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

Margahny H. Mohamed, Hosam E. Refaat

Abstract:

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.10, No.11