

A hybrid grey based K-means and feature selection for bank evaluation

Mohammad Emami and Farshad Faezy Razi*

Department of Industrial Management, Semnan Branch, Islamic Azad University, Semnan, Iran

CHRONICLE

Article history:

Received October 15, 2013
 Received in revised format
 March 2 2014
 Accepted May 1, 2014
 Available online
 May 7 2014

Keywords:

Banking industry
 K-means
 Ranking

ABSTRACT

Performance measurement plays essential role on improving the performance of business units and their efficiencies. During the past few years, there have been tremendous development in banking systems and the primary focus of many managers is to improve the quality of services for market retention. Performance measurement in banking industry is normally involved with various qualitative as well as quantitative criteria, which leads to the implementation of multiple criteria decision making techniques. This paper presents a hybrid grey relational analysis and K-means to cluster and measure the performance of banking system. The proposed study uses different criteria, clusters banks into various segments and ranks 43 different banks in city of Semnan, Iran.

© 2014 Growing Science Ltd. All rights reserved.

1. Introduction

Performance measurement plays essential role on improving business units' performance and their efficiencies. During the past few years, there have been tremendous development in banking systems and the primary focus is to improve the quality of services as an objective for market retention. Performance measurement in banking industry is normally involved with various qualitative as well as quantitative criteria, which leads to the implementation of multiple criteria decision making techniques. Data mining is the result of applying sophisticated modeling techniques from the diverse fields of statistics, artificial intelligence, and database management (Yuantao & Siqin, 2008; Han & Kambert, 2001). Data mining has been widely used to determine marketing trend (Kaefer et al., 2005), customer detection (Kim & Nick Street, 2004), fraud detection (Farvares & Sepehri, 2010), etc.

Today, the ability to detect the profitable customers, building a long-term loyalty in them and expanding the existing relationships is the primary key and competitive factors for a customer-oriented organization. The prerequisite for having such competitive factors is the existence of a very

* Corresponding author.

E-mail addresses: farshadfaezy@yahoo.com (F. Faezy Razi)

powerful customer relationship management (CRM). The precise evaluation of customers' profitability is one of the most important reasons that lead to a successful CRM programs. RFM is a technique, which scrutinizes three properties, namely recency, frequency and monetary for each customer and scores customers based on these properties. Zalaghi and Abbasnejad Varzi (2014) presented a method, which obtains the behavioral traits of customers using the extended RFM approach and having the information associated with the customers of a firm. It then classifies the customers based on K-means algorithm and finally scores the customers in terms of their loyalty in each cluster. In their method, first the customers' records are clustered and then the RFM model items are specified through selecting the effective properties on the customers' loyalty rate based on the multipurpose genetic algorithm. Next, they are scored in each cluster based on the effect that they have on the loyalty rate.

2. The proposed study

2.1 K-means clustering

K-means clustering is a popular data mining clustering method, which aims to partition N observations into K clusters in which each observation belongs to the cluster with the nearest mean. Normal assessment of a proper K is accomplished by minimizing the inner-cluster variation and maximizing the among-cluster variation, simultaneously. K-means clustering is normally sensitive to outliers, so, outliers must be removed before completing clustering (Ying & Feng, 2008; Cheng & Chen, 2008; Farvaresh & Sepehri, 2010). According to Edwards (2003) and Kantardzic (2011), the K-means method used in this paper has the following steps,

1. Choose a primary part of K categories including samples that were randomly selected and calculate the mean of each pair,
2. Create a new section of each part by determining the nearest center core,
3. Calculate the new batches as the main centers,
4. Repeat step 2 and step 3 until the algorithm reaches termination criteria.

2.2. Grey Relational Analysis

Grey relation analysis proposed in this paper has the following steps (Deng, 1989; Hsia et al., 2004; Huang et al., 2008; Razi et al., 2013):

Consider X_0 as reference and N alternatives with k criteria as follows,

$$\begin{aligned}
 X_0 &= \{X_0(1), X_0(2), \dots, X_0(j), \dots, X_0(k)\} \\
 X_1 &= \{X_1(1), X_1(2), \dots, X_1(j), \dots, X_1(k)\} \\
 &\vdots \\
 &\vdots \\
 X_i &= \{X_i(1), X_i(2), \dots, X_i(j), \dots, X_i(k)\} \\
 &\vdots \\
 &\vdots \\
 X_N &= \{X_N(1), X_N(2), \dots, X_N(j), \dots, X_N(k)\}
 \end{aligned} \tag{1}$$

Grey relational coefficient are calculated as follows,

$$\gamma_{0i} = \frac{\Delta \min + \xi \Delta \max}{\Delta X_{0i} + \xi \Delta \max}, \tag{2}$$

where ΔX_{0i} is the absolute difference between X_0 and X_i in k th criterion, $\Delta X_{0i} = |X_0(k) - X_i(k)|$. In addition, $\Delta \max = \max_i \max_j \Delta X_{0i}$ and $\Delta \min = \min_i \min_k \Delta X_{0i}$. Finally, grey relational degree is calculated as follows,

$$\Gamma_{0i} = \sum_{j=1}^k w_j \gamma_{0i}, \tag{3}$$

where w_j is the weight of criterion j and we may use $W_j = \frac{1}{k}$. Finally, all relationships must be normalized as follows,

$$x_i^*(j) = \frac{x_i(j) - \min_j x_i(j)}{\max_j x_i(j) - \min_j x_i(j)}, \tag{4}$$

$$x_i^*(j) = \frac{\max_j x_i(j) - x_i(j)}{\max_j x_i(j) - \min_j x_i(j)}. \tag{5}$$

Grey relational analysis has been widely used in various industries. Gupta and Kumar (2013), for instance, presented optimization of performance characteristics in unidirectional glass fiber reinforced plastic composites using Taguchi method and Grey relational analysis. Performance characteristics such as surface roughness and material removal rate in this paper were optimized during rough cutting operation. Salardini (2013) applied AHP and grey relational analysis to offer a method for portfolio management. They used a statistical sample consists of 16 firms whose shares were trading during the fiscal year of 2010 on Tehran Stock Exchange and used AHP and grey relational analysis to assign weight to each firm.

The proposed study of this paper uses a hybrid of Grey relational analysis as well as K-means for clustering 43 banks in city of Semnan, Iran based on 24 criteria.

3. The results

In this section, we present details of our findings on clustering 43 banks based on 24 different criteria defined in Table 1 as follows,

Table 1
The criteria used for clustering banks

Item	Description	Item	Description
1	Number of saving accounts (Type 1)	13	Number of cheques cached by banks
2	Number of saving accounts (Type 2)	14	Other types of loans excluded read state
3	Number of short term investment account	15	All Islamic contracts
4	Number of transactions between banks	16	The number of returned checks issued by a branch
5	Number of active point of sales	17	Short term investments
6	The amount of investments	18	Temporary creditors
7	Amount of investments on decentralized systems	19	Number cheques sold to customers
8	Long term investments	20	Other liabilities
9	Bills paid by system	21	Number of ATM card issues
10	Interest free loans excluding real-state	22	Other investments
11	Interest based loans	23	Other loans given to customers
12	Loans paid for real-state	24	Number of long term accounts created

As we can observe from the results of Table 1, there are relatively large numbers of criteria and we use feature selection to reduce the number of criteria from 24 to 15. Table 2 shows the input data for the reduced numbers of criteria.

Table 2
The input data

Branch	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
1	-0.47	-0.70	-0.80	-0.42	-0.41	-0.29	-1.25	-0.64	-1.04	-1.27	-1.07	-1.00	-1.24	-1.06	-0.95
2	-0.34	2.04	-0.74	1.03	-0.33	-0.27	0.50	-0.31	0.23	1.19	-0.49	-0.63	-0.54	0.52	0.39
3	-0.15	0.65	-0.47	0.69	-0.11	-0.23	2.07	-0.16	-0.40	0.23	1.07	0.54	-0.37	3.21	1.31
4	-0.06	1.74	1.44	0.30	0.51	-0.05	1.44	0.51	1.59	2.19	2.13	0.83	1.72	0.83	1.03
5	-0.11	0.09	0.04	-0.25	0.11	-0.11	-0.42	-0.20	-0.02	0.18	0.35	-0.31	-0.07	-0.29	-0.55
6	-0.14	-0.39	0.24	-0.34	-0.20	-0.24	-0.32	-0.44	-0.51	-0.36	-0.31	-0.68	-0.28	-0.35	0.04
7	-0.31	-0.57	-0.62	-0.33	-0.18	-0.26	-0.68	-0.38	-0.43	-0.63	0.33	-0.49	-0.66	-0.61	-0.27
8	-0.43	-0.62	-0.69	-0.34	-0.36	-0.28	-0.93	-0.51	-0.63	-0.63	-0.56	-0.60	-0.76	-0.82	-0.54
9	-0.32	-0.43	-0.71	0.37	0.06	-0.20	-0.02	-0.11	-0.24	0.26	2.00	0.04	-0.30	-0.18	0.02
10	0.30	-0.40	-0.45	-0.40	-0.30	-0.26	0.18	-0.48	-0.74	-0.43	-0.42	0.13	-0.09	0.70	-0.08
11	-0.34	-0.41	-0.56	-0.38	-0.30	-0.14	-0.68	-0.39	-0.30	-0.45	-0.46	-0.13	-0.53	-0.78	0.24
12	6.17	5.10	2.92	5.88	0.54	6.42	2.64	5.28	4.29	3.01	3.63	3.40	3.38	2.19	1.86
13	-0.27	-0.45	0.29	-0.40	-0.29	-0.18	-0.54	-0.23	0.03	-0.66	-0.21	-0.17	0.35	-0.25	-0.41
14	-0.19	-0.09	0.93	0.44	-0.24	0.17	-0.02	0.00	-0.12	-0.37	-0.10	0.60	0.17	0.27	-0.13
15	-0.06	0.02	2.26	-0.39	-0.17	-0.13	-0.40	0.03	-0.26	-0.35	-0.58	0.20	0.12	-0.41	-0.32
16	-0.20	-0.21	0.09	-0.40	0.52	-0.16	-0.22	0.05	0.35	-0.02	0.29	0.10	-0.14	-0.29	-0.31
17	-0.32	-0.22	0.21	-0.38	-0.22	-0.14	0.06	0.01	0.23	-0.32	-0.01	1.12	0.28	0.03	0.36
18	-0.21	0.24	0.04	0.71	0.97	-0.14	-0.38	-0.16	0.30	-0.39	-0.24	0.05	0.16	-0.44	-0.67
19	-0.09	0.38	2.14	-0.36	-0.05	-0.06	0.43	0.58	1.58	0.33	0.11	0.27	0.30	0.23	-0.01
20	0.59	0.63	2.47	1.18	-0.32	0.09	0.96	0.16	1.25	0.19	-0.03	-0.05	-0.11	-0.10	2.26
21	-0.35	-0.33	-0.42	-0.33	-0.14	-0.22	-0.83	-0.37	-0.22	-0.70	-0.53	-0.34	-0.26	-0.66	-0.47
22	-0.36	-0.54	0.26	-0.36	-0.34	-0.18	-0.46	-0.45	-0.74	-0.91	-0.41	-0.14	1.30	-0.64	-0.43
23	-0.22	-0.42	-0.07	-0.32	-0.30	-0.20	-0.69	-0.32	-0.38	-0.36	-0.72	-0.38	-0.29	-0.52	-0.45
24	0.17	0.39	-0.12	0.35	0.34	0.02	0.56	-0.11	0.25	0.65	0.85	0.13	0.15	0.08	0.39
25	-0.15	0.08	-0.42	0.34	6.17	-0.18	-0.48	-0.01	1.26	-0.34	-0.49	-0.06	-0.36	-0.51	-0.50
26	0.52	0.00	2.06	0.04	-0.31	0.39	2.40	1.41	1.59	2.23	2.41	0.98	1.83	2.23	-0.10
27	-0.33	-0.57	-0.74	-0.42	-0.36	-0.24	-0.58	-0.41	-0.59	-0.68	-0.55	-0.72	-0.71	-0.53	-0.60
28	-0.22	-0.55	-0.72	-0.42	-0.40	-0.21	-0.40	-0.47	-0.71	-0.71	-0.75	-0.86	-0.54	-0.54	0.12
29	-0.42	-0.65	-0.75	-0.39	-0.36	-0.26	-1.08	-0.56	-0.80	-0.93	-0.98	-0.82	-0.94	-0.90	-0.71
30	-0.19	-0.50	-0.75	-0.40	-0.25	-0.24	-0.65	-0.39	-0.68	0.09	-0.21	-0.70	-0.54	-0.48	-0.54
31	-0.20	-0.14	0.39	-0.37	-0.39	-0.26	0.05	-0.57	-0.70	-0.69	-0.53	-0.99	-1.05	-0.93	4.91
32	0.79	1.30	0.62	0.13	0.34	0.20	3.46	2.84	2.07	3.32	1.91	4.39	3.43	3.52	0.21
33	-0.19	-0.45	-0.69	-0.39	-0.16	-0.23	-0.50	-0.30	-0.42	-0.26	-0.47	-0.49	-0.63	-0.43	-0.52
34	-0.27	-0.48	-0.68	-0.36	-0.30	-0.22	-0.71	-0.23	-0.16	-0.35	-0.47	-0.08	-0.37	-0.33	-0.57
35	-0.32	-0.39	-0.45	-0.41	-0.18	-0.20	0.07	-0.11	-0.23	-0.27	-0.13	-0.34	0.02	-0.02	-0.52
36	-0.20	-0.57	-0.67	-0.42	-0.34	-0.22	-0.73	-0.35	-0.52	1.14	0.37	-0.45	-0.75	-0.59	-0.69
37	-0.30	-0.14	-0.73	-0.31	-0.06	-0.17	0.09	0.03	-0.05	0.02	-0.27	0.33	0.76	-0.04	-0.07
38	-0.40	-0.48	-0.15	-0.40	-0.31	-0.24	-0.23	-0.44	-0.74	-0.49	-0.62	-0.59	-0.43	-0.21	-0.20
39	-0.33	-0.41	-0.54	-0.42	-0.24	-0.16	-0.27	-0.33	-0.38	-0.18	-0.53	-0.31	0.63	-0.30	-0.25
40	0.12	-0.58	-0.74	-0.39	-0.34	-0.24	-0.63	-0.06	-0.33	-0.70	-0.56	-0.47	-0.70	-0.52	-0.36
41	-0.29	-0.45	-0.65	-0.33	-0.31	-0.02	0.17	-0.16	-0.49	-0.05	-0.71	0.18	-0.36	0.58	-0.47
42	0.83	0.60	-0.60	0.09	-0.30	-0.26	-0.16	-0.40	-0.63	-0.44	-0.47	-0.38	-0.85	-0.10	-0.27
43	-0.31	-0.58	-0.67	-0.30	-0.35	-0.07	-0.18	-0.37	-0.75	-0.28	-0.66	-0.43	0.27	-0.16	-0.50

All computations have been accomplished on Clementine®12 and the results of clustering are summarized in Fig. 1. In order to have an efficient clustering, we also calculated the average silhouette coefficients for various clusters and Table 3 demonstrates the results of our survey.

Table 3
The summary of the average silhouette coefficient

Cluster	2	3	4	5	6	7	8	9	10
Rank	0.3124	0.4067	0.2466	0.1979	0.2158	0.2704	0.1384	0.3085	0.3208

As we can observe from the results of Table 3, the highest value belongs to third cluster and based on this cluster, we rank different banks and the results are summarized in Table 4-6 as follows,

Table 4
The results of ranking different banks

Branch	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Cluste	1	1	3	3	1	1	1	1	3	1	1	2	1	1	1
Rank	0.41	0.57	0.59	0.73	0.48	0.46	0.45	0.43	0.46	0.48	0.45	0.82	0.47	0.52	0.52

Table 5

The results of ranking different banks

Branch	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Cluster	1	1	1	3	3	1	1	1	3	1	3	1	1	1	1
Rank	0.497	0.517	0.509	0.538	0.621	0.452	0.480	0.455	0.489	0.647	0.820	0.436	0.443	0.423	0.448

Table 6

The results of ranking different banks

Branch	42	43	31	32	33	34	35	36	37	38	39	40	41	42	43
Cluster	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1
Rank	0.486	0.457	0.567	0.588	0.449	0.454	0.468	0.470	0.496	0.453	0.470	0.447	0.475	0.486	0.457

As we can observe from the results of ranking, out of 43 banks, 34 has been located in the first cluster while the second cluster only includes two banks.

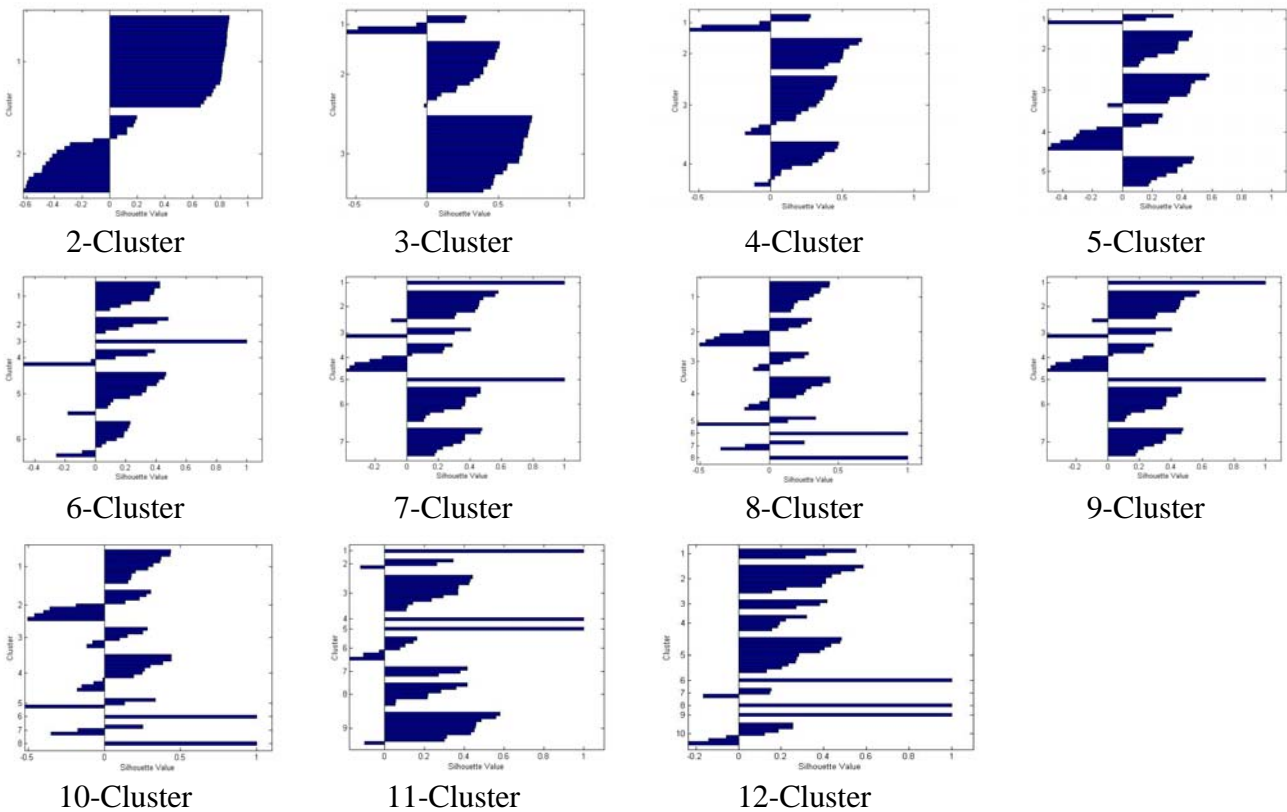


Fig. 1. The results of clustering

4. Conclusion

In this paper, we have investigated the relative efficiencies of banks in one of Iranian cities called Semnan. The proposed study has applied K-means clustering for ranking various banks based on 15 criteria. The results of ranking can be compared with some other alternative performance measurement methods and we leave it for interested researchers as future studies.

Acknowledgement

The authors would like to thank the anonymous referees for constructive comments on earlier version of this paper.

References

- Cheng, C., & Chen, Y. (2008). Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications*, 4176–4184.
- Deng, J. L. (1989). Introduction to grey system theory. *The Journal of Grey System*, 1(1), 1-24.
- Edwards, D. (2003). Data mining: Concepts, models, methods, and algorithms. *Journal of Proteome Research*, 2(3), 334-334.
- Farvareh, H., & Sepehri, M. M. (2010). A data mining framework for detecting subscription fraud in telecommunication. *Engineering Applications of Artificial Intelligence*, 24(1), 182-194.
- Gupta, M., & Kumar, S. (2013). Multi-objective optimization of cutting parameters in turning using grey relational analysis. *International Journal of Industrial Engineering Computations*, 4, 547-558.
- Han, J., & Kambert, M. (2001). *Data Mining: Concepts and Techniques*, Morgan Kaufmann. San Francisco.
- Hsia, K. H., Chen, M. Y., & Chang, M. C. (2004). Comments on data pre-processing for grey relational analysis. *Journal of Grey System*, 7(1), 15-20.
- Huang, S. J., Chiu, N. H., & Chen, L. W. (2008). Integration of the grey relational analysis with genetic algorithm for software effort estimation. *European Journal of Operational Research*, 188(3), 898-909.
- Kaefer, F., Heilman C.M., & Ramenofsky, S. D. (2005). A neural network application to consumer classification to improve the timing of direct marketing activities. *Computer & Operations Research*, 32(10), 2595–2615.
- Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons.
- Kim, Y. & Nick Street, W. (2004). An intelligent system for customer targeting: A data mining approach. *Decision Support Systems*, 37, 215–228.
- Razi, F. F., Danaei, A., Ehsani, A., & Dolati, C. (2013). Selecting an optimal mixed products using grey relationship model. *Management Science Letters*, 3(6), 1825-1832.
- Salardini, F. (2013). An AHP-GRA method for asset allocation: A case study of investment firms on Tehran Stock Exchange. *Decision Science Letters*, 2, 275-280.
- Yuantao, J. & Siqin, Y. (2008). Mining e-commerce data to analyze the target customer behavior. *First International Workshop on Knowledge Discovery and Data Mining*, 406-409.
- Ying, L. & Feng, L. (2008). Customer segmentation analysis based on SOM clustering. *IEEE International Conference on Service Operations and Logistics, and Informatics*, 1, 15-19.
- Zalaghi, Z., & Abbasnejad Varzi, Y. (2014). Measuring customer loyalty using an extended RFM and clustering technique. *Management Science Letters*, 4, 905-912.