# An Efficient Approach for Incremental Association Rule Mining through Histogram Matching Technique

*Ajay Kumar, Department of Computer Science & Engineering, Jaypee University of Engineering & Technology, Guna, India*

*Shishir Kumar, Department of Computer Science & Engineering, Jaypee University of Engineering & Technology, Guna, India*

*Sakshi Saxena, Department of Computer Science & Engineering, Jaypee University of Engineering & Technology, Guna, India*

## ABSTRACT

*The objective of the work being presented is to propose an approach for obtaining appropriate association rules when the data set is being incrementally updated. During this process raw data is clustered by K-mean Clustering Algorithm and appropriate rules are generated for each cluster. Further, a histogram and probability density function are also generated for each cluster. When Burst data set is coming to the system, initially the histogram and probability density function of this new data set are obtained. The new data set has to be added to the cluster whose histogram and probability density functions are almost similar. The proposed method is evaluated and explained on synthetic data.*

*Keywords: Association Rule Mining, Histogram, K-Means Clustering, Kurtosis, Skewness*

## 1. INTRODUCTION

In present scenario computers are being used in a variety of applications. Large volumes of data sets are being collected based on the type of application in use. These data contain useful information but requires huge analysis. Capability of humans for handling and analyzing data to obtain information is very limited. To overcome this limitation data mining techniques may be used. It has been observed that analysis of past stored data can provide useful information and may lead in improvement in quality decisions making process. To fulfill this requirement there is a need to design efficient algorithms that will find useful information from such data sets (Kamber & Han, 2000). Sufficient literature is available for dealing with such of the problem.

DOI: 10.4018/ijirr.2012040103

Association rules mining (Agrawal, Imielinski, & Swami, 1993; Agrawal & Srikant, 1994; Han, Pei, & Yin, 2004) are one of the most important data mining tasks to find association of data in a given data set. Some popular association algorithms like *Apriori* (Agrawal & Srikant, 1994), *FP growth* (Park, Chen, & Yu, 1995) and *DHP* (Ben-David, Gehrke, & Kifer, 2004) for the market basket type of association rule discovery have also narrated in the literature (Agrawal et al., 1993; Agrawal & Srikant, 1994; Han et al., 2004). In the present era of dynamic information environment the issue is to study the effect in changes of state after addition of the new data set, or there may be needed to modify or delete some or all the existing set of data during the process of data mining.

A lot of work has been done on this problem (incremental Association rule mining). Detecting changes in the data stream (Ben-David et al., 2004) and finding contrast patterns incrementally in changing data (Bailey & Loekito, 2010) is an important issue. The concept of temporal association rules in order to solve the problem of handling time series by including time expressions into association rules Earlier user has to repeat the whole procedure, which is time-consuming, its lack in efficiency also. Actually, temporal databases are continually appended or updated so that the discovered rules need to be updated. Rerunning the temporal mining algorithm every time is ineffective since it neglects the previously discovered rules, and repeats the work done previously (Gharib, Hassar, Taha, & Abraham, 2010).

Association rule process generates a huge number of candidate item-set in each iteration. Candidate item-set needs to be regenerated after every updating and making earlier generated set as fritter away. Through this paper an incremental data mining algorithm will be proposed, using the fundamental of association rule mining for generation of rules along with clustering analysis in order to find groups within the data set. A new similarity measure (Skewness and Kurtosis) has been used to find the appropriate cluster or class of incremental data set by histogram matching technique.

## 2. PRELIMINARY DEFINITIONS

Data mining is the process of extracting interesting (non-trivial, implicit, previously unknown and potentially useful) information or patterns from large information repositories and it is the core process of Knowledge Discovery in Database (KDD) (Zhao & Bhowmick, 2003). Data mining techniques include association rule mining, classification, clustering, mining time series, and sequential pattern mining, to name a few, with association rules mining receiving a significant research attention (Huang, Dai, & Chen, 2007).

Let at any given instant of time 't' there is a transaction dataset D. Here D consists of a set of transactions and each transaction is identified with a transaction identifier, TID. Now let us assume that at time 't+l', the database is updated such that a set of transaction $D^+$ is added to the original database D and a set of transaction $D^-$ is deleted from the dataset. So the resulting modified database D* can be represented as:

$$D^* = (D \cup D^+) - D^-.$$

The incremental model algorithm can be defined as to find the mining model corresponding to the updated database D* from the mining model which corresponds to the database D, without building the mining model from scraps.

### 2.1. Association Rule

Association rule mining was first introduced by R. Agrawal, T. Imielinski, and A. Swami in 1993, where the goal is to discover interesting patterns among items in a given transaction dataset.

A mathematical model was proposed by R. Agrawal, T. Imielinski, and A. Swami in 1993 to address the problem of mining association rules. Let $I = \{i_1, i_2 \ldots I'm\}$ being a set of literals, called items. Let D be a set of transactions, where each transaction T is a set of items such that $T \subseteq I$. It may be noted here that the quantities of items bought in a transaction are not considered, meaning that each item is a binary

Integrating Technical Advance in Mobile Devices to Enhance the Information Retrieval in Mobile Learning

Zhaozong Meng and Joan Lu (2014). *International Journal of Information Retrieval Research (pp. 61-85).*

www.igi-global.com/article/integrating-technical-advance-in-mobile-devices-to-enhance-the-information-retrieval-in-mobile-learning/113333?camid=4v1a