Volume 8, No. 5, May-June 2017
ISSN No. 0976-5697
International Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info

A Survey on Association Rule Mining Algorithms Used in Different Application Areas

PandyaJalpa P.
Asst. Professor,
UCCC & SPBCBA& SDHG College of BCA & IT
Udhna, Surat, Gujrat, India

MorenaRustom D.
Professor,
Department of computer Science,
VNSG University, Gujrat India

Abstract: Data mining is a collection of techniques to extract hidden and potentially useful information from large databases of various business domains. For identifying the interesting patterns and co-relation and to get benefits from the repository data, Association Rule Mining (ARM) methods are used. ARM is one of the most popular techniques of data mining, which is helpful in extracting interesting knowledge. In this paper, the effort is done to survey the research papers associated with real life applications like stock market, health care, Supermarket etc. Various ARM algorithms applied by the different researchers are studied. The main goal of this paper is to identify recent issues and techniques used in different application areas specially related with ARM algorithms in data mining.

Keywords: Data Mining, Data Mining Techniques, Association Rule Mining, ARM algorithms

I. INTRODUCTION

Data Mining is also called as "knowledge Mining" which refers to extracting or mining knowledge from large amounts of data. Data mining carry different meanings such as knowledge extraction, data or pattern analysis. It is also known as knowledge discovery from large data. Mining can be done on huge database systems such as stock market, time-series data, spatial databases, multimedia databases, World Wide Web, text databases, medical databases, criminal databases or some specific application oriented databases. Data mining techniques include clustering, classification, prediction, and link analysis (associations) [26][27].

Data mining is a collection of techniques to extract hidden and potentially useful information from large databases from business domain point of view. For identifying the useful patterns and co-relation between different attributes Association Rule Mining (ARM) technique is used. ARM is utilized in various business domains now a day. We are trying to discuss recent issues and algorithms used in different areas for ARM in data mining.

There are mainly six steps for knowledge discovery [26].

1) Data cleaning: To remove noise and inconsistent data

2) Data integration: Where multiple data sources may be combined.

3) Data Selection: Where data relevant to the analysis task are retrieved from the database.

4) Data Transformation: Where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operation for instance.

5) Data Mining: It is an essential process where intelligent methods are applied in order to extract data pattern.

6) Rule Finding: It is final step which produce meaningful rules.

A. *Data Mining Techniques*

Data classification, clustering, prediction, outlier analysis, association rule mining and characterization-cum-discrimination are the different Data Mining Techniques which are explained in brief below [26][27][19].

1) Data Classification

Classification is the most commonly used supervised data mining technique, whose task is to formulate models or functions that describes and distinguish data classes. These models are used to predict or generate class information for data where class labels are not known. A classification process in which classes have been pre-defined needs a method that will train the classification system to allocate objects to classes. The training is based on training sample, as set of sample data where for each sample the class is already known. Decision tree or neural network-based classification algorithms are the commonly used methods to represent classification models.

2) Data Clustering

Clustering is used to group similar objects or data and associated it with the cluster labels. Clustering is an unsupervised data mining technique. It accepts descriptive model, which identifies the patterns or relationships in data and explores the properties of the data examined. It works on the principle of maximizing intra-cluster and minimizing inter-cluster similarities, resulting in clusters. Clustering and classification are important techniques that partition objects that have many attributes into meaningful disjoint subgroups so that objects in each group are more similar to each other in the values of their attributes than they are to objects in other groups. But major difference between cluster-analysis and classification is that in supervised classification, the classes are predefined, and clustering we does not know what classes or clusters exist in prior.

3) Data Prediction

Prediction is the supervised learning data mining technique that is used to predict mission or unavailable data. This technique comes under predictive model. Prediction

technique is used only for numerical data which is different than classification that predicts class labels. In data prediction most preferable method is regression analysis.CART (Classification and Regression Trees) decision tree algorithm can be used to build both classification trees and regression trees. Neural networks can be used to create both classification and regression models.

4) Association Rule Mining

Association Rule Mining (ARM) is the process of finding relationships amongst data items. It is the process of extracting association rules that identify frequently related data items. The best example of ARM is market basket analysis in which frequently related items are identified in a transaction. An example of frequently related items are (bread, jam) → butter, implying the fact that transactions involving sales of items, bread and jam, also followed by the butter.

5) Outlier analysis

Outlier analysis is the data mining technique that deals with identification of outliers. This technique is used in the identification of fraudulent usage of credit and ATM cards, where one unusual behaviour or outlier could be the withdrawal of a huge amount as opposed to the general withdrawal trend of the customer.

One of the most important data mining applications is mining association rules. Association rules, first introduced in 1993, are used to identify relationships among a set of items in a database.

B.*Association Rule Mining*

The problem of mining association rules between sets of items in a large database of customer transactions is firstly introduced by agrawal in 1993 . The formal statement is "Let I = {i1,i2,…..in} be a set of n binary attributes called items. Let D = {t1,t2,…..tn} be a set of transactions called the database. Each transaction in D has a unique transaction ID and contains a subset of the items in I. A rule is defined as an implication of the form $X \Longrightarrow Y$ where X, $Y \subseteq I$ and $X \cap Y=0$. The sets of items (for short itemsets) X and Y are called antecedent (left-hand-side or LHS) and consequent (right-hand-side or RHS) of the rule". Each transaction consists of items purchased by a customer in a visit. He discovered the concept of Support and confidence. Def^n on support is "The support (s) of an association rule is the ratio (in per-cent) of the records that contain X and Y to the total number of records in the database." Def^n on confidence is "For a given number of records, confidence (α) is the ratio (in percent) of the number of records that contain X and Y to the number of records that contain X."[1][7].

This problem is partitioned into two sub-problems.

1. Finding all itemsets, called large itemsets, that represent in at least s%(given support percentage) of transactions.

2. Generating rules from each large itemsets that useitems from the large itemsets.

C. *Association Rule Mining Algorithms*

Association Rule Mining algorithms are commonly used for Frequent Pattern Mining. An algorithm for finding all association rules, henceforth referred to as the AIS algorithm, was proposed by Agrawal. The AIS algorithm suggested that the large itemsets need not be in memory during a pass over the database and can be disk-resident. In 1994 agrawal[2] has presented two new algorithms, Apriori and AprioriTid that differ fundamentally from these algorithms. Apriori is a candidate generation algorithm and proceeds in a level-wise fashion. It is by far the most well-known association rule algorithm. The algorithm terminates when no further successful extensions are found. Apriori uses breadth-first search and a hash tree structure to count candidate item sets efficiently. The fundamental differences of this algorithm from the AIS algorithms are the way of generating candidate itemsets and the selection of candidate itemsets for counting. This algorithm suffers from limitations of a huge number of repeated input scans. However, variations of apriori exist which minimize the overhead of repeated input scans. One of the variations is algorithm aprioriTID emphasizing the fact that transactions in the database are replaced by candidate item-sets that occur in that transaction. This performs better at lower level since the entry might be longer than the corresponding transaction that evolved the concept of apriori-hybrid which combined the advantages features of apriori and aprioriTID [4]. Another limitation of the apriori-based approach is the generation of candidate sets which can become cumbersome and time consuming when the frequent 1 item sets is large. For eliminating this kind of problems several efficient algorithms such as Frequent pattern (FP) tree growth is introduced. FP-growth algorithms use pattern growth approach that generates patterns on the fly and avoids the candidate pattern logic of apriori. The size of the candidate item-sets can be reduced by adopting a depth first logic as done by FP-growth algorithm. The FP-growth algorithm stores the transactions of the database in a tree format and every item has a linked list that goes through all transactions that contain that item. In 2000, Han proposed the FP-growth algorithm. Because it only scans the database twice and does not use the candidate set, the algorithm became very well-known and efficient. The basic idea is that first sweep the transactional database to find frequent 1-item sets, and then construct the FP-tree. At last it discovers conditional pattern base to mine frequent pattern based on the FP-tree [5].

II. RELATED WORK

SanjeevRao et al. [14] have presented the use of an ARM (Association rule mining) driven application to manage retail businesses that provide retailers reports regarding prediction of product sales trends and customer behaviour. They studied two algorithms specially Apriori and FP-growth on the dataset of super market and concluded that FP-growth is faster than other association mining algorithms and is also faster than tree- Researching. The algorithm reduces the total number of candidate item sets by producing a compressed version of the database in terms of an FP-tree.

Rajesh V. Argiddi et al.[8] has worked for analysing the behaviour of the stock market data and based on this data to predict the future trading of the stock market. They used dataset of BSE, different companies such as Infosys, TCS and Oracle etc. from Yahoo Finance to find the association among the large scale. IT companies and small scale IT companies. Their aim in this research is to find dependencies among different IT companies in the stock market and generate their rules. In this paper they used high values of the shares and applied fragment based mining algorithm & generate some useful rules which influences in

the behaviour of the stock market.There are some weaknesses in the previous FITI(First Intra then Inter) approaches such as time and space involvedin processing the data is more. In FITI approach it is difficult to process an information table with many attributes and long intervals for intertransaction associations. This results into large amount of time and cost in processing the data.

By some experimental analysis they found Fragment Based approach generated more generalized rules as compared to FITI approach. Fragment based mining groups all the attributes once and performs the operation group wise instead of single attribute, which results into more generalized rules. Also time needed to process the data is less as we reduced the size of the input table. The rules generated from fragment based approach can be recommended to the customers who invest their money in the stock market [10].

Kanti and Argiddi [9] have investigated the application of Inter-transaction association rules mining in stock price predication and the possibility of generalizing this method to futures market. They have compared EH-Apriori and FITI algorithms according to their algorithmic structures and itemsets used;they found that FITI is much better than EH-Apriori. FITI generates many extra and meaningless rules and makes the process complex. Thus they have stated another technique called granule -based transactions to have efficient mining process. It uses sliding window setting on decision attribute or constraints and uses SUM Measure. Due to sliding window set only for decision attribute less memory requires. It also scans the data twice same as FITI. It also scans the data twice same as FITI. It is more applicable in real industry in case of large database.

Mahmood A. Rashid et al. [13] has proposed Clinical State Correlation Prediction (CSCP) which extracts data from patients' healthcare database, transforms the OLTP data into a Data Warehouse by generating association rules. The CSCP system helps reveal relations among the diseases. The CSCP system predicts the correlation(s) among primary disease (the disease for which the patient visits the doctor) and secondary disease/s (which is/are other associated disease/s carried by the same patient having the primary disease) carried by a patient using the healthcare repository. In this paper, we have implemented Association Rules mining based a novel idea for finding co-occurrences of diseases. The improved Apriori Algorithm is used for Association Rule Mining to find out frequent dataset that satisfy the predefined minimum support and confidence from a given database. A set of diseases obtained by each patient presented is obtained by Algorithm 1 – CountDisease Pseudo code to count diseases carried by any patient. The number of diseases counted by the Algorithm 2-FindSupport. They have succeeded to investigate correlation among diseases for patients of different age & sex groups providing the outcome in statistical as well as in graphical format.

Goswami D.N. and et al. [11] has proposed three different frequent pattern mining approaches Record filter, Intersection and Proposed Algorithm are given based on classical Apriori algorithm. In these approaches Record filter approach proved better than classical Apriori Algorithm, Intersection approach proved better than Record filter approach and finally proposed algorithm proved that it is much better than other frequent pattern mining algorithm. In Record filter approach they count the support of candidate set only in the transaction record whose length is greater than or equal to the length of candidate set, because candidate set of length k, cannot exist in the transaction record of length k-1, it may exist only in the transaction of length greater than or equal to k. Intersection approach has taken very less time as compared to classical apriori. In modified approach, to calculate the support we count the common transaction that contains in each elements of candidate set, by using the intersect query of SQL. This approach requires very less time as compared to classical apriori. They have considered those transaction that contain at least k items, not less than k in process of support counting for candidate set of k length. This approach requires very less time as compared to all other approaches.

Sheenu Toms et al. [12] have proposed Correlation Analysis provides an alternative framework to support-confidence for finding interesting relationships or to improve understanding of meaning of some association rules i.e. it is a lift of an association rule. The proposed method will effectively mine XML. Support-confidence framework is not so efficient in association rule mining because it identifies a rule (A => B) as strong even if the occurrence of B is so high. Correlation provides extra information about the association rule.

In this paper, Lin Lin, Mei-Ling Shyu [18] proposed a video semantic concept detection framework that uses ARM together with a novel correlation based interestingness measure is obtained from applying Multiple Correspondence Analysis (MCA) to capture correlation between the features and concept classes and to bridge the semantic gap between low-level-features and high-level concepts. This new correlation based interestedness measure is first used in the rule generation stage and then reused and combined with the inter-similarity and intra-similarity values to select the final rule set for classification. Selected rules are used for classification that the concept class is determined by the majority class of the matched rules. To evaluate their proposed framework, the high-level concepts and videos from TRECVID 2007 and 2008 are used and the performance is compared with the well-known decision tree, support vector machine, neural network, k^{th} nearest neighbour. Proposed framework is very accurate and performs all six classifiers.

In this paper Lijuan Zhou and et al.[15] has studied the FP-growth algorithm and proposes a parallel linked list-based FP-growth algorithm based on MapReduce programming model, named as the PLFPG(Parallel Linked List-based FP-Growth) algorithm. This algorithm improves the shortcomings of the traditional FP-growth algorithm. First, it describes item-sets space theory, the basic idea of FP-Growth algorithm and the basic components of the Hadoop platform, including HDFS framework and MapReduce programming model. Then, it describes the PLFPG algorithm design ideas. Finally, the algorithm was validated by varying the size of the data set. The results show that the PLFPG algorithm compared with the traditional FP-growth algorithm has a higher operating efficiency and better scalability and extensibility. It can effectively analysis and deal with large data sets.

According to Jiangping Chen [16] puts a new method of mining spatial association rules based on taking account of

the spatial autocorrelation with a cell structure theory. It defines spatial autocorrelation with an algebra data structure than the autocorrelation of the spatial data can be calculated in algebra. In spatial data mining, we mine rules in the spatial database. They can get the spatial frequency items from the autocorrelation of the spatial data. This replaces the repeated scanning of the database by the measure of the spatial autocorrelation. They proposed algorithm called AR-Miner, shown in below, consist of two faces. In the first phase, we join the algebraic structure table of the graph & their attribute table. In the second phase, mining for the association rules are introduced.

Praveen Pappula and et al. [17] has tried to search the result on dataset of weather forecasting using apriori and GSP(Generalized Sequential Pattern) algorithms. GSP is also looks like theapriori algorithm but one main difference is however the generation of candidate sets. Let us assume that: A → B and A → C are two frequent 2-sequences. The items involved in these sequences are (A, B) and (A, C) respectively. The candidate generation in a usual apriori style would give (A, B, C) as a 3-itemset, but in the present context we get the following 3-sequences as a result of joining the above 2- sequences A → B → C, A → C → B and A → BC.

Ubon [27] has reviewed the literatures on data mining applications, especially applications that applied to solve the crimes. They also tried to show research gaps and challenges of crime data mining. In this paper they also tried to provide show inside into the data mining for finding the patterns and trends in crime to be used. They analysed the crime data from various sources to find patterns and trends in crimes and added the idea to increase efficiency in solving the crimes faster and automatically.

Xi-Zheng et al. [20] focuses on the customer personalization system in the field of E-commerce. In this research analysis of system and design is presented. This system uses association rule mining and classification. Customer requirements are extracted from text documents and transformed into a set of significant phrases. As customer requirements are usually expressed by natural languages where many common words occur which contribute nothing to information retrieval. To extract the requirement information, semantic analysis is adopted. Association rules are mined using Apriori algorithm. They proposed CBA-CB algorithm to produce the best rules out the whole set of rules. After the test best classifiers and validation of the rules are generated and predict the item labels for new customer requirements.

Amir Michai [21] discovered library reuse patterns in existing applications. They introduced the concept of inheritance hierarchy in generalized association rules in library usage in a way that takes into account inheritance relationships. They work on the problem of discovering librarlasses and member functions that are typically reused in combination by application classes. The proposed system explains the concept with the tool, CodeWeb, by demonstrating characteristic ways in which applications reuse classes in the KDE application framework.

Hu Min and et. al. [22] overviewed characteristics of image mining and analyse the overall process and discuss the main technology of image mining. Technologies for image mining like classification and clustering, association rule mining are discussed with respect to image processing. They presented some future research directions and problems of image mining. The aim of this paper is to find valuable hidden information from large amounts of image data.

Lin Lu and et. al. [23] analysed a classical mining algorithm Apriori to produce association rules and found Apriori a very inefficient because it scans the database many times. To overcome this problem they introduced a new improved Apriori Algorithm and proved through their experiments their improved algorithm is very time and space efficient. Thus, this paper has designed and implemented the OOO algorithm for the supermarket's consumer knowledge mining. They proved that OOO algorithm is more stable only needs to scan database once, has no additional candidate items, and does not need pruning.

Peter Wlodarczak and et. al. [22] focused on some of the deep learning techniques that have been used in recent research for multimedia data mining. Large amount of Multimedia data is available on Internet. Multimedia Data Mining extracts relevant data from multimedia files such as audio, video and still images to perform similarity searches, identify associations, entity resolution and for classification. This paper focuses on the subfield of Deep Learning (DL) in Artificial Neural Networks (ANN). Deep learning is a branch of machine learning and has been used among other on Smartphones for face recognition and voice commands. Deep learners are a type of artificial neural networks with multiple data processing layers that learn representations by increasing the level of abstraction from one layer to the next. These methods have improved the state-of-the-art in multimedia mining, in speech recognition, visual object recognition, natural language processing and other areas such as genome mining and predicting the efficacy of drug molecules.

AyhamOmary and et. Al. [3] introduced a weather prediction model, which is the base of other prediction models. Those models depend on collection historical weather data and collect all possible attributes that may impact the current and future amounts of yearly falling rain in Jordan. The information is collected from local and web resources. Data mining techniques and AI algorithms are used for future precipitation forecasting based on historical data. Data mining and statistical methods are used to predict future forecasting and possible climate change.

Monica Gandhi and et. al. [24] focused on knowledge abstraction by using data mining methods for prediction of heart disease. Data mining methods specially Naive Bayes, Neural network, Decision tree algorithm are analysed on medical data sets in this paper. They tried to find out hidden patterns for making decision in healthcare organizations. The aim of this paper is to find out the aspects of use of healthcare data for aid of people by method of machine learning furthermore data mining procedures. The main goal is to produce an automated system for diagnosing heart diseases by taking into account earlier information and data.

III. DISCUSSION AND ANALYSIS

We have observed that ARM is a very popular method for finding frequent patterns and association rules and it can be applied in many different domains of business. The researchers have used variousalgorithms, methods and frameworks in their research work.Performance evaluation

generally based on different parameters like execution time, database scan, memory requirement, quality of frequent patterns, data structure used in algorithms or methods. We have observed that from survey, many researchers have used classical apriori and improvisedapriori but it generally adopts candidate generation theory [4][25]. FP-growth is also a very demanding algorithm which can remove the On the basis of execution time the comparison is done and we found AIS requires larger execution time than SETM and Apriori. Apriori beats AIS and SETM. AprioriTID is slower for larger problem size then Apriori. AprioriHybrid is very slow when the minsup is smaller than apriori. FPgrowth is faster than Apriori and all other algorithms.

Regarding database scan AIS and SETM requires several database scan. Apriori also requires too many database scan and save reading effort. AprioriTID does not use database after first pass and use own method then after. AprioriHybrid uses candidate generation in earlier pass as Apriori and in later stage it uses method used by AprioriTID. FPgrowth only does two times database scan in overall execution period.

deficiency of apriori by using decision tree concept. We have compared some basic algorithms for finding frequent patterns and compared performances based on different parameters.

We have compared the basic algorithms based on performance parametersas shown in Table I.

Table I. Comparison of basic algorithms based on different parameters

Algorithm (Author & Invention Year)	*Execution Time*	*Database Scan*	*Memory Requirement*	*Quality of Frequent Patterns*	*Data structures used*	*Overall Performance*
AIS (AgrawalImielinski Swami) R. Agrawal, T. Imielinski, and A. Swami (1993)	Requires larger execution time than SETM &Apriori	Scan database several times.	Requires large space	Unnecessary itemsets are generated	Simple sequential structure is used.	AIS is the first algorithm for finding association rules. Requires more space and wastes efforts. AIS is better than SETM.
SETM M. Houtsma and A. Swami (1993)	SETM is very much slow then AIS	Same as AIS	Requires large space	Same as AIS	TID of generating transaction is saved with candidate itemsets in a sequential structure.	It generates candidate itemsets and counted on fly like AIS. The aim of SETM is to use SQL operations. It is very slow.
Apriori RakeshAgrawalRamakrishnan (1994)	Apriori beats AIS and SETM	Subsequent too many database scan are required and save reading effort.	Requires large space	Larger Frequent patterns are generated	Candidate Itemsets as stored in hash tree. Nodes or hash tree contains list of itemsets or hash tables.	It is fast, Candidate Generation algorithm and proceeds in a level-wise fashion. Too many database scan. Database should be shorted in lexicographical order Apriori works better in scarcity. It uses breadth first strategy towards the issue of frequent patterns. It is better than AIS and SETM.
AprioriTID RakeshAgrawalRamakrishnan (1994)	Slow for larger problem size then Apriori.	Data base D is not used after first pass. But it use its own method for its task.	Same as Apriori.	Same as Apriori	Each set of candidate itemsets Ck is kept in an array indexed by the IDs of the itemsets in Ck. Each Ckis stored in a sequential structure.	It generates candidate itemsets along with corresponding TID. After the first pass it does not use the database for counting support. It uses apriori's candidate generating function. It is slow for large items then Apriori.
AprioriHybrid RakeshAgrawalRamakrishnan (1994)	It is very slow when the misup is smaller than Apriori	In earlier pass it uses candidate generation algo. and then it only scan Ck(candidate set k)	Same as Apriori.	Frequent patterns are large.	Used concept of Apriori and AprioriTID both.	AprioriHybrid is invented with the concept of combining the ability of both the algorithms. In earlier passes, Apriori is better than AprioriTid and in latter passes AprioriTid is better than Apriori. So the Cost of switching is occurred and it does not give any advantage. It is speedy than Apriori and AprioriTID.

FPgrowth J. Han (2000)	It is faster than Apriori.	Only two pass are required.	Large Memory required to store FP-tree.	Larger itemsets are produced.	Compact data structure FP-tree is used.	It avoids repeated scan of database. It does not generate candidate sets but only tests. It uses depth first logic towards the issue of frequent patterns. It uses compact tree structure for frequent patterns. It uses compact tree structure for frequent patterns. It is better than Apriori where transaction database is huge and minimum support is low.

Concerning to the memory requirement, AIS, SETM, Apriori, AprioriTID and AprioriHybrid requires larger memory space. FPgrowth requires larger memory than Apriori to store FP-tree structure in memory.

Regarding the quality of frequent patterns, AIS and SETM generates itemsets in which unnecessary itemsets are included which wastes time. Apriori is able to generate larger frequent patterns. AprioriTID and AprioryHybrid is same as Apriori. FPgrowth is also able to generate larger itemsets.

Regarding the data-structure, day-by-day technology is improved. In AIS and SETM simple sequential structure is used. Apriori uses hash tree for storing itemsets. AprioriTID uses array indexed and store itemsets by IDs of transaction and also used sequential structure for candidate set Ck. AprioriHybrid used concept of both the algorithmsApriori and AprioriHybrid. FP-growth introduced new FP-tree structure which is able to avoid costly candidate generation.

After comparing different basic algorithms, we foundthat AIS is better than SETM. Apriori requires repeated database scan but it is faster than AIS and SETM. Apriori-TID and AprioriHybrid are improved version of Apriori and does the work for reduced repeated scan but performance is same as apriori and still uses candidate generation method. FP-growth avoids costly candidate generation by introducing FP-tree structure and overall performance is better than all other algorithms.

From the above comparison it is very much clear that the basic algorithms are generally designed toidentify only frequent patterns not to generate association rules in proper format. There are number of different frameworks and new models are proposed according to the problem requirement. The concept of correlation measurement and lift concept with the support and confidence is also applied. Machine learning algorithms like neural network is also applied for association rule mining.

During this survey, we could know the different research trends and issues problems. The main sources for the research in data mining is dataset of any one domain, method or algorithm to apply, do the research and experiments and get the result and make the comparative study based on the performance criteria. Rather than applying the existing algorithm the researcher can improve existing algorithm by removing its drawbacks, the researcher also can propose a framework by combining different processes or combining more than two algorithms.

IV. CONCLUSION

From last few years, large amount of data have been generated by various large companies. We can analyse those stream of data to mine useful frequent patterns or co-related rules according to the business domain. ARM is one of the most popular techniques of the data mining process which is used to find association rules. We have surveyed some papers written by various researchers and analysed them for focussing on methods, algorithms as well as results carried out by them for association rule mining in different application areas like prediction in stock market, analysed the customer behaviour in retail or super store, association needed for XML mining, correlation among diseases for patients in health care, video semantic concept detection framework, the spatial autocorrelation, weather forecasting and etc. We have observed that the main steps required by the research is dataset of any one domain, method or algorithm to apply, do the research and experiments, get the result and analyse them and do the comparative study based on the performance criteria like execution time, memory requirement, database scan, data structure used etc.Researcher can applythe existing algorithm or improve existing algorithm by removing its drawbacks. The researcher also can propose a framework by combining different processes or combining more than two algorithms. One more thing we found during the survey is that when we are talking about association rule mining, everyone discusses about finding the frequent patterns and algorithms and methods related with that but no one has discussed about finding the association rules in proper format except the Agrawal in 1994 while he developed apriori. Our paper's intension is to highlight recent trends and issues and also to provide proper directions to other researchers in the context of selection of the research domain, methodology, and performance evaluation parameters for association rule mining.

V. REFERENCES

[1] RakeshAgrawal, Tomasz Imielinski, and Arun N. Swami, Mining AssociationRules Between Sets of Items in Large Databases, Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207-216, Washington, D.C., May1993.

[2] RakeshAgrawal and Ramakrishnan Srikant, Fast Algorithms for Mining Association Rules in Large Databases, Proceedings of the Twentieth International Conference on Very Large Databases, pp. 487-499, Santiago, Chile, 1994.

[3] AyhamOmary; Ahmad Wedyan; Ahmed Zghoul; Ahmad Banihani; IzzatAlsmadi, "An interactivepredictivesystem for weatherforecasting", Computer, Information and Telecommunication Systems (CITS), 2012 International Conference, IEEE Conference Publications, pp.1 - 4, 2012

[4] Yanxi Liu, "Study on Application of AprioriAlgorithm in DataMining", Second International Conference on Computer Modeling and Simulation,2010, Volume: 3,IEEEConference Publications, pp.111 - 114, DOI: 10.1109/ICCMS.2010.398

[5] Jiawei Han, Jian Pei, Yiwen Yin, "Mining frequent patterns without candidate generation ", Proc. of ACM SIGMOD International Conference on Management of Data, pp. 1- 12, Volume 29 Issue 2, June 2000.

[6] UbonThongsatapornwatana, "A Survey of Data Mining Techniques for Analysing Crime Patterns", Second Asian

Inter Transactions", International Journal of Current Engineering and Technology, Vol.4, No.4 (Aug 2014), E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

[10] Rajesh V. Argiddi, Sulabha S. Apte, " An Evolutionary Fragment Mining Approach To Extract Stock Market Behaviour For Investment Portfolio", (IJCET) ISSN 0976 – 6367(Print),ISSN 0976 – 6375(Online) Volume 4, Issue 5, September – October (2013), pp. 138-146

[11]Goswami D.N., ChaturvediAnshu. ,Raghuvanshi C.S., "An algorithm for Frequent Pattern Mining Based on Apriori.", (IJCSE) International Journal on Computer Science and Engineering, Vol. 02, No. 04, 2010, 942-947, ISSN : 0975-3397

[12] Sheenu Toms, Deepa John," Rules Extraction in XML Using Correlation", International Journal of Scientific and Research Publications, Volume 3, Issue 6, June 2013 1 ISSN 2250-3153

[13] Mahmood A. Rashid, MdTamjidulHoque, Abdul Sattar, "Association Rules Mining Based Clinical Observations ",Griffith University Nathan, QLD, Australia, {m.rashid, t.hoque, a.sattar}@griffith.edu.au, 2010

[14] SanjeevRao, Priyanka Gupta, ―*Implementing Improved Algorithm over APRIORI Data Mining Association Rule Algorithm*‖, ISSN: 0976-8491 (Online) | ISSN: 2229-4333 (Print) IJCST Vol. 3, Issue 1, Jan. - March 2012.

[15] Lijuan Zhou, Xiang Wang, " Research of the FP-Growth Algorithm Based on Cloud Environments", JOURNAL OF SOFTWARE, VOL. 9, NO. 3, MARCH 2014

[16] Jiang ping Chen, "An Algorithm About Association Rule Mining Based On Spatial Auto Correlation", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B6b. Beijing 2008

[17] Praveen Pappula, Ramesh Javvaji, " Experimental Survey on Data Mining Techniques for Association rule mining", International Journal of Advanced Research in Computer Science and Software Engineering 4 (2), February - 2014, pp. 566-571

[18] Lin Lin, Mei-Ling Shyu, "Association rule mining with a correlation-based interestingness measure for video semantic concept detection", Int. J. Information and Decision Sciences, Vol. x, No. x, xxxx, 2009

[19] Rahul Thakkar, "Data Mining Techniques And Stock Market",International Journal of World Research, Vol: I Issue XIII, December 2008, Print ISSN: 2347-937X

Conference on Defence Technology (ACDT), IEEE Conference Publications, pp. 123-128, 2016

[7] Ziauddin, ShahidKammal, KhaiuzZaman Khan, Muhammad Ijaz Khan," Research on Association Rule Mining", (ACMA) 226 Vol. 2, No. 1, 2012, ISSN 2167-6356

[8] Rajesh V. Argiddi, Sulabha S. Apte (2012) "Fragment Based Approach to ForecastAssociation Rules from Indian IT Stock Transaction Data" IJCSIT, Vol 3(2), 3493-3497

[9] Kranti M. Jaybhay,V.ArggiddiÀ, " A Comprehensive Overview of ARM Algorithms in Real Time

[20] Xi-zhengzhang , "Building personalized recommendation system in e-commerce using association rule-based mining and classification", proceedings of the sixth international conference on machine learning and cybernetics, Volume 7, IEEE Conference Publications. pp. 4113 – 4118, Year 2007

[21] Amir Michai, "Data Mining Library Reuse Patterns using Generalized Association Rules", ICSE 2000 Limerick Ireland, Copyright ACM 2000, 1-581 13-206-9/00/6 ... $5.00

[22] Hu Min, Yang Shuangyuan, "Overview of imageminingresearch", 5th International Conference on Computer Science and Education (ICCSE), 2010, IEEE Conference Publications, pp. 1868 - 1870, DOI: 10.1109/ICCSE.2010.5593813

[22] Peter Wlodarczak, Jeffrey Soar; Mustafa Ally, "Multimediadataminingusingdeeplearning", Fifth International Conference on Digital Information Processing and Communications (ICDIPC), 2015, IEEE Conference Publications, pp. 190 - 196, DOI: 10.1109/ICDIPC.2015.7323027, ISBN: 978-1-4673-6832-2©2015

[23] Lin Lu, Pei-qi Liu, "Study on an improved apriorialgorithm and its application in supermarket", 3rd International Conference on Information Sciences and Interaction Sciences (ICIS), 2010,IEEE Conference Publications, pp. 441 - 443, DOI: 10.1109/ICICIS.2010.5534789

[24] Monika Gandhi; Shailendra Narayan Singh, "Predictions in heartdiseaseusingtechniques of datamining", International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), 2015, IEEE Conference Publications, pp. 520 - 525, DOI: 10.1109/ABLAZE.2015.7154917

[25]Chirag A. Mewada, Rustom D. Morena, "Model using Improved Apriori Algorithm to generate Association Rules for Future Contracts of Multi Commodity Exchange (MCX)", International Journal of Advanced Research in Computer Science, Volume 8, No. 3, ISSN No. 0976-5697, March – April 2017

[26] N.P. Gopalan, B. Sivaselvan, "Data Mining Techniques and Trends", PHP Learning Private Limited, New Delhi-110001, ISBN-978-81-203-3812-B, 2009

[27] G.K. Gupta, "Introduction To Data Mining With Case Studies", Second Edition, PHP Learning Private Limited, New Delhi-110001, ISBN-978-81-203-4326-9, 2011

