Fuzzified FCM for Mining Sales Data and Establishing Flexible Customer Clusters

Toly Chen and Yu-Cheng Wang

Department of Industrial Engineering and Systems Management
tolychen@ms37.hinet.net

Abstract

RFM model is an important method in customer clustering. Some past studies have proposed fuzzy RFMs model to overcome the shortcomings of traditional RFM models. However, there are some problems unsolved in these approaches. To deal with these problems and to enhance the flexibility in customer clustering, the traditional fuzzy c-means (FCM) method is fuzzified to deal with R, F, and M scores that are expressed in fuzzy values. A fuzzified RFM model is then established by incorporating the fuzzified FCM approach, which is based on the inherent structure of the data itself. The number of customer clusters can be arbitrarily specified in advance, considering the scarcity of marketing resources and the diversification of marketing strategies. Besides, exploring the content of each customer cluster provides the business with many meaningful suggestions that could be usefully employed to establish target marketing programs. An example is adopted to demonstrate the application of the proposed methodology and to make some comparisons.

Keywords: *RFM model, Fuzzy c-means, Customer clustering*

1. Introduction

In the field of customer clustering, the RFM model is a very famous approach, which classifies customers according to the different buying power and magnitude of each customer [1-9]. Typical RFM models include the models of Stone [8], Miglautsch [7], Sung [9], etc. Chiu and Su [6] summarized some shortcomings of these RFM models as follows:

(1) The way of dividing a dimension (R, F, or M) into equal sections might not be suitable for all products.

(2) Scoring is too subjective and inflexible.

(3) It is difficult to explain the contents of all customer clusters, or to discriminate customers with the same RFM total scores.

Trying to overcome these shortcomings, Chiu and Su proposed a fuzzy RFM model. However, there are some problems still unsolved in Chiu and Su's approach: the number of customer clusters cannot be specified in advance; the inherent structure of customer data is not taken into account in clustering; customer importance is still evaluated with a crisp-valued RFM score. To deal with these problems, a fuzzified RFM model is proposed in this study by incorporating the fuzzified fuzzy c-means (FFCM) approach, which is based on the inherent structure of the data itself. The number of customer clusters can also be arbitrarily specified in advance, considering the scarcity of marketing resources and the diversification of marketing strategies. Besides, exploring the content of each customer cluster provides the business with many meaningful suggestions that can be usefully employed to establish target marketing programs. It is also possible to consider the case in which the R, F, and M scores are fuzzy values.

2. Methodology

FCM is an extension of the classic k-means approach using the concepts of fuzzy logic [3-5]. The FCM approach performs clustering by minimizing the following objective function:

$$\text{Min } \sum_{k=1}^{K}\sum_{i=1}^{n}\mu_{i(k)}^{m}e_{i(k)}^{2}, \tag{1}$$

where *n* is the number of entities; *K* is the required number of clusters; $\mu_{i(k)}$ represents the membership that entity *i* belongs to cluster *k*; $e_{i(k)}$ measures the distance from entity *i* to the centroid of cluster *k*; $m \in (1, \infty)$ is a parameter to increase or decrease the fuzziness. Higher values of the fuzziness will make the result fuzzier, i.e. it becomes more difficult to classify an entity into a cluster absolutely. For normal data an $m = 2.0$ can be used most of the times.

The procedure of applying the FCM approach is as follows:

(1) Determine an initial clustering result arbitrarily.

(2) (Iteration) Calculate the centroid of each cluster as:

$$\bar{x}_{(k)} = \{\bar{x}_{(k)j}\}, \tag{2}$$

$$\bar{x}_{(k)j} = \frac{\sum_{i=1}^{n}\mu_{i(k)}^{m}x_{ij}}{\sum_{i=1}^{n}\mu_{i(k)}^{m}}, \tag{3}$$

where $\bar{x}_{(k)}$ is the centroid of cluster *k*. The formula of calculating the membership $\mu_{i(k)}$ is:

$$\mu_{i(k)} = \frac{1}{\sum_{l=1}^{K}\left(\frac{e_{i(k)}}{e_{i(l)}}\right)^{\frac{2}{m-1}}}, \tag{4}$$

and the distance to the cluster centroid $e_{i(k)}$ is calculated as:

$$e_{i(k)} = \sqrt{\sum_{all\ j}(x_{ij} - \bar{x}_{(k)j})^{2}}\ . \tag{5}$$

(3) Re-measure the distance of each entity to the centroid of every cluster, and then recalculate the corresponding membership.

(4) If the following condition is satisfied, stop; if else, return to step (200).

$$\max_{k}\max_{i}\left|\mu_{i(k)}^{(t)} - \mu_{i(k)}^{(t-1)}\right| < d \tag{6}$$

where $\mu_{i(k)}^{(t)}$ is the membership of entity *i* belonging to cluster *k* after the *t*-th iteration; *d* is a real number representing the threshold of membership convergence.

The FCM approach has been universally applied to data clustering. Chiang, et. al., [2] applied the FCM approach to classify airline customers into three groups, according to the personal factors affecting these customers' choices of airlines. Then services with different contents can be provided for different customer groups to satisfy their requirements.

As to the determination of the best number of clusters, from the viewpoint of data clustering, Xie & Beni [10] proposed the separate distance test (*S* test) that can be applied for this purpose [11-15]:

$$\text{Min } S \tag{7}$$

s.t.

$$J_m = \sum_{k=1}^{K} \sum_{i=1}^{n} \mu_{i(k)}^m e_{i(k)}^2, \tag{8}$$

$$e_{\min}^2 = \min_{p \neq q} \left(\sum_{all\ j} (\bar{x}_{(p)j} - \bar{x}_{(q)j})^2 \right), \tag{9}$$

$$S = \frac{J_m}{n \times e_{\min}^2}, \tag{10}$$

$$K = 1, 2, \ldots \tag{11}$$

The *K* value when *S* is minimized determines the best number of clusters. If x_{ij} is a fuzzy value such as the fuzzy R, F, or M score, then equation (4) becomes

$$\tilde{x}_{(k)} = \{\bar{\tilde{x}}_{(k)j}\} \tag{12}$$

The centroid of each cluster is also a fuzzy value. The distance from $\tilde{x}_{ij}$ to $\tilde{x}_{(k)}$ is

$$\tilde{e}_{i(k)} = \sqrt{\sum_{all\ j} (\tilde{x}_{ij} - \bar{\tilde{x}}_{(k)j})^2}, \tag{13}$$

The membership that entity *i* belongs to cluster *k* is

$$\tilde{\mu}_{i(k)} = \frac{1}{\sum_{l=1}^{K} \left(\frac{\tilde{e}_{i(k)}}{\tilde{e}_{i(l)}} \right)^{\frac{2}{m-1}}}, \tag{14}$$

It is also uncertain, which gives us more flexibility in clustering customers. For example, when the economy is good, the upper bound of the membership, $\mu_{i(k)3}$, can be adopted, and each category will contains more customers. Conversely, when the economy is not good, we should reduce marketing expenses. Adopting the lower bound of the membership, $\mu_{i(k)1}$, can effectively reduce the number of customers in each category. To solve this problem, the concept of α-cut operations is used, and we derive $\tilde{\mu}_{i(k)}$ by repeating FCM a sufficient number of times until the following results converge:

$$\tilde{\mu}_{i(k)} = (\min_t \mu_{i(k)1}(t), \underset{t}{\mu_{i(k)2}}, \max_t \mu_{i(k)3}(t)) \tag{15}$$

where *t* indicates the *t*-th replication of FCM.

3. A Demonstrative Example

In this section, the example in Chiu and Su [6] is taken to demonstrate the application of the RBF-FCM approach. First, the data are normalized, and then we apply MATLAB 6.0 to implement FCM to classify the customers into three clusters. After 63 iterations of improvement, the minimal objective function value is 0.704, and the clustering result is shown in Table 1. Take customer #7 as an example, the membership of belonging to cluster #2 is 0.55, while that to cluster #1 is much lower (0.26). As a result, if we classify customers according to the highest values of membership, then customer customers #7 and #9 should be assigned to the same cluster (cluster #2), while customers #2, #4, and #10 belong to another cluster (cluster #1), and customers #1, #3, #5, #6, and #8 are members of the third cluster.

Table 1. The Classification Result Using FCM (*K* = 3)

Customer #	Membership□(□)			Classification result (□ ≥ 0.5)
	Cluster #1	Cluster #2	Cluster #3	
1	0.03	0.01	0.96	3
2	0.60	0.13	0.27	1
3	0.21	0.07	0.72	3
4	0.96	0.01	0.03	1
5	0.18	0.04	0.78	3
6	0.15	0.07	0.77	3
7	0.26	0.55	0.19	2
8	0.16	0.09	0.74	3
9	0.07	0.89	0.04	2
10	0.78	0.12	0.10	1

In this case, S is minimized (with a minimum of 0.255) when K is 4. As a result, the best number of customer clusters for the demonstrative example is four.

Subsequently, if the R, F, and M scores are fuzzy values (see Table 2), then the classifications results will be more uncertain. For example, the membership that customer #7 belongs to each cluster is:

Cluster #1: (0.07, 0.26, 0.81)

Cluster #2: (0.10, 0.55, 0.88)

Cluster #3: (0.05, 0.19, 0.57)

If the economy is good, then we should choose customers according to the upper bound of the membership value. Assuming cluster #3 is targeted and the membership threshold is 0.5, then customer #7 should also be targeted by this action since 0.57>0.5.

Table 2. The Fuzzy R, F, and M Scores

Customer #	R_1	R_2	R_3	F_1	F_2	F_3	M_1	M_2	M_3
1	7	8	10	4	5	6	3929	4500	4605
2	10	10	11	10	10	12	2446	3000	3277
3	19	19	20	3	3	3	1357	1500	1556
4	25	25	27	6	7	8	4846	6000	6725
5	4	5	5	5	6	6	2971	3500	3990
6	1	1	1	5	5	6	8270	10000	10151
7	42	51	53	4	4	5	13826	15000	15815
8	12	14	16	2	2	2	7210	8500	9427
9	31	34	38	7	9	10	17491	21500	23331
10	37	40	47	8	8	8	4642	5000	5074

4. Conclusions

RFM continues to provide the foundation for customer classification. A fuzzified RFM model is established in this study for sales data analysis and customer clustering, which is important to the formulation of marketing strategies. To enhance the flexibility in customer clustering, the traditional FCM method is fuzzified to deal with R, F, and M scores that are expressed in fuzzy values. An example is also used to illustrate the applicability of the proposed methodology and make comparison with that by the existing method. The experimental results supported the usefulness of the proposed methodology. In the future, other clustering methods can be fuzzified in similar ways for the same purpose.

Acknowledgements

This work is partially supported by the National Science Council of Taiwan.

References

[1] T. Chen and Y. C. Wang, “A study on the construction of a ubiquitous customer relationship management system with RFID – a case study”, International Journal of Information and Communication Technology, vol. 4, **(2012)**, pp. 13-26.

[2] I. W.-Y. Chiang, G. -S. Liang and S. Z. Yahalom, “The fuzzy clustering method: applications in the air transport market in Taiwan”, Database Marketing & Customer Strategy Management , vol. 11, **(2003)**, pp. 149-58.

[3] T. Chen, “Incorporating fuzzy c-means and a back-propagation network ensemble to job completion time prediction in a semiconductor fabrication factory”, Fuzzy Sets and Systems, vol. 158, **(2007)**, pp. 2153-68.

[4] M. H. Zarandi and Z. S. Razaee, “A fuzzy clustering model for fuzzy data with outliers”, International Journal of Fuzzy System Applications, vol. 1, **(2011)**, pp. 29-42.

[5] T. Chen and Y. C. Wang, “Estimating the job cycle time in wafer fabrication with distributed sensors”, International Journal of Hybrid Information Technology, vol. 5, **(2012)**, pp. 81-88.

[6] H. P. Chiu and C. Y. Su, "A flexible fuzzy RFM model for CRM and database marketing", Journal of Electronic Bu siness, vol. 6, **(2004)**, pp. 1-26.

[7] J. Miglautsch, "Thoughts on RFM scoring", Journal of Database Marketing, vol. 8, no. 1, **(2000)**, pp. 1-7.

[8] B. Stone, “Successful Direct Marketing Methods”, NTC Business Books, **(1989)**.

[9] H. H. Sung and C. P. Sang, "Application of data mining tools to hotel data mart on the Intranet for database marketing", Expert Systems with Applications, vol. 15, **(1998)**, pp. 1-31.

[10] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering", IEEE Transactions of Pattern analysis and Machine Intelligence, vol. 13, **(1991)**.

[11] S. Dehuri and S. Cho, "Learning fuzzy network using sequence bound global particle swarm optimizer", International Journal of Fuzzy System Applications, vol. 2, no. 1, **(2012)**, pp. 54-70.

[12] T. Chen, "A PCA-FBPN approach for job cycle time estimation in a wafer fabrication factory", International Journal of Fuzzy System Applications, vol. 2, no. 2, **(2012)**, pp. 50-67.
[13] P. Singamsetty and S. Panchumarthy, "Automatic fuzzy parameter selection in dynamic fuzzy voter for safety critical systems", International Journal of Fuzzy System Applications, vol. 2, no. 2, **(2012)**, pp. 68-90.
[14] T. Chen, "An intelligent mechanism for lot output time prediction and achievability evaluation in a wafer fab", Computers & Industrial Engineering, vol. 54, **(2008)**, pp. 77-94.
[15] T. Chen, "Evaluating the mid-term competitiveness of a product in a semiconductor fabrication factory with a systematic procedure", Computers & Industrial Engineering, vol. 53, **(2007)**, pp. 499-513.

Authors

Toly Chen received the B. S. degree, the M S. degree, and the Ph. D. degree in industrial engineering from National Tsin Hua University. He is now a Professor in the Department of Industrial Engineering and Systems Management of Feng Chia University.

Yu-Cheng Wang received the B.S. degree the M S. degree, in industrial engineering from Feng Chia University. And now he is a Ph. D. student in the Department of Industrial Engineering and Systems Management of Feng Chia University.