Parallel Computing to Predict Breast Cancer Recurrence on SEER Dataset using Map-Reduce Approach

Abstract

Due to the late overpowering development rate of large scale data, the advancement of handling faster processing algorithms with optimal execution has turned into a critical need of the time. In this paper, parallel Map-Reduce algorithm is proposed, that encourages concurrent participation of various computing hubs to develop a classifier on SEER breast cancer data set. Our algorithm can prompt supported models whose speculation execution is near the respective baseline classifier. By exploiting their own parallel architecture the algorithm increases noteworthy speedup. In addition, the algorithm don't require singular processing hubs to communicate with each other, to share their data or to share the knowledge got from their data and consequently, they are powerful in safeguarding privacy of computation also. This paper utilized the Map-Reduce framework to implement the algorithms and experimented onSEER breast cancer data sets to exhibit the execution as far as classification accuracy and speedup.

References


22. B. Panda, J. Herbach, S. Basu, and R. J. Bayardo, “Planet: Massively parallel learning


Index Terms

Computer Science  Biomedical

Keywords

Breast cancer; Big dataanalytics, Classification; Parallel Computing; MapReduce, SEER.