Solving the NP-Hard computational problem in Bayesian Networks using Apache Hadoop MapReduce

Nipat Jongsawat, Wichian Premchaiswadi
Graduate School of Information Technology
Siam University
Bangkok, Thailand
nipatj@yahoo.com, wichian@siam.edu

***Abstract*— The problem of exact probabilistic inference in an arbitrary Bayes network is NP-hard. The process is time consuming and complex. To speed up the processing, we need to run parts of the subnetwork in parallel. This work addresses the application of a MapReduce based distributed computing framework, Hadoop, to Bayesian network model to speed up the Bayesian update and inference processes. We present an analytical framework for understanding the transformation of Bayesian network model to Map and Reduce tasks. Computer-based Patient Case Simulation System (422 nodes) is chosen as a case study for the transformation.**

Keywords— Hadoop MapReduce, Hadoop Distributed File System, Bayesian network, Bayesian Inference.