iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IX **Month of publication:** September 2018

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IX, Sep 2018- Available at www.ijraset.com

A Survey of Need of Utility Based Data Mining For Product Recommendation

Iqura Khan[1], Megha Singh[2]
[1]*Research Scholar- COE-CSE,* [2]*Asst. Professor-COE-CSE, Dr. APJ Abdul Kalam University Indore*

Abstract: ***Utility based item set mining is the significant task of data mining. Most of the algorithms does not consider transaction addition and deletion. Data structure needs to be rebuild in every update and after it complete database needs to be scanned. Data is collected in convenient manner in our work. In this work multi dimensional support will be provided which is in three dimensions and it significantly focused on frequency. On users interest and dimensions algorithms are evaluated. Interest of users is based on the dimension on which he selected. This work will decrease sparsity of data. High efficiency is evaluated with justified weight. It also increases popularity and frequency of item and compared it with existing work. Searching can be avoided by balanced B+ tree which is used in our work because it saves time consumption and it is easy to storage and have easy sorting.***

Keywords : B+ tree, High utility item set mining, transactions and multi-dimensions.

I. INTRODUCTION

This work is based on super market based utility mining with high utility item set mining and it performs better in marketing and promotion of business and organizations with classifying analysis and evaluating hypermarkets. Database stores all the item sets using high utility based item set mining. When comparing it with minimum threshold there utilities are not less. Those items which are generated in database are known as high utility item set. HUIM algorithm is generated on the basis of static database with not considering any type of addition and deletion of transaction. Whenever database required to be updated it means that data structure is rebuild by scanning the complete database of high utility item set mining. After this complete work the previous data will not be used and only the data which is updated will be used. A tree structured is used to overcome the limitation of updating database which is called as IHUP-Tree. The proposals are increasing because of interactive HUIM. Whenever operations are performed then IHUP will efficiently adjusted with the incremental of IHUP-Tree on HUIM.

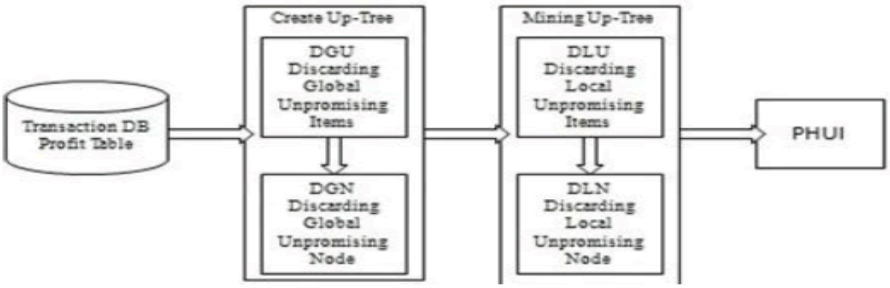


Figure 1: High Utility Itemset Mining

IHUP-Tree algorithm is based in two phases: in first phase over-estimation is adopted and calculated for the itemset database utility by upper bound. In second phase, database scanning are verified with the limitation of large database with consuming more time for the verification of large data.

II. RELATED WORK

A. Study of Base Paper

Shiming Guo et al. In[1] described about significant role of data model with using high utility item set mining. They explores that most of the research does not considers any type of changes like modification or editing in analytical model. Utility mining mainly supports lead generation and the outdated transactions it can become wrong. To resolve with this issues author proposed IHUP-Tree in spite of conventional HUIM.

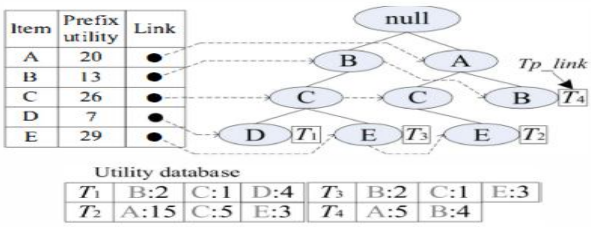


Figure 2: Existing work showing proposed tree

B. Related Work

H. Yao et al. In[2] proposed about high utility itemset mining in database. Item set mining is evaluated in this approach with customized approach is also discovered. Mining is performed on high itemset, with not considering some operations like addition and deletion of transaction items. Data structure needs to be updated every time after build and then database is scanned. Data sparsity issue is overcome using collaborative filtering. In this work, data collection is convenient.

Y. Liu et al. [3] used some datasets and evaluated it experimentally. The uses FIMI repository to obtain real dataset and synthetic dataset. With there id datasets are stored. Analysis of dataset with its quantity and quality is stored.

V.S. Tseng et al. [4] implemented IHUIL-Tree based algorithm with adopting pattern growth methodology, it is based on bottom-up approach. In database utility itemset is stored with header table as a prefix. High utility item set computes utility items with there IP.

M. Liu et al. In[5] proposed high utility item set mining application which is beneficial in marketing and promoting. In high utility item set mining database, itemsets are stored. HUIM algorithms are generated for static database with not considering addition and deletion of transactions. Everytime database needs to be update whenever it rebuild and scan the complete database after it complete database requires to be update. This issue needs to be overcome with using IHUP-Tree.

III. PROBLEM DOMAIN

A. Problem Statement

On the basis of dimensions frequency and popularity are calculated in the complete work.

1) Numbers of occurrence are ignored in existing work with focusing only on weight (net profit) and volume.
2) Existing system works on only 2 dimensions.
3) Tree generated in existing work is complex because of use of list data structure.
4) Unbalanced tree is used in existing work.
5) Graphical representation is worked which creates difficult searching.
6) Worked on M*N dimensional.
7) Frequency which is significant is neglected.

Limitations overcome in this work

a) This work overcomes the issue of data sparsity by using collaborative filtering.
b) Existing work used unbalanced and unweighted tree data structure with creating issues of searching and sorting.

IV. PROPOSED WORK

Complete work describes methodology used and analyzed on the basis of user interest.

Multi-dimensional System on the basis of 3D system is used with concentrating on frequency and popularity. Impact is calculated on the basis of multidimensional system.

1) B+ tree will be used in proposed work with replacing tree used in existing work.
2) In this work frequency will be count as one of the dimension.
3) M*N*L, where,

M= Volume
N= Net weight
L= Frequency

4) Product weight is justifies one of the dimension.
5) One is considered as item id and another as weight. B+ tree is generated by weight.

B+ Tree is an efficient tree which stores data in an convenient way. Sorting is easy and simple using B+ tree with easy search. B+ tree decreases input-output operations with easy search.

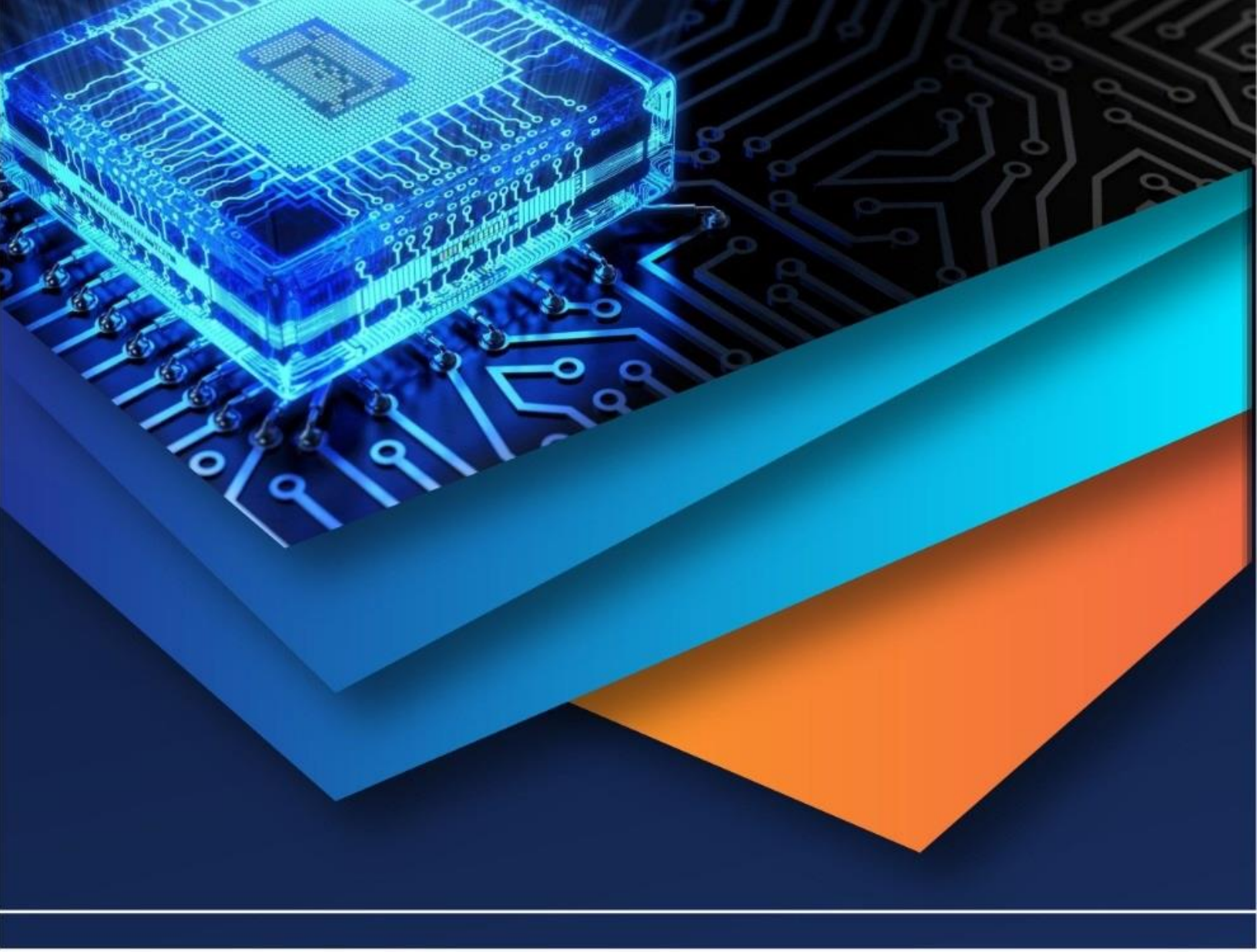
V. CONCLUSION AND FUTURE WORK

Entire work concludes that the interest of user is more significant as it works on multi-dimensional data. In this work issue of data sparsity degree is decreased. It provides with judgment of variety of data. Highly efficient result is evaluated with justified weight, frequency and popularity as compared to existing work.

Proposed work implemented three dimensional system with focusing popularity and frequency. Impact is used to calculate multi dimension system.

REFERENCE

[1] Shiming Guo, Hong Gao, "An Efficient Algorithm for Incremental and Interactive High Utility Itemset Mining," published in International Conference on Image, Vision and Computing, 2017 pp 996-1001
[2] H. Yao, H.J. Hamilton, and C.J. Butz, "A foundational approach to mining item set utilities from databases," In SDM2004, pp. 482-486.
[3] Y. Liu, W. Liao, and A. Choudhary, "A two-phase algorithm for fast discovery of high utility of item sets," In PAKDD 2005, pp. 689-695.
[4] V.S. Tseng, C.W. Wu, and B.E. Shie et aI., "UP-Growth: an efficient algorithm for high utility itemset mining," In KDD 2010, pp. 253-262.
[5] M. Liu, and J. Qu, "Mining high utility itemsets without candidate generation," In CIKM 2012, pp. 5 5-64.
[6] C.F. Ahmed, S.K. Tanbeer, and B.S. Jeong et aI., "Efficient tree structures for high utility pattern mining in incremental databases," In TKDE 2009 Vol. 21(12) , pp. 5 5-64.
[7] M. Zihayat and A. An, "Mining top-k high utility patterns over data streams," In Inf. Sci. 2014 Vol. 285, pp. 138-161.
[8] Ramanathan Narayana, Berkin Ozisikyilmaz, “A Benchmark suit for Data Mining Workloads” NU-MineBench: A Data Mining Benchmark Suite, published in “International Symposium on workload Characterization,” 2006 IEEE, pp. 182-188.
[9] D. Goldberg, D. Nichols, B. M. Oki and D. Terry, "Using Collaborative Filtering to Weave an Information Tapestry," Communications of the ACM, vol. 35, no. 12, pp. 61-70,1992.
[10] B. Sarwar, G. Karypis, 1. Konstan and 1. Riedl, "Item-based Algorithms," Filtering Recommendation Collaborative In:Proceedings of the 10th International Conference on World Wide Web, pp.285-295, 2001.
[11] J. Bobadilla, F. Serradilla and 1. Bernal, "A New CollaborativeFiltering Metric that Improves the Behavior of Recommender Systems," Knowledge-based Systems, vo1.23, no.6, pp.520-528, 2010.
[12] Y. Yu and G. H. Qiu, "Research on Collaborative Filtering Recommendation Algorithm by Fusing Social Network," New Technology of Library and Information Service, vol. 28, no. 6, pp. 54-59,2012.
[13] Zhou, Ling, and Stephen Yau. "Efficient association rule mining among both frequent and infrequent items." Computers & mathematics with applications54, Vol no. 6,pp737-749, 2007.
[14] Sudip Bhattacharya1 , DeeptyDubey, ―High Utility Itemset Mining, International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, Vol. 2, Issue 8, August 2012.
[15] J. Hu. And A. Mojsilovic, ―High-utility pattern mining: A method for discovery of high-utility item sets, Pattern Recognition 40 ISSN No:3317 – 3324, 2007.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY