Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 12, December 2015, pg.273 – 281

A Review on Modern Approach: New Parameter for Recent Improvement of Apriori Algorithm

VIJAITA KASHYAP

COMPUTER SCIENCE, L.N.C.T, BHOPAL, INDIA

Vijaita.kashyap@gmail.com

ATUL KUMAR GUPTA

COMPUTER SCIENCE, L.N.C.T, BHOPAL, INDIA

Atulkumar.2k9@gmail.com

Abstract — In recent years, the rapid growth and large volume of data increasingly requires of Internet, Web search has been taken an important role in our ordinary life based on Association rule mining is an important data analysis method to discover associated web pages. Lots of algorithms for mining association rules and their mutations are proposed on starting point of Apriori algorithm, however conventional algorithms are not proficient to remove information from a database and deal with different application domains for estimating a future value.

Index Terms— Data Mining, Association Rule Mining, Apriori Algorithm, styling.

I. INTRODUCTION

Data Mining is the process of discovering new patterns from large data sets involving techniques at the communication of Artificial Intelligence (AI), Machine Learning, Statistic and Database Systems. The over goal of data mining is to extract knowledge from large source of database. The actual data mining task is the automatic or partially-usual examination of huge amount of data to removed up to that time indefinite remarkable outlines such as collections of data records i.e. cluster analysis, remarkable records i.e. anomaly detection and dependencies i.e. association rule mining.

Data Mining (DM) is an essential area in the field of data and knowledge based schemes, which have been prompted by an interesting new field called Knowledge Discovery in Databases (KDD).

The relative and iterative Knowledge Discovery in Databases (KDD) steps that were described in [1] are shown in:

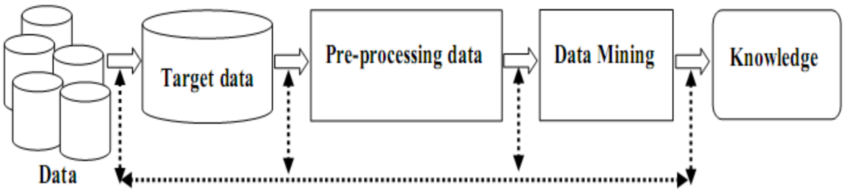

Figure 2.1 Generic steps of the KDD. [1]

Discovered knowledge can approach in various forms for example: Association Rules, Correlations, Sequences, Episodes, Classifiers, Clusters and many more. Many people treat data mining as a synonym for another popularly used term, "Knowledge Discovery in Databases", or KDD. On the other hand, previous examination data mining as simply an essential step in the process of knowledge detection in databases. Knowledge discovery as a development and consists of an iterative series of the following measures:

- data cleaning i.e. to eliminate noise or inappropriate data,
- data integration i.e. where various data sources may be unite,
- data selection i.e. where data appropriate to the study job are get back from the database,
- data transformation i.e. where data are changed or combined into structures suitable for mining by presenting outline or aggregation process,
- data mining i.e. an fundamental procedure where intellectual techniques are useful to facilitate removed data patterns,
- pattern evaluation i.e. to recognize the accurately attractive patterns demonstrating knowledge based on some attractiveness evaluates; and
- Knowledge presentation where visualization and knowledge demonstration techniques are exploited to here the mined knowledge to the consumer.

To best apply these advanced techniques, they must be fully integrated with a data warehouse as well as flexible interactive business analysis tools.

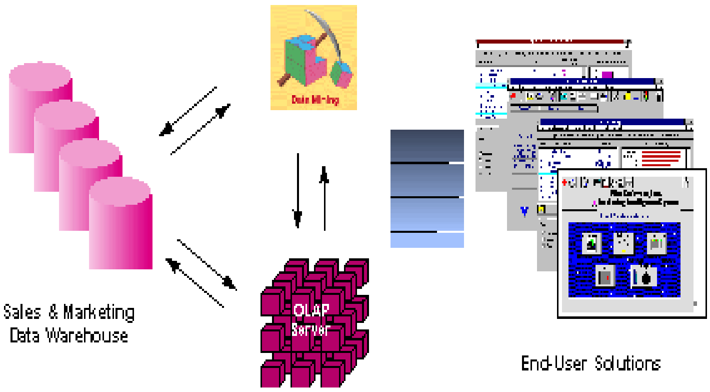

Figure 1 illustrates architecture for advanced analysis in a large data warehouse in Data Mining Architecture.

Many data mining tools currently operate outside of the warehouse; necessitate further actions for removing, introducing and analyzing the data. Furthermore, when new insights require operational performance incorporation with the warehouse make simpler the application of results from data mining. The resulting analytic data warehouse can be applied to improve business processes throughout the organization, in areas such as advertising operation supervision, fraud detection, novel manufactured goods level, and so on.

Association rule mining is one of the most important tasks in DM, which plays a vital role in business applications such as, market basket for manufacturing well-built laws from common things in a operational database [2]. Aprioi algorithm which used for finding frequent item set and uses this item to generate association rules. The benefits of these rules are detecting unknown relationships, producing results which can be used for decision making and prediction. There are numerous of researches and projects that exploit Apriori algorithm and how to improve its efficiency.

.

II. THEORETICAL BACKGROUND

In today's multimedia environment enormous quantities of knowledge is accumulated contained by many datasets and databases regularly the defaulting arrangement of this data signifies that the information within is not instantly available, but moderately has to be mined and removed. This needs programmed implements and they require being efficient and proficient. The responsibility of data mining is easy and has been explained as "removing knowledge from huge quantities of data" [3]. Association rule mining is one of the governing data mining knowledge. It is that move toward to acquiring knowledge stored with existing datasets a database which comprises common patterns and association rules between the articles / characteristics of a dataset with show discrepancy levels of strong point. On the other hand, this is also association rule mining's problem; the amount of regulations that can be originated is more often than not very large. With the intention of efficiently exploit the association rules and the information inside the amount of rules requires being reserved convenient, thus it is essential to have a technique to decrease the number of association rules. On the other hand, they do not desire to be unable to find information through this procedure. For most of the effort completed in extending association rule mining, the most important meeting point has been on the effectiveness of the approach i.e. how rapidly can it originate the rules and to a smaller amount the value and the estimation procedure / determines to find out the feature of the developed regulations has been highlighted. Frequently for a dataset, a vast amount of regulations can be obtained, but many of them can be unnecessary to other rules and thus are ineffective in perform. The enormously huge amount of regulations makes it easier said than done for the end customers to understand and consequently successfully to apply the discovered rules and thus considerably decreases the efficiency of rule mining algorithms. If the mined knowledge can't be efficiently exploited in explain authentic world difficulties the attempt of removing the information is value small. This is a serious difficulty but not until now explained adequately.

III. ASSOCIATION RULE MINING AND ITS TECHNIQUES

In this part association rule mining will be commenced and then enclosed in strength and feature counting an evaluation of multi-stage and cross-stage association rule mining. Many unusual move towards and algorithms will then be seemed at and evaluation. Association rule mining was primary accessible by Agrawal in [4]. Association rules are usually in the form X => Y, where both X and Y are items or itemsets that are completely contained within a dataset or database and $X \cap Y = \emptyset$. X is the antecedent and Y is the consequent. This regulation entails that on every occasion X is here, Y will also be present or that X implies Y.

For the reason that most datasets are huge and the user does not always feel like all the rules but only those that are of concentration or significance determines are required so that the uninteresting rules can be eliminated. Conventionally, there are two essential evaluates used in association rule mining support and confidence. Support is to compute that describes the percentage or fraction of records or entries in the dataset that contain X U Y to the entirety amount of records/entries. The essential technique for influential the support does not take into explanation the amount of the item with a record and the support value provides as an arithmetical consequence of the rule [5]. Confidence is to compute that describes the percentage or fraction of records or entries in the dataset that contain X U Y to the entire number of records/entries that holds just X. The confidence value provides as to determine of the strong point or precision of the rule [5]. Several different approaches to association rule mining will be accessible, preliminary with the conventional approaches, followed by multi-level and cross-level approaches. After an evaluation of the unusual approaches other concerns associated to association rule mining will be offered.

A. Single Level Association Rule Mining

In this section the evaluation will initiate numerous of the additional conventional association rule mining algorithms. These approaches are intra-transactional algorithms more often than not for a particular abstract level and 1-D datasets.

AIS Algorithm: The AIS algorithm is the former approach that was suggested for association rule mining. It was firstly introduced by Agrawal, Imielinski & Swami [4] and here it paying attention on improving database feature and initiated functionality required to procedure decision support queries and the discovering of relations/associations inside the data. The effectiveness of the AIS algorithm was developed by the calculation of an evaluation scheme to prune candidate itemsets that had no possibility of being on top of the support threshold [5].

Apriori Algorithm: The AIS algorithm was a simple approach but had numerous disadvantages that required enhancement ahead. The Apriori algorithm was that progress and is measured a most important step in association rule mining [5].

FP-Tree Algorithm: Because the Apriori approach has two major drawbacks, work to fix these drawbacks has been conducted. One of the results from this work is the designing of tree structures for use in association rule mining. Frequent Pattern Tree (known as FP-Tree) was first introduce by Han & Pei and is an approach that requires only two passes/scans through a database/dataset to generate the frequent itemsets and does so without the need to generate candidate itemsets [5]

RARM (Rapid Association Rule Mining) Algorithm: Rapid Association Rule Mining (RARM) is an approach that also uses a tree structure to represent the database/dataset and does not utilise a candidate generation process. It was first introduced by Das, Ng & Woon [6] with the focus of being faster than the existing algorithms. The approach taken in developing RARM was to build 1-itemsets and 2-itemsets lists quickly and without needing a candidate generation process to get the frequent 2-itemsets.

Non-Derivable Itemset Algorithm: Another approach to association rule mining has been proposed in the form of non-derivable itemsets and rules. In this approach, itemsets are removed if their support can be derived [7], since the derivability is monotone. The Non-Derivable Itemset (NDI) approach is based on the Apriori approach, however it does not try to find all of the frequent itemsets. It instead focuses on obtaining a complete set of deduction type rules in order to derive what is called 'tight' times on the support value of a candidate itemset.

Closed Itemset Algorithm: The conception of closed frequent itemsets has its starting points in the mathematical theory of Formal Concept Analysis initiated in the early 1980s'. An itemset is said to be closed if and only if no proper superset of this itemset has the same support that this itemset has. For a given support threshold, knowing all frequent closed itemsets is sufficient to produce all the frequent itemsets and they sustains without right to using to the dataset. The use of frequent closed itemsets presents a understandable assure to decrease the amount of removed rules and also provide a concise representation of association rules [8].

B. Multi-Level & Cross-Level Association Rule Mining

Traditionally, association rule mining has been performed at a particular conception or abstract stage which was frequently either a low abstract/primitive level or at a high abstract/concept level. It is widely accepted that single level rule mining has two major problems; firstly it is difficult to find well-built associations at a small or primitive stage suitable to the sparseness of data and secondly, mining at high levels may result in common knowledge rules being presented which are already known and are of little use or interest [3]. It is quite possible that a given database, which can be mined by a single level algorithm, is not in fact flat, but contains data in a hierarchical format. While this structure may be present, it has been argued that few algorithms use or take advantage of this type of structure. Therefore alternatives were investigated and both multi-level rule mining and cross-level rule mining came about.

One of the major arguments for the use of multi-level or cross-level rule mining is that it has the potential for undiscovered knowledge to be discovered. Such knowledge may possibly not be established by the single level move toward and this novel information may be extremely appropriate or remarkable to a given user. Multi-level rules span multiple levels of concept but the items inside one law come from the same concept or abstract level. That means they can be at dissimilar stages and have more universal or more precise information than single level rules and the intermediate results from high levels of abstraction can be used to help mine lower abstract levels and refine the process [3]. However the use of results from one concept/abstract level to draw conclusions at another level needs to be done carefully as there are problems with this approach.

One of the advantages of the association rule discovery is extracting explicit rules that are practically important for the user/human expert to understand the application domain. Thus, this can be making possible to alter or extend the rules manually with additional domain knowledge which is complicated to get with other mining techniques [9].

IV. SEQUENTIAL SEQUENCE MINING TECHNIQUES

Sequential sequence [10] is defined as: The data set is a set of sequences, named as data-sequences. Each one data-sequence is a set of operations. Each operation is a set of truthful called items or events. Typically there is a transaction time connected with each operation. The sequential series mining discovers all sequential sequences with a client distinct minimum support.

Apriori-based Techniques: The first and simplest family of sequential sequence mining algorithms is Apriori-based algorithms and their most important feature is that they use Apriori principle. The problem of sequential sequence mining was initiated together with additional three Apriori-based algorithms i.e. AprioriAll, AprioriSome and DynamicSome [10]. At each step k, a set of candidate frequent sequences C_k of size k is generated by performing a self-join on L_{k-1}; L_k consists of all those sequences in C_k that satisfy a minimum support threshold. The efficiency of support counting was progressed by using a hash-tree arrangement.

A related move toward, GSP (Generalized Sequential Patterns) was developed that exploits time restrictions in addition to the window restrictions. This was proved to be more efficient than its predecessors. Determining all frequent sequential series in huge databases was an extremely difficult job since the search space was large. For the database with m attributes and length of k frequent sequence, there are O(mk) potentially frequent ones. Increasing the number of objects might lead to a high computational cost. Apriori-based algorithms utilize a bottom-up search lists every single frequent sequence. To produce a frequent sequence of length *l*, all 2*l* subsequences have to be generated. It can be easily worked out that this exponential complexity is restricting all the Apriori-based algorithms to discover only short sequences, since they only implement subset infrequency pruning by removing any candidate sequence for which there be presents a subsequence that does not be in the right place to the set of frequent sequences.

Tree-based Techniques: A faster and more proficient candidate creation can be accomplished by means of a tree-like formation. The traversal is made in a depth-first exploration method. It is functional such that all the candidate series is appropriating both compartment infrequency and superset occurrence pruning. To begin with the above scheme was initiated for mining frequent itemsets, but then it was extended for sequential orders. Ayres engaged a well-organized move toward in SPAM [11].

SPAM generated sequence enumeration tree to generate all the candidate frequent sequences. The stage k of the tree haves the entire set of sequences of size k (with each node representing one sequence) that occurs in the database. The nodes of each level are generated from the nodes of the earlier stage using two kinds of expansions:

(1) Itemset expansion i.e. the last itemset in the series is comprehensive by adding together one more item to the set of itemset,

(2) Sequence expansion i.e. a sequence is extensive by adding mutually a novel itemset at the closing stages of the sequence.

The candidate sequences are specified by traversing the tree using depth-first search. If the sequence is found infrequent, the sub tree of the node representing that sequence is pruned. If the sequence is found to be frequent, then all its subsequences have to be frequent, thus the tree nodes representing those sequences are skipped. For efficient support counting, the database is represented by a bitmap, which further improves performance over the lattice-based approaches [12] discussed in next method.

Lattice-based Techniques: Lattice structure was an additional set of sequential sequence mining algorithms was suggested a lattice based technique to get detail the candidate sequences proficiently. In actual fact, a lattice give the impressions to be a “tree-like” structure where each node may have additional than one parent node. A node on the lattice characterizes a sequence s, is associated to all the two of a kinds of nodes on the earlier stage that can be attached to form s. This is exposed in the example: let s = {d, (bc), a}, then all the following nodes should be associated to s on the lattice: {(bc), a}, {d, b, a}, {d, (bc)}, {d, c, a}, since all pairs of these subsequences can be connected to structure s. SPADE [12] used greater than arrangement to proficiently identify the candidate sequences. The essential features of SPADE were:

(1) Vertical illustration of the database using id-records, where each series is connected with a record of database sequences in which it happens.

(2) To employ lattice-based approach to crumble the unique investigate space into smaller subspaces.

(3) Each sub-lattice two dissimilar investigates strategies i.e. breadth-first and depth-first search were used for receiving frequent sequences.

cSPADE was the expansion of SPADE was suggested in [12], which permits a set of restrictions to be positioned on the extracted sequences. These limitations are:

(1) Length and width restrictions
(2) Gap and window restrictions
(3) Item restrictions
(4) Class restrictions

Regular Expression based Techniques: enormous popular of the earlier algorithms paying attention the detection of frequent sequential sequences based on simply a sustain entrance which edges the consequences to the most frequent. Consequently be deficient in consumer prohibited center of attention in the sequence mining development can be identified that may a moment or two show the way to huge volume of useless sequences. A solution to this problem was proposed in [13], where the mining process was restricted by a support threshold and user-specified constraints modeled by regular expressions. Later on the series of SPIRIT [13] algorithms were introduced, where a set of constraints C was pressed into the mining procedure together with a sequence database. Consequently, the minimum support constraint and a set of extra consumer particular restrictions were useful concurrently which confine the set of candidate sequences created during the mining procedure. To accomplish these two unusual types [13] of pruning methods were exploited.

First was based on constraint and second was based on support value. The first technique used a relaxation C_0 of C ensuring that during each pass of the candidate generation, all the candidate sequences satisfy C_0. The subsequent method, make an effort to ensure that all the subsequences of a candidate series satisfy C_0 are present in the current set of discovered frequent series.

An additional feature of the SPIRIT [13] algorithms were related to anti-monotonicity. Consider a given set of candidates C and a relaxation C_0 of C. In fact C_0 was a weaker constraint which was less restrictive. In such case, support-based pruning was maximized, since support information for every subsequence of a candidate sequence in C_0 could be used for pruning. In addition, if C_0 was not anti-monotone, the efficiency of both support-based and constraint-based pruning depends on the relaxation C_0.

Prefix-based Techniques: Other techniques of sequential sequence mining algorithms include the prefix-based [14]. In this method, the database is projected with respect to a frequent prefix series and supported on the result of the projection novel frequent prefixes are recognized and employed for an additional projections in anticipation of the support entry limitation is contented.

The main steps of a prefix-based algorithm are following:
(1) Scanning of the database for the frequent 1-sequences.
(2) Project the database with respect to s for each frequent 1-sequences found in the previous step.
(3) Scan the projected database for local frequent items.
(4) Add each new frequent item to the end of the prefix and project the database with respect to the new prefix.
(5) Repeat steps 3-4 for each new prefix, until the projected database is of size less than the support threshold.

Closed Sequential Sequences Techniques: In addition to mine the complete set of frequent sequences including their subsequences, the closed frequent sequence techniques were proposed. They use special techniques to limit the number of frequent sequences and finally keep only the closed ones.

CloSpan [15] used the candidate maintenance-and-test approach, i.e. it first produces a set of secured sequence candidates which is accumulated in a hash-indexed tree structure and then prunes the search space using Common Prefix and Backward Sub sequence pruning. However the drawback of CloSpan is that it consumes much remembrance when there are several closed frequent progressions in view of the fact that sequence finish confirming show the ways to a enormous search space. Therefore, it does not scale well with respect to the number of closed sequences. To overcome this limitation, BIDE employed a BIDirectional Extension paradigm for mining closed sequences, where a forward directional extension is used to grow the prefix sequences which checks their closure and a backward directional extension. It is used to check the closure of a prefix sequence and prune the search space. Overall, It is seen that BIDE has high efficiency, regarding speed (an order of magnitude faster than CloSpan [15]) and scalability with respect to database size.

Time interval Sequence Mining Techniques: Up to this point, the events were considered to be instantaneous. There were several techniques on discovering intervals that occurred frequently in a transactional database [16]. In most cases, the intervals were not labelled and no relations were between them considered. In time interval sequential mining, the time between events is considered.

V. LITERATURE SURVEY

Zoghby [17] this work introduces a new system developed to discover soft-matching association rules using a similarity dimensions based on the origin characteristic of the Arabic language. As well, here author try to shows the new characteristics by means of Frequent Closed Item-sets (FCI) idea in mining the association rules to a certain extent than Frequent Itemsets (FI).

Najadat and Maolegi et al [18] indicate the limitations of the original Apriori algorithm of wasting time for examining the complete database investigating on the frequent itemsets, and shows an enhancement on Apriori by exhausted time depending on examining only some operations. They showed by experimental results that applied on the original Apriori and the improved Apriori that the new Apriori reduces the time consumed by 67.38% in comparison with the original Apriori and constructs the Apriori algorithm more proficient and less time consuming. The results show that the improved Apriori algorithm that scan only some transactions instead of the whole database reduce the consumed time.

Rao and Gupta [19] present a new scheme for finding the rules out of transactional datasets which improve the original Apriori in terms number of database scans, memory consumption, and the interestingness of the rules. It also avoids scanning the database again and again. So, they use Frequent Pattern (FP) Growth ARM (Association rule mining) algorithm that is more efficient to mine patterns when database grows.

In this paper [20], author has afforded out two types of developed algorithms i.e. N Painting-Growth algorithm and Painting-Growth algorithm. N Painting-Growth algorithm constructs two-item permutation sets to discover association sets of all frequent items and then mines up all the frequent item sets according to the association sets.

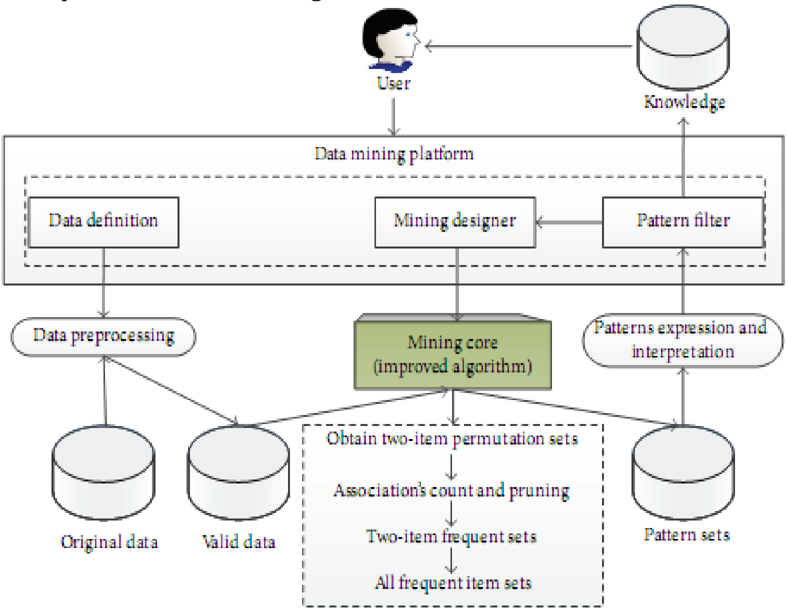

Figure 2: Association rules mining system model [22]

Painting-Growth algorithm constructs an association image based on the two-item permutation sets to discover association sets of all frequent items and then mines up all the frequent item sets according to the association places. From the above procedures it can be distinguish that the N Painting-Growth algorithm is the eliminating of painting steps description of Painting-Growth. Both of the developed algorithms examining the database only on one occasion improving the transparency of scanning database two times in conventional FP-Growth algorithm and implementation the mining only in proportion to two-item permutation sets, consequently have the benefits of organization earlier taking up undersized space in memory, having low complication and individual effortless to sustain. It is understandable that get better algorithms make available a position for next association rules mining investigate.

In this paper author [21], proposed a new scheme based on transactional matrix is offered to discover the frequent itemsets from a huge transactional database. In this scheme a transactional matrix is produced straightforwardly from the database and then frequent itemsets and also they support of each one frequent itemset is produced straightforwardly from the transactional matrix. Here it is originate that the novel recommend move toward to discover the frequent itemsets more proficiently. The presentation of novel technique is evaluated with that of Apriori algorithm.

Authors	**Technique**	**Benefit**
Suhani Nagpal [22].	-Temporary Tables for scanning. -Logarithmic Decoding	-Low system overhead and good working presentation -Effectiveness advanced than Apriori Algorithm.
Jaishree Singh, Hari Ram [23].	-Variable Size Of Transaction on the basis of which Transactions are reduced.	-Reduces the I/O cost. -Reduce the size of Candidate Item sets (Ck)

Jaio Yabing [24].	-Double Pruning method is used. -States that before C_k come out, prune L_{k-1}	-For large datasets, it saves time and cost and increases the efficiency
Sunil Kumar [25].	-Probability Matrix has been used. -Uses Bottom Up approach.	-Reduced Execution time than Apriori Algorithm

In this review we have focused on content, collaborative and hybrid recommender systems, along with a look at recommender systems based on association rules. Several different approaches to recommender/recommendation system will be presented and reviewed, along with some of the issues/problems that these systems suffer from will also be presented and discussed.

Recommendation Approach	**Recommendation Techniques**	
	Heuristic-based	**Model-based**
Content-based	TF-IDF (Information Retrieval) Clustering (Nearest Neighbour eg. kNN)	Linear & Bayesian Classifiers Clustering Decision Trees Artificial Neural Networks Relevance Feedback (eg. Rocchio's Algorithm)
Collaborative	Nearest Neighbour (Cosine) Clustering Graph Theory	Bayesian Classifiers Clustering Artificial Neural Networks Linear Regression Probabilistic Models
Hybrid	Linear combination of predicted ratings Voting Schemes Incorporating one component as a part of the heuristic for the other	Incorporating one component as a part of the model for the other Build a single unifying model

VI. CONCLUSION

Various Classical Algorithms have been examine and review the literature in the area of Data Mining techniques to discover the modern methods i.e. Fuzzy Logic, association rule mining, associative classification, feature selection technique and application areas. To demonstrate the advantages of feature selection techniques for dropping elevated dimensional data and extend an efficient characteristic technique that can considerably be replicated on the prediction power. To demonstrate the competence and usefulness of the prediction model by appropriating it to unusual case studies and standard data sets of different application domains to guarantee generalization of the data mining representation.

REFERENCES

[1] Mitra, S., Pal, S. K. and Mitra, P. Data mining in soft computing framework: A survey. Neural Networks, IEEE Transactions on, 13 (1), 3-14, 2002.

[2] Hahsler, M., Grun, B. and Hornik, K. Rules–A computational environment for mining association rules and frequent item sets. Journal of Statistical Software, 14 (15), 1-25, 2005.

[3] Han, J., & Kamber, M. (2001). Mining Association Rules in Large Databases. In D. D. Cerra (Ed.), *Data Mining: Concepts and Techniques* (pp. 225-277). San Francisco,

[4] Agrawal, R., Imielinski, T. and Swami, A. Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data. Washington, D.C., United States, pp. 207-216, 1993.

[5] Q. Zhao & S. S. Bhowmick, Association Rule Mining: A Survey, 2003.

[6] A Das, WK Ng, YK Woon, "Rapid association rule mining" Proceedings of the tenth international conference on, 2001

[7] T Calders, B Goethals, "Non-derivable itemset mining" - Data Mining and Knowledge Discovery, 2007

[8] Pasquier, N., Taouil, R., Bastide, Y., Stumme, G., & Lakhal, L. Generating a Condensed Representation for Association Rules. *Journal of Intelligent Information Systems, 24*(1), 29-60, 2005.

[9] Gedikli, F. and Jannach, D. (2010) Neighborhood-restricted mining and weighted application of association rules for recommenders. In: the 11th International Conference on Web Information Systems Engineering (WISE 2010). Hong Kong, China: Springer Berlin / Heidelberg, Vol. 6488, pp. 157-165.

[10] R. Agrawal and R. Srikant, "Mining Sequential Patterns", Proc. 1995 Int‟l Conf. Data Eng. (ICDE "95), Pages 3-14, Mar. 1995.

[11] J. Ayres, J. Gehrke, T. You and J. Flannick, "Sequential Pattern Mining Using a Bitmap Representation", Proc. ACM SIGKDD Int‟l Conf. Knowledge Discovery and Data Mining (SIGKDD "02), Pages 429-435, July 2002.

[12] M. Zaki, "SPADE: An Efficient Algorithm for Mining Frequent Sequences", Machine Learning, Vol. 40, Pages 31-60, 2001.

[13] Garofalakis, M., Rastogi, R., and Shim, K.,"Spirit: Sequential pattern mining with regular expression constraints", In Proc. of Int‟l Conf. on Very Large Databases (VLDB), Pages 223–234,1999.

[14] M.C., "Prefixspan: Mining sequential patterns efficiently by prefixprojected pattern growth", In Proc. Of IEEE Int‟l Conf. on Data Engineering (ICDE), Pages 215–224, 2001.

[15] X. Yan, J. Han, and R. Afshar, "CloSpan: Mining closed sequential patterns in large datasets", Proceedings of the Int. Conf. SIAM Data Mining, 2003.

[16] Lin, J. L.,"Mining maximal frequent intervals. Technical report", In Proc. of Annual ACM, Symposium on Applied Computing (SAC), Pages 624–629, 2002.

[17] A.-Z. A, "Mining Arabic Text using Soft-Matching Association Rules," Computer Engineering & Systems, ICCES'07. International Conference, pp. 421-426, 2007.

[18] B. A. Mohammed Al-Maolegi, "An Imporoved Apriori Algorithm for Association Rules," International Research Journal of Computer Science and Application, vol. 1, pp. 1-8, 2013.

[19] P. G. Sanjeev Rao, "Implementing Improved Algorithm Over Apriori Data Mining Association Rules Algorithm," International Journal of Computer Science and Technology, vol. 3, pp. 489-493, 2012.

[20] Yi Zeng, Shiqun Yin, Jiangyue Liu, andMiao Zhang, "Research of Improved FP-Growth Algorithm in Association Rules Mining" Volume 2015.

[21] Harpreet Singh and Renu Dhir, "A New Efficient Matrix Based Frequent Itemset Mining Algorithm with Tags" International Journal of Future Computer and Communication, Vol. 2, No. 4, August 2013.

[22] Suhani Nagpal, Improved Apriori Algorithm using logarithmic decoding and pruning, International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622, Vol. 2, Issue 3, May-Jun 2012, pp.2569-2572.

[23] Jaishree Singh, Hari Ram, Dr. J.S. Sodhi, "Improving Efficiency of Apriori Algorithm using Transaction Reduction" International Journal of Scientific and Research Publications, Volume 3, Issue 1, January 2013 ISSN 2250-3153

[24] Jiao Yabing, "Research of an Improved Apriori Algorithm in Data Mining Association Rules", International Journal of Computer and Communication Engineering, Vol. 2, No. 1, January 2013.

[25] Sunil Kumar , Shyam Karanth , Akshay K , Ananth Prabhu,Bharathraj Kumar M, Improved Apriori Algorithm Based on bottom upapproach using Probability and Matrix, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012,ISSN (Online): 1694-0814.

