

# Review on Mining High Utility Patterns Decreasing Candidates

Satyajeet Bankar<sup>1</sup>, Krushna Barve<sup>2</sup>, Rohit Dusane<sup>3</sup>, Tejas Desale<sup>4</sup>

Department of Computer Engineering,  
 Loknete Gopinathji Munde Institute of Engineering Education and Research, Nasik

## Abstract:

Utility mining is a latest development of data mining technology. Among utility mining issues, utility mining with the item set share structure is a hard one as no anti-monotonicity property hold with the interestingness measure. Prior work on this problem all use a two-phase, candidate generation approach with one exception that is however in efficient and not scalable with big databases. This paper plan sa novel algorithm that find shigh utility patterns in a single phase with out generating candidate

## 1: Introduction:

Utility mining has rise drecently to address the limitation of frequent pattern mining by see in g the user’s prospector objective as well as the raw data. Utility mining with the itemset share structure for example, discovering combination so products with high profit or revenues, is much harder than other categories of utility mining problems, for in stance, weighted itemset mining and objective-oriented utility-based association mining.

Data mining is defined broadly as a process to extract implicit, people do not know in advance, but is potentially useful information and knowledge from a lot of noisy, uncertain, stored in various forms or incomplete large data sets.

The experiment is the number of candidates can be huge, which is the scalability and proficiency hold up.

## 2.1: Comparative Analysis:

Although a lot of effort has been made to reduce the number of candidates generated in the first phase, the test still persists when the raw data contains many long transactions or the least utility threshold is small.

## 2: Related Works:

High utility pattern mining problem is thoroughly related to frequent pattern mining, with constraint-based mining. In this section, we briefly review prior works both on frequent pattern mining and on utility mining, and discuss how our work connects to and differs from the prior works.

Paper Name	Auther	Doma in	Issues	Solution	Algorit hm	Implementation	Conclusion
Trust-but-Verify: Verifying Result Correctness of Outsourced Frequent Itemset Mining in Data-Mining-As-a-Service Paradigm	1)Boxin g Dong 2) Hui Wang	Data Mining	how can the client of weak computational power verify that the server returned correct mining result?	Before sending the dataset D to the server, the client constructs an authenticated data structure.	Bilinear algorithm	We implemented a prototype of our probabilistic approach in Java,	In this paper, we present two integrity verification approaches for outsourced frequent itemset mining.
A Study on Effective Chart Visualization and Interaction Techniques for	1)James Walker 2) Rita Borgo	Data Mining	Pilot study results were positive and revealed trends in the data	regular basis (every couple of weeks). They provide us with a constant source	Layout algorithm	We have implemented all of the methods in our software in a consistent way.	In this paper, we comparatively evaluate existing methods for exploring time-series data.

Time-Series Data			supporting our initial hypothesis,	of feedback.			
A Framework for Categorizing and Applying Privacy-Preservation Techniques in Big Data Mining	1)LeXu 2) Yan chen	Data Mining	A wider perspective than data mining, we investigated a range of approaches across the entire KDD process.	To protect sensitive information in mined data, researchers need a variety of ongoing work.	Sanitation algorithm	To find that agreement, we developed a game model, based heavily on earlier $k$ -anonymity algorithm to the gathered data.	In this paper, we comparatively work has broaden the horizons for early assessment and prediction.
Mining Disease Sequential Risk Patterns from Nationwide Clinical Databases for Early Assessment of Chronic Obstructive Pulmonary Disease	1)Yi-Ting Cheng	Data Mining	To our best knowledge, this is the first work that addresses the important issue of early assessment on COPD through mining	propose a novel approach for early assessment on COPD by mining COPD-related sequential risk patterns from diagnostic records.	CBS (classify by sequence) algorithm	We can effectively classify sequential risk patterns for prediction. To support our work of COPD assessment and analysis.	The approach and corresponding techniques presented in this work has broaden the horizons for early assessment and prediction on the disease
Review on State of Art Data Mining and Machine Learning Techniques for Intelligent Airport Systems	1)Jain Bo 2) Chamta Aryawana	Data Mining	The challenge faced by airports today is the complexity of players and processes.	The airport ecosystem must become more 'intelligent' to optimize its supply chain.	FP-TREE algorithm	The starving management bodies of airports have always tried to implement isolated solutions	A system with high recalls but low precision returns many results, but most of its predicted labels are incorrect
Accelerating Graph Mining Algorithms via Uniform Random Edge Sampling	1) Ruohan Gao 2) Pili Hu	Data Mining	The random subgraphs resulted from URE sampling can often be used as substitutes for the original graphs .	These graph mining tasks can yield well-approximated solutions for the original graph with substantial reduction.	Apriori algorithm	We demonstrate how to make use of URE sampling to accelerate common graph mining	This work is an attempt to use Uniform Random Edge (URE) sampling to accelerate graph mining algorithms.
An Approach Based On Association Rules Mining To Improve Road Safety In Morocco	1)Garnat hi 2)R. Agrawal	Data Mining	we find three problems like, sorting, selection, and arrangement. In this context of an important number of extracted rules.	resulting from an accident using real data obtained from the Ministry of Equipment and Transport of morocco,	Machine learning algorithm	The implementation of the case provides a preference relationship between rules and profiles.	In this paper, we have discussed the usefulness and relevance problem issued from a KDD process.
Use of Reality Mining Dataset for Human Behavior Analysis – A Survey	1) Sunita K v n 2) Rama prasad	Cloud Computing	This system addressed various issues like who should receive an incoming notification	The sensor data collected from smart phones is experimented with survey data	Significant Subgraph algorithm	Implement numerous applications and systems, such as a mobile app that helps smokers quit smoking.	Smart phones of current and future generations will play a vital role in everyone's life.
Mining the Enriched	1) Pieter Meysm	Data Mining	A common problem is the	Detect personalised and	DP-based	This implementation	In this paper, we have presented a novel type

Subgraphs for Specific Vertices in a Biological Graph	an 2) Kris Laukens		discovery of frequent subgraphs in a graph data set.	abnormal behaviours of internet users, We propose Sequential topic patterns.	algorithm	allows the user to set any of the introduced interestingness .	of subgraph mining approach for subgroup discovery subgraph.
Mining User – Aware Rare Sequential Topic Pattern in Document Stream	)Yunkun Hu 2) Gurumit Baggo	Cloud Computing	In this paper we formally define and a group of algorithm and design to combine systematically solves this problem.	its algorithm performance directly affects the efficiency of data mining and the integrity, effectiveness of ultimate data mining results.	Apriori algorithm	its algorithm performance directly affects the efficiency of data mining and the integrity, effectiveness of ultimate data mining results.	Mining URSTP in publish document stream on the internet is the significant and challenging problem.
							.

### 2.1: Frequent Pattern Mining:

Frequent pattern mining was first projected by Agrawal, which is to discover all patterns whose supports are no less than a user-defined minimum support threshold. Frequent pattern mining employs the anti-monotonicity property: the provision of a superset of a pattern is no more than the provision of the pattern. Algorithms for mining frequent patterns as well as algorithms for mining high utility patterns tumble into three classes, breadth-first search, depth-first search, and hybrid search.

### 2.2: Constraint-Based Mining:

Constraint-based mining is a landmark in evolving from frequent pattern mining to utility mining. Works on this are mainly focus on how to push constraints into frequent pattern mining algorithms. Pei et al. discussed constraints that are related to (normalized) weighted drops.

L. DeRaedt inspected how standard constraint programming techniques can be applied to constraint-based mining difficulties with constraints that are monotone, anti-monotone, and adaptable.

### 2.3: Some Categories of Utility Mining:

Interestingness measures can be branded as objective measures, subjective measures, and semantic measures. Objective measures, such as support or confidence are built only on data; Subjective measures, such as abruptness or novelty, take into account the user's domain information; Semantic measures, also known as utilities, consider the data as well as the user's expectation.

### 3. High Utility Pattern Growth:

The universal approach to mining high utility pattern is to check on each subset  $X$  of  $I$ , and test if  $X$  has a utility over the threshold. However, a comprehensive enumeration is infeasible due to the huge number of subsets of  $I$ , and hence it is dangerous to employ strong pruning techniques.

### 4. Association Rule Mining Algorithm:

Apriori algorithm, although its implementation process is relatively simple, but each generation of candidate set containing different number of items needs to scan the transaction database, when the candidate set larger, space-time overhead of Apriori algorithm becomes relatively large; directory in  $N_k$  contains all the candidate  $k$ -dimension item sets in the transaction.

### 5. Performance Analysis of Association Rule Mining Concept:

At present, there are many frequent item sets generation algorithms, and when generating frequent  $k$ -item sets these algorithms will scan each transaction of the database for statistics of the support of  $k$ -item sets, and according to the determined minimum support find all frequent  $k$ -item sets at the first  $k$  iteration. However, because that the size of the databases usually very large, so the above method consumes too much time.

### 6. Conclusion and Future Scope:

This paper proposes a new algorithm, d2HUP, for utility mining with the item sets are framework, which finds high utility patterns without candidate generation. Our influences contain: A linear data structure, CAUL, is proposed, which marks the root cause of the two-phase, candidate generation approach adopted by prior algorithms, that is, their data structures cannot keep the unique utility information.

### Acknowledgment:

In the future, we will work on high utility sequential pattern mining, parallel and distributed algorithms, and their application in big data analytics. With the development of information technology, data mining technology has obtained widespread concern, which prompted the scientific community more in-depth study of this technology. Data mining technology has many research fields, in which association rule mining is one of the important research directions, the in-depth study on it not only has important theoretical significance, but also has a very important application value.

## 8: References

- [1]R.Agarwal, C.Agarwal, and V.Prasad, “Depth first generation of long patterns,” in Proc. ACMSIGKDD Int. Conf. Knowl. Discovery Data Mining, 2000, pp.108–118.
- [7]Dickson, 2014. A Survey of Predictive Analytics in Data Mining with Big Data I Dickson Lam.
- [8]Eagle N., Greene K., “Reality Mining – Using Big Data to Engineer a Better World”, Association of American Publishers, MIT Press.
- [9]C.K.Chui and B.Kao, “A decremental approach for mining frequent itemsets from uncertain data,” in Proc. PAKDD’08, 2008, pp.64–75.
- [2]R.Agrawal, T. Imielinski, and A.Swami, “Mining association rules between sets of items in large databases,” in Proc. ACMSIGMOD Int. Conf. Manage. Data, 1993, pp.207–216.
- [3]R.Agrawal and R.Srikant, “Fast algorithms for mining association rules,” in Proc. 20th Int. Conf. Very Large Databases, 1994, pp.487–499.
- [4] L.DeRaedt, T.Guns, and S.Nijssen, “Constraint programming for itemset mining,” in Proc. ACMSIGKDD, 2008, pp.204–212.
- [5]Jiawei Han, Hong Cheng, Dong Xin, Xifeng Yan. Frequent Pattern Mining: Current Status and Future Directions[J], Data Mining and Knowledge Discovery, 2007, volume(15): 55-86.
- [6] HAN J, PEI J, YIN Y. Mining Frequent Patterns Without Candidate Generation[C] , Proc of the 2000 ACM SIGMOD Internal Conference on Management of Data. Dallas, Texas: ACM Press, 2000: 1-12.
- [10]M. J. Zaki, “Efficient enumeration of frequent sequences”, Journal of Machine learning, vol. 42, no. 1-2, pp.31-60, 2001
- [11] C. T. Li, Emerging Digital Forensics Applications for Crime Detection, Prevention and Security, IGI-Global, 2013.