



AN EFFICIENT FUZZY COMPUTATIONAL FRAMEWORK FOR CUSTOMER SEGMENTATION MODEL IN CREDIT ANALYSIS

Femina Bahari T

Research Scholar, Division of Computer Science &
Engineering
Cochin University of Science & Technology
Kochi, Kerala, India 682022

Dr Sudheep Elayidom M

Associate Professor, Division of Computer Science &
Engineering
Cochin University of Science & Technology
Kochi, Kerala, India 682022

Abstract: In this paper we propose an efficient fuzzy computational framework for customer segmentation model in credit analysis. Normally segmentation methods cannot perform complex analysis so as to obtain the customer segments with high value. If the knowledge of experts in the data domain can be imparted to the generation of segments, this can bring in better results in the performance of classification models. In our approach customer attributes are selected after knowledge experts analysis and are segmented based on the limits set by them on the real numerical values. For each segment we generate the segmentation rules with definition of fuzzy basic linguistic term set. Each linguistic term set is assigned to a fuzzy membership function to generate the segmentation function. Combining the segmentation rules and generated functions the real valued numerical attributes are converted to fuzzified values in the interval $[0, 1]$. Both linguistic and numeric information are aggregated by a series of computations and a 2-tuple linguistic value is generated for each attribute in the database. The same term after a series of computations can be used in many decision making problems as it suffers no loss of information.

Keywords: Fuzzy, BLTS, symbolic aggregation, customer segmentation, credit analysis.

I. INTRODUCTION

Customer segmentation can be defined as a process by which customers can be divided among various subgroups based on their attributes or various other features such as behaviour, character, value, needs, loyalty, morale etc. [1]. Identifying and grouping customers with unique characters will benefit the organizations by managing them with more fruitful outcomes in products, services and marketing. Segmentation can be seen as a simplification of the messy complexity of dealing with numerous individual customers, each with distinct needs and potential value [2].

The outcome of segmentation depends mainly on the knowledge imparting variables, which can be gathered from demography, market, region etc. [3]. This outcome is very rare to be fully data driven or automatic. Many important decisions have to be identified like the selection of segments, size of the segments, criteria for identifying each segment etc. These input variables can be further divided into subgroups in demographic, geographic and life style [4].

Credit Analysis is essential to identify the risks involved in the decision making regarding any lending situation. It assess the repayment capacity of the customer, the morale and trustworthiness shown by them during their previous interactions, the risks involved in making decision in view of the collateral and the capacity of the customers. The credit analysis is essential to build credit score models which categorize applicants as either accepted or rejected with respect to the applicants' characteristics such as age, income, and marital condition[5]. The benefits of credit scoring involve reducing the credit analysis cost, enabling faster credit decisions, closer monitoring of existing accounts and prioritizing credit collections. Credit analysis enables the officers to accept credit application provided that the applicant is reliable and expected to replay the financial obligation.

Fuzzy approach in segmentation is capable of analyzing the uncertainty and vagueness that characterize the features,

experiences and perceptions of customers in banking [6]. Fuzzy approach in segmentation method assigns each observation to each segment with a certain degree of membership [7]. The fuzzy theoretic approach which includes fuzzy numbers, fuzzy sets, and fuzzy computations is capable to cope with the imprecision, uncertainty and vagueness that characterize each aspect and experience [8].

In our proposed framework we attempt to segment the numerical attributes of the customers in our dataset by generating linguistic terms and member functions with fuzzified approach. In section II a brief description about the fuzzy set theory and linguistic approach is given. Section III describes the general model developed for segmentation in credit analysis. In section IV we have shown an extended Fuzzy computational framework for customer segmentation model which incorporates the fuzzy linguistic approach using membership function computations. Section V we draw our conclusions.

II. FUZZY SET THEORY AND FUZZY LINGUISTIC APPROACH

Fuzzy Set Theory which was first introduced by Zadeh [9] deals with vagueness of human thoughts. In certain decision making problems this vagueness in thoughts yields better ideas and judgments for tracing a solution. Fuzzy Set Theory is a very powerful tool in dealing with such problems where human perceptions give better ideas and thoughts to reach out to a better conclusion or solution. It provides a better way to represent the knowledge of decision maker's especially vague knowledge. Also the use of linguistic variables in scenarios where human thoughts are applied yields better outcome. Linguistic variable values are not mere numbers but words or it can be sentences in a natural or artificial language [10] which the human beings can understand easily. The value is represented in linguistic terms. Before describing the Linguistic

computation models it is appropriate to give the basic definitions for fuzzy sets and fuzzy numbers.

A fuzzy set can be defined as a pair (A, μ) where A is a set and $\mu: A \rightarrow [0, 1]$ is a function. For each $x \in A$, $\mu(x)$ is called grade of membership of x in (A, μ) or membership function of x in (A, μ) . Here we define a trapezoidal fuzzy number which is given with a membership function $\mu_A(x)$ which is defined as follows:

$$\mu_A(x) = \begin{cases} \omega \frac{x-c}{a-c}, & c \leq x \leq a \\ \omega, & a \leq x \leq b \\ \omega \frac{x-d}{b-d}, & b < x \leq d \\ 0, & \text{otherwise} \end{cases}$$

and the trapezoidal member function is as given in Fig. 1.

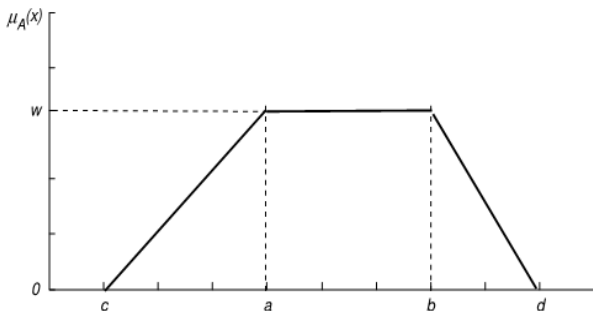


Fig. 1. Trapezoidal membership function.

Linguistic Computation Models or Computing with words and their applications has been widely used in decision making problems [11]. Initial models were based on membership functions and ordinal scales. The drawbacks of such models were in terms of accuracy due to loss of information. Hence the linguistic model [10] with symbolic translations aimed at improving accuracy was computed. Computations were performed on the index of the linguistic terms. In our approach also we performed the symbolic translations based on the index of the linguistic terms. Here we deal with a single term set which is also considered as the Basic Linguistic Term Set (BLTS). Typically we select values with odd number of terms such as 3,5,7,9 etc. The midterm represents an assessment of value 0.5 and rest of the terms is placed symmetrically around it. For example for a set of 7 terms we can select the term set as

$$S = \{S_0 = EL, S_1 = VL, S_2 = L, S_3 = M, S_4 = H, S_5 = VH, S_6 = EH\}$$

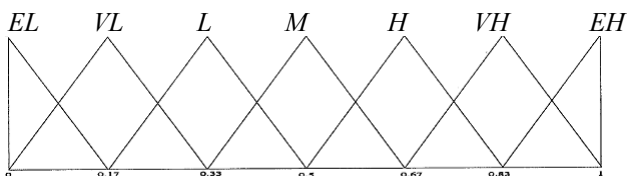


Fig. 2. Example of a term set with 7 labels for translation.

Consider the linguistic assessment of a case whose membership functions are triangular, i.e., a case where $a=b$ in Fig. 1. The membership functions can be represented by a 3-tuple (a,b,c) . Fig. 2. shows an example whose functions are defined as follows.

$$\begin{aligned} EL &= (0,0,0.17) & VL &= (0,0.17,0.33) & L &= (0.17,0.33,0.5) \\ & & & & M &= (0.33,0.5,0.67) \\ H &= (0.5,0.67,0.83) & VH &= (0.67,0.83,1) & EH &= (0.83,1,1) \end{aligned}$$

Each numerical attribute is aggregated with a linguistic attribute generated in the Basic Linguistic term set using the rules, functions and computations involving fuzzy linguistic term. Following definitions give the special functions used in the computation procedure used in the fuzzy framework.

Definition 1. Let $v \in R$ be the set of numerical attributes and $S = \{S_0, \dots, S_g\}$ be the linguistic term set defined, v is aggregated with S by function T as $T(v) = \{(s_0, w_0), \dots, (s_g, w_g)\}$, $s_i \in S$ and $w_i \in [0, 1]$, such that $w_i = \mu_{s_i}(v)$ where μ_s is the fuzzy membership function defined [11].

$T(v) = \{(s_0, w_0), \dots, (s_g, w_g)\}$ be the aggregation representation of the numerical value $v \in R$ over the linguistic term set $= \{S_0, \dots, S_g\}$, a numerical value in the interval $[0, g]$ by means of the function Ψ is obtained as follows.

$$\Psi(T(v)) = \Psi((s_j, \omega_j) \ j = 0, \dots, g) = \frac{\sum_{j=0}^g j \omega_j}{\sum_{j=0}^g \omega_j} = \beta$$

Definition 2: Let $S = \{S_0, \dots, S_g\}$ be a linguistic term set and $\beta \in [0, g]$ be a value representing the result of a symbolic aggregation operation, then we can generate the 2-tuple that expresses the equivalent information to β using the following function:

$$\begin{aligned} \Delta: [0, g] &\rightarrow S \times [-0.5, 0.5] \\ \Delta(\beta) &= (S_i, \alpha), \text{ where } \begin{cases} S_i, & i = \text{round}(\beta) \\ \alpha = \beta - i, & \alpha \in [-0.5, 0.5] \end{cases} \end{aligned}$$

Where round is the usual round operation, S_i has the closest index label to " β " and α is the value of the symbolic translation [11]. Δ is a one to one mapping and $\Delta^{-1}: [0, g]$ is defined by

$$\Delta^{-1}(S_i, \alpha) = i + \alpha = \beta.$$

Let us suppose a symbolic aggregation operation over labels assessed in $S = \{S_0, S_1, S_2, S_3, S_4\}$ that obtains as its result $\beta = 1.6$, then the representation of this counting of information by means of a 2-tuple will be

$$\Delta(1.6) = (S_2, -0.4)$$

Graphically, it is represented as in Fig. 3.

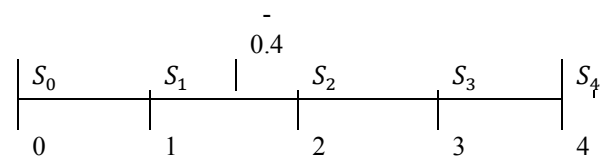


Fig. 3. Example of a symbolic translation computation.

III. CUSTOMER SEGMENTATION MODEL IN CREDIT ANALYSIS

In data mining models segmentation plays a major role in predictions related to function and value analysis, risk analysis, promotional features and various other tasks. In any of these cases the major components in the segmentation model are the customer attributes, their conceptual definition, the data mining models selected and the tasks defined in the prediction process. These segmentation strategies completely support the customer acquisition, profiling, validation, satisfaction, promotion and other dynamic management strategies.

Concept definitions aid in mapping the attributes to the rules in a more effective manner. Rules can be created more effectively with segment models associated. Trained segments or models can be used further in redefining the rules created for more accurate functional or value assessments.

A general model using segmentation rules and functions that is to be applied on the customer attributes is presented with focus on credit analysis. The segmentation model is formed based on the rules created in segmenting the customers using their attributes. An expert knowledge background will aid in creating rules and functions in such situations. The segments thus formed will play major role in various customer value assessments. This model shown in Fig. 4. is used as the base for our proposed fuzzy linguistic framework in segmentation for credit analysis.

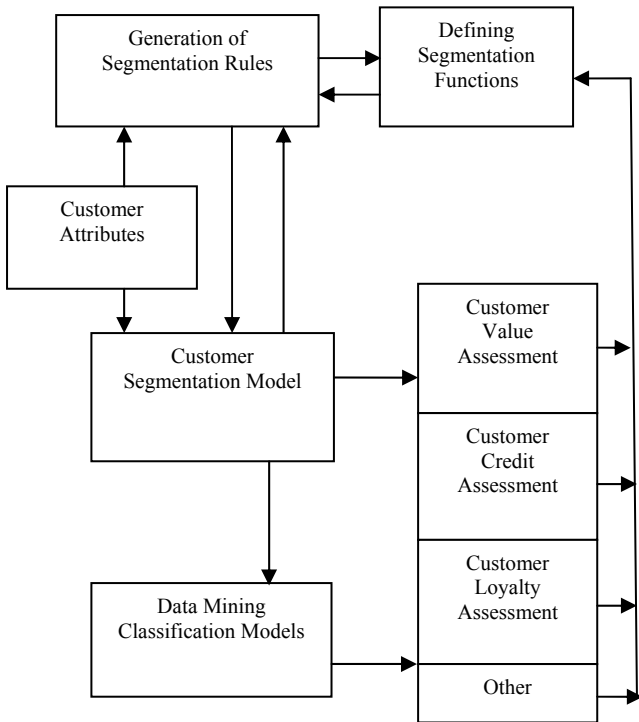


Fig.4. Customer segmentation model in credit analysis

IV. FUZZY COMPUTATIONAL FRAMEWORK FOR SEGMENTATION MODEL IN CREDIT ANALYSIS

The proposed model with basic linguistic terms and fuzzy membership functions segment the customers based on their attributes. A fuzzy model with linguistic approach and computations involving numeric and linguistic attributes are combined with the segmentation rules and the associated member functions. The framework is formed with a series of steps involved in transforming the customer attributes to rules, membership functions and function and value analysis. The steps involved in the computational framework to develop the model are as follows:

1) Select the customer data attributes relevant to the credit analysis. This is performed after an interaction and detailed discussion and analysis with knowledge experts at managerial level who are in the forefront of business. Datapreprocessing can remove the irrelevant attributes from the dataset. Manual intervention can also be done for the removal of attributes with less significance as per expert’s observation, since our plan is to focus on real numerical attributes of the customers in the proposed framework.

2) Generate the Basic Linguistic Term Sets for the customer attributes. The term sets generated should fall within the defined segments. For each of the attributes, term sets are defined and as mentioned earlier in our linguistic approach we select values with odd number of terms such as 3,5,7,9 etc.

3) Perform Segmentation of the selected customer data attributes based on the generated BLTS. Here the limits are set so as to clearly distinguish each of the defined segments. The membership functions are defined based on the limits set in each of the segments. The rules for segmentation can be varied based on the feedback from knowledge experts.

4) Define the membership functions for each term in the BLTS. We can use an appropriate membership function for the definition as per our requirement. A triangular member function in general yields better outcome of results and with more accuracy.

5) Generate membership degree for each term in the term set. Membership degree values are assumed to be between 0 and 1. Greater the membership degree of an instance in a given segment, greater is the confidence in assigning that instance to the segment. It represents the level of uncertainty or vagueness with which each instance belongs to the segment.

6) Aggregate the value generated. Aggregation function is used to represent the linguistic information by a linguistic 2-tuple which consisted of a pair of values (S, α) . S_i is the linguistic term and α is the numeric value representing the symbolic translation [10]. The symbolic translation of a linguistic term, S_i , is a numerical value assessed in $[0.5, 0.5]$ that supports the “difference of information” between a counting of information $\beta \in [0, g]$ obtained after a symbolic aggregation operation and the closest value in $\{0, \dots, g\}$ that indicates the index of the closest linguistic term in $S(i = \text{round}(\beta))$.

To illustrate the steps mentioned above we consider the attribute age from our bank customer dataset under study. As per experts analysis rules are formulated so as to form the required segments. Limits are defined by the experts to categorize them into young age, middle age and old age. It is possible to categorize the age into more number of term sets like 5, 7, 9 etc. Initially we have applied the rules to a subsection of the entire dataset for study purpose. This can be extended to the entire dataset with slight modifications in the range of values selected for defining the membership functions.

For example the old age category in our dataset can be defined as customers who falls within the range 40 and 60. A triangular membership function as shown below is used in our study to define the membership function for old_age.

$$\mu_{S_2}(x) = \begin{cases} 0, & x < 40, \quad x \geq 60 \\ \frac{x - 40}{10}, & 40 \leq x < 50 \\ \frac{60 - x}{10}, & 50 \leq x < 60 \end{cases}$$

According to our definition of age, a customer of age 46 will be considered under the category old_age, linguistic term S_2 and membership degree computed for each of the linguistic term will be $\{(S_0, 0)(S_1, 0.4)(S_2, 0.6)\}$ $(S_2, 0)$ represents the linguistic 2-tuple, betavalue (Δ) computed as $(S_1, 0.6)$ and Δ^{-1} of the numeric 2-tuple $(S_2, -0.4)$.

Fig.5. shows the fuzzy computational framework for customer segmentation in credit analysis.

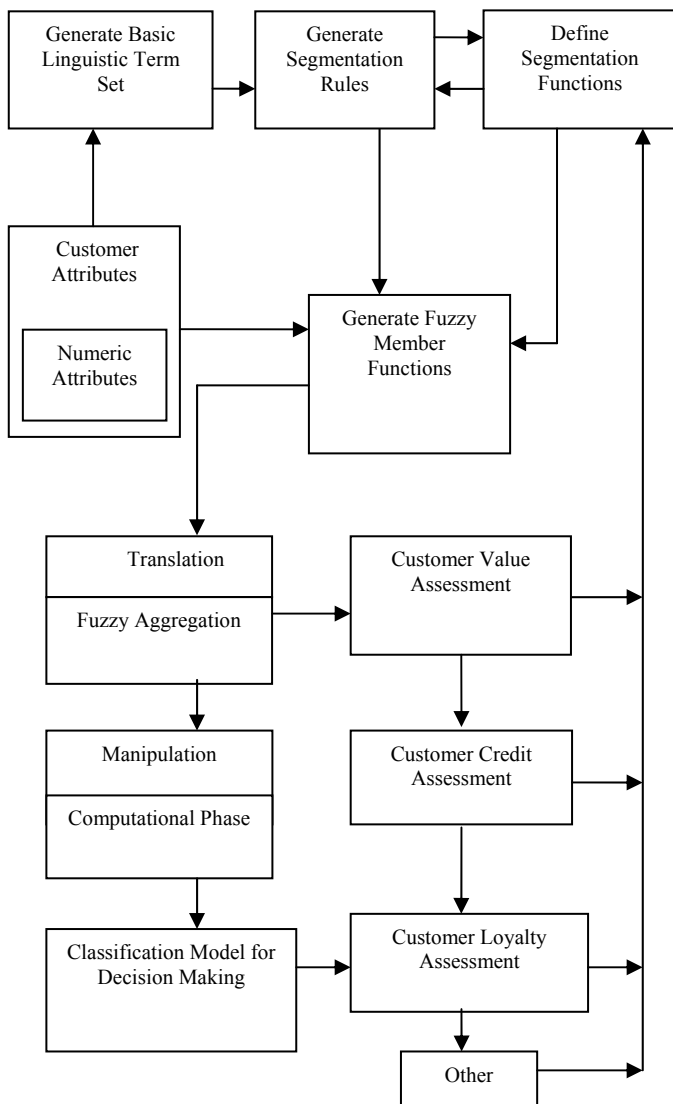


Fig.5. Fuzzy computational framework for customer segmentation model in credit analysis

The model generated can be used with any membership function that supports the semantics of the linguistic terms. We have used triangular membership functions in our example as it produces more accurate results. The main application field of this model is decision analysis and decision making [10]. This model has associated computational models and several aggregation functions based on the previous functions Δ and Δ^{-1} .

V. CONCLUSION

Fuzzy computational framework offers a better solution in areas where human perceptions and ideas play a major role. The ideas can be formulated as rules and functions with experts aid in this framework. In the segmentation approach, the rules are framed by experts in the organization who have better insight into the customer attributes. This results in more

reliable segments which will give better outcome for computing the desired values of the customers. Even though customer segments are formed with defined range or values as selected by the experts, membership function definitions and the aggregation values provide us basic methods to compute the equivalent real values. There exists a series of computation steps and methods to select our preference order to compute these real values to be used in decision making problems.

VI. REFERENCES

- [1] Konstantinos Tsipitis and Antonios Chorianopoulo, "Data Mining Techniques in CRM: Inside Customer Segmentation," Wiley and Sons Ltd, 2009.
- [2] Hugh Wilson, Elizabeth Daniel and Malcolm McDonald, Factors for Success in Customer Relationship Management (CRM) Systems, Journal of Marketing Management, 2002, 18, 193-219.
- [3] Jan-Benedict E.M. Steenkamp, and FrenkelTerHofstede. "International market segmentation: issues and perspectives." International Journal of Research in Marketing 19, no. 3 (2002).
- [4] Brito, Pedro Quelhas, Carlos Soares, Sérgio Almeida, Ana Monte, and Michel Byvoet. "Customer segmentation in a large database of an online customized fashion business." Robotics and Computer-Integrated Manufacturing , Elsevier ,2015.
- [5] Cheng-Lung Huang , Mu-Chen Chen, Chieh-Jen Wang, Credit scoring with a data mining approach based on support vector machines. Expert Systems with Applications 33 (2007) 847–856, Elsevier,2007.
- [6] Pierpaolo D’Urso, Marta Disegna , Riccardo Massari, Linda Osti, Fuzzy segmentation in postmodern consumers Bozen Economics & Management Paper Series NO 20 / 2014.
- [7] Tuma, M. N., Decker, R., & Scholz, S. W. (2011). A survey of the challenges and pitfalls of cluster analysis application in market segmentation. International Journal of Market Research, 53 (3), 391–414,2011.
- [8] Wang, Y., Ma, X., Lao, Y., & Wang, Y. (2014). A fuzzy-based customer clustering approach with hierarchical structure for logistics network optimization. Expert Systems with Applications, 41 (2), 521–534.regarding real-life (Y. Wang, Ma, Lao, & Wang, 2014)
- [9] L.A. Zadeh, The concept of a linguistic variable and its applications to approximate reasoning, Part I, Information Sciences 8(1975) 199–249, Part II, 8, 301–357; Part III, 9, 43–80.
- [10] L.A. Zadeh, Fuzzy logic¼computing with words, IEEE Transactions on Fuzzy Systems 4 (2) (1996) 103–111
- [11] F. Herrera and L. Martínez, "An approach for combining linguistic and numerical information based on 2-tuple fuzzy linguistic representation model in decision-making," Int. J. Uncertainty, Fuzziness, Knowledge-Based Syst., vol. 8, no. 5, pp. 539–562, 2000.
- [12] F. Herrera, E. Herrera-Viedma, and L.Martínez, "A fusion approach for managing multi-granularity linguistic terms sets in decision making," Fuzzy Sets Syst., vol. 114, no. 1, pp. 43–58, 2000.