Discovering Coherent Association Rules using Propositional Method
Geetha G1 Veena K R2 Dr Prashanth C M3
1P.G. Student 2Assistant Professor 3Professor & HOD
1,2Department of Computer Science & Engineering
1,2,3SCE Bangalore, India

Abstract— Another system for Information mining through which fascinating association rules called reasonable guideline can be found. Lucid guidelines are those rules that can be mapped to sensible comparability as per propositional method. In the event that positive intelligent standards are found, then negative rational principles between the nearness and nonattenuance of the same thing sets won’t happen and the other way around. The pseudo ramifications of equivalences can be further characterized into an idea called Reasonable Guidelines. The disclosure of coherent rules from exchange records sort of dataset is called Market Basket Analysis. The proposed work provides a framework for discovering coherent rules that overcomes the limitations associated with existing methods and enables the finding of association rules among the presence and absence of a set of items without a preset minimum support threshold.

Key words: Information Mining, Association Rule, Coherent Rule, Propositional Method, Minimum Support, Market Basket Analysis

I. INTRODUCTION

With the introduction of the PC, scientists were given the force and comfort to investigate and find intriguing and non-evident information from expansive databases. The procedure of getting this learning through the PC is known as Information Mining. The prominence and significance of information mining has its roots in two causes: the always expanding volume of in-formation and calculation power. The measure of data on the planet duplicates like clockwork. Business exercises, for in-stance, keep on producing an expanding stream of information which is put away in bigger and less expensive information stockpiling [1]. Meanwhile, the computational force accessible keeps on expanding. The result of the expanding volume of information and computational force is a chance to make information mining applications taking into account calculations to find intriguing learning from extensive volumes of information.

A. Market Basket Analysis

Various capacities are utilized as a part of information mining including, for instance, join examination, expectation and perception. A connection examination commonly finds the information of “what runs with what” and “what takes after what”. The last is called arrangement investigation and recognizes a grouping of occasions, while the previous is known as fondness examination [2]. A case of fondness examination in the retail division is the Business sector Wicker container Investigation (MBA). Given an arrangement of retail exchange records, a MBA discovers relationship between the diverse things that clients place in their shopping market crate. A few things are frequently bought together and different things are most certainly not. For instance, thing A is frequently acquired together with thing B. Finding these associations portrays clients’ purchasing propensities. Knowing such associations helps a retailer to devise successful promoting techniques. An advancement to build the offer of any one thing inside an association could expand the offers of another thing.

One great way to deal with finding the examples that go together is by means of the backing and certainty system proposed by Agrawal, Imieliński and Swami. Utilizing this structure, designs that can be watched every now and again in an arrangement of exchange records are recognized. To distinguish these examples, the system requires a client to preset an edge that isolates as of-ten as possible watched designs from rare examples. This limit is known as a minimum support. Later, arrangements of things that have seemed together over this minimum support are looked. Decides that interface two arrangements of much of the time watched things that have seemed together over a support are found and a second measure of interestingness, for example, certainty [3].

B. Motivation

Finding a complete arrangement of association rules in information mining. The unfavorable impacts of settling on choice taking into account deficient data can be exorbitant to an association. The unfavorable impacts are an outcome of the accompanying reasons:

1) It is misleading to report an inadequate arrangement of tenets and in the meantime make a feeling that all accessible standards have been found. This circumstance deludes a chief into imagining that just these guidelines are accessible which thusly will lead a leader to dissuade fragmented data. Dissuading fragmented data while not realizing that it is deficient may prompt wrong choices [4].

2) Because of the substantial measure of tenets accessible, a client regularly designs an association rule mining calculation to yield just the most grounded guidelines. It is dangerous to make examination in view of the reporting of the most grounded accessible standards from the computational pursuit that does not cover a complete arrangement of guidelines [5]. There is no assurance that the most grounded association rules found are in reality the most grounded when different standards that might be covered up are considered. It is conceivable that the most grounded tenet lies among the shrouded rules. This circumstance can again prompt a basic leadership unconsciously reaching mistaken determinations about the relationship among things in a dataset.

3) Reporting an association that disregards the nonattendance of things in a given exchange record amid the information mining procedure is deluding. For instance, to report that thing A is connected with thing B is deceiving if a more grounded association can be
found between thing A and the nonattendance of thing B. Once more, improper choices might be made as an outcome[6].

The above reasons spurred to look for a complete arrangement of association decides that incorporate association rules including rarely watched things and truant things in every exchange record.

II. LITERATURE SURVEY

1) To create the association standards to dodge loss of principles (both positive and negative tenets are found). Coherent tenets are found from the propositional rationale so it doesn't require the area master. The coherent principles are found without knowing the estimation of least edge. Along these lines, there is no need of choosing the limit esteem. The outcomes are contrasted and consequences of Apriori calculation:
   - Taking diverse minimum support values on X pivot and number of principles produced on Y hub.
   - Comparison results demonstrate that: The quantity of principles created by coherent hunt calculation is free of minimum support.
   - Issues with apriori calculation are:
     - The calculation utilizes minimum support to find the rules.
     - Minimum support value to be chosen precisely.
     - The generated principles must have support more than the minimum support.

2) Which takes minimization of XOR-genuine worth into thought and creates high certainty and dependable numeric co-herent rules:
   - They proposed a High Numeric Coherent Association Rule Mining (HNCARM) calculation. This contains two stages:

A. Pre-Handling Step
   In preprocessing step, numeric characteristics are changed over into Boolean properties

B. Post Handling Step
   In post handling step, rules, having specific all out class property in its resulting, are extricated.

The proposed methodology is implemented with two benchmark datasets and generates encouraging results with strong and efficient numeric coherent rules.

3) Feature selection method and pruning technique to reduce the complexity of coherent algorithm to discover interesting positive and negative rules for business which are discovered based on the properties of propositional logic and thus do not require the minimum support threshold.

   - There is no chance of neglecting minimum support threshold as the large number of association rules generated missed some interesting rules discovered.
   - The procedure of basic leadership with these tenets may prompt undesirable and sudden results.
   - Minimum support hence assumed a basic part in the whole process.
   - To expel this reliance on minimum support value.

4) Association rules which we get thus incorporate thing sets that are often and rarely seen in set of exchange records. There is no loss of any rules. Notwithstanding that likewise we found exam-ple in light of produced rules which are more effective.

I) Real Troubles
   - Threshold must be exact so area sends out examination is im-portant.
   - The information of interest more likely than not happened much of the time at any rate equivalent to the limit.
   - To recognize the learning looked for by an expert the single threshold is utilized.

II) Highlights
   - Flexible Database Similarity
   - Finds the characteristic threshold.
   - Ability area individual is not required for setting minsupport.
   - ased on created standard we found Productive Example
   - Holes Recognized

1) The client needs to give a precise minimum support value to get the right association rules from the dataset.

2) Negative rules are not excluded which are additionally imperative standards.

III. PROPOSED SYSTEM

A. System Architecture

   ![System Architecture Diagram]

Fig. 1: System Architecture

To generate the both positive and negative association rules for the transaction dataset can be categorized into different phases

1) Frequent item set generation module
2) Generating association rules for the transaction dataset by using the existing algorithm.
3) Generating coherent association rules for the same transaction dataset.

B. Frequent Pattern Mining

The FP-Growth Algorithm, proposed by Han in, is an efficient and scalable method for mining the complete set of frequent patterns by pattern fragment growth, using an extended prefix-tree structure for storing compressed and crucial information about frequent patterns named frequent-
pattern tree (FP-tree). The FP-Growth Algorithm is an alternative way to find frequent itemsets without using candidate generations, thus improving performance. The core of the method is the usage of a special data structure named frequent-pattern tree (FP-tree).

1) Algorithm- Frequent Item Mining

Step 1: Generate I-itemset
FOR each Item ID in TransactionVertical
Step 2: Item = TransactionVertical.get(ItemID)
Support = Item.getTransactions().Count()
IF Support >= MinimumSupport
THEN FrequentItemSet.add(Item)
END IF
Step 3: AllFrequentItemSets.add(FrequentItemSet)
KFrequentItemSet = NULL
Step 4: Support = K-Itemset.getTransactions().Count()
IF Support > MinimumSupport
THEN FrequentItemSet.add(K-Itemset)
END IF
Step 5: IF KFrequentItemSet = NULL
THEN
Save(AllFrequentItemSets)
RETURN
END IF
Step 6: AllFrequentItemSets.add(KFrequentItemSet)
KFrequentItemSet = NULL
K = K + 1
END FOR

C. Association Rule Mining

Association rule mining is a procedure which is meant to find frequent patterns, correlations, associations, or causal structures from data sets found in various kinds of databases such as relational databases, transactional databases, and other forms of data repositories. The current module finds the association rules from the frequent itemsets generated, for the rules satisfying minimum support and confidence threshold specified by the user [7].

Support: The support value of \( X \) with respect to \( T \) is defined as the proportion of transactions in the database which contains the item-set \( X \).

In formula:

\[
supp(X) = \frac{\text{Occurrences of } X}{\text{Total no of transactions}}
\]

Confidence: The confidence value of a rule, \( X \Rightarrow Y \), with respect to a set of transactions \( T \), is the proportion of the transactions that contains \( X \) which also contains \( Y \) [8].

Confidence is defined as:

\[
\text{conf}(X \Rightarrow Y) = \frac{\text{supp}(X \cup Y)}{\text{supp}(X)}
\]

1) Algorithm- Association Rule Mining

Step 1: Confidence = Support(RuleAntecedent U RuleConsequent) / Support(RuleAntecedent)
Step 2: IF Confidence >= MinimumConfident
THEN AssociationRules.add (rule)
END IF
Step 3: Sort(AssociationRules)
FOR index 1 to N
START
Step 4: Print(AssociationRules[index])
END FOR

D. Coherent Rule Mining

Coherent rules are those associations that can be mapped to logical equivalences according to propositional logic. Both Positive and negative rule can be generate without any minimum support factor into consideration [9].

1) Algorithm- Coherent Rule Mining

Step 1: Initialize the coherent Rules to null
Step 2: Find the frequent items from the transaction data
Step 3: Find the items in the unique items but not in the consequent items
Step 4: Generate the powerset and convert to binary representation
Step 5: Generate the +ve Rules for the item set which satisfies minimum confidence value
Step 6: Generate the –ve rules for the item set which also satisfies minimum confide.

IV. RESULTS

Figure 1 and figure 2 to are the inferences from the outputs of the proposed work. The existing association rule mining algorithms only able to generate the first graph which represents the +ve association rule generation. Where as in the proposed algorithm both +ve and –ve rules are going to generate which could by represented by the below figures.

ACKNOWLEDGEMENT

I am thankful to ‘Ms. Veena K R’, Assistant Professor, Dept of CSE for her valuable advice and support extended without which I could not have been able to complete the paper. I express deep thanks to ‘Dr. Prashanth C M’, Head of Department (CS&E) for warm hospitality and affection towards me. I thank the anonymous referees for their
reviews that significantly improved the presentation of this paper. I take this opportunity to express my sincere thanks to all staff members of CS&E department of SCE for the valuable suggestion.

V. CONCLUSION

In this proposed work, the new algorithm for finding the coherent association rules. The previous algorithms only discovers the positive rules only. But in this algorithm both positive and negative rules are discovered by using the propositional method. Hence, the loss of rules is reduced to very small. The accuracy in decision making can also be improved.

REFERENCES


