IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE)
e-ISSN: 2278-1676,p-ISSN: 2320-3331
PP 41-48
www.iosrjournals.org

Improving Utility Mining Based on Dimensionality Reduction

Sreenithya S[1], Sumesh R[2], Bindu S J[3]
[1]*(Department of Computing, Curtin University, Australia)*
[2]*(Department of Equipment & Commodities solutions, Z&R Techno-Commercial Solutions Pty Ltd, Australia)*
[3]*(Department of Electrical and Electronics Engineering, College of Engineering Perumon, India)*

***Abstract**: Lately, utility mining has grown as a developing area in Data Mining field. It remains computationally costly both as far as runtime and memory utilization. Data type, quality, and dimensionality of the dataset are some of the factors which aﬀect the computational cost of utility mining. However, high dimensionality of the datasetcauses more difficulties in managing runtime and memory utilizationduring mining process as compared to the other two factors. A best possible solution could be the dimensionality reduction of the dataset in such a way that it should retain the significant information present in the dataset while discarding the noise.Hence it is a significant challengefor data mining researchers to design a more proficientmethod for extracting high utility itemsets. We develop a novel dimensionality reduction method in this paper by addressing the above issue for improving the efficiency of utility itemset miningwith restricted memory space and processing time.*

I. Introduction

Recently, growth in the database technology leads to a large volume of digital data. This high-dimensional data is a challenge for humans to extract valuable information. Data mining technique helps this task the easiest one by searching relevant information in a high volume of datasets put away in numerous databases, data warehouses or some other data vaults. This method is a profoundly interdisciplinary precinct spreading over from extent disciplines like Statistics, Machine Learning, Databases, Pattern Recognition and others. Numerous data mining methods were technologically advanced for the universe of business, for example for client relationship administration.Traditional data mining systems, have focused to a great extent on distinguishing the statistical relationshipsamong objects which are mostrecurrent in vast transaction databases without plunging the size[1]. Such data mining methods are called as frequent itemset mining[2]; where objects which seem all the most much of the time isaddedmore significance to the customer from the business point of view. Frequent itemsets mining is a central and vital problem in various data mining applications [2].In most of the decision-making areas such as business exchanges, medicinal, security, deceitful exchanges, and retail groups, the frequent itemset mining cannot itself be able to provide the relevant information for making decisions. Consider the instance of a general store;customers buy ovens or washing machines infrequently when contrasted with items like milk, eggs, cleanser, etc. However, the previous transactions return more benefit to the general store. Additionally, the morebenefit frequentthingsare observedas exceptionally helpful in numerous applications. Consider anapplication in the medical field, the unusualgrouping of frequent signs could give helpful bits of knowledge to specialists. In most well-established companies, the manager might be keen on recognizing its most efficient employer i.e. who can contribute a noteworthy portion of the overall business benefit. All of these applications consider utility commitments that can allude to the importance or the price or the look of the items in transactions measured on profit, benefit, deals or whatever other user inclinations more than the frequency of occurrence of things in transactions during the mining process. Such kind of data mining is called Frequent Utility Itemset Mining that can cover all characteristics of financial values and thus helps in the extraction of things that are frequent and have high utility value.

The datasets in this data mining environment are usually large. Modern large datasets are often viewed as large matrices or high dimensional matrices. Mining such datasetscause lots of difficulties in mining process and also consumes much processing cost and time.Reducing the size of such a large dataset without losing significant information is the best solutionfor overcoming such issues. For that, we need to apply a dimensionality reduction algorithm before starting the data mining processing. Dimensionality reduction is a vital study area in numerous applications,for example, pattern recognition, artificial intelligence, data mining and statistics. The primary objective of dimensionality reduction is to lessen the size of thedatasetwith the end goal that the significant information contained in the data should be preserved and the reduced dataset should convey all necessary information. In this paper,a novel concept for reducing the dimensionality of the dataset is presented that includes utility values. This methodology can be extremely valuable much of the time, for example, for making suggestions about which book or DVD to purchase, for searching profoundly hidden mineral residues without penetrating, for investigating the protein structure, for recognizing doubtful messages or calls, and for making sense of what information a document set focussed on.Rest of the sectionsare sorted out in this fashion:the foundation to see how dimensionality drop helps in enhancing the proficiency of utility data mining is disclosed in section 2. Section 3 exhibits the hypothetical foundation to comprehend the proposed calculation. The proposed algorithm of dimensionality reduction in utility data is stated in section 4. Section 5 clarifies the proposed algorithm with illustrations. Experimental evaluation is explained in Section 6.At last, the conclusions and future works are commented in Section 7.

II. Literature Review

Data mining[3]is defined as a process that extracts new, interesting, valuable and potentially useful information and actionable knowledge that is implicit in large databases[4]. Many data-mining applications were developedand designed for the world of business transactions. Traditional applications extract only more frequentitems from large transaction databases without considering the items' utility values such as significance, worth or profit. In most of the recent applications mainly in business and medical, it is necessary to consider those utility values for making decisions in addition to identifying more frequent happenings. By including the utility values and frequency of occurrences of itemsets, a new topic has emerged called frequent utility mining. The datasets in such environment are high dimensional. This kind ofhigh dimensional dataset can badly influence the efficiency of data mining process;that is getting to be the recent most critical issues to solve. It also consumes much processing cost and time. The solution is to includea dimensionality reduction algorithmin the data mining process.In this section, a brief overview, ideas, and methodologies of different utility mining methodsare presentedthat are characterized in variousresearch publications[5].High utility itemset mining has numerous applications, for example, finding collections of products in store transactions that yield the most benefit to the seller [4]. A utility database is defined as a data storein which eachobjectare storedin their amounts and cost per unit. Despite the fact that these methods are regularly exhibited with regards to market basket investigation, there exist different applications.

Yao et al. in [6] characterizes the issue of utility mining appropriately by discovering all itemsets with utility qualities higher than the minimum utility threshold in a transaction database.This laida foundation for future utility mining methods. A frequenthigh utility itemset mining method was illustrated by J. Hu et alin [7] that can recognize a collection of high utility objects,rather than the conventional association rule and frequent itemset mining methods.In the paper [8]**,** H.F. Li presented two proficientone-pass processes for mining high utility itemsets from data streams, MHUI-BIT (Mining High-Utility Itemsets based on BITvector) and MHUI-TID (Mining High-Utility Itemsets based on TIDlist). Two powerful representations of data and an expanded lexicographical tree-based synopsis information structure were produced to enhance the proficiency of mining high utility itemsets. V.S. Tseng et al.[9]developed a novel strategy, to be specific THUI (Temporal High Utility Itemsets),for mining temporal high utility itemsets from data streamsproductively and adequately. Another sort of patterns in amulti-database environment, called Rare Utility Itemsets, developed by G.C.Lan et al.[10]. They proposed a mining method based on profits and quantities as well as normal existing periods and branches of data.TP-RUI-MD (Two-Phase Algorithm for Mining Rare Utility Itemsets in Multiple Databases)[11], another mining approach was developed to find uncommon utility itemsets effectively.HUI-Miner [7] is an another method for finding highutility itemsets containing utility information[10]. High utility itemset mining is a more troublesome issue than frequent itemset mining. Along these lines, highutility itemset mining calculations are for the most part slower than frequent itemset mining calculations. A standout amongst the most productive algorithm for high utility itemset mining[10] is HUI-Miner mining algorithm. Subsequently, in this paper, we apply HUI-Miner calculation for separating high utility itemsets. Be that as it may, as of late the FHM algorithm [10] was appeared to be up to six times quicker than HUI-Miner, particularly for scanty datasets. All the more as of late, the EFIM calculation (2015) was proposed and was illustrated in[10], to beat FHM (2014)[12], HUI-Miner (2012)[13], HUP-Miner (2014)[14].

In this segment, we introduce a short review of different dimensionality reduction designs that have been characterized in different publications[5], [15], [16], [17], [18], [19], [20]. Dimensionality reduction strategies convert the first high-dimensional space into a lower-dimensional space[21]. Thus it can makea big effect, and its outcomes are so specifically appropriate to the discovery of the high utility items. The two most pertinent dimensionality lessening calculations with regards to Utility mining[20] arePrincipal Component Analysis (PCA)[18] and Singular Value Decomposition (SVD)[19]. These methods can be utilized as a preprocessing step during data mining process. PCA is an established measurable strategy to discover patterns in high-dimensional datasets. Another vital dimensionality diminishment calculation in the context of utility mining is SVD. It is a specific understanding of the Matrix Factorization methodology, and it is along these lines additionally identified with PCA. Our contribution is to develop aproficient dimensionality reduction technique forHigh Utility Itemset Mining.L. M. Eric, J. Herik[22] proposed an empirical comparison of all dimensionality reduction methods on five artificial datasets and five original datasets and observed that nonlinear dimensionality reductionmethods are not proficientfor beating conventional linear methods.Many dimensionality reduction methods are currently available. Most of them cannot explain full or nearly comprehensive information about original itemsets. The linear time Closed Itemset Miner (LCM) [11] is known to be the state-of-art algorithm for accomplishing this reduction with no data loss. It can explain wholly or nearly fully information. Since high utility itemset mining is a more troublesome issue than traditional frequent itemset mining[23], so they are slower than frequent itemset mining calculations [23]. If we apply a reduced utility dataset for extracting high utility itemset, then it can improve the speed of high utility itemset mining process. Our proposal focussed on a lin**e**ar dimensionality reduction to reduce high dimensional utility dataset to low dimensional. Thusit can accomplish reduction without any data loss. The reduced dataset we obtained by applying linear dimensionality reduction method can efficiently enhance the execution of high utility itemset mining calculations.

III. Theoretical Background For Dimensionality Reduction Of High Utility Itemset Mining

An itemset is termed as a frequent itemset if it frequently occurs in a transaction database[24]. In numerous applications (particularly in dense database applications)like gene expression studies, network intrusion, web content and usage mining, and so on with long, frequent patterns computation are infeasible. There are two current explanations for the long high-dimensional dataset in themining process. One is to mine only the maximal frequent itemsets[25]. A frequent itemset is said to be maximal on the off chance that it has no frequent superset[26].Be that as it may, help the long, frequent patterns computation easier in dense domains, however, they prompt data loss. Mine only the frequent itemsets that are closed and not maximal is the other solution for making the data mining process of the longhigh-dimensional dataset easier. Frequent closed itemsets extraction can help to perform a dimensionality reduction more efficiently without losing information than maximal frequent itemsets extraction.A frequent itemset is said closed if it has no superset with the same frequency[27]. The proposed approach for implementing dimensionality reduction on utility mining used a concept of finding frequent closed itemsets to reduce dataset size by grouping common items together to form factor items. These factor items are extensively smaller than the original datasets. It can explain full or almost complete information ofrealdatasets. It will take less time to generate rules when compared to frequent itemsets.Our contribution in this paper is to implement the concept of generating frequent closed itemset for reducing transactional utility database thereby enhancing the execution of high utility itemset mining. It is important to audit a few definitions to clarify about high utility itemset mining [17], [28], [29], [30]–[39].

Definition 1: An itemset is an unordered arrangement of dissimilar objects.

Definition 2: The utility of an itemset is the summation of the utility of its items.

Definition 3: The utility of an itemset in a database is the summation of its utility in all transactions where it shows up[17].

Definition 4: A high utility itemset is an itemset with the end goal that its utility is greater than the minimumthreshold utility.

In the accompanying segment, we demonstrate an algorithm for reducing the size of the input transaction database in light of the idea of theextracting frequent closed itemset.

3.1 Dimensionality Reduction Algorithm Based On Discovering Frequent Closed Itemsets

Consider a transaction without any utility information, T={t1, t2, t3, t4, t5, t6}, represents a set of transactions, where each transaction is explained as t1={1, 2, 4, 6, 7, 8}, t2={3, 5, 9}, t3 ={1, 2, 7}, t4={0, 1, 2, 4, 7, 9}, t5={1, 2, 4, 6, 7, 8}, t6={0, 3, 4, 5, 9}[23]. The support of an item is defined as the number of transactions that contain that item[23]. Since itemset {1, 2}appears in four transactions, it has a support value of four. Consider a minimum threshold support value of 40 %, then the frequent closed itemsets from the above transaction database are {1 2 7}, {1 2 4 6 7 8}, {3 5 9}, {0 4 9}. Thus we can replace the transaction dataset T with a factorized transaction F={R1, R2, R3, R4, R5, R6} with its factor items B={ B1, B2, B3, B4}where each factorized transaction set[40]is explained as R1={B2}, R2={B3}, R3={B1}, R4={B1,B4}, R5={B2}, R6={B3,B4}where factor items B1, B2, B3 and B4 are the set of frequent closed itemsets that are defined as B1={1, 2, 7}, B2={1, 2, 4, 6, 7, 8}, B3={3, 5, 9} and B4={0, 4, 9}. Thus the original transaction dataset T is reduced to R1={2}, R2={3}, R3={1}, R4={1, 4}, R5={2}, R6={3, 4}. To computethe degree of reduction of original transaction $T1,...,Tm$ by factorized transaction $R1,..., Rm$provided that $Ti \supseteq \cup Ri$ holds, an approximation degree formula [40] is defined of $T1,...,Tm$ by $R1,...,Rm$as

$$\frac{\sum_{i=1}^{m} |\cup \text{Ri}|}{\sum_{i=1}^{m} |\text{Ti}|} \qquad (1)$$

Usually, approximation degrees are expressed as percents. So 90% approximation degree means that R1, R2, …., Rn contains 90% of items contained in T1, T2, …., Tn. The degree of reduction for the above example is 100%. That means there is no loss of data after reduction. If we used frequent itemsets instead of frequent closed itemsets for reduction, then there will be 15-factor items instead of 4-factor items. It shows that the frequent itemsets setare much bigger than the frequent closed itemsets set. Discovering such frequent closed itemsets set creates a much smaller dataset with no loss of data[23].In the next section, an algorithm is presentedthat can augmentthe performance of high utility mining algorithm by reducing the size of the input transaction database that contains utility information.

IV. Proposed Algorithm For Dimensionality Reduction Of High Utility Itemset Mining

Normally in data mining process to have not just a dataset with elements that characterize a high dimensional space but have extremely sparse data in that. Dimensionality reductions strategies, changing the first high-dimensional space into a lower-dimensionality, conquer this issue. Applying dimensionality diminishment has such an impact, and its results are so clearly germane to the extraction of high utility itemsets.

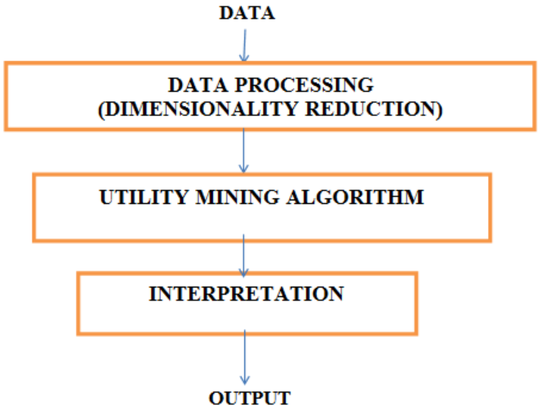

Figure 1 Flowchart for Dimensionality Reduction of High Utility itemset mining

Algorithm:
Given a Transaction Utility Database
Step 1:Apply aUtility Mining algorithm called HUI-Miner[23]on the given database. The result of this mining process is a collection of itemset**s** that have utility greater than theminimum threshold utility specified by theuser.
Step 2: Apply theproposed dimensionality reduction algorithmwith the concept of generating frequent closed itemsets for reducing the size of transactional utility database.
Step 3: Apply the Utility Mining algorithmcalled HUI-Miner that we used in step 1on the result obtained from step 2.
Step 4: Compare the high utility itemsets obtained from step 1 and step3
Step 5: Evaluate the performance of step 1 and step 3.
In the next section, we explain the algorithm in detail with examples.

V. Explanation Of The Proposed Algorithm With Examples

Consider for example a sample retailer transaction database[41]shown in TABLE 1 with theexternal utility value of each productshown in TABLE 2. Eachvalue in everyrowdemonstrate the amount of productspurchased in a transaction, and thelast column contains the transaction utility with the aggregatetransaction utility of the database shown in the last row[42].

Table 1 A Sample Transaction Database of a retailer

TID\ITEM ID	A	B	C	D	E	F	Transaction Utility
T1	2	0	1	1	0	0	80
T2	2	1	1	0	0	0	195
T3	0	0	1	1	10	0	110
T4	0	1	0	0	15	0	225
T5	1	0	1	1	0	1	72
T6	2	0	0	1	10	0	105
T7	2	0	0	0	8	1	62
T8	1	1	0	1	2	0	205
T9	1	0	0	1	10	0	95
T10	1	1	0	0	5	0	185
Total	120	600	100	210	300	4	1334

Table2 Unit Profit for each Items

PRODUCT ID	PRICE($)
A	10
B	150
C	25
D	35
E	5
F	2

Thus TABLE 1can be represented in the format as item ids: transaction utility: item utilities so T1 is equivalent to 1 3 4:80:20 25 35. Similarly all other transactions are represented as 1 2 3:195:20 150 25, 3 4 5:110:25 35 50, 2 5:225:150 75, 1 3 4 6:72:10 25 35 2, 1 4 5:105:20 35 50, 1 5 6:62:20 40 2, 1 2 4 5:205:10 150 35 10, 1 4 5:95:10 35 50 and 1 2 5:185:10 150 25. For identifying the quantity of each items in each transactions, the transaction database in TABLE 2are represented with quantity of each item in bracket. So T1 is defined as 1(2) 3(1) 4(1). Similarly all other transactions are defined as follows 1(2) 2(1) 3(1), 3(1) 4(1) 5(10), 2(1) 5(15), 1(2) 3(1) 4(1) 6(1), 1(2) 4(1) 5(10), 1(2) 5(8) 6(1), 1(1) 2(1) 4(1) 5(2), 1(1) 4(1) 5(10) and1(1) 2(1) 5(5). In the next paragraphs we explain each steps of the algorithm based on TABLE1.

Step 1: Apply HUI-Miner –High Utility Itemset Mining Algorithmon the transactional database shown in TABLE 1.

By considering the diversevaluations of each item as utilities, utility mining concentrates on recognizing the itemsets with high utilities. As " downward closure property"[43] doesn't matter to utility extraction, the candidate itemsetsgenerationis the most expensive as far as time and memory space. In this paper, we applyan HUI-Miner calculation to prune down the quantity of candidate itemsets effectively and can decisively acquire the complete collection of high utility itemsets. In the primary stage, we propose a model that applies the " transaction-weighted downward closure property" on the pursuit space to speed up the candidate itemsets extraction. Transaction-Weighted downward closure property shows that any superset of a low transaction weighted usage itemset is low in weighted transaction utilization. Transaction weighted Utility of an itemset X, indicated as TWU(X) and is the total of the Transactionutilities of the considerable number of Transactions containing X. For a given itemset X, X is a high Transaction weighted utilization itemset if TWU(X)≥ ε', where ε' is the minimum threshold utility specified by theuser.

Table 3 Transaction Weighted Utility

ITEM INDEX	1	2	3	4	5
ITEM SET	E	A	B	D	C
PROFIT	5	10	150	35	25
QUANTITY	60	12	4	5	4
TWU	987	964	810	565	422

In the next stage, one extra scan is madefor recognizing high utility itemsets. Thus without reducing the dimensionality we got the output of HUI-Miner as

2:600
2 5:560
2 1:490

i.e.,the set of high utility itemsetsthat have utility greater than the minimum threshold utility**(** *min_utility)* specified by theuser(Assume *min_utility* threshold as 400). In retailer point of view, the product B, the combination of productsB and E, and the combination of products A and B give high profit to the seller after comparing with other products and other combination of products.

Step 2: Apply the proposed dimensionality reduction method in light of the idea offrequent closed itemset generation for reducing the size of transactional utility database.

This process is performed in 2 phases:

Phase1: Dimensionality reduction of utility transaction database shown in TABLE 1 without considering their utility values.

Based on the dimensionality reduction algorithm of discovering frequent closed itemsets explained in section 3.1, the original transaction utility database shown in TABLE 1 without considering their utility values are reduced to a smaller databaseas shown in Fig. 1. Factor items used for this reduction are: B1= {1}, B2= {1,4,5}, B3= {1,2}, B4= {1,2,3}, B5= {1,3,4,6}.

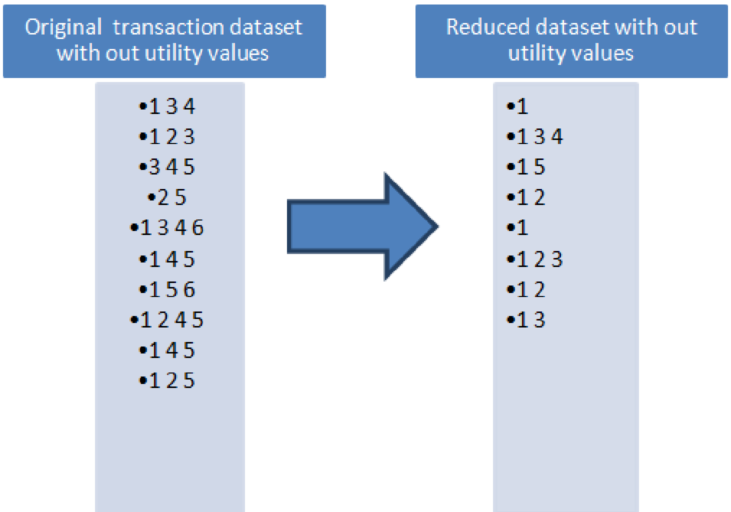

Figure2 Reduction without utility values

In the nextphase, we reduce the utility information associated with each transaction set.
Phase 2: Dimensionality reduction of the utility values associated with each transaction set.
Considereach row of reduced dataset andfollow the steps below:

- Find the union of factor items obtained in phase 1for each row of the reduced dataset to identify their original dataset.Fore.g. consider third row {1, 5}, the factor items are 1= {1}, 5= {1, 3, 4, 6}. Their union is {1, 3, 4, 6}, that means this row is the reduced representation of 5th transaction set in the original database, i.e. {1, 3, 4, 6}.
- Obtain the reduced utility information for each row. For obtaining the reduced utility data, multiply each element in the factor items of that row with their quantity and profit. Fore.g. consider third row {1, 5},their reduced utility values can be obtained as 1x2x10, 1x2x10+3x1x25+4x1x35+6x1x2. So the reduced utility value of {1, 5} is 20, 247. Transaction utility of reduced set is obtained by adding their individual utilities. So transaction utility of this reduced transaction set {1, 5} is 267.

So the original transaction, 1 3 4 6:72:10 25 35 2 is reduced to 1 5:279:20 259.
Thus, we can reduce the original transaction database with utility valuesof a retailer shown in TABLE 1 to a reduced databaseillustrated in Fig. 4.

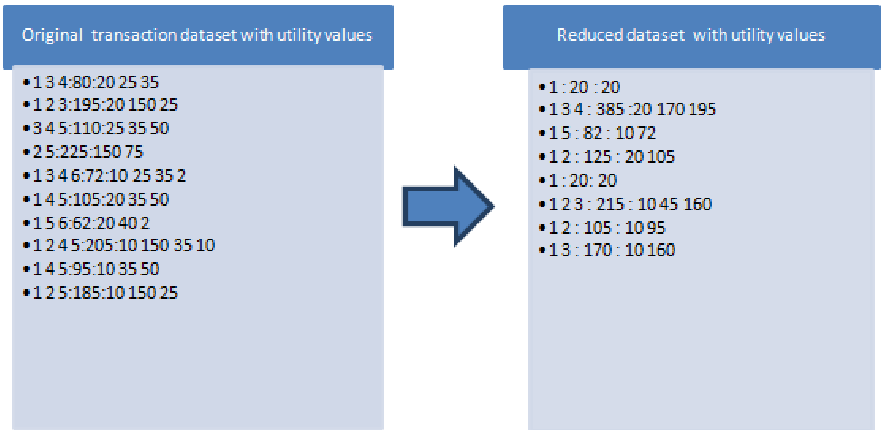

Figure 3 Reduced Database

Step 3: Apply **HUI-Miner** that we used in step 1 on the result obtained from step 2.
By considering the " transaction-weighted downward closure property " to speed up the itemsets generation, we acquire the product of HUI-Miner as
4 3:365
4 3 1:385
3:490

3 1:530 that means the factor items B3, B4, B1= {1}, B3= {1, 2}, B4= {1, 4, 5} ie, the items A, B and E as itself or its combinations gives a high profit to the retailer.

Step 4: Compare the high utility itemsets obtained from step 1 and step3

The output of step 3 and step 1 are same. That means there is no loss of relevant data during reduction.

Step 5: Evaluate the performance of step 1 and step 3.

VI. Experimental Evaluation

To assess the contrasts between the utility itemsets created from a vast high dimensional database environment with the one from a dimensionality reduced database environment, we conducted a number of experiments in various user defined parameters. Likewise, the proficiency of the proposed calculation is assessed by differing different parameters**.** The modelis executed in J2SDK 1.5.0. The proposed algorithm is conductedona machine with 3.0 GHz CPU and 8 GB memory. The experimental data used for evaluation is availableat theutility itemset mining implementations repository,FoodMart2000, Microsoft Developer Network (MSDN),NU-MineBench version 2.0 dataset and technical report.

VII. Conclusion

Our work introduces a novel way to deal with the dimensionality reduction of utility data for enhancing the execution of utility mining in whicheach item's utility values are permitted to be dynamic in a predefined timeframe, not at all like conventional methodologies where these values are static within the timeframe. Also as a future work, our methodology can be further improved by incorporating a fuzzy model where utilities are defined as fuzzy values. Accordingly, it can build up a proficient and related technique to real-lifeinformation and can catch real-worldconditions in fuzzy utility mining.

References

[1]. S. Bhattacharya and D. Dubey, "High Utility Itemset Mining," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 8, pp. 476–481, 2012.

[2]. M. Y. Khot and A. P. M. M. Kulkarni, "Survey on High Utility Itemset Mining from Large Transaction Databases."

[3]. "Data Mining - Term Paper." [Online]. Available: http://www.termpaperwarehouse.com/essay-on/Data-Mining/110035. [Accessed: 15-Sep-2016].

[4]. P. Drineas and R. Kannan, "Pass efficient algorithms for approximating large matrices.," in *SODA*, 2003, vol. 3, pp. 223–232.

[5]. S. Arora, E. Hazan, and S. Kale, "A fast random sampling algorithm for sparsifying matrices," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, Springer, 2006, pp. 272–279.

[6]. H. Yao, H. Hamilton, and C. Butz, "A Foundational Approach to Mining Itemset Utilities from Databases," in *Proceedings of the 2004 SIAM International Conference on Data Mining*, 0 vols., Society for Industrial and Applied Mathematics, 2004, pp. 482–486.

[7]. J. Hu and A. Mojsilovic, "High-utility pattern mining: A method for discovery of high-utility item sets," *Pattern Recognit.*, vol. 40, no. 11, pp. 3317–3324, Nov. 2007.

[8]. H.-F. Li, H.-Y. Huang, and S.-Y. Lee, "Fast and memory efficient mining of high-utility itemsets from data streams: with and without negative item profits," *Knowl. Inf. Syst.*, vol. 28, no. 3, pp. 495–522, Jul. 2010.

[9]. C.-J. Chu, V. S. Tseng, and T. Liang, "An efficient algorithm for mining temporal high utility itemsets from data streams," *J. Syst. Softw.*, vol. 81, no. 7, pp. 1105–1117, Jul. 2008.

[10]. G.-C. Lan, T.-P. Hong, and V. S. Tseng, "Mining Rare-Utility Itemsets in a Multi-Database Environment."

[11]. G.-C. Lan, T.-P. Hong, and V. S. Tseng, "A novel algorithm for mining rare-utility itemsets in a multi-database environment," in *Proceedings of the 26th workshop on combinatorial mathematics and computation theory*, 2009, pp. 293–302.

[12]. P. Fournier-Viger, C.-W. Wu, S. Zida, and V. S. Tseng, "FHM: Faster high-utility itemset mining using estimated utility co-occurrence pruning," in *International Symposium on Methodologies for Intelligent Systems*, 2014, pp. 83–92.

[13]. M. Liu and J. Qu, "Mining high utility itemsets without candidate generation," in *Proceedings of the 21st ACM international conference on Information and knowledge management*, 2012, pp. 55–64.

[14]. S. Zida, P. Fournier-Viger, J. C.-W. Lin, C.-W. Wu, and V. S. Tseng, "EFIM: a highly efficient algorithm for high-utility itemset mining," in *Mexican International Conference on Artificial Intelligence*, 2015, pp. 530–546.

[15]. I. K. Fodor, *A survey of dimension reduction techniques*. Technical Report UCRL-ID-148494, Lawrence Livermore National Laboratory, 2002.

[16]. B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Application of Dimensionality Reduction in Recommender System - A Case Study," Jul. 2000.

[17]. J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[18]. I. Jolliffe, "Principal Component Analysis," in *Wiley StatsRef: Statistics Reference Online*, John Wiley & Sons, Ltd, 2014.

[19]. L. De Lathauwer, B. De Moor, and J. Vandewalle, "A Multilinear Singular Value Decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, Jan. 2000.
[20]. M. E. Wall, A. Rechtsteiner, and L. M. Rocha, "Singular Value Decomposition and Principal Component Analysis," in *A Practical Approach to Microarray Data Analysis*, D. P. Berrar, W. Dubitzky, and M. Granzow, Eds. Springer US, 2003, pp. 91–109.
[21]. A. M. Roumani, D. B. Skillicorn, and others, "A Dimensionality Reduction Technique for Collaborative Filtering," 2006.
[22]. L. Van Der Maaten, E. Postma, and J. Van den Herik, "Dimensionality reduction: a comparative," *J Mach Learn Res*, vol. 10, pp. 66–71, 2009.
[23]. P. Fournier-Viger, J. C.-W. Lin, A. Gomariz, T. Gueniche, A. Soltani, Z. Deng, and H. T. Lam, "The SPMF Open-Source Data Mining Library Version 2," in *Machine Learning and Knowledge Discovery in Databases*, B. Berendt, B. Bringmann, É. Fromont, G. Garriga, P. Miettinen, N. Tatti, and V. Tresp, Eds. Springer International Publishing, 2016, pp. 36–40.
[24]. C.-W. Lin, T.-P. Hong, G.-C. Lan, J.-W. Wong, and W.-Y. Lin, "Efficient updating of discovered high-utility itemsets for transaction deletion in dynamic databases," *Adv. Eng. Inform.*, vol. 29, no. 1, pp. 16–27, Jan. 2015.
[25]. M. J. Zaki, "Mining Closed & Maximal Frequent Itemsets," *NSF CAREER Award IIS-0092978 DOE Early Career Award -FG02-02ER25538 NSF Grant EIA-0103708*, 1.
[26]. O. R. Zaiane, M. El-Hajj, Y. Li, and S. Luk, "Scrutinizing Frequent Pattern Discovery Performance," in *21st International Conference on Data Engineering (ICDE'05)*, 2005, pp. 1109–1110.
[27]. J. W. Joe and S. P. S. Ibrahim, "A Novel Approach for High Utility Closed Itemset Mining with Transaction Splitting," in *Proceedings of the 3rd International Symposium on Big Data and Cloud Computing Challenges (ISBCC – 16')*, V. Vijayakumar and V. Neelanarayanan, Eds. Springer International Publishing, 2016, pp. 307–315.
[28]. T. P. Hong, C. H. Lee, and S. L. Wang, "An Incremental Mining Algorithm for High Average-Utility Itemsets," in *2009 10th International Symposium on Pervasive Systems, Algorithms, and Networks*, 2009, pp. 421–425.
[29]. W. Gan, J. C.-W. Lin, P. Fournier-Viger, and H.-C. Chao, "More Efficient Algorithms for Mining High-Utility Itemsets with Multiple Minimum Utility Thresholds," in *Database and Expert Systems Applications*, S. Hartmann and H. Ma, Eds. Springer International Publishing, 2016, pp. 71–87.
[30]. Q.-H. Duong, B. Liao, P. Fournier-Viger, and T.-L. Dam, "An efficient algorithm for mining the top-k high utility itemsets, using novel threshold raising and pruning strategies," *Knowl.-Based Syst.*, vol. 104, pp. 106–122, Jul. 2016.
[31]. P. Fournier-Viger, J. C.-W. Lin, C.-W. Wu, V. S. Tseng, and U. Faghihi, "Mining Minimal High-Utility Itemsets," in *Database and Expert Systems Applications*, S. Hartmann and H. Ma, Eds. Springer International Publishing, 2016, pp. 88–101.
[32]. S.-J. Yen and Y.-S. Lee, "Mining High Utility Patterns in Different Time Periods," in *Industrial Engineering, Management Science and Applications 2015*, vol. 349, M. Gen, K. J. Kim, X. Huang, and Y. Hiroshi, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 779–789.
[33]. H. Yao, H. J. Hamilton, and L. Geng, "A unified framework for utility-based measures for mining itemsets," in *Proc. of ACM SIGKDD 2nd Workshop on Utility-Based Data Mining*, 2006, pp. 28–37.
[34]. P. S. M. Tsai, "Mining frequent itemsets in data streams using the weighted sliding window model," *Expert Syst. Appl.*, vol. 36, no. 9, pp. 11617–11625, Nov. 2009.
[35]. H. Ryang and U. Yun, "High utility pattern mining over data streams with sliding window technique," *Expert Syst. Appl.*, vol. 57, pp. 214–231, Sep. 2016.
[36]. M. Liu and J. Qu, "Mining High Utility Itemsets Without Candidate Generation," in *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, New York, NY, USA, 2012, pp. 55–64.
[37]. M.-Y. Lin, T.-F. Tu, and S.-C. Hsueh, "High utility pattern mining using the maximal itemset property and lexicographic tree structures," *Inf. Sci.*, vol. 215, pp. 1–14, Dec. 2012.
[38]. H.-F. Li, "MHUI-max: An efficient algorithm for discovering high-utility itemsets from data streams," *J. Inf. Sci.*, vol. 37, no. 5, pp. 532–545, Oct. 2011.
[39]. G.-C. Lan, T.-P. Hong, and V. S. Tseng, "Discovery of high utility itemsets from on-shelf time periods of products," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5851–5857, May 2011.
[40]. P. Krajca, J. Outrata, and V. Vychodil, "Using Frequent Closed Itemsets for Data Dimensionality Reduction," in *2011 IEEE 11th International Conference on Data Mining*, 2011, pp. 1128–1133.
[41]. A. Erwin, R. P. Gopalan, and N. R. Achuthan, "A Bottom-up Projection Based Algorithm for Mining High Utility Itemsets," in *Proceedings of the 2Nd International Workshop on Integrating Artificial Intelligence and Data Mining - Volume 84*, Darlinghurst, Australia, Australia, 2007, pp. 3–11.
[42]. T. Nadu, "ADVANCE MINING OF HIGH UTILITY ITEMSETS IN TRANSACTIONAL DATA."
[43]. Y. Liu, W. Liao, and A. Choudhary, "A Fast High Utility Itemsets Mining Algorithm," in *Proceedings of the 1st International Workshop on Utility-based Data Mining*, New York, NY, USA, 2005, pp. 90–99.