Management Science Letters 4 (2014) 2319–2324

Contents lists available at GrowingScience

Management Science Letters

homepage: www.GrowingScience.com/msl

Ranking banks using K-Means and Grey relational method

Seyed Abdoreza Tahami Pour Zarandi[a], Vahid Shahabi[b*], Mohammad Emami[b], Atefeh Rezaee[c] and Mehdi Hassanloo[c]

[a]*Deputy Manager, Bank Qarzol Hasane Resalat, Tehran, Iran*
[b]*Department of Management, Semnan Branch, Islamic Azad University, Semnan, Iran*
[c]*Department of Management, Science and Research Branch, Islamic Azad University, Semnan, Iran*

C H R O N I C L E

Article history:
Received January 20, 2014
Accepted 30 August 2014
Available online
September 1 2014

Clustering
K-Means
Grey relational
Banks
Ranking

A B S T R A C T

Research has been done on evaluating the performance more emphasis either on the technical aspects of the performance evaluation or considers only one aspect of banks such as sales, revenue, etc. This paper presents an empirical investigation to measure the relative efficiency of different banks using a hybrid of K-means and Grey relational analysis. The study uses eight criteria such as resources, customer satisfaction, etc. for clustering different banks and using Grey relational analysis provides ranking of different branches of a bank named Bank Qarzol Hasane Resalat located in city of Tehran, Iran. The preliminary study has indicated that the proposed model of this paper may capable of providing some promising results.

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 powerful customer relationship management (CRM). The precise evaluation of customers'

*Corresponding author.
E-mail addresses: Vahidshahabi1366@yahoo.com (V. Shahabi)

doi: 10.5267/j.msl.2014.9.002

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. Emami and Faezy Razi (2014) presented a hybrid grey based K-means and feature selection for bank evaluation. They presented a hybrid grey relational analysis and K-means to cluster and measure the performance of banking system and applied their method for selected banks in city of Semnan, Iran. The proposed study of this paper tries to re-examine the implementation of this method for a newly established bank in city of Tehran, Iran.

2. The proposed study

2.1 K-means clustering

The K-means method has become an effective in generating good clustering results for many real-world case studies (Aggarwal et al., 2003; Jain & Dubes, 1988; Jain et al., 1999; Han & Kambert, 2001). K-means clustering is a well-known data mining clustering technique, which attempts to partition N observations into K clusters where each observation is assigned to the cluster with the closest mean. Normal evaluation of an appropriate K is executed by minimizing the inner-cluster variation and maximizing the among-cluster variation, concurrently. K-means clustering is often sensitive to outliers, therefore, outliers have to be removed before accomplishing clustering (Ying & Feng, 2008; Cheng & Chen, 2008; Farvaresh & Sepehri, 2010). Edwards (2003) and Kantardzic (2011) described the K-means method with the following steps,

1. Select a primary part of K categories including samples randomly selected and compute the mean of each pair,
2. Generate a new section of each part by computing the nearest center core,
3. Measure the new batches as the main centers,
4. Repeat step 2 and step 3 until the algorithm approaches termination criteria.

2.2. Grey Relational Analysis

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

Let X_0 be the reference and N alternatives with k criteria as follows,

$$
\begin{aligned}
&X_0 = \{X_0(1), X_0(2), \ldots\ldots., X_0(j), \ldots.X_0(k)\}\\
&X_1 = \{X_1(1), X_1(2), \ldots\ldots.., X_1(j), \ldots.X_1(k)\}\\
&.\\
&.\\
&X_i = \{X_0 = \{X_0(1), X_0(2), \ldots\ldots., X_0(j), \ldots.X_0(k)\} \qquad (1)\\
&.\\
&.\\
&X_N = \{X_i(1), X_i(2), \ldots\ldots.., X_i(j), \ldots.X_i(k)\}
\end{aligned}
$$

Grey relational coefficient are computed 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, $\Delta X_{0i} = |X_0(k) - X_i(k)|$. Besides, $\Delta \max = \max_i \max_j \Delta X_{0i}$ and $\Delta \min = \min_i \min_k \Delta X_{0i}$. The computations of Grey relational degree is 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 apply $W_j = \frac{1}{k}$. Finally, all relationships have to 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 applied in different industries. Gupta and Kumar (2013), for example, used optimization of performance characteristics in unidirectional glass fiber reinforced plastic composites by applying Taguchi method and Grey relational analysis. Performance characteristics such as surface roughness and material removal rate in this paper were optimized using rough cutting operation. Salardini (2013) implemented analytical hierarchy process and grey relational analysis to offer a method for portfolio management. They implemented a statistical sample with 16 firms and applied analytical hierarchy process as well as gray relational analysis to assign weight to each firm.

The proposed study of this paper applies a hybrid of Grey relational analysis as well as K-Means for clustering 26 banks in city of Tehran, Iran based on different criteria.

3. The results

In this section, we present some details of our results on clustering 26 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) for individuals	6	Customer satisfaction
2	Number of saving accounts (Type 2) for individuals	7	Expenditure
3	Number of saving accounts (Type 1) for firms	8	Deferred
4	Number of saving accounts (Type 2) for firms	9	High ratio deposits
5	The amount of resources		

3.1. K-Means implementation

The implementation of K-Means has been accomplished on Clementine®12 yields two clusters shown in Table 2 and Table 3 as follows,

Table 2
The summary of standard data for the first cluster

	S1	S2	S3	S4	S5	S6	S7	S8	S9
P1	1.38	0.0654	0.843	0.3024	1.951	1.641	0.3418	2.3585	1.6163
P2	0.502	0.1309	1.7448	0.3024	1.8703	-0.6634	3.4166	-0.635	0.2586
P3	0.224	-0.1459	2.5636	1.6328	1.6575	0.1047	0.0094	1.3607	1.7234
P4	0.502	-0.4335	0.3072	-1.0281	1.3808	3.5614	1.0482	-0.635	2.0981
P5	2.396	-0.6943	-0.3727	0.3024	0.9715	-0.2793	-0.9878	-0.635	-0.0918
P6	0.039	0.157	0.8524	-1.0281	0.6525	0.104	1.0898	-0.635	-0.0577
P7	2.073	-0.7633	-0.5987	-1.0281	0.6016	0.4888	-0.4061	-0.635	0.268
P8	0.5482	-0.1317	0.0218	1.6328	0.2374	0.4888	-0.323	2.3585	0.1369
P9	0.317	-0.7483	-0.4667	0.3024	-0.0066	-0.2793	0.3834	0.3629	0.1321
P10	-0.7457	-0.7527	-0.9091	0.3024	-0.2944	0.1047	-0.323	0.3629	0.127
P11	-0.0987	-0.5578	-0.4805	-1.0281	-0.3206	-0.6634	-0.9047	-0.635	-0.7342
P12	-0.3298	-0.6926	-0.9425	0.302	-0.5195	-0.6634	-0.3645	0.3629	-0.2475
P13	-0.7919	-0.7633	-0.9929	1.6328	-0.5869	-0.2793	0.0094	-0.635	-0.2865
P14	-1.0691	0.3913	-0.2706	-1.0281	-0.7948	-0.2793	-0.6969	-0.635	1.8986
P15	0.4096	0.5113	-0.1194	0.3024	-0.7977	0.8729	-0.6138	-0.635	-0.812
P16	-0.8381	1.0805	0.2415	-1.0281	-0.8052	-0.6634	-0.8631	-0.635	-0.778
P17	-0.5146	-0.0402	-0.394	-1.0281	-0.8064	-0.6634	-0.1983	0.3629	-0.194
P18	-1.207	1.3085	0.3723	1.6328	-0.8346	-0.6634	-0.6554	1.3607	-0.924
P19	-0.56-	0.0415	-0.5828	0.3212	-0.8346	-0.6634	0.9651	-0.635	-0.8704
P20	1.0229	3.5637	1.5901	0.3024	-0.8449	-0.6634	-0.4665	-0.635	-1.25
P21	-1.3002	-0.7633	-1.1898	-1.0281	-0.8485	-0.6634	-1.0709	-0.635	-0.8996
P22	0.0861	-0.7633-	-1.2173	-1.0281	-1.028	-0.2793	-0.323	-0.635	-1.1137

Table 3
The summary of standard data for the second cluster

	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**
P1	-0.6473	-1.1619	-1.1619	0.866	0.9083	-1.1619	-0.5179	-1.1619	0.2284
P2	-0.4221	0.3873	0.3873	0.866	0.7823	-0.3873	1.1098	-0.3873	1.059
P3	1.4915	1.1619	1.1619	-0.866	-0.5865	1.1619	0.5179	1.1619	-1.3499
P4	-0.4221	-0.3873	-0.3873	-0.866	-1.1041	0.3873	-1.1098	-0.3873	0.0623

3.2. GRA results

The implementation of grey relational analysis has been performed on two clustering data and Table 4 and Table 5 shows the summary of ranking.

Table 4
The summary of ranking for the first cluster

Branch	1	2	3	4	5	6	7	8	9	10	11
GRA	0.8175	0.5044	0.7651	0.7368	0.5971	0.4669	0.5332	0.7235	0.4834	0.4873	0.4842
Branch	12	13	14	15	16	17	18	19	20	21	22
GRA	0.4740	0.4808	0.5434	0.5426	0.4886	0.4583	0.5961	0.3943	0.4517	0.4270	0.4101

Table 5
The summary of ranking for the second cluster

Branch	23	24	25	26
GRA	0.2318	0.7087	1.0271	0.6859

4. Conclusion

In this paper, we have measured the relative efficiencies of 26 branches of Bank Qarzol Hasane Resalat in city of Tehran, Iran. The proposed study has applied K-means clustering along with gray relational analysis for ranking various branches of Bank Qarzol Hasane Resalat. The implementation of K-Means clustering method has yielded two clusters and they were ranked using gray relational method. The preliminary results of this survey have indicated that the proposed study of this paper may provide a quick approach for ranking various similar units.

Acknowledgement

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

References

Aggarwal, C. C., Han, J., Wang, J., & Yu, P. S. (2003, September). A framework for clustering evolving data streams. In *Proceedings of the 29th international conference on Very large data bases-Volume 29* (pp. 81-92). VLDB Endowment.

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.

Dubes, R. C., & Jain, A. K. (1988). Algorithms for clustering data.

Edwards, D. (2003). Data mining: Concepts, models, methods, and algorithms. *Journal of Proteome Research*, *2*(3), 334-334.

Farvaresh, H., & Sepehri, M. M. (2010). A data mining framework for detecting subscription fraud in telecommunication. *Engineering Applications of Artificial Intelligence*, 24(1), 182-194.

Emami, E., & Faezy Razi, F. (2014). A hybrid grey based K-means and feature selection for bank evaluation. Decision Science Letters, 3(3), 269-274.

Guha, S., Mishra, N., Motwani, R., & O'Callaghan, L. (2000). Clustering data streams. In *Foundations of computer science, 2000. proceedings. 41st annual symposium on* (pp. 359-366). IEEE.

Guha, S., Meyerson, A., Mishra, N., Motwani, R., & O'Callaghan, L. (2003). Clustering data streams: Theory and practice. *Knowledge and Data Engineering, IEEE Transactions on*, *15*(3), 515-528.

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.

Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall, Inc.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, *31*(3), 264-323.

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.