# Realizing Behavioral Patterns using Fuzzy Logic in Market Basket Analysis

W. Sarada
Assistant Professor, Dep't. of Computer Science
RBVRR Womens College, Hyderabad, Telangana,
Research Scholar, RayalaSeema University,
Kurnool, A.P., India

Dr. P. V. Kumar
(Retd. Professor in Osmania University,
Dept. of CSE, Hyd., Telangana)
Professor, Department of CSE in Acharya Institute of
Technology, Bangalore, India

*Abstract*—**Data mining is an area of research and study within a computer science discipline involving to make out the meaning and interpret the information or data, something that repeats in a predictable way which refer to a design or to customary behavior through any type of calculation that includes both arithmetical and non-arithmetical steps and follows a well-defined model, for example an algorithm, that integrates technology with a plan or intention or an idea or invention to help sell or publicize a commodity in view of such as AI, database systems, ML and statistics. In this paper, the focus is on a new novel approach which scans the market basket database and finds those occurrences of items which are of noteworthy and prunes consecutive item sets which satisfy the support and confidence threshold, interest and generates association rules which are used to locate the way in which two or more things are connected in the vast database that realizes the buy conduct of the client and helps in the expanding of the deals in the grocer's stores or super market and also in choosing the precise territory and the accurate period of gathering in crops, farming and helps in increasing the produce.**

*Keywords—Data mining, item sets; market basket; algorithm; Association rules;*

## I. INTRODUCTION

Data mining is a series of actions or steps taken in finding out or in determining patterns from a big set of information in order to achieve a particular end. It is a repetitive process. Once found, the extracted knowledge can be extended to the user and various measures are used for evaluating, refining, transforming and integrating the data to get the accurate result
by removing unnecessary data and combining from a common source using various DM tools, analyzing the data and making decision, retrieving the relevant information using various techniques and transforming the data into appropriate form through mapping and creating or generating the code, identifying repeated ones based on some measure of quality, value or effect of something, representing using various appropriate tools and produce reports, tables, various types of rules(classification, characterization, discriminant)

## II. LITERATURE REVIEW

In the data mining field, Association Rule Mining (ARM) is apprehensive with uncovering appealing patterns in binary valued data sets. Standard (classical) ARM stand in need of that all elements are binary valued (yes-no, true-false, 0-1, etc.). Of course, in real life, not all fields in the data sets we want to apply ARM are binary valued. The quantitative approach of data about numeric values allows an item either to be member of an interval or not. This leads to, as to perceive (someone or something) as having a lower value, quantity, worth, etc, than what it actually has or is to judge too highly of importance that are accessible to the boundary of such breakable sets. To get over this situation, the process of fuzzy association rules has been developed.

## III. PROPOSED FRAMEWORK

We propose a framework to discover domain knowledge report as fuzzy association rules. Fuzzy association rules are discovered using fuzzy apriori approach without the requirement of the domain knowledge compared to those rules discovered based on the properties of propositional logic called as coherent rules. Fuzzy approach allows the intervals to overlap, making the set fuzzy instead of crisp. Elements of sets are able to appear a restricted belonging to more than one set, overwhelming the "cutting edges issue". . The membership of an item is defined by a membership function (whose value is always limited to between 0 and 1) and fuzzy set theoretic operations are incorporated to calculate the quality measures of discovered rules. Using this approach, rules can be discovered that might have got lost with the standard quantitative approach by finding the frequent item sets with a maximum length by generating Types and Sub types which enables a user to find what are the general purchases and analyze the FMI (i.e. frequent item sets) and SMI (i.e. item sets which are not frequent) depicted through a graph periodically and plan for their future purchase requirement accordingly..

The main difficulties or complexities leading for regularly or frequently behave in a particular way or have a certain characteristic to be are

- The relentless standard of something as measured against other things of a similar kind and pattern of behavior or uniformity among the current methodologies for a Market Basket Analysis is not remarkably huge.

- At present, a large section of the available techniques are not changeable in nature.
- The schedule taken for extracting the appropriate data is huge.
- There exists a determined dismissal of insignificant things in the query items.

## IV. DOMAIN-DRIVEN DATA MINING REQUIREMENTS

By considering this new approach in finding data pattern, a solution towards fulfilling domain-driven data mining requirements can be made by

- Finding the frequent item sets (The data mining process consists of two phases. In the first phase, all candidate item sets (combinations of some items) are found, and support is calculated for each of them. Those whose support is above a certain threshold (minimum support) are called "frequent item sets", and used to find larger item sets) with a maximum length by generating Types (Customers and items)and Sub types:-Under Customer type (a) Top 10 Customers (b)Regular Customers(i.e. who often visits the store), under Type Items- sub types are: (a) FMI (i.e., Fast Moving Items) (b) SMI (i.e., Slow Moving Items)]
- Generate fuzzy association rules (In the second phase of data mining, association rules are formed from the frequent item sets: for each frequent item set, the confidence value is computed for all the combinations of the prefix and postfix of the rule (A and B respectively, which both are distinct subsets of the large item set). The rules that are above the minimum confidence limit are shown as interesting AR's. The significance of every rule is determined by its support and confidence.
- The support is the percentage of records in the database where both A and B occur together. The rules are called strong association rules when they meet or exceed a minimum confidence (min conf) (The confidence is the proportion of documents in the database having B given conf(X -> Y) = supp(X -> Y)/supp(X) = P(X and Y)/P(X) = P(Y | X) Confidence is defined as the probability of seeing the rule's consequent(The final or "then" part of a fuzzy rule) under the condition that the transactions also contain the antecedent(The initial or "if" part of a fuzzy rule.).Confidence is lead and gives dissimilar values for the rules X -> Y and Y -> X. Measuring how many times more often X and Y occur together than expected if they where statistically independent which is called as Lift. The goal is to not only discover interesting relationships between retail products in order to help retailers in identifying cross-sale opportunities but also to ensure easy access.

According to the goal we would like to realize with data processing, in this area or division, many data processing ways are there to settle on from. These ways are loosely classified in supervised learning, unsupervised learning and

market basket analysis. Many input variables in this area or section want to build models that predict a given output variable. Supervised learning strategies either enable just one single or many output attributes. Unsupervised learning doesn't have any output variable however rather tries to seek out structures within the information by grouping the instances into completely different categories.

The designation of market basket analysis is to seek out regularities in information so as to explore client behavior, not like in supervised learning, we have a tendency to don't wish to predict a particular output here, however rather discover unknown structures that exist among a knowledge set. This method orders the instances of a knowledge set into bunches with similar attributes and values. This bunches of things in this area or part is known as clusters. By clusters we have a tendency to mean subsets of the information set that's being deep-mined. Clusters area or territory created within the mining method while not apriori information of cluster attributes.

This method can be used to divide quantitative attributes into fuzzy sets, which deals with the problem that it is not always easy to define the sets apriori. Apriori algorithm is applied on the Fuzzy data. The pseudo code of the algorithm to demonstrate fuzzy association rules is as follows:

```
Main Apalgo(m_sp, m_cf, Db)
1) I=Search (Db);
2) (C₁, T) =modify (Db, I);
3) L=1;
4) (Ck Fq) =scan (Ck, T, m_sp);
5) While (|Ck|≠Φ) do
6) Init
7) inc (L);
8) If L==2 then
9)    Ck=add1 (Ck-1);
10) Else Ck =add1 (Ck-1);
11) Ck=Prune (Ck);
12) (Ck, Fq) =scan (Ck, T, m_sp);
13) F=FU Fq;
14) End;
15) Rules generated(F, m_cf);
```

Fk -Set of frequent k-item sets (having k items) after generating the candidate item sets, the transformed database is scanned in order to evaluate the support and after comparing the support to the predefined minimum support, the items which don't meet the criteria are deleted.

The frequent item sets FK will be created from the candidate items sets CK.

Finally, the association rules are generated from the discovered frequent itemsets providing the following functionality:

- Generate the association rules out of the frequent item sets
- Evaluate discovered or presumed rules with fuzzy support and fuzzy confidence values and display through a graph.

The main purpose of fuzzy sets is to overcome the sharp boundary problem, it is not necessary to be able to enter a single membership function for every fuzzy item set. It is sufficient enough to know where the borders of the fuzzy sets lie.Algorithms used are bubble sort, dijikstra, Brute force, fuzzy c-means clustering, converting crisp data set to

fuzzy dataset, fuzzy apriori. The maximum number of frequent sets and fuzzy association rules that may be generated is given as 5, 00,000 and 10,000.Example data set is properties.txt is used to act as an example.The maximum value for confidence and support is given as 100.0.To ease output, attributes are identified by labels which tend to convey meaning.This is done by loading an output schema files (fruits, groceries etc.) in a database, which is simply one per line for each attribute.The user selects type as customer or items and their sub types- if

customers (a) top 10 customers or (b) regular customers

who often visits the store. The algorithms or modules which have been developed and named as* .java files are (1) MainClass.java depending upon the users input or selection it will retrieve the data accordingly (2) Test.java (3) CombinedXYPlotDemo1.java for displaying graph

depicting the SMI or the infrequent itemsets are the slow moving itemsets and the FMI or the frequent itemsets or fast moving itemsets (4) Helper.java (5) ConnectionHelper.java to establish database connection and other related java files.

The 3 phases of methods of measure are specified below:

Phase I. PAA approach for finding out the predominant consortium of rules, recast the ordering to restore AR's, retrieves without human intervention , an apportionment of accepted principles, reveal and put to use this interrelation between supervision found in step one, scrutinize the intent unaffectedly and partition the classification of exchange of instructions achieved in the above rules of step two. Phase II. PAARMA approach which link evaluation to deals put forward for consideration. PAARMA constitutes into 2 sections, PAARMA-1 is used for controlling the minimum support count and discover the rules with the highest supports. The minimum support check meets the littlest measure as substitution to enable a level of quality or attainment amongst the point prevailing the action or process making it bigger size or amount of transactions. PAARMA-2 supports and extracts rules just for one attainable thing which distinguishes from CBA-RG(Classification Based on Associations-Rule Generator), within a way to ease, extract diverse levels of quality internal to a certain span which strives to, convey successive leads for attaining maxRulenum in advance, it just ends its implementation and returns the instructions it has extracted and Phase III. MAARMA approach makes the consumer more convenient in making constructive purchases by incorporating the equivalence amid rules as well as confidence, the purchasers will Place the most related items within the appropriate places. Therefore, it will expand the sales as well as moreover it will be more comfortable for the consumers.

To survey proposed approaches, various datasets such as Synthetic, Departmental, Real, Mushroom, supermarket, Production-crop datasets are utilized.

WEKA has also been used for numerical examination. The performances of the proposed methods are evaluated in view of the following elements. They are AUROCC- for checking, visualizing and representing a degree of separability, if it is 1, better is the predicted model and vice-versa otherwise, Accuracy-Sensitivity and Specificity- inversely corresponds to one another depending up on the increase or decrease in the threshold, and Execution time.For ex:-

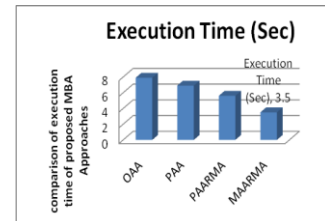| Proposed MBA Approaches | Execution Time (Sec) |
|---|---|
| OAA | 7.9 |
| PAA | 6.9 |
| PAARMA | 5.6 |
| MAARMA | 3.5 |



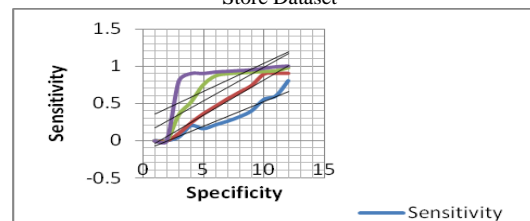Table 1.0 and Fig.1.0 Comparison of Execution Time in Departmental Store Dataset



Fig.1.1 Comparison of AUC value within Departmental Dataset

From the above table 1.0, fig 1.0 and fig.1.1, it can be observed that in the Departmental stores dataset, the AUC value of standard OAA approach is low, whereas PAA is fairly adequate, the AUC value of PAARMA approach is good and that of the third proposed MAARMA approach is excellent.

These elements are central within Market Basket Analysis since one can choose about the execution of the proposed method within association amid exchange approach. These are the basic cause of changes within the market basket evolution or improvement and are elucidated expeditiously.

## V. CONCLUSION

This exploration rules effective proposal of the approach for differentiating the repeatedly acquired things and its suitability for deals in grocer's shop and during harvesting when used to perform the operations simultaneously. The recommended approach is useful for both the stores and the horticulture. In future if combined with big data or cloud using tools may prove to be effective in spite of their limitations.

## REFERENCES

[1] An, X. & Wang, W. (2010). Knowledge management technologies and applications: A literature review. IEEE, 138- 141. doi:10.1109/ICAMS.2010.5553046

[2] Md. Zahid Hasan, Mohiuddin Ahmed and Md. Elias Mollah Green University of Bangladesh, Data Mining Research Challenges in ECommerce. International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 3, May - 2012

[3] G. Wang, J. Hao, J. Ma and L.Huang, " A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering", Elseviers journal of Expert Systems with Applications, Volume 37, Issue 9, page 6225-6232, September 2010.

[4]  K. Sudheer Reddy, G. Partha Saradhi Varma and S. Sai Satyanarayana Reddy Understanding the Scope of Web Usage Mining & Applications of Web Data Usage PatternsIEEE International Conference

[5]  S. Venkata Lakshmi, K. Hema, 2015, Applications of data mining in knowledge management, international journal of engineering research & technology (IJERT) NCACI – 2015 (Volume 3 – Issue 18)

[6]  Shivali, Joni Birla, Gurpreet, 2015, Knowledge Discovery in Data-Mining, international journal of engineering research & technology (IJERT-2018) NCETEMS – 2015 (Volume 3 – Issue 10),

[7]  P . Madhura, M . Padmavathamma, 2014, A Study on Datamining Techniques in Personalised Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT-2018 July) NCDMA – 2014 (Volume 2 – Issue 15),

[8]  J. Hsu,"Rise of Data Mining: Current and Future Application Areas", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011 ISSN (Online): 1694- 0814.

## AUTHORS PROFILE

W.Sarada is a research scholar in Computer Science at Rayalaseema University, Kurnool, Andhra Pradesh,India., working as an Assistant professor, in the Dept. Of computer science, RBVRR Womens college, Narayanaguda, Hyderabad,Telangana,India. Her interests include Data Mining, Computer Networks and Software Engineering.

Dr .P.V.Kumar is a retd. professor at University College of Computer Science and Engineering, Osmania University, Hyderabad. He is currently working as Professor in Department of CSE in Acharya Institute of Technology, Bangalore.He has vast experience in teaching, guiding and in administration. He is a research supervisor/guide to M.Tech, M.Phil and PhD students.