

Classifying Association Rules with Minimized Sets using Fuzzy-Apriori Algorithm

Sonam Jain
Computer Science Engg
Shri Vaishnav institute Technology & Science
Indore

Anand Rajawat, PhD
Computer Science Engg
Shri Vaishnav Institute Technology & Science
Indore

ABSTRACT

Using Association Rule mining is extremely well-organized method for getting strong relation between correlated data or information. The correlation of data provides significance complete taking out progression. For the mining of positive and negative rules, a range of algorithms are utilized for example Apriori algorithm and tree based algorithm. A numeral of algorithms is be unsure presentation but manufactures huge number of negative association rule and also goes through from multi-scan difficulty. The proposal of is to get rid of these difficulties and decrease huge amount of negative rules. Here an efficient technique is implemented for the classification of association rules generated using Fuzzy-Apriori algorithm and classification of these rules can be done supervised learning such as Naïve Bayes Algorithm. The proposed methodology implemented here provides efficient results as compared to the existing technique implemented for the generation of association rules.

Keywords

Association Rules, CART, Naïve Bayes, Decision Tree, Frequent Item sets, In-frequent Item Sets, Positive Rules, Negative Rules.

1. INTRODUCTION

The mining association rule is a data mining undertaking that aspires to determine connections between things in a transactional record. This assignment has been considered extensively in the literature for its advantage in many application areas, for example Web usage mining, recommender methods, and intrusion detection. It arrangements with the market basket investigation to discover frequent patterns and produce association rules of the form $A \Rightarrow B$, which can be utilized to calculate that “If a customer buys itemset A, he/she will a large amount possible buy itemset B as well”, where A and B are frequent itemsets in a operational record and $A \cap B = \emptyset$. The rule of the outline $A \Rightarrow B$ is measured a strong rule if its support (s) and confidence (c) assure minimum support (mins) and minimum confidence (minc) thresholds; these are called positive association rules (PARs). Besides to PARs, the association characterized as “customers that buy itemset A do not buy itemset B” pass on to a negative relation between two itemsets [1].

Data mining is the method of removing useful information from records. Many come within reach of to sequential information mining have been suggested to remove useful information, for example time series analysis, sequential association rules mining, and sequential pattern detection. Numerous important methods that are used in data mining explain the kind of mining and data recovery process. Data

mining or machine learning method is used to calculate group membership for data occurrence. Such as, an important person may possibly objective to use classification to forecast whether the weather on a particular day will be “sunlit”, “wet” or “hazy”. Well-liked categorization methods comprise decision trees and neural networks. For the categorization of web log information and study the importance consumers from them. Due to the participation of unresponsive users in the web log, the original log cannot be used as a procedure in the web usage mining technique for the reason that of this, the first phase or period of the web log data is renewed, to force from the concerned data and after that to continue with the removed information. During this stage, the definite size of the record will be reduced to confident amount. The second phase consists of separating the data using CART algorithm [1].

The most important mining algorithm found on association rule, Apriori not only influenced the association rule mining group of people, but it affected other data mining fields also. Genetic Algorithms are a people of computational representations motivated by progression. These algorithms are set a potential solution to a definite difficulty on a uncomplicated genetic material-like data structure and be appropriate recombination operatives to these arrangements as to protect important in sequence. Genetic algorithms are consistently examined as function optimizer, even though the varieties of predicaments to which genetic algorithms have been useful are moderately extensive. In this paper they are proposing Apriori and Genetic Algorithm (GA) based frequent itemsets mining for network traffic.

Association rule mining is a technique to recognize the concealed information in huge instances record and illustrate interventions on how separations of items manipulate the subsistence of other subsets. Association rule mining aspires to determine well-built or attractive relation between attributes. All simplified frequent pattern sets are not very proficient for the reason that a segment of the frequent pattern sets are surplus in the association rule mining. This is why this algorithm constructs some uninteresting rule along with the attractive rule. This difficulty can be defeat with the help of genetic algorithm. A large amount of the data mining methods use the greedy algorithm on behalf of genetic algorithm. Genetic algorithm is formed by optimized consequence as evaluate to the greedy algorithm for the reason that it completes a widespread investigate and enhanced attributes interaction [3]. In genetic algorithm population development is created. Genetic algorithm is an organic method which uses gene as a constituent on which explanations (individuals) are influenced. Usually association rule is used to discovery positive relationship involving the data set. Negative association rule is also very

important in investigation of intelligent information. Negative association rule mining is assumed where a domain has too several issues and large amount of infrequent pattern sets in operation record. Negative association rule mining occupations in turn around way and it describe decision making competence, whether which one is significant as an alternative of confirming all rules. On the other hand difficulty with the negative association rule is it utilizes enormous space and can take more time to manufacture the rules as evaluate to the predictable mining association rule. In the simplified association rule folder is examinee once and operation is altered into space decreased arrangement. The association rule mining difficulty can be decayed in statistical and categorical attributes in a record. The application of association rule mining is utilized to examine a mixture of circumstances like market basket analysis, banks, weather prediction, pattern reformation, and multimedia information.

1.1 Frequent Item sets Mining

Frequent patterns are itemsets, subsequences, or substructures that appear in a data set with frequency no less than a user-specified threshold. For example, a set of items, such as milk and bread that appear frequently together in a transaction data set is a frequent itemset. A subsequence, such as buying first a PC, then a digital camera, and then a memory card, if it occurs frequently in a shopping history database, is a (frequent) sequential pattern. A substructure can refer to different structural forms, such as subgraphs, subtrees, or sublattices, which may be combined with itemsets or subsequences. If a substructure occurs frequently in a graph database, it is called a (frequent) structural pattern. Finding frequent patterns plays an essential role in mining associations, correlations, and many other interesting relationships among data. Moreover, it helps in data indexing, classification, clustering, and other data mining tasks as well. Thus, frequent pattern mining has become an important data mining task and a focused theme in data mining research.

Frequent itemsets mining is a center part of data mining and differences of association analysis, like association-rule mining and sequential-pattern mining. Removal of frequent itemsets is a center walk in numerous association analysis methods. An itemset is known as frequent if it at hands in a huge an adequate amount of portion of the dataset. This frequent happening of item is communicated in conditions of the support count. Consequently, it requires make difficult methods for hiding or improvement users' private information for the duration of a data collecting procedure. Additionally, these methods should not submit the exactness of mining results [4].

The frequent itemsets are patterns or items like itemsets, foundations, or subsequences that come out in a data set regularly or quickly. Such as, some frequent words or information that replicated frequently in a data set can be pleased as frequent itemset for that data set. A subsequence as purchase a digital camera trailed by Akash tablet and then a memory card if it happens frequently in a shopping record. It is known as (frequent) sequential prototype. In the same way underneath is submitting to different structural appearances, like sub-trees, sub-graphs or sub-lattices, which may be combined with itemsets or subsequences. If a foundation happens repeatedly, it is called a (frequent) structured model. Detection such frequent pattern participate a significant responsibility in

mining relations, associations, and many other application relationships down with data. Moreover, it assists in data clustering, classification, and other data mining tasks also [5].

The process of finding out attractive and unforeseen rules from big or huge data sets is called as association rule mining. It is a suggestion or if-then-rule which is sustained by information. The association rule difficulty [6] was first discovered and was generally used in investigation in super markets, called the difficulty of market-basket. The beginning difficulty was the subsequent given that a set of items and an enormous group of sales records, which consists of operation date, the items that bought in the operation, the occupation is to discover connections between the items enclosed in the different transactions [7].

1.2 Positive and Negative Association Rules

Association rule mining is a data mining undertaking that determines affiliations between items in a transactional record. Association rule study is the assignment of determining association rules that happen normally in a given data set. A distinctive illustration of association rule mining application is the market basket investigation. In this procedure, the behavior of the customers is considered when obtaining different item for consumptions in a shopping store. The detection of interesting models in this set of data can show the way to significant marketing and management considered decisions. Such as, if a customer buys bread, what is the chance that he/she buys milk as well? Depending on the probability of such an association, marketing human resources can enlarge better arrangement of the projection space in the store or can found their concession approach on such associations/correlations found in the information.

Negative Association Rules is described take a broad view negative association rule, a rule that controls a negation of an item (i.e. a rule for which its antecedent or its resulting can be created by a combination of existence or nonexistence of terms). An illustration for such association would be as follows: $A \wedge \neg B \wedge \neg C \wedge D \rightarrow E \wedge \neg F$. To the most excellent of our acquaintance there is no algorithm that can resolve such type of associations. Deriving such an algorithm is not a simple difficulty in view of the fact that it is familiar that the itemset creation in the association rule mining development is an exclusive one. It would be essential not only to think all items in an operation, but also all probable items deficient from the operation. There could be a significant exponential expansion in the candidate generation stage. This is particularly true in datasets with extremely associated attributes. That is why it is not viable to expand the attribute space by accumulation the negated attributes and use the subsisting association rule algorithms.

A strong positive association is referred to as a positive relation between two data sets. Negative relation implies a negative rule between the two pattern sets. However, strong negative associations disclose only the existence of negative rules in an unseen representation and do not give the real negative rules. Unlike existing mining techniques, this paper extends conventional associations to include association rules of forms $A \rightarrow B$, $\neg A \rightarrow B$, and $\neg A \rightarrow \neg B$, which indicate negative associations between pattern sets. Lets illustrate rules of the form $A \rightarrow B$ positive rules, and rules of the another forms negative rules [1]. While positive association rules are useful in marketing analysis, negative association rules also play key roles in marketing analysis.

For example: there are four types of products each product has profit gain $P=20\%$, $Q=30\%$, $R=60\%$, $S=80\%$. The goal of the market analysis team is to consider only those product which has high profit gain, positive rules help in determine which product can be adopt, negative association rules help in determine which product can be ignored, The goal of market analysis team is to ensure a fair and efficient trading for all suppliers through a signal system assume that each piece of evidence P , Q , R and S , can cause a signal of profit and loss, if having rules in the form $P \rightarrow \neg X$, $Q \rightarrow \neg X$, $R \rightarrow X$ and $S \rightarrow X$, the team can make the decision of profit gain when P, Q occurs, in other words, signal caused by P, Q can be ignored.

2. LITERATURE SURVEY

G. Sathyadevi [11] has proposed a new application of using cart algorithm in Hepatitis disease diagnosis. This effort was inspired by area-ambitious data mining. This effort challenges to take full advantage of the effectiveness of domain proficient (oracles) in energetic learning process. CART Algorithm was used to observe hepatitis disease diagnosis. From the given training datasets, only appropriate attributes are decide on using decision tree algorithm CART. Recognition and selection of appropriate attributes that make a payment to hepatitis disease is a difficult job. Thus experiential studies with UCI hepatitis patient datasets give you an idea about that the suggested active learning algorithm is more efficient than the other modern algorithms for energetic learning. She [8] proposes a CART decision tree Algorithm adjacent to the biomedical hepatitis patient datasets and evaluates the results with other data mining methods. The decision tree which is generated by other two algorithms may have two or more child. In addition, in value of accuracy and time complication CART algorithm carry outs enhanced than the other two algorithms.

S.Vijayalakshmi et al [9] studied an extensive choice of techniques of Extracting Sequential Access Pattern from Pre-processed Web Logs. Sequential pattern mining is a significant data mining difficulty with extensive purposes, as well as the examination of client acquire performance, patterns of web right of entry, technical experiments, and behavior of disease, ordinary disaster, and protein arrangements. They [9] increase the possibility of frequent sequence pattern mining by investigating a well-organized algorithm for mining frequent and scientifically investigate a pattern-growth move toward for well-organized mining of sequential patterns in huge sequence record. Initially scrutinize an uncomplicated pattern growth technique, free Span for Frequent pattern-projected Sequential pattern mining, which help out to reduce the hard work of candidate subsequence construction. Prefix-projected Sequential pattern mining, which offers ordered growth and decrease projected records. The Prefix Span put away a much less important memory space in evaluation with GSP.

Alka Gangrade et al [10] paying attention on the evaluation of the methods for privacy preserving categorization of data mining under multi-party situation. The presentation examination of classification replica and make safe multi-party computation algorithms has been determined in connection with the classification. Classification Rule Mining (CRM) algorithms are found on fundamental data model that is all data is accumulated into a single site.

Hidenao Abe [11] illustrated a classification rule mining structure by integration the two models i.e., sequential pattern extraction and rule mining. This framework has been

extended for mining if-then rules enclosing of sequential patterns in left hand side of the regulations. The right hand side helps us to calculate both of significant occurrences and temporal patterns of significant index. Sequential data mining requires temporal data as input. Output rules are if-then rules, which have temporal patterns or/and ordinal sections, characterized in $A = x$, $A _ y$, and $A > z$ arrangements of removed patterns and/or ordinal clauses can be acquired as if-then rules by a rule induction algorithm. To put into operation the situation they evaluated temporal data mining structures.

Then they identified procedures for pattern extraction as data pre-processing, rule induction as mining, and evaluation of rules with visualized rule as post-processing of mined result. The system provides these procedures as commands for users. At the same time, they designed a graphical interface, which include data processing, validation for patterns on elemental sequences, and rule visualization as charts [11]. This integrated time-series data mining environment combines the following major functional components: time-series data pre-processing, mining, post-processing for mined results, and other database operators to validate data and results of every phase.

Farah Hanna AL-Zawaidah et al [12] suggested An Improved Algorithm for Mining Association Rules in Large records. They show aggression the association rule mining by an apriori based move toward distinctively proposed for the optimization in exceptionally huge transactional records. The extended mining comes up to call Feature Based Association Rule Mining Algorithm (FARMA). The buildup come within reach of assumes the thinking of Apriori approach with some alterations with the intention of condense the time execution of the algorithm. First, the suggestion of producing the feature of items is used and; second, the power for each candidate itemset is determined to be used during dealing out. The feature array data arrangement is built by accumulating the decimal correspondent of the position of the item in the operation. In other words alter the operation record into the feature matrix. Altering here indicates reordering and changing a large database into convenient arrangement to accomplish two purposes: (a) reducing the number of I/O right of entries in data mining, and (b) speeding up the mining procedure. There is one compulsory conditions for the altering technique, that the process confirmation should be read only once inside the complete life cycle of data mining. By storing the showing feature of each concentration item as a constricted vector independently, the size of the database to be right of entry can be decreased very much [12].

In year 2013, Ruchi Bhargava and Shrikant Lade presented Effective Positive Negative Association Rule Mining Using Improved Frequent Pattern. They presented review for association rules generation and positive negative rules finding from large data. A new algorithm is presented to generate all positive and negative class association rules using improved FP-tree and to build an accurate classification. The scheme has several features like (a) Its classification is performed based on positive and negative class association rules, which leads to better overall classification accuracy; (b) It prunes contradictory positive and negative class association rules effectively based on correlation between itemsets. (c) An improved FP tree applies for mining association rules. This algorithm extracts all promising frequent item set with no producing the provisional FP tree. It also makes available the occurrence of

frequent items, which is utilized to approximation the aspiration association rules. This method expected that the algorithm would be highly effective at classification and has better average classification accuracy and efficiency [13].

In recent times 2013; Luca Cagliero and Paolo Garza suggested a new method based on Infrequent Weighted Itemset Mining using Frequent Pattern Growth. Here they concentrate on the detection of infrequent and weighted itemsets, i.e., the Infrequent Weighted Itemsets (IWI), from operational weighted datasets. To concentrate on this problem the IWI-support determines is described as a weighted frequency of incidence of an itemset in the investigated data. Happening weights are originated from the weights associated with items in each operation by affecting a given expenditure function. They generally centers on following: (i) The IWI-support-min calculate, which relies on a lowest cost function, i.e., the amount of an itemset in a given operation is weighted by the weight of its smallest amount importance item, (ii) The IWI-support-max measure, which relies on a maximum cost function, i.e., the occurrence of an itemset in a given transaction is weighted by the weight of the most interesting item [14].

R. Uday Kiran and P. Krishna Reddy suggested a new method to decrease exploration space in Multiple Minimum Supports-Based Frequent Pattern Mining Algorithms [15]. They explored come within reach of to decrease the investigate space while removing frequent patterns and suggested two extra pruning methods which extensively diminish the search space by keep away from comprehensive search at the same time as removing frequent patterns from a tree structure. In general, they have recommended a wide-ranging algorithm by utilizing four pruning methods to proficiently mine frequent patterns. To mine frequent patterns containing both frequent and rare items, “multiple minsups framework” was proposed in the literature. By considering “multiple minsups framework,” CFP-growth algorithm has been proposed to extract frequent patterns. They proposed an improved CFP-growth algorithm, called CFP-growth++, by commencing the subsequent pruning methods: smallest amount least support, unconfirmed minsup, restrictive closure property and uncommon leaf node pruning. By performing researches on both artificial and real-world datasets, they have exposed that the recommended algorithm get betters the presentation considerably over the stimulating move towards. At high minsup value, rare item sets are missed, and at low minsup value, the number of frequent item sets explodes. The technique in which minsup of each item is fixed or based on support difference reduces both “missing rule and “explosion rule”. To improve the performance of extracting itemsets involving rare items, an approach known as Multiple Minimum Support Apriori (MSApriori).It was observed MSApriori still suffer from the “rare item problem” if item support vary widely [15].

3. PROPOSED METHODOLOGY

Here the proposed methodology is based on the combinatorial method of rules generation and classification. The proposed methodology works in the following phases:



Figure 2.1. Architecture of the proposed methodology

Here two types of dataset are used real life dataset and synthetic dataset. Here we use ARFF version of the dataset means attribute relation file format. Since the base work is for infrequent item sets, here we take both frequent and infrequent and weighted and un-weighted datasets.

Here for the generation of frequent sets and association rules for a given dataset and hence a minimum threshold is decided with support and confidence.

Step 1: A Data structure is maintained based on the given dataset and generate a FP- tree using the given dataset.

Step 2: The Generated Tree is then used for the computation of frequent sets and hence on the basis of which rules are generated. The generation of In-frequent item sets is given below:

- a) On the basis of Generated FP-Tree traverse the tree and for each of the items in the Tree compute Support at each level of the tree.
- b) Maintain a separate data structure for the storage of frequent and In-frequent item sets.
- c) Traverse each of the two data structures which contains Frequent and In-Frequent item sets and sort each of the data structure in increasing or decreasing order on the basis of support.
- d) Each of the node in the data structure contains an item sets and their counter value.
- e) On the basis of these nodes and their respective counters 1 transaction is read every time and mapping is done.
- f) As soon as the transaction which is traversed right now is used to check the prefix of the item set and hence the counter is incremented.
- g) Each of the nodes containing items sets is pointed containing same item sets and the remaining items are stored in the memory.
- h) Finally Frequent and In-frequent item sets are extracted from the generated Tree.

3.1 Naïve Bayes Algorithm

Abstractly, the probability model for a classifier is a conditional model.

$$p(C / F_2, \dots, F_n)$$

over a dependent class variable C with a small number of outcomes or *classes*, conditional on several feature variables F_1 through F_n . The problem is that if the number of features n is large or when a feature can take on a large number of values, then basing such a model on probability tables is infeasible. We therefore reformulate the model to make it more tractable.

Using Bayes' theorem, this can be written

$$p(C / F_2, \dots, F_n) = \frac{p(c)p(F_1, \dots, F_n / C)}{p(F_1, \dots, F_n)}$$

In plain English the above equation can be written as

$$\text{Posterior} = \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}}$$

In practice, there is interest only in the numerator of that fraction, because the denominator does not depend on C and the values of the features F_i are given, so that the denominator is effectively constant. The numerator is equivalent to the joint probability model.

$$p(C, F_1, \dots, F_n)$$

which can be rewritten as follows, using the chain rule for repeated applications of the definition of conditional probability:

Now the "naive" conditional independence assumptions come into play: assume that each feature F_i is conditionally independent of every other feature F_j for $j \neq i$ given the category C . This means that

$$p(F_i / C, F_j) = (p(F_i / C))$$

This means that under the above independence assumptions, the conditional distribution over the class variable C is:

$$p(C / F_1, \dots, F_n) = \frac{1}{Z} p(C) \prod_{i=1}^n p(F_i / C)$$

where Z (the evidence) is a scaling factor dependent only on F_1, \dots, F_n , that is, a constant if the values of the feature variables are known.

4. RESULT ANALYSIS

The proposed methodology implemented here is tested on various popular datasets including:

1. **Compendium Dataset:** The dataset is specially designed to investigate the various health services and research on medical education.

2. **StemCells Dataset:** The dataset is based on Pluripotent cells which generates cell in the human body such as HESC.

3. **Yeast Dataset:** The dataset contains the various predicted attributes of the protein localization. It contains 1484 number of instance values with 8 attributes.

The figure shown below is the comparison of total frequent item sets generates on Step Cells Dataset. The Result is compared on Stem Cells Dataset for various support and confidence values.

The existing and proposed work is compared on various values of support and confidence. The rules generated for the existing and proposed work for support and confidence.

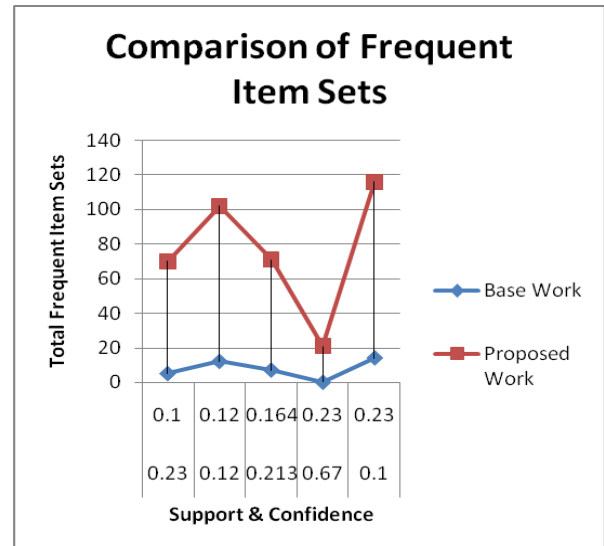


Figure 4.1 Comparison of Frequent Item Sets

The figure shown below is the comparison of total In-frequent item sets generates on Step Cells Dataset. The Result is compared on Stem Cells Dataset for various support and confidence values.

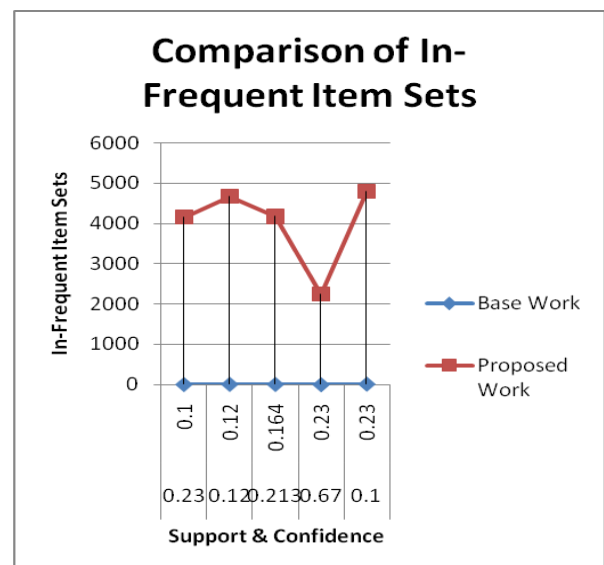


Figure 4.2 Comparison of In-Frequent Item Sets

The figure shown below is the comparison of total Positive rules generates on Step Cells Dataset. The Result is

compared on Stem Cells Dataset for various support and confidence values.

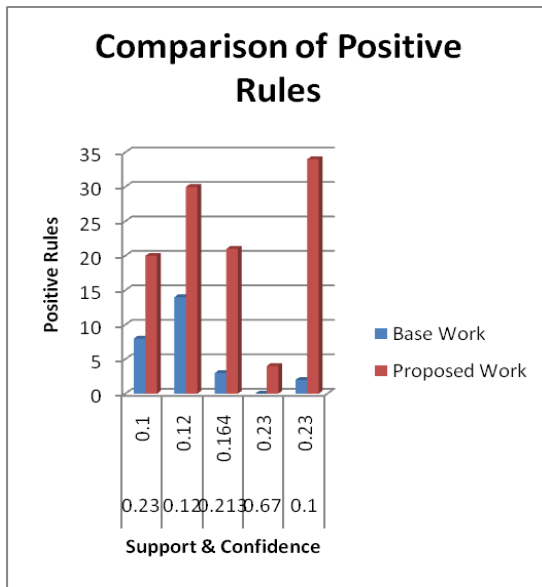


Figure 4.3 Comparison of Positive Rules

The figure shown below is the comparison of total Negative rules generates on Step Cells Dataset. The Result is compared on Stem Cells Dataset for various support and confidence values.

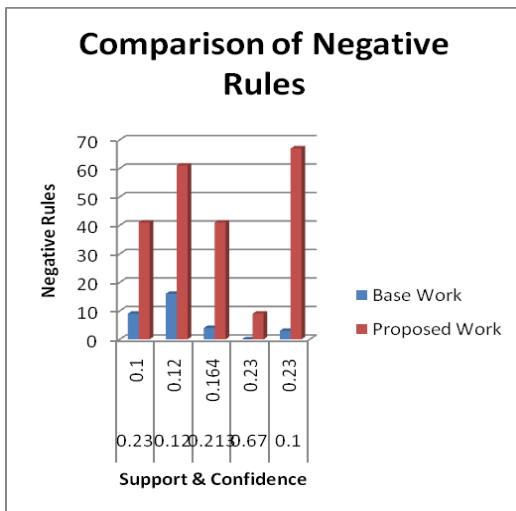


Figure 4.4 Comparison of Negative Rules.

5. CONCLUSION

Here proposed a new technique for optimization of concentration on positive and negative association rule using mixture combinatorial technique of Fuzzy FP-Growth and Naïve bayes Classifier is utilized. The examination is that when transform the examine procedure of operation, making of rule is quick. With more imperatives promising it involve there should be a method for organization their huge numbers. The big produced rule is optimized with Fuzzy Fp-Growth. We tentatively conformational a relation between in the neighborhood huge and worldwide large patterns that is used for pruning at each level to diminish the exploration candidates. We developed a locally huge threshold using a worldwide set minimum remember threshold. Pruning accomplishes a decrease in the number of investigated

candidates and this decline has a comparative force on the decreasing of huge amount of negative rules.

The proposed methodology implemented here for the analysis of various datasets using Fuzzy FP-Growth and Naïve Bayes classifier is an efficient technique which generated positive and negative rules in a more classified manner, so that the analysis can be done easily and quickly. The experimental results are performed on various datasets and results are compared with the existing Rule generation techniques. The proposed methodology implemented has better performance as compared to the existing technique.

6. REFERENCES

- [1] Idheba Mohamad Ali O. Swesi, Azuraliza Abu Bakar, Anis Suhailis Abdul Kadir "Mining Positive and Negative Association Rules from Interesting Frequent and Infrequent Itemsets", *9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012)*, pp. 650 – 655, 2012.
- [2] A. K. Santra, and S. Jayasudha "Classification of Web Log Data to Identify Interested Users Using Naïve Bayesian Classification", *International Journal of Computer Science Issues (IJCSI)*, Vol. 9, Issue 1, No 2, January 2012.
- [3] Pengfei Guo, Xuezhi Wang and Yingshi Han "The Enhanced Genetic Algorithms for the Optimization Design", *3rd International Conference on Biomedical Engineering and Informatics (BMEI)*, Vol. 7, pp. 2990 – 2994, 2010.
- [4] Xin Li, Xuefeng Zheng, Jingchun Li, and Shaojie Wang "Frequent Itemsets Mining in Network Traffic Data", *Fifth International Conference on Intelligent Computation Technology and Automation*, pp. 394-397, 2012.
- [5] Soumadip Ghosh, Sushanta Biswas, Debasree Sarkar, Partha Pratim Sarkar "Mining Frequent Itemsets Using Genetic Algorithm", *International Journal of Artificial Intelligence & Applications (IJAA)*, Vol.1, No.4, pp. 133 – 143, October 2010.
- [6] Agrawal R., Imielinski T. and Swami A. "Mining Association rules between sets of items in large databases, *In the Proc. of the ACM SIGMOD Int'l Conf. on Management of Data (ACM SIGMOD '93)*, pp. 207-216, Washington, USA, 1993.
- [7] G. Vijay Bhasker, K. Chandra Shekar, V. Lakshmi Chaitanya "Mining Frequent Itemsets for Non Binary Data Set Using Genetic Algorithm", *International Journal Of Advanced Engineering Sciences And Technologies (IAEST)*, ISSN: 2230-7818, Vol. 11, Issue No. 1, pp. 143 – 152, 2011.
- [8] G. Sathyadevi, "Application of cart algorithm in Hepatitis disease diagnosis", *IEEE International Conference on Recent Trends in Information Technology (ICRTIT-2011)*, pp. 1283 - 1287, June 2011.
- [9] S. Vijayalakshmi, Dr. V.Mohan, M. S. Sassirekha, O.R. Deepika, "Extracting Sequential Access Pattern from Pre-processed Web Logs", *Proceeding in IEEE International Conference on Process Automation, Control and Computing (PACC-2011)*, pp. 1- 6, 2011.

- [10] Alka Gangrade, Durgesh Kumar Mishra and Ravindra Patel ;“Classification Rule Mining through SMC for Preserving Privacy Data Mining: A Review”, *International Conference on Machine Learning and Computing IPCSIT*, vol. 3, pp. 431- 434, IACSIT Press Singapore, 2011.
- [11] Hidenao Abe “Development of a Classification Rule Mining Framework by Using Temporal Pattern Extraction”, *New fundamental technologies in data mining*, pp. 493-504, January 2011.
- [12] Farah Hanna AL-Zawaidah, Yosef Hasan Jbara and Marwan AL-Abed Abu-Zanona “An Improved Algorithm for Mining Association Rules in Large Databases”, *World of Computer Science and Information Technology Journal (WCSIT)*, ISSN: 2221-0741, Vol. 1, No. 7, pp. 311-316, 2011.
- [13] Ruchi Bhargava and Shrikant Lade “Effective Positive Negative Association Rule Mining Using Improved Frequent Pattern”, *International Journal of Modern Engineering Research (IJMER)*, ISSN: 2249-6645, Vol.3, Issue.2, pp-1256-1262, March-April. 2013.
- [14] Luca Cagliero and Paolo Garza “Infrequent Weighted Itemset Mining using Frequent Pattern Growth”, *IEEE Transactions on Knowledge and Data Engineering*, pp. 1- 14, 2013.
- [15] R. Uday Kiran and P. Krishna Reddy “Novel Techniques to Reduce Search Space in Multiple Minimum Supports-Based Frequent Pattern Mining Algorithms”, *Proceedings of the 14th International Conference on Extending Database Technology*, pp. 11 – 20, 2011.