International Journal of Computational Engineering Research||Vol, 03||Issue, 4||

Association Rule Mining by Using New Approach of Propositional Logic

[1,]Prof. M.N.Galphade, [2,]Pratik P. Raut, [3,]Sagar D. Savairam, [4,]Pravin S. Lokhande, [5,]Varsha B. Khese

[1] *Prof. M.N. Galphade, Computer Engineer, Sinhgad Institute of Technology, Lonavala, Pune India.*
[2] *Pratik P. Raut, Computer Engineer, Sinhgad Institute of Technology, Lonavala, Pune India.*
[3] *Sagar D. Savairam, Computer Engineer, Sinhgad Institute of Technology, Lonavala, Pune India.*
[4] *Pravin S. Lokhande, Computer Engineer, Sinhgad Institute of Technology, Lonavala, Pune India.*
[5] *Varsha B. Khese, Computer Engineer, Sinhgad Institute of Technology, Lonavala, Pune India.*

Abstract :

In Data mining we collecting, searching, and analyzing a large amount of data in a database, as to discover rules and patterns. In data mining field association rules are discovered according to minimum support threshold. The accuracy in setting up this threshold directly influences the number and the quality of association rules. The algorithms which are already exist required domain experts to set up the minimum support threshold but if single lower minimum threshold is set too many association rule is generated and if threshold set high then those association rules involving rare items will not be discovered. The risks associated with incomplete rules are reduced because in proposed algorithm association rules are created without the user having to identify a minimum support threshold. . The proposed algorithm discovers the natural threshold based on observation of data set .We propose a framework to discover domain knowledge report as coherent rules. Based on the properties of propositional logic, and therefore the coherent rules are discovered.

Keywords: *data mining, association rules, support, confidence a minimum support threshold, material implication, patterns.*

I. INTRODUCTION

Association Rule Mining (ARM) is a unique technique. It has the advantage to discover knowledge without the need to undergo a training process. It discovers rules from a dataset, and each rule discovered has its importance measured against interesting measures such as support and confidence [2].To describe the associations among items in a database the association rule mining technique is used. It is also useful to identify domain knowledge hidden in large volume of data efficiently. The discovery of association rules is typically based on the support and confidence framework. To start the discovery process a minimum support (min sup) must be supplied. A priori is one of the algorithm which is based on this framework. Association rules can be discovered without this threshold specified,because the procedure to discover the rules will quickly exhaust the available resources.Nonetheless, having to constrain the discovery of association rules with a preset threshold, in turn, requires in-depth domain knowledge before the discovery of rules can be automated. The use of min support generally assumes that:

- Threshold value accurately provided by the domain experts.
- The knowledge of interest must have occurred frequently at least equal to the threshold.
- To identify the knowledge sought by an analyst the single threshold is used.

In practice, there are cases where these assumptions are not appropriate and rules reported lead to erroneous actions. In this paper, we propose a novel framework to address the above issues by removing the need for a minimum support threshold. The proposed algorithm discovers the natural threshold based on observation of data set. Associations are discovered based on logical implications. The principle of the approach considers that an association rule should only be reported when there is enough logical evidence in the data. To do this, we consider both presence and absence of items during the mining. An association such as bread => milk will only be reported if we can also find that there are fewer occurrences of ¬ bread => milk and bread =>¬ milk but more of ¬ bread =>¬milk. This approach will ensure that when a rule such as hair => milk is reported, it indeed has the strongest statistical value in the data as comparison was made on both presence and

absence of items during the mining process. In addition, the inverse case of customer not buying bread and customer not buying milk should have statistics that support the rule being discovered due to the logic properties of an equivalence.

1.1 Minimum Support Threshold-

Before association rules are mined, a user needs to determine a support threshold in order to obtain only the frequent item sets. Having users to determine a support threshold attracts a number of issues. We propose an association rule mining framework that does not require a pre-set support threshold [2].

II. ASSOCIATION RULE MINING

An association rules describes association relationships among set of data.Association rules are statements of the form {X1,X2……..Xn) ⇒ Y , meaning that if we Find all ofX1,X2…… Xn in the market basket, then we have a good chance of finding Y . The probability ofFinding Y for us to accept this rule is called the confidence of the rule. We normally would searchonly for rules that had confidence above a certain threshold. We may also ask that the confidence beSignificantly higher than it would be if items were placed at random into baskets. For example, we might find a rule like (milk, butter)⇒bread simply because a lot of people buy bread. However, theBeer/diapers story asserts that the rule (diapers)⇒ beer holds with confidence significantly greater than the fraction of baskets that contain beer.We propose a novel association rule mining framework thatcan discover association rules without the need for aminimum support threshold. This enables the user, in theory,to discover knowledge from any transactional record withoutthe background knowledge of an application domainusually necessary to establish a threshold prior to mining.

This section starts with the distinction between an association rule and the differentmodes of an implication as defined in propositional logic.The topic of implication from logic is raised because ourproposed mining model is based on an association rule'sability to be mapped to a mode of implicationIf an association can be mapped to an implication, then there isreason to report this relation as an association rule.An implication having a rule where the left-hand side is connected to the right-handside correlates two item sets together. This implication existsbecause it is true according to logical grounds, follows aspecific truth table value, and does not need to be judged tobe true by a user. The rule is reported as an interestingassociation rule if its corresponding implication is true.

2.1. An Implication-

In an argument, the truth and falsity of an implication (alsoknown as a compound proposition) (→) necessarily rely onlogic. Each implication, having met specific logicalprinciples, can be identified each has a set of different truth valuesWe highlight here that an implication isformed using two propositions p and q. These propositionscan be either true or false for the implication's interpretation.From these propositions, we have four implications

1. p→q,
2. p→¬q,
3. ¬p→q and
4. ¬p→¬q.

Each is formed using standard symbols "→" and "¬" Thesymbol "→" implies that the relation is a mode of implication in logic, and "¬" denotes a false proposition.The truth and falsity of any implication is judged by "anding"(^) the truth values held by propositions p and q.
In a fruit retail business where no bread is sold, the implicationthat relates p and q will be false based on the operationbetween truth values; that is, 1 ^ 0 = 0. The second implicationbased on the operation will be true because 1 ^ 1 = 1.Hence, we say that the latter implication p →¬q is true, butthe first implication p →q is false. Each implication has itstruth and falsity based on truth table values alone.We highlight two modes of implication and their truthtable values in the next two sections.

2.1.1. Material Implication-

A material implication (⊃) meets the logical principle of acontraposition. A contrapositive (to a material implication) iswritten as ¬q →¬p. For example, suppose, if customers buyapples, that they then buy oranges is true as an implication.The contrapositive is that if customers do not buy oranges,then they also do not buy apples. If an implication has thetruth values of its contrapositive, ¬(p ∧¬q) it is a materialimplication. That is, p ⊃ q iff¬(p∧¬q).[1]The truth table for a material implication is shown inTable 1.

p	q	P ⊃ q
T	T	T
T	F	F
F	T	T
F	F	T

Table1. Truth Table for a Material Implication

2.1.2. An Equivalence-

An equivalence (=) is another mode of implication. Inparticular, it is a special case of a material implication. Forany implication to qualify as an equivalence, the followingcondition must be met p = q iff¬(p xor q) wheretruth table values can be constructed in Table 2.

p	q	p= q
T	T	T
T	F	F
F	T	F
F	F	T

Table 2 Truth Table for an Equivalence

One of many ways to prove an equivalence is to show thatthe implications p →q and ¬p →¬ q hold true together. Thelatter is also named an inverse.Suppose, if customers buy apples, that they then buy oranges is a trueimplication. The inverse is that if customers do not buyapples, then they do not buy oranges.[1]

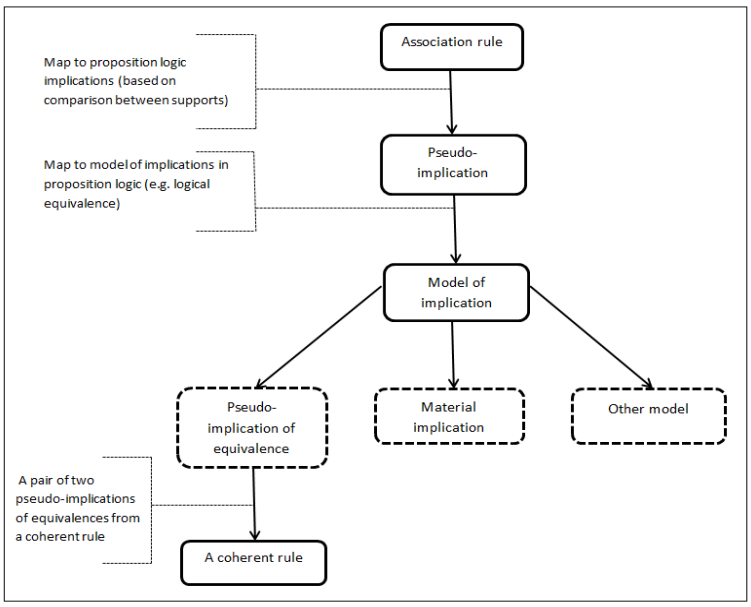

Fig 1. A generalized framework of association rules that based on pseudo implications.

We summarize in this section that a typical statement of the format "if ... then" is a conditional or a rule. If this conditional also meets specific logical principles with a truth table, they are an implication. Among many modes of implications, a material implication relates propositions together. An equivalence is a special case of the former, where propositions are necessarily related together all the time and are independent of user knowledge. In other words, equivalence is necessarily true all the time and judged purely based on logic. We are interested in finding association rules that map to this equivalence. By mapping to this equivalence, we can expect to find association rules that are necessarily related with true implication consistently based on logic. These are the association rules deemed interesting. In addition, the process of finding such association rules will be independent of user knowledge because the truth and falsity of any implication is based purely on logical grounds.

III. CH-SEARCH ALGORITHM

We propose coherent rules by using properties of positive and negative rules on the condition set (positive rule) >set (negative rule) at preselected consequence item set.

We already write a basic algorithm to generate coherent rule in Fig: 2.

Input: D – a database, Y – a consequence item set
Output: CR – a set of coherent rules

[1] $CR \leftarrow \emptyset$
[2] $I \leftarrow$ find a set of unique items from D
[3] Let $A = I - Y$
[4] $Y.count \leftarrow$ total counts of Y in D
[5] $O_{P(A)} \leftarrow$ virtually map the power sets of A to the indices of a binary system
[6] For each i-th element of the power sets of A in order of O_i,
(i) $X \leftarrow \{P_i : i \in P(A)\}$
(ii) $S(X,Y) \leftarrow XY.count$
(iii) $S(\neg X,Y) \leftarrow Y.count - S(X,Y)$
(iv) if $S(X,Y) > S(\neg X,Y)$,
if equation (2) is met, $CR = CR \cup (X,Y)$
Loop [6] until $i = |P(A)|$
(v) remove all power sets of A having the i-th element
[7] return CR

* For example, given 3 items, the first item set *null* – a member in the power sets of X, item set $X_{i=1\_}$ is indexed using binary number '0', item set $X_{i=2\_}$ is indexed using '1_2', and item set $X_{i=3\_}$ is indexed using '10_2'.

Fig 2. A simple search for coherent rules algorithm (ChSearch)

IV. FEATURES OF CH-SEARCH

- Ch-search algorithm does not require preset minimum support to find out association rules. Coherent rule found based on logical equivalence. And these rule further used as association rule.
- In ch-search, no need to generate frequent item set and also there is no need to generate association rule within each and every item set.
- In apriori algorithm, negative rules not found there. But in ch-search algorithm, we found negative rules and use them to implement both positive and negative rules found. (In apriori, database required in binary format and results are contra-dictionary.)

V. PATTERNS

We discovered pattern based on generated rule which are more efficient.

- **Positive Rules**

  When we got association rules some of them consider only itemsenumerated in transactions, such rules are referred to as positive associationrule.

  Ex. bread => milk

- **Negative Rules**

  Negative association rules also consider the same items, but in additionconsider negated items.

  Ex.¬bread =>¬milk.

Algorithm which is presented in paper extends the support-confidence framework with a sliding correlation coefficient threshold.In addition to finding confident positive rules that have a strong correlation,the algorithm discovers negative association rules with strong negative correlation between found the strongest

correlated rules, followed by rules withmoderate and small strength values.After finding the association rules we found that patterns are more efficient than the rules.In association rules only those attributes are considered which are strongly responsible to find the result.

In case of the patterns all the attributes are considered.

Ex.milk >=1 and aquatic >=1 and predator >=1 and toothed >=1 and backbone >=1 and breaths >=1 and fins >=1 and tails >=1 and cat size>=1 and hair=0 and feathers=0 and eggs=0 and airborne=0 and venomous=0 and legs=0 and domestics=0 => MAMMAL
Patterns are more efficient than rules

Features-

- Flexible Database Compatibility
- Discovers the natural threshold.
- Expertise domain person is not required for setting min support value.
- Based on generated rule we discovered Efficient Pattern

VI. CONCLUSION

The proposed algorithm discovers the natural threshold based on observation of data set. Association rule which we get as a result include item sets that are frequently and infrequently observed in set of transaction records. There is no loss of any rule. In addition to that also we discovered pattern based on generated rule which are more efficient.

REFERENCES

[1] Alex Tze Hiang Sim, Maria Indrawan, Samar Zutshi, Member, IEEE, and Bala Srinivasan:"Logic-Based Pattern Discovery"
[2] Alex TzeHiangSim, Maria Indrawan, BalaSrinivasan."A Threshold Free Implication Rule Mining"
[3] Yanqing Ji, Hao Ying, Senior Member, IEEE, Peter Dews, Ayman Mansour, John Tran, Richard E. Miller, and R. Michael Massanari" A Potential Causal Association Mining Algorithm for Screening Adverse Drug Reactions in Postmarketing Surveillance"