ISSN 2250-0987
Kavitha et al, UNIASCIT, Vol 3 (2), 2013, 387-393

AN ADEQUATE APPROACH FOR ACTIONABLE PATTERN USING COMBINED AND COMPOSITE ASSOCIATION RULE MINING

K. Kavitha[#1], E. Ramaraj[*2]

[#]*Assistant Professor, Department of Computer Science, Dr. Umayal Ramanathan College for Women, Karaikudi.*

[*]*Professor, Department of Computer Science and Engineering, Alagappa University, Karaikudi.*

***Abstract*— Data mining promises to discover valid and potentially useful patterns in data. Association rule mining often produces large numbers of association rules and sometimes it is difficult for users to understand such rules, therefore it cannot satisfy the needs of real world completely. Though numerous of techniques have been proposed for data mining with the aspiration to discover knowledge. Among those techniques, an association rule learning algorithm named (IP-FPM) Intensified Priority based Frequent Pattern Mining is most famous. It generates the frequent item set among the specified items that are given as input, not for all the data items presented in the data set. However, it is more effective in mining it suffers from number of generating rules and numbers of frequent items used. In this paper, we proposed an adequate approach for combined mining and composite mining approach for generating actionable association rules. Moreover, combined association rules cannot find any rules. Consequently, we generate meta-learning to deliberate the framework for classification tasks and use the combined association rule mining to extract additional actionable patterns and superlative combined association rule mining with composite items. Experimental results reveal the efficacy of the adduced methodology as compared to the related works.**

***Keywords*— Actionable knowledge, association rules, combined mining, composite mining, data mining, pattern mining.**

I. INTRODUCTION

Data mining is a major step in the (KDD) Knowledge Discovery in Databases process consisting of applying computational techniques that under acceptable computational efficiency limitations produce a particular enumeration of patterns over the data. Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. These tools can include statistical models, mathematical algorithms, and machine learning methods. The progress in data mining research has made it possible to implement several data mining operations efficiently on large databases. The mined information is typically represented as a model of the semantic structure of the dataset, where the model maybe used on new data for prediction or classification. Data mining techniques have been successfully applied in many different fields including marketing, manufacturing, process control, fraud detection, and network management. Moreover, actionable Patterns can not only afford important grounds to business decision-makers for performing appropriate actions, but also deliver, expected outcomes to business. In data mining area, in order to discover the knowledge, the general framework suggested is called as (KDD) knowledge discovery in databases. In our framework KDD extract rules in the forms of association rules, which is used frequently.

Association rule mining is a main method to produce patterns. However, as large numbers of association rules are often produced by association mining algorithm, sometimes it can be very difficult for decision makers to not only recognize such rules, however also find them a useful source of knowledge to apply to the business processes. The association rules can only present limited knowledge for possible actions. Consequently, there is a strong and challenging need to mine for more informative and comprehensive knowledge for decision-making in the real world. A complete and general approach for discovering informative knowledge in complex data is suggested. To present associations in an interesting and effective way, and in order to find actionable knowledge from resultant association rules, the idea of combined patterns is used. Combined patterns comprise combined association rules, combined rule pairs and combined rule clusters. The resultant combined patterns provide more interesting knowledge and more actionable results than traditional association rules. The issue of mining association rules with composite items was proposed several years ago. Moreover, algorithms designed to mine association rules with composite items are also described. The algorithms allow large itemsets to contain composite items. This enables them to discover certain useful rules which cannot be found out by other existing algorithms without composite items. Thus, general profit can be employed as the measurement criteria in real life applications. To this end, in order to generate actionable patterns, interesting measures including objective measure such as support and confidence, and subjective measure such as expert knowledge and domain knowledge should be introduced in actionable combined association rule mining with composite items. In particular, in the mining process expert knowledge and domain knowledge should be involved at first. Then, the patterns generated should be evaluated.

We present different approach, the first step is used to employ the specialist knowledge and domain knowledge to ascertain the composite items accomplish from itemset. Then the association rule algorithm to find the rules with composite items. Moreover, we present combined association rule mining integrating the rules generated.

The remainder of this paper is structured as follows, Section 2 confers about the related works, Section 3 summarizes our proposed Actionable pattern for combined and composite mining, Section 4 discussed the experiments and results achieved. Finally, we present the conclusion and future enhancements of the adduced work.

II. Related Work

Some of the notable research efforts have been made in previous works. A main contribution of this work is the introduction of actions as a way to address this issue. They presented a new concept of actionability and the algorithms for its discovery. In a real world application domain, our approach demonstrates the effectiveness of finding useful patterns and applying pruning strategies [15]. Zhao et al. [17] Proposed combined association rule mining, which create through further extraction of the learned rules. Moreover, present the associations in an effectual way and categorize to discover actionable knowledge from resultant association rules.

The paper [4] shows the flexibility and instantiation capability of combined mining in discovering more informative and actionable patterns in complex data. Furthermore, it presents combined patterns in dynamic charts, a novel pattern presentation method reflecting the evolution and impact change of a cluster of combined patterns and supporting business to take action on the deliverables for intervention. In [11] proposed an Actionable Knowledge Discovery Applications they are Customer Relationship Management, Supplier Selection, Crime Identification and Business Intelligence. (CRM) Customer Relationship Management consists of four dimensions: Customer identification, customer development, customer retention and customer attraction. Then, customer satisfaction is the central concern for customer retention. The supplier selection is one of the major parts in supply chain management. Strategic partnership with best performing suppliers should be integrated into the manufacturing to improve the performance in various aspects including reducing costs by eliminating waste, reducing lead time at distinct stages of the manufacturing and continuously improving quality to accomplish zero defects. Crime detection for credit applications is so popular in the banking industry.

An actionable knowledge discovery (AKD) from the system and decision-making perspectives is presented in [1]. AKD is a closed optimization problem solving process from problem definition, a framework or model design for actionable pattern discovery and is designed to deliver operable business rules that can be seamlessly integrated or associated with business processes and systems. Moreover, actionable knowledge discovery is critical in attribute and releasing the productivity of data mining and knowledge discovery for smart business operations and decision making. Both SIGKDD and ICDM panelists pointed it out as one of the greatest challenges in developing the next generation KDD methodologies and systems [14]. Lovici and Braha [6] develop a decision-theoretic framework for evaluating data mining systems that employ a classification approach in term of their efficacy in decision making. The decision-theoretic model affords an economic aspects of the value of extracting knowledge its payoff to the organization and suggests a wide range of decision problems that arise from this point of view. Then the relation between the attribute of a data mining system and the amount of investment that the decision maker is concerned to make is formal. A high level picture of combined mining and many novel aspects of pattern relation analysis and combined patterns are given in [2]. Pattern combination dimensions, pattern relation, pattern combination criteria, pattern paradigms and pattern structures, which are important for constructing combined patterns and for discovering actionable knowledge in complex data. Pattern ontology and the pattern dynamic chart also acquaint to present combined patterns. The authors [5] discuss about the new idea about pruning the association rules before making the rule clusters or rule pairs. The domain driven concept gives the idea of selection of data features before generating the frequent patterns. Through domain driven user can select the customer characteristics from static customer data. The rule pairs can also be generated from pruned frequent patterns, maximum frequent patterns or closed frequent pattern. Consequently, the number of rules generated get condensed and also readable and understandable to the user. This new method giving the cluster of rules for similar type of customers with their change of class as the transaction characteristic changes.

In [9] described the combined Mining approach using Irule parameter for generating actionable association rules. The approach is tested on a survey data set that consists of multiple related data items collected independently. A new techniques were proposed in [12] that were used to utilize the decision making to drive the mining process. They introduce extrinsic measure which evaluates patterns with respect to the decisions made. Moreover, to include decision making as a part of Knowledge Discovery in Databases (KDD) process and come up with a general architecture to measure the pattern's usefulness with respect to the decision made. The authors [13,8] discuss about the performance of Apriori and FP-growth algorithms among the various algorithms in Association Rule Mining. The major difference between the two approaches is that the Apriori-like techniques are based on bottom-up generation of frequent itemset combinations and the FP-growth based ones are divide-and-conquer and partition-based, methods. Furthermore, the Apriori algorithms generate the candidate itemsets but FPgrowth does not generate the candidate itemset.

A theoretical connection between the process of frequent itemset discovery and learning of generative models was given in [10]. Moreover, proposed a class of generative models called Itemset Generating Models (IGMs) and associated each itemset with an individual IGM from this class. A connection between data likelihood under the IGM and the frequency of the associated itemset and showed that frequency ordering among itemsets is preserved as likelihood ordering among the associated IGMs. The authors [3] discuss about a combined mining as a universal approach to mining for informative patterns combining components from either multiple data sets or multiple features or by multiple methods on demand. They summarize general frameworks processes for multimethod combined mining, multifeature combined mining, and multi feature combined mining. The authors [16] discuss about a Novel Frequent-Pattern tree for finding large composite items first. Furthermore, then how to measure the reliability of these discovered rules with composite items in order to find out the most reliable association rules is discussed.

III. Proposed Work

In our proposal, we present for mining actionable patterns at the mining and decision stages of the Knowledge Discovery in Database process. The data mining stage tries to optimize several intrinsic measures such as accurately that pattern evaluation evaluates the pattern for its actionabilty based on the number of decision theoretic framework. Therefore, model selection stage uses to utilize the pattern in decision making, since we use ideas from the field of Meta-learning. We look at the utility in decision making as the performance measure that is optimized using the meta-learning.

We present different approach, the first step is used to employ the specialist knowledge and domain knowledge to ascertain the composite items accomplish from itemset. Then the association rule algorithm to find the rules with composite items. Moreover, we present combined association rule mining integrating the rules generated.

A. Association Rules

Association rule uses the if-then rules to produce extracted information into the form transaction statements. Such rules have been created from the dataset and it is obtained with the help of support and confidence of rule that illustrate the rate of occurrence of a given rule. Support,confidence and lift are three main purpose indices, which have constructive applications to evaluate the association rules. As the combined association rule mining more extracts from the simple learned rules. Support, confidence and lift are major metrics of combined and composite association rule mining.

1. Interestingness measurement

Interestingness measurements and constraint based mining are employed in data mining to filter out such redundancy and uninteresting patterns. The most effective way of reducing the volume of discovering pattern is so called interestingness measures. There are two types of such measures namely, objective and subjective measures. Objective measures are those that depend only on the structure of a pattern such as the confidence, lift and support the measure and usually quantified by using statistical methods. Subjective measures are further divided into actionable, unexpected and the novel. A pattern is actionable which means that decision makers can get benefit from discovered patterns. Therefore, in order to obtain the actionable knowledge for the decision making effectively, the balance between objective and subjective measures should be both involved in the mining process. For mining combined association rules with composite items, the key point is how to build the objective and subjective measures between combined association rule mining and mining association rules among composite items.

As the database is enormous and others be concerned about only those frequently bought items, typically thresholds of support and confidence are predefined by users to drop those rules that are not so useful. The three thresholds are called minimal support and minimal confidence and lift respectively, furthermore constraints of interesting rules also can be recognized by the users.

Input : $A_I \rightarrow$ Frequent Actionable Pattern, $S_C \rightarrow$ Support Counts

Output : The complete set of Frequent Patterns

begin

  for each A_I // $A_I \rightarrow$ Frequent actionable patterns itemsets

    identify F_I with A_{IT} // $F_I \rightarrow$ Frequent itemsets

      for each A_{IT} // $A_{IT} \rightarrow$ Target class

        record S_{CT} // $S_{CT} \rightarrow$ Support count

        compute conditional CT // $CT \rightarrow$ Support ConSup

    if $(ConSup\,(CT) > MinSup)$ then

      identifying three itemsets

      for each A_{ITF}

        record the support count S_{CTF}

        compute $Support : ConSup\,(CF)$

        compute $Conf, Lift$ and $ConfLift$

    endif

      if $(Conf \geq min_s \,\&\, Lift \geq min_l \,\&\, ConLift \geq min_{sl})$ then

        Adding the mined frequent itemsets to the rule set

      endif

end

3.1.2 Algorithm for Frequent Pattern Generation

1.1 Support

Support of an association rule is defined as the fraction of the data that hold A U B to the total number of records. The count for every item is increased by one all time the item is coming across in dissimilar transaction S in database Q during the scanning process. Support count is calculated through equation (1).

$$Support\ (A\ B) = \frac{Support\ Count\ of\ AB}{Total\ Numbberof\ Transaction\ in\ Q} \quad (1)$$

1.2 Confidence

Confidence of an association rule is defined as the fraction of the number of transactions that cover A U B to the total number of records that have A, where if the percentages go above the threshold of confidence an interesting association rule A→B can be formed. Confidence is computed from the equation (2).

$$Confident(A/_B) = \frac{Support\ (AB)}{Suppport\,(B)} \quad (2)$$

Correspondingly to make certain that the interestingness of rules recognized minimum confidence is also pre-defined by users. The popular association rules Mining is to mine strong association rules that convince the user's individual both minimum support threshold and confidence threshold. That is, minconfidence and min support.

1.3 Lift

Association of rule confidence to the predictable rule confidence is carried out through the lift. The lift can be calculated from the equation (3).

$$Lift\left(\frac{A}{B}\right) = \frac{Sup\ (AB)}{Sup\ (B) \times Sup\ (B)} \quad (3)$$

Mining association rules is particularly useful for discovering relationships among items from large databases. A standard association rule is a rule of the form A→B which says that if A is true of an instance in a database, so is B true of the same instance, with a certain level of significance as measured by two indicators, support and confidence. The goal of standard association rule mining is to output all rules whose support and confidence are respectively above some given support and coverage thresholds. The mining process of association rules can be divided into two steps.

- Frequent Itemset Generation: produce all set of items that have support greater than a certain threshold called min support.
- Association Rule Generation: from the frequent itemsets, create all association rules that have confidence greater than a certain threshold called minconfidence.

2. Algorithm for Frequent Pattern Generation

Algorithm 1, express the steps taken to discover the frequent patterns. The first step is to find single rule composed of frequent itemsets. The second step is to extract interesting combined and composite association rule sets. In order to compute the interestingness measures, the support count of all frequent itemset is recorded in the frequent itemset generation step.

B. Composite Items Mining

In our proposal, we present the concept of composite items. Let us consider $C = \{x_1, \ldots\ldots x_m\}$ be the set of accurate atomic items. A composite item is composed by combining numerous atomic items. The common form of a composite item is determined as $x_1 \vee \ldots \vee x_n$, where $x_i \in C$ for $1 \le i \le n$ and $x_j \neq x_i$ for $j \neq i$. A database consists of transactions and every transaction has a set of atomic items. Then, the transaction incorporates a composite item if the transaction contains at least one of the atomic items that form the composite item. Moreover, atomic items and composite items will be associate to use items in common. Subsequently, the item is large if the number of transactions includes the item minimum support.

An itemset is a set of items such that none of the items in the set have general items. For example, (W, X$\cup Y$) is a valid itemset. Though, (W, W$\cup X$, Y$\cup Z$) is not an accurate itemset as X and X$\cup$Y has ordinary item X. Then, the association rule is an assumption of the form A$\rightarrow B$ where A, B and A$\cup$B are itemsets, $A \neq \emptyset, B \neq \emptyset$ and $A \cap B = \emptyset$ in [16].

C. Combined Mining

Support, confidence and lift are major purpose indices that have effective applications to estimate the association rules. The combined association rule mining further extracts from the simple learned rules. Support, confidence and lift are major metrics of combined association rules. Moreover, subjective metrics such as expert knowledge and domain knowledge are not involved in combined association rules.

Combined association rule mining is to determine the association rule mining among the attribute value. For the convenience of description, we call an attribute value pair an item. Suppose item set $B \subseteq I_B, I_B$ is the item set of any items with attribute, itemsets $A \subseteq I_A, I_A$ Its a term of attributes and S is 1-itemsets and combined association rule set is represented as

$$\begin{cases} D + A_1 \rightarrow E_{k1} \\ \quad\quad . \\ \quad\quad . \\ \quad\quad . \\ \quad\quad . \\ D + A_i \rightarrow E_{ki} \end{cases}$$

Usually, there are a lot of tasks of data mining, such as classification, clustering and rule mining. Among of these tasks, the mining of patterns for discerning relationships between data items in large databases is a well studied technique. In order to introduce the key research, more details of association rule mining, combined association rule mining and association rule mining with composite items will be presented in this section.

ISSN 2250-0987
Kavitha et al, UNIASCIT, Vol 3 (2), 2013, 387-393

D. *Reliability with Combined and Composite Items*

An essential characteristic of data mining is the discovered knowledge should be appealing. In [16] proposed how to measure the interestingness of association rules to find out best and optimal rules in the database. Moreover, there are two kinds of measure namely, subjective measures and objective measures.

Subjective measures use to take into account the users domain knowledge and goals. Objective measures the support, confidence, entropy and correlation, it asses the interestingness of rule in terms of rule generation and rule structure. There are numerous key properties that can be used to preferable the right objective measures for given purpose. Therefore, the results can be applied to the case that association rule with composite items while treating every composite item as an independent item like the atomic items.

IV. EXPERIMENTAL RESULTS

In order to evaluate the effectiveness of our adduced work, we organized several executions on the synthetic sequence of tasks. The efficiency of the proposed methodology is manipulated for two different datasets as shown in Table 1. Both the zoo and mushroom databases retrieve from the UCI repository of machine learning database.

Table 1 : Details of Dataset

Data Set Name	Number of Transactions	Number of Items
Zoo	101	15
Mushroom	8124	119

Figure 1 depicts the time taken by both the proposed and IP-FPM algorithm to generate the actionable pattern for zoo dataset by varying the minimum weight value. Both IP-FPM and CCARM data set having the threshold value for minimum weight.

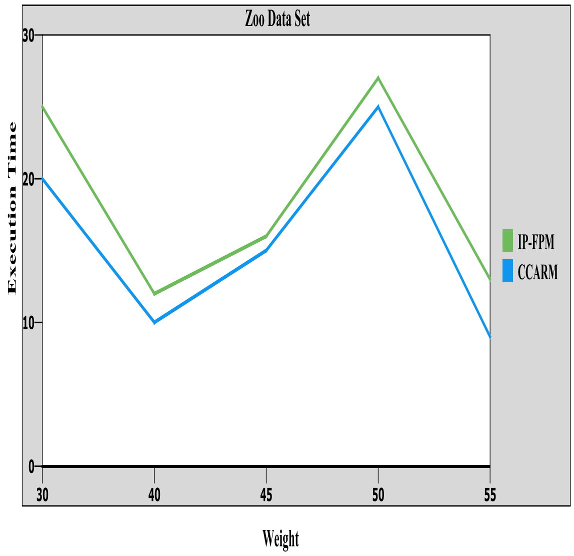
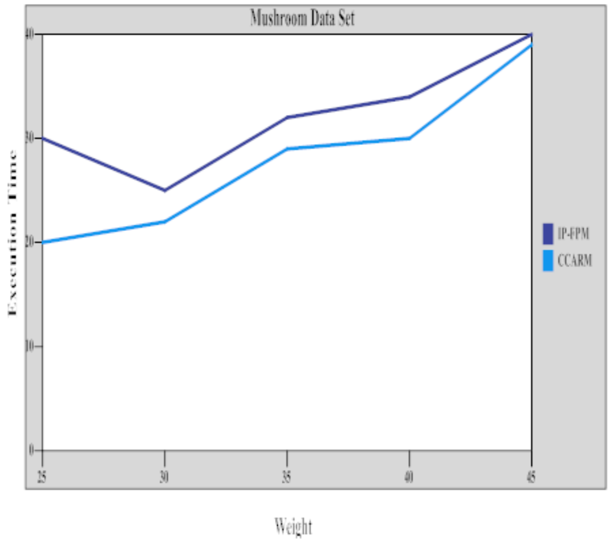


Figure 1 : IP-FPM vs CCARM Algorithms Execution Time for Zoo Dataset

Figure 2 illustrates the time taken by proposed and existing algorithm to create the actionable pattern mining for mushroom dataset by varying the minimum weight value. It contains a massive number of transactions and items than the zoo dataset.



IP-FPM vs CCARM Algorithms Execution Time for Mushroom Dataset

From the figure 1 and 2 it is obvious that the execution time for the proposed algorithm is lesser than the existing algorithm. This is because of the limited number of transactions prior to the generation of actionable pattern that made the proposed algorithm to discover the actionable pattern faster than the existing technique.

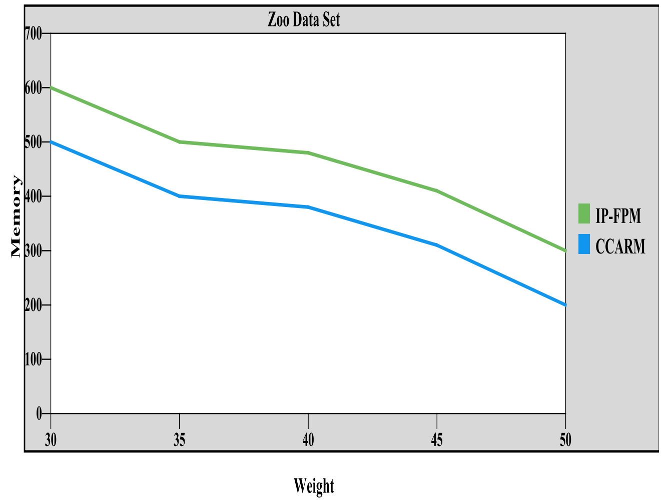


Figure 3 : IP-FPM vs CCARM Algorithms Memory Requirement for Zoo Dataset

Memory usage determines the required memory to execute the techniques. Figure 3 expresses the utilization of main memory during execution of both proposed and existing techniques. It generates the actionable pattern for the zoo dataset by varying the minimum weight value.

Figure 4 depicts the required memory for proposed and existing algorithm to produce the actionable pattern for mushroom dataset by varying the minimum weight value. It contains an enormous number of transactions and items than the zoo dataset. Consequently, it requires much memory than zoo data set.

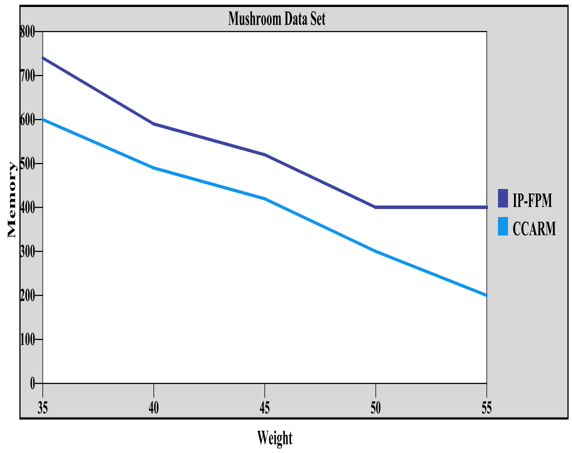


Figure 4 : IP-FPM vs CCARM Algorithms Memory Requirement for Mushroom Dataset

Figure 3 and 4 expresses the memory consumption of the proposed method for the zoo dataset and for the mushroom dataset are lesser than the existing algorithm.

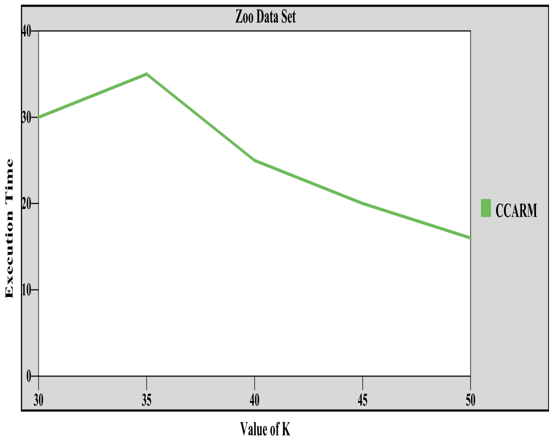


Figure 5 : Execution time for Zoo dataset for various K Values

Execution time required by the proposed algorithm for each K process is analyzed for two different datasets by fixing the threshold value of weight as 50%. Figure 5 depicts the execution time for intermediate values of K, which required for the zoo dataset. Moreover, figure 6 illustrates the mushroom data set value is analyzed their execution time for different K values.

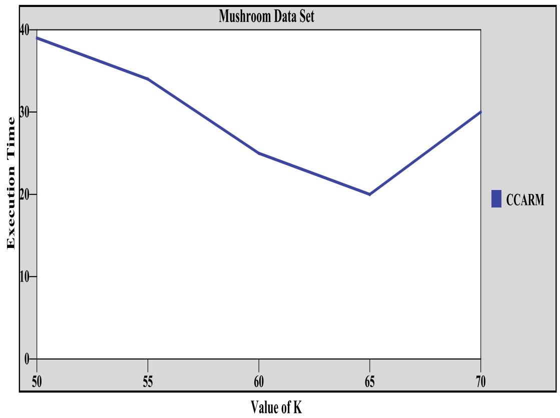


Figure 6 : Execution time for Mushroom dataset for various K Values

Figure 5 and 6 shows that the execution time for the intermediate K values is lesser in both the dataset. Therefore, selection of K values should have more carefulness otherwise the execution time for the proposed will be increased.

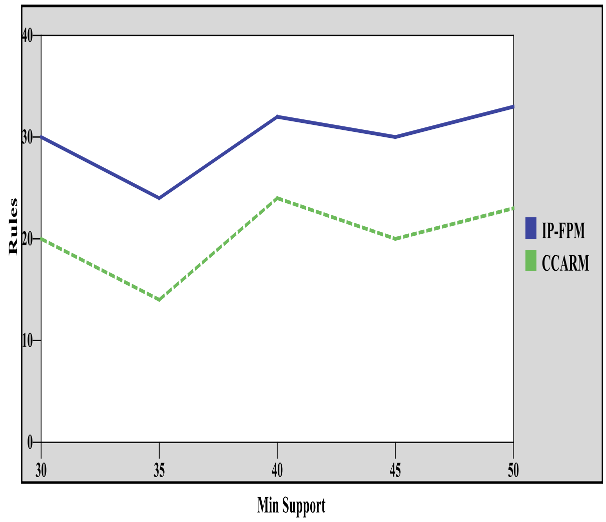


Figure 7 : Min Support for Existing and Proposed

Figure 7 depicts the minimum support for existing and proposed method. We compare the Min support of CCARM with IP-FPM algorithms. According to the results of our proposed work, we find out that CCARM uses less time to complete all rules because CCARM method assigns mine strong association rules that convince the user's individual for minimum support threshold.

This clearly denotes that proposed method is faster than the existing method. Therefore, the result expresses that our proposed method retrieves the actionable pattern efficiently and also effectively than the existing method.

V. CONCLUSION

In this paper, we derived an adept methodology for an approach for combined mining and composite mining approach for generating actionable association rules. We proposed, generic framework that uses utility in decision making to drive the data mining process and we use concepts from meta-learning that uses decision theory for formulating a utility measure, to concentrate the framework for mine actionable combined patterns with composite items. Moreover, this paper proposes an adequate approach of mining actionable association rules with composite items integrating combined association rules and association rules with composite items. The algorithm is experimenting with two different datasets namely Zoo and Mushroom that are retrieved from the UCI repository of machine learning database. The experimental results show that our proposed method predicts the relevant memory requirement and also reveals that it consumes less execution time than the existing method. Our proposed algorithm outperforms the existing method.

REFERENCES

[1] L. Cao, "Actionable Knowledge Discovery," 2009.
[2] L. Cao, "Combined Mining: Analyzing Object and Pattern Relations for Discovering Actionable Complex Patterns," 2012.
[3] L. Cao, H. Zhang, Y. Zhao, D. Luo, and C. Zhang, "Combined mining: discovering informative knowledge in complex data," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, vol. 41, pp. 699-712, 2011.
[4] L. Cao, H. Zhang, Y. Zhao, and C. Zhang, "General Frameworks for Combined Mining: Discovering Informative Knowledge in Complex Data."
[5] A. Deshpande and A. Mahajan, "Domain Driven Multi-Feature Combined Mining for Retail Dataset."
[6] Y. Elovici and D. Braha, "A decision-theoretic approach to data mining," Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, vol. 33, pp. 42-51, 2003.
[7] F. Han, "Mining Actionable Combined Patterns with Composite Items," Journal of Convergence Information Technology, vol. 6, 2011.
[8] I. D. Hunyadi, "Performance algorithms in generating association rules."
[9] P. Kanikar and K. Shah, "Extracting Actionable Association Rules from Multiple Datasets," International Journal of Engineering Research and Applications, vol. 2, pp. 1295-1300, 2012.
[10] S. Laxman, P. Naldurg, R. Sripada, and R. Venkatesan, "Connections between mining frequent itemsets and learning generative models," in Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on, 2007, pp. 571-576.
[11] J. Li and J. Dang, "A Survey on Actionable Knowledge Discovery Applications," in Intelligent Systems and Applications (ISA), 2010 2nd International Workshop on, 2010, pp. 1-3.
[12] B. Ravindran, "Mining Actionable Patterns," 2010.