Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 05, May - 2017

A Fuzzy Possibilistic Framework for Segmentation of Customer Data for Behavioural Analysis in Banking Sector

Femina Bahari T
Research Scholar,
Division of Computer Science & Engineering
Cochin University of Science and Technology, Kochi,
Kerala - 682022, India

Sudheep Elayidom M
Associate Professor,
Division of Computer Science & Engineering
Cochin University of Science and Technology, Kochi,
Kerala - 682022, India

***Abstract*— Banking plays a crucial role in the economic development of any country. Every banking organization considers their customers as inevitable part of the development of their firm. Behaviour of customers in their various interactions with firm can be closely analyzed and motivates them for better relationship with the firm. Customer Segmentation is important among many Customer Relationship Management strategies. In this paper a Fuzzy possibilistic theory is proposed for a better analysis of customer segmentation in banking. With Fuzzy logic, the use of linguistic terms and variables are made possible. As a result we obtain a more human oriented querying process. The segmentation strategies we adopt in banking depend on various features of the customers. Also segmentation based on specific criteria's are applied to customer dataset. Application of fuzzy logic incorporating linguistic information yields better segmentation approaches for customer data analysis. In this paper we discuss the fuzzy possibilistic approach in segmentation of demographic features of bank customers in order to divide the customer database into targeted groups. This approach in segmentation can be made effectively useful in adopting various service oriented strategies for the targeted customer group. Also this approach acts as a strong foundation for application in fuzzy expert systems.**

Keywords*— *Customer Relationship Management, Customer Segmentation, Fuzzy logic, Possibilistic theory, Membership function, Linguistic variables.

I. INTRODUCTION

Customer Relationship Management (CRM) consists of a set of processes and enabling technologies supporting a business strategy to build long term, profitable relationships with specific customers [1]. Customers and the study of their behaviour forms the centre of any organization as the changes in behaviour of customer plays a key role in predicting the profitability of the organization. Various service oriented policies, marketing strategies and planning are designed in view of the customer centric behavioral aspects. Customer segmentation plays a crucial role in the framework of Customer Relationship Management. A successful CRM strategy is built upon the foundation of customer data and information technology (IT) tools [2]. An organization to achieve greater profitability, its relationship with customers should be continuously and actively monitored for behavioral fluctuations. There is no universally accepted definition of CRM, even though CRM is widely accepted as a major business approach [3].

Customer segmentation can be based on customer's value, behaviour, loyalty, socio-demographic features, life stages etc [4]. Targeted customers with specified features are produced as an outcome of segmentation. Earlier in our study a segmentation framework based on demographic features to enhance the outcome of direct marketing is proposed [5].

Fuzzy set theory was first proposed by Lofti Zadeh as a precise mathematical tool which deals with object classes without any precisely defined criteria of membership [6]. Fuzzy logic deals with decisions with uncertain phenomena or partial truth and the rules used to make those decisions cannot be acquired automatically [7]. A general learning method was proposed as a framework to derive membership functions automatically by Hong, et al. The method also derives fuzzy if-then rules from a set of given training examples. Based on the membership functions and the fuzzy rules derived, an expert system prototype is developed to run the fuzzy inference procedure [8].

The proposed framework in our paper uses the possibilistic theory approach in fuzzy logic to study the customer demographic features and the segmentation strategies adopted. Section 2 describes the concepts of fuzzy set theory, fuzzy logic and possibilistic theory. Section 3 defines the customer segmentation in banking and section 4 briefs the proposed fuzzy possibilistic framework for customer segmentation in banking based on the defined concepts. Section 5 describes a case study of district bank with selected features of bank customers.

II. CONCEPTS

A. *Fuzzy Set Theory*

Fuzzy Set Theory was first proposed by Zadeh in 1965 and was concerned with reasoning using natural language in which many words have ambiguous meanings [6].The observed gap between natural human reasoning and classical set theory was the motivation for Zadeh's work. Fuzzy set theory also deals with uncertainty and noise. It is frequently used in various applications and expert systems because of its similarity to human reasoning. In many cases crisp sets could not define many classes that we encounter in the real world. For human beings the perception of world is not in terms of concepts that are defined with sharp cutoffs and limits. The classic sets are inappropriate for such kind of human perceptions. The fuzzy set offers the advantage of defining the concepts in a more acceptable and realistic manner as perceived by human

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 05, May - 2017

reasoning. Sharp cutoffs are replaced with gradual membership functions in fuzzy set theory for a more intuitive representation of reality. Hence fuzzy sets can be looked upon as a natural extension of classical sets [9].

The membership of elements in a classical set is defined in terms of binary values 1 and 0. The value 1 means an element belongs to the set and the value 0 denotes that the element does not belong to a set. The Fuzzy Set allows the elements to have varying degrees of membership, in a range that lies between 0 and 1 [6]. A fuzzy set is built from a reference set which is called universe of discourse and this reference set is never fuzzy. Assume that U ={x_1,x_2,,…x_n} is the universe of discourse, then a fuzzy set A in U (A∈U) is defined as a set of ordered pairs {(x_i ,$\mu_A(x_i)$)} where $x_i \in U$, $\mu_A : U \rightarrow [0, 1]$ is the membership function of A and $\mu_A(x_i) \in [0,1]$ is the degree of membership of x in A. Fig 1. Shows a simple triangular membership function with lower limit a, upper limit b, and value m where a<m<b

$$\mu_A(x) = \begin{cases} 0, & x \le a \\ \frac{x-a}{m-a}, & a<x\le m \\ \frac{b-x}{b-m}, & m<x<b \\ 0, & x\ge b \end{cases}$$

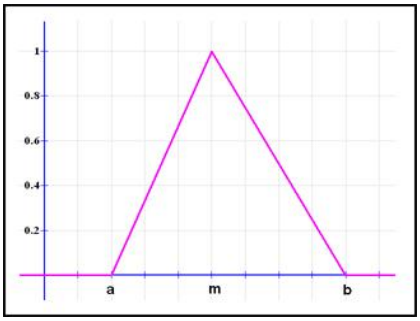

Fig 1.Triangular membership function defined by a lower limit ‘a’, an upper limit ‘b’, and a value ‘m’, where a < m < b

B. Fuzzy Logic and Possibility Theory

Fuzzy logic performs modeling of expert knowledge with the help of linguistic variable that takes linguistic values. For each linguistic variable seven levels can be considered. The value of the linguistic variable can be given as a sub function that shows the degree of dependency of each value to this level. Expert knowledge is required to develop membership functions for the members in a set. For a fuzzy set a membership value is defined as the degree to which an element belongs to this fuzzy set. Membership degree can also be interpreted as degree of truth, certainty factor, degree of possibility, degree of satisfaction [9]. Zadeh extended the fuzzy set theory to a possibility theory in which the degrees of possibility and the membership values are considered as same. Natural language is more possibilistic in nature rather than probabilistic which gives emphasis to possibility theory in fuzzy. Here more focus is on the meaning of the information. This theory in turn opens the door to fuzzy reasoning by which natural language can be represented and manipulated in a more better and effective manner. But due to complexity of the problems only approximate linguistic expressions are used. Probability is degree of likelihood whereas possibility is degree of feasibility [9].

III. CUSTOMER SEGMENTATION IN BANKING

Segmentation in Banking can be performed based on the features and behaviour of the customers. The customer demographic features can be segmented based on the requirement of the application studied. A close analysis of the segmented features can be done to assess the customer’s value, behaviour, loyalty, etc [5]. Segmentation often produces targeted customers with specified features. In our proposed framework a fuzzy possibilistic theory approach is applied to the customer sociodemographic features and linguistic variables are used for segmenting the datasets in order to effectively segment the bank customer database. The methodology adopted again depends to large extend on business and data understanding, preparation and identification of valuable segments.

Segmentation in banking can be used to create a basic frame work to define the customer features, their behaviours and then the grouping of them in view of their differences identified. The customers could be classified in the same category as well in view of the segmented approach. Segmentation based on the purpose for which loan is sanctioned, income of the applicants, period for which the loan is approved etc are few of the features upon which the credit risk analysis concentrate [10]. Fuzzy logic method is used to compute membership functions and hence the degree of certainty or possibility of each member value. This method helps in generating fuzzy inference rules with help of fuzzy extension principle. Fuzzy expert systems are built on such rules in many decision making systems.

IV. FUZZY POSSIBILISTIC FRAMEWORK FOR CUSTOMER SEGMENTATION IN BANKING

The proposed framework highlights the role of linguistic variables and the associated fuzzy logic theory applied to the customer segmentation strategies. The customer profile can be categorized more effectively based on the domain expert knowledge applied in segmentation. The customer behaviour can also be grouped into classes of interest for further application of classification model. The selection of qualifying attributes, the introduction of customer classes and the choice of membership functions that are appropriate are important design issues [9].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 05, May - 2017

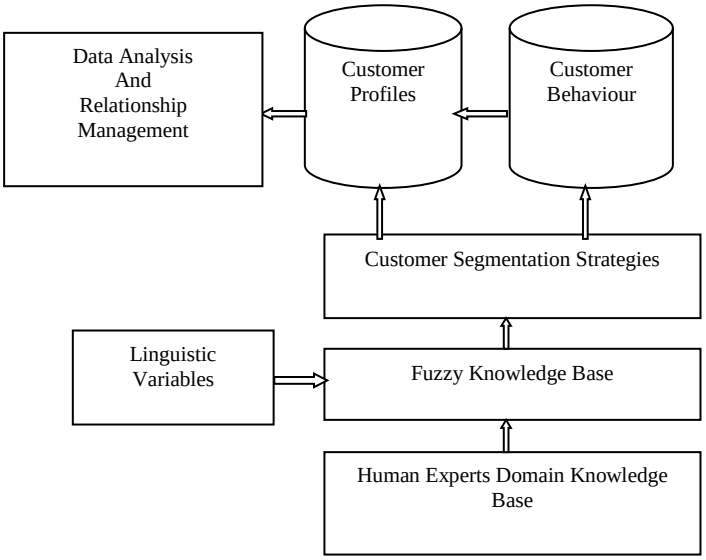

Fig 2. Fuzzy Possibilistic Framework for Customer Segmentation

Using the knowledge of experts of the domain is one method for effectively forming the groups of customers that tend to be useful in the design of expert systems or classification models. The complexity of data due to abundance in information which in turn give rise to ambiguity in reasoning can be alleviated to some extend with this proposed approach. Also the fuzzy rules based on domain expert give more clarity to the class of problems and the associated solutions.

V. LINGUISTIC APPROACH IN CUSTOMER SEGMENTATION: A DISTRICT BANK CASE STUDY

A district bank case study is done on bank's customer data of three years for a period from 2013 to 2015 (the name of the bank is altered to meet the terms of privacy). Out of 20404 records collected only 11113 records were available to perform the segmentation studies after removal of missing and duplicate valued records. The data analysis of the segmented features help banks to assess the behavior of customers and segment them effectively in order to assess the financial risk associated with granting loan to the customers. There are various other applications in banking field that could make use of the segmentation strategies on customer database.

According to the description from crisp to fuzzy sets, from our bank dataset we identified five attributes of fuzzy properties. These attributes are age, asset value, loan period, loan amount and income. As mentioned in the linguistic variable part, seven levels can be considered for each of the fuzzy features. Let us consider for example the attribute age. The range of value for the linguistic variable age of customers [C(age)] is from18 to 92 years old, which is considered the universe of discourse of fuzzy set to characterize the age, [U[18,92]]; (C(age) ∈U[18,92]). Seven levels and the intervals specified for the attribute age is shown here. The levels are listed as {extreme low (EL), very low (VL), low (L), medium (M), high (H), very high (VH), extreme high (EH)}. Even though we have defined seven possible levels we can always concentrate to smaller number of levels as required by our application and context of usage of features. In many application areas we only need to categorize the data into three or five levels based on the value of the feature for yielding better segmentation results.

$$Age(x) = \begin{cases} if\ 10 \le x < 33, & "EL" \\ if\ 20 \le x < 43, & "VL" \\ if\ 33 \le x < 55, & "L" \\ if\ 43 \le x < 67, & "M" \\ if\ 55 \le x < 78, & "H" \\ if\ 67 \le x < 90, & "VH" \\ if\ 78 \le x < 100, & "EH" \end{cases}$$

Membership functions can be defined for each of the seven linguistic variables in age category. The sub function defined for "EL", "VL" as triangular membership function is given below.

$$\mu_{EL}(x) = \begin{cases} 0, & x \le 10 \text{ or } x \ge 32 \\ \frac{x-10}{20-10}, & 10 < x \le 20 \\ \frac{32-x}{32-20}, & 20 < x < 32 \end{cases}$$

$$\mu_{VL}(x) = \begin{cases} 0, & x \le 20 \text{ or } x \ge 42 \\ \frac{x-20}{32-20}, & 20 < x \le 32 \\ \frac{42-x}{42-32}, & 32 < x < 42 \end{cases}$$

For each of the variables the membership functions can be defined in a similar manner. The same approach can be made applicable to the other selected features in the dataset. Table1. Shows the range of values in each level for linguistic variable age.

TABLE.1. AGE INTERVAL AND SEVEN LEVELS OF LINGUISTIC VARIABLE AGE

Level	Linguistic Term	Age Interval
L1	EL	10-33
L2	VL	20-43
L3	L	33-55
L4	M	43-67
L5	H	55-78
L6	VH	67-90
L7	EH	78-100

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 05, May - 2017

Therefore each level has a sub function which is a membership function defined on the fuzzy set and which shows the degree of dependency of each value to this level [11]. The linguistic values are described by a membership function. Each level has a triangular membership function as shown in fig 3.

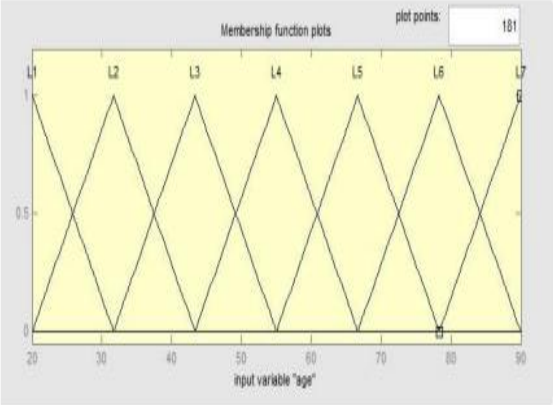

Fig 3. Membership function plot for linguistic variable 'age'.

Seven levels are considered in this attribute feature segmentation. The number of levels varies based on the problem and different criteria imposed on the situation. The same criteria are applied to the other features mentioned above.

The membership functions of the sub functions are defined to appropriately categorize the customer features. The result of the segmentation yields a better categorization with fuzzy knowledge base. This approach in segmentation adds to the knowledge base in fuzzy logic systems used for decision making purposes. The data can be effectively used in fuzzy expert systems for fuzzy rule generation.

CONCLUSION

Segmentation of the customers in banking based on their contributing features will aid in data analysis and effective data model development. In decision making problems where human experts knowledge can contribute more to better classification, the idea of incorporating fuzzy possibilistic approach to this knowledge base can yield much better results. With expert knowledge base the segmentation strategies applied in the database will be used in our fuzzy expert systems for the prediction of customer behaviour. We hope that this strategy can yield better outcome compared to other classification models, even though this may depend to large extend on the context of application and concepts defined. The same can be used in fuzzy logic systems with the predefined fuzzy knowledge base and rules generated.

REFERENCES

[1] Ling, R., & Yen, D. C., "Customer relationship management: An analysis framework and implementation strategies", Journal of Computer Information System, vol. 41, pp 82–97, 2001.

[2] Ngai, E. W. T., "Customer relationship management research (1992–2002):" An academic literature review and classification. Marketing Intelligence, Planning, vol. 23, and pp 582–605, 2005.

[3] Ngai, E. W. T., Xiu, L. and Chau, D. C. K., "Application of data mining techniques in customer relationship management: A literature review and classification", Elsevier: Expert Systems with Applications, vol. 36, pp 2592–2602, 2009.

[4] Konstantinos Tsiptis and Antonios Chorianopoulo. Data Mining Techniques in CRM: Inside Customer Segmentation, Wiley & Sons LTD, 2009. 6. D. H. Payle, "Bank Risk Management", Working Paper No.272, University of California., Berkeley Research Program in Finance Working Papers.No: RPF-272. July, 1997.

[5] Femina Bahari T, Sudheep Elayidom M, "A Data Driven Framework for Customer Segmentation and Prospective Customer Prediction in Direct Marketing", International Journal of Research in Computer Science and management, Vol.. 3, issue 2, pp 15-18 July 2016.

[6] Zadeh, L.A. Fuzzy Sets, Information and Control. 1965.

[7] Umoh, A. U., Nwachukwu, E. O., Obot, O. U and Umoh, A. A., "Fuzzy Inference System for effective Control of profitability in paper Recycling Plant" America Journal Science and Industrial Research, vol. 2, issue 4, pp 552-558, 2011.

[8] Hong, Tzung-Pei and C. Lee, Induction of fuzzy rules and membership functions from training examples, Fuzzy Sets and Systems, Vol. 84, pp. 33-47, Nov.

[9] N.Werro. Fuzzy Classification of Online Customers. PhD thesis, University of Fribourg, 2008.

[10] Femina Bahari T, Sudheep Elayidom M, Segmentation Approach in Predicting Credit Worthiness of Customers in Banking", International Journal of Research in Computer Science and management, Vol.. 4, issue 1, pp 17-20 Dec 2016

[11] M. Mukaidono. Fuzzy Logic for Beginners. World Scientific Publishing, London, 2001.