IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 8, August 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 8, August 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0908009 |

Improving Utility Mining Based On Principal Component Analysis: A Review

Swati Nigam[1], Ruchika Pachori[2]

Department of Information Technology, Mahakal Institute of Technology, Ujjain, RGPV Bhopal, 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 principal component analysis 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 challenge for data mining researchers to design a more proficient method for extracting high utility itemsets. We develop a novel principal component analysis method in this paper by addressing the above issue for improving the efficiency of utility itemset mining with 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 utilityvalue.

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 principal component analysis algorithm before starting the data mining processing. principle component analysis is a vital study area in numerous applications,for example, pattern recognition, artificial intelligence, data mining and statistics. The primary objective of principle component analysis 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,formakingsuggestionsaboutwhichbookorgadget

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 8, August 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0908009 |

topurchase,forsearchingprofoundlyhiddenmineral residues without penetrating, for investigating the protein structure, for recognizing doubtful messages or calls, and for making sense of what information a document set focused on.

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 differentapplications.

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] isanother 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 FHMalgorithm[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 for High 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],

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 8, August 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0908009 |

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 focused on a linear dimensionality reduction to reduce high dimensional utility dataset to low dimensional. Thus, it 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 miningcalculations.

III. PROPOSED METHODOLOGY

Principal Component Analysis, or PCA, is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

Reducing the number of variables of a data set naturally comes at the expense of accuracy, but the trick in dimensionality reduction is to trade a little accuracy for simplicity. Because smaller data sets are easier to explore and visualize and make analyzing data much easier and faster for machine learning algorithms without extraneous variables to process.

So, to sum up, the idea of PCA is simple; reduce the number of variables of a data set, while preserving as much information as possible.

STEP 1: STANDARDIZATION

The aim of this step is to standardize the range of the continuous initial variables so that each one of them contributes equally to the analysis.

More specifically, the reason why it is critical to perform standardization prior to PCA, is that the latter is quite sensitive regarding the variances of the initial variables. That is, if there are large differences between the ranges of initial variables, those variables with larger ranges will dominate over those with small ranges (For example, a variable that ranges between 0 and 100 will dominate over a variable that ranges between 0 and 1), which will lead to biased results. So, transforming the data to comparable scales can prevent this problem.

Mathematically, this can be done by subtracting the mean and dividing by the standard deviation for each value of each variable.

$$z = \frac{value - mean}{standard\ deviation}$$

Once the standardization is done, all the variables will be transformed to the same scale.

STEP 2: COVARIANCE MATRIX COMPUTATION

The aim of this step is to understand how the variables of the input data set are varying from the mean with respect to each other, or in other words, to see if there is any relationship between them. Because sometimes, variables are highly correlated in such a way that they contain redundant information. So, in order to identify these correlations, we compute the covariance matrix.

The covariance matrix is a $p \times p$ symmetric matrix (where p is the number of dimensions) that has as entries the covariances associated with all possible pairs of the initial variables. For example, for a 3-dimensional data set with 3 variables x, y, and z, the covariance matrix is a 3×3 matrix of this from:

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 8, August 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0908009 |

$$\begin{bmatrix} Cov(x,x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Cov(y,y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Cov(z,z) \end{bmatrix}$$

Covariance Matrix for 3-Dimensional Data

Since the covariance of a variable with itself is its variance (Cov(a,a)=Var(a)), in the main diagonal (Top left to bottom right) we actually have the variances of each initial variable. And since the covariance is commutative (Cov(a,b)=Cov(b,a)), the entries of the covariance matrix are symmetric with respect to the main diagonal, which means that the upper and the lower triangular portions are equal.

What do the covariances that we have as entries of the matrix tell us about the correlations between the variables?

It's actually the sign of the covariance that matters:

- if positive then: the two variables increase or decrease together (correlated)
- if negative then: One increases when the other decreases (Inversely correlated)

Now, that we know that the covariance matrix is not more than a table that summaries the correlations between all the possible pairs of variables, let's move to the next step.

STEP 3: EIGENVECTORS AND EIGENVALUES

Eigenvectors and eigenvalues are the linear algebra concepts that we need to compute from the covariance matrix in order to determine the principal components of the data. Before getting to the explanation of these concepts, let's first understand what do we mean by principal components.
Principal components are new variables that are constructed as linear combinations or mixtures of the initial variables. These combinations are done in such a way that the new variables (i.e., principal components) are uncorrelated and most of the information within the initial variables is squeezed or compressed into the first components. So, the idea is 10-dimensional data gives you 10 principal components, but PCA tries to put maximum possible information in the first component, then maximum remaining information in the second and so on, until having something like shown in the scree plot below.

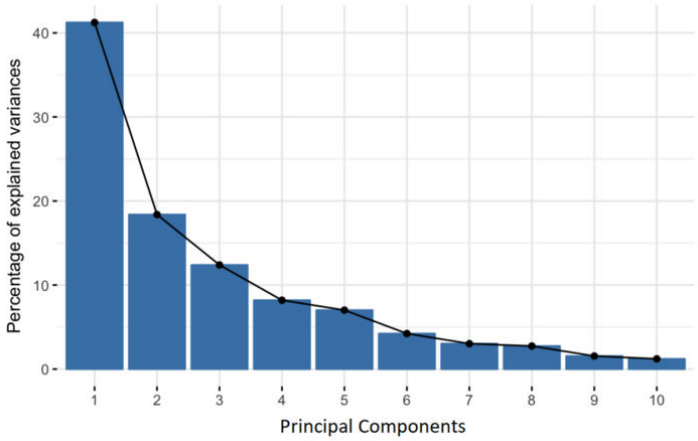

Fig1. Percentage of Variance (Information) for each by PC

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 8, August 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0908009 |

Organizing information in principal components this way, will allow you to reduce dimensionality without losing much information, and this by discarding the components with low information and considering the remaining components as your new variables.

An important thing to realize here is that, the principal components are less interpretable and don't have any real meaning since they are constructed as linear combinations of the initial variables.

Geometrically speaking, principal components represent the directions of the data that explain a **maximal amount of variance**, that is to say, the lines that capture most information of the data. The relationship between variance and information here, is that, the larger the variance carried by a line, the larger the dispersion of the data points along it, and the larger the dispersion along a line, the more the information it has. To put all this simply, just think of principal components as new axes that provide the best angle to see and evaluate the data, so that the differences between the observations are better visible.

As there are as many principal components as there are variables in the data, principal components are constructed in such a manner that the first principal component accounts for the **largest possible variance** in the data set. For example, let's assume that the scatter plot of our data set is as shown below, can we guess the first principal component? Yes, it's approximately the line that matches the purple marks because it goes through the origin and it's the line in which the projection of the points (red dots) is the most spread out. Or mathematically speaking, it's the line that maximizes the variance (the average of the squared distances from the projected points (red dots) to the origin).

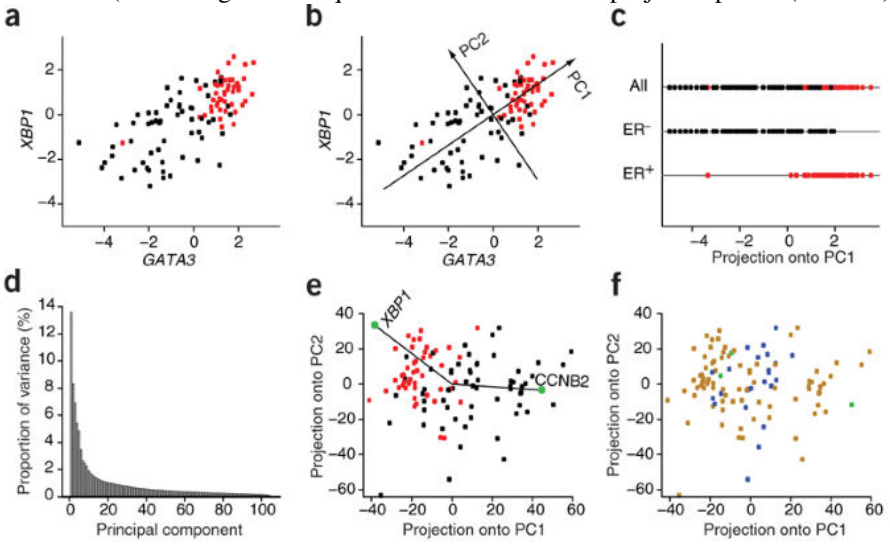

Fig2. Plot of Principle Components

The second principal component is calculated in the same way, with the condition that it is uncorrelated with (i.e., perpendicular to) the first principal component and that it accounts for the next highest variance.

This continues until a total of p principal components have been calculated, equal to the original number of variables.

Now that we understood what we mean by principal components, let's go back to eigenvectors and eigenvalues. What you firstly need to know about them is that they always come in pairs, so that every eigenvector has an eigenvalue. And their number is equal to the number of dimensions of the data. For example, for a 3-dimensional data set, there are 3 variables, therefore there are 3 eigenvectors with 3 corresponding eigenvalues.

Without further ado, it is eigenvectors and eigenvalues who are behind all the magic explained above, because the eigenvectors of the Covariance matrix are actually the directions of the axes where there is the most variance(most information) and that we call Principal Components. And eigenvalues are simply the coefficients attached to eigenvectors, which give the amount of variance carried in each Principal Component.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 8, August 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0908009 |

By ranking your eigenvectors in order of their eigenvalues, highest to lowest, you get the principal components in order of significance.

Example:
let's suppose that our data set is 2-dimensional with 2 variables x,y and that the eigenvectors and eigenvalues of the covariance matrix are as follows:

$$v1 = \begin{bmatrix} 0.6778736 \\ 0.7351785 \end{bmatrix} \qquad \lambda_1 = 1.284028$$

$$v2 = \begin{bmatrix} -0.7351785 \\ 0.6778736 \end{bmatrix} \qquad \lambda_2 = 0.04908323$$

If we rank the eigenvalues in descending order, we get λ1>λ2, which means that the eigenvector that corresponds to the first principal component (PC1) is v1 and the one that corresponds to the second component (PC2) isv2.

After having the principal components, to compute the percentage of variance (information) accounted for by each component, we divide the eigenvalue of each component by the sum of eigenvalues. If we apply this on the example above, we find that PC1 and PC2 carry respectively 96% and 4% of the variance of the data.

STEP 4: FEATURE VECTOR

As we saw in the previous step, computing the eigenvectors and ordering them by their eigenvalues in descending order, allow us to find the principal components in order of significance. In this step, what we do is, to choose whether to keep all these components or discard those of lesser significance (of low eigenvalues), and form with the remaining ones a matrix of vectors that we call Feature vector.
So, the feature vector is simply a matrix that has as columns the eigenvectors of the components that we decide to keep. This makes it the first step towards dimensionality reduction, because if we choose to keep only p eigenvectors (components) out of n, the final data set will have only p dimensions.

Example:
Continuing with the example from the previous step, we can either form a feature vector with both of the eigenvectors v1 and v2:

$$\begin{bmatrix} 0.6778736 & -0.7351785 \\ 0.7351785 & 0.6778736 \end{bmatrix}$$

Or discard the eigenvector v2, which is the one of lesser significance, and form a feature vector with v1 only:

$$\begin{bmatrix} 0.6778736 \\ 0.7351785 \end{bmatrix}$$

Discarding the eigenvector v2 will reduce dimensionality by 1, and will consequently cause a loss of information in the final data set. But given that v2 was carrying only 4% of the information, the loss will be therefore not important and we will still have 96% of the information that is carried by v1.

So, as we saw in the example, it's up to you to choose whether to keep all the components or discard the ones of lesser significance, depending on what you are looking for. Because if you just want to describe your data in terms of new variables (principal components) that are uncorrelated without seeking to reduce dimensionality, leaving out lesser significant components is not needed.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 8, August 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0908009 |

STEP 5: RECAST THE DATA ALONG THE PRINCIPAL COMPONENTS AXES

In the previous steps, apart from standardization, you do not make any changes on the data, you just select the principal components and form the feature vector, but the input data set remains always in terms of the original axes (i.e, in terms of the initial variables).

In this step, which is the last one, the aim is to use the feature vector formed using the eigenvectors of the covariance matrix, to reorient the data from the original axes to the ones represented by the principal components (hence the name Principal Components Analysis). This can be done by multiplying the transpose of the original data set by the transpose of the feature vector.

$$FinalDataSet = FeatureVector^T * StandardizedOriginalDataSet^T$$

IV. CONCLUSION

Our work introduces a novel way to deal with the principal component analysis 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 utilitymining.

REFERENCES

[1]. S.BhattacharyaandD.Dubey,"HighUtilityItemsetMining,"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.DrineasandR.Kannan,"Passefficientalgorithmsforapproximatinglargematrices.,"inSODA,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 SystemA 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.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 8, August 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0908009 |

[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," inA 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 SystAppl., 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," in2011 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.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details