Abstract

The problem of motif discovery has been studied extensively over the last few decades. Many sequential and parallel algorithms have been proposed and studied. A significant runtime is still required for several challenging instances of the motif search problem. This paper studies parameter spaces to find an optimal point for load balancing between the master and the worker processes, which are collaboratively and concurrently searching for motifs. Extensive experiments have been carried out on the state-of-the-art TACC Stampede System. The results demonstrated that a workload from the parallel motif discovery problem is best divided between the master and worker processes by having the master process worked on the first \(((l-4)/2)\) nucleotides in the DNA sequences, where \(l\) is the total length of the input DNA sequences, before passing the remaining work to the worker process. In addition, the results also shown that the latency tolerance techniques used in the implementation of this work is effective because of the almost linear speedup obtained.

Index Terms

Computer Science
Algorithms
Keywords

Latency Tolerance, Load Balancing, Message Passing, Motif Discovery