Customers Fuzzy Clustering and Catalog Segmentation In Customer Relationship Management

Conference Paper · December 2009

DOI: 10.1109/IEEM.2009.5372997

Michel Fathi
University of North Texas

Kamran Kianfar
University of Isfahan

Amir Sadeghi
Islamic Azad University, South Tehran Branch

Customers Fuzzy Clustering and Catalog Segmentation In Customer Relationship Management

Mehdi Fathi [1], Kamran Kianfar [2], Amir Hasanzadeh [3], Amir Sadeghi [4]
[1] Department of Industrial Engineering Iran University of Science and Technology, Tehran, Iran
[2] Department of Industrial Engineering ,Isfahan University of Technology, Isfahan, Iran
[3] Department of Industrial Engineering, AmirKabir University of Technology, Tehran, Iran
[4] Department of Industrial Engineering ,Tehran Payamenoor University, Tehran, Iran
(mfathi@iust.ac.ir)

Abstract - **This work is concerned with the fuzzy clustering problem of different products in *k* variant catalogs, each of size *r* products that maximize customer satisfaction level in customer relationship management (CRM). The satisfaction degree of each customer is defined as a function of his/her needed product number that exists in catalog and also his/her priority. To determine the priority level of each customer, firstly customers are divided to three clusters with high, medium and low importance based on his/her needed products list. Then, all customers have been ranked based on their membership level in each of the above three clusters. In this paper in order to cluster customers, fuzzy c-means algorithm is applied. The proposed problem is firstly modeled as a bi-objective mathematical programming model. The objective functions of the model are to maximize the number of covered customers and overall satisfaction level results of delivering service. Then, this model is changed to a single integer linear programming model by applying fuzzy theory concepts. Finally, the efficiency of the proposed solution procedure is verified by using a numerical example.**

Keywords - **Catalog Segmentation, Customer Clustering, Fuzzy C-means, Customer Relationship Management**

I. INTRODUCTION

Nowadays, one of the business and marketing organizations goals is customer attraction. Catalog design has known as the most common way in customer attraction, satisfaction, and retention cycle of customer relationship management. Meanwhile one of the challenges that companies using catalogs face is the optimization problem of available products; this means by the catalog which products are covered and for which cluster of customers is designed. The catalog optimization has an effective role in satisfying customers' requirements and company profitability.

With the fast growing number of products offered by retailers and e-retailers alike, the design of one catalog that can contain all products is not feasible. In many cases, some customers are interested in only a small portion of the products the company carries. In today's global competition, the catalog design has changed as one of the competitive tools for marketing among companies. The latest Benchmark Survey on Critical Issues and Trends conducted by Catalog Age magazine in 2003 indicates that catalog companies that participated in the survey spent a mean 26.1% of their revenue on marketing expenses and that print and postage costs account for nearly half of those expenses. The number of catalogs that companies send to customers and potential buyers is increasing at a fast pace [1].

Efficiency and profitability of a catalog is evaluated by the number of available products of interest to customers that companies send to them. The companies seek optimum decision making on the customers clusters, not optimum decision making as individual.

The classical catalog segmentation problem was introduced by Kleinberg et al.[8]. This problem consists of designing K catalogs, each of size r products that maximize the number of covered customers. In this problem, requirements of a customer are satisfied if at least the specified number of products of interest to her\him is covered in one of the catalogs.

In general, the customer-oriented catalog segmentation problem is a customer clustering problem where:

- The task is to determine K clusters of customers where each cluster is defined by a set of products of interest to the customers and each customer is assigned to the cluster (equivalent to a catalog) that contains the largest number of products of interest to him/her.
- An acceptable clustering should satisfy two constraints:
  1) Each cluster includes r products,
  2) The need of a customer can be satisfied (i.e., covered) by a catalog only the catalog contains at least k products of interest to the customer.

Amiri [2] presented a mathematical model that consists of designing K catalogs, each of size r products and maximizes the number of covered customers. In this paper, we have modeled the catalog design problem as a fuzzy mathematical programming model and efficiency of the proposed model has been verified through a numerical example.

The remainder of the paper is organized as follows. In Section 2, we give the definition and formulation of the problem. In Section 3, circumstance of the priority degree determination on each customer and in section 4 clustering algorithm of *fuzzy c-means* is described briefly. In section 5, a bi-objective mathematical programming model is proposed. This model has been changed to a single integer linear programming model by applying fuzzy theory concepts in section 6. In the following section, a numerical example has been used for better

978-1-4244-4870-8/09/$26.00 ©2009 IEEE

understanding of the problem. Finally, section 8 has been addressed to present a general conclusion of the paper.

II. Problem definition

Assume that customers and their product interests are known. The main input data for catalog segmentation is the customer interest database that contains the set of products that each customer is interested in. The product interests of a customer can be obtained either by aggregating all purchase transactions of the customer or by obtaining his/her explicit preferences from a set of products. The customer-oriented catalog segmentation problem consists of designing *K* catalogs, each of size *r* products that satisfy the defined objective function of the model at the best possible circumstance. It's assumed the need of a customer is satisfied if at least the specified minimum number (t) of the products of interest to him/her is in one of the catalogs.

In this paper, the customers fuzzy clustering and the catalog segmentation problem are modeled as a bi-objective programming model and then are solved. In this model, first objective function shows the number of customers that we want to satisfy their needs through the catalogs set and second objective function implies the marginal utility results of allocating the catalogs to the customers. In order to build second objective function, at first it's necessary that a priority degree is defined per customer.

III. Determining priority degree of each customer

In order to determine the priority degree of the customers, initially they are divided into three clusters with high, medium and low importance based on their needed products list. Because of fuzzy nature of the problem and the fact that each of the customers may partially belong to each of three clusters noted above, *fuzzy c-means* algorithm has been used for clustering the customers. More details related to this algorithm have been addressed in section 4. If m_{ij} show the membership degree of the customer i in cluster j and c_j show the importance coefficient of cluster j, the priority degree of the customer i will be calculated as the following equation:

$$p_i = \sum_{j=1}^{3} m_{ij}.c_j \qquad (1)$$

This problem can be modeled as what will be explained in section (5).

IV. Fuzzy c-means clustering algorithm

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method (developed by Dunn in 1973 and improved by Bezdek in 1981) is frequently used in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2 \qquad (2)$$

$$1 \le m < \infty$$

Where m is any real number greater than 1, u_{ij} is the membership degree of x_i in (delete the) cluster j, x_i is the ith of d-dimensional measured data, c_j is the d-dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center.

Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by:

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m .x_i}{\sum_{i=1}^{N} u_{ij}^m} \qquad (3)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|}\right)^{\frac{2}{m-1}}} \qquad (4)$$

This iterative process will stop when $\| U^{(k+1)} - U^{(k)} \| < \delta$

Where δ is a termination criterion between *0* and *1*, whereas k is the number of iteration steps. Finally, this procedure converges to a local minimum or a saddle point of j_m. The algorithm is composed of the following steps:

1. *Initialize* $U=[u_{ij}]$ *matrix,* $U^{(0)}$.
2. *At step k: calculate the centers vectors* $C^{(k)} = [c_j]$ *using* $U^{(k)}$.
3. *Update* $U^{(k)}$, $U^{(k+1)}$.
4. *If* $\| U^{(k+1)} - U^{(k)} \| < \delta$ *then STOP; otherwise return to step 2.*

V. Mathematical programming model

The following variables and indices are used in the mathematical model definition.

A. Sets and indices

Set of customers	$k \in N$	$k = \{1,\ldots,N\}$
set of products	$i \in M$	$M = \{1,\ldots,M\}$
set of catalogs	$j \in T$	$T = \{1,\ldots,T\}$

B. *Model parameters*

Parameter's notation	Description
r	*Number of available products on each catalog*
t	*Minimum customer interest threshold (product number)*
p_i	*Priority degree of customer ith*

C. *Decision variables*

Variable	Description
x_k	*1 if catalog k is chosen, otherwise 0.*
v_{kj}	*1 if need of customer k is covered by catalog j, otherwise 0.*
y_{ji}	*1 if catalog j includes product i, otherwise 0.*
a_{ki}	*1 if customer k needs to product i, otherwise 0.*

The mathematical programming model is presented as follows:

$$Max\ Z_1 = \sum_{k\in N} X_k \tag{5}$$

$$Max\ Z_2 = \sum_{k\in N} p_k X_k \tag{6}$$

$$s.t.$$

$$\sum_{i\in M} Y_{ji} = r \quad \forall j \in T \tag{7}$$

$$tV_{kj} \le \sum_{i\in M} a_{kj} Y_{ji} \quad \forall k \in N, j \in T \tag{8}$$

$$X_k \le \sum V_{kj} \quad \forall k \in N \tag{9}$$

$$X_k, Y_{ji}, V_{kj} \in \{0,1\} \quad \forall k \in N, j \in T, i \in M$$

(10)

The existing relations in the model can be described as follows:

Objective function (Z_1) maximizes the number of customers covered by the catalogs. Objective function (Z_2) maximizes overall satisfaction level results of the catalogs. Constraints (7) ensure that each catalog includes r products. Constraints (8) guarantee that the need of each customer should be satisfied by at least one catalog. Constraints (9) state that if the customer need is satisfied by a catalog, this catalog needs to be chosen. Constraints (10) impose the binary nature on sets of decision variables. Note that if the sizes of the catalogs are different, then the right-hand sides of constraints (7) should be set. As following, the above multi-objective mathematical programming model will be changed to a single **ILP** model by applying decision making methods based on fuzzy theory.

VI. Fuzzy Multi-objective Decision Making (FMODM)

In the recent decades, Fuzzy Multi-objective Decision Making (FMODM) methods are considered by researchers. In these methods, instead of using an optimization criterion, several optimization criteria are used. Two major areas have evolved, both of witch concentrate on decision making with several criteria: Multi Objective Decision Making (MODM) and Multi Attribute Decision Making (MADM). Fuzzy set theories can be used beside each of them. In MODM problems, we don't have a single optimal solution, what we have is a set of efficient solutions that we should choose one of them as an optimal compromise solution with respect to the preference that decision maker attributes to different objective functions.

In order to model a multi objective problem on fuzzy method, function h_i can be defined as follows:

$$h_i : R \rightarrow [0,1]$$

Where h_i is a function of real numbers set on domain [0,1]. In this function, value of $h_i(t)$ implies the satisfaction degree of the decision maker from function ith for a specific value t. One of the applied functions used in function definition h_i is as follows:

$$h_i = (f_i(x)) = 1 - \frac{M_i - f_i(x)}{M_i - m_i} \tag{11}$$

Where M_i and m_i respectively represent maximum and minimum values of the function on space of the variable decisions. In order to obtain values of M_i and m_i , the problem should be solved by considering only objective function ith and model constraints. Following figure shows the satisfaction degree of objective function ith. It's noted that maximization of the objective function is considered.

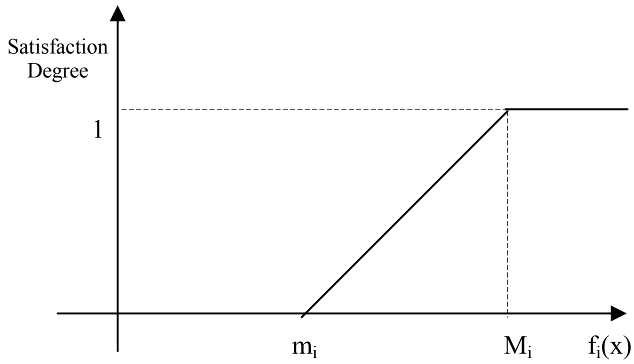

Fig. 1. Utility function related to decision variables

In Fig. 1. , it's obvious that the satisfaction degree equals *1* for some solutions $f_i(x) \ge M_i$ (ideal solution for objective function ith) and the satisfaction degree equals *0* for some solutions $f_i(x) \le m_i$ (non-ideal solution for objective function ith). Also the satisfaction degree in interval (m_i,M_i) tend to 1 in term of objective function bigness. Thus a good compromise solution for a multi objective problem can be defined as "a solution that best serves our cause with respect to set of objective functions".

Another one of applied functions that also used in this paper is as follows:

$$h_i(t)=\begin{cases}1 & t \geq R_i \\ u_i(t)=1-\dfrac{R_i-f_i(x)}{R_i-r_i} & r_i \leq t < R_i \\ 0 & t < r_i\end{cases} \qquad (12)$$

Where

$R_i \leq M_i = \max\{f_i(x)\}\ \ x \in X$

$r_i \geq m_i = \min\{f(x)\}\ \ x \in X$

r_i , R_i respectively represent at least satisfaction level and perfect satisfaction level of objective function *i*th. This is shown in Fig. 2:

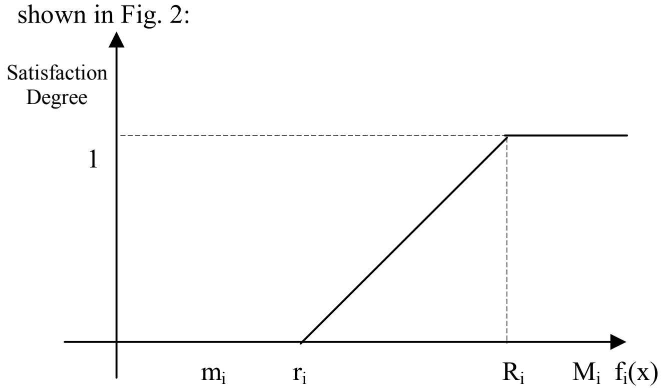

Fig. 2. Revised utility function

In the next stage, the objective functions are changed to a unique utility function by using one of integration operators. In this paper, standard integration operator of *max-min* is used. Here, we should point out that instead of the operator *max-min*, each other combination of *T-norms* and *S-norms* can be used; but if other operators are used, the model constraints are changed to non-linear form that will increase the complexity of the model.

The multi-objective mathematical programming model can be changed to an ILP model as explained below:

$$\begin{aligned}
&Max\quad \alpha \\
&s.t. \\
&\alpha \leq 1-\frac{R_i-\sum_{k\in N} X_k}{R_i-r_i} \\
&\alpha \leq 1-\frac{R_i-\sum_{k\in N} p_k . X_k}{R_i-r_i} \qquad (12) \\
&\alpha \leq 1 \\
&\sum_{i\in M} Y_{ji}=r \quad \forall j \in T \\
&tV_{kj} \leq \sum_{i\in M} a_{kj} Y_{ji} \quad \forall k \in N, j \in T \\
&X_k \leq \sum V_{kj} \quad \forall k \in N \\
&X_k, Y_{ji}, V_{kj} \in \{0,1\} \quad \forall k \in N, j \in T, i \in M
\end{aligned} \qquad (13)$$

In above equation, λ shows the marginal utility results of solving the model that always its value is between [0,1].

VII. Numerical Example

A supplier with 10 different products and 15 customers is considered. The sale price for each of these products is presented in table 1. In table 2 the products of interest to each customer and in addition value summation of the products of interest to him/her are shown.

TABLE 1
The prices of available products

Product	1	2	3	4	5	6	7	8	9	10
Price	4100	6200	5400	4700	5800	6900	4000	5200	3900	5000

TABLE 2
The products of interest to each customer and in addition value summation of them

Customer	Interest products	Value of interest products	Customer	Interest products	Value of interest products
1	1,10	9100	9	1,4,6,7,9	23600
2	2,4	10900	10	2	6200
3	1,3,4,5,7	24000	11	4,9	10600
4	1,2,3,4,6,7,9,10	40200	12	1,2,5,7	20100
5	2,9	10100	13	4,5,10	15500
6	1,2,4,6,9	25800	14	5,6,7	16700
7	3,5	11200	15	1,2,3,7	19700
8	5,6,7,8	21900			

After clustering the customers by fuzzy c-means algorithm, the membership degree of each customer on each of three clusters was obtained as shown table 3. In the next stage, the priority of each customer was calculated according to (1) as table 4.

TABLE 3
The membership degree of the customers in each group

Customer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Low	0.99	1	0.02	0	1	0.06	0.99	0	0.02	0.93	1	0.03	0.58	0.39	0.05
Medium	0.01	0	0.96	0	0	0.88	0.01	1	0.98	0.06	0	0.96	0.39	0.59	0.94
High	0	0	0.02	1	0	0.07	0	0	0.01	0.01	0	0.01	0.03	0.03	0.01

TABLE 4
The priority degree of each customer

Customer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Priority degree	1.04	1.02	4.98	9	1	5.04	1.04	5	4.98	1.35	1	4.9	2.77	3.58	4.85

Assume that we want to design three catalogs which each of them include 5 different products. The results of solving the integer linear programming model is given as table 5.

TABLE 5
The results of solving the integer linear programming model

Catalog number	Available products in catalog
1	1-2-4-6-9
2	2-5-6-7-8
3	1-3-4-5-7

If it's given that at least necessary utility level for satisfying the customers need equals 75% of products needed by each customer, according to catalogs achieved from solving the mathematical programming model, the need of 13 customers are satisfied. Note the marginal utility degree is equal to 82%. To solve the above model, a *PC* with processor *P4-2.4 GHz* was used and integer linear programming problem was modeled and solved by using *Lingo optimization software*.

VIII. CONCLUSION

In the paper, we have studied the fuzzy clustering problem of different products in k variant catalogs, each of size r products that maximize the customer satisfaction level in customer relationship management. The proposed problem is firstly modeled as a bi-objective mathematical programming model. Then this model is changed to a single integer linear programming model by applying the fuzzy theory concepts. Finally, the efficiency of the proposed solution procedure is examined by using a numerical example.

REFERENCES

[1] S. Chiger, "Benchmark survey on critical issues and trends", *Catalog Age* 20 (13) (2003), pp. 32–37.

[2] A. Amiri, "Customer-oriented catalog segmentation: Effective solution approaches", *Decision Support Systems* 42, 2006, pp.1860–1871

[3] M. Ester, R. Ge, W. Jin, Z. Hu, "A Microeconomic Data Mining Problem: Customer-Oriented Catalog Segmentation", *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining (Seattle, Washington)*, 2004, pp. 557–562.

[4] M. Turkey, "A mixed-integer programming approach to the clustering problem with an application in customer segmentation", *European Journal of Operational Research* 173, 2006, pp. 866–879

[5] M.R. Rao, "Cluster analysis and mathematical programming", *Journal of the American Statistical Association* 66, 1971, pp. 622–626.

[6] P.S. Bradley, U.M. Fayyad, O.L. Mangasarian, "Mathematical programming for data mining: Formulations and challenges", *Journal on Computing* 11 1999, pp. 217–238.

[7] C. Rygielski, J.C. Wang, D.C. Yen, "Data mining techniques for customer relationship management", *Technology in Society* 24, 2002, pp. 483–502.

[8] J. Kleinberg, C. Papadimitriou, P. Raghavan, "Segmentation problems", *Proceedings of the Thirtieth Annual ACM Symposiumon Theory of Computing (Dallas, Texas)*, 1998, pp. 473–482.

[9] H. J. Zimmermann, *Fuzzy Set Theory and Its Applications*, Kluwer academic publishers, 3rd Edition, Boston/ Dordrecht/ London, 1996.