FREQUENT PATTERN MINING TECHNIQUESFOR VARIOUS FORMS OF PATTERNS IN DATA ANALYSIS

[1]Mrs. E. SHEEBA SUGANTHARANI & [2] Dr. A.SUBRAMANI
[1]Associate Professor, [2]Assistant Professor
[1]Department of Computer Science, Lady Doak College, Madurai,
Tamil Nadu, India- 625002
[2]Department of Computer Science, M.V. Muthiah Govt. Arts College for Women,
Dindigul, Tamil Nadu, India - 624001.

I. ***Abstract:*** Data mining involves identification of important trends or patterns through huge amounts of data. Advanced statistical techniques such as cluster analysis, artificial intelligence and neural network techniques are used in the data analysis processes. Data mining helps in better analysis of geographical data, Genome and medical sector. Classification is used for predicting outcomes and association is used to find rules affiliated with items having co-occurrence. Frequent Itemset Mining (FIM) is an approach to discover association rules in datasets. Frequent Pattern Mining (FPM) is used for finding relationships among the items in a large database obtained from the cloud environment. Association rule mining is applied for obtaining the frequent patterns. Association rule mining and frequent itemset mining are two popular and widely studied data analysis techniques for a wide range of applications such as market basket analysis, healthcare, web usage mining, bioinformatics, personalized recommendation, network optimization, medical diagnosis. This paper reviews different frequent pattern mining algorithms with weighted, interesting pattern and uncertain databases. A brief comparison of various mining algorithms based on their metrics, dataset , inferences of their work with few drawbacks were summarized. According to the reviewed papers, it was observed that uncertain database requires larger storage space and it was a time consuming process. Moreover, various challenges include checking accuracy and efficiency with time bound, setting the threshold criteria, choosing the appropriate datastructure and number of transactions containing the itemset.

***IndexTerms* -** Frequent Pattern Mining, uncertain databases, Weighted frequent itemset mining, interesting patterns, BFI-forest.

I. INTRODUCTION

Pattern mining is a subfield of Data Mining as suggested by Dr. KanakSaxena, D.S. Rajpoot,(2009). Pattern mining concentrates on developing data mining algorithms for finding interesting, unexpected and useful patterns in databases. The various data types supported by these algorithms are strings, sequence database, graphs, and spatial data. Various types of patterns can be discovered such as sequential patterns, periodic patterns, subgraphs, association, indirect association, lattices.
The main focus of this paper is to review regarding interesting patterns and uncertain databases with reference to Dr. HuiXiong, (2006). In a database if a pattern appears frequently then it is called as interesting. Sometimes patterns that have high confidence and top patterns may also be called as interesting pattern. AnyFI (Anytime Frequent Itemset) mining algorithm for data streams and BFI (Buffered Frequent Itemset)-forest has been proposed by PoonamGoyal, JagatseshChalla, ShivinShrivastava, NavneetGoyal, (2017). This BFI-forest is proficient in treating transactions with varying inter-arrival rate. AnyFI maintains itemsets in BFI-forest in such a way that it can give a mining result almost immediately when time allowance to mine is very less and can refine the results for better accuracy with increase in time allowance.
In some cases, while mining frequent item sets having dense type database may results in redundant of information. To overcome this problem a weighted frequent itemsets mining algorithm is applied in many referred proposals. These algorithms concentrate on two combined measures such as support and weight. To reduce the number of candidate itemsets the methods used are pruning and hash table. Pruning reduces the complexity of the final classifier, and hence improves predictive accuracy by the reduction of over fitting. The hash table is used to store weighted non-frequent binomial set.

The main purpose of this review is to present the taxonomy of the existing Frequent Pattern Mining along with the significant advantages and issues. The main contributions of this review are stated as: -

- Providing the basic concepts and terminologies used in the field of Frequent Pattern Mining.
- Discussing about the two main types of Patterns: interesting pattern, uncertain databases.
- Presenting which problem, they solved and how they solved it.
- Analyzing the constraints under them tested their solutions.
- Summarizing how their solution mechanisms are better than the previous work and supported by which parameters.
- Highlighting the result of the paper.
- Concluding with the drawback in their papers.
  The remaining sections in the paper are systematized in the following order: Section III includes comparison of various Mining algorithms. Survey of Existing papers is presented in Section IV. Section V comes up with the conclusion of the review.

Paper 1:
Xuejian Zhao, Xinhui Zhang, Pan Wang, Pan Wang, Song Chen and Zhixin sun (2018), “A Weighted Frequent Itemset Mining Algorithm for Intelligent Decision in Smart Systems”. Instead of considering binary data model they have analyzed with uncertain data. In a transaction, the existence of an item is determined by a measure or probability, and then it is an uncertain data. They lead to more accurate analytical results. Even though they have drawbacks such as larger storage, more complicated and time consuming, in reality, all items have different value and importance. But, in uncertain data it has same importance. When mining based on only occurrence frequencies without taking importance or values of items into account is insufficient to identify meaningful patterns. Items with high importance will be discovered based on the weights of the item. This means that, infrequent itemset may have a frequent superset. Therefore WD-FIM algorithm is proposed to narrow the searching space, improve time efficiency by using the property weighted judgment downward closure and existence property for ensuring the item discovery.

III. TABULATION

Comparison of various mining algorithms

Paper	Author	Year	Proposed Algorithm	Metrics	Dataset	Related works/Ref. algorithms	Inferences	Drawback
WFI Mining Algorithm for IDSS[2]	Xuejian Zhao, Xinhui Zhang, Pan Wang, Pan Wang, Song Chen and Zhixin sun	2018	WD-FIM algorithm	(i) Completeness (ii) Time efficiency (iii) Memory consumption. (iv) searching space	(i)Mushroom (ii) retail (iii)T10I4D100K	(i)Frequent itemset mining in uncertain databases (ii)Weighted frequent itemset mining in uncertain databases (iii)Uapriorian d HEWI-Uapriori algorithm	(i) The weight judgment downward closure property is used to narrow the searching space of weighted frequent itemsets. (iii) Existence property to ensure the discovery (ii) Improve time efficiently.	(i) uncertain databases requires larger storage space, more complicated and time consuming.. (ii) uncertain data has same importance value so an additional attribute weight is used . (iv) Time efficiency of this algorithm depends on the number of candidate and it is faster only when the support threshold is relatively large. (iii) Uapriori algorithm requires less memory than the proposed WD-FIM algorithm
An efficient approach for mining uncertain frequent patterns using dynamic data structure without false positives[30]	Arunnya Radhakrishnan , Vijayakumar R	2018	Linked List based Uncertain Frequent Pattern Mining Algorithm (LUFPA)	(i) Memory (ii) Usage, (iii) Runtime, (iv) Scalability,	(i) Accidents (ii) Connect (iii) Kosarak (iv) Pumsb (v) Mushroom	LUNA[31]	(i) Effectively store given uncertain data without any false positives (ii) Perform uncertain frequent pattern mining operations with less runtime and memory resources (iii) If the current threshold is set lower, the scalability becomes better	The drawback is that only if the current threshold is set lower, the scalability of LUFPA becomes better than the other data structures and various mining techniques. As Linked List is used additional memory to store the link is required. Setting the threshold criteria is not mentioned.
AnyFI : An Anytime Frequent Itemset Mining Algorithm for Data Streams [1]	PoonamGoyal, Jagat SeshChalla	2017	(i) AnyFI, Anytime Frequent itemset mining algorithm (ii) BFI-forest (Buffered Frequent Itemset Forest)	(i)varying inter-arrival rate (ii) Speed (iii) accuracy	(i) Retail (ii) MSNBC		(i)Handling varying inter arrival rate of transactions (ii) Giving the best possible result according to the available time allowance.	As this algorithm uses tree data structures, insertion and deletion of tree nodes require additional traversal process which will result in excess time bounds. Accuracy is related to increased time allowance.
Review of Algorithm for Mining Frequent Patterns from	LiwenYue	2015	SUF-growth algorithm	(i) Size of the Database (ii) execution time /running time (iii) system consumption (iv) accuracy	(i)IBM (ii)FIMI (iii)UCI	(i)FP-Growth Algorithm[20] (ii) U-Apriori	Trimming module, pruning module and patch up module.	The drawback is that it is not suited for maximal frequent patterns, frequent closed patterns and constraint based. And they haven't focused on the efficiency and application area.

Uncertain Data[3]								
An Uncertainty-based Approach : Frequent Itemset Mining from Uncertain Data with Different Item Importance, Knowledge-Based Systems [23]	G. Lee, U. Yun, H. Ryang,	2015	U-WFI	(i) items' existential probability and weight factors (ii) quickly access with less memory (iii) more efficient and scalable	(i)Connect (ii)Kosarak (iiii)Pumsb	(i)WFIM[24] (ii)MWS[25] (iii)WMFP-SW[26] (iv)MCWP[27] (v)WEP[28] (vi)U-Apriori (vii)MBP (viii)IMBP (ix)UF-Growth (x)UH-Mine[19] (xi)CUFP-Mine[29] (xii)AT-Mine[16] (xiii)U-WFI	(i)Propose a new uncertain itemset mining algorithm considering importance of items such as weight constraints (ii) Selectively obtain more meaningful itemsets with high importance and existential probabilities	In U-WFI they employed own tree and list data structures for mining process. If the uncertain database size becomes extremely large, it is impossible to load the data in the main memory during the mining process. This algorithm focuses on static uncertain database, but in real world we will have dynamic DataStream too.
Weighted frequent itemset mining over uncertain databases [38]	Jerry Chun-Wei Lin,Wensheng Gan1 ,Philippe Fournier-Viger,Tzung-Pei Hong,,VincentS,Tseng	2015	HEWI-Uapriori(High Expected Weighted Itemset	(i) Performance and (ii) scalability	(i) Retail (ii)food mart (iii)muhroom (iv)T10I4D100K	Apriori-like two-phase approach	(i) Proposed a property high upper-bound expected weighted downward closure (HUBEWDC) to early prune the search space and unpromising itemsets. (ii) Proposed to consider both the weight and existential probability constraints (iii) Revealed more interesting and meaningful Information(importance of each item is considered in an uncertain database).	This algorithm consider only weighted and existential probability constraints and also with static database.
Comparative Study of Various Frequent Pattern Mining Algorithms[39]	Amit Mittal1, Ashutosh Nagar, Kartik Gupta, Rishi Nahar	2015	Review of Apriori FPGrowth DIC	(i)execution time (ii)value of support (iii) importance	(i)Medicine (ii)Letrecog (iii)Nursery (iv)Retail	(i)Apriori (ii)FPGrowth (iii)DIC	(i)The fastest algorithm FP-growth followed by Apriori. (i)DIC takes more time as compared to other algorithms for same datasets various parameters of importance.	-
Tightening upper bounds to the expected support for uncertain frequent pattern mining [12]	Carson K. Leung*, Richard Kyle MacKinnon , Syed K. Tanbeer	2014	TPC-tree	(i)number of transactions (ii) Execution time (iii)space	(i)IBM (ii)Almaden (iiii)Research (iv)Data, (v)kosarak, (vi)mushroom and (vii)retail	(i)UF-tree (ii) UFP-tree (iii) PUF-tree	(i) Scalable with respect to the number of transactions (ii) Can mine high volumes of uncertain data within a reasonable amount of time.	The TPC-growth scans the uncertain data a third time to compute the true expected support and to eliminate the small number of false positives. And it takes up more (e.g. 50%) space than the existing PUF-growth algorithm as shown in results.
Sliding window based weighted maximal frequent pattern mining over data streams[32]	GanginLee ,Unil Yun , Keun Ho Ryu	2014	WMFP-SW	Performance in terms of runtime, memory usage, and scalability over the sliding window-based data streams.	(i)Accidents (ii)Connect (iii)Kosarak (iv)Pumsb (v)Mushroom (vi)T10L1000N10000 (vii)T10I4D100 K series	(i) FP-Growth [33] (ii) IWFP algorithm [34] (iii) FPmax, (iv) MFP mining methods[35] (v) CPS-tree[36]	(i) Obtained weighted maximal frequent patterns reflecting recent information over data streams. (ii) Outperforms previous algorithms in terms of runtime, memory usage, and scalability.	This WMFP-SW algorithm proved the performance without checking the accuracy constrain.

Research of the FP-Growth Algorithm Based on Cloud Environments[40]	Lijuan Zhou, Xiang Wang	2014	Parallel FP-Growth algorithm (PLFPG)	higher operating efficiency and better scalability and extensibility	Transaction	(i)FP-Growth algorithm, (ii)HDFS framework [42] (iii)MapReduce programming model[41].	(i)Implemented parallel FP-Growth algorithm based on the linked list, (ii)Discovered valuable knowledge model from large-scale data in a fast and efficient way.	For experimental results they used only one database Transaction on parallel system.
Mining frequent itemsets using the N-list and subsume concepts[43]	Bay Vo ,Tuong Le , FransCoenen,Tzung-Pei Hong,	2014	N-list and Subsume-based algorithm(NSFI)	runtime and memory usage	(i)Accidents (ii)Chess (iii)Mushroom (iv)Pumsb_star (v)Retail	(i)PrePost[45] (ii)dEclat[44]	(i) Use of a hash table to enhance the process of creating the N-lists associated with the frequent 1-itemsets and (ii) An improved intersection function to find the intersection between two N-lists. algorithm improves the runtime and memory usage.	NSFI does not improve runtime and memory over PrePost respect to sparse but only with dense datasets.
AT-Mine: An Efficient Algorithm of Frequent Itemset Mining on Uncertain Dataset[16]	Le Wanga,Lin Fenga,b, and Mingfei Wu	2013	AT-Mine CUFP-Mine	(i)time &performance (ii)memory usage (iii)decreasing of the minimum expected support threshold.	(i) T20I6D (ii) kosarak (iv) connect (v) mushroom	(i)FIM algorithms (ii)U-Apriori , (iii) MBP [17] and (iv)IMBP [18]	(i)AT-Mine has a better performance than UF-Growth, CUFP, (ii)Number of candidate itemsets generated by MBP algorithms in terms of memory usage.	It requires two sets of scanning process of datasets. The number of transactions (D) tree nodes(I) used in the two datasets(sparse, dense) having varied count.
Mining maximal frequent patterns by considering weight conditions over data streams[37]	UnilYun ,Gangin Lee , Keun Ho Ryu	2013	MWS for mining WMFPs over data streams	(i)performance in terms of runtime, memory usage (ii) scalability	(i)Accidents (ii)Connect (iii)Kosarak (iv)Pumsb (v)Mushroom (vi)T10L1000N10000 (v)T10I4D100 K series	(i)FP-Growth[33] (ii)IWFP algorithm[34] (iii)FPmax, (iv)MFPminingmethods[35] (v)CPS-tree[36]	(i)More efficiently extract WMFPs streams with one database scan, and it could conduct mining operations more quickly and effectively (ii)Applying WMFP-array, pre-pruning useless patterns by weight conditions, and processing single-paths	Performance testing not based on accuracy and number of itemsets. They have applied only in the field of Maximal frequent pattern mining.

Significant differences between WD-FIM algorithm and U-Apriori algorithm: -The WD-FIM algorithm mines the weighted frequent itemsets from an uncertain database using the candidate generate-and-test paradigm.

(i) U-Apriori can only be used to discover frequent itemsets in uncertain datasets. But WD-FIM algorithm is proposed for mining weighted frequent itemsets in uncertain datasets.

(ii) Downward closure property is used directly to narrow the searching space of frequent itemsets in U-Apriori algorithm. But the basis of the proposed WD-FIM algorithm is the aforementioned weight judgment downward closure property and existence property of weighted frequent subsets.

Performance analysis: According to the pseudo code of the WD-FIM algorithm to narrow the searching space of weighted frequent itemsets by deleting itemsets which are definitely not weighted frequent. The time efficiency of a weighted frequent itemset mining algorithm mainly depends on the number of candidate weighted frequent itemsets. According to the analysis the HEWI-Uapriori algorithm have to identify actual high expected weighted itemsets among high upper-bound expected weighted itemsets by performing an additional round of database scan. But WD-FIM algorithm is more time efficient than HEWI-Uapriori algorithm on proved condition.

Experimental result: The three groups of experiments are carried out to show the performance of the proposed WD-FIM algorithm in terms of runtime, number of patterns and memory consumption. The sizes of the dataset are fixed. The runtime of the proposed WD-FIM algorithm and the compared algorithms are first analyzed. The runtimes of Uapriori algorithm, HEWI-Uapriori algorithm and WD-FIM algorithm all decline gradually with the increase of minimum expected weighted support threshold. The WD-FIM algorithm is faster than Uapriori algorithm and HEWI-Uapriori algorithm when the minimum expected weighted support threshold is relatively large. This is because the number of candidates can be effectively reduced by the proposed weight judgment downward closure property. For analyze the pattern for the same dataset, the itemsets produced by the compared three algorithms have the same distribution trend.

Inferences: i) The weight judgment downward closure property is used to narrow the searching space of weighted frequent itemsets.

(ii) Existence property to ensure the discovery

(iii) Improve time efficiently.

Drawback: (i) uncertain databases require larger storage space, more complicated and time consuming.

(ii) Uncertain data has same importance value so an additional attribute weight is used.

(iii) Time efficiency of this algorithm depends on the number of candidate and it is faster only when the support threshold is relatively large.

(iv) Uapriori algorithm requires less memory than the proposed WD-FIM algorithm.

Paper2: Arunnya Radhakrishnan1,Vijayakumar R2, (2018) "An efficient approach for mining uncertain frequent patterns using dynamic data structure without false positives". This paper focus on uncertain pattern mining. The concept of uncertain pattern mining is to discover interesting pattern information from uncertain databases (existential probability values). When an algorithm has to be considered efficient, it has to guarantee faster runtime, smaller memory usage, and better scalability compared to state-of-the-art techniques, without mention its accuracy. If time to mine interesting patterns becomes too longer, it can cause fatal problems and memory overflow. With reference to this problem, a new approach called "List Based Uncertain Frequent Pattern Mining Algorithm (LUNA) from GaninLee, Unil Yun,(2016), which is used for mining uncertain frequent patterns based on novel data minimum structures, but it has limitations in runtime performance. Manipulation with Array List is slow because it internally uses array. If any element is removed from the array, all the bits are shifted in memory.

Inferences:

(i) Effectively store given uncertain data without any false positives

(ii) Perform uncertain frequent pattern mining operations with less runtime and memory resources

(iii) If the current threshold is set lower, the scalability becomes better

Drawback:

The drawback is that only if the current threshold is set lower, the scalability of LUFPA becomes better than the other data structures and various mining techniques.

As Linked List is used additional memory to store the link is required. Setting the threshold criteria is not mentioned.

Paper 3: AnyFI , An Anytime Frequent Itemset Mining Algorithm for Data Streams, PoonamGoyal, JagatseshChalla, ShivinShrivastava, NavneetGoyal, (2017) .Accuracy Relevant to time are the two metrics focus in this algorithm. To handle varying inter-arrival rate, a novel data structure, BFI-forest (Buffered Frequent Itemset Forest) is proposed here. The experimental results show better accuracy with increase in time allowance. According to the already existing FIM algorithms the core drawbacks are (i) Fixed Stream speed (ii) No Immediate result.

(i) Fixed Stream speed: According to the application domain the speed of the streams varies. For example, Retail shop the rate of transaction is high in rush hours and vice versa.

(ii) No Immediate result: Even with the compromising of accuracy the existing algorithm is not able to produce an immediate result.

BFI is used to insert the incoming transaction as tree nodes

Inferences :(i)Handling varying inter arrival rate of transactions (ii) Giving the best possible result according to the available time allowance.

Drawback: As this algorithm uses tree data structures, insertion and deletion of tree nodes require additional traversal process which will result in excess time bounds. Accuracy is related to increased time allowance.

Paper 4: LiwenYue, (1994) "Review of Algorithm for Mining Frequent Patterns from Uncertain Data ". In this paper, the author reviewed the Apriori algorithm and FP-growth algorithm, and then analyzed the SUF-growth algorithm for mining frequent patterns form of uncertain data and data streams. Apriori-based depend on a generate-and-test paradigm to mine frequent patterns from transaction databases of precise data by first generating candidates and then checking their actual support (i.e., occurrences) against the database as discussed by R.Agrawal , R.Srikant ,(1994). To improve algorithmic efficiency, tree-based frequent pattern mining algorithms as discussed by J.Han, J.Pei, Y.Yin, (2000) construct an extended prefix-tree structure (FP-tree) which is discussed in J. Han, J. Pei and Y. Yin, (2000) to capture the contents of the database and perform the mining process using a restricted test-only approach. These tree-based algorithms do not generate candidates, only test for support. Algorithm scans the transaction database every time to determine frequent item sets among the candidates, when the database is very large, candidate sets will generate large numbers of subsets. Apriori has low efficiency especially when the database very big and a large number of item sets exist. The FP-Growth Algorithm discussed by B. Chikhaoui, S. Wang, & H. Pigot, (2011) is an efficient and scalable method for mining the complete set of frequent patterns by pattern fragment growth, using an extended prefix-tree structure for storing compressed and crucial information about frequent patterns named frequent-pattern tree (FP-tree).

Inferences: Trimming module, pruning module and patch up module.

Drawback: The drawback is that itis not suited for maximal frequent patterns, frequent closed patterns and constraint based. And they haven't focused on the efficiency and application area.

Paper 5: Carson K. Leung, Richard Kyle MacKinnon, Syed K. Tanbeer,(2014) "Tightening upper bounds to the expected support for uncertain frequent pattern mining" . In this paper, they have proposed (i) a compact tree structure for capturing uncertain data, (ii) a technique for using tree structure to tighten upper bounds to expected support, and (iii) an algorithm for mining frequent patterns based on tightenedbounds. In this paper they find out the key solutions such as: - (i) to make the resulting tree as compact as the FP-tree which is discussed in J. Han, J. Pei and Y. Yin, (2000)(ii) to mine frequent patterns from such a tree. And the solutions they derived as

1. The concept of the tightened prefixed pattern cap (TPC);
2. A tightened prefixed-capped uncertain frequent pattern tree (TPC-tree) structure, which can be as compact as the original FP-tree while capturing uncertain data;
3. A tightened prefixed-capped uncertain frequent pattern-growth mining algorithm—called TPC-growth—which is guaranteed to mine all and only those frequent patterns (i.e., no false negatives and no false positives) from uncertain data.

They have compared the performance of three Existing tree-based frequent pattern mining algorithms such as UF-growth which is proposed by Leung CK, Mateo MAF, BrajczukDA,(2008) , UFP-growth which was discussed by Aggarwal CC, Li Y, Wang J, Wang J ,(2009) and PUF-growthas discussed byTong Y, Chen L, Cheng Y, Yu PS ,(2012) and formulated a new algorithm as TPC- Tree.

Inferences :(i) Scalable with respect to the number of transactions
(ii) Can mine high volumes of uncertain data within a reasonable amount of time.

Drawback : The TPC-growth scans the uncertain data a third time to compute the true expected support and to eliminate the small number of false positives. And it takes up more (e.g. 50%) space than the existing PUF-growth algorithm as shown in results.

Paper 6: LeWanga,LinFenga, Mingfei Wu, (2013) proposed an array-based tail node tree structure (namely AT-Tree) to maintain transaction itemsets, and a pattern-growth based algorithm named AT-Mine for FIM over uncertain dataset. AT-Tree is created by two scans of dataset and it is as compact as the original P-Tree. AT-Mine mines frequent itemsets from AT-Tree without additional scan of dataset. The sparse and dense datasets were used to prove the performance of the algorithm.

(i) A new tree structure named AT-Tree (Array based Tail node Tree) for maintaining important information related to an uncertain transaction dataset.

(ii) An algorithm named AT-Mine for FIM over uncertain transaction datasets based on AT-Tree.

(iii) Both sparse and dense datasets are used in the experiments to compare the performance of the Le Wanga's proposed algorithm against the state-of-the-art Algorithms based on level-wise approach and pattern-growth approach. The algorithm MBP is the state-of-the-art algorithm employing the level-wise approach, UP Growt has discussed from C.C. Aggarwal, Y. Li, J. Wang, and J. Wang, (2009) is the state-of-the-art algorithm employing the pattern-growth approach and CUFP-Mine as discussed by Le Wanga,LinFenga, Mingfei Wu, (2013) is a new proposed algorithm. They have compared AT-Mine with the algorithms UF-Growth, CUFP-Mine and MBP on both types of datasets: the sparse transaction datasets and dense transaction datasets.

Inferences :(i)AT-Mine has a better performance than UF-Growth, CUFP,
(ii)Number of candidate itemsets generated by MBP algorithms in terms of memory usage.

Drawback: It requires two sets of scanning process of datasets.

The number of transactions (D) tree nodes (I) used in the two datasets(sparse, dense) having varied count.

Proposed Work :

The mining algorithms for frequent itemsets within the stream's constrained environment having limited time and memory. They are not capable of handling varying inter-arrival rates of streams and instant results. If there is a need of accuracy then it will results in increase in time allocated.ImplementingAnyFI, anytimefrequent itemset mining algorithm for data streams. And also propose a novel data structure, BFI-forest (Buffered Frequent Itemset) is proficient of treating transactions with varying inter-arrival rate. AnyFI maintains itemsets in BFI-forest in such a way that it can give a mining result almost immediately when time allowance to mine is very less and can refine the results for better accuracy with increase in time allowance.

In some cases while mining frequent item sets having dense type database may results in redundant of information. To overcome this problem a weighted frequent itemsets mining algorithm is applied in this proposal. This algorithm concentrates on two measures support and weight, together they are applied at the same time.

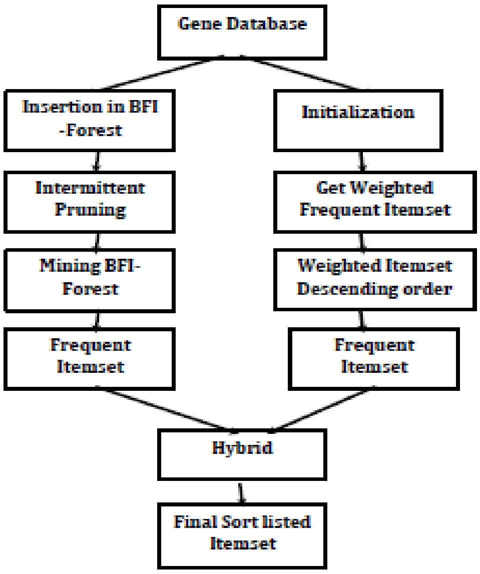

Fig. 1 Implementation flow

To reduce the number of candidate itemsets the methods used are pruning and hash table. The pruning uses property weighted effectively extension. The hash table is used to store weighted non frequent binomial set.

	Runtime			**Pattern Analysis**			**No. of K-itemsets**			**Memory consumption(0.002) thershold(MB)**		
Algorithm	*Mushroom*	*Retail*	*T1014D*	*Mushroom*	*Retail*	*T1014D*	*Mushroom*	*Retail*	*T1014D*	*Mushroom*	*Retail*	*T1014D*
Uapriori	240	3500	9000	14000	1400	14000	6000	1000	9500	60	900	300
HEWI-Uapriori	150	4500	7000	5000	500	6000	2500	400	4000	95	1400	450
WD-FIM	265	5500	1500	5000	500	6000	2500	400	4000	80	200	500

Table 1.1 The performance of the WD-FIM algorithm is analyzed

Memory consumption(0.01)threshold(MB)		
Mushroom	Retail	T1014D
40	400	100
60	700	300
45	1500	400

Table 1.2 The Memory consumption

To find the interest pattern in dataset in order to find its importance, in case if it is an uncertain database then in a weighted frequent itemsets mining the probability of both existential and importance are considered into account. In WFI when using downward closure property is applied which leads to poor time efficiency. So along with the downward closure, the existence property is also applied. These two properties are applied in order to narrow the searching space of the WFI and to improve the time efficiency.

Based on the BFI-forest (Buffered Frequent Itemset) and weighted frequent itemsets mining algorithm, a hybrid Frequent Pattern Mining algorithm is derived in order to increase the efficiency and performance in Frequent Itemset with constrained streams of limited time and memory.

V. Conclusion

The main purpose of this review is to present the features of the existing frequent pattern mining along with the significant advantages and issues. In this paper few popular algorithms for frequent pattern mining were reviewed, all of them were based on U-Aprior or tree structure UF-growth or WD-FIM or AnyFI or BFI or SUF -growth or improvement of them. Especially, during the literature review, the main focus is given to AnyFIAnytime Frequent itemset mining algorithms for data streams. Accuracy and relevant to time are the two metrics focused in this review. To handle varying inter-arrival rate, a novel data structure, BFI-forest (Buffered Frequent Itemset Forest) were reviewed. WD-FIM algorithms which were used of mining frequent pattern form uncertain data. The efficiency of US-streaming and SUF-growth algorithm was checked by setting the Minimum support threshold and size of database respectively. From this review, various drawbacks has been analyzed and observed to find better solutions for future.

VI. Reference

[1] PoonamGoyal, JagatseshChalla, ShivinShrivastava, NavneetGoyal, Department of Computer Science, Pilani Campus, Brila Institution of Technology & Science, Pilani, "AnyFI : An Anytime Frequent Itemset Mining Algorithm for Data Streams" Published in : IEEE International Conference on Big Data(BIGDATA), 2017 IEEE Conference.

[2] Xuejian Zhao, Xinhui Zhang, Pan Wang, Pan Wang, Song Chen and Zhixin sun, "A Weighted Frequent Itemset Mining Algorithm for Intelligent Decision in Smart Systems" Published in : IEEE Access, Vol 6, pp. 29271 - 29282, May 2018.

[3] LiwenYue, "Review of Algorithm for Mining Frequent Patterns from Uncertain Data ", Published in : IJCSNS International Journal of Computer Science and Network Security, VOL.15 No.6, pp. -21, June 2015.

[4] R. Agrawal& R. Srikant, Fast algorithms for mining association rules, in VLDB 1994, pp. 487–499.

[5] J. Han, J. Pei & Y. Yin, Mining frequent patterns without candidate generation, in ACM SIGMOD 2000, pp. 1–12.

[6] B. Chikhaoui, S. Wang, & H. Pigot, A frequent pattern mining approach for ADLs ecognition in smart environments, in IEEE AINA 2011, pp. 248–255.

[7] C.-K. Chui, B. Kao & E. Hung, Mining frequent itemsets from uncertain data, in PAKDD 2007, pp. 47–58.

[8] Chui C K, Kao B.A decremental approach for mining frequent itemsets from uncertain data[C] LNAI 5012：PAKDD，2008：64-75.

[9] Leung C K S，Mateo M AF，Brajczuk D A.A tree-based approach for frequent pattern mining from uncertain data[C] LNAI 5012：PAKDD，2008：653-661.

[10] Leung C K S，Carmichael C L，HaoB.Efficient mining of frequent patterns from uncertain data[C] Proc IEEE ICDM Workshops,2007：489-494.

[11] Leung C K S，BrajczukAD.Efficient mining of frequent itemsets from data streams[C] LNCS 5071：BNCOD，2008：2-14.

[12] Carson K. Leung∗ , Richard Kyle MacKinnon, Syed K. Tanbeer, Department of Computer Science, University of Manitoba, Winnipeg, MB, R3T 2N2, Canada, "Tightening upper bounds to the expected support for uncertain frequent pattern mining ", Published in :ScienceDirectProcedia Computer Science 35 ,pp. -328 – 337, 2014.

[13] Aggarwal CC, Li Y, Wang J, Wang J. Frequent pattern mining with uncertain data. In: Proceedings of the ACM KDD 2009. ACM; 2009, p. 29–37.

[14] Tong Y, Chen L, Cheng Y, Yu PS. Mining frequent itemsets over uncertain databases. PVLDB 2012; 5(11):1650–1661.

[15] Leung CK, Mateo MAF, Brajczuk DA. A tree-based approach for frequent pattern mining from uncertain data. In: Proceedings of the PAKDD 2008. Springer; 2008, p. 653–661.

[16] Le Wanga,LinFenga, MingfeiWu,"AT-Mine: An Efficient Algorithm of Frequent Itemset Mining on Uncertain Dataset", Published in : Journal of Computers, VOL. 8, NO. 6, pp. - 1417 -1426, JUNE 2013.

[17] L. Wang, D.W. Cheung, R. Cheng, S. Lee, and X. Yang, "Efficient Mining of Frequent Itemsets on Large Uncertain Databases," IEEE Transactions on Knowledge and Data Engineering, no.99(PrePrints), 2011.

[18] X. Sun, L. Lim and S. Wang, "An approximation algorithm of mining frequent itemsets from uncertain dataset," International Journal of Advancements in Computing Technology, Vol.4, no.3, pp.42-49, 2012.

[19] C.C. Aggarwal, Y. Li, J. Wang, and J. Wang, Frequent pattern mining with uncertain data, in 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09). 2009, pp.29-37.

[20] J. Han, J. Pei and Y. Yin, Mining frequent patterns without candidate generation, in ACM SIGMOD International Conference on Management of Data. 2000, pp.1-12.

[21] C.K. Leung, M.A.F. Mateo and D.A. Brajczuk, A treebased approach for frequent pattern mining from uncertain data, in 12th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2008). 2008, pp.653- 661.

[22] C.K. Leung, C.L. Carmichael and B. Hao, Efficient mining of frequent patterns from uncertain data, in International Conference on Data Mining Workshops (ICDM Workshops 2007). 2007, pp.489-494.

[23] G. Lee, U. Yun, H. Ryang," An Uncertainty-based Approach: Frequent Itemset Mining from Uncertain Data with Different Item Importance", published in: Knowledge-Based Systems, (2015)

[24] U. Yun, "On pushing weight constraints deeply into frequent itemset mining", Intelligent Data Analysis, vol. 13, no. 2, pp. 359-383, 2009.

[25] U. Yun, G. Lee, and K.H. Ryu, "Mining maximal frequent patterns by considering weight conditions over data streams", Knowledge Based Systems, vol. 55, pp. 49-65, 2014.

[26] G. Lee, U. Yun, and K. H. Ryu, "Sliding Window based Weighted Maximal Frequent Pattern Mining over Data Streams", Expert Systems with Applications, vol. 41, no. 2, pp. 694-708, 2014.

[27] U. Yun and K.H. Ryu, "Efficient Mining of Maximal Correlated Weight Frequent Patterns", Intelligent Data Analysis, vol. 17, no. 5, 2013.

[28] G. Lee, U. Yun, and H. Ryang, "Mining Weighted Erasable Patterns by using Underestimated Constraint-based Pruning Technique", Journal of Intelligent and Fuzzy Systems, vol. 28, no. 3, pp. 1145-1157, 2015.

[29] C. Lin, and T. Hong, “A new mining approach for uncertain databases using CUFP trees”, Expert Systems with Applications, vol. 39, no. 4, pp. 4084-4093, 2012.
[30] ArunnyaRadhakrishnan1 ,Vijayakumar R2,”An efficient approach for mining uncertain frequent patterns using dynamic data structure without false positives”, Published in :IPASJ International Journal of Computer Science (IIJCS), VOL. 6,Issue 6, NO. 6, JUNE 2018.
[31] Gangin Lee, Unil Yun, (2016) “A new efficient approach for mining uncertain frequent patterns using minimum data structure without false positives”, Wiley Publishing, Incorporated-India, pp. 89-110.
[32] GanginLee ,Unil Yun , Keun Ho Ryu ,”Sliding window based weighted maximal frequent pattern mining over data streams “, Published in :Expert Systems with Applications, VOL.41, pp. 694–708, 2014.
[33] Han, J., Pei, J., Yin, Y., & Mao, R. (2004). Mining frequent patterns without candidate generation: A frequent-pattern tree approach. Data Mining and Knowledge Discovery, 8(1), 53–87.
[34] Ahmed, C. F., Tanbeer, S. K., Jeong, B. S., Lee, Y. K., & Choi, H. J. (2012). Single-pass incremental and interactive mining for weighted frequent patterns. Expert Systems with Applications, 39(9), 7976–7994.
[35] Grahne, G., & Zhu, J. (2005). Fast algorithms for frequent itemset mining using FPtrees. IEEE Transactions on Knowledge and Data Engineering, 17(10), 1347–1362.
[36] Tanbeer, S. K., Ahmed, C. F., Jeong, B. S., & Lee, Y. K. (2009b). Sliding window-based frequent pattern mining over data streams. Information Sciences, 179(22), 3843–3865.
[37] UnilYun ,Gangin Lee , Keun Ho Ryu, “Mining maximal frequent patterns by considering weight conditions over data streams”, Published in :Knowledge-Based Systems, VOL.Mode 5g, 2013.
[38] Jerry Chun-Wei Lin,Wensheng Gan1 ,Philippe Fournier-Viger,Tzung-Pei Hong,,VincentS,Tseng “Weighted frequent itemset mining over uncertain databases”, Published in :Springer Science+Business Media New York,2015.
[39] Amit Mittal1, Ashutosh Nagar, Kartik Gupta, Rishi Nahar, “Comparative Study of Various Frequent Pattern Mining Algorithms”,Publishedin:International Journal of Advanced Research in Computer and Communication Engineering,Vol. 4, Issue 4,pp.-550-553, April 2015.
[40] Lijuan Zhou, Xiang Wang, “Research of the FP-Growth Algorithm Based on Cloud Environments”, Published in:JOURNAL OF SOFTWARE, VOL. 9, NO. 3, pp.-676-683, MARCH 2014
[41] Le Zhou, ZhiyongZhong, Jin Chang, Junjie Li, Huang, J.Z., ShengzhongFeng, “Balanced parallel FP-Growth with MapReduce”, Information Computing and Telecommunications (YC-ICT), 2010 IEEE Youth Conference on, Page(s): 243 -246 ,2010.
[42] Shulan Zhao, “Typical Hadoop cloud computing”,Electronic Industry Press, Beijing, 2013.
[43] Bay Vo ,Tuong Le , FransCoenen,Tzung-Pei Hong, “Mining frequent itemsets using the N-list and subsume concepts”,Published in: Springer-Verlag Berlin Heidelberg, pp.-253–265, April 2014.
[44] Zaki MJ, Hsiao CJ (2005) Efficient algorithms for mining closed itemsets and their lattice structure. IEEE Trans Knowl Data Eng 17(4):462–478
[45] Vo B, Le T, Coenen F, Hong TP (2013) A hybrid approach for mining frequent itemsets. In: Proceeding of the IEEE SMC’13, pp 4647–4651
[46] Dr. KanakSaxena, D.S. Rajpoot, “A Way to Understand Various Patterns of Data Mining Techniques for Selected Domains”,(IJCSIS) International Journal of Computer Science and Information Security, Vol. 6, No. 1, pp. - 186-191, 2009.
[47] Dr. HuiXiong. Rutgers University, Association Analysis: Basic Concepts and Algorithms. Chapter 6, 2006.