IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

VOLUME : 2 ISSUE : 2 Print / Issue Publication Date: 27-Jan-2017

ISSN : 2455-2143

Indexed In

Google SCHOLAR OPEN ACCESS doi Crossref

WWW.IJEAST.COM

editor@ijeast.com

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 2, Issue 2, ISSN No. 2455-2143, Pages 17-21
Published Online December-January 2016 in IJEAST (http://www.ijeast.com)

FILTERS AND PRUNING TECHNIQUES ON ASSOCIATION RULES IN DOMAIN ONTOLOGY RELATIONSHIP

B.Sakthi Karthi Durai,
Assistant Professor
Department Of Computer Science and Engineering
PSNA College of Engineering and Technology
Kothandaraman Nagar, Dindigul-624622
Tamil Nadu, India

Abstract- Association rule mining plays an important role in data mining. But the rules discovered by association rules are huge amount. This makes user to reduce the interaction to deliver interesting rules for a fuzzy association. The major drawbacks are redundancy and uninteresting to user to handle large dataset and frequent itemsets. In Data Mining, the usefulness of association rules is strongly limited by the huge amount of delivered rules. To overcome this drawback, several methods were proposed like Apriori algorithm, tree pruning, itemsets concise representations, redundancy reduction. This Paper propose Apriori algorithm, post processing steps and filtering rules. First, we propose Apriori algorithm for handling large dataset to extract the rules fuzzy database. Secondly, post-processing steps proposes an interactive process of rule discovery to the user. Finally, by propose Pruning and Filtered rules, to reduce the number of rules. This three processing steps will produce an effective association rules to the user. The algorithm which is described in proposed system is evaluated.

Keywords: Data Mining, Association Rules, Fuzzy Logic, Interaction Processing, Knowledge Discovery Management.

I. INTRODUCTION

Generally, association rule mining will generate a large number of association rules. In this paper we use fuzzy database which will be more complicated than normal database. Fuzzy: - means uncertainty. In existing system, they did not mention about the fuzzy data. This paper proposes fuzzy database, which has large amount of item and those items are not described in a certain way. For example consider a stock market which has the details of stock item details, supplier details, customer details, purchase details, payment details which will be presented properly and which has unique id. If we search from our own database or online database, they don't have unique id (or) not well presented in database. Then while searching the result would produce redundancy result.

The representation of user knowledge is an important issue. The more the knowledge is represented in a flexible, expressive, and accurate formalism, the more the rule selection is efficient. In the Semantic Web field, ontology is considered as the most appropriate representation to express the complexity of the user knowledge, and several special languages were proposed.

Reduction number of association rules by closed [5], nonredundant rules [2] and pruning techniques [4] and optimal itemsets[7] will use pruning, summarizing, grouping and visualizations. Moreover the rule discovery programs have been classified into those that find association rules and those that find the qualitative rules and qualitative laws. MAFIA: A Maximal Frequent Itemsets algorithm was designed to support counting the combines a vertical bitmap representation of the data and filtering technique with efficient bitmap compression scheme [10].

Definition 1. [3] **Bayes *theorem***: relates the conditional and marginal probabilities of event X and Y, and provide that the probability of Y does not equal to zero

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)}$$

Bayes theorem is used to find the probability of occurrence in the data items between them.

Definition 2. [12]. A Fuzzy database relation, R is a subset of the set of cross product

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 2, Issue 2, ISSN No. 2455-2143, Pages 17-21
Published Online December-January 2016 in IJEAST (http://www.ijeast.com)

$$2^{\times 1} \times 2^{\times 2} \times \ldots\ldots \times 2^{\times m},\ where\ 2^{\times j} = 2^{\times j} - \alpha.$$

A fuzzy tuples is a member of a fuzzy database relation as follows.

Definition 3. [12].Let $R \leq 2^{\times 1} \times 2^{\times 2} \times \ldots\ldots \times 2^{\times m}$ be a fuzzy database relation. A fuzzy tuples t (with respect to R) is an element of R.

Even though the fuzzy relational database consider components of tuples as set of data values from the corresponding domains, by applying the concept of equivalence, it is possible to define a notion of redundancy which is similar to classical relational database theory.

II. SYSTEM DESCRIPTION

Increase in number of available ontology covers a wide domain of applications which causes a great advantage in an fuzzy association rules in domain ontology in large database. This paper consists of several steps to reduce the number of association rules. Thus, we extend the specification language proposed by Claudia et al. Interactive Post mining process (IPP), General Impressions (GI) and Reasonably Precise Con will be used in ontology concepts.

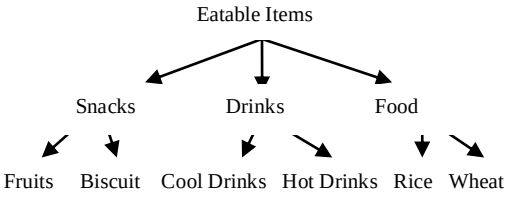

Fig 1. Food Items in Supermarket item taxonomy

Example: In case of General Impressions, the user might believe that there exist some association between c*ool drinks* or *hot drinks*, Snacks item and food (assume that the user uses the taxonomy in Fig. 1). User could specify the user beliefs using General Impressions:

$$gi(<\{hot\ drinks\ or\ cool\ drinks\}^*, Snacks+,\ food>).$$

The examples of association rules that are conformed to the specification rules are as follows.

$$biscuit \rightarrow food$$
$$fruit, biscuit, rice \rightarrow cooldrinks.$$

While working with these database items, developing specification becomes a complex work. Moreover, this will be very useful for the user to introduce in the GI language. For example, in the market case, it would be easy to find the customers buying snacks, also buy food. In order to select this type of rules, the user should be able to create RPC:

$$rpc(<Snacks \rightarrow Food>),$$

where Snacks and Food represents, the set of item related to snacks and those item which are produced in food, thus this concept will not be possible in using taxonomies.

Starting from Eatable Items in Fig. 1, (that we developed an ontology based on the earlier onee organize Boolean type specification that snacks (Issnacks),and those that are food (IsFood). Designing ontology, allows concept definition using restriction on properties. Snacks are defined as a restriction on Food hierarchy using the data property. Similarly we define *food* concept. In our example, *oil food* and *vegetable* items are *food* and *fruit*, *cool drinks* and *vegetable* items are *snacks*. In Fig. 2, we present the structure of the ontology resulting after applying a reasoner.

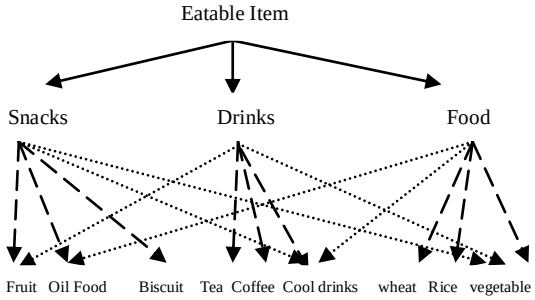

Fig 2. Generating ontology in Supermarket

2.1Fuzzy Association Rule and Post mining Steps

The Proposed work is based on Association Rule mining with fuzzy rule and Post processing steps from user goals. At first, user set the normal knowledge and goals. From domain knowledge view over the user knowledge in database domain and user expectations express. The user will set the prior knowledge over the discovered rules. In this we will get many number of fuzzy association rules. To reduce this association rule we post processing step, in this step pruning technique we consider tree pruning (is detailed in) as one of the best solution for our method. In our process we consider the following steps to be important Formalizing user knowledge and goals, a set of filters, Interactive Post Processing steps with tree pruning process. The following steps are suggested to the user shown in the Fig. 3.

1. *User Driven Fuzzy Database*: Describes how the data Items are set in the database and relationship between them. Where fuzzy is uncertainty;

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 2, Issue 2, ISSN No. 2455-2143, Pages 17-21
Published Online December-January 2016 in IJEAST (http://www.ijeast.com)

2. *User Ontology knowledge*: It starts from the database or selects from the existing database for the user developing ontology on database itemsets;
3. *Applying operator and visualizing the results*: First the operators are applied over the rule schema and then filtered association rules are proposed to the user.
4. *Filters and Pruning Techniques*: first the filters are applied over the rules whenever the user needs them with the main goal of reducing the number of rules In Pruning technique we use tree pruning method in order not to take the unwanted data. This was not found in our result. So other results are considered and the technique searches from them. The Apriori algorithm, (step 1) is used to find the frequent 1-itemsets, L1. (In step 2 to 10), L_{x-1} is used to generate candidates C_x in order to finds L_x in which $x \geq 2$.

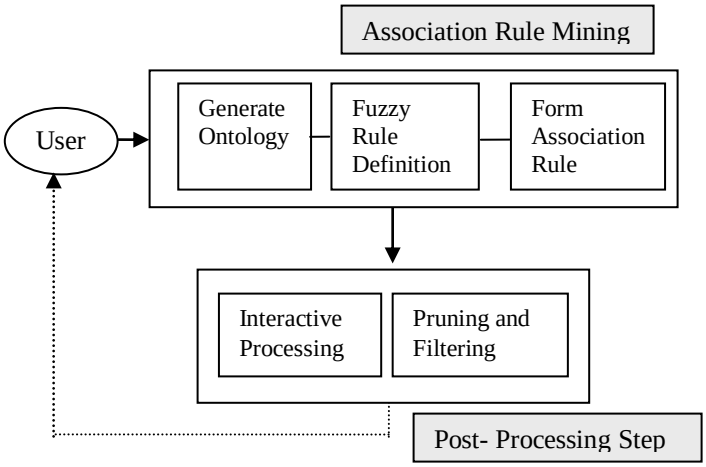

Fig 3. Interactive Fuzzy Post Process Description

The apriori_gen procedure generate the candidates and then uses the Apriori property to eliminate those having a subset that is not frequent in (step 3) . Once all of the candidates have been generated, the database is scanned (step 4) .

1) $L_1 = \{ frequent1-itemsets\}$
2) $For(x = 2; L_{x-1} \neq \Phi; x++)\, do\, again$
3) $C_x = apriori - gen(L_x - 1);$
4) $Forall\, transaction\ t \in D\, do\, begin$
$t = add - ancestor(t, T)$
5) $C_t = subset(C_x, t)$
6) $Forall\, candidates\, c \in c_1\, do$
7) $c.count++;$
8) *end*
9) *end*
10) $L_x - \{c \in C_x | c.count \geq \min \sup$
11) *END*
$Answer = \bigcup_x L_x;$

Fig 4.Apriori Algorithm in our process

For each transaction, a subset function is used to find all subsets of the transaction that are candidates (step 5). And the count for each of these candidates is accumulated (steps 6 and 7). Finally all of those candidates satisfying minimum support (step 9) form the set of frequent itemsets, L (step11).
A Procedure can then be called to generate association rules from the frequent itemsets. For selecting a maximal frequent itemsets algorithm which used pruning techniques such as pruning and checking to reduce the search space. All confidence, any confidence and bond constraint are limited one by selection according to the user searching items.

III. EXPERITMENTAL RESULTS

1) In our experiment, let us consider a medical data search which will have different company products but those products will produce the same tablet (or) syrup for the disease. If we search to get different tablets or syrups or injection dose for curing any disease, we can get the particular medicine. The database which doesn't have proper definition will lead to huge association rule. To reduce this association rule, let us introduce the post processing steps. At first admin will generate ontology and set the fuzzy rule definitions.
While talking about generating ontology admin will enter all the records with no proper definition. In fuzzy rule it will set true or false. At the time of scanning, if keyword is found, it will give 'yes' result otherwise it will give 'no result. The result found in the fuzzy rule is said to be

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 2, Issue 2, ISSN No. 2455-2143, Pages 17-21
Published Online December-January 2016 in IJEAST (http://www.ijeast.com)

association rule. After the association rule formation it will let to Pruning Rule Schema (PRS) and Filtering Rule Schema (FRS). PRS is a filtering process after pruning the result. The unwanted data can be reused for filtering. While FRS is a filter process in which unwanted data cannot be reused for filtering again. In the medical data, while searching for a name of a medicine for a particular disease the result given is association rule. This is got from the huge amount of data. Pruning will select if it is a tablet or a syrup or an injection dose. And result of our process will reduce the target specification and the time required is shown in the Fig 5.

2) Tree Pruning Process

These methods address this problem of over the association rule fitting the data. Such methods typically use statistical measures to remove the least reliable branches. An unpruned tree (Fig 6) and a pruned version of it are shown in (Fig .7). Pruned trees tend to be smaller and less complex and thus easier to comprehend. Searching process in a pruned tree is usually faster and better than in an unpruned decision tree. In the Prepruning approach, a tree is pruned by halting its earlier construction by deciding not to split subset of training tuples at a given node. Post Pruning removes subtrees from a fully grown tree to eliminate the unwanted data.

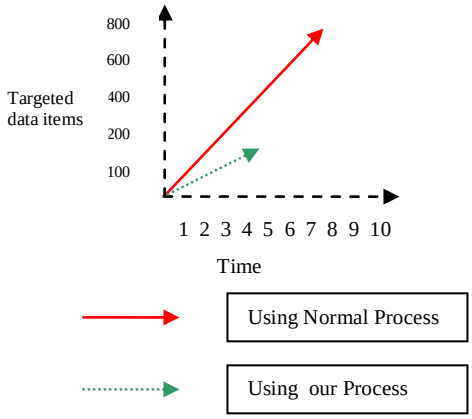

Fig 5. The Measured Output

A subtree at a given node is pruned by removing its braches and replacing it with a leaf.

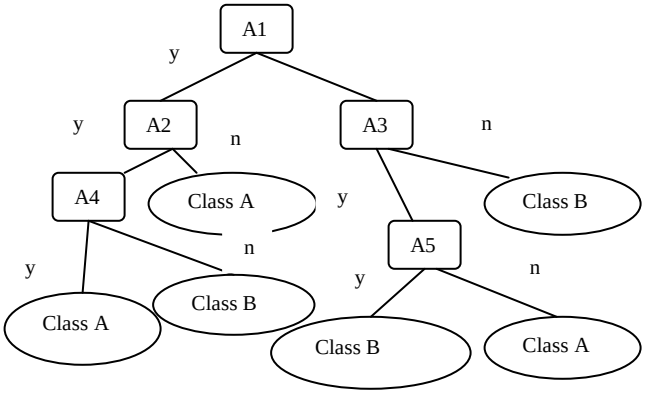

Fig 6.An unpruned decision tree

Generating all significant association rules between items in the database was presented in efficient algorithm for buffer management, novel estimation, filtering and pruning in large database. While saying about objective measures it say about practical item sets and subjective measure says about the theoretical item sets. Algorithm used here is automatic using fuzzy rule set. Pruning techniques are used to reduce the redundancy of data in database. Pruning can be classified into two types 1) Post Pruning: Let the user search the full database and discard unwanted item sets. 2) Pre Pruning: will stop the searching from database when the information becomes unwanted. Post Pruning is preferred in practice to search the data, prepruning data stop earlier. Fuzzy concepts do not prepare a clear orientation.

In this paper we provided a solution to solve fuzzy association rule, using the algorithm and post mining process. And output of our result shown in (Fig .5) describes the x axes as time and y axes as targeted data items using our algorithms, fuzzy association rules, user ontology ,filter and pruning (tree pruning) steps to reduce the data items and reduce the time. The output clearly shows the normal process which searches more time and data items and large amount of fuzzy association rules, considering the processing steps which get less amount of time to process less amount of data items.

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 2, Issue 2, ISSN No. 2455-2143, Pages 17-21
Published Online December-January 2016 in IJEAST (http://www.ijeast.com)

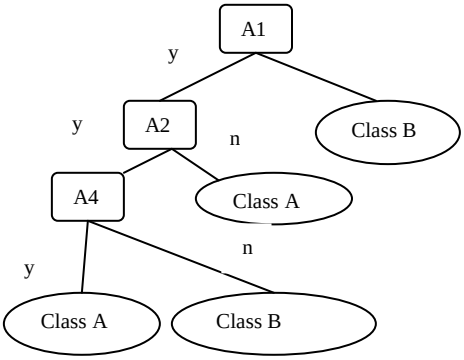

Fig 7.A Pruned Version

IV. CONCLUSION

By applying the new approach over the large database, will allowed to integrate the domain expert knowledge in the post processing step in order to reduce the number of rules to several dozens. Moreover, the quality of the filtered rules was validated by the expert throughout the interactive process.

V. REFERENCES

[1] R. Agrawal, T.Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases," – 1993 Proceedings of the 12th ACM SIGMOD, *pages* 207-216.

[2] Claudia Marinica and Fabrice Guillet "Post-Processing of Discovered Association Rules using Ontologies". 2008. Proc.IEEE International Conference on Data Mining Workshops ICDMW'08. Pages: 126-133.

[3] A. Maedche and S. Staab, "Ontology Learning for the Semantic Web," 2001. IEEE Intelligent Systems pages. 72-79,

[4] Key-Sun Choi and KAIST, Daejeon," IT Ontology and Semantic Technology" Natural Language Processing and Knowledge Engineering, 2007. **page(s):** 14 – 15.

[5] Jiuyong Li Hong Shen Topor, R." Mining the smallest association rule set for predictions " Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference" on 2001. **page(s):** 361 – 368.

[6] Mazid, M.M. Ali, A.B.M.S. Tickle, K.S." A Comparison Between Rule Based and Association Rule Mining Algorithms" Network and System Security, 2009. NSS '2009. **page(s):** 452 – 455.

[7] Klemettinen, M., H. Mannila, P. Ronkainen, H.Toivonen" Finding Interesting Rules from Large Sets of Discovered Association Rules."1999. *International Conference on*
(CIKM), pages 401-407.

[8] Jiang, D. Pei, J. Zhang, A. An interactive approach to mining gene expression data. 2005. Knowledge and Data Engineering, IEEE Transactions on **page(s):** 1363 – 1378

[9] Qi Yang Weining Zhang Chengwen Liu Jing Wu Yu, C. Nakajima, H. Rishe, N.D." Efficient processing of nested Fuzzy SQL queries in a fuzzy database "Knowledge and Data Engineering, IEEE Transactions on 200. **page(s):** 884 – 901.

[10] Feng Hao Daugman, J. Zielinski, P." A Fast Search Algorithm for a Large Fuzzy Database" 2008. **page(s):** 203 – 212.

[11] Bayardo, R.J. Jr. and R. Agrawal "Mining the most interesting rules" 1999. *Proc.in Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Pages:* 145 – 154.

[12] Rolly Intan and Masao Mukaidoro "Conditional Probability Relations in Fuzzy Relational Database" Rough Sets and current Trends in Computing in 2001 Pages: 251 to 260

IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

ABOUT IJEAST

International Journal of Engineering Applied Science and Technology (IJEAST) is a peer-reviewed, open access journal that publishes high-quality research papers in the field of Engineering, Applied Science and Technology.

IJEAST aims to provide a platform for researchers, academicians, and professionals to share their innovative ideas, research findings, and practical experiences with the global scientific community.

FOCUS AREAS

- Engineering
- Applied Science
- Technology
- Innovation & Development
- Interdisciplinary Studies

PEER REVIEWED
All submissions are rigorously peer reviewed to ensure quality.

OPEN ACCESS
Free and unrestricted access to research for all.

GLOBAL REACH
Connecting researchers and professionals worldwide.

TIMELY PUBLICATION
We ensure a swift and efficient publication process.

For more information, visit our website
www.ijeast.com

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

editor@ijeast.com
www.ijeast.com
India

2455-2143