International Journal of Electrical and Electronics Engineers
Vol. No.8 Issue 01, January-June 2016
www.arresearchpublication.com
IJEEE
ISSN 2321 - 2055

PROPOSED APPROACH FOR CUSTOMER SEGMENTATION USING FUZZY APRIORI TECHNIQUE

Mayur Paithankar[1], Somnath Rashinkar[2], Ajinkya Padawal[3], Prof. J.E.Nalavade[4]

[1,2,3,4] *Department of Computer Engineering, Sinhgad Institute of Technology, Lonavala, India*

ABSTRACT

Customer's requirement is the jumping-off point of productive activities of enterprises, whose satisfaction degree is a key to the success or failure of an enterprise. However, the enterprises may satisfy the customers' diversified and individualized requirements up to the satisfaction level only if the requirement is analyzed accurately. Marketers are constantly looking for ways to improve the analysis techniques which will in turn improve the exactness of the requirement. Marketers try to target customers eligible for particular offers so that they can be attracted back to the store and spend some more time and money on their next visit. This process is called as market segmentation. Market segmentation helps the marketers to understand the needs of the target audience and adopt specific marketing plans accordingly. In this paper, a system is proposed for customer segmentation using association rule mining. Association rule mining is an important approach for Market Basket Analysis. In this paper association rule mining is performed on fuzzy data because the fuzzy partitions are made by considering the nature of itemsets.

Keywords: Association rule mining, Cluster-based fuzzy set table, Customer Segmentation, Fuzzy set, FCM, Fuzzy frequent itemset,Market Analysis.

I. INTRODUCTION

A set-up where two or more parties (also called buyers and sellers) are engaged in transaction of goods and services in exchange of money is called a market. At the market place the sellers sell their goods to the consumers (buyers) in exchange of money. Consider the following examples: 1] Nokia offers wide range of handsets for both males as well as females. 2] The handset for females would be sleeker and more colorful as compared to sturdy handsets for males. Males generally do not prefer stylish handsets. The organizations can't have similar products for all individuals. 3] Perfumes and deodorants for females have a sweet fragrance whereas perfumes for males have a strong fragrance. A marketer can't have similar strategies for all consumers. The process of creating small segments comprising of like minded individuals within a broad market refers to market segmentation. Market segmentation helps in the division of market into small segments including individuals who show inclination towards identical brands and have similar interests, attitudes and perception. Not all individuals have similar needs. A male and a female would have varied interests and liking towards

different products. A kid would not require something which an adult needs. A school kid would have a different requirement than an office goer. Market Segmentation helps the marketers to bring together individuals with similar choices and interests on a common platform.

In most of the business areas, knowledge discovery is the basic necessity to obtain useful knowledge from data stored in large repositories. Since data represent a certain real-world domain, patterns that hold in data help to get interesting relations that can be used to improve the understanding of the particular domain. Data mining is the step in the knowledge discovery process that attempts to discover novel and meaningful patterns in data. One of the best studied models for data mining is that of association rules.

In early days, with the use of association rule mining the frequent itemsets are determined based on the customer transactions of binary data. But now-a-days, fuzzy data are used to determine the frequent itemsets because it provides the nature of frequent itemset ie., it describes whether the frequent itemset consists of only highly purchased items or medium purchased items or less purchased items or combination of all these based on the fuzzy partitions correspond to quantity purchased. In the proposed system fuzzy frequent itemset mining is used for knowledge discovery. The fuzzy frequent itemset mining creates fuzzy partitions for numerical attributes and selects the fuzzy partitions to construct the fuzzy records and create the cluster-based fuzzy set table. Then, it uses cluster-based fuzzy set table, finds the fuzzy frequent itemset and reduces the size of the cluster-based fuzzy set table iteratively. Finally, it concludes with the large fuzzy frequent itemset.

II. LITERATURE SURVEY

The implementation of Mass Customization successfully is challenging as the important question the enterprises encounter is to understand and analyze accurately customer's requirements for the customized products. At present the study of requirement is mostly confined to the requirement investigation and requirement acquisition method, etc., but little has been found on requirement clustering analysis. From the angle of counter market segmentation theory and the production mode, the necessity of the clustering analysis of product requirement is analyzed. In paper [3], with a certain type product as an example, the product requirement has been clustered using the method of fuzzy clustering, whereby a theoretical guidance can be provided for the enterprise to grasp the customers' diversified and individualized requirements accurately.

Fuzzy methods have been applied to data mining and to databases of customer information for marketing. In paper [4], authors Steve Russell, Weldon Lodwick, have explored fuzzy clustering approaches to telecommunications database marketing. Fuzzy clustering methods can be used to mine Telco customer and prospect databases to gain residential and business customer market share. In paper [5], two-step-classification methods are proposed. The aim of this proposed system is to adopt two-stage classification methods, and to apply fuzzy clustering analysis for mining data in the credit market in order to reflect the characteristic type knowledge of common nature of the similar things and different type characteristic knowledge of dissimilar things.

International Journal of Electrical and Electronics Engineers
Vol. No.8 Issue 01, January-June 2016
www.arresearchpublication.com
IJEEE
ISSN 2321 - 2055

III. PROPOSED SYSTEM

The proposed system covers all the steps starting from collecting of raw data to the required analysis results. The whole data mining process includes following steps –

1. Gathering raw data
2. Normalizing
3. Preprocessing raw data
4. Performing Association Rule mining on the pre-processed data.

The block diagram shown in fig 1 shows the flow of generation of frequent itemset using fuzzy logic. The following diagram represents the architecture diagram of proposed system.

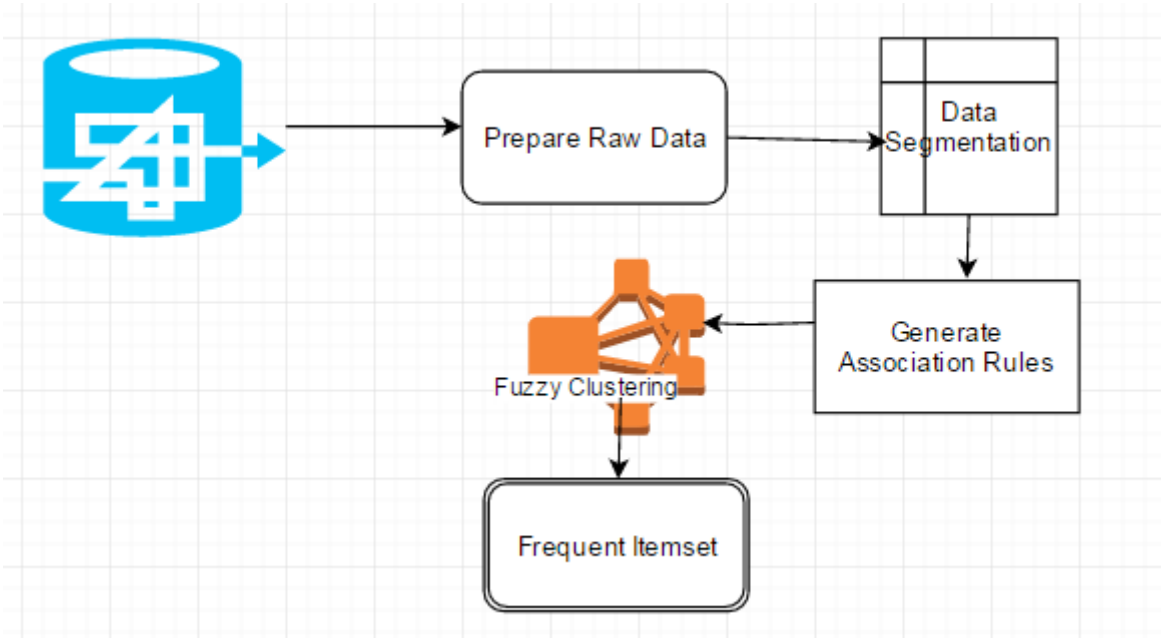

Fig 1. Block Diagram of Proposed System

The diagram in fig 2 shows detailed flow of steps carried out to generate frequent itemset on fuzzy data. Initially after the dataset is loaded, fuzzy partition is created. Then depending upon the number of clusters the partition matrix is calculated. If the sum of the partition is greater than the threshold then fuzzy table is generated. From fuzzy table intersection are analyzed. If the fuzzy intersection is greater than threshold then the particular item item is included in the frequent itemset.

International Journal of Electrical and Electronics Engineers
Vol. No.8 Issue 01, January-June 2016
www.arresearchpublication.com
IJEEE
ISSN 2321 - 2055

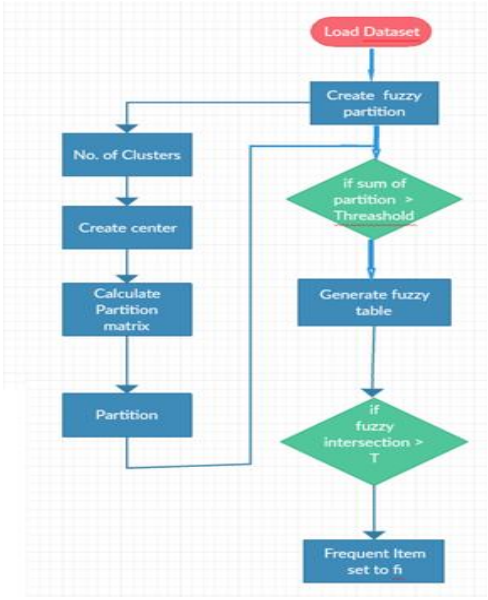

Fig 2. Flow Diagram of Association Rule Mining on Fuzzy Data

IV. TECHNIQUES USED

The proposed system makes use of fuzzy apriori technique to generate frequent itemset. Fuzzy Apriori technique and classical apriori technique are similar as both the techniques count the support of each itemset for finding frequent itemset. The only difference between two techniques is that Fuzzy Apriori technique makes use of fuzzy data. Fuzzy sets are of extension of the classical crisp sets. A fuzzy set A on a universe X is characterized by X − [0, 1] mapping, also called the membership function of A. Therefore, it is necessary to convert the crisp dataset into fuzzy dataset. For converting crisp dataset into fuzzy dataset, first fuzzy partitions are created for each numerical attribute then; multiple fuzzy records are created for each record in the crisp dataset. There is no standard method to create fuzzy partitions. But, FCM is a very popular and established algorithm for fuzzy clustering in various domains. Another difference between Fuzzy Apriori technique and classical apriori technique is that in Fuzzy Apriori technique the support for any itemset is the sum of membership functions over the whole fuzzy dataset.

The proposed system considers only the numerical attributes for frequent itemset generation. Following are the steps for frequent itemset generation.

Step1: Fuzzy partitions are generated using FCM algorithm.

Step2: Fuzzy records are created by considering one of the selected partitions of each numerical attribute and joining them. The partition can be selected by finding the sum of each partition separately and choosing the maximum among them. Then, fuzzy frequent 1-itemset is determined by using the created fuzzy records and simultaneously cluster-based fuzzy set table with count >1 is created by eliminating the fuzzy records whose number of fuzzy value satisfying threshold equal to 1.

Step3: Frequent large itemset is determined by following the modified Apriori algorithm along with cluster based fuzzy set table.

Following is the pseudo code for the frequent itemset generation in the proposed system.

Input: database D, min-support, threshold –membership value

Output: fuzzy frequent large itemset

Main ()

begin

1. Call FCM1

2. Call Modified Apriori ()

End

Procedure FCM1

{

D1= Null;

candidate= { set of numerical items}

for candidate's numerical attribute = 1 to n do

{

For each transaction T ϵ D do

{

1. create fuzzy partitions using FCM

2. sum each partition separately

}

select any one partition with greater sum

D1=D1+selected partition

}

//create fuzzy_cltable

Fuzz_cltable={}

for each transaction T ϵ D1 do

{

candidate.supp is calculated based on threshold membership value

if (no of membership of T >1)

add T into fuzzy_cltable

}

L_1= {candidate | candidate.supp ≥ min-support}

D1=fuzzy_cltable

}

Procedure Modified Apriori ()

{

number of items in a combination (c1) = 1

k:=2

while (L_{k-1} != φ)

{

call_generate_new (c1)

c1++

```
        k++
}
}
Procedure call_generate_new (c1)
{
generate candidates (L_{k-1})
fuzzy_cltable= {}
for each generated candidates c ϵ C_k
        {
        if (c1=1) call frequent(c)
        else
        {
                get all partial subset( c)
                if (all partial subset(c) ϵ L_{k-1} )
                        call frequent(c)
        }
        D1=fuzzy_cltable
        }
}
Procedure frequent(c)
{
c.supp=0
for each fuzzy transaction T ϵ D1
{
        find fuzzy intersection(c) // during fuzzy intersection,
        check each value against threshold -membership value; if value is < threshold -membership value, then
        skip that transaction
        if ( fuzzy intersection(c) ≥ threshold –membership value )
        {
                c.supp++
        }
        add no of membership of T >k to fuzzy_cltable
}
L_k = {c | c.supp ≥ min-support}
}
```

International Journal of Electrical and Electronics Engineers
Vol. No.8 Issue 01, January-June 2016
www.arresearchpublication.com
IJEEE
ISSN 2321 - 2055

V. SYSTEM RESULTS

Following table shows comparison between existing system and proposed system on various parameters.

Analysis Value	Existing System	With Fuzzy Clustering
Number of Rules	426	312
Number of Valid Transactions	300	300
Minimum Support(%)	5,333	5566
Maximum Support(%)	53	55
Minimum Confidence(%)	50	51
Maximum Confidence(%)	91,429	92,567

Table 1. Comparison between Existing System and Proposed System

The graph shown in figure 3 shows comparison between existing system and proposed system by plotting confidence on X-axis and support on Y-axis.

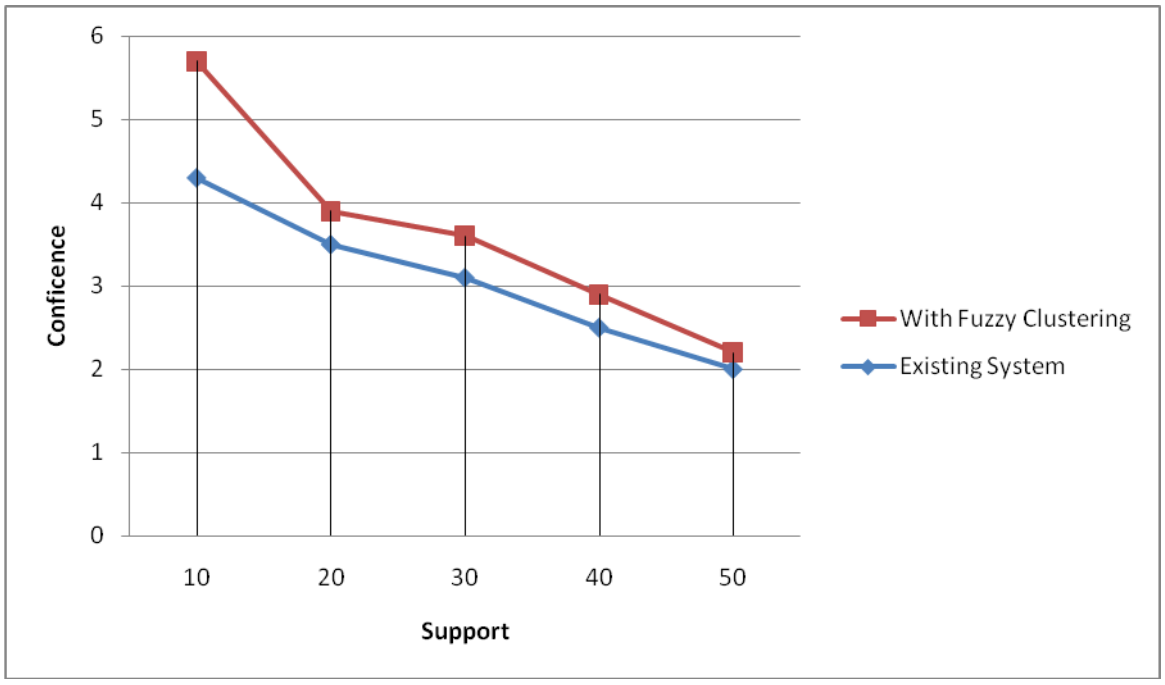

Fig 3. Graph for association rule generation with support and confidence

VI. CONCLUSION

The proposed system performs customer segmentation by using the most important data mining method named as association rule mining. The main purpose of this method is to determine correlations among the sales of items using a set of customer transactions on items. Association rule mining is also known market basket analysis. Market basket analysis helps to understand about the sets of items that are likely to be purchase together. The proposed system will help to explore questions such as "Which products are commonly purchase together?". Thus the dependency between two products X & Y can be analyzed. The results gathered from this analysis can be used for store layout, promotions, discounts, catalog design, etc. The gathered results can be related to customers' data such as age, gender, income, marital status.

REFERENCES

1] Sabri Serkan Güllüoğlu, "Segmenting Customers With Data Mining Techniques", IEEE 2015.

2] R. Prabamanieswari, "A Combined Approach for Mining Fuzzy Frequent Itemset", International Journal of Computer Applications (0975 – 8887).

3] Xuefeng He, Yuewu Jing, "Fuzzy Clustering Analysis of Mass Customized Product Requirement", IEEE 2008.

4] Steve Russell, Weldon Lodwick, "Fuzzy Clustering in Data Mining for Telco Database Marketing Campaigns", IEEE 1999.

5] YE Qian, "Two Stage Fuzzy Clustering Based on Latent Knowledge Discovery and Its Application in the Credit Market", IEEE 2006.

6] Ashish Mangalampalli, Vikram Pudi, "FPrep: Fuzzy Clustering driven Efficient Automated Pre-processing for Fuzzy Association Rule Mining", IEEE Intl Conference on Fuzzy Systems, July 2010.

7] Ashish Jaiswal, Nitin Janwe, "Fuzzy Association Rule Mining Algorithm to Generate Candidate Cluster: An Approach to Hierarchical Document Clustering", IJCSI International Journal of Computer Science Issues March 2012.

8] Fokrul Alom Mazarbhuiya, Muhammad Abulaish, "Clustering Periodic Frequent Patterns Using Fuzzy Statistical Parameters", International Journal of Innovative Computing, Information and Control March 2012.