number of variables was 4, 8, 12, 16, and 20 criteria, the firefly algorithm provided higher optimality in the first objective. When the number of variables was increased to 24 and 27 criteria, the genetic algorithm provided higher optimality in the first objective. About the second objective, when the number of variables was 4, 8, 12, 16, and 20 criteria, the genetic algorithm revealed better optimality. For 12 and 27 variables, the firefly algorithm provided better optimality compared with that of genetic algorithm.

8. Conclusion

Formation of stock portfolio is one of the best known models in the field of multi-objective optimization. In the stock portfolio optimization problem, the decision-maker analyzes the Pareto risk and rank combination to select candidate shares through a single- or multi-objective mathematical programming model. In this category of decision problems, given the multiplicity of decision variables, the problem is naturally considered as a NP-Hard decision problem. To deal with these types of optimization problems, application of meta-heuristic algorithms can be regarded as a main approach. In the present study, the conditions related to the selection and formation of Pareto optimal stock portfolio was based on the mentioned view. In this study, the studied shares were clustered using two-step clustering model.

IBM SPSS Modeler 14.2 was used for clustering the studied shares. To calculate the optimal number of clusters in two-step clustering model, the Auto Cluster node in IBM SPSS Modeler 14.2 was used. It was found that in the studied shares, the optimal number of clusters was 2. Accordingly, it was found that, rather than placing all shares in one cluster, two clusters should be created; thus, the decision-maker can select shares from two clusters. In this study, two initial stock portfolios were created using two-step clustering. Each cluster created from candidate shares was ranked by the two-step clustering algorithm based on the studied indicators by the grey relational analysis technique. The advantage of using this technique to assess the studied shares is that it performs complete rankings of shares. Then, a zero-one two-objective programming model was designed for the Pareto analysis of the ranking resulted from grey relational analysis risk and the beta factor per share. To deal with the model, the firefly algorithm and the genetic algorithm were used. To check the validity of the model, the number of properties related to stock selection was increased, and shares were clustered respectively with 8, 12, 16, 20, 24, and 27 criteria with application of two-step clustering.

The newly established clusters were re-ranked by grey relational analysis. The results showed that when the number of variables was 4, 8, 12, 16, and 20 criteria, the firefly algorithm provided higher optimality with respect to the first objective. When the number of variables was increased to 24 and 27 criteria, the genetic algorithm provided higher optimality with regard to the first objective. Considering the second objective, when the number of variables was 4, 8, 12, 16, and 20 criteria, the genetic algorithm presented better optimality. For 12 and 27 criteria, the firefly algorithm provided better optimality compared with the genetic algorithm.

References

Alinezhad, A., Zohrehbandian, M., Kian, M., Ekhtiari, M., & Esfandiari, N. (2011). Extension of Portfolio Selection Problem with Fuzzy Goal Programming: A Fuzzy Allocated Portfolio Approach. *Journal of Optimization in Industrial Engineering, Volume 4*(9), 69-76.

Bagheri, Ahmad, Peyhani, Hamed Mohammadi, & Akbari, Mohsen. (2014). Financial forecasting using ANFIS networks with Quantum-behaved Particle Swarm

Optimization. *Expert Systems with Applications, 41*(14), 6235-6250.

Baykasoğlu, A., Yunusoglu, M. G., & Özsoydan, F. B. (2015). A GRASP based solution approach to solve cardinality constrained portfolio optimization problems. *Computers & Industrial Engineering, 90*, 339-351.

Belton, Valerie, & Stewart, Theodor. (2002). *Multiple criteria decision analysis: an integrated approach*: Springer Science & Business Media.

Cabrera-Paniagua, D., Cubillos, C., Vicari, R., & Urra, E. (2015). Decision-making system for stock exchange market using artificial emotions. *Expert Systems with Applications, 42*(20), 7070-7083.

Chang, Jui-Fang, Yang, Tsai-Wei, & Tsai, Pei-Wei. (2014). Stock Portfolio Construction Using Evolved Bat Algorithm *Modern Advances in Applied Intelligence* (pp. 331-338): Springer.

Cheng, Shou-Hsiung. (2013). A stock selective system by using hybrid models of classification *Intelligent Information and Database Systems* (pp. 127-134): Springer.

Chiu, Tom, Fang, DongPing, Chen, John, Wang, Yao, & Jeris, Christopher. (2001). *A robust and scalable clustering algorithm for mixed type attributes in large database environment.* Paper presented at the Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining.

Duxbury, Darren, Hudson, Robert, Keasey, Kevin, Yang, Zhishu & Yao, Songyao. (2013). How prior realized outcomes affect portfolio decisions. *Review of Quantitative Finance and Accounting, 41*(4), 611-629.

Faezy Razi, F. (2015). A Grey-Based Fuzzy ELECTRE Model for Project Selection. *Journal of Optimization in Industrial Engineering, 8*(17), 57-66.

Hamzaçebi, C., & Pekkaya, M. (2011). Determining of stock investments with grey relational analysis. *Expert Systems with Applications, 38*(8), 9186-9195.

Huang, C.-Y., Chiou, C.-C., Wu, T.-H., & Yang, S.-C. (2015). An integrated DEA-MODM methodology for portfolio optimization. *Operational Research, 15*(1), 115-136.

Ince, Huseyin. (2014). Short term stock selection with case-based reasoning technique. *Applied Soft Computing, 22*, 205-212.

Kantardzic, Mehmed. (2011). *Data mining: concepts, models, methods, and algorithms*: John Wiley & Sons.

Kavousi-Fard, Abdollah, Samet, Haidar, & Marzbani, Fatemeh. (2014). A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting. *Expert systems with applications, 41*(13), 6047-6056.

Khan, Zahid A, Siddiquee, Arshad N, Khan, Noor Zaman, Khan, Urfi, & Quadir, GA. (2014). Multi response optimization of wire electrical discharge machining process parameters using Taguchi based grey relational analysis. *Procedia Materials Science, 6*, 1683-1695.

Mehdizadeh, E., & Moghaddam, R. T. (2008). Fuzzy Particle Swarm Optimization Algorithm for a Supplier Clustering Problem. *Journal of Industrial Engineering, 1*, 17-24.

Merceron, Agathe, & Yacef, Kalina. (2005). Clustering students to help evaluate learning *Technology Enhanced Learning* (pp. 31-42): Springer.

Momeni, Mansor. (2011). Data Clustering. *momeni.*

Naderi, B. (2013). The project portfolio selection and scheduling problem: mathematical model and algorithms. *Journal of Optimization in Industrial Engineering, 6*(13), 65-72.

Baykasoğlu, Adil, Mualla Gonca Yunusoglu, and F Burcin Özsoydan, (2015). A GRASP based solution approach to solve cardinality constrained portfolio optimization problems. *Computers & Industrial Engineering* 90:339-351.

Belton, V, & Theodor S. (2002). Multiple criteria decision analysis: an integrated approach: Springer Science & Business Media.

Cabrera-Paniagua, D. (2015). Decision-making system for stock exchange market using artificial emotions. *Expert Systems with Applications* 42(20):7070-7083.

Faezy Razi, F. (2015). A Grey-Based Fuzzy ELECTRE Model for Project Selection. *Journal of Optimization in Industrial Engineering* 8(17), 57-66.

Hamzaçebi, C. & Mehmet P. (2011). Determining of stock investments with grey relational analysis. Expert Systems with Applications 38(8), 9186-9195.

Huang, Chun-Y, (2015). An integrated DEA-MODM methodology for portfolio optimization. *Operational Research,* 15(1), 115-136.

Ince, H. (2014). Short term stock selection with case-based reasoning technique. *Applied Soft Computing* 22, 205-212.

Khan, Zahid A, (2014). Multi response optimization of wire electrical discharge machining process parameters using Taguchi based grey relational analysis. *Procedia Materials Science* 6, 1683-1695.

Poorbagheri, Tahereh, (2015). Vendor Managed Inventory of a Supply Chain under Stochastic Demands. *Journal of Optimization in Industrial Engineering* 8(18), 47-60.

Sun, Y. (2016). Multi-period portfolio optimization under probabilistic risk measure. Finance Research Letters.

Xiao, Z. (2012). The trapezoidal fuzzy soft set and its application in MCDM. *Applied Mathematical Modelling* 36(12), 5844-5855.

Rezaie, Kamran, Dalfard, Vahid Majazi, Hatami-Shirkouhi, Loghman, & Nazari-Shirkouhi, S. (2013). Efficiency appraisal and ranking of decision-making units using data envelopment analysis in fuzzy environment: a case study of Tehran stock exchange. *Neural Computing and Applications, 23*(1), 1-17.

Sarstedt, Marko, & Mooi, Erik. (2014). A Concise Guide to Market Research. *The Process, Data,*

Shen, Kao-Yi, & Yan, Min-Ren, & Tzeng, Gwo-Hshiung. (2014). Combining VIKOR-DANP model for glamor stock selection and stock performance improvement. *Knowledge-Based Systems, 58*, 86-97.

Simon, Dan. (2013). *Evolutionary optimization algorithms*: John Wiley & Sons.

Sonsino, Doron, & Shavit, Tal. (2014). Return prediction and stock selection from unidentified historical data. *Quantitative Finance, 14*(4), 641-655.

Sun, Chia Chi. (2014). Combining grey relation analysis and entropy model for evaluating the operational performance: an empirical study. *Quality & Quantity, 48*(3), 1589-1600.

Sun, Y., Aw, G., Teo, K. L., Zhu, Y., & Wang, X. (2016). Multi-period portfolio optimization under probabilistic risk measure. *Finance Research Letters.*

Xiao, Z., Xia, S., Gong, K., & Li, D. (2012). The trapezoidal fuzzy soft set and its application in MCDM. *Applied Mathematical Modelling, 36*(12), 5844-5855.

Tzeng, Gwo-Hshiung, & Huang, Jih-Jeng. (2011). *Multiple attribute decision making: methods and applications*: CRC Press.

Upadhyay, P, Kar, R, Mandal, D, & Ghoshal, SP. (2014). A new design method based on firefly algorithm for IIR system identification problem. *Journal of King Saud University-Engineering Sciences*.

Yu, Huanhuan, Chen, Rongda, & Zhang, Guoping. (2014). A SVM stock selection model within PCA. *Procedia Computer Science, 31*, 406-412.

Zhang, Xiangzhou, Hu, Yong, Xie, Kang, Wang, Shouyang, Ngai, EWT, & Liu, Mei. (2014). A causal feature selection algorithm for stock prediction modeling. *Neurocomputing, 142*, 48-59.

Zhou, Xiaocong, Nakajima, Jouchi, & West, Mike. (2014). Bayesian forecasting and portfolio decisions using dynamic dependent sparse factor models. *International Journal of Forecasting, 30*(4), 963-980.

This article can be cited: Faezy Razi, F. & shadloo, N. (2017). A Hybrid Grey-Based Two-Step Clustering and Firefly Algorithm for Portfolio Selection. *Journal of Optimization in Industrial Engineering.*10 (22), 49-59.

URL: http://qjie.ir/article_276.html

DOI: 10.22094/joie.2017.276