# A Report on Haui-Miner and Ehaupm Algorithms on Pattern Mining with Upper Limits

**K. Anuradha, V. Srilakshmi, Madhuri Bandla**

***Abstract: Utility-mining is the present developing discipline of information-mining. Utility-mining combines different structures such as High relevant item-set mining, Relevant successive item-set mining, Negative relevant item-set mining, Uncommon high relevant item-set mining and so forth. Each procedure of these item-sets mining doesn't acknowledge length of item-sets. An ongoing improvement in the field of Utility-mining is high normal utility item-set mining. The normal Utility-mining deals with length of item- sets alongside the utility of item-sets. Here few calculations are introduced to recover high average relevant item-sets present in the database. Primary target of the present work was to look at the three High Normal Utility Models calculations:1)High Normal Utility Models (HAUP) calculation, 2)High Normal Utility Item-Set-Excavator (HAUI-Miner) Calculation and 3)Productive High Normal Utility Pattern-Mining (EHAUPM) calculation. The execution-time and memory-space are examined as achievement measures for correlation. The EHAUPM calculation is more efficient compared to other calculations; this is discovered from the performed analysis.***

***Keywords: Average high utility item-sets, EHAUPM calculation, HAUI miner calculation, HAUP calculation, High normal Utility-mining, Length of item-sets, Utility-mining.***

## I. INTRODUCTION

**U**tility-mining grabs the high utility item-sets from the databank [1]. The significant measure of an item is termed as utility based on the domain nature [2, 3]. Utility-mining joins different structures such as High Utility Item-Set Mining, Utility Regular Item-Set Mining, Negative Utility Item-Sets Mining, Uncommon High Relevant Item-Set Mining, Arrangement High Relevant Item Mining, and High Relevant Affiliation Mining and so on. The primary target of all these structures is to fetch high utility item-sets (HUI) by considering the minimal threshold-esteem. Models with utility esteem which are higher than threshold-esteem is known as HUI .Each of these structures does not think about the length of the models. The above is the disadvantage Observed from the conventional Utility-mining approaches [4, 5] because the length of the models is also having

Significant effect on utility of models .The utilities of item-set probably have greater utility for item-sets which have more items. Consider instance, utility of the two item-sets is not exactly the utility of three item-sets. By considering the problem and for solving the problem, high normal relevant item-set extraction was introduced. This approach acknowledges the both quantity and utility of item-set. Thus, this approach is the best for calculating utility of item- sets in the real-time operations.

Customary relevant item-sets mining procedures calculates the utility estimation of a pattern by multiplying the interior and exterior utility esteem. Interior utility of model is defined as some huge proportion of the model. The unit measure of the model is defined as the Exterior utility of a model. A few utility measures are characterized in conventional Utility-mining calculations to recover high utility models from a database. They are Transaction Utility, Transaction weighted utility, High exchange weighted utility item-sets, utility of item-set [1−4]. Every one of these measures considers just the utility estimations of the models, not the length of models. The length of models builds, the utility of models additionally increments. Thus, high average utility item-set mining (HAUIM) was proposed.

The normal utility is characterized as the addition of the utilities of the item-set in exchanges where they are appeared, partitioned by the quantity of items that it contains. This measure conquers the disadvantage of customary Utility-mining calculations. The principle goal of this work is to think about three High normal utility models calculations HAUP calculation, HAUI-mineworker calculation and EHAUPM calculation.

## II. LITERATURE SURVEY

Affiliation rule mining and relevant item-set mining are central information mining undertakings [1]. Mostly these approaches are acknowledged because of their performance in revealing item-sets which have high event densities in databases, which are used in numerous applications. Affiliation rules ARs are primarily produced by the Apriority approach. This algorithm is performed in dual stages. Firstly, it separates the arrangement of relevant item-sets (FIs) considering the client- specified minimal support limit.

Revised Manuscript Received on October 15, 2019

**Dr. K. Anuradha** is working as Professor in CSE dept, GRIET, Hyderabad, India. Completed her Ph.D.
E-mail:kodali.anuradha@yahoo.com
**V. Srilakshmi** is working as Assistant Professor in CSE dept, GRIET, Hyderabad, India. E-mail: potlurisrilakshmi@gmail.com
**Madhuri Bandla** is currently pursuing master's degree program in Computer science and engineering in GRIET, Hyderabad, India. E-mail: Info.bandlamadhuri@mail.com

Secondly, The ARs set is formed by combining the acquired FIs and by considering the client-specified minimal confidence the respective FIs. The incremental high-utility pattern (IHUP) approach [2] is implemented for incrementing and interacting with mined HUIs depending upon the tree representation which is identical to FP-tree approach. A HUI-Miner approach is introduced depending on the utility-list representation for mining HUIs by ignoring produced entrants and depth-first search is also utilized in this approach.

## III. STARTERS IN HIGH NORMAL UTILITY MODELS

Suppose let us consider B = {b1, b2, b3… ….bm} be a lot of items. Let E = {E1, E2, E3 En} be a Database DB, where Ei denotes exchanges. Let O = { O(b1), O(b2) …. O (bm)} be the unit of items B.

Give the A, a chance to be item-set consisting items bm and l is the length of item-set. The length of item-set is the quantity of items in A. Let µ be the base normal utility edge. Fundamental illustrations for figuring utility estimations of the item-sets are defined below [1−8]:

**Illustration 1:** An inward utility of the item is an amount of item in an exchange database.

**Illustration 2:** An outside utility of an item is a unit benefit of the item.

**Illustration 3:** Utility capacity c is the result of the inward and outer utility

**Illustration 4:** Utility of the item in exchange B is the utility capacity of the item in that specific exchange.

**Illustration 5:** Utility of item-set S in exchange B is characterized whole of utility of items in that item-set in an exchange.

**Illustration 6:** Utility of item-set S in database DB is characterized as the addition of utility of item- sets in all exchanges.

**Illustration 7:** Utility of an exchange B is characterized as the addition of utility of every item in an exchange.

**Illustration 8:** Item-set B will become high utility item-set only if its utility is more prominent than least utility limit.

## IV. RELATED WORKS

Krishnamoorthy [3] introduced a two-stage calculation for HUI mining. Firstly, the database is searched and resulted entrants are eliminated utilizing the upper limit condition. Secondly, the database is searched one more time and the absolute utility esteem of remaining entrants are evaluated. This approach then results the HUIs from entrants. This approach has a drawback of reducing entrants and time for researching database.

In [4], the creators mainly considered the comparison of high utility approaches. The creators examined UP-growth+, efficient fast item-set mining (EFIM) calculation, fast high Utility- mining (FHM) calculation, high utility item-set digger (HUI-Miner) calculation, direct discovery of high utility pattern (D2HUP) calculations. By considering implementation time and memory space consumed by these approaches are acknowledged as efficiency measures. D2HUP is the best approach when compared to the other available approaches by considering implementation time as an efficiency measure. HUI-Miner approach consumes little space compared to other approaches and it is the best approach when memory space is considered as an efficiency measure. Most of the calculations present in research publications used individual normal minimal utility edge for retrieving HAUI.

In [5] creators presented the theory of various minimal normal utility edges. Each item (t) of this method has distinct edges. Likewise, improved estimated utility co-occurrence pruning system (IEUCP) and pruning before count system (PBCS) methodologies were proposed for pruning item- sets.

In [6], creators explained different types of Utility- mining like negative Utility-mining, on rack Utility-mining, visit Utility-mining ideas, high normal utility item-set mining, shelf high utility item-set mining and so forth. The creators using instances well described the utility evaluations. The greater part of Utility-mining approaches can deal with fixed databases.

Dynamic databases are utilized mostly in the actual applications. The modifications are performed in those databases by entering the new exchanges. The HUI can be retrieved from the variable databases by using the few approaches present now. Those approaches are called as incremental high-utility item-set mining (IHUIM) approach. The IHUIM approach is well described by the creators [8].

In Hong et al. [9] the creators utilized normal utility upper-limit for pruning obtained entrants, also created the list containing item-sets whose upper-limit esteem is larger or equivalent to edge. The list consisting (r+1)item-set are produced with the combination r-item-sets and 1-item-sets having high normal utility upper-limit.

In [10], creators proposed, HAUI-Miner approach. They introduced an effective normal utility (AU) - list arrangement for evaluating high normal utility item-sets more proficiently. This approach analyses the inquiry space by ignoring the entrants' production. To mine rapidly, additionally an effective pruning methodology is executed for reducing the inquiry space.

In [11] the creators introduced EHAUPM approach. The HAUIM approach can be enhanced by introducing the two new more tightly upper- limits methods in the present work. To mine HAUIs these methods are considered as substitute to the conventional normal utility upper-limit (AUUB) method. The lower upper-limit model (lub) by considering the existing-maximal utility of exchanges decreases the upper-limit esteems of the utilities of item-sets.

In second upper-limit method (rtub), to additionally tight an upper-limit the unnecessary items of exchanges are ignored. Three pruning methodologies are likewise developed.

In [12] Lin et al and other authors introduced the approach for revealing HAUIs. This approach is the projection dependent hierarchical method. Creators once again additionally introduced the novel upper-limit method using prefix theory. By Comparing with the TMU method, the current approach decreases the insignificant entrants count.

## V. ALGORITHMS

### A. High Normal Utility Item-Set Pattern Tree (HAUP Tree) Calculation:

Lin et al. [12] introduced a tree structure known as high normal utility pattern tree (HAUP tree) representation is used for storing utility data of item-sets. The HAUP-development calculation is introduced for mining high normal utility item-sets present in the tree representation. The present calculation incorporates both the high normal Utility-mining calculation.

FP-tree-like methodology .The calculation introduced the consolidated tree representation to effectively determine high normal utility models. This tree representation is known as HAUP tree [12]. The tree consists on many points, each point should collect the average utility upper limit of the item in the point and additionally the lengths of its previous items in the route. After the tree is developed, the items from the base are stored in Header-Table step by step. Each point that is processed presently with item in the HAUP tree collectively produces the item-sets. If an item-set is produced from more than one way, combine the amounts from the two exhibits by expansion. At that point, the real normal utility estimation of each blended item-set is determined. Check now if the genuine normal utility esteem of each produced normal utility item-set is greater than or equivalent to the base normal utility esteem. If it is, the items are high normal utility item-sets.

### B. High Normal Utility Item-Set-Excavator (HAUI Miner) Algorithm

HAUI-Miner calculation utilizes a normal utility AU- list structure for storing the data required during mining procedure. In addition, a calculation called HAUI-Miner was created for mining HAUIs most proficiently compared to past approaches [12]. The current approach includes the exchange greatest utility descending conclusion TMUDC characteristic. The current method adequately decreases the pursuit area [12], consequently permits pruning unfavorable entrants first.

Consequently, the calculation decreases pursuit area for finding the real high normal utility models [12] effectively. The exchange greatest utility of an exchange is characterized as the most extreme utility of items in an exchange.

The normal utility upper limit of an item-set (auub) is described as the addition of the exchange most extreme utility of exchanges. An item-set A is called high normal utility upper limit item-set if normal utility upper limit of the item-set isn't exactly base normal utility. The HAUI-Miner calculation filters the database two times to figure out normal utilities of entrant item- sets. While performing the principal database analysis, the arrangement of the high normal utility upper-limit 1-item-sets were fetched [12]. At that point AU-lists of 1-item-sets were built. While performing the second database analysis, unfavorable item-sets whose AUUB is not exactly the base normal utility sum were eliminated .At that point, the database is reconsidered by eliminating every unfavorable item. Presently utility qualities are determined utilizing AU-list arrangement.

### C. Efficient High Normal Utility Pattern Mining (EHAUPM) Calculation:

The calculations introduced till now used the AUUB method to overrate the normal utility of item-sets. The execution of HAUI digger can be enhanced through new approach. Lin et al. [12] introduced two new more tightly upper-limits method. The looser upper-limit method (LUB) acknowledges the existing most extreme utility in exchanges to diminish the upper-limits on the utility of the item-sets. The next upper-limit method overlooks unimportant items in exchanges to additionally fix the upper-limit.

Three eliminating systems were further executed to diminish the inquiry space. The AUUB esteems are incredibly influenced by most utility of the items in an exchange. Despite, this esteem is extremely free on the upper limit of the normal utility of item. Thus, the item might not be high normal items while performing the mining procedure.

Subsequently two new upper limits are established for diminishing the quantity of unfavorable item- sets. An adjusted normal utility (MAU)-list structure is produced to prevent from implementing different database checks, additionally stores necessary data in memory. By considering inquiry capacity in proposed calculation for excavating high normal models are known as a list-tree [10], [12].

In an event the lub(A) estimation of the item-set A cannot be exactly the base high normal utility check, each expansion B of A isn't the HAUI. Subsequently, complete expansions from the item- set A could be overlooked during guaranteeing that all HAUIs could be discovered even now, in this manner saving the fulfillment and rightness of the structured calculation.

## VI. TEST RESULTS

The calculations are executed in programming language java and examinations are done in the PC containing an Intel(R) Core(TM) i7-2600 3.40GHz processor with 8 GB of fundamental memory, running the 64 bit Microsoft Windows 8 operating system. The calculations are performed utilizing RETAIL and CHESS datasets. The implementation time, count of item-sets recovered, including memory space is considered as execution parameters. The outcomes of several minimal high average utility limit esteems of two data sets are represented in the figures below.

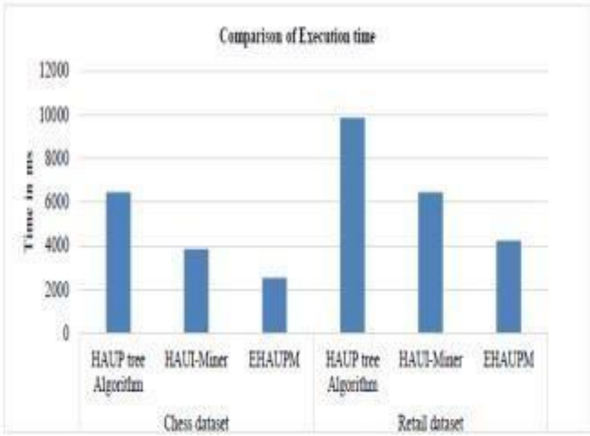


**Fig 1: Comparison of execution time in CHESS and RETAIL dataset**

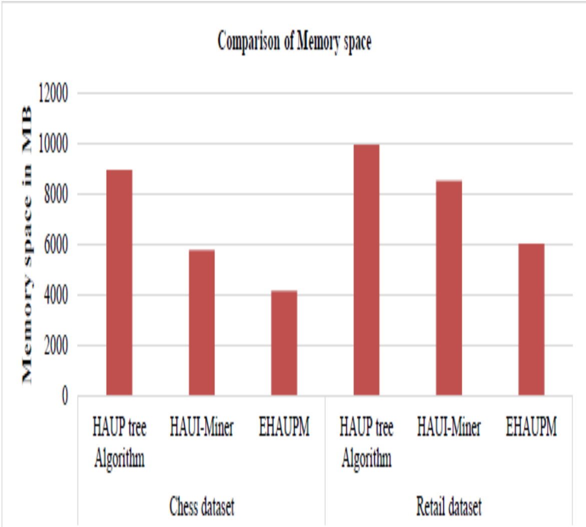


**Fig 2: Comparison of memory space in CHESS and RETAIL dataset**

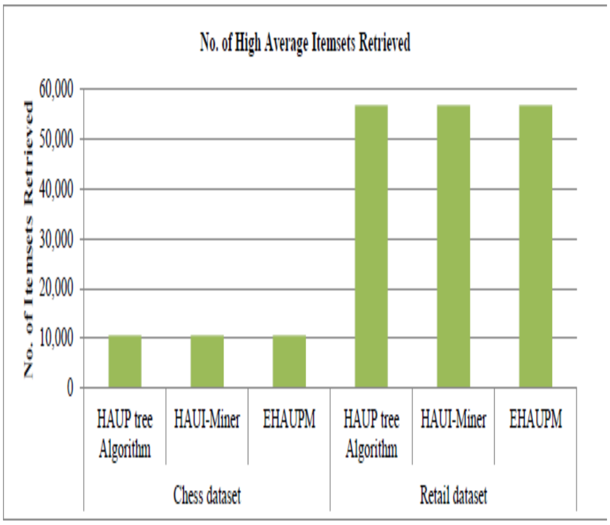


**Fig 3: Comparison of number of item-sets retrieved in CHESS and RETAIL dataset**

From the Figures 1-3, it was discovered that EHAUPM was effective regarding execution time and memory space, when distinguished with the available alternative calculations. This approach devours small memory space and is implemented faster than the other algorithms available.

## VII. CONCLUSION

Customary high-relevant item-set mining considers just the noteworthy relevant estimations of items. Because the relevance of larger item-set is more prominent than the relevance of smaller item-set, conventional relevant mining calculations endure a few disadvantages. Along with these, the customary relevant measure is certainly not a reasonable estimation in genuine applications. For solving this problem, the method of high normal relevant item- set mining has been developed. This method obtains a great deal of consideration because it gives the valuable elective intriguing quality measure to assess the discovered models. By comparing the EHAUPM, HAUI-Miner and HAUP tree approaches it is discovered that EHAUPM approach is efficient. It produces the results effectively in aspects of execution time, memory, scalability.

## REFERENCES


1. Lin CW, Hong TP, Lu WH. Proficiently mining high normal utility item-sets with a tree structure. In Asian gathering on canny data and database frameworks 2010 (pp. 131-9). Springer, Berlin, Heidelberg.
2. Agrawal R, Imieliński T, Swami A. Mining affiliation runs between sets of items in expansive databases. In ACM SIGMOD record 1993 (pp. 207-16). ACM.
3. Ahmed CF, Tanbeer SK, Jeong BS, Lee YK. Proficient tree structures for high utility pattern mining in steady databases. IEEE Transactions on Knowledge and Data Engineering.2009; 21(12):1708-21.
4. Krishna moorthy S. Pruning techniques for mining high utility item-sets. Master Systems with Applications. 2015; 42(5):2371-81.
5. Han J, Pei J, Yin Y, Mao R. Mining successive models without hopeful age: a regular pattern tree approach. Information Mining and Knowledge Discovery. 2004; 8(1):53- 87.
6. Hong TP, Lee CH, Wang SL. Mining high normal utility itemsets. In universal meeting on frameworks, man and artificial intelligence 2009 (pp. 2526-30). IEEE.
7. Lu T, Vo B, Nguyen HT, Hong TP. Another strategy for mining high normal utility item-sets. In IFIP universal meeting on PC data frameworks and mechanical administration 2014 (pp. 33-42). Springer, Berlin, Heidelberg.
8. Lin JC, Li T, Fournier-Viger P, Hong TP, Zhan J, Voznak
9. M. A proficient calculation to mine high average-utility itemsets. Propelled Engineering Informatics. 2016; 30(2):233- 43.
10. Sivamathi C, Vijayarani S. Execution examination of Utility-mining calculations. In worldwide meeting on innovative calculation advances 2016 (pp. 1-4). IEEE.w
11. Lin JC, Li T, Fournier-Viger P, Hong TP, Su JH. Effective mining of high normal utility itemsets with various least limits. In mechanical meeting on information mining 2016 (pp. 14-28). Springer, Cham.
12. J. C.-W. Lin, W. Gan, P. Fournier-Viger, T.-P. Hong, J. Zhan, "Efficient mining of high-utility itemsets using multiple minimum utility thresholds", Knowl.-Based Syst., vol. 13, pp. 100-115, Dec. 2016.
13. J. C.-W. Lin, T. Li, P. Fournier-Viger, T.-P. Hong, J.-H. Su, "Efficient mining of high average-utility itemsets with multiple minimum thresholds", Proc. Ind. Conf. Data Mining, pp. 14-28, 2016.
14. R. Martinez, N. Pasquier, C. Pasquier, "GenMiner: Mining non-redundant association rules from integrated gene expression data and annotations", Bioinformatics, vol. 24, no. 22, pp. 2643- 2644, 2008.
15. B. E. Shie, V. S. Tseng, P. S. Yu, "Online mining of temporal maximal utility itemsets from data streams", Proc. ACM Symp. Appl. Comput.,1622-1626,2010. C. W. Wu, B. E. Shie, V. S. Tseng, P. S. Yu, "Mining top-k high utility itemsets", Proc. ACM SIGKDD Conf. Knowl. Discovery Data Mining, pp. 78-86, 2012
16. C. W. Wu, B. E. Shie, V. S. Tseng, P. S. Yu, "Mining top-k high utility itemsets", Proc. ACM SIGKDD Conf. Knowl. Discovery Data Mining, pp. 78-86, 2012.
17. H. Yao, H. J. Hamilton,C. J. Butz, "A foundational approach to mining itemset utilities from databases", Proc. SIAM Int. Conf. Data Mining, pp. 482- 486, 2004

18. H. Yao, H. J.Hamilton, L. Geng, "A unified framework for utility-based measures for mining itemsets", Proc. Int. Workshop Utility- Based Data Mining, pp. 482-486, 2004. C.-H. Yun, M.-S. Chen, "Using pattern-join and purchase-combination for mining Web transaction models in an electronic commerce environment", Proc. Comput. Softw. Appl. Conf., pp. 99-104, 2000.7
19. S. Zida, P. Fournier-Viger, J. C. W. Lin, C. W. Wu, V. S. Tseng, "EFIM: A highly efficient algorithm for high-, 2015 utility itemset mining", Proc. Mexican Int. Conf. Artif. Intell., pp. 530-546.
20. Y. C. Lin, C. W. Wu, V. S.Tseng, "Mining high utility itemsets in big data", Proc. Pacific Asia Conf. Knowl. Discovery Data Mining, pp. 649-661, 2015.

## AUTHORS PROFILE

**Dr.K.Anuradha** is working as Professor in Computer Science and Engineering in GRIET since 2007. She is also the Dean of Academic Affairs since 2014 in GRIET. She has over 31 years of academic and research experience in Gokaraju Rangaraju Institute of Engineering and Technology and VNRVJIET in Hyderabad. She completed her Ph.D from JNTU, Anantapur in 2011 & Ph.D in Mathematics from JNTU, Hyderabad in 2006. Prior to PhD, she studied Master of Technology in Computer Science from Birla Institute of Technology, Ranchi.Her research interests are Data Mining, Big Data Analytics, Machine Learning and Software Engineering in which she has more than 35 publications in various reputed journals and conferences.. She is top performer in "Data Mining" NPTEL course by IIT Madras. She is also Life Member of ISTE.

E-mail: kodali.anuradha@yahoo.com

**V.Srilakshmi** is working as Assistant Professor in CSE department at GRIET. She is pursuing Ph.D in Computer Science and Engineering from JNTU college of Engineering, Anantapur. She completed M.Tech(Computer Science & Engineering) from TRR College of Engineering, Hyderabad in the year 2011. Her area of research interest includes: Text Mining, Machine Learning, Data Science and Natural Language Processing. She has 12 years of teaching experience. She has published papers in international conferences and journals. She has been certified as Oracle Certified Java Standard Edition 6 Programmer (OCJP) and Oracle Certified Associate (OCA).

She stood one among the Top 1% securing 92% in Cloud Computing course in NPTEL courses. She got qualified in Faculty Eligibility Test (FET) conducted by JNTUH in the year 2013 and State Eligibility Test (SET) conducted by Osmania University in the year 2015

E-mail: potlurisrilakshmi@gmail.com

**Madhuri Bandla** is currently pursuing master's degree program in Computer science and engineering department in GRIET, Hyderabad, India. She has completed her Bachelor of Technology in computer science and engineering department in Swami Ramananda Tirtha Institute of Science and Technology, Nalgonda, India.. Her research areas include DataMining and Utility Mining.

E-mail: Info.bandlamadhuri@mail.com