



REVIEW ON FREQUENT PATTERN MINING

¹M. PUSHPALATHA, ²Dr. A. VIJAYA

¹Assistant Professor, ²Assistant Professor and Head,

¹Department of Computer Science, ²Department of Computer Application,

¹Padmavani Arts & Science College for Women, ²Government Arts College (Autonomous)

^{1,2}Salem – 636 007.

ABSTRACT: Conceptual Frequent pattern mining is a basic data mining task, with an objective of finding information as rehashed patterns. The idea of frequent pattern mining is stretched out to dynamic database mining and the data streams today. The investigation of existing algorithms is expected to plan the efficient algorithms and data structure in the idea of frequent pattern mining is significant for the scientists. Henceforth, in this paper we checked on the idea of frequent pattern mining and review of different algorithms for the extraction of frequent patterns dependent on data processing models.

Keywords: [Frequent pattern mining, dynamic database, data processing.]

1. INTRODUCTION

Frequent pattern mining has been a significant topic in data mining from numerous years. An amazing advancement in this field has been made and bunches of efficient algorithms have been intended to look through frequent patterns in a transactional database. Data Mining has pulled in a lot of consideration in the data business and in the public eye overall lately, because of the wide accessibility of tremendous measures of data and the approaching requirement for transforming such data into helpful data and information. Association rules were introduced by R. Agrawal and others in 1993, it is a significant exploration issue in data mining. Mining association rules targets finding the correlation between the various things in a database. It tends to be utilized to discover the buy patterns of clients, for example, how the exchange of getting a few merchandise will

affect on the exchange of purchasing others. Association rules are made out of the accompanying two stages:

1. Locate the huge thing sets that have exchange uphold over a base help by utilizing frequent itemset mining algorithms.
2. From the found enormous itemsets, produce the ideal association rules dependent on some standard measures.

Frequent patterns are those things, groupings or foundations that repeat in database exchanges with a client indicated recurrence. An item set with recurrence more noteworthy than or equivalent to least limit will be considered as a frequent pattern. For instance in market based examination if the base limit is 30% and bread shows up with eggs and milk multiple occasions or possibly multiple times then it will be a frequent item set. Frequent pattern mining can be utilized in an assortment of certifiable applications. It very

well may be utilized in general stores for selling, item arrangement on racks, for advancement rules and in text looking. It tends to be utilized in wireless sensor networks particularly in keen homes with sensors connected on Human Body or home use objects and different applications that require checking of client condition cautiously that are dependent upon basic conditions or risks, for example, gas break, fire and explosion. These frequent patterns can be utilized to screen the exercises for dementia patients. It very well may be viewed as a significant methodology with the capacity to screen exercises of everyday life in brilliant condition for following practical decrease among dementia patients.

2. CHALLENGING ISSUES

The issues identified with ease, flexibility and reusability for producing association rules. Writing additionally uncovers that more investigations have been done on frequent pattern generation. As referenced before, distinguishing frequent itemsets is computationally costly process. Checking the events of itemsets requires a lot of processing time. As an outcome, quantities of efficient algorithms are proposed in writing for mining the frequent itemsets. It is seen that, the majority of the algorithms for finding frequent patterns accessible in the writing require numerous disregards the database bringing about countless disk reads and putting a gigantic weight on the I/O subsystem. This requires the presentation of mining algorithms that offers single database examine.

Although, different algorithms are accessible that can help uncover patterns and connections, it doesn't tell the client the worth or centrality of these patterns.

Most of the algorithms accessible in the writing don't offer flexibility for testing the legitimacy of Meta rules.

Algorithms are accessible for keeping up the association rules because of expansion or erasure of exchanges in the database. Nonetheless, algorithms are not accessible for

mining gradual rules because of expansion of more things.

There is a prerequisite for the improvement of equal and additionally dispersed algorithms so as to accelerate the calculation movement as distinguishing proof of frequent itemsets is computationally costly task. This will bring about generally speaking execution improvement.

Most of the algorithms accessible in the writing for mining frequent itemsets don't offer flexibility for reusing the calculation during mining process.

Much research is as yet expected to considerably lessen the size of determined patterns and enhance the nature of held patterns (conservative excellent pattern set).

Some genuine restrictions of momentum association rule mining algorithms and a few related issues spur preceded with studies and examination in to this zone. One of the solid persuading factors for the exploration here is to enhance ease, flexibility, efficiency and reusability during mining process.

3. LITERATURE REVIEW

[1]. **S. Vijayalakshmi, V.Mohan, S.Suresh Raja** proposed another frequent sequence pattern strategy called AWAPT (Adaptive Web Access Pattern Tree), for FSP mining. An AWAPT joins for efficient capacity of the apparent multitude of sequences that contain a given item. It disposes of recursive reproduction of transitional WAP tree during the mining by allotting the double codes to every hub in the WAP Tree. Web access pattern tree (WAP-tree) mining is a sequential pattern mining strategy for web log access sequences, which first stores the first web access sequence database (WASD) on a prefix tree, like the frequent pattern tree (FP-tree) for putting away non-sequential data. WAP-tree algorithm at that point, mines the frequent sequences from the WAP-tree by recursively re-developing transitional trees, beginning with suffix sequences and finishing with prefix sequences. An endeavor has been made to AWAPT approach for improving

efficiency. AWAPT absolutely kills the need to participate in various reproductions of middle of the road WAP-trees during mining and impressively decreases execution time.

[2]. **D.Vasumathi, A .Govardhan** explore another sequence pattern procedure called BC-WAPT (Binary coded Web Access pattern Tree).it dispenses with recursive reproduction of halfway WAP tree during the mining by allotting the twofold codes to every hub in the WAPTtree. Sequential Pattern mining is the process of applying data mining strategies to a sequential database for the reasons for finding the correlation connections that exist among an ordered rundown of occasions. Web access pattern tree (WAPTtree) mining is a sequential pattern mining strategy for web log access sequences, which first stores the first web access sequence database on a prefix tree, like the frequent pattern tree (FP-tree) for putting away non-sequential data. WAP-tree algorithm at that point, mines the frequent sequences from the WAPTtree by recursively re-building transitional trees, beginning with suffix sequences and finishing with prefix sequences. An endeavor has been made to adjust WAP tree approach for improving efficiency. BCWAPT thoroughly kills the need to take part in various recreation of middle of the road WAP-trees during mining and impressively lessens execution time.

[3]. **C. Gomathi, M. Moorthi, Dr. K. Duraiswamy** propose an efficient sequential pattern mining strategies for KASC web log access sequences known as CS-WAP Tree. This proposed algorithm adjusts the WAP tree approach for improving efficiency. The proposed algorithm thoroughly disposes of the need to participate in various reproductions of halfway WAP trees and impressive decreases execution time. The aftereffects of trials show the efficiency of the improved algorithm. The following key point is to think about WAP algorithms. This algorithm utilizes the pre-order linking of header nodes to store all occasions in a similar suffix tree intently

together in the linkage, making the search process more efficient.

[4]. **AshinAraBithi , Abu Ahmed Ferdaus** proposed an efficient Sequential Pattern Tree Mining algorithm which can create the total arrangement of frequent sequential patterns from a Sequential Pattern Tree without producing any up-and-comer sequence and any middle extended tree that decrease the both existence unpredictability. It doesn't produce extended trees during mining by finding the first occurrence nodes from the suffix trees of prefix sub sequences. Likewise, it lessens the exertion of continued examining of database because of capacity of include in the tree's hub that help us to enhance the presentation of our algorithm. This methodology first produces a Sequential Pattern Tree from the sequence database which stores both frequent and non-frequent items. So that, it requires just one sweep of sequence database to make the tree alongside Header Table which likewise lessens the tree development time extensively. It decreases the utilization of memory by putting away just basic data in the tree and by not creating extended trees during mining.

[5]. **Jay Ayres, Johannes Gehrke, TomiYiu, and Jason Flannick** present another algorithm for mining sequential patterns. Algorithm is particularly efficient when the sequential patterns in the database are long. They present a novel depth-first search strategy that incorporates a depth-first traversal of the search space with compelling pruning components. Their execution of the search strategy consolidates a vertical bitmap representation of the database with efficient help checking. A notable element of algorithm is that it gradually yields new frequent item sets in an online manner.

[6]. **JOTUN HEIN** presented O(lk) algorithm is conceivable by utilizing the mathematical tails of the p-functions to make an algorithm, that lone includes 0 or 1 character of the leaf sequences all at once. Doing this in the overall

case is marginally more specialized and won't be sought after here. The present strategy might be made to work for four to seven sequences by various techniques for corner cutting permitting the measure of calculation to be decreased. For bigger number of sequences it appears to be that markov chain monte carlo (MCMC) techniques would be the undeniable decision.

[7]. **Yu Hirate, Hayato Yamana** proposed a novel algorithm for summing up sequential pattern mining with time intervals. The speculation incorporates three focuses; (an) a capacity to deal with two sorts of item interval estimation, item hole and time interval, (b) an ability to deal with broadened sequences which are characterized by embeddings pseudo items dependent on the interval itemization function, and (c) receiving four item interval requirements. Summed up sequential pattern mining can substitute a wide range of customary sequential pattern mining algorithms with item intervals. Summed up sequential pattern mining with time interval can extricate interval expanded sequences that incorporate time interval with variable division size. What's more, by adjusting our sorts of limitation identified with time intervals, it additionally avoids extraction of interval expanded sequences with time intervals in which the client isn't intrigued.

[8] **Bin Gu and Victor S. Sheng** proposed another GSP algorithm. That the GSP is significant from a hypothetical and useful perspective. First, the GSP gives a bound together and vigorous execution for a broad number of PQP problems, where an enormous number of PQP problems actually have no solution path algorithms. Second and all the more critically, the GSP is anything but difficult to be utilized by researchers or clients as a result of its bound together framework. They dissected the limited union and the time multifaceted nature of the GSP. The exploratory outcomes on an assortment of data sets not just affirm the identity between the

GSP and a few existing solution path algorithms, yet additionally show the predominance of our GSP over the current solution path algorithms on both speculation and power. At last, a commonsense organization of utilizing the GSP to take care of two significant learning problems, i.e., summed up error path and Ivanov SVM.

[9] **JU-DONG REN, YIN-BO CHENG, LUNG-LUNG YANG** propose another enhanced strategy dependent on Prefixspan, called EPSpan. EPSpan mines summed up sequential patterns with time constraints, time windows or potentially scientific categorization efficiently. Its overall thought is to locate all biggest sequences which every data-sequence support and eliminate sliding windows, time constraints. It is fascinating to mine sequential patterns with other more muddled constraints and improve the exhibition of sequential patterns mining.

[10] **Wen-Chung Shih** dissected partaking understudies' activity log data in the App. They proposed to utilize a blockly visual programming instructive App for college understudies to examine their learning practices. Visual programming environments joined with problem-solving are valuable for advancing understudies' inspiration of learning programming. The motivation behind this work is to utilize a blockly visual programming instructive application for college understudies and break down their learning behaviour patterns. The investigation utilized a sequential pattern mining tool to mine the understudies' standards of conduct. The preliminary outcomes demonstrated that the proposed strategy can discover personal conduct standards of high-accomplishment understudies and low accomplishment understudies.

[11] **A. Sirisha, Suresh Pabboju and G. Narsimhasaid** the significant disadvantages of the customary methodologies for mining sequential patterns is that all components in a

sequence are dealt with consistently, while progressively various items have various significances and they create countless patterns with low least support. They proposed a way to deal with mine additionally intriguing sequential patterns by considering the time interval between the components in the sequences. Algorithm TIWSP centers around mining the time interval weighted frequent patterns dependent on the prefix extended sequential pattern development approach. The analysis results have demonstrated that it is more efficient than Prefix length algorithm in creating fascinating pattern.

[12] **Amik Singh, Mohit Gupta and Manoj Misra** propose a novel Parallel Progressive Sequential Pattern (PPSP) mining algorithm to address the adaptability problem of the progressive sequential pattern mining. At each point in time the up-and-comer sequential patterns of all sequences are figured on the host CPU. At that point, utilizing all up-and-comer sequential patterns as the information data, the portion, which runs on GPUs, collects the occurrence frequencies of competitor sequential patterns in the current POI and reports frequent. sequential patterns to clients. The proposed algorithm efficiently utilizes both the CPU and the GPU to circulate remaining task at hand to accomplish profoundly upgraded and versatile solution to progressive sequential pattern mining problem.

[13] **YalingZhang, Ting Wang, Shangping Wang PBEK** is proposed to take care of the problem of efficiency decrease of the privacy-preserving data mining algorithms causes by including privacy-protection mechanism. PBEK presents parcel and the earlier imperative dependent on likelihood. At last, the aftereffects of the investigation show that the efficiency of PBEK is higher than that of BEMASK for the situation that the support degree error isn't extended, and the impact of

the quantity of division on the efficiency of the algorithm is talked about.

[14] **Aditya S P, Hemanth M, Lakshmikanth C K and Suneetha K R** said Apriori algorithm is a persuasive data mining algorithm which produces a rundown of most frequent web pages visited. Because of quick changing substance of database one needs an algorithm which is continuous. The significant downside of Apriori algorithm is that, it needs to check primary database every single time to produce frequent patterns which brings about more utilization of memory and execution time, thus so as to lessen execution time and use of memory a ton of research has been led to improve Apriori Algorithm. Towards improving Apriori, an adjusted rendition is proposed in this paper to create frequent patterns. This empowers discovering patterns as opposed to returning to the database at each pass. This restricts the quantity of sweep and furthermore the quantity of complete blends is brought down from $2N$ to $2N-2$. This impressively lessens memory utilization just as execution speed and makes ongoing pattern disclosure conceivable.

[15] **Avinash Kadimisetty, C. Oswald and B. Sivaselvan** presented a novel, efficient and straightforward sequence mining based lossy picture compression algorithm for pictures. They have demonstrated utilizing the trial results that our technique is acceptable in mining the ideal pixel sequences it actually beats its rivals in compression proportion with a satisfactory visual quality. And, after its all said and done, their compression algorithm is considered as a model that requirements further improvement. They plan to outfit proposed algorithm with better bunching strategy to improve the nature of pictures utilizing the properties of pixel neighborhood. To consider mining sub networks rather than sub sequences that would give a huge improvement in the compression proportion. Expansion to that, they would expect to explore with broad reenactments on huge

picture size with more standard datasets, including the time to mine the sequence, which is by all accounts extremely encouraging.

CONCLUSION

Frequent pattern mining is a significant task in association rule mining. It has been discovered helpful in numerous applications like market basket analysis, financial forecasting and so forth. This paper presents the audit on different research papers pertaining to utilizations of frequent patterns mining and association rule mining. It gives information about different frequent pattern mining algorithm and augmentations of the equivalent. It likewise clarifies about the distinctive application zones where these frequent patterns can be utilized. We sure that this paper will support the researches and data excavators to acquire information and uncovers the benefits of applying frequent pattern mining algorithm alongside rule mining in different fields.

REFERENCES

- [1]. S.Vijayalakshmi, V.Mohan, S.Suresh, "MINING OF USERS' ACCESS BEHAVIOUR FOR FREQUENT SEQUENTIAL PATTERN FROM WEB LOGS", International Journal of Database Management Systems (IJDMS) Vol.2, No.3, August 2010, DOI: 10.5121/ijdms.2010.2304.
- [2]. D.Vasumathi, A .Govardhan, "BC-WASPT: Web Access Sequential Pattern Tree Mining", IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.6, June 2009.
- [3]. C. Gomathi, M. Moorthi, Dr. K. Duraiswamy, "Web Access Pattern Algorithms in Education Domain", Computer and Information Novel, VOL.1 No.4, November 2008, www.ccsenet.org/journal.html
- [4]. AshinAraBithi , Abu Ahmed Ferdaus, "Sequential Pattern Tree Mining", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 15, Issue 5 (Nov. - Dec. 2013), PP 79-89 www.iosrjournals.org
- [5]. Jay Ayres, Johannes Gehrke, TomiYiu, and Jason Flannick, "Sequential PATTERN Mining using A Bitmap Representation", NSF grants IIS-0084762 and IIS-0121175, by a gift from Intel Corporation, and by the Cornell Intelligent Information Systems Institute.
- [6]. JOTUN HEIN, "AN ALGORITHM FOR STATISTICAL ALIGNMENT OF SEQUENCES RELATED BY A BINARY TREE", Pacific Symposium on Biocomputing 6:179-190 (2001).
- [7]. Yu Hirate, HayatoYamana , "Generalized Sequential Pattern Mining with Item Intervals", JOURNAL OF COMPUTERS, VOL. 1, NO. 3, JUNE 2006.
- [8]. Bin Gu and Victor S. Sheng, "A Solution Path Algorithm for General Parametric Quadratic Programming Problem", **Published in: IEEE Transactions on Neural Networks and Learning Systems (Volume: PP, Issue: 99)**,
- [9]. JU-DONG REN, YIN-BO CHENG, LUNG-LUNG YANG **Published in: Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826), Date Added to IEEE Xplore: 24 January 2005, Print ISBN: 0-7803-8403-2**
- [10]. Wen-Chung Shih, "Mining Sequential Patterns to Explore Users' Learning Behavior in a Visual Programming App", 2018 42nd IEEE International Conference on Computer Software & Applications, 0730-3157/18/\$31.00 ©2018 IEEE DOI 10.1109/COMPSAC.2018.10216
- [11]. A. Sirisha, Suresh Pabboju and G. Narsimha, "An approach to mine Time Interval based Weighted Sequential Patterns in Sequence Databases", **Published in: 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Electronic ISBN: 978-1-5386-4283-2 Print on Demand (PoD) ISBN: 978-1-5386-4284-9**
- [12]. Amik Singh, Mohit Gupta and Manoj Misra, "Parallel Progressive Sequential Pattern (PPSP) Mining", **Published in: 2018 IEEE 8th Annual Computing and**

Communication Workshop and Conference (CCWC), **Electronic ISBN:** 978-1-5386-4649-6, **Print on Demand (PoD) ISBN:** 978-1-5386-4650-2

[13]. YalingZhang,Ting Wang, Shangping Wang, “An Efficient Algorithm for Frequent Pattern Mining based on Privacy-preserving”, **Published in:** 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), **Publisher:** IEEE, **Electronic ISBN:** 978-1-5386-2165-3, **USB ISBN:** 978-1-5386-2164-6, **Print on Demand (PoD) ISBN:** 978-1-5386-2166-0

[14]. Aditya S P, Hemanth M, Lakshmikanth C K and Suneetha K R, “Effective Algorithm for Frequent Pattern Mining”, **Published in:** 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), **Publisher:** IEEE, **Electronic ISBN:** 978-1-5386-1887-5, **Print on Demand (PoD) ISBN:** 978-1-5386-1888-2

[15]. AvinashKadimisetty, C. Oswald and B. Sivaselvan, “Frequent Pattern Mining approach to Image Compression”, **Published in:** 2016 22nd Annual International Conference on Advanced Computing and Communication (ADCOM), **Publisher:** IEEE, **Electronic ISBN:** 978-1-5386-2389-3, **Print on Demand(PoD) ISBN:** 978-1-5386-2390-9.