International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-7 Issue-4S, November 2018

Periodicity Mining, "a Time Inference over High Utility Item set Mining" – A study

Arunkumar M. S, Suresh P, Gunavathi C, Preethi S.

Abstract: Inthe past few years, mining oftemporal frequent patterns from transactional database has gathered momentum. Numerous works and algorithms have been proposed for FIM [1,2,3,4], but the same models cannot be implemented to mine temporal patterns as none of the models are built to find patterns that consider periodicity of its occurrence in a database. The importance of an itemset really rests upon its utility rather than its participation count. Works over utility mining [5, 6 and 7] have gathered more momentum in this last decade and many research works have been carried out. In this paper, a survey is conducted on i) the works that led to periodic pattern mining, ii) the works over periodic pattern mining and iii) the extended and enhanced works of Periodic pattern mining.

Keyword: Temporal Mining, Utility Mining, Periodic Mining, High Utility Itemset mining, Sequential Pattern Mining.

I. INTRODUCTION

Data mining is a subject area aids to extract useful information from plenty of unprocessed data that are gathered for a certain time period. Frequent PatternMining is one such subfield of data mining that identifies the frequency of an itemset in transaction database. Even though copious works and algorithms are available for Frequent Itemset Mining FIM [1, 2, 3, 4], the main drawback of conventional FIM algorithms is that it doesn't find periodic patterns in a database. **A**nother important constraint is that the frequency metric alone cannot be considered as a useful measure of decision making, as it only reveals the count of transactions that contains a particular itemset[8].

The importance of an itemset really rests upon its utility rather than its participation count. Works over utility mining [5, 6 and 7] have gathered more momentum and many researchers are doing their research in the field of utility based mining. One such extended work is termed as High Utility Itemset (HUI) mining [9, 10, 11, 12 and 13]. In this paper, periodicity based mining framework, which contains time series mining and periodic pattern mining over transaction databases are discussed. Then, an overview of extensions of periodic mining models like sequential HUI Mining framework and periodic high utility itemsets are presented and further a detailed review about algorithms used to mine HUI are also presented.

Revised Version Manuscript Received on 25 November, 2018.

Arun Kumar M. S, Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Sathyamanglam (Tamil Nadu), India.

Suresh P, Department of Information Technology, Kongu Engineering College, Erode (Tamil Nadu), India.

Gunavathi C, School of Information Technology and Engineering, VIT University, Vellore (Tamil Nadu), India.

Preethi S, Infinijth Technologies Pvt Ltd, Gobichettipalayam (Tamil Nadu), India

II. TIME SERIES DATAMINING

The measurement of any analysis in the scientific field is carried out with respect to time. The identifications over a set of ordered data based on time is referred as time series.

The mining paradigm with time series tries to discover the essential meaningful knowledge from the gathered data in time bound manner. The sequential observations carried out with respect to time results in a set of values are referred as time series. The research contributions over the past one and a half decades in the area of time series was certainly overwhelming irrespective of the complex nature associated with it.

One of the earlier works, Agrawal et al [14] has introduced an index based technique to process similarity queries bound in a time series. They used a map structure based on Discrete Fourier Transform (DFT) to connect the frequency domain and the time sequences. They also observe that only the first set of frequencies is strongly mapped in the sequence and also measured the distance relation in terms of Fourier transforms in time or frequency domain.

Berndt et al. [15] carried out works with a dynamic programming approach to the time series based mining problem. They tried to detect double top peak patterns and express the same in higher-level relations. They proposed Dynamic Time Wrapping (DTW) algorithm based on independent time series mapped segments with a template. It then segments the time series based on the template and applies the parallel matching process. They also tried to control the computations based on time series and then distributed it along multiple processors.

Chan Kin et al. [16] tried to capitalize on the advantages of Discrete Wavelet Transform (DWT), as the DWT helps to represent the signal in form of multi-resolution. This multi-resolution attribute helps in representing the locations in terms of both frequency and time. This certainly tends to provide more advantage by bearing more information than Discrete Fourier Transform (DFT). While the DFT just denotes the time sequence extracted from lower harmonics, the DWT encapsulates the original time sequence with the respective coefficients. Through the experiments, they proved that the Haar transformed domain can be used to preserve the Euclidean distance and also it outperformed the conventional DFT conveniently. They also proposed a model to handle vertical shift of time series problem and the model is based on the wavelet method that uses the query range and modifies the range using the Euclidean distance preservation property. Yi et al.

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication

[17] in their work studied the issue of performance reduction due to the “dimensionality curse” phenomenon. The issue generally arises because of certain limitation that was prevalent on earlier works. The limitations include i) multi-modality support and ii) Feature Extraction. They proposed a segment based feature extraction model by segmenting the time sequence in to fixed size. These segments are used to form feature vector. The mathematical influence of the model helps in decreasing the search space considerably without compensating the accuracy of the results.

Yongwei Ding et al [18] in their work proved the importance of mining data that occur recurrently over certain time period. The importance of the time series data largely rests on its nature of easy understanding and easy implementation. The approach uses indexing method based on piece liner representation then it compresses and measures the time series. Further, they proposed a novel segmentation method based on Piece-wise Linear Representation (PLR) for slicing the time series in the data mining. They tried to implement their model in the real stock market and the results showed better efficiency.

Yin-Fu Huang & Po-LunLiou [19] has proposed a technique that incorporates the time series data in a simple organized collection of XML document. The XML document has a complexity due to the tree-structure representation, as the clustering of XML and tree-structured documents based on time series is always a challenge. They tried to overcome this complexity by mapping Self-Organized Map (SOM) with Jaccard coefficients to match the XML documents. They provided a sequential mining method to mine maximum frequent sequences. Using these sequences they constructed the frequent structures in the cluster.

Liping Liu &Ninghai Cui [20] came up with a method of mining association rules with time series. The time series was first subjected to anti-season pretreatment and the algorithm was used to accomplish the pattern-sequence that rest on the treated time series. This model minimized the effort on mining. The approach was well suited for higher precision models where the changes are carried out in anti-season using moving average method.

Uday et al. [21] classified periodic patterns based on the properties of time series as regular patterns and recurring patterns. Discovering recurring patterns is a significant application due to i) its temporal inclination, ii) its non anti-monotonic nature, and iii) its importance in real-life events. The authors tried to provide a model to mine recurring patterns by considering all the above mentioned facts. They tried to revamp the transactional database into transactions ordered according to a particular timestamp, which is arrived by taking into account the time series as a time-based sequence. Periodic-support, periodic-interval, and recurrence are the three novel procedures proposed to measure the dynamic periodic behavior of recurring patterns. Further, they proposed an efficient pruning strategy based pattern growth algorithm (RP-growth) to mine the recurring patterns. The result showed that the algorithm mines the recurring patterns on data obtained from real-life applications.

A. Periodicity Mining Over Time Sequence Data

Periodicity mining on time sequence data is an instrument that supports in identifying the behavior of time series data. In this section, few works over time series periodicity mining is discussed.

C.Faloutsos et al. [22] had contributed to one of the earlier works in mining in time series data, as they tried to discover subsequences within a collection of sequences. The concept they devised was to associate every subsequence into a multidimensional rectangles in feature space. Using R* tree spatial access model they indexed the rectangles, further by employing sliding window to extract the features as trails from data sequence. The trials were divided into sub trials using a proposed efficient algorithm and the sub trails are presented as Minimum Bound Rectangles (MBR). The model was examined in both synthetic and real data. The model outperformed sequential scanning, the only competitor at that period of time.

Indyk et al. [23] in their work formalized various problems pertaining to the time series identification as different representative trends. Based on certain properties, they categorized an interval of observation as a representative trend in a time series. Further, they devised an algorithm to analyze large time series datasets for representative trends; they tried to prove their algorithm’s efficiency against data processing and handling. They employed dimensionality reduction technique for identifying the representative trends, the technique employs a mechanism to replace each interval in the time space with “sketch”; these sketches are polynomial convolutions that are pre-computed through different sets of algorithms proposed by the authors. The model effectively computed the representative trends, using the pre-computed sketches. Through their results,they showed considerable performance gains against other models.

Mohamed et al. [24], classified periodicities into segment as well as symbol periodicity. In the first method, termed as segment periodicity they considered the entire time series. In the second concept termed as symbol periodicity they considered the periodicities symbols in the time series. Both type of periodicities were examined and analyzed through their proposed convolution-based algorithm. Further, they proposed an extension of symbol periodicity technique that identifies the periodic patterns of periods which are unknown, referred as obscure periodic patterns. They tried to discover periodicity pace and also frequent periodic patterns, simultaneously.

III. PERIODIC PATTERN MINING

Ozden, et al. [26] has introduced a model earlier for mining periodic patterns. They studied the performance drawback involved in mining association rules that vary cyclically over a period of time. They proposed two algorithms to clear up the problem of cyclic association rules and focused their work over the fact that the association rule can never abide to the minimum threshold over absolute

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

time series, but essentially the confidence of the association rule may remain far above the minimum threshold in certain time gaps. The first algorithm, named as sequential algorithm treated both the cycles and association rules independently, a new pruning method called cycle pruning was employed to reduce the time to discover cyclic association rules.Cycle pruning and optimization methods are used in the second algorithm named interleaved algorithm to discover cyclic association rules. Through series of experiments they proved that interleaved algorithm has significant performance against sequential algorithm.

Romani et al [28] proposed climate and remote sensing association patterns miner, a novel approach to locate patterns from heterogeneous time series.The proposed system, named RemoteAgri comprised of climate database, topographical images of low resolution captured through remote sensing satellites, an image processor and a module for time series extraction and mining. In their preprocessing phase, the images were subjected to geometric corrections to map land and agricultural applications. Through graphical interface, they designed more user friendly time series extraction module. Using sliding window concept, they mapped the patterns of other series with the transferred symbolic representation series with a view to discover patterns in multi-temporal satellite images.

Ma et al. [29] involved in discovering P-Patterns, a nomenclature given for partially periodic temporal patterns. The P-Patterns happen to occur every 30 seconds in a network with a loop of port-up event and followed by port-down event and then followed with a random gap. They employed a two step process, first finding the period lengths followed by temporal associations. The temporal associations were mined using a variation of Apriori algorithm. Their work focused on discovering the periods by employing chi-squared test and examined the model's execution with the existence of noise. The authors devised two algorithms to mine the p-patterns based on the discovered periods and temporal associations. Through experiments over synthetic data they proved that the association based algorithm has better tolerance over noisy data and period based algorithm has better computational efficiency.

Berberidis et al.[30]investigated partial periodic patterns for unsuspected periodicities. They proposed a novel model for discovering the periodicities over a time series without prior familiarity over the data. They employed Fast Fourier Transform to measure the Circular Autocorrelation Function, the function helps in generating candidate period lengths for each letter in character sequence. And in the refining step they employed Han's algorithm over the candidate period length to find partial periodic patterns. The algorithm gained linear superiority over both total number of time points and the alphabet size in character sequence.

Cheng-KuiGu& Xiao-Li Dong [31] proposed a model to substitute the position list produce algorithm which was the foundation of many earlier asynchronous periodic patterns mining algorithms. Their model was based on dot product to mineasynchronous periodic patterns from time series. They designed a mapping scheme that was based on binary representation and proposed an enhanced dot product algorithm to discover the event positions in the time series. Theirintention was to bring down the calculation time by replacing the existing series calculation method with parallel calculation and proved with their results showing that the proposed model was considerably accurate than other algorithms.

Yang et al. [32] put forward a model to discover asynchronous periodic patterns. The idea of mining asynchronous periodic patterns was found on the assumption that the patterns might be discovered only in certain subsequence due to disturbances which may shift the entire pattern occurrences. The model employed max_disand min_rep parameters to validate the qualification of patterns. They assumed that the pattern should reoccur in a certain time gap (min_rep) to represent its periodicity and the disturbance between two segments are bound to a higher limit (max_dis). The authors devised a two-phase algorithm where first phaseemploys distance based pruning to emulate potential periods. The generated periods are then subjected through a repeated series of steps to gather and recognize candidate patterns thereby the longest reliable subsequence is founded.

IV. SEQUENTIAL PATTERN MINING

R.Agrawal and R. Srikant[33] proposed on mining sequential patterns from massive transaction database. The authors proposed three algorithms to mine sequential patterns, they are i) AprioriSome, ii) AprioriAll and iii) AprioriDynamicSome. They divided the concern of mining sequential patterns into five phases as 1) sort phase – the transaction database is reconstructed as customer sequences, 2) L-itemset phase – the set of all L itemsets are discovered and from that the set containing large L-sequences are identified, 3) Transformation phase – the customer sequence is transformed into alternate representation and finds the large sequences that are contained in customer sequence.4) Sequence phase – three different algorithms with different approaches are employed to find desired sequences, and 5) Maximal phase - the set of substantial sequences are analyzed and the maximal sequences are discovered. The experimental results showed that AprioriSome and AprioriAll scaled linearly with number of customer transactions.

M.J.Zaki [34] proposed SPADE (**S**equential **PA**ttern**D**iscovery using **E**quivalence) classes to address the problem of making repeated database scans for mining sequential patterns. The models prior to SPADE adopted complex hash structures that had limited locality, the SPADE divides the problem into smaller sub-problems by employing combinatorial properties; these smaller sub-problems were solved by using a techniques like lattice search and join operations in main-memory. The model reduced the total number of database scans to just three scans. The experimental result showed that SPADE has considerable advantage over other models by ratio with present-processed data. Further, it was also proved that the model is linear scalable with respect to input sequences and other database parameters.

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*

Yong-Gui& Hong Yu [36] studied moving sequential pattern mining. The authors tried to overcome the flaws in models like PrefixSpan, which was not suitable to spatial applications, where the algorithms tend to generate large number of duplicate projected datasets. They designed Sequential Mining of Moving Patterns Based on Spatial Constraints in Mobile Environment (shortly SMPM) to overcome the drawbacks of PrefixSpan method. The proposed model has spatial constraint property and stores the initial place of the suffix in the sequence, this helps in avoiding the repeated pruning of same database and physical projections.

HaifengLi [38] proposed a model to face the realistic character of the data stream, that it cannot be stored in secondary storage for multiple scans. The authors designed a multiple level sequential pattern mining method; the approach had been built with the stream property in mind. The model employed conventional mining algorithms to come across sequential patterns from data streams that are over in-memory. The approach divided the memory to hold sequential patterns with different supports and the model proved its effectiveness through a series of experiments carried out.

V. PERIODIC FREQUENT PATTERN MINING

A class of frequent patterns, the periodic frequent pattern mining had garnered large research groups. To categorize a frequent pattern as a periodic-frequent pattern, it should occur periodically throughout the database.

Surana et al. [41] put forward an explanation for "rare item problem", an issue that occurs when trying to mine the periodic frequent patterns that contain both rare and frequent items. This dilemma arises usually when the mining models depend on "single constraints", which was prevalent in basic periodic pattern mining algorithms. Further, they also seek to confront the "rare item problem" by providing an approach that is based on "multiple constraints", which makes the approach not to comply to closure property. Though this makes the approach computationally expensive, yet it contemplates by generating some interesting patterns as periodic patterns. The authors inspired by this fact have designed another efficient approach that is based on multiple constraints and also satisfy the downward closure property. The results are effective when compared with other basic models.

Uday and Reddy [43] proposed an extended work of Tanbeer et al., but they added the real-world scenario to the model, as such the frequent patterns seldom occur throughout the database without any gaps. By adding an interesting measure that appear on certain gap throughout the database, the authors recommended an improved viewpoint and further a pattern-growth algorithm is used to find the periodic frequent patterns from the transactional database. The experiment'soutcome showed that the preferred model performs well against its counterparts.

Rashid et al. [44] in their work tried to emphasize the significance of temporal regularity measure of patterns, an important measure equivalent to the measure occurrence frequency of patterns. They proposed a model to mine regularly frequent patterns. They employed variance of interval time for measuring the temporal regularity and to find the regularly frequent patterns. They used pattern-growth approach with user-specified threshold. The experimental results of the model were time and memory efficient in discovering the repeatedly frequent patterns.

VI. HIGH UTILITY ITEMSET MINING

Over the last few years, Utility mining is evolved as one of the key areas in data mining. Unlike the frequent itemset mining, HUIM does not hold anti-monotone property [45, 46, 47]. Anti-monotone property states that supersets of infrequent itemsets are infrequent and subsets of frequent itemsets are frequent. This lack of anti-monotone property results complexity in handling candidate generation process.

Chan et al. [12], put forward a model that utilizes the profit of individual itemsets during the mining process. Here, the utility of an item are accessed using two measures called external utility and local transaction utility . The local utility is calculated from the database and the external utility values are supplied by the user. The transaction database and the external utility table provided by the user are used to mine itemsets that could satisfy the expectation of the user. As per the results of experiment, it is proved that user specified threshold value is not required in the proposed method and it is also successful in practice.

Liu et al. [48], introduced an algorithm with two phase based on transaction weighted utilization for mining high utility itemsets.The downward closure property is addressed in generating high utility itemsetsin the first phase. The second phase prunes the itemsets with over estimated utilities [31]. The downside of this method is its iterative nature which prevents the algorithm from scaling upwards. The approach showed considerable memory and computational efficiency when implemented over both synthetic and real databases that are even large. It also showed considerable memory and computational efficiency when implemented over both synthetic and real databases that are even large.

Lan et al. [45], put forward an approach for efficient utility mining, named as PB algorithm to mine high utility itemsets that implements a projection technique along with a new indexing structure. The use of indexing structure is to subject the conventional projection based algorithms to generate the sub-databases for mining. It lessened the unpromising candidates by employing a pruning strategy and theoutcome of the experiments showed that the PB algorithm outstands against Two-phase [48] and CTU-PRO. [49].

Li et al. [11], proposed a strategy named Isolated Items Discarding Strategy (IIDS), that can be used in utility mining approachof any level-wise and it uses a different tactic to diminish the number of candidates generated. The IIDS, have been integrated with the existing ShFSM and DCG models through FUM and DCG+ methods. The experimental result over both synthetic and real databases showed that FUM and DCG+ has competitive performance against Sh FSM and DCG models.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Chowdhury et al. [50], proposed multiple models that use tree structure to mine incremental and interactive high utility patterns. They also tried to maximize the performance by employing tree structures that could avoid multiple data structure design for multiple threshold values supplied by the user. They designed three tree structures i) Incremental HUP Lexicographic Tree ($IHUP_L$-Tree)- arrange items in a lexicographic order, ii) IHUP transaction frequency tree ($IHUP_{TF}$-Tree) and iii) IHUP-transaction-weighted utilization tree- used to build a compact representation of items by arranging the items based on their frequency ($IHUP_{TWU}$-Tree)- used to limit the mining time. The proposed algorithm utilizes "build once mine many" property.

Kavitha and Geetha [51] introduced a method to mine high utility itemsets with cross selling features. The external utilities of itemsets that sway potential high utility itemsets are redefined in the process of mining. It is connected with 'good transaction contains good itemset' approach. The outcome of experiment showed that the model has generated new set of unknown patterns that reflect cross-selling attributes.

VII. HIGH UTILITY SEQUENTIAL RULE MINING

High-utility itemset mining has wider application and proved to be an important data mining task. It identifies the itemsets that yield greater profits in a transaction databases. Following HUIM, High-Utility Sequential Pattern (HUSP) had developed to find itemsetsthat produce higher profit in sequences of customer transaction.

C.F.Ahmed et al. [53], put forward a model for mining high utility sequential patterns. They suggested that considering only the participation count of an item in a sequence does not add any essentials to the transaction and does not reflect real world scenarios. Their novel approach tried to extract real-life information through non-binary representation of items in a transactional database. They designed two algorithms i) UtilityLevel – a level wise approach that generates candidate sequence patterns from the database and ii) UtilitySpan – a pattern growth approach for mining HUSP. The experimental results showed that both the algorithms are efficient and scalable.

Yin et al. [54], incorporated the utility factor, they designed USpan, an algorithm to find high utility sequential patterns. They introduced a new tree based on lexicography named as lexicographic quantitative sequence tree, using this tree structure they were able to generate high utility sequence patterns and they also proposed two pruning strategies, one to find the utility of a node through concatenation and other strategy to find the utility of the node's children. The experimental results showed that the USpan model generates HUSP from large database with minimum utility.

O.K.Alkan et al. [55] tried to address the challenge of large search space that affects the efficiency of any algorithm while eliminating the candidate patterns. They proposed a model to diminish the search space. The envisioned algorithm was based on Cumulated Rest of Match (CRoM) based upper bound. This upper bound value gives more stringent utility of candidates to be pruned there by limiting the number of candidates generated. The authors have further proposed High Utility Sequential Pattern Extraction (HuspExt), an algorithm to measure the utility of the child patterns depending on their parents. Experimental result on real and synthetic data showed that the proposed model generates HUSP in more efficient manner.

G.C.Lan et al. [56] stated that the utility calculation cost for measuring the subsequence utilities in sequences has always proved to be a challenging task. They tried to resolve this bottleneck by supplementing a measure called "maximum utility measure". The measure states that the score of a subsequence in a sequence will always be one. With this assumption the measure simplifies the utility calculation of subsequences in the database. The authors designed an upper bound model to maintain the information without any loss during mining and also devised a pruning strategy based on projection to derive accurate utility of subsequences. They also came up with an indexing strategy in an effort to reduce the search time. The result showed that the advised method has better pruning accuracy and execution efficiency.

S.Zida et al. [57] addressed an inherent problem in sequential patterns, which is they are not bound under any confidence or support measure. They proposed an algorithm HUSRM (High-Utility Sequential Rule Miner) and various optimization methods to mine HUSP from transaction database. The results of their approach over four datasets showed that their model was 25 times faster and consumed memory up to 50%.

VIII. PERIODIC HIGH UTILITY ITEMSET MINING

Philippe et al. [59],proffered the basic model to mine period based high utilities. They analyzed the drawback associated with HUIM that it does not considerthe recurring patterns in the database. They counteracted this issue by proposing periodic high utility itemset mining (PHUIM). They focused to spot the items that yield higher profit and are bought together periodically by a customer. They proposed PHM (Periodic High-utility itemset Miner) to mine all periodic HUIs from the database. The analysis of the model showed that the algorithm filters out non-periodic patterns and generates real PHUIs that yield higher profits.

Ismail et al. [60] foundthat periodic high utility itemsets provide in deciding the profitability of any business enterprise. They emphasized that traditional model of mining HUIs would not be apt to discover the periodic buying patterns and the relationship between those items. They proposed a new method to address this limitation and generate profitable high utility periodic patterns from transactional database.

The outcome of their experimentshowed that the proposed model generates certain patterns that are previously not discovered but profitable.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

IX. CONCLUSION

This paper has imparted a survey of various contributions over and related to periodicity pattern mining. The main types of algorithms for finding periodic patterns are discussed in the paper. Further, the paper has given prime enhancement works of the periodic pattern mining problems. Also, the paper has looked into other extended works connected with periodic pattern mining for instance periodic frequent pattern mining, sequential HUI mining and periodic HUI mining. The insight idea of each model has been briefly described in each section. Also, each section gave the basic application to extended application of the respective technology.

REFERENCES

1. C. C. Aggarwal, J. Han, “Frequent pattern mining”, Springer, 2014.
2. Jian Pei, Jiawei Han, Hongjun Lu, ShojiroNishio, Shiwei Tang Dongqing Yang (2007), “H-Mine: Fast and space-preserving frequent pattern mining in large databases”, IIE Transactions, 39:6, 593-605.
2. 3.RakeshAgrawal , Tomasz Imieliński , Arun Swami, “Mining association rules between sets of items in large databases”, Proceedings of the 1993 ACM SIGMOD international conference on Management of data, p.207-216, May 25-28, 1993, Washington, D.C., USA.
1. Minato S., Uno T., Arimura H. (2008), “LCM over ZBDDs: Fast Generation of Very Large-Scale Frequent Itemsets Using a Compact Graph-Based Representation”. In: Washio T., Suzuki E., Ting K.M., Inokuchi A. (eds) Advances in Knowledge Discovery and Data Mining. PAKDD 2008. Lecture Notes in Computer Science, vol 5012. Springer, Berlin, Heidelberg.
2. H. Yao, H. J. Hamilton, L. Geng, “A unified framework for utility-based measures for mining itemsets”. In Proc. of ACM SIGKDD 2nd Workshop on Utility-Based Data Mining, pp. 28--37, USA, Aug., 2006.
3. Hong Yao, Howard J. Hamilton, “Mining itemset utilities from transaction databases”, Data & Knowledge Engineering, v.59 n.3, p.603-626, December 2006.
4. Yao, H., Hamilton, H. J., and Butz, C. J. 2004, “A foundational approach for mining itemset utilities from databases”. In Proceedings of the SIAM International Conference on Data Mining. Orlando, pp. 482-486.
5. Erwin A., Gopalan R.P., Achuthan N.R. (2008), “Efficient Mining of High Utility Itemsets from Large Datasets”. In: Washio T., Suzuki E., Ting K.M., Inokuchi A. (eds) Advances in Knowledge Discovery and Data Mining. PAKDD 2008. Lecture Notes in Computer Science, vol 5012. Springer, Berlin, Heidelberg.
6. C. F. Ahmed, S. K. Tanbeer, B. Jeong and Y. Lee, “Efficient Tree Structures for High Utility Pattern Mining in Incremental Databases”, in IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 12, pp. 1708-1721, Dec. 2009.
7. Yu-Chiang Li , Jieh-Shan Yeh , Chin-Chen Chang, “Isolated items discarding strategy for discovering high utility itemsets”, Data & Knowledge Engineering, v.64 n.1, p.198-217, January, 2008.
8. Raymond Chan , Qiang Yang , Yi-Dong Shen, “Mining High Utility Itemsets”, Proceedings of the Third IEEE International Conference on Data Mining, p.19, November 19-22, 2003.
9. Ying Liu , Wei-keng Liao , AlokChoudhary, “A fast high utility itemsets mining algorithm”, Proceedings of the 1st international workshop on Utility-based data mining, p.90-99, August 21-21, 2005, Chicago, Illinois.
10. Agrawal R., Faloutsos C., Swami A. (1993), “Efficient similarity search in sequence databases”. In: Lomet D.B. (eds) Foundations of Data Organization and Algorithms. FODO 1993. Lecture Notes in Computer Science, vol 730. Springer, Berlin, Heidelberg.
11. Donald J. Berndt, James Clifford, “Using dynamic time warping to find patterns in time series”, Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, July 31-August 01, 1994, Seattle, WA.
12. Chan, K., Fu, W., “Efficient Time Series Matching by Wavelets”, Proceedings of the 15th International Conference on Data Engineering, p.126, March 23-26, 1999.
13. Byoung-KeeYi , Christos Faloutsos, “Fast Time Sequence Indexing for Arbitrary Lp Norms”, Proceedings of the 26th International Conference on Very Large Data Bases, p.385-394, September 10-14, 2000.
14. Yongwei Ding, Xiaohu Yang, A. J. Kavs and Juefeng Li, “A novel piecewise linear segmentation for time series”, 2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE), Singapore, 2010, pp. 52-55.
15. Y. Huang and P. Liou, “Retrieving Representative Structures from XML Documents Using Clustering Techniques”, 2011 European Intelligence and Security Informatics Conference, Athens, 2011, pp. 332-339.
16. L. Liu and N. Cui, “A mining algorithm based on time series association rules”, 2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), Yichang, 2012, pp. 786-789.
17. R. U. Kiran, H. Shang, M. Toyoda, M. Kitsuregawa, “Discovering recurring patterns in time series”, EDBT, pp. 97-108, 2015.
18. Christos Faloutsos , M. Ranganathan , YannisManolopoulos, “Fast subsequence matching in time-series databases”, Proceedings of the 1994 ACM SIGMOD international conference on Management of data, p.419-429, May 24-27, 1994, Minneapolis, Minnesota, USA.
19. PiotrIndyk , Nick Koudas , S. Muthukrishnan, “Identifying Representative Trends in Massive Time Series Data Sets Using Sketches”, Proceedings of the 26th International Conference on Very Large Data Bases, p.363-372, September 10-14, 2000.
20. M. G. Elfeky, W. G. Aref and A. K. Elmagarmid, “Periodicity detection in time series databases”, in IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 7, pp. 875-887, July 2005.
21. B. Ozden, S. Ramaswamy and A. Silberschatz, "Cyclic association rules," Proceedings 14th International Conference on Data Engineering, Orlando, FL, USA, 1998, pp. 412-421.
22. Jiawei Han, Guozhu Dong and Yiwen Yin, “Efficient mining of partial periodic patterns in time series database”, Proceedings 15th International Conference on Data Engineering (Cat. No.99CB36337), Sydney, NSW, Australia, 1999, pp. 106-115.
23. L. A. S. Romani et al., “A New Time Series Mining Approach Applied to Multitemporal Remote Sensing Imagery”, in IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 1, pp. 140-150, Jan. 2013.
24. Sheng Ma and J. L. Hellerstein, “Mining partially periodic event patterns with unknown periods”, Proceedings 17th International Conference on Data Engineering, Heidelberg, Germany, 2001, pp. 205-214.
25. Berberidis, C., Vlahavas, I.P., Aref, W.G., Atallah, M.J., Elmagarmid, A.K., 2002. “On the discovery of weak periodicities in large time series”. In: Proceedings of the 6th European Conference on Principles and Practice of Knowledge Discovery in Databases, pp. 51–61.
26. Cheng-KuiGu and Xiao-Li Dong, “Dot product based time series asynchronous periodic patterns mining algorithm”, 2009 International Conference on Machine Learning and Cybernetics, Hebei, 2009, pp. 178-182.
27. Jiong Yang, Wei Wang and P. S. Yu, “Mining asynchronous periodic patterns in time series data”, in IEEE Transactions on Knowledge and Data Engineering, vol. 15, no. 3, pp. 613-628, May-June 2003.
28. RakeshAgrawal , RamakrishnanSrikant, “Mining Sequential Patterns”, Proceedings of the Eleventh International Conference on Data Engineering, p.3-14, March 06-10, 1995.
29. Mohammed J Zaki. (2001), “SPADE: An efficient algorithm for Mining requent Sequences”, Journal Machine Learning, Volume 42 Issue 1-2, January-February 2001 Pages 31-60.
30. Yong-GuiZou and Hong Yu, “Moving sequential pattern mining based on Spatial Constraints in Mobile Environment”, 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems, Xiamen, 2010, pp. 103-107.
31. D. Fricker, Hui Zhang and Chen Yu, “Sequential pattern mining of multimodal data streams in dyadic interactions”, 2011 IEEE International Conference on Development and Learning (ICDL), Frankfurt am Main, 2011, pp. 1-6.
32. Haifeng Li, “A stream sequential pattern mining model”, Proceedings of 2011 International Conference on Computer Science and Network Technology, Harbin, 2011, pp. 704-707.
33. C. Chang, H. Chueh and Y. Luo, “An integrated sequential patterns mining with fuzzy time-intervals”, 2012 International Conference on Systems and Informatics (ICSAI2012), Yantai, 2012, pp. 2294-2298.

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication

34. Tanbeer S.K., Ahmed C.F., JeongBS., Lee YK. (2009), "Discovering Periodic-Frequent Patterns in Transactional Databases". In: Theeramunkong T., Kijsirikul B., Cercone N., Ho TB. (eds) Advances in Knowledge Discovery and Data Mining.PAKDD 2009. Lecture Notes in Computer Science, vol 5476. Springer, Berlin, Heidelberg.
35. Surana A., Kiran R.U., Reddy P.K. (2012), "An Efficient Approach to Mine Periodic-Frequent Patterns in Transactional Databases". In: Cao L., Huang J.Z., Bailey J., Koh Y.S., Luo J. (eds) New Frontiers in Applied Data Mining. PAKDD 2011. Lecture Notes in Computer Science, vol 7104. Springer, Berlin, Heidelberg.
36. .UdayKiran R., Krishna Reddy P. (2010), "Towards Efficient Mining of Periodic-Frequent Patterns in Transactional Databases". In: Bringas P.G., Hameurlain A., Quirchmayr G. (eds) Database and Expert Systems Applications. DEXA 2010. Lecture Notes in Computer Science, vol 6262. Springer, Berlin, Heidelberg.
37. Rashid M.M., Karim M.R., JeongBS., Choi HJ. (2012), "Efficient Mining Regularly Frequent Patterns in Transactional Databases". In: Lee S., Peng Z., Zhou X., Moon YS., Unland R., Yoo J. (eds) Database Systems for Advanced Applications. DASFAA 2012. Lecture Notes in Computer Science, vol 7238. Springer, Berlin, Heidelberg.
38. Lan, GC., Hong, TP. and Tseng V.S., "An efficient projection-based indexing approach for mining high utility itemsets", KnowlInfSyst (2014) 38: 85.
39. 42. G.-C. Lan, T.-P. Hong, V.S. Tseng, "An efficient gradual pruning technique for utility mining", International Journal of Innovative Computing Information and Control, 8 (2012), pp. 5165-5178.
40. J. Liu, K. Wang and B. C. M. Fung, "Direct Discovery of High Utility Itemsets without Candidate Generation", 2012 IEEE 12th International Conference on Data Mining, Brussels, 2012, pp. 984-989.
41. Liu Y., Liao W., Choudhary A. (2005), "A Two-Phase Algorithm for Fast Discovery of High Utility Itemsets". In: Ho T.B., Cheung D., Liu H. (eds) Advances in Knowledge Discovery and Data Mining. PAKDD 2005. Lecture Notes in Computer Science, vol 3518. Springer, Berlin, Heidelberg.
42. Alva Erwin , Raj P. Gopalan , N. R. Achuthan, "A bottom-up projection based algorithm for mining high utility itemsets", Proceedings of the 2nd international workshop on Integrating artificial intelligence and data mining, p.3-11, December 01-01, 2007, Gold Coast, Australia.
43. C. F. Ahmed, S. K. Tanbeer, B. Jeong and Y. Lee, "Efficient Tree Structures for High Utility Pattern Mining in Incremental Databases", in IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 12, pp. 1708-1721, Dec. 2009.
44. Kavitha V, Geetha B G, (2016), "High utility itemset mining with influential Cross selling items from transactional Database", International Journal of Advanced Engineering Technology, vol 8, pp. 820 -826.
45. Vincent S. Tseng , Cheng-Wei Wu , Bai-En Shie , Philip S. Yu, "UP-Growth: an efficient algorithm for high utility itemset mining", Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, July 25-28, 2010, Washington, DC, USA.
46. C.F. Ahmed, S.K. Tanbeer, B.-S. Jeong, "A novel approach for mining high-utility sequential patterns in sequence databases", ETRI J., 32 (5) (2010), pp. 676-686.
47. JunfuYin ,ZhigangZheng , Longbing Cao, "USpan: an efficient algorithm for mining high utility sequential patterns", Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, August 12-16, 2012, Beijing, China.
48. O. K. Alkan and P. Karagoz, "CRoM and HuspExt: Improving Efficiency of High Utility Sequential Pattern Extraction", in IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 10, pp. 2645-2657, 1 Oct. 2015.
49. Guo-Cheng Lan, Tzung-Pei Hong, Vincent S. Tseng, Shyue-Liang Wang, "Applying the maximum utility measure in high utility sequential pattern mining", Expert Systems with Applications, Volume 41, Issue 11, 2014, Pages 5071-5081.
50. Zida S., Fournier-Viger P., Wu CW., Lin J.CW., Tseng V.S. (2015), "Efficient Mining of High-Utility Sequential Rules". In: Perner P. (eds) Machine Learning and Data Mining in Pattern Recognition. MLDM 2015. Lecture Notes in Computer Science, vol 9166. Springer, Cham.
51. Fournier-Viger P., Lin J.CW., Duong QH., Dam TL. (2016), "PHM: Mining Periodic High-Utility Itemsets". In: Perner P. (eds) Advances in Data Mining. Applications and Theoretical Aspects.ICDM 2016. Lecture Notes in Computer Science, vol 9728. Springer, Cham.
52. W. Ismail, M. M. Hassan and G. Fortino, "Productive-associated Periodic High-utility itemsets mining", 2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC), Calabria, 2017, pp. 637-642.
53. J. C. Lin, Jiexiong Zhang, P. Fournier-Viger, T. Hong, Chien-Ming Chen and J. Su, "Efficient mining of short periodic high-utility itemsets", 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, 2016, pp. 003083-003088