Hori-Vertical Distributed Frequent Itemsets Mining Algorithm on Heterogeneous Distributed Shared Memory System

Margahny H. Mohamed[1] and Hosam E. Refaat[2]

Dept. of Computer Science, Faculty of Computers and Information,
Assuit University, Assuit, Egypt
margami@gmail.com
[2] Dept of Computer Science, community collage,
King Khalid University– Kingdom of Saudi Arabia
hosam.refaat@yahoo.com

Abstract

The big challenge in discovering association rules is to find the largest frequent itemsets. Sequential algorithms do not have analytical ability, especially in terms of run-time performance, for such very large databases. Therefore, we must rely on high performance parallel and distributed computing. We present a new parallel algorithm for frequent itemset mining, called HoriVertical algorithm. The algorithm passes the database only one time and starts a new stage with the finished itemsets while some other itemsets in the same stage have not been finished yet. Also, the new algorithm is based on partitioning the database vertically and horizontally. We present the result on the performance of our algorithm on various databases, and compare it against well known algorithms.

Keywords: *Parallel Systems, Distributed shared memory, data mining, Association rule, Linda system, Tuple-space, Jini, JavaSpace.*

Published In:

IJCSNS International Journal of Computer Science and Network Security, VOL.1 No. , November 2010

References

A Fast Parallel Association Rule Mining Algorithm Based on the Probability of Frequent Itemsets

Margahny H. Mohamed[1] and Hosam E. Refaat[2]

Dept. of Computer Science, Faculty of Computers and Information,
Assuit University, Assuit, Egypt
margami@gmail.com
[2] Dept of Computer Science, community collage,
King Khalid University– Kingdom of Saudi Arabia
hosam.refaat@yahoo.com

Abstract

Frequent itemset finding is the most costly processing step in analyzing large transactional databases. At each stage in discovering frequent itemset a huge number of candidate itemsets are produced. Then, if we predict which candidate itemset will be frequent and which will not, we can reduce wastage of time in the processing unfrequent itemsets. In this paper we propose a new parallel algorithm for frequent itemset mining, called probability of frequent itemset (PFI) mining algorithm. The PFI algorithm can predict frequency of the candidate based on the probability of its subset and makes priority between candidate itemsets base on it's probability. Moreover, the PFI algorithm passes the database only one time by dividing the database horizontally and distributes it over the system nodes. Also, while finding the k-itemsets, the algorithm can start a new stage (finding k+1 itemsets) with the discovered frequent k-itemsets while some other itemsets in the same stage have not been finished yet. Moreover, we introduce a method for reducing the number of transactions. We present the result on the performance of our algorithm on various datasets, and compare it against well known algorithms.

Keywords: *Parallel Systems, Distributed shared memory, data mining, Association rule, Linda system, Tuple-space, Jini, JavaSpace.*

Published In:

IJCSNS International Journal of Computer Science and Network Security, VOL.1 No. , November 2011

References

[1] A. A. Name, and B. B. Name, Book Title, Place: Press, Year.
[2] A. Name, and B. Name, "Journal Paper Title", Journal Name, Vol. X, No. X, Year, pp. xxx-xxx.
[3] A. Name, "Dissertation Title", M.S.(or Ph.D.) thesis, Department, University, City, Country, Year.
[4] A. A. Name, "Conference Paper Title", in Conference Name, Year, Vol. x, pp. xxx-xxx.

A New Algorithm for Parallel Association Rule Mining in Distributed Shared Memory System.

Marghny H. Mohamed[1] and Hosam E. Refaat[2]

Dept. of Computer Science, Faculty of Computers and Information,
Assuit University, Assuit, Egypt
margami@gmail.com
[2] Dept of Computer Science, community collage,
King Khalid University– Kingdom of Saudi Arabia
hosam.refaat@yahoo.com

Abstract

Finding the frequent itemset is the most important problem in Association Rule Mining(ARM) because it is the most time costly step in ARM. In the case of huge number of transaction and items in the database, it is important to investigate efficient distributed algorithm for mining association rules. The efficient distributed algorithm must be scalable, easy partitioned and distributed of centralized database, minimize the calculation and communication. In this paper we present a new algorithm for finding frequent itemset, that is called HVPFI. This algorithm is an extension of our previous work. We then analyze the algorithm and compare it with other published algorithms.

Keywords: *Parallel Systems, Distributed shared memory, data mining, Association rule, Linda system, Tuple-space, Jini, JavaSpace.*

Published In:

CiiT International Journal of Data Mining and Knowledge Engineering, Vol 3, No 10, August 2011

References

[1] A. A. Name, and B. B. Name, Book Title, Place: Press, Year.
[2] A. Name, and B. Name, "Journal Paper Title", Journal Name, Vol. X, No. X, Year, pp. xxx-xxx.
[3] A. Name, "Dissertation Title", M.S.(or Ph.D.) thesis, Department, University, City, Country, Year.
[4] A. A. Name, "Conference Paper Title", in Conference Name, Year, Vol. x, pp. xxx-xxx.

A New Parallel Association Rule Mining Algorithm on Distributed Shared Memory System

Marghny H. Mohamed[1] and Hosam E. Refaat[2]

Dept. of Computer Science, Faculty of Computers and Information,
Assuit University, Assuit, Egypt
margami@gmail.com
[2] Dept of Computer Science, community collage,
King Khalid University– Kingdom of Saudi Arabia
hosam.refaat@yahoo.com

Abstract

Frequent itemset finding is the most time consuming step in analyzing large transactional databases. The using of sequential algorithms can't give analytical ability for such very large databases especially in terms of run-time performance. Therefore, we must rely on high performance parallel computing. In this paper, we present a new parallel algorithm for frequent itemset mining, called "HorVertical" algorithm. This algorithm introduces a new database partitioning that called "HorVertical" partitioning. This technique in partitioning the database reduces the dependency in the parallel computation and gives new properties to reduce the computations. The algorithm passes the database only one time and starts a new stage with the finished itemsets while some other itemsets in the same stage have not been finished yet. We present the result on the performance of our algorithm on various databases, and compare it against well known algorithms.

Keywords: *Parallel Systems, Distributed shared memory, data mining, Association rule, Linda system, Tuple-space, Jini, JavaSpace.*

Published In:

Int. J. Business Intelligence and Data Mining, VOL. No. , January 2012

References

[1] A. A. Name, and B. B. Name, Book Title, Place: Press, Year.
[2] A. Name, and B. Name, "Journal Paper Title", Journal Name, Vol. X, No. X, Year, pp. xxx-xxx.
[3] A. Name, "Dissertation Title", M.S.(or Ph.D.) thesis, Department, University, City, Country, Year.
[4] A. A. Name, "Conference Paper Title", in Conference Name, Year, Vol. x, pp. xxx-xxx.