ECSS: EFFICIENT CONSTRAIN BASED SOFT SET APPROACH FOR ASSOCIATION RULE MINING

Ujala Gupta[1], Guide Name[2]
[1]Research Scholar, [2]Head of Department
Computer Science and Engineering, Shri Satyasai Institute, Sihore (M.P.), India.

Abstract: An extensive research has been taken on association rule mining to improve the performance of the process. The traditional approach to handle the uncertainty present in the dataset and finding association rule is based on Rough set approach which operates on huge item sets available in the dataset which is very large[1,8,16]. Since the size of dataset is huge and it contains large number of items that rarely occur and also contains items that are false frequent in nature. These item take lots of time and take large amount of system memory, due to which the efficiency of traditional approach is affected which makes it inefficient. To overcome these problem and in order to increase the efficiency of previous approach we proposed a new algorithm which is known as Efficient Constraints based Soft set (ECSS) approach. In the proposed approach, first find the item that is rarely occurring and false frequent in nature with the help of initial reduced support threshold. After finding these items we can remove these items from dataset and then apply the Soft set on this reduced dataset and find the association rule. Removal of these items from the dataset in early steps, reduces the time and memory space at every step of the algorithm and hence, the final result has less time and space complexity which is more efficient and effective than previous methods and produce results with more accuracy.

Key Terms: Associations rule, Data mining, soft set, Apriory Algorithm.

I. INTRODUCTION

Association rule is one of the most popular data mining techniques and has received considerable Attention, particularly since the publicationof the AIS and Apriori algorithms [2,3]. They are particularly effective and useful for finding relationships among data in huge databases and applicable to many different domains including market area and risk analysis in commercial environments, epidemiology, clinical medicine, fluid dynamics, astrophysics, and crime prevention. The association rules are considered effective if it satisfies certain constraints, i.e. Predefined minimum Support (min_sup) and minimum confidence (min_conf) thresholds.For Rule X→Y their support and confidence is calculated as:

$$Sup(X \to Y) = \frac{(XUY).Count}{N}$$

N = Total number of transaction

$$Conf(X \to Y) = \frac{(XUY).Count}{X.count}$$

In this X is antecedent and Y is consequent. The rule X→Y has support s% in the transaction set D if s% of transactions in D contain X∪Y. The rule has confidence c% if c% of transactions in D that contain X also contain Y. The goal of association rule mining is to find all the rules with support and confidence exceeding user specified thresholds. Many algorithms of association rulesmininghave been proposed. The association rules method was developed particularly for the analysis of transactional databases.A huge number of association rules can be found from a transactional dataset. The rules that satisfy the minimum support threshold and minimum confidence threshold is called the strong association rules and rest of the rules is discarded. Soft set theory [7], proposed by Molodtsovin 1999, is a new general method for dealing with uncertain data. Soft setsare called (binary, basic, elementary) neighborhood systems. As for standard soft set, it may be redefined as the classification of objects in two distinct classes, thus confirming that soft set can deal with a Boolean-valued information system. Molodtsov [7] pointed out that one of the main advantages of soft set theory is that it is free from the inadequacy of the parameterization tools, unlike in the theories of fuzzy set [8].Since the „„standard" soft set (F,E) over the universe U can be represented by a Boolean-valued information system, thus a soft set can be used for representing a transactional dataset. Therefore, one of the applications of soft set theory for data mining is formining association rules. However, not manyresearches have been done on this application. Moreover Association rule mining is defined as a pair (F,E) is called a soft set over U, where F is a mapping given by:

$$F: E \to P(U)$$

In other words, a soft set over U is a parameterized family of subsets of the universe U. For e belongsE, F(e) may be considered as the set of e-elements of the soft set (F,E) or as the set of e-approximateelements of the soft set. Clearly, a soft set is not a (crisp) set. To illustrate this idea, let we consider the following example.

Example . Let we consider a soft set (F, E) which describes the "attractiveness ofhouses" that Mr. X is considering to purchase.Suppose that there are six houses in the universe U under consideration,

U = { h1, h2, h3, h4, h5, h6}and

E = { e1, e2, e3, e4, e5}

is a set of decision parameters, where e1stands for the parameters "expensive", e2stands for the parameters "beautiful", e3stands for the parameters "wooden", e4stands for the parameters "cheap", e5stands for the

parameters "in the green surrounding".Consider the mapping from equation
F: E → P (U),
Given by "houses (.)", where (.) is to be filled in by one of parameters ebelongs toE.Suppose that
F (e1) = {h2, h4}
F (e2) = {h1, h3}
F (e3) = {h3, h4, h5}
F (e4) = {h1, h3, h5}
F (e5) = {h1}
ThereforeF(e1)means "houses (expensive)", whose functional value is the set { h2 ,h4}. Thus, we can view the soft set (F, E) as a collection of approximations as below

$$(F,E)=\begin{Bmatrix}\text{expensive houses}=\{h_2,h_4\},\\ \text{beautiful houses}=\{h_1,h_3\},\\ \text{wooden houses}=\{h_3,h_4,h_5\},\\ \text{cheap houses}=\{h_1,h_3,h_5\},\\ \text{in the green surrounding houses}=\{h_1\}\end{Bmatrix}$$

Fig 1:Soft set Example

Each approximation has two parts, a predicate p and an approximate value set v. For example, for the approximation "expensive houses = { h2 ,h4 } ", we have the predicate name of expensive houses and the approximate value set or value set { h2 ,h4 } .Thus, a soft set (F, E) can be viewed as a collection of approximations below: (F,E)= { p1 = v1, p2 = v2 ,p3 = v3,….., pn = vn },Binary representation can be represented as:

U	e1	e2	e3	e4	e5
h2	1	0	0	0	0
h3	0	1	1	1	0
h4	1	0	1	0	0
h5	0	1	1	0	0
h6	0	0	0	0	0

Fig2. Soft set in Boolean system

Now here we summarize our paper Second section describe the review of literature. Section 3 describe our proposed approach and section 4 describe our implementation and result of proposed ECSS approach and section 5 conclusion and feature work.

II. REVIEW OF LITRATURE

Number of extensive research effort has been conducted on Association rule mining here we briefing some previous approach for rule mining. In the previous work a soft set approach for association rule mining there are direct applicability of soft set on the Boolean valued information system that contains large number of false frequent item and also contains rare items whose support is less than initial support. Due to the presence of such items in database the previous approach is slow in result generation. These false frequent item and rare item is neither be frequent and no interesting rule is generated with the help of these items. These items is removed when we generated the frequent pattern latter in the process with the help of min_sup. If these item not deleted from input transaction then time complexity and space complexity of the approach is increased. Therefore previous approach has high time and space complexity [1]. In the previous app methods proposed to found out association rule from the transaction dataset. These method is based on Rough set [16] to find association rule. In these method rough set is used to find the association rule on the basis of decision table .In these methods first of all find the conditional attribute and on the basis of which we construct the decision table. This decision table is used to find the association rulesin the IF-THEN context. With the help of Rough set for association rule we find rule with less response time than traditional techniques [14, 15] of association rule mining. But in the rough set based approach the decision table is maintain and then association rule is derived from that decision table is also time consuming in rule generation

III. PROPOSED EVBOR SYSTEM FOR 3-D OBJECT RETRIVAL

A. Proposed Architecture

In our proposed approach we reduce the size dataset with the help of initial red_sup.Due to this the false frequent items and rare items is eliminated or removed from the input transaction dataset. Due to this response time of rule generation is increased .The overall algorithm step of our proposed approach isdepicted as.

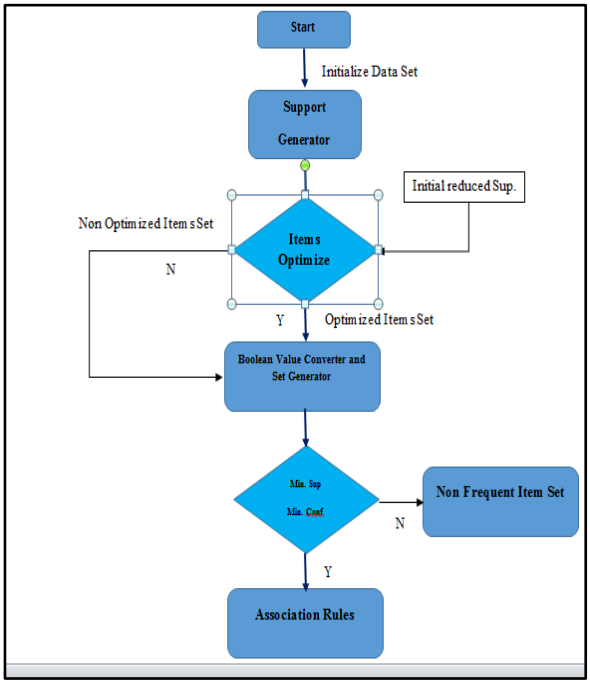

Fig 3. Architecture of Proposed ECSS Algorithm for Association rule mining

Algorithm Steps:
Step 1 :- Scan the dataset D for all transaction 1 to N and calculate the Support for all items.
Step 2 :- for all items in dataset
If initial reduced sup is greater than item support than remove that item from transaction dataset. Therefore we get the more accurate and reduced form of dataset.
Step 3:- Convert the reduced dataset obtained into Boolean valued information system S=(U,A,V{0,1},F) and generate item sets.
Step 4:- Apply the soft set (F,E) on the Boolean valued information system S .
Step 5:- Apply the principle of parameter co-occurrence (min sup and min conf) and calculate the count of various itemsets.
Step 6:- Generate the association rule from the frequent pattern and check with min_conf threshold to find out the rule is strong or not.
Step:-7 End.

IV. RESULT ANALYSIS

In this section, we compare the proposed ECSS method for association rules mining with the algorithm of [1]. The proposed approach ECSS and Previous approach is executed on derived dataset. The algorithm of the proposed approach is implemented in MATLAB. A Dataset derived from the widely used Reuters-21578.It contains 30 transactions with TIDs 1 to 30 and contains 10 items labelled P1 to P10.Now we show the execution time graph between ECSS approach and existing Soft set approach.In execution graph the X-axis indicate the 6 function of approaches and Y-axis indicate time in second.After which we show the bar graph of memory used between ECSS approach and Soft set approach and ultimately we give the table that compare execution time differences as the Min_sup and Min_conf threshold is change.

Figure 4: Graph show the execution time difference between previous soft set approach and our proposed approach ECSS when min_sup=2 and min_conf=0.6

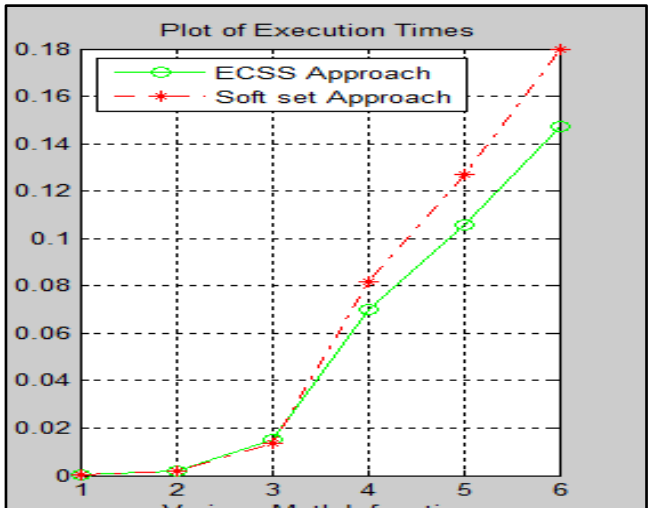

Figure 5: Graph show the execution time difference between previous soft set approach and our proposed approach ECSS when min_sup=3 and min_conf=0.6

Minimum support threshold	Minimum confidence threshold	soft set approach (execution time in sec)	ECSS approach (execution time in sec)
2	0.6	0.283	0.0625
3	0.6	0.05314	0.03952
4	0.6	0.04404	0.02727
5	0.6	0.0508	0.0216

Table 5.1: Table show result of both soft set approach and ECSS approach when taking different supports and generate result in term of execution time in sec.

V. CONCLUSION ANDFUTURE WORK

ECSS approach for association rule mining is a novel method for finding association rule .With the help of ECSS we can handle the uncertainty present in the dataset. The existing approach has more time and space complexity and also has chances of some inaccurate result due to the presence of some false frequent items and rare items that never be frequent. In our proposedapproach firstly we reduce these items from input transaction dataset with the help of initial red_sup and then convert that reduced dataset in to Boolean valued information system.In the next step we apply soft set to handle uncertainty of information system.Now we apply the parameter co-occurrence on the soft set to generate the count of various itemset and then generate the resulting strong association rule.In our approach due to removal of

false frequent items and rare items the space and time complexity is reduced and the generate the result with less time and take less memory space and same accurate than previous approach[1].In feature we can automate it for generating association rule based on optimum auto selected confidence and market support.

REFERENCES

[1] T. Herawan, M.M. Deris A soft set approach for association rule mining Knowledge-Based Systems 24 (2011) 186–195.

[2] R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large databases, in: Proceedings of the ACM SIGMOD International Conference on the Management of Data, 1993, pp. 207–216.

[3] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), 1994, pp. 487–499.

[4] M. Mat Deris, N.F. Nabila, D.J. Evans, M.Y. Saman, A. Mamat, Association rules on significant rare data using second support, International Journal of Computer Mathematics 83 (1) (2006) 69–80.

[5] A.H.L. Lim, C.S. Lee, Processing online analytics with classification and association rule mining, Knowledge-Based Systems 23 (3) (2010) 248–255.

[6] Y.L. Chen, C.H. Weng, Mining fuzzy association rules from questionnaire data, Knowledge-Based Systems 22 (1) (2009) 46–56.

[7] D. Molodtsov, Soft set theory-first results, Computers and Mathematics with Applications 37 (1999) 19–31.

[8] L.A. Zadeh, Fuzzy set, Information and Control 8 (1965) 338–353.

[9] P.K. Maji, A.R. Roy, R. Biswas, An application of soft sets in a decision making problem, Computers and Mathematics with Applications 44 (2002) 1077– 1083.

[10] D. Chen, E.C.C. Tsang, D.S. Yeung, X. Wang, The parameterization reduction of soft sets and its applications, Computers and Mathematics with Applications 49 (2005) 757–763.

[11] A.R. Roy, P.K. Maji, A fuzzy soft set theoretic approach to decision making problems, Journal of Computational and Applied Mathematics 203 (2007) 412–418.

[12] Y. Zou, Z. Xiao, Data analysis approaches of soft sets under incomplete information, Knowledge Based Systems 21 (2008) 941–945.

[13] Z. Kong, L. Gao, L. Wang, S. Li, The normal parameter reduction of soft sets and its algorithm, Computers and Mathematics with Applications 56 (2008) 3029–3037.

[14] R. Feldman, Y. Aumann, A. Amir, A. Zilberstein, W. Klosgen, Maximal association rules: a new tool for mining for keywords cooccurrences in document collections, in: The Proceedings of the KDD 1997, 1997, pp. 167–170.

[15] A. Amir, Y. Aumann, R. Feldman, M. Fresco, Maximal association rules: a tool for mining associations in text, Journal of Intelligent Information Systems 25 (3) (2005) 333–345.

[16] J.W. Guan, D.A. Bell, D.Y. Liu, The rough set approach to association rule mining, in: The Proceedings of the Third IEEE International Conference on Data Mining (ICDM‟03), 2003, pp. 529–532.

[17] P.K. Maji, R. Biswas, A.R. Roy, Soft set theory, Computers and Mathematics with Applications 45 (2003) 555–562. [18] Y. Bi, T. Anderson, S. McClean, A rough set model with ontologies for discovering maximal association rules in document collections, Knowledge-Based Systems 16 (2003) 243–251.

