ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 12, December 2013

A Speculative Study of Apriori with Inferencifiction(SSAI)

Ruchi Gupta[1], Pooja Saharan[2]

Assistant Professor, CSE Dept, ABES Engineering College / Mahamaya Technical University, Ghaziabad, U.P, India [1]

Abstract: Frequent Pattern discovery is one of the major problem in areas of Data Mining and Business Intelligence, because it reveals associations, correlations, and many other interesting relationships among items in the transactional databases. FPD is a process of searching for patterns such as set of features/ items that appear in data frequently. Our idea is to introduce inference rules in association with association rules that reduces the use of join operation while applying Apriori algo to a transactional database to discover the frequent patterns.

Keywords: Apriori, Inferencifiction, Frequent Pattern Discovery (FPD), Association Rule mining, Inference Based Pattern Discovery.

I. INTRODUCTION

Data Mining is the nontrivial process of extracting implicit, previously unknown and potentially useful information from data [2]. The basic goal of data mining is to discover patterns occurring in the databases, such as associations, classification models, sequential patterns, and so on. In this paper , the focus is on Frequent pattern discovery . Apriori Algorithm has been introduced in this area to resolve this problem. But , I tried to introduce Inference Laws and reducing the use of join operation in Apriori Algorithm.The problem of Frequent Pattern Discovery is defined as the process of searching for patterns such as sets of features or items that appear in data frequently [1]. It is an important phase of Association Mining because , it is a difficult task to search all possible patterns. I pay attention to the designing of frequent pattern discovery by applying the concept of Inference Rules.

Data mining is defined as the process of discovering patterns in data. Data mining is the analysis of data and the use of software techniques for finding patterns and regularities in sets of data . Data mining categorised in two types :recommendation system and web personalisation. Web Personalisation is the process of customizing the content and structure of a Web site to the needs of each specific user or set of users, taking advantage of the knowledge acquired through the analysis of the user's navigational behavior [3]. Recommendation systems are implemented in commercial and non-profit web sites to predict the user preferences .The main functions of recommendation systems include analyzing user data and extracting useful information for further predictions[4] For commercial web sites, accurate predictions may result in higher selling rates. the substantial increase in the number of websites presents a challenging task for webmasters to organize the contents of the websites to cater to the needs of users. The recommendation quality is measured by how many given recommendations are correct, as well as how many clicks the recommendation allows users to skip.Data Mining is only one part of the knowledge discovery process.

II. MOTIVATION

To discover the association rule from a huge database is a very challenging task in today's Query formulating environment. Traditionally, the analysis of data employs machine learning and statistical techniques to find hidden patterns and correlations in the data [1]. Every day we have new type types of queries and to resolve these queries, we require the patterns that we have obtained from past years or from the past records continuously. Let us take an example, a person who is just a newcomer in Teaching- Profession is asked to evaluate the students and categorise them into three categories only as given below :

a) Average Student(B-category)
b) Weak Student(C-category)
c) Above an Average Student(A-category)

then he first try to get a pattern on the behalf of a criteria as follows:

1) How many marks obtained by a student in Test T1?
2) How many marks obtained by a student in Test T2?
3) How many assignments he has submitted?

From the above questions, he got that a student 'X' got 60/100 in test T1 , 70/100 in test T2, 3/5 assignments have been submitted. So, if we calculate an average of these three we get the result as follows:

$$R = (P1/P2)$$

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 12, December 2013

- P1= (Marks obtained in T1 + Marks obtained in T2 + No. Of assignment submitted)
- P2= (Max. Marks in T1 + Max. Marks in T2 + Max. No. Of Assignment)

R = (60+70+3)/205 = 64.87~ 65

From all the above facts, he get a pattern that if (T1,T2 ,A) has (60,70,3) as values then that student is an average student. From these he formulate a rule :

(60, 70, 3) => Avg. Student

The next task is to get the patterns so that our task of computation get reduced. How we can reduce the task of that new teacher ????? just think.....Okay !! yes using a "PREDICATE" . Predicate is a function which takes some input arguments and returns back an output result. It seems very simple that to judge a student we just feed their marks in to a predicate (or function), named as "Evaluation" and get the status of that student. If roughly speaking, then we can design a predicate "Evaluation" as follows :

Evaluate(t1, t2, a)
{
(R>=60 and R<=74) => Average student.
(R<=59) => poor student.
(R>=75) => Above an average student.
}

Here , We pass the data of 6 students like E1(60,70,3); E2(70,60,3); E3(65, 65, 3);E4(60,60,3);E5(70,70,2),E6(60,60,2). For E1 to E5 , it gives an average student but for E6 it gives poor student.
So , what pattern u r finding here???? Just think for a minute ..... YES..
(1) If third value is same and first two get interchange the result will not be affected.
(2) If the first two values are greater than 60 and third value is greater than or equal to 3 then will never give a poor student.
(3) If the first two values are equal to sixty and third value is equal to 2 it results in a poor student.
So, why are forming these patterns ?? Obivously to reduce our task of computation. We will just see the data and go through the trends and can easily categorise each and every student on the basis of that pattern.
I have introduced here a new term "Inferencifiction"i.e. A fiction that inference rules can be used to resolve the queries related with frequent pattern. I am trying to apply the inference as an fiction(is an imaginative creativity which donot show reality but has been invented.) to solve the problem of that teacher.

Similarly , he finalise three categories as:

CATEGORY –I : if 60 <= R <=74 implies an average student.
CATEGORY –I : R <=59 implies a poor student.
CATEGORY –I : if 75 <= R implies above an average student.

PROBLEM STATEMENT

Association Rule Generation process is composed of two major phases:

- Frequent Itemset Mining
- Rule Generation [1].

My focus is to apply the rule of inferences after forming the rules and then deduce the desirable patterns. My work is to introduce inference laws in Apriori algorithm and to get the same results. The methodology that i am opting here reduces the join operation in Apriori algorithm. In this paper, I proved this with the help of an example.

PROPOSED APPROACH (SSAI):

Our approach is a two -step approach

- Frequent Pattern Generation

- Generate all patterns whose min_sup. is satisfied

- Rule Generation

- After join operation a candidate set {I1, I2} can be formulated as either I1->I2 or I2->I1.
- Conjugate two rules if
  - They have same support and
  - Either one of antecedent or consequent is same

⦿ Hypothetical Syllogism
((p->q)∧(q->r))-> (p->r)
⦿ Modus Ponens
(p ∧(p->q))->q
⦿ Conjunction
((p->q)∧(p->r))->(p->q∧r)

Rule Constraint :
-There will be only one variable as antecedent and consequent may be in combination but should not be Ø.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 12, December 2013

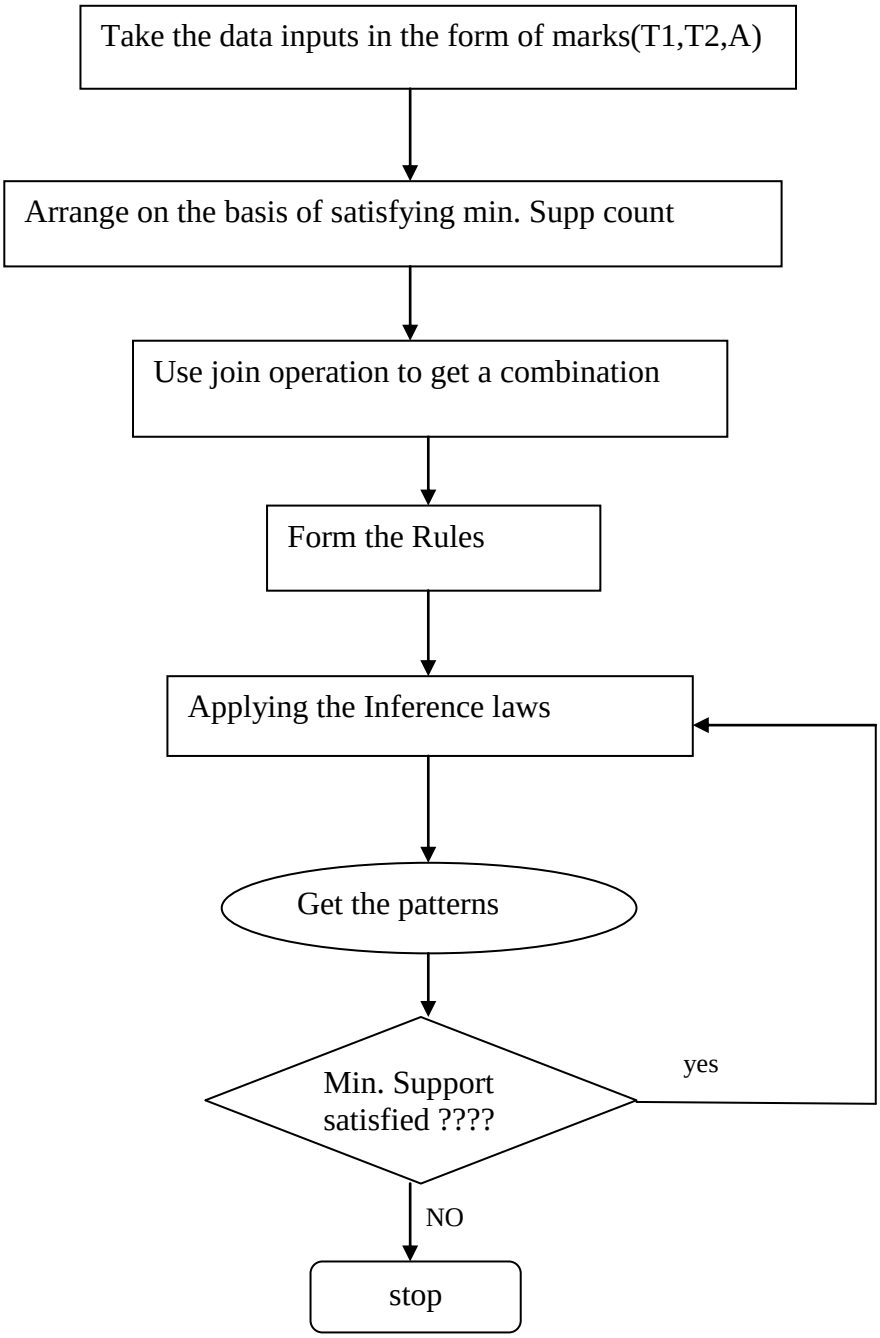

Fig 1. Flow-chart of the Proposed approach(SSAI)

FUTURE SCOPE

Frequent Pattern discovery is one major problem in the areas of data mining, business Intelligence. I suggest that the SSAI approach can be further extended by involving more inference laws like-Modus pollens, Modus Tollens , Disjunctive Syllogism and many more. One more suggestion is that the SSAI aaproach can be efficiently and concisely implemented with high-level declarative language such as – PROLOG. I focus my future research on the introducing some more inference laws to get the frequent discovered patterns and then to design some more constraints and then their implementation.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 12, December 2013

REFERENCES

[1] *Nittaya Kerdprasop and kittisak Kerdprasop, Data Engineering Research unit, Suranaree University of Technology,Thailand*."*Recent Researches in Computational Techniques, Non-linear Systems and control.*"

[2] W.J. Frawley, G. Piatetsky-Shapiro, and C.J. Matheus, /knowledge discovery in databases: An overview, AI Magazine, vol. 13, no. 3,1992, pp.57-70.

[3] Bellogín, A., Cantador, I., Castells, P., & Ortigosa, Á. (2008). "Discovering Relevant Preferences in a Personalised Recommender System using Machine Learning Techniques". Work, 82-96.

[4] Shankar, R., & Pudi, V. (2010). "Evolutionary Clustering using Frequent Itemsets". Database, 3-8

BIOGRAPHIES

Ms. Ruchi Gupta, is working as an Assistant Professor in Computer Science & Engineering Department in ABES Engineering College, affiliated to Uttar Pradesh Technical University (Formerly MTU) Ghaziabad, U.P. , India. She has around more than once of year teaching experience at renowned institutions like Mewar University, Rajasthan. She has completed her graduate and masters in Engineering with HONOURS. She has been one of top ten students of the UNIVERSITY in M.Tech,CSE. She has authored two more research papers in International Journals related to "Multicasting in Networks".

Pooja Saharan is an Assistant Professor of the Department of Computer Science & Engineering at ABES Engineering College, Ghaziabad, affiliated to Uttar Pradesh Technical University, U.P , India. She has received degrees in Computer Science and Engineering from UIET, Maharshi Dayanand University, Rohtak . She has two and half years of teaching experience. She has completed her graduation and masters in Engineering with HONOURS and among one of top five students of the Maharshi Dayanand University, Rohtak, Haryana. She has authored two more research papers in International Journals related to "DATA MINING".