Abstract

An MPI based parallelization technique for improving the scalability of the global sequence alignment algorithm on clusters of workstation is presented. We propose the parallel implementation of the Wavefront algorithm based on a chunk size transformation to handle large dataset with message passing model. Molecular biologists frequently align DNA sequences of entire genomes to detect important matched and mismatched regions. Even though efficient dynamic programming algorithms exist for this problem, the required computing time is still very high due to the size of these sequences. Because the number of sequenced organisms is increasing rapidly, fast and accurate solutions are of highest importance to research in this area. We show that an appropriate choice of the number of processes and chunk size has great impact on the overall system performance on cluster system. We have conducted the experiments on real-life DNA samples of house mouse mitochondrion and the DNA of rabbit mitochondrion obtained from the public database GenBank [GenBank, http://www.ncbi.nih.gov] in our experiment to measure the algorithm behavior appropriately. The results obtained from performed experiments, demonstrate that developed parallel Wavefront algorithm exposes high speedup and scales linearly with the increasing number of processes. Also the communication among processes and memory requirements are kept at minimum to achieve high efficiency. The experiments were performed on cluster which consists
Parallelizing and Analyzing the Behavior of Sequence Alignment Algorithm on a Cluster of Workstations for Large Datasets of two workstations of 12 core each with multithreading environment.

References

- Martins, W. S., del Cuvillo, J. B., Cui, W., Gao, G. R., "Whole Genome Alignment using a Multithreaded Parallel Implementation," Proceedings 13th Symposium on
Parallelizing and Analyzing the Behavior of Sequence Alignment Algorithm on a Cluster of Workstations for Large Datasets

- J. C. Setubal, J. Meidanis. Introduction to Computational Molecular Biology.
- C. Quammen. Introduction to programming shared-memory and distributed memory parallel computers, Crossroads 12, 2005.
- http://www.niu.edu/mpi

Index Terms

Computer Science
Artificial Intelligence
Keywords
Sequence Alignment  MPI  Cluster Computing  Parallel Wavefront algorithm