UACEE International Journal of Artificial Intelligence and Neural Networks – IJAINN
Volume 3 : Issue 2 [ISSN 2250 – 3749]
Publication Date : 05 June 2013

A Novel Approach for Dynamic Classification in CRM using Fuzzy Logic

Shabnoor Shaikh
Information Technology Dept.
SVMIT,
Bharuch, India
kitty.141191@gmail.com

Ekta Patel
Information Technology Dept.
SVMIT,
Bharuch, India
ektapatelnnm@yahoo.in

Falguni Soni
Information Technology Dept.
SVMIT,
Bharuch, India
soni.falguni11@gmail.com

Khushbu Tapiawala
Information Technology Dept.
SVMIT,
Bharuch, India
khushbut2@gmail.com

Rutvij Jhaveri
Computer Science & Engg./I.T. Dept.
SVMIT,
Bharuch, India
rhj_svmit@yahoo.com

***Abstract*—Customer Relationship Management (CRM) analyzes, classifies, examines and manages customer's needs according to their interest. In this paper, we propose an algorithm for grouping customers using fuzzy logic. The term fuzzy means not clear or blurred and fuzzy logic means representing variation or imprecision in logic. We use fuzzy logic, and not other mining and statistical methods, to dynamically classify the customers. As a result, we dynamically classify customers into more than one groups at the same time. The algorithm calculates the customer's chance of belonging to a particular group on the basis of probabilities without using any prior information. Fixed number of groups is not given as input because groups are dynamically generated according to the requirement.**

Keywords—Customer Relationship Management, fuzzy logic, probability, dynamic classification.

I. Introduction

Classification is a form of data examination that is used to mine models describing important data classes [1].We may build classification approach to categorize customers. Researchers use many classification methods in pattern reorganization and statistics. Most algorithms occupy more memory and typically assuming small data size. The basic techniques for classification are Bayesian, tree classifier and rule-based classification [1]. Bayesian classification can predict class bias probabilities such as the probability that a given record belongs to a particular class [2, 3]. The major inadequacy of such classification is that the fix number of groups is given as an input [4, 5]. Though we need to classify customers for which fixed groups are not defined as we classify customers according to their interest. In this paper, we propose an algorithm that is an extension to Bayesian as it removes the mentioned drawback of Bayesian.

The proposed algorithm calculates probability of customer and predicts the class label for that customer. The customer is classified to the class to which the higher probability belongs to or it is assigned to a new class if it is less than the predefined fuzzy value. The algorithm includes fuzzy logic for dynamically generating a new group for customers whose matching class is not found.

The rest of the paper is organized as follows. In section II, we discuss various existing algorithms that are usually used for the purpose of grouping. These algorithms include various classification algorithms such as Bayesian, and also clustering algorithm such as k-means and k-medoid. In section III, we discuss the proposed algorithm. This algorithm does dynamic classification without requiring any training dataset. In section IV, we discuss the explanation of proposed algorithm with real example. In section V, we conclude with the advantages of the proposed algorithm compare to existing algorithms.

II. Related Work

There are many systems based on CRM that groups customers with the use of various clustering algorithms such as k-means and k-medoids [1]. Clustering is the task of separating data points into consistent classes or clusters so that items in the same class are as related as possible and items in different classes are as unrelated as possible [1]. k-means is a unsubstantiated partition based algorithm which classifies different data objects into *k* number of groups based on resemblance [1].

In k-means number of clusters and initial centroid are defined and given in advance [6, 7]. On the other side, our algorithm does not give fixed number of groups. Moreover, k-means does not work for categorical data and is not suitable for arbitrary shaped clusters [8]. k-medoids is also similar to k-means but only difference is that it uses median to calculate centroid [9, 10]. But the problem remains the same as k-means. We cannot use clustering as it cannot work with one customer, it needs more than one customer to initiate the clustering. If we consider classification then we can also group categorical data, we can use Bayesian classification for it; but customers can be classified into one of the given groups. In CRM, customers' demands and interests change frequently. Therefore, belongings of a group may also change and new group generation is required. Also, the limitation for both classification and clustering is that fixed number of groups is given as input. Therefore, we propose an algorithm which overcomes the limitations of exiting classification algorithms by including fuzziness in them.

TABLE I. VARIOUS APPROACHES STUDIED FOR ALGORITHM

Approaches	Limitations
Bayesian algorithm[11]	It can classify numeric as well as categorical data but it works for fixed group given as input and hence fuzziness cannot be included.It needs training dataset to be given as input.
k-medoids algorithm[12, 13]	It also clusters numeric data and does not work well for large dataset.
k-means algorithm[14, 15]	It clusters only numeric data and fix number of groups must be given as input.

III. Proposed Algorithm

The algorithm is based on FCC, that uses the fuzzy value (0.5) to decide whether the value of probability is acceptable or not for a particular group. So here classification is not on static value but is done through fuzzy value.

A. *Algorithm*

Input: Data of new customer

Output: Dynamic classification of customers

Step- 1: Calculate the total no. of existing customers in database.

IF (total no. of customers=0)
Then Go to Step-2
Else Go to Step-3

Step- 2: Create the first group for that customer with group's ID=1.

Step- 3: Select the entire column of the numeric data (Budget) from the database.

Step- 4: Compare each value of budget with customer's entered value.

Considering current budget value from entire list =a and Customer's entered value=b.

IF (b + 15% of b)
Then Go to Step- 4.1

4.1: Select the budget and group's ID from the database where user's ID is the ID of the current budget (a).

4.2: Add these budgets and group's ID to the list. This is the list of such budget whose values are nearer to the customer's entered budget.

4.3: For each value in the list perform:

4.3.1: Select group's ID for each value and store in variable say group.

4.3.2: Count the total no. of entries in that group.

4.3.3: For each categorical value find the probability of belonging to that group.

Probability for mode P (M):(no. of entries with selected mode in group) / (total count in group).

Probability for state P (S): (no. of entries with selected state in group) / (total count in group).

Probability for city P (C): (no. of entries with selected city in group) / (total count in group).

Total probability P = (P (M) + P (S) + P (C)) / 3.

4.4: Repeat Step- 4.3 for all the budgets selected in Step- 4.2.

Step- 5: Find the maximum probability (MAX) from all the probabilities found after Step- 4.

Step- 6: IF (MAX > 0.5)
Then Go to Step- 7
Else Go to Step- 8

Step- 7: Add the new customer with the group's ID to which this MAX probability belongs.

Step- 8: Create new group for that customer with group's ID= max (group's ID) + 1.

B. *Flowchart*

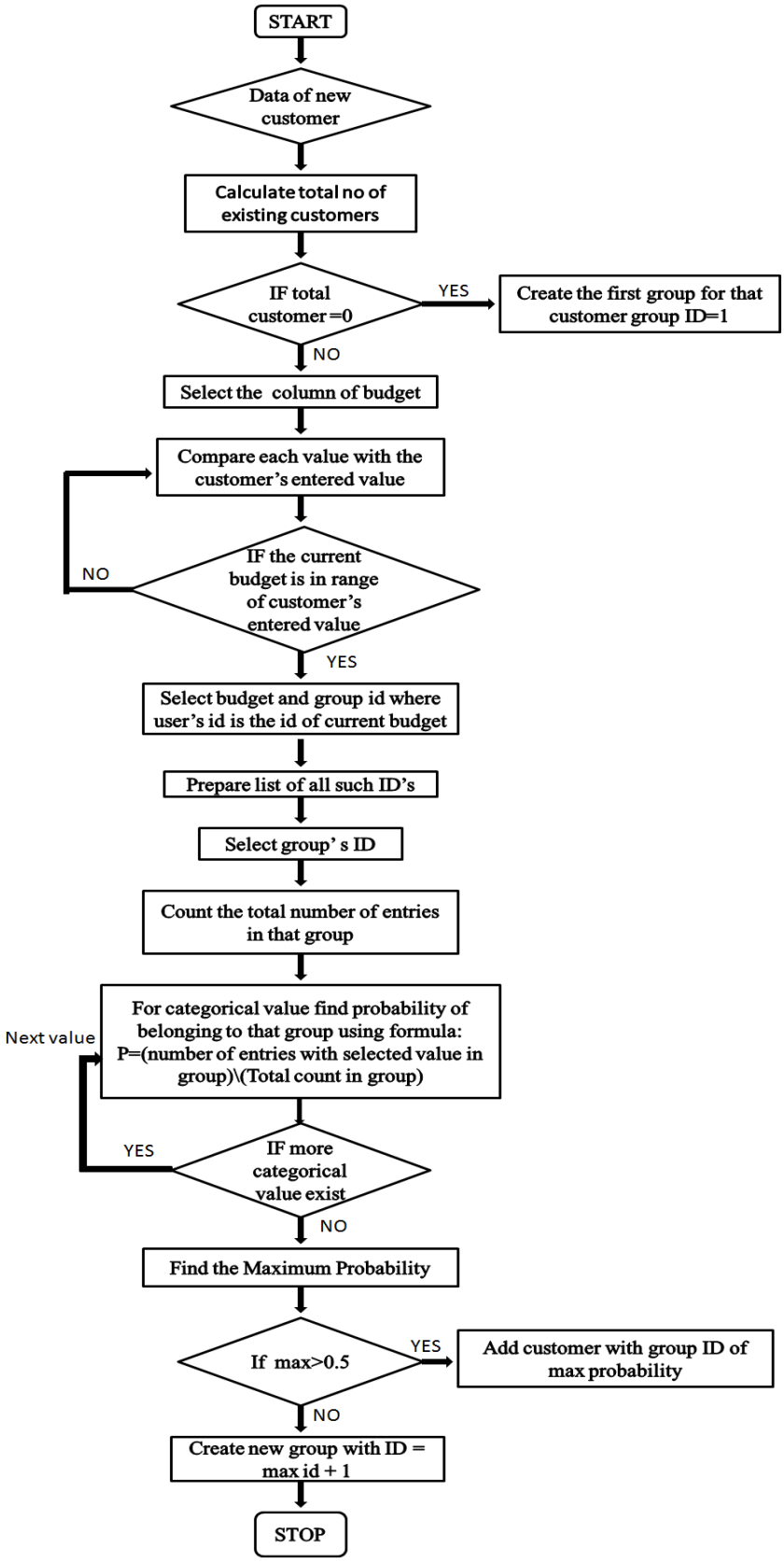


Figure 1. Flowchart of proposed algorithm.

IV. Exploring Algorithm

tblUser: Query(h...ATA\DATABASE.MDF)

UserID	UserName	Email	ContactNo	Budget	Mode	State	City	GroupID
512	shabnoor123	kitty.141191@g...	9908765432	26000	Bus	Gujarat	Bharuch	1
513	ekta123	ektapatelnnm...	9908765432	28000	Bus	Gujarat	Baroda	1
514	fali123	falguni.soni@g...	9908765432	29000	Bus	Maharastra	Panchgani	2

Figure 2. Database before applying algorithm.

As shown in Figure 2 and Figure 3, customers are entered into the database, according to this data new customer's probability is calculated and group id is assigned to the customer. For the first customer, no training dataset is present and further dynamically dataset is used to generate the group id for new customer.

Say customer with the following details is to be added: Budget=30,000, mode = bus, state = Maharashtra and city = Mumbai.

Step- 1: Calculate the total no. of existing customers in database.

Total count = 3 (from above database).
IF (total no. of customers=0)
Then Go to Step-2
Else Go to Step-3, here total no. of customer=3 > 0

Step- 2: Create the first group for that customer with group's ID=1.

Step- 3: Select the entire column of the numeric data (Budget) from the database.

Step- 4: Compare each value of budget with customer's entered value.

Let say, current budget value from entire list =a and Customer's entered value=b.
IF (b + 15% of b)
Then Go to Step- 4.1

4.1: Select the budget and group's ID from the database where user's ID is the ID of the current budget (a).

4.2: Add these budgets and group's ID to the list. This is the list of such budget whose values are nearer to the customer's entered budget.

4.3: For each value in the list perform:

4.3.1: Select group's ID for each value and store in variable say group.

4.3.2: Count the total no. of entries in that group.

4.3.3: For each categorical value find the probability of belonging to that group.

Probability for mode P (M):(no. of entries with selected mode in group) / (total count in group).
Probability for state P (S): (no. of entries with selected state in group) / (total count in group).
Probability for city P (C): (no. of entries with selected city in group) / (total count in group).
Total probability P = (P (M) + P (S) + P (C)) / 3.
Result after computation of step3:
Here b= 30,000

TABLE II. COMPUTATION FOR STEP4

A	Group's ID	Count (group)	P (M)	P (S)	P (C)	Total (P)
26000	1	2	1	0	0	0.33
28000	1	2	1	0	0	0.33
29000	2	1	1	1	0	0.66

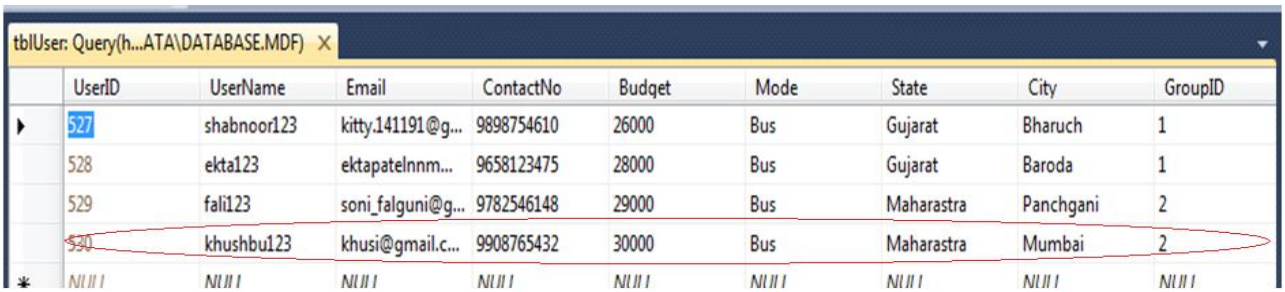
4.4: Repeat Step- 4.3 for all the budgets selected in Step- 4.2.

Step- 5: Find the maximum probability (MAX) from all the probabilities found after Step- 4.
It is clear from above table that MAX=0.66

Step- 6: IF (MAX > 0.5)
Then Go to Step- 7
Else Go to Step- 8

Step- 7: Add the new customer with the group's ID to which this MAX probability belongs.
(0.66 > 0.5)
MAX belongs to group's ID =2, so the customer is added with the group's ID= 2.

Step- 8: Create new group for that customer with group's ID= max (group's ID) + 1.


tblUser: Query(h...ATA\DATABASE.MDF)

UserID	UserName	Email	ContactNo	Budget	Mode	State	City	GroupID
527	shabnoor123	kitty.141191@g...	9898754610	26000	Bus	Gujarat	Bharuch	1
528	ekta123	ektapatelnnm...	9658123475	28000	Bus	Gujarat	Baroda	1
529	fali123	soni_falguni@g...	9782546148	29000	Bus	Maharastra	Panchgani	2
530	khushbu123	khusi@gmail.c...	9908765432	30000	Bus	Maharastra	Mumbai	2
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Figure 3. Database After Applying Algorithm.

As shown in Figure 3, the new customer is added with group id 2 as the requirement of this customer is same as the existing customer of this group. If the customer's requirement does not match with any of the groups then entire new group will be created.

v. Conclusion

The existing algorithms do not function without training dataset; they classify customers from the fixed no of classes given as input. In this paper, we introduce a novel Fuzzy Customer Classification algorithm to classify customers. The proposed algorithm can dynamically perceive the clusters of customers which overcomes the limitations of the known algorithms. Experimental outcome demonstrates that the proposed FCC algorithm significantly reduces the unwanted results in customer classification. Also, the algorithm is relatively independent of the type of added noise because the training dataset is not used, and as significance, in the deficiency of prior knowledge of the noise. The algorithm can be used in CRM systems and may prove its significance for researchers working in this area.

Acknowledgement

The authors would like to thank Asst. Prof. Vaisali Patel of SVMIT, Bharuch for her valuable suggestions and guidance.

References

[1] Clau dio moraga, "Introduction to fuzzy Logic", Facta universitatis(NIS), vol.18, No.2, pp.319-328, August 2005.
[2] John Yen, "Fuzzy Logic-A Modern perspective", IEEE, pp.1-25, October 1998.
[3] Han,j., and M.kamber, 2006. Data Mining: Concepts and Techniques. Morgan kaufman publishers, 2nd edn., New Delhi, ISBN: 978-81-312-0535-8.
[4] Leon Versteegan, "The simple Bayesian classifier as a Classification algorithm". http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.55.8281&rep=rep1&type=pdf.
[5] P.domingos, and M.pazzani, "On the optimality of the simple Bayesian Classifier under Zero-One Loss", Kluwer Academic publishers, pp. 103-130, 1997.
[6] Jiangtao Ren, Sau Dan Lee, Xianlu Chen, Ben Kao, Cheng R, and Cheng D, " Naive bayes Classification of uncertain Data", IEEE computer society, pp. 944-949, Dec 2009.
[7] Biao Qin, Yuni Xia, Shan Wang, and Xiaoyong Du, "A Novel Bayesian Classification Technique for Uncertain Data", pp. 2-22. http://citeseerx.ist.psu.edu/~yxia/publications/KBS-Bayesian.pdf
[8] Tapas Kanungo, David m.mount, Nathan S.Netanyahu, Christine D.Piatko,Ruth Silverman, and Angela Y.Wu, "An Efficient k-means Clustering Algorithm: Analysis and Implementation", IEEE Transactions on pattern analysis and machine intelligence, Vol 24, No.7, pp. 881-892, July 2002.
[9] Vance Faber, "Clustering and Continuous k-means Algorithm", Cos Alamos Science, No.22, pp. 138-144, 1994.
[10] K. A. Abdul Nazeer, and M. P. Sebastian "Improving the Accuracy and Efficiency of the K-means Clustring Algorithm", Proceedings of the World Congress on Engineering, Vol. I, 2009.
[11] T.Velmurugan, T.Santham, "Computational Complexity Between K-means and K-medoids Clustering Algorithms for normal and uniform Distributions of Data Points", Vol 27, No.1, pp. 19-30, May 2011.
[12] Park, H.S., J.S.Lee, and C.H. Jun, "A K-means Like Algorithm for K-medoids Clustering and its Performance". http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.90.7981&rep=rep1&type=pdf.
[13] Pat Langly, Wayne Iba, and Kevin Thompson, "An Analysis of Bayesian Classifiers", In proceedings of the Tenth national conference on Artificial Intelligence, pp. 224-228, 1992.
[14] T.Velmurugan, T.Santham, "A Comparative analysis between K-medoids, and fuzzy c-means clustering algorithms for Statistically Distributed Data Points", Journal of theoretical and applied Information Technology, Vol 27, No.1, pp. 19-30, May 2011.
[15] Ji Wentian, Guo Qingju, Zhong Sheng and Ehou En, "Improved K-medoids clustering algorithm under semantic web", Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering(ICCSEE), pp. 731-733, 2013.
[16] Mu-Chun Su, Chien-Hsing Chou, "A Modified Version of the K-means Algorithm with a Distance Based on Cluster Symmetry", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 23, No.6, pp. 674-680, June 2001.
[17] Chih-Tang Chang, Jim Z.C.Lai and Mu-Der Jeng, "A Fuzzy K-means Clustering Algorithm using cluster center displacement", Journel of Information Science and Engineering 27, pp. 995-1009, Sep 2011.