Abstract

The performance of job scheduling algorithms in campus-wide PC-cluster distributed computing environment may be influenced by several input variables (factors) such as sum of the job sizes of all the jobs in the workload, number of PCs in the cluster and even on the type of scheduling algorithm being used. Response surface methodology (RSM) based statistical
The Framework for Performance Modeling and Evaluation of Parallel Job Scheduling Algorithms

regression techniques build empirical model for performance prediction of the scheduling system by means of mathematical equation that relate the scheduler performance (response) to the input process parameters. Artificial neural networks (ANNs) can also be successfully employed for modeling of complex non-linear prediction problems. Feed-forward ANN models viz. multilayer perceptron (MLP) and radial basis functions (RBF) are trained with empirical data to approximate the makespan response of job scheduling algorithms and they can be generalized to predict the new large instances of same problem class. Overall predictive capabilities of these modeling techniques are also measured with various statistical goodness-of-fit standards. This paper will focus on comparing the performance of RSM and ANN based modeling schemes to predict makespan values of job scheduling algorithms in PC-cluster based distributed computing environment. Performance of three space-sharing scheduling algorithms namely First-come-first-serve, Fit-processors-first-served and Largest-job-first is also compared in this paper.

Reference

clusters”, Proceedings of the 2006 Winter Simulation Conference.
- SPSS 17.0 User’s guide http://www.hks.harvard.edu accessed on October 2011.
- Design Expert Software version 8.0 user’s guide 2009.

Index Terms

Computer Science

Distributed Computing

Key words

PC-cluster

Job scheduling

Response

surface methodology

Multilayer perceptron

Radial basis functions