Improvements of Execution-time Estimation Model Construction for Heterogeneous Clusters

Graduate Adviser : Shuichi Ichikawa 011011 Yuu Kawai

1 Introduction

The performance of a parallel application for homogeneous environment is degraded by load imbalance on a heterogeneous cluster, which consists of various processor elements (PE). It is thus necessary to distribute workloads considering the performance of each PE. Though load imbalance can be solved by invoking multiple processes on fast PEs, it is difficult to determine the optimal number of processes on each PE.

Takahashi and Ichikawa [1] constructed the execution-time estimation models of four scientific applications from measurement results, and reported that the (sub-)optimal configurations were actually estimated. However, their model requires homogeneous PEs, and the precision of the model is degraded by measurement deviation. This study presents execution-time estimation models that are constructed from heterogeneous cluster itself. A method that prevents the precision degradation are also examined.

2 Execution-time Estimation Model

Let N be the size of the problem. G_i stands for a group of PEs comprised of equivalent PEs in heterogeneous cluster. P_i is the number of PEs actually used for the job in G_i, while M_i is the number of processes on each PE in G_i. The total number of processes in the cluster is given by $P = \sum_i P_i M_i$. The execution time of G_i is designated by T_i, which is parameterized by N, P and M_i. The total execution time T is estimated by $max_i T_i$. The estimation function of T is also represented as *execution-time estimation model* in this study. Optimal configurations are estimated by examing the models of all possible configurations (P_i, M_i).

In case of HPL, T is estimated by Equation (1). Equation (1) is transformed to Equation (2) for each M_i, using parameters N and P. Takahashi named Equation (2) as NP-T model [1]. Constant factors k_0, ..., k_9 are determined by least squares method, using the measurement results of two or more homogeneous PEs.

$$T(N, P) = \frac{1}{P} \cdot O(N^3) + P \cdot O(N^2) + O(N^2) \quad (1)$$

$$T_i(N, P)|_{M_i} = \frac{1}{P} \cdot (k_0 N^3 + k_1 N^2 + k_2 N + k_3) + P \cdot (k_4 N^2 + k_5 N + k_6) + k_7 N^2 + k_8 N + k_9 \quad (2)$$

For Equation (1), at least three measurement results of different P are required to extract parameters. Since the number of PEs have to be changed to measure the results of various P, four or more homogeneous PEs are required. The errors in measurements may cause failure of parameter extraction (model failure), which leads to negative estimation time. This often results in poor estimation results.

3 Improvement methods

This study attempts to give some improvements on model construction methods. Following two methods are independent each other, and may be used together.

3.1 Constructing models from heterogeneous clusters

In this study, we assume that the execution time of a heterogeneous cluster that includes slow PEs is approximately equal to the execution time of the cluster comprised of the same number of the slowest PEs. With this assumption, the models can be constructed from heterogeneous cluster, except to the fastest PEs.

3.2 Extraction by non-negative least squares

The negative execution time might be estimated, when k_i ($i = 0, ..., 9$) include negative numbers. To prevent negative execution time from estimation, k_i have to be limited to non-negative numbers. In this study, Non-negative least squares method (**nnls**) is examined to avoid the degradation of estimation.

Table 1: Evaluation environment

	G_1	G_2	G_3
CPU	Pentium4 3.6GHz	Xeon 2.8GHz	CeleronM 1.5GHz
OS	FedoraCore4	RedHatLinux9	FedoraCore5
Complier,Library	icc 9.0, ifc 9.0, mpich 1.2.7p1		
P_i	$0 \le P_1 \le 2$	$0 \le P_2 \le 4$	$0 \le P_3 \le 2$
M_i	$0 \le M_1 \le 3$	$0 \le M_2 \le 2$	$0 \le M_3 \le 1$

Table 2: Problem Size (N)

Benchmark	Measurement	Evaluation
HPL	400〜6400 (9 sets)	400〜9600 (11 sets)
Himeno BMT	32〜192 (9 sets)	32〜256 (11 sets)
hpcmw-solver-test	60〜442 (7 sets)	60〜600 (12 sets)
FFTE	2^{12}〜2^{20} (9 sets)	2^{12}〜2^{23} (12 sets)

4 Evaluation

In this study, four benchmark programs, HPL, Himeno BMT, hpcmw-solver-test and FFTE, are examined on the heterogeneous cluster showin in Table 1. The problem sizes for measurements and evaluations are summarized in Table 2. The models of G_3 are constructed by G_1, G_2 and G_3, and the models of G_2 are constructed by G_1 and G_2. The models of G_1 that consists of the fastest PEs are constructed from uniform PEs (G_1).

Figure 1 summarizes the measured execution times of the estimated optimal configurations and actual optimal configurations. For HPL and hpcmw-solver-test, the precisions of the model constructed from heterogeneous cluster (**hetero**) and that from homogeneous PEs (**homo**) were approximately equal. For Himeno BMT, though the estimation was degraded at $N \le 96$ and $N \ge 224$ with least squares method (**ls**), (sub-)optimal configurations were estimated by non-negative least squares method (**nnls**). For FFTE, though the estimation was degraded at $N \le 2^{16}$ and $N \ge 2^{21}$ with least squares method, the estimation was improved by non-negative least squares method. For Himeno BMT and FFTE, the precisions of the models constructed from homogeneous PEs and that from heterogeneous PEs are different with least squares method. By using non-negative least squares method, both homo and hetero models showed an approximately similar precision.

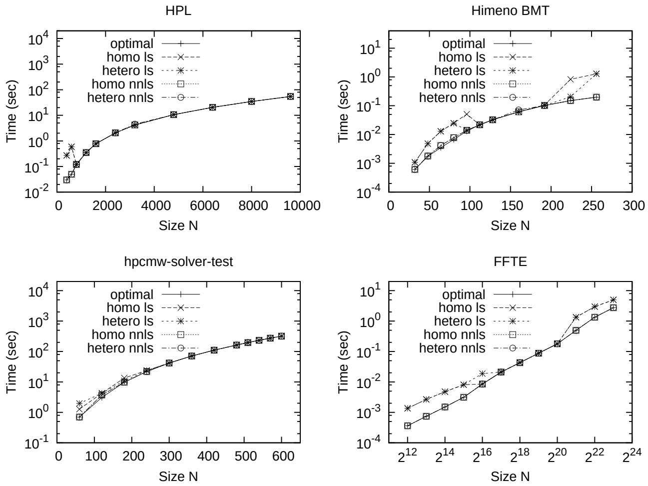

Figure 1: Evaluation results

5 Conclusion

This study examined two methods to construct models from a heterogeneous cluster itself, and showed that both methods give comparable accuracies if model failures were avoided. The degradation of precision was prevented by non-negative least squares method, and (sub-)optimal configurations were estimated.

References

[1] S. Takahashi, S. Ichikawa “Application and Evalution of Optimal Configuration Estimation Scheme for Heterogeneous Clusters,” *IPSJ SIG Notes 2006-HPC-105*, pp. 97–102 (2006).