International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 2, Issue 1, ISSN No. 2455-2143, Pages 113-115
Published Online November-December 2016 in IJEAST (http://www.ijeast.com)

PARALLEL ALGORITHM FOR FINDING TOP K HIGH UTILITY ITEMSET FROM BIG TRANSACTION DATA

Aniruddha Prabhu B P
Department of CSE,
JNNCE, Shimoga, Karnataka, India

***Abstract*— High utility itemsets (HUIs) mining is a rising subject in data mining which alludes to finding all itemsets having an utility meeting a client determined least limit known as least utility. However setting least utility suitably is a troublesome issue for clients. On the off chance that base utility is set too low, excessively numerous High Utility Itemsets (HUIs) will be produced, which may bring about the mining procedure to be extremely wasteful. Then again, if least utility is set too high, it is likely that no HUIs will be found from big transaction data. Here the executed structure for top-k high utility itemset mining, where k is the fancied number of HUIs to be mined will address the above issue. The effective calculation named TKU (mining Top-K Utility itemsets) will mine such itemsets without the need to set least utility, this calculation is effectively contrasted with the earlier techniques since client simply need to indicate the quantity of required itemset without determining the limits.Since Big data requires fast computation this method executes in parallel fashion that reduces the execution time of finding HUIs.**

Keywords— **High Utility Itemset, Frequent Itemset Mining**

I. INTRODUCTION

Data mining is the process of extracting the knowledge from the set of raw data generated from the vast computation. Data Mining alludes to mining or extraction from the database or Big Data which stores extensive measure of information. Both Data mining and revelation of information in the information stockpiling is another interdisciplinary field, measurable thought blending, machine learning, databases and parallel processing. There are many types of Data mining technique such as Associative rule mining, probabilistic mining and static mining. The customary Frequent Itemset Mining may find a lot of incessant yet low-esteem itemsets,however may not highlight the data on important itemsets having low offering frequencies and high benefit. Henceforth, it can't fulfill the necessity of clients who yearning to find itemsets with high utilities. So there is a need to mine top K high useful itemset where K is the desired number of Itemsets of high utility. Here the end user need not to specify the minimum_utility value .This is a scalable solution to mine Big Transactional Data.

II. PROPOSED ALGORITHM

TOP K High Utility Algorithm For Big Data –

The generalized system architecture that explains overall process that takes place in hadoop distributed computing environment .Big Data is put away in Hadoop Distributed File System (HDFS) which is disseminated to the Map Nodes for parallel handling.The Map Nodes execute with respect to the huge information doled out and passes the middle of the path result to the Reduce Nodes.The Reduce Nodes gather the middle of the path results from the Map Nodes and process them. The aftereffect of the Reduce Nodes can be composed to the HDFS or can be gone on to the following Map Reduce Job.

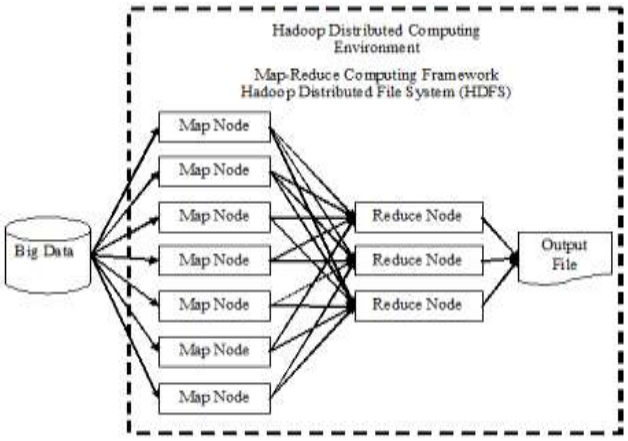

Fig. 1. Generalized Big Data Execution Model

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 2, Issue 1, ISSN No. 2455-2143, Pages 113-115
Published Online November-December 2016 in IJEAST (http://www.ijeast.com)

On the basis of such considerations, the HUI algorithm uses a different approach to mine from the big transaction data instead of database.

Initially Potential 1 High utility itemset which contains all possible Potential itemsets are generated. From the classified Potential 1 HUI,initial minimum utility of each itemset is estimated and the classified itemset UP-tree will be built. UP-tree is parsed and utility of each itemset is checked for condition Itemset Utility > minimum utility, if true sent to next step else it will be Ignored and next itemset from the UP-tree will be fetched. So obtained Itemset will be added to final potential most useful itemset list and the final most useful itemset list requires user specified k value. High useful itemset will be fetched from the final list.

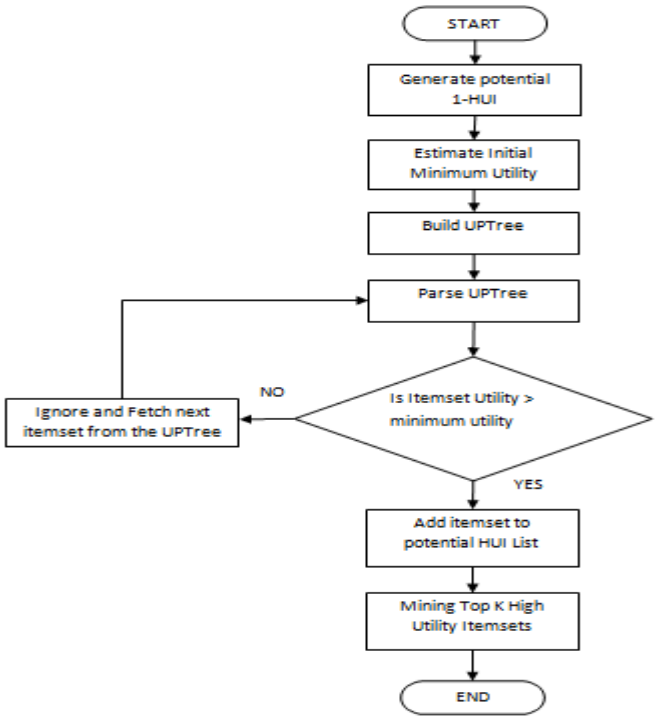

Fig. 2. Flow Diagram For Top K HUI Algorithm

III. Experiment and Result

The test set for this evaluation experiment of Top k HUI from big transaction used random transaction from internet. Open source Hadoop software was used on ubuntu platform to perform the experiment. The PC for experiment is equipped with an Intel i4 2.20GHz Personal laptop and 6GB memory.

The big transaction is given as input which will be processed to produce first phase output from which the algorithm will generate the Candidate Itemset .Here some of the Itemset which have the potential to be opted as high utility itemset are selected based on the utility of the dataset given as input to the next phase, finally Top K HUI list of itemset is extracted.

Fig. 3. Big Transaction Data File As Input

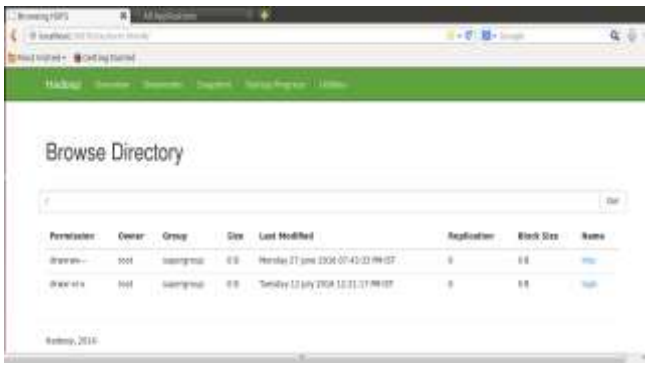

Fig. 4. HDFS Before Running The Application

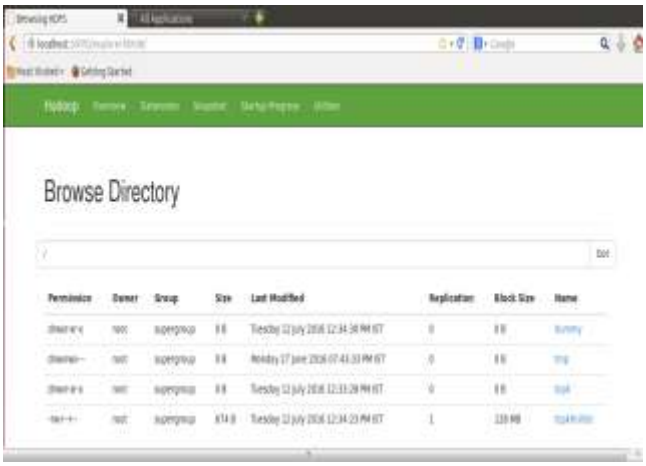

Fig. 5. HDFS After Running HUI Miner Application

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 2, Issue 1, ISSN No. 2455-2143, Pages 113-115
Published Online November-December 2016 in IJEAST (http://www.ijeast.com)

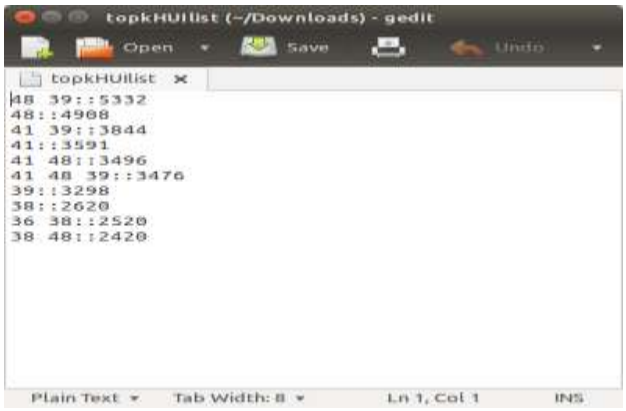

Fig. 6. Top K HUIs

Table 1 show the candidate key generation experiment results of TKU for bid transaction data and TKU_{Base} algorithm applied for big transaction data.This clearly shows that TKU algorithm is more effective than earlier algorithm.

Table -1. Experiment Result

K	TKU	TKU_{Base}
1	**1450**	**2,500,489**
10	**1559**	**2,529,557**
100	**2650**	**2,630,329**
1,000	**40329**	**2,682,322**

IV. CONCLUSION

This framework works for the issue of top-k high useful itemsets mining, where k is the required number of itemsets to be mined which will be high useful to the end user. This proficient calculation of TKU (Mining Top-K Utility itemsets) is executed for mining such itemsets without setting least utility edges. TKU is two-stage calculation for mining top-k high useful itemsets, which fuses five techniques to viably raise the outskirt least utility edges and further prune the inquiry space.This framework have great versatility on vast datasets and the execution of this framework greatly contrasted with ordinary mining strategy.

V. REFERENCE

[1]. Tseng, V.S.; Bai-En Shie ; Cheng-Wei Wu ; Yu, P.S., *"Efficient Algorithms for Mining High Utility Itemsets from Transactional Databases"*, IEEE Transactions on Knowledge and Data Engineering, June 2013, pp. 1772 – 1786.

[2]. Fumarola, F.; Malerba, D., *"A parallel algorithm for approximate frequent itemset mining using MapReduce"*, Int. Conf. on High Performance Computing & Simulation (HPCS), Jul 2014, pp. 335 – 342.

[3]. Chun-Wei Lin; Wensiieng Gan ; Tzung-Pei Hong ; Chien-Ming Chen, *"Maintaining high-utility itemsets in dynamic databases"*, Int. Conf. on Machine Learning and Cybernetics (ICMLC), July 2014, pp. 469 – 474.

[4]. Sheng-Hui Liu; Shi-Jia Liu ; Shi-Xuan Chen ; Kun-Ming Yu, *"IOMRA - A High Efficiency Frequent Itemset Mining Algorithm Based on the MapReduce Computation Model"*, IEEE Int. Conf. on Computational Science and Engineering (CSE), Dec 2014, pp. 1290 – 1295.

[5]. Lin, Jerry Chun-Wei; Gan, Wensheng ; Fournier-Viger, Philippe ; Hong, Tzung-Pei ; Tseng, Vincent S., *"Mining high-utility itemsets with various discount strategies"*, IEEE Int. Conf. on Data Science and Advanced Analytics (DSAA), Oct 2015, pp. 1 – 10.