

Matlab Implementation of Parallel Apriori Algorithm for Reduced Space Time Complexity

Gaurav Kumar Das¹, Prof. Vijay Kumar²

^{1,2}*Kautilya Institute Of Technology & Engineering, Jaipur,*

Abstract— Data mining has gained wide spread usage & acceptance in the last decade. There are numerous tasks & fields in which data mining provides knowledge discovery solution one of the primary advents of data mining is determination of frequent of patterns or frequent patterns & their count. Another of data mining techniques are available for frequent pattern search but apriori algorithm is highly preferred & proved choice & caters to small as well as large data sets & database. This work is aimed at optimizing the time & space consumed by apriori algorithm without sacrificing the basic apriori construct by employing parallel computing services of matlab. Unlike other work carried to enhance apriori algorithm by modifying its iterations or traversal, our proposed system keeps the basic apriori proposal intact & enhance the apriori algorithm by using parallel computing to utilize all cores of the CPU or the network.

The proposed work also support processing nodes spread over internet, thus various remote processing serves can handle users apriori search, thus improving user experience user modifier while monitoring apriori integrity.

Keywords— Frequent Patterns, Parallel Computing, Enhanced Apropri, Matlab Parallel Computing Server, Transactional Data Base Mining.

I. INTRODUCTION

Information mining, or learning revelation, is the PC helped procedure of burrowing through and investigating colossal arrangements of information and after that separating the importance of the information. Information mining devices foresee practices and future patterns, enabling organizations to make proactive, learning driven choices. Information mining devices can answer business addresses that customarily were excessively tedious, making it impossible to determine. They scour databases for shrouded designs, finding prescient data that specialists may miss since it lies outside their desires. Information mining gets its name from the likenesses between scanning for significant data in a huge database and digging a mountain for a vein of profitable mineral. The two procedures require either filtering through a monstrous measure of material, or astutely testing it to discover where the esteem dwells.

Generally, information mining is tied in with handling information and distinguishing examples and patterns in that data so you can choose or judge. Information mining standards have been around for a long time, at the same time, with the approach of huge information, it is considerably more common. Enormous information caused a blast in the utilization of more broad information mining methods, incompletely on the grounds that the span of the data is substantially bigger and in light of the fact that the data has a tendency to be more differed and broad in its exceptionally nature and substance. With vast informational collections, it is not any sufficiently more to get moderately basic and clear insights out of the framework. With 30 or 40 million records of itemized client data, realizing that two million of them live in one area isn't sufficient. You need to know whether those two million are a specific age gathering and their normal income with the goal that you can focus on your client needs better. These business-driven necessities changed straightforward information recovery and measurements into more perplexing information mining. The business issue drives an examination of the information that assembles a model to portray the data that eventually prompts the production of the subsequent report.

II. OBJECTIVES OF STUDY

1. Design and development of an enhance apriori algorithm for data mining application(s) and simultaneously optimizing the number of database scan(s) required for the same.
2. The proposed system will integrate apriori algorithm with artificial intelligence, to achieve meaningful dataset in less number of database scan(s) and thus provides less computationally intense and faster results.
3. Probable use of Bio Inspired algorithms such as Bacteria Foraging and Particle Swarm Optimization in conjunction with apriori algorithm will improve database travesty time requirement in large dataset(s).
4. The proposed system can be customized to be applied to event analysis, critical event separation, spatial - temporal pattern matching.
5. The proposed system can be integrated with earth mapping data, and can be applied to massive data mining requirement(s) such as weather prediction and forecasting, sea water movement analysis, sensor data refinement etc, which require global and spatial reference.
6. The proposed system will result in significant decline in space or time complexity of existing apriori technique, thus making Big data mining efficient and fast.
7. The proposed technique can further be optimized at a negligible or no expense of accuracy by application of constraints on apriori variable such as iteration. This method of con straining can be very beneficial especially in cases of large data sets, where averaging does not incur an error of more the few decimal places.

I. LITERATURE REVIEW

Ashish Shah et al Information Mining is a field of software engineering that is worried about extricating valuable data from shifted sources. In a period where data has turned into the characteristic need of individuals, its expanded importance and handiness has taken concentration as need of great importance. The most imperative piece of this affiliation manage mining is the mining of thing sets that are visit. Market crate investigation is finished by organizations keeping in mind the end goal to recover itemsets that are visit and regularly utilized together by clients. Apriori calculation is a broadly utilized procedure with a specific end goal to discover those mixes of itemsets. Be that as it may, when any of these successive itemsets increments long, the calculation needs to go through much cycle and, therefore, the execution radically diminishes. In this paper, we propose a change to the apriori calculation by utilizing a hash work which partitions the regular thing sets into containers. Further, we propose a novel procedure to be utilized as a part of conjunction with the apriori calculation by disposing of rare thing sets from the applicant set. In this best down approach, it finds the incessant thing sets without experiencing a few emphases, in this manner sparing time and space. By finding a huge maximal regular thing set right on time in the calculation, every one of its subsets are likewise visit subsequently we never again need to filter them. Obviously, the proposed strategy has preference over the current apriori calculation when the most regular thing set's length is long. see that the apriori calculation works in a base up, expansiveness first inquiry technique. The calculation begins from the littlest arrangement of continuous thing sets and moves upward till it achieves the biggest regular thing set. The circumstances the calculation passes a database is equivalent to the biggest size of the regular thing set. At the point when a thing set turns out to be longer and it is visit then calculation turns out to be slower and takes an execution hit. We proposed an alteration to the current apriori calculation which diminished the quantity of cycles required henceforth expanding execution. Further, we proposed another method to build productivity and diminish the time taken for creating incessant thing sets. This best down approach keeps up a rundown of competitor thing set to decrease the quantity of database examines. Through this system we could essentially decrease the time required to gain visit thing sets. Henceforth expanding productivity of the whole calculation.

III. METHODOLOGY

A. System Block Diagram of Operation

- Single Core Setup (Basic Apriori)

Apriori is intended to work on databases containing exchanges (for instance, accumulations of things purchased by clients, or points of interest of a site frequentation). Different calculations are intended for discovering affiliation administrators in information having no exchanges (Winepi and Minepi), or having no timestamps (DNA sequencing). Every exchange is viewed as an arrangement of things (an itemset). Given an edge C , the Apriori calculation recognizes the thing sets which are subsets of at any rate C exchanges in the database. Apriori utilizes a "base up" approach, where visit subsets are expanded one thing at once (a stage known as applicant age), and gatherings of competitors are tried against the information. The calculation ends when no further effective expansions are found.

- Multicore Setup (Basic Apriori)

Multiple setups in basic apriori we have implemented two processor in automatic job sequencer in basic apriori and most of can find the better result.

- Multicore Setup- Modified Apriori

Multiple setups in basic apriori we have implemented two processor in automatic job sequencer and automatic job sequencer is feeding to parallel processor in basic apriori and most of can find the better result.

B. Apriori Implementation

In similar flow chart we started the apriori and set elements of similar array and classified the number of element and we can return to the path and stop.

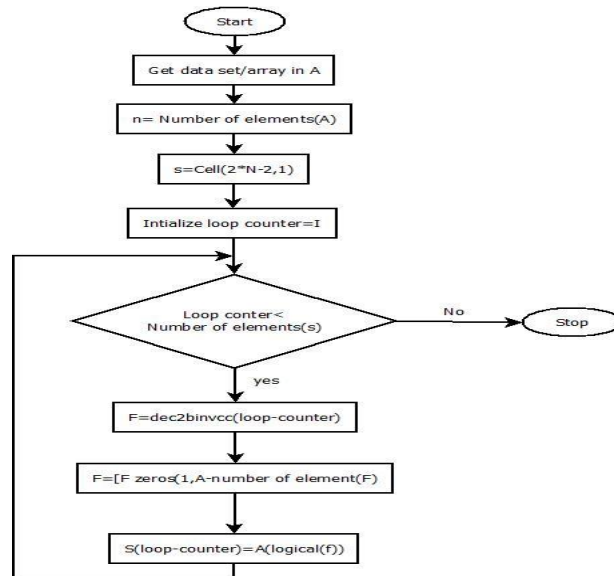


Fig 1. Apriori Implementation

A. Single Core Basic Apriori Process Flow

Start the program and load data files & star data in matrix T and display data using cell disp T and get the value of minimum support threshold & minimum confidence threshold in MST and MCT and start timer and load basic apriori and stop timer and exaction time and read out the final rules and stop it.

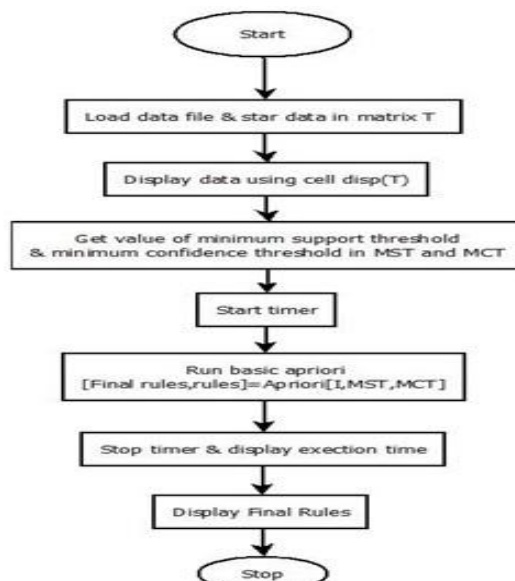


Fig 2. Single Core Basic Apriori Process Flow

B. Multi Core Modified Apriori

Start the program and load data files & star data in matrix T and display data using cell disp T and get the value of minimum support threshold & minimum confidence threshold in MST and MCT and start timer and load modified basic apriori and stop timer and exaction time and read out the final rules and get processor pool availability of pool size is == remain zero and initialize matlab and optimize two core and stop the loop processor.

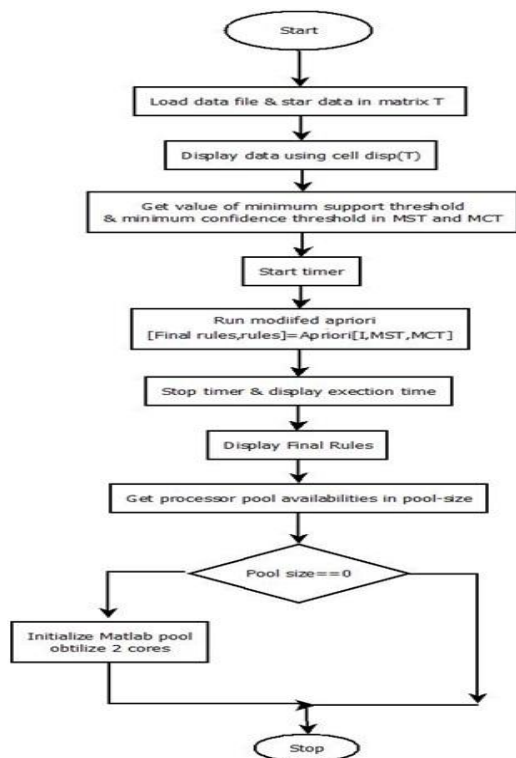


Fig 3. Multi Core Modified Apriori

IV. RESULTS

Operational Screenshots & Menu Options

For sending signal to all and hybrid data and missing algorithm and parallel computer technique and system is initializing. In processing load data in file and display data run apriori in single core and run apriori datasets in single core without any code implementation. In processing load data in file and display data run apriori in single core and run apriori datasets in single core without any code implementation and with code implementation it.

In processing load data in file and display data run apriori in single core and run apriori datasets in single core with code implementation and upload the data to the apriori algorithm.

Time Computation

Table 1. Time Computation

S.No	Dataset	Single Core Basic Apriori	Multi Core Basic Apriori	Multi Core Modified Apriori
1.	Data 9	0.3420sec	0.2492 sec	0.9388 sec
2.	Data 100	0.4199 sec	0.4174 sec	0.6144 sec
3.	Data 500	2.1051 sec	2.1800 sec	1.8181 sec
4.	Data 2000	8.3890 sec	8.4274 sec	6.8097 sec

CONCLUSIONS

The propose technique, matlab parallel computing & modified apriori, reduce the search time considerably. As shown by the results the performance separation of single core & multi core operations becomes more & more evident with growing data base size. Also comparative time consumption of single core execution, multicore unaltered apriori & multicore modified apriori are presented in results to validate the proposed technique. A average reduction of Apprx 0.3 Seconds & second highest reduction for Lengthy Data of 1.58 Seconds & Average 10% & maximum 80% reduction in execution time was observed. Thus it is evident from the results that display the final rules, that performance of apriori algorithm has been significantly improved without modifying the basic construct, as was desired.

FUTURE SCOPES

As demonstrated by the results, a enhanced ariori algorithm is present that improves perform manifolds especially on large data sets, but as data mining & parallel computing are emerging technologies, the proposed system must be adapted to future trends and demands. One of the sought advent cements is integrator of measurement of CPU core load or network load in conjunction with artificial intelligence to divide job to be processed. Another important can be derived by storing various variables for a transaction data base EOF record, so that if searched again, the adoptive apriori algorithm should search the new records only & combine them with pre saved results.

REFERENCES

- I. Association Rule Mining with Modified Apriori Algorithm Using Top Down Approach Ashish Shah Dept. Of Computer Science And Engineering Manipal Institute Of Technology, Manipal University, Manipal, India Ashish.Shah1512@Gmail.Com
- II. Mining Optimized Positive And Negative Association Rule Using Advance Abc Algorithm li.Berin Jeba Jingle, 2j.Jeya A.Celinjournal Of Theoretical And Applied Information Technology assistant Professor, Noorul Islam University, Department Of Cse, Tamilnadu, India Tamilnadu, India
- III. A Fast Advanced Reverse Apriori Algorithm For Mining Association Rules In Web Data Bina Bhandari, Bhaskar Pant, R H Goudar Cs/It Department, Graphic Era Hill University, 510, Society Area, Clement Town Dehradun, India. Kotiyalbina@Gmail.Com 2cs/It Department, Graphic Era University, Dehradun, India.
- IV. Optimization Of Association Rule Mining Using Genetic Algorithm International Journal Of Innovations & Advancement In Computer Science Ijiacs Issn 2347 – 8616 Volume 5, Issue 6 June 2016 Umesh Kumar Patel School Of Ict, Gautam Buddha University, Greater Noida, Gautam Budha Nagar(U.P.),India
- V. Privacy Preserving In Association Rule Mining Zahra Kiani Abari1 , Mohammad Naderi Dehkordi department Of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran Zahrakiani@Sco.Iaun.Ac.Iracsij Advances In Computer Science: An International Journal, Vol. 4, Issue 1, No.13 , January 2015 department Of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran Naderi@Iaun.Ac.Ir
- VI. Frequent Pattern Mining Using Parallel Architecture Of Artificial Bee Colony Sanjay Patel K. Kotecha Vice Chancellor, Parul University, Waghodia, Vadodara, Pin-391760, India Computer Engineering Department, Government Engineering College, Katpur, Patan, Gujarat. Pin -384265, India
- VII. The Concept Of Data Mining, Frequent Pattern Mining And Actionable Pattern Mining Techniques International Journal Of Advance Research In Computer Science And Management Studies Research Article / Survey Paper / Case Study Available Online At: www.Ijarcsms.Com Summarizing Dr. K. Kavitha Assistant Professor, Department Of Computer Science Mother Teresa Women's University Kodaikanal – India
- VIII. An Enhanced Apriori Algorithm For Discovering Frequent Patterns With Optimal Number Of Scansudhir Tirumalasetty, Aruna Jadda And Sreenivasa Reddy Edara Department Of Computer Science & Engineering, Vasireddy Venkatadri Institute Of Technology Guntur, Andhra Pradesh 522508, India
- IX. Improvements In Classical Apriori Algorithm And Generation Of Association Rules Suresh Gyan Vihar University, Jaipur International Journal Of Converging Technologies And Management (Ijctm) Volume 1, Issue 3, 2015 Issn : 2455-7528 Lovedeep Varinder Kaur Atri Department Of Computer Science & Engineering Guru Nanak Dev University Regional Campus, Jalandhar, India
- X. Graph-Based Clustering For Detecting Frequent Patterns In Event Log Data Erika Sy, Sam Ade Jacobs, Aldo Dagnino And Yu Ding