Lattice Structure With CHARM-L To Improve Efficiency of Generating Frequently Closed Item Sets

Bhumi Patel[1], Mr. Ishan K Rajani [2]
Department Of Computer Engineering
[1]PG Student,Silver Oak College of Engineering and Technology
[2]Assistant Professor,Silver Oak College of Engineering and Technology

Abstract-*Efficient algorithms for mining frequent item-sets are crucial for mining association rules as well as for many other data mining tasks. Many algorithms are available for frequent mining like Apriori, Fp-Growth, CountTableFI, BinaryCounTableFI,etc. Proposed methods also compress crucial information about all itemsets, maximal length frequent itemsets,avoid expensive, and repeated database scans.I try to resolve the problem of Incremental Maintenance of Frequent Itemsets (IMFI) and apply lattice theory to solve the problem of computationally complexity of frequent itemsets. Concept lattice is an effective tool and platform for data analysis and knowledge discovery such as classification or association rules mining.As real data sets for data mining are very large, concept lattice structure suffers from its complexity issues on such data.In order to increase the efficiency of concept lattice-based algorithms in data mining, it is necessary to make use of an efficient algorithm to build concept lattices.So we need to compare the existing lattice algorithms and develop more efficient algorithm.*

Keywords-Lattice,Efficient,Itemsets,ScallingNextClosure

I. INTRODUCTION

This section briefly reviews the background related to this work, existing frequent itemset mining algorithms working based on sparse and dense dataset. Many of algorithms are not suitable for large dataset.There are many algorithms like Apriori,CHARM,FP-Growth,closet,etc.But there is a problem with these algorithms it has some limitations.We try to resolve the problem of Incremental Maintenance of Frequent Itemsets (IMFI) and apply lattice theory and combinatorial number theory to solve the problem of computationally complexity of frequent itemsets.We have to improve this efficiency of algorithm and we will try to reach to more accurate results.

II. METHODOLOGY

Lattice structure

Concept lattice is an effective tool and platform for data analysis and knowledge discovery such as classification or association rules mining.

As real data sets for data mining are very large, concept lattice structure suffers from its complexity issues on such data.

Keeping in mind the end goal to expand the productivity of idea grid based calculations in information mining, it is important to make utilization of a proficient calculation to fabricate idea lattices.So we have to analyze the current cross section calculations and grow more effective calculation.

ScallingNextClousure

We propose another calculation ScalingNextClosure which breaks down the pursuit space and manufactures all ideas of each inquiry sub-space. For each inquiry sub-space, we utilize a similar technique (NextClosure calculation) to create the con-cepts. So we can create all ideas of each pursuit sub-space in parallel, as the hunt sub-spaces are autonomous.

ScallingNextClosure calculation has two stages: determin-ing the allotments (see Algorithm 1) and creating all con-cepts of each parcel (see Algorithm 2). For the initial step of the calculation (deciding the standard titions), we can choose the extent of parcel by a param-eter DP of our calculation as indicated by the measure of information and our needs. For the genuine information, we can give an estimation of DP($0 < DP < 1$). DP is utilized to decide the position of the start and the finish of each segment

Algorithm 1 The first step of ScalingNextClosure algorithm: determining the partitions

1: input a parameter DP($0<DP<1$)

2:generate the ordered data context (saving the order of attributes for ordered data context in an array)
3: output the order of attributes of the ordered data context
4: m= cardinal of the attribute set of the ordered data context
5: min:=m
6: k:=0
7: while(min>=1)do {determining partition}
8: k++
9: Pk:=min
10: output Pk
11: min:=int(min*DP)
12: end while
13: T:=k / is the number of the partitions

Algorithm 2 The ScalingNextClosure algorithm to find all concepts in each partition

1: input the order of attributes of the ordered data context
2: input Pk and Pk+1 //input the partition
3: a □ {aPk}
4: END □ aPk+1
5: stop:=false
6: while(!stop)do
7: A□ generate the next closure of : for the ordered data context
8: if END (- A when searching the next closure then
9: stop:= true
10: end if
11:end while

Here we demonstrate a case of utilizing ScalingNextClo-beyond any doubt calculation to discover all ideas: First, we require not to produce an information petition for requested information setting, the request of characteristics is just put away in the principle memory.

The or-dered property set of the requested information setting for this ex-adequate is: a1a2a3a4a5a6a7a8. And afterward, we give an estimation of the parameter to decide the parcels, for exam-ple, DP = 0:5.We utilize ScalingNextClosure calculation to get 4 allotments: [a8, a4[, [a4, a2[, [a2, a1[and[a1). At last, we discover all idea aims in each segment.

III. USE OF SIMULATION SOFTWARE

There are numbers of software available which can mimic the process involved our research work and can produce the possible result. One of such type of software is Eclipse.We can readily find files related to our research work on internet or in some cases these can require few modifications. We can also use Weka Tool for implementation.

IV. PROPOSED WORK

Based on CHARM algorithm we have proposed new algorithm CHARM-L algorithm.Itemset mining is done by CHARM-L lattice algorithm for efficient mining.It enumerates closed sets using lattice structure,using an closure efficient search that skips many level. It uses a fast hash-based approach to remove any "nonclosed" sets found during computation.CHARM-L explicitly generates the frequent closed itemset lattice.Multiple data structure combines and find their data correlation.It will increase execution speed of algorithm.Reduction of memory in middle level computation in process of finding match item sets.

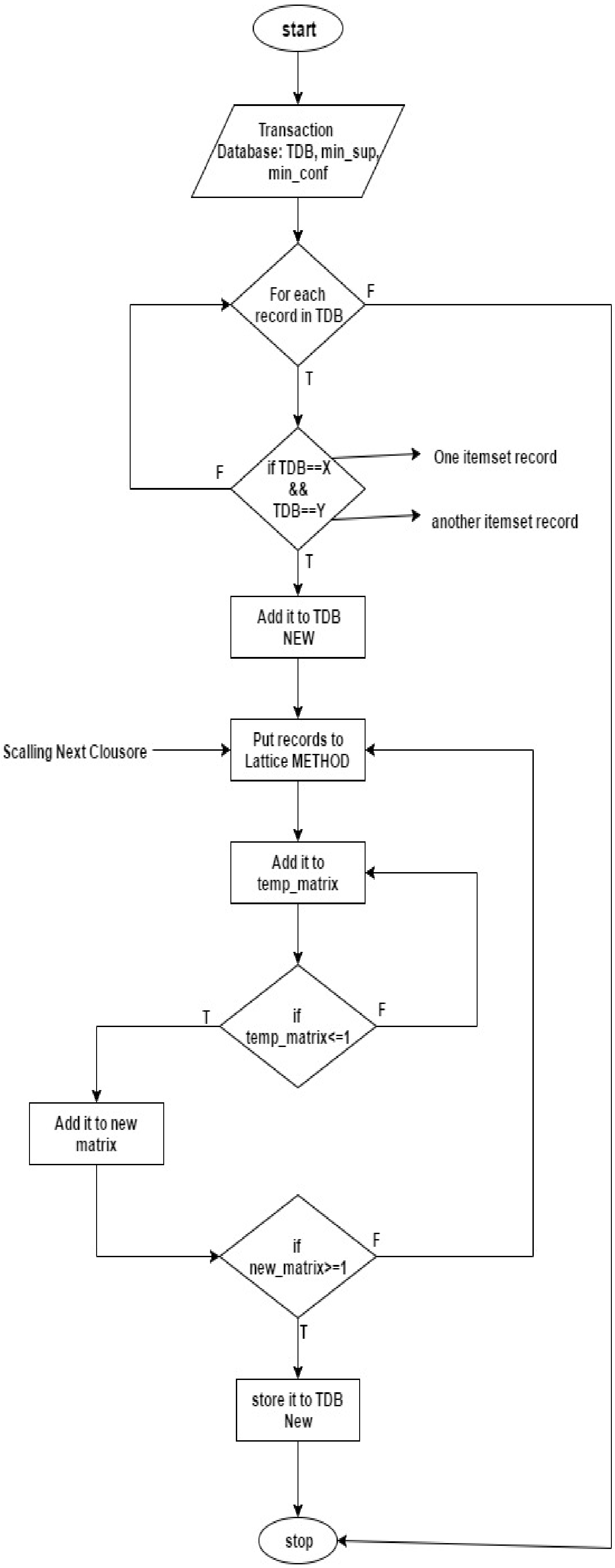
Steps of Proposed Algorithm:

Input: Transaction Database TDB,min_sup,min_conf
Output: Closed Frequent Itemset I

Step 1: For each record in TDB
if TDB==X && TDB==Y
then add it to TDB New
else go to again step 1 & scan next record
Step 2: Put records in Lattice Method
Step 3: Add records in temp_matrix
Step 4: If temp_mat <= 1
then add it to new_matrix
otherwise go to step 3
Step 5: If new_mat >= 1
then store it to TDB New
else go to step 2

Advantages over Existing algorithm:

- Fast and better than CHARM algorithm.
- Suitable for large dataset.
- Solves memory inefficient problem.
- To minimize duplication of item sets in process.
- Reduction of memory in middle level computation.
- To generate the finalized dataset with unique records.
- Multiple data structure produces data correlation from frequent itemset.



V. RESULTS

Number of Frequent closed Item sets Found & Time Efficiency Achieved

Dataset	Total records/attributes	F.C.I	Time efficiency
Pumsb(census)	32661/8	129	O(n)
Mashroom	8124/22	24	O(n)

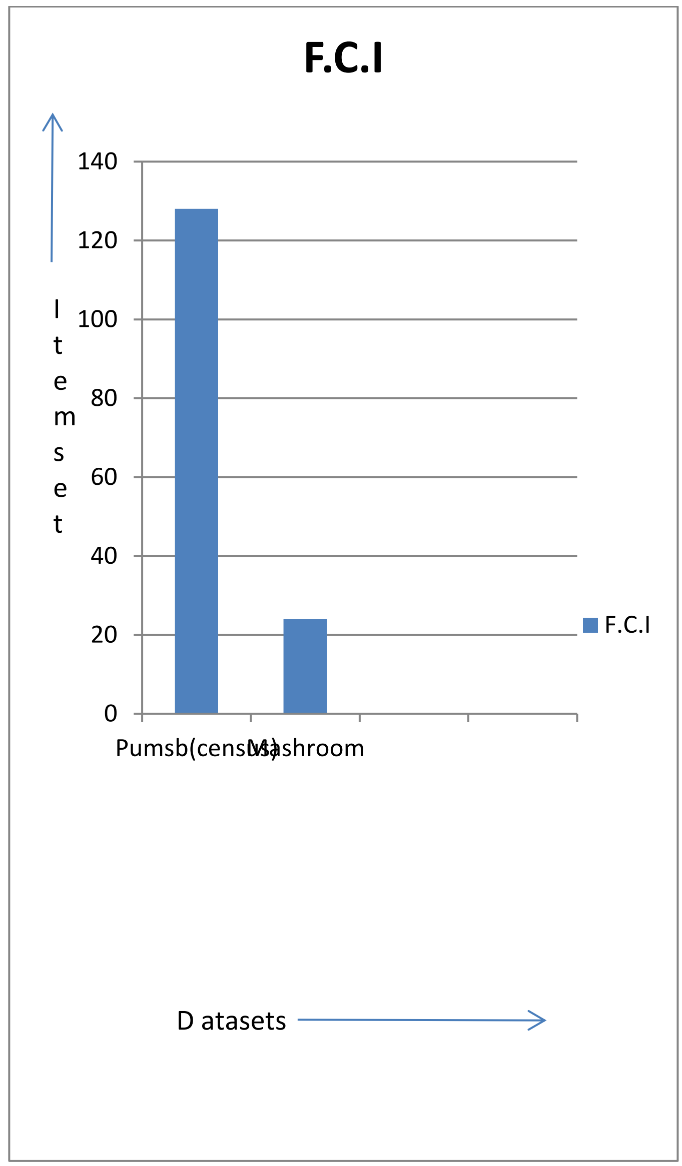


Fig. Number of Frequent Closed Item Sets Found & Time Efficiency

Comparison Between Base Paper Algorithms of Running Time of Finding F.C.I

Datasets	BCTFI	Closet	Proposed Algo
chess	10.22	12.24	9.84
pumsb	10.48	19.06	10.08
mashroom	0.5	0.79	0.44
T10I4D100K	0.23	0.42	0.2

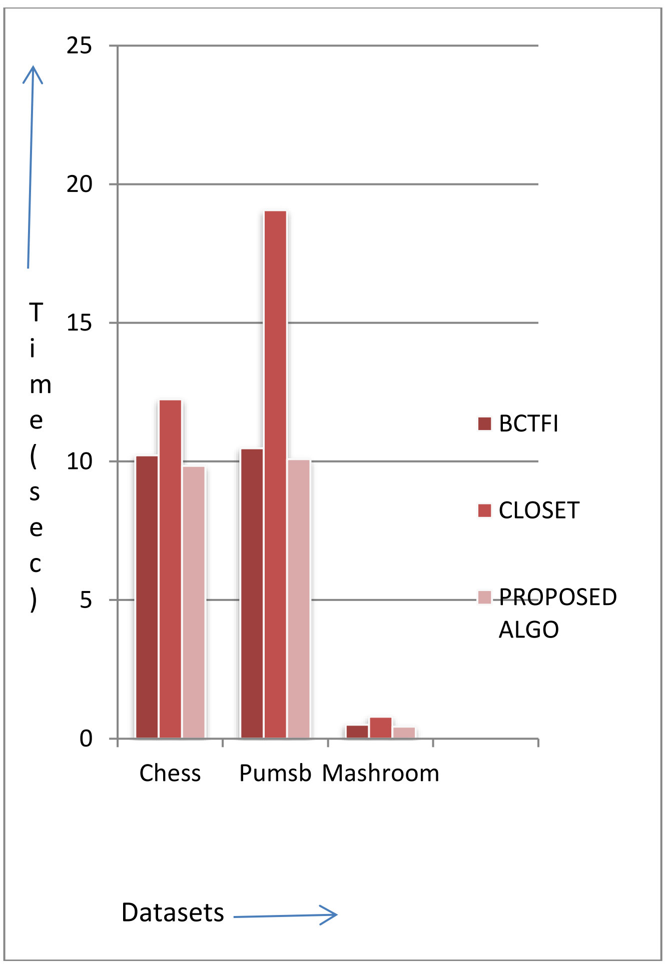


Fig. Comparison between Base Paper Algorithms of Running Time of Finding F.C.I

VI. CONCLUSION

CHARM- L can give requests of size change over existing techniques for mining shut itemsets.CHARM- L is a cutting edge calculation that creates the continuous shut itemset cross section.These calculations all the while investigate both the itemset space and tidset space utilizing the new IT-tree structure, which permits a novel hunt strategy that skips many levels to rapidly distinguish the shut regular itemsets, rather than enumerating numerous nonclosed subsets.Moreover, since we pruned nonclosed item sets timely, we reduced the search space. And with some optimized operations that reduce work and time, our algorithm also runs faster.

REFERENCES

[1] Mohamed, M. H., & Darwieesh, M. M. (2014). Efficient mining frequent itemsets algorithms. International Journal of Machine Learning and Cybernetics, 5(6), 823-833.

[2] Bodon, F. (2010, April). A fast apriori implementation. In Proceedings of the IEEE ICDM workshop on frequent itemset mining implementations (FIMI’03) (Vol. 90).

[3] Kosters, W. A., & Pijls, W. (2003, November). Apriori, A Depth First Implementation. In FIMI.

[4] Lucchese, C., Orlando, S., & Perego, R. (2004, November). DCI Closed: A Fast and Memory Efficient Algorithm to Mine Frequent Closed Itemsets. In FIMI.

[5] Zaki, M. J., & Hsiao, C. J. (1999). CHARM: An efficient algorithm for closed association rule mining (Vol. 10). Technical Report 99.

[6] Dandu, S., Deekshatulu, B. L., & Chandra, P. (2013). Improved Algorithm for Frequent Itemsets Mining Based on Apriori and FP-Tree. Global Journal of Computer Science and Technology, 13(2).

[7] Toivonen, H. T., Onkamo, P., Vasko, K., Ollikainen, V., Sevon, P., Mannila, H., ... & Kere, J. (2000). Data mining applied to linkage disequilibrium mapping. The American Journal of Human Genetics, 67(1), 133-145.

[8] Han J, Kamber M (2006) Data mining: concepts and techniques,2nd edn. Morgan Kaufmann, San FranciscoData Mining –Concepts and Techniques – JiaWei Han and MichelineKamber

[9] Yadav, N. (2014). A Review paper for mining Frequent Closed Itemsets. International Journal, 2(1).

[10] Zaki, M. J., & Hsiao, C. J. (2002, April). CHARM: An efficient algorithm for closed itemset mining. In Proceedings of the 2002 SIAM international conference on data mining (pp. 457-473). Society for Industrial and Applied Mathematics.

[11] Zaki, M. J., & Hsiao, C. J. (2005). Efficient algorithms for mining closed itemsets and their lattice structure. IEEE transactions on knowledge and data engineering, 17(4), 462-478.

[12] Li, L., Zhai, D., & Jin, F. (2003, April). A graph-based algorithm for frequent closed itemsets mining. In Systemsand Information Engineering Design Symposium, 2003 IEEE (pp. 19-24). IEEE.Ye, X., Wei, F., Jiang, F., & Cheng, S. (2015, October).

[13] An Optimization to CHARM Algorithm for Mining Frequent Closed Itemsets. In Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on (pp. 226-235). IEEE.