Chapter 2

Review of Literature

This chapter incorporates the review of relevant literature; the scholar has produced since 1980. There is few research papers belonging to trie data structure in the year 1959-1960 were also reviewed. The period of this research work is purposely selected because the present research work can justified its finding only when sufficient time has been elapsed after effective implement of concept hierarchy for generating the association rules at multiple levels.

The succeeding review of the research papers published by the eminent researchers was helpful in crystallizing the research plan and it provides support for designing the methodology of the present study. This review work explores the groundwork of fundamental concepts about the association rule mining and a survey on the existing association rule mining techniques.

This chapter has been divided into four sections. First section includes some basic concepts related to data mining, its techniques and specifically association rules mining which are helpful for carrying out present research work. In second section single-level association rules mining approaches are validated, third section explore the literature related to multiple-level association rules mining algorithms and last section includes miscellaneous research paper which helps the researcher to carrying out the research work directly or indirectly on “design of an improved multiple level association rule algorithm for discovery of frequent patterns”. These sections are elaborated in the upcoming paragraphs.

This chapter is organized as follows: Section 2.1 explores review on basic data mining concepts, Section 2.2 discussed literature on single-level association rules mining algorithms, Section 2.3 examines review on multiple-level association rules mining algorithms, Section 2.4 presented the miscellaneous literature reviewed which may relate to present research work directly or indirectly and Section 2.5 provides the outcome of chapter.

2.1 Section-I

First section includes the research work carried out by many researchers proposed various encoding methods with their pros and cons. A brief study is given as below:

To facilitate finding of multiple level association rules, it is indispensable to offer data at multiple levels of abstraction. The data at multiple levels of abstraction can be provided with the help of concept hierarchy of items. These hierarchies describe the relationships among
the items and the grouping of data at numerous levels of abstraction. Author has classified these hierarchies into four major categories. These are Schema hierarchy, set-grouping hierarchy, operation-derived hierarchy and rule based hierarchy. These different types of hierarchies are used for different type of applications.

To support multiple-level association rules a positional encoding scheme of concept hierarchy has been presented by researchers (Han & Fu, 1995). In positional encoding method, the nodes of any concept hierarchy have been represented by string of numbers as 1, 2, 3..., n. This sequence proceeds from left to right with respect to their arrangement in the hierarchy. The encoding of hierarchy has been performed at the time of significant data collection for mining task. With the help of this encoding, item can be represented by fewer bits as compare to their equivalent object identifier. In encoded hierarchy, actual items have been placed on leaf nodes of the tree. The internal nodes characterize classes or notions created from lower level. Positional encoding scheme is appropriate to illustrate the relationship along with items at various levels concept hierarchy. It is also helpful to mine multiple level association rules.

It has been learned that post-order, binary, positional are the traversal encoding method of the concept hierarchy. Post-order traversal encoding has been proposed by Wang and Iyer (Wang & Iyer, 1997) which is performed as finding the post-order traversal of tree. According to researcher, labeling of nodes is started from left most descendent. For example, a node has label y and its descendent with smallest value of label has label x. Then y has exactly (y-x) descendents and all descendents have labels value less than y. Therefore all descendents of y have labels as integer values from x to (y-1). Post-order traversal encoding scheme is useful for the drill-down operation in OLAP (Chamberlin, 1996).

In addition to previous stated traversal methods, researcher (Lu, 1997) had explored a new general purpose encoding scheme. This scheme is appropriate for any functionalities of data mining. In this encoding, each node of concept hierarchy has been assigned by a unique binary code which consists of n numerals where n represents the node level in hierarchy. Further Lu (1997) also supported the research work carried out by the earlier stated researchers that the encoded concept hierarchies perform superior than without-encoded hierarchies. The encoded concept hierarchies take less storage and disk access time as compare to without-encoded hierarchies. The concept hierarchy encoding algorithm proposed by author is practical and proficient particularly for concept generalization.
2.2 Section-II

The second section includes the research work carried on various single-level association rules mining algorithms with their problems and findings. A concise study is given as below:

Mining of useful association rules from huge amount of data has been one of the most significant areas of research in the field of data mining. It was introduced in year of 1993. The purpose was to extract interesting patterns, fascinating correlations, casual structures or associations between sets of items in the transactional databases, warehouses and other data repositories. The mined rules are helpful in a variety of domains such as networks, telecommunication, retail store settings and layout, market analysis, business forecasting and risk management etc. There are several association rule-mining algorithms proposed in the research literature. This section provides a review of a representative set of the major single level association rules mining algorithms.

The first association rule mining algorithm was proposed by researchers in 1993 known as AIS algorithm (Agrawal, et al., 1993). This algorithm repeatedly scans the large database to discover large itemsets. These large itemsets are useful for discovery of association rules in form of \( X \rightarrow Y \) and \( Y \rightarrow X \), if these rules fulfill the condition of minimum confidence. A drawback of this algorithm was that there were too many passes and too many candidate itemset generations.

Two fast algorithms Apriori and Apriori-TID were explored by the authors (Agrawal & Srikant, 1994). Apriori algorithm is based on a special property i.e. ‘Apriori property’. According to this property, if any k-itemset is frequent then it’s all subsets (k-1) must be frequent. Further, if any subset of a candidate set is infrequent then remove the candidate itemset. Consequently this property is used to prune the candidate items. Although pruning diminishes the candidate items, but several scans of database are required. Here, the number of scans of the transaction databases is equal to the maximum number of items in a candidate itemset. In the algorithm Apriori-TID, the database is scanned and a hopefully smaller database Ti is created. This algorithm is faster than the Apriori algorithm since Ti is usually smaller than the original database. It is proved that Apriori works well in earlier stages as compare to Apriori-TID algorithm. On the other hand, Apriori-TID algorithm becomes more proficient than Apriori in the later phases.

Subsequently, in the year 1995, a new algorithm SETM (Houtsma & Swami, 1995) was proposed, which uses SQL (Structured Query Language). This algorithm uses join operation to generate the candidate itemset. The researcher demonstrates that by the use of general
query languages like SQL, a number of data mining aspects can be carried out. Here the disadvantage was of unnecessary generation of too many candidate itemset.

The investigators introduced a Direct Hashing and Pruning (DHP) algorithm (Park, et al., 1995). This algorithm employs a hash-based method to reduce the number of candidate itemsets produce in the first phase (i.e., candidate-2 itemsets). To find the occurrence of all candidate-1 itemsets, this algorithm takes only one scan of complete database. Parallely all possible 2-itemsets in the current transaction are mapped to the hash table. This mapped hash table is used in the subsequent pass to reduce the number of candidate itemsets. This algorithm performs better if large-2 itemsets are comparatively less than the candidate-2 itemsets generated by Apriori. The transactions without any frequent k-itemsets cannot have any frequent (k+1)-itemsets. So, these transactions can be eliminated from the database. With this efficient pruning the algorithm works much faster than Apriori.

Partition algorithm (Savasere, et al., 1995) was also introduced. This was investigated for mining large volume of data. The algorithm logically divides the database into a number of independent partitions. Local frequent itemsets are calculated for each partition and they are merged to find the global frequent itemsets. This algorithm scans the database at most two times. It is suitable for very large transactional databases.

In 1996, the researcher had proposed a novel association rules mining algorithm with some distinct features i.e. Fast Distributed Mining (FDM) algorithm (Cheung, et al., 1996). Initially, in this algorithm candidate sets has been generated in same way as in Apriori. To perform distributed mining, it associates the local frequent itemsets and the global frequent itemsets with some relationship. As a result of this relationship at each iteration, less number of candidate itemsets has been generated and therefore, number of massages passing for local and global frequent itemsets generation also gets reduced. On the other hand, two pruning methods were used in this algorithm. First was a local pruning method which trims the local candidate sets at every individual site. Second was global pruning method, which got the overall reduced candidate sets. In this algorithm, only O(n) number of massages were required to verify whether a candidate set is frequent or not. Here n is the total number of sites. So FDM algorithm performs better in contrast to Apriori because it required O(n^2) messages.

In this research work a new approach Generalized Sequential Patterns (GSP) has been was anticipated by the author (Srikant & Agrawal, 1996). The downward-closure property of sequential patterns has been employed in this algorithm. This algorithm is more flexible then Apriori in terms of candidate set generation and test approach.
The researchers (Brin, et al., 1997) proposed another algorithm called Dynamic Itemset Counting (DIC). This algorithm reduced the number of scans of the transaction databases by combining the counting for number of itemsets. In this algorithm the transaction database was divided into a number of partitions. The efficiency of algorithm depends on size of partition (M). The advantage of this algorithm is that it needs only two scans of the database.

In the subsequent year i.e. 1998, a new approach, Pincer-search algorithm (Lin & Kedem, 1998), for mining maximal frequent itemset was recommended. This algorithm is based on both top-down and bottom-up searching to discover frequent itemsets fruitfully. It classifies the data source into three classes which are frequent, infrequent, and unclassified data. In this, the Bottom-up approach is same as the Apriori. The top-down search uses a new set called Maximum-Frequent-Candidate-Set (MFCS). It also uses another set called the Maximum Frequent Set (MFS) which contains all the maximal frequent itemsets identified during the process. Any itemset that is classified as infrequent in bottom-up approach is used to update MFCS. Any itemset that is classified as frequent in the top-down approach is used to reduce the number of candidates in the bottom-up approach. When the process terminates, both MFCS and MFS are equal. This algorithm involves more data source scans in the case of sparse data sources.

Further in 1999, a Boolean algorithm was explored (Wur & Leu, 1999). In this algorithm, the logical operations AND, OR and XOR were employed to identifies the frequent patterns. The algorithm scans the database only once and it avoids the candidate itemset generation. They proposed two efficient implementations of the Boolean algorithm, the Bit Stream approach and the Sparse- Matrix approach. This algorithm is also suitable for large transactional databases.

A novel technique named as Continuous Association Rule Mining Algorithm (CARMA) was introduced in 1999 (Hidber, 1999). With the help of this algorithm frequent itemsets can be generated continuously with a decreasing minimum support value for each itemset. In this algorithm, the value of minimum support can be changed by user anytime during first competent investigation of pattern. For small support values, CARMA performs comparable to Apriori and DIC Algorithms.

In 1999 the researcher presented an algorithm for Closed Association Rule Mining (CHARM, here “H” is complimentary) (Zaki & Hsiao, 1999). This powerful algorithm was proposed for generating all frequent closed itemsets. In this algorithm, a dual itemset-Tidset search tree has been used as closed sets and with the help of this hybrid technique several
searching phases can be avoided. The experimental evaluation on various real world and synthetic datasets proved that the CHARM algorithm extensively outperforms preceding algorithms.

In 2000, the DepthProject (Agrawal, et al., 2000) approach was developed for mining maximal frequent itemsets. In this algorithm, to find frequent itemsets, a hybrid approach of breadth-first and depth-first traversals was used on itemset lattice. The subset infrequent itemsets reduction and superset frequent itemsets reduction properties were employed in this algorithm. The bitmap format has been used to represent the complete database. In bitmap database each row signifies to a transaction and each column represents to an item. So, the number of columns is equivalent to the number of items in bitmap dataset and the number of rows is equivalent to the number of transactions. Due to use of cautiously designed counting methods, the DepthProject algorithm drastically decreases the time to calculating the support of items.

The researchers discovered an algorithm called FP-Growth (Han, et al., 2000). Basically, this method was proposed for generating effective and scalable long and short frequent patterns. FP-tree is anticipated as a compressed data structure that symbolizes the database in tree structure. In this algorithm, each transaction has been scanned and then mapped onto a link in the FP-tree. This mapping has been performed for all transactions. Different transactions with common subsets make no change in FP-tree to remain compressed because their links overlie. In worst case scenario each transaction has a unique itemset and creates a new link of nodes in FP-tree. As a result, the tree takes more space in memory as compare to original database, because of extra space required to store tree pointers between nodes and also the counters for each item.

Eclat is an algorithm proposed by (Zaki, 2000) for discovering frequent itemsets from a transaction database. The first scan of the database builds the TID_set of each single item. Starting with a single item \((k = 1)\), the frequent \((k+1)\)-itemsets grown from a previous \(k\)-itemset can be generated according to the Apriori property, with a depth-first computation order similar to FP-growth (Han, et al., 2000). The computation is done by intersection of the TID_sets of the frequent \(k\)-itemsets to compute the TID_sets of the corresponding \((k+1)\)-itemsets. This process repeats, until no frequent itemsets or no candidate itemsets can be found.

During the mining of frequent patterns, researchers face the problem of several passes of a large database. The study presented a novel algorithm PASCAL (Bastide, et al., 2000) for quick discovery of frequent patterns. The PASCAL algorithm introduced as an optimization
of Apriori algorithm. In this algorithm pattern counting inference approach has been utilized. The concept of key patterns has been used in similar classes to reduce the number of patterns counted and database scans. The experimental outcomes prove that PASCAL is very proficient when data are dense or correlated.

The researcher has been proposed an algorithm SPADE for mining frequent sequential patterns from a sequence database proposed by (Zaki, 2001). The conjunctional properties were used to decompose the original problem into smaller sub-problems. These sub-problems can be resolved with competent lattice searching methods, and by using the simple join operations. These operations have been performed in main-memory. All frequent sequences can be generated in only three database scans.

The researchers (Ayres, et al., 2002) offered a novel algorithm SPAM for mining sequential patterns. This algorithm was offered to handle lengthy sequential patterns efficiently in the database. The researcher used the innovative depth-first search approach that incorporate a depth-first traversal of the search space with proficient pruning methods. A vertical bitmap demonstration of the database with efficient support calculation was implemented in this algorithm and it was also combined with depth-first search strategy.

In 2003, (Zaki & Gouda, 2003) presented a novel vertical data representation which differentiates the tids of a candidate itemsets from its generating frequent itemsets. The proposed method was named as Diffsets. This research work demonstrates that Diffsets algorithm is significantly reduced (by orders of magnitude) the size of memory required to maintain intermediary results.

A novel single-pass Data Stream Mining (DSM) algorithm has been implemented for generating frequent itemsets in 2004 (Li, et al., 2004). This algorithm was designed to identify all large itemsets over the history of data streams.

In 2004, an accomplished algorithm PRICES (Wang & Tjortjis, 2004) was developed which identify all the frequent itemsets which were used to create association rules. This algorithm takes only one pass of entire database and find out all frequent itemsets using logical operations. Due to this, PRICES algorithm takes less execution time as compare to Apriori algorithm.

The investigator presented an approach for extracting absolutely sporadic association rules. This approach named as Sporadic Rules Algorithm proposed by authors in 2005 (Koh, and Rountree, 2005). The sporadic rules defined as the rules which have lower value of support but higher value of confidence. In this algorithm, Apriori-Inverse method has been used by
researchers to discovery of sporadic rules. The sporadic rules can be excavated by ignoring all candidate itemsets exceeding a maximum support value.

For taking out the IGB association rules from transactional database an Informative and Generic Basis (IGB) method has been developed by investigator (Gasmi, et al., 2005). This was an integrated approach of compactness and the information lossless of the generic basis offered to the user. This integrated approach illustrates a new informative generic basis and an inclusive self-evident system permitting the derivation of all the association rules. To facilitate advance quality of exploitation of the knowledge, a new classification of "factual" and "implicative" rules has been presented.

A technique GenMax (Gouda & Zaki, 2005) was developed by researchers. It was based on backtrack searching for finding maximal frequent itemsets. In this algorithm, several optimization methods were used to prune the search space. The progressive focusing strategy was used to execute maximality checking, and diffsets propagation to achieve fast frequency calculation.

This research work has been explored a method called FPMax (Frequent Maximal Item Set) was developed which was based on FP-Tree (Grahne & Zhu, 2005). FPMax algorithm used a set of transactional data items which recognized from relational data model. Subsequently, minimum support and minimum confidence values required to generate frequent itemsets by applying array based structure than tree structure. Moreover, the FPMax algorithm was a deviation of the FP-growth algorithm, for mining maximal frequent item sets. In this algorithm a frequent itemset can be a subset only of a previously revealed MFI because FPMax is a depth-first algorithm.

The Fuzzy Healthy Association Rule Mining Algorithm was developed by investigators (Khan, et al., 2006). In this approach, edible attributes were filtered from transactional input data by rejections and are then converted to Required Daily Allowance (RDA) numeric values. The averaged RDA database is then converted to a fuzzy database that contains normalized fuzzy attributes comprising different fuzzy sets.

An innovative method H-Mine was proposed for discovering frequent itemsets from a transaction database (Pei, et al., 2007). In this algorithm researcher proposed a simple and novel data structure using hyper-links, H-struct. The H-Mine algorithm takes advantage of this data structure and dynamically adjusts links in the mining process. A distinct feature of the proposed method is that it has a very limited and precisely predictable main memory cost and runs very quickly in memory-based settings. Moreover, it can be scaled up to very large databases using database partitioning.
An algorithm FHSAR was developed which hides the sensitive association rules (Weng, et al., 2008). The algorithm can completely hide any given SAR by scanning database only once, which significantly reduces the execution time. The conducted results show that FHSAR outperforms previous works in terms of execution time required and side effects generated in most cases.

Reverse Apriori was presented to use in association rules mining for frequent pattern production (Kamrul, et al., 2008). The proposed approach generates large frequent itemsets only if it satisfies user specified minimum item support. It then gradually decreases the number of items in the itemsets until it gets the largest frequent itemsets.

Researchers discovered an inventive approach Distributed Trie-based Frequent Itemset Mining (Ansari, et al., 2008). This algorithm was proposed for a multicomputer environment and it is revised with some FDM algorithm ideas for candidate generation step. The proposed algorithm shows that Trie data structure can be used for distributed association rule mining not just for sequential algorithms.

The GIT-tree is a tree structure which was developed in 2009. It mines the frequent itemsets in a hierarchical database with the aim to reduce the mining time (Vo & Le, 2009). This algorithm scans the database only one time and use Tidset to compute the support of generalized itemset faster.

This work provides an Enhanced scaling Apriori algorithm for association rule mining (Prakash & Parvathi, 2010). This approach proposes an improved Apriori algorithm to minimize the number of candidate sets while generating association rules by evaluating quantitative information associated with each item that occurs in a transaction, which was usually, discarded as traditional association rules focus just on qualitative correlations. The proposed approach reduces not only the number of itemsets generated but also the overall execution time of the algorithm.

To generate the maximal frequent itemsets with minimum effort, an innovative method has been designed (Rajalakshmi, et al., 2011). This method was based on a different concept of partitioning the database as compare to other methods (Lin & Kedem, 1998). In this method, the database has been divided in to several segments and after that, these segments have been used for determining maximal frequent itemsets. It takes only two scans over the transactional database. So, candidate set generation time has been reduced by this method. This algorithm follows three steps to find out the MFS from a database. Firstly, the algorithm performs segmentation of the transactional database. After that, carry out the prioritization of the segments and finally mining frequent itemsets from these segments.
CMRules is an algorithm which mine sequential rules from a sequence database (Fournier-Viger, et al., 2012). The proposed algorithm proceeds by first finding association rules to prune the search space for items that occur jointly in many sequences. Then it eliminates association rules that do not meet the minimum confidence and support thresholds according to the time ordering. The tested results show that for some datasets CMRules is faster and has a better scalability for low support thresholds.

Frequent Pattern Growth Association Rule Mining approach was proposed for extracting association rules (Rao & Gupta, 2012). This approach focuses on the number of database scans, memory consumption, the time and the interestingness of the rules. They used a FIS data extracting association algorithm to remove the disadvantages of APRIORI algorithm which is efficient in terms of the number of database scan and time.

An approximate algorithm developed by (Fournier-Viger & Tseng, 2012) which aims to mine the top-$k$ non-redundant association rules that the named as TNR (Top-$k$ Non-redundant Rules). It was based on a recently proposed approach for generating association rules that is named “rule expansions”, and adds strategies to avoid generating redundant rules. An evaluation of the algorithm with datasets commonly used in the literature shows that TNR has excellent performance and scalability.

In the year 2013, an algorithm named ClaSP was proposed which is a method for mining frequent closed sequence proposed by (Gomariz, et al., 2013). This algorithm uses several efficient search space pruning methods together with a vertical database layout.

2.3 Section- III

Second section includes the research work carried on various multiple levels association rule mining methods with their problems and findings. A brief study is given as below:

A variety of approaches mentioned in previous section has been used to generate single level association rules, which hold enormously common information. To discover extra understandable and interpretable facts, multiple level association rules were came into existence. From last decade, several researchers have been paying attention to mining association rules at multiple level of abstraction. A comprehensive survey, of research methods for taking out the frequent itemsets and association rules at multiple levels of abstraction has been presented in this section. Initially, the conception of association rules mining, at multiple levels of abstraction was introduced in 1995 (Han & Fu, 1995). Authors built up a new algorithm ML_T2L1 with several extensions named as ML_T1LA, ML_TML1 and ML_T2LA. The author implemented the top-down progressive deepening approach to develop theses algorithms. This research explores the existing single-level
association rule mining algorithms and realizes techniques for sharing information structures.

In the next year, Investigators presented a competent algorithm named as MLUp (multiple level association rules update) (Cheung, et al., 1996). This algorithm was developed for revising the of exposed multi-level association rules. This algorithm is appropriate for those databases, which are updated frequently or occasionally by insertions of new transactions.

Subsequently in the year 1998, a novel algorithm PRUTAX was introduced by authors for extracting frequent itemsets at multiple levels (Hipp, et al., 1998). The algorithm PRUTAX is based on the concept of hash tree. In this algorithm, the count of only those candidate itemsets has been calculated where all subset itemsets are frequent. The investigation's outcomes explain that numbers of support computations has been reduced by this method. A relatively different approach was introduced in 1999 by (Han & Fu, 1999). The researchers had analyzed the methods proposed in (Han & Fu, 1995), and presented enhanced and optimized algorithms. Starting with the topmost level of the taxonomy, the rules on all level were extracted with decreasing support value as going down in the taxonomy. The extraction of rules, including items of different levels is not possible by this method.

The association graph construction (AGC) algorithm has been produced by building of an association graph (Jane & Arbee, 2001). This algorithm is not suitable in case of large database because relevant information can not fit in the main memory. Along with this algorithm, researchers had proposed the MLAPG (Multiple-Level Association Pattern Generation) algorithm which discovers itemsets at all levels of abstraction (Jane & Arbee, 2001). In this approach, only one scan of database has been required to construct the association graphs of that database. After that all frequent itemsets have been generated by traversing of that association graph. Additionally, the infrequent items at the prior concept level have been pruned in order to reduce the size of database. At the last level, no further reduction of the database is required.

The AprioriNewMulti algorithm was presented by investigator (Rajkumar, et al., 2003) for mining multiple-level association rules where the minimum support for itemsets depends upon the length of itemset. A new notion of multi minimum support was introduced in this algorithm. In this algorithm, the different value of minimum support was used for different lengths of the itemset. The large itemsets of various lengths has been discovered with the help of multi minimum support. This algorithm did not use the idea of concept hierarchy.
Later, in the year 2005, the researchers had developed a competent approach for extraction of positive and negative association rules (Sharma, et al., 2005). In this approach, the researcher applied the multiple level spatial mining schemes to mine the interesting patterns in spatial or non-spatial predicates.

An efficient hierarchical online rule mining algorithm was developed to optimize the time requirements of the earlier reported algorithm HORM (Kumar & Jotwani, 2006). The researchers included two particular enrichments in the designed algorithm. First is hierarchy-aware counting and second is transaction reduction. This algorithm generates the hierarchical association rules by natural modification in new algorithm. This algorithm avoids the building of a precise adjacency lattice.

The researchers (Cao, et al., 2007) devised a fast algorithm FAMML_FPT which was based on frequent pattern tree to extract the association rules at multiple levels. This method introduced the notion of repaired items and the cross-level repaired items, which encourage to generate FP-tree from lower levels to higher levels. In this algorithm, association rules discovering time has been reduced because there is no need of generating the candidate itemsets.

This research work has been presented a fresh technique ADA-AFOPT (Mirela & Popescu, 2008). This algorithm was proposed to resolve the problem associated with mining of frequent itemsets at multiple levels. This algorithm adopted both, depth-first and top-down, traversals to give more precise frequent itemsets. The items in database have been stored according to their support value in increasing order. To provide more accurate frequent itemsets the concept of non-uniform minimum support has been used in this algorithm. First of all, this algorithm scans the original database to find frequent items and arranges them in increasing order of occurrence. After that, an AFOPT structure is created in the second scan which is to signify the conditional databases of the frequent items. This AFOPT structure is used to compress the complete information regarding the transactional database. Later the itemset generation procedure is preceded by recursively traversing of the AFOPT structure.

The Level Wise Filter Table algorithm (LWFT) was established by researcher in order to reduce memory requirements (Vishav, et al., 2010). This algorithm was implemented for mining multiple level association rules from large transactional databases. This method extends the scope of mining frequent patterns at different levels, by using a top-down dynamic developing procedure. In this algorithm, at each concept level the number of database scans been reduced by applying an innovative itemsets counting implication approach. This approach has been based on the conception of key patterns of
correspondence class of itemset. The LWFT algorithm also makes the use of reduced database at each concept level, which has been provided by filtration of encoded table at each level. Experimental results have been proved that LWFT algorithm performed superior in terms of execution time and memory.

This study investigated a Boolean matrix based new model Multi Level Boolean Matrix (MLBM), for mining multilevel association rules (Gautam & Pardasani, 2010). This method identifies the frequent itemsets at lower levels by implementing boolean vector relation calculus. This algorithm takes only one scan of database to convert it in Boolean matrix. After that, Boolean logic operations have been used on that Boolean matrix to produce the multilevel association rules.

A different approach was presented which was based on FP-tree structure (Shrivastva, et al., 2010). A hybrid strategy has been implemented on this algorithm. In this approach the FP-Growth tree has been combined with COFI (co-occurrence frequent item) to create multiple level association rules. This algorithm constructs COFI-tree to make use of the memory space in efficient way. After that a straightforward traversal of COFI-tree has been required to generate frequent patterns. So, there is no requirement of repetitive mining process.

In order to improve the efficiency, a multi-level association rule mining method was presented, which applied a secondary storage structure (Tang, et al., 2011). In this algorithm, hash table was used to enhance the efficiency of searching items. The experimental results have been proved the significance of this algorithm.

AC tree, a hybrid algorithm for mining multilevel association rules was presented by inventers (Ramanaiah, 2011). The algorithm AC tree has combined Apriori and COFI tree to overcome the shortcoming of these conventional algorithms and make it appropriate for mining frequent itemsets at multiple levels. This method generally performs 2 key steps. During the initial step, Apriori algorithm is used to find out the frequent 1-itemsets by scanning the dataset. After that in next step, FP header table is created by those 1- frequent itemset, subsequently these FP trees for each element is considered for mining frequent itemset at each and every level.

Subsequently, in the year 2012, researchers presented an innovative multilevel fuzzy association rule mining approach for extraction of hidden knowledge (Kousari, et al., 2012). This approach uses the concept of different minimum support threshold at all levels and different membership function for each item. This method includes the method fuzzy boundaries instead of sharp boundary intervals. As a result of this, it presents the extracted rules which are more close to reality.
A different method for multi-level association rules mining was proposed in 2013 (Han, et al., 2013). Proposed framework has been used to provide multisource geo-knowledge in form of association rules. This Multi-Level Association Rules (MLAR) algorithm has been presented as an enhancement of FP-Growth model, which performs better in frequent itemset mining.

To speed up the discovery of multiple levels association rules and to reduce the unnecessary computation, a new genetic-based algorithm was introduced (Xu, et al., 2014). In this algorithm, a tree encoding schema is used, to decrease the association rule exploration space. The empirical results demonstrate the efficiency and competence of the proposed method in big data.

**2.4 Section-IV**

Previous segment explore the methods and techniques used to mine association rules at multiple-levels of abstraction. This section includes reviewed papers which are not suited to above three sections.

Different kind of researchers had proposed various improved algorithms using diverse data structures. In 2003, Researcher presented a frequent itemsets mining algorithm, COFI which is based on COFI-tree structure (El-Hajj & Zaoiane, 2003). The efficiency of the algorithm was accomplished by implementing four new notions. First, it uses a compact memory base data structure for mining. Second, for each frequent item dispersed, a comparatively small independent tree builds the summarizing co-occurrences. Third, the search space significantly gets reduced due to intellectual pruning. All locally non-frequent itemsets has been removed by applying significant pruning method. The COFI-tree has been constructed with only locally frequent items. Finally, a straightforward and non-recursive technique applied for mining process which reduced the memory requirements and all frequent itemsets can be generated with a traversal of the COFI-tree. The author concluded that proposed method go one better than the FP-Growth algorithm in memory requirements, and execution time. As a result larger transactional databases can be processed with smaller main memory available. The COFI algorithm focused on the discovery of reduced candidates during the mining. This algorithm has been avoided the recursion that FP-growth uses.

An innovative disk-based association rule mining method named as Inverted Matrix was developed (El-Hajj & Zaoiane, 2003). The algorithm accomplished its competence by converting transactional data into a new Inverted Matrix layout. This inverted matrix avoided the multiple scans of the database during the mining phase. This Inverted Matrix
algorithm outperforms FP-Tree mainly in case of extracting frequent patterns from very large transactional databases with a very large number of unique items. This random access disk-based algorithm is mainly useful in a repetitive and interactive situation.

In this research paper author had presented an advanced version of trie for solving association rules mining task (Bodon & Ronyai, 2003). The author proposed an innovative approach instead of classical approaches. An advance data structure trie was used instead of a hash tree. A trie is a rooted, labeled tree. In this implementation, the value of each node has been taken as the support count of that itemset. Empirical results have been demonstrated that at high support values proposed algorithm perform about to the performance of hash-trees. On the other hand at lower support values trie-based algorithm outperforms the hash-tree based ones. Furthermore, tries are particularly appropriate for candidate generation because collection of items that create candidates have the same parents. Thus, candidates have been obtained simply by the scan of that trie. Last, maybe the most important feature of trie is their self-adjusting behavior. Author showed that if the parameters of the hash-tree are not close to the optimal values, then the hash-tree based algorithm can give very poor results. Consequently, efficient use of the hash-tree based algorithm requires users to understand the technical details. In contrast, no tuning is needed for tries, and hence, the algorithm is easier to use.

In the year 2005, an efficient way was proposed in which three data structures: prefix tree, bitmap and array lists for the calculation were combined, so that the combination gives the best performance (Uno, et al., 2005). In this paper, author proposed a new data structure for frequent itemsets mining algorithms and applied to LCM algorithm. By computational experiments it shows that purposed method performs batter at any minimum support.

The author recommended an innovative approach in mining association rules (Huang, et al., 2007). Researchers proposed a new algorithm Gradational Reduction Approach (GRA). In this algorithm, three mechanisms have been implemented to amplify the performance of association rules mining. Initially, GRA algorithm has employed a hash based procedure, Hash MAP, which is used to enhance the access efficiency. Secondly, this proposed algorithm has realized a mechanism to filter infrequent itemsets. Finally, with the purpose of reducing the size of database, this algorithm has applied gradational reduction method. In order to remove the infrequent items from database at every phase, this gradational reduction method has used the frequent itemsets as the information of filtration mechanisms. Consequently GRA algorithm has reduced large number of non-frequent itemsets and enhances the usefulness of memory.
The majority of offered algorithms deal with traditional transactional databases that contain precise data for discovery of frequent. Despite the fact that, there are many real-life circumstances in which one needs to deal with uncertain data. To handle these circumstances, author presented a tree-based mining algorithm (Leung et al., 2008). This tree-based algorithm has been offered to handle uncertain data where each item in the transactions is associated with an existential probability. In this work, the researcher has been adopted two key concepts: first is the UF-tree to efficiently capture the content of uncertain transactional databases and other is a tree-based mining method named as UF-growth to proficiently discover frequent patterns from UF-trees. This algorithm also presented two improvements to UF-growth are the rounding of probable support values and the purging of UF-trees for projected databases for non-singleton patterns. Due to these improvements this algorithm takes less amount of memory and takes less time in mining process. Consequently, with the help of this tree-based method, user can mine frequent itemsets from uncertain data successfully.

In this work researcher has presented an innovative feature selection algorithm (Xie, et al., 2009). Author has intended an adaptive feature selection strategy using Apriori algorithm for generating association rules. Investigational outcomes have been revealed that the proposed feature selection algorithm is capable to achieve a smaller and acceptable feature subset as compare to existing feature selection algorithms. However it presents a significant improvement in reducing the feature number, because of using Apriori algorithm directly to mining association rules, the time complexity of the algorithm is quite high.

A Performance Based Transposition Algorithm (PBTA) for frequent itemsets generation is offered in research work (Sharma & Ugrasen, 2011). According to author this algorithm is used to improve the efficiency of Apriori algorithm and it is observed to be very fast. Proposed algorithm is not only efficient but also very fast for finding association rules in large databases. The proposed algorithm drastically reduces the I/O overhead associated with Apriori algorithm and retrieval of support of an itemset is quicker as compared to Apriori algorithm. This algorithm may be useful for many real-life database mining scenarios where the data is stored in boolean form.

In this paper investigator described text mining technique for automatically extracting association rules from collections of textual documents (Bhujade & Janwe, 2011). The technique called, Extracting Association Rules from Text (EART). It depends on keyword features for discover association rules amongst keywords labeling the documents. EART system focused on the words and their statistical distribution in documents rather than
paying attention to the order in which the words occurs. The system based on Information Retrieval scheme (TF-IDF) for selecting most important keywords for association rules generation. It consists of three phases: Text Preprocessing phase (transformation, filtration, stemming and indexing of the documents), Association Rule Mining (ARM) phase (applying our designed algorithm for Generating Association Rules based on Weighting scheme GARW) and Visualization phase (visualization of results). Experiments applied on Online WebPages related to the cryptography. The extracted association rules contain important features.

An improvement of Frequent Pattern Tree (FP-Tree) has been presented in this research (Abdulla, et al., 2014). Basically FP-Tree is a compact data structure of representing frequent itemsets. The FP-Tree requires two scans of database, first to construct and sort the frequent itemsets on the basis of prior knowledge of support threshold and second to make its prefix paths. Thus, two times database scanning is a major drawback in implementation of FP-Tree. Therefore, the author proposed scalable Trie Transformation Technique Algorithm (T3A) to convert predefined tree data structure, Disorder Support Trie Itemset (DOSTrieIT) into FP-Tree. Experiment results exposed that the proposed T3A generates FP-Tree up to 3 magnitudes faster than that the standard FP-Growth.

2.5 Outcome

On the basis of literature reviewed, challenges faced by the rules mining domain are highlighted. This literature review is related to association rules mining at different levels of abstraction. Single level association rules are providing the general knowledge, which is not helpful to get concise and specific knowledge.

Existing algorithms that accomplish the task of mining association rules at multiple levels of abstraction with Apriori based methods are not efficient to achieve time efficiency. The repetitive numbers of database scan are required to generate the candidate sets. Proposed methods that employ FP-growth approach are inefficient in the way that they use a large amount of computer memory to store conditional trees for frequent itemsets mining. However most kind of methods produce redundant multiple levels association rules in association rule discovery process.

To encounter these problems associated with single/multiple levels, an efficient approach is required which will overcome the problem of repetitive database scans and space requirement.
However, few work has been done in the field of “design of an improved multiple level association rule algorithm for discovery of frequent patterns”. Therefore, this research work will try to fill up this gap.