Abstract

Big Data has an increasing impact on the use of bioinformatics software. One way to deal with this challenge is through parallel computing. Using the program Structure as a case study, this paper investigates ways in which to counteract the challenges created by the growing datasets. This paper proposes an OpenMP-MPI hybrid parallelization of the MCMC steps, which are an integral part of Structure, and analyses the performance under various scenarios. The results indicate that the parallelization produce significant speedups over the serial version in all scenarios tested. This allows for the use of the hardware in a more efficient manner, by adapting the program to the parallel architecture. This is important because not only does it reduce the time required to perform existing analyses, but also opens the door to the analysis of previously impractically large datasets.

References

- Jonathan K. Pritchard, Matthew Stephens, and Peter Donnelly. Inference of Population
A Hybrid OpenMP-MPI Parallelization of Structure Software


Index Terms

Computer Science

Software Engineering

Keywords

MPI OpenMP parallelization Structure MCMC SPRNG SHARCNET speedup Big Data High Performance Computing