不均一クラスタ上での実行時間予測モデルとその評価

指導教官：市川 周一　　学籍番号 003101 岸本 芳典

1 はじめに

性能・構成が異なる要素プロセッサ (PE) で構成された計算機クラスタを**不均一クラスタ**と言う．既存の科学技術計算応用の多くは計算負荷を PE に均等に割り当てるため，不均一クラスタ上で実行すると負荷不均衡による性能低下を生ずる．不均一クラスタ上で負荷を均衡化する手法の一つに，高速 PE 上に複数のプロセスを起動する手法 (**マルチプロセス法**) がある．マルチプロセス法は実現が容易で，幅広い応用に適用可能である．

一般に，不均一クラスタ内の全てのプロセッサを使用しても，通信時間の発生により実行時間が最小になるとは限らない．マルチプロセス法で実行時間を最小化するには，(1) 最適な PE セットと，(2) 各 PE 上の最適なプロセス数を求める必要がある．本研究では，最適な実行方法 (構成) を求める問題を組合せ最適化問題としてモデル化する．モデル化には，構成から実行時間を予測する近似式が必要である．

以下，HPL (High Performance Linpack)[1] を応用例として，実行時間の実測値から予測モデルを構築し，(準) 最適な PE 構成およびマルチプロセス数の予測が可能であることを示す．また，実測値の測定方法とモデルの精度の関係についても検討する．

2 実行時間予測モデル

HPL のテストケースについて実行時間の測定を複数行ない，その測定結果から実行時間予測モデルを構築する．HPL のアルゴリズムから実行時間の近似式が求まるので，実測値を最小二乗法で処理して係数項を求める．このようなモデル化は実装や応用に依存しないため幅広い応用に適用可能である．

本研究では，(1) 通信時間が通信相手に依存しないと仮定し，(2) 等価な PE のマルチプロセス数は同一として，可能な構成の組み合わせ数を減らす．このような単純化が妥当か否かは，実測値との比較によって検証する (3 章)．以後，等価な PE をグループ G_i で表す．G_i で使用するプロセッサ数を $0 \leqq P_i \leqq |G_i|$，各 PE 上のマルチプロセス数を M_i とする．予測モデルは $N, P(=\sum P_i M_i), M_i$ から G_i 上の各プロセスの実行時間 T_i を予測する．不均一クラスタ全体の実行時間 T は $T = \max_i(T_i)$ で近似できる．

サイズ N, 全プロセス数 P に対する実行時間は，HPL の計算量・通信量のオーダより式 (1), 式 (2) で見積もる．式中の添字 P および M_i は，G_i の特定の構成 $[P, M_i]$ および $[M_i]$ に対するモデルであることを表す．以後，式 (1) を N-T モデル，式 (2) を P-T モデルと呼ぶ．

$$T_i(N)|_{P,M_i} = k_0 N^3 + k_1 N^2 + k_2 N + k_3 \quad (1)$$

$$T_i(N,P)|_{M_i} = k_4 P \cdot T_i(N)|_{P,M_i} + k_5 \frac{1}{P} \cdot T_i(N)|_{P,M_i} + k_6 \quad (2)$$

N-T モデル中の係数 k_0〜k_3 は 4 つ以上の異なる N の実測値から求める．可能な全ての構成 $[P, M_i]$ について N-T モデルを構築するとモデル数が非常に多くなるので，P をパラメータとして N-T モデルを統合する (P-T モデル)．P-T モデルの係数 k_4〜k_6 は，P の異なる 3 つ以上の N-T モデルより求めることができる．但し，G_i での PE 単独実行時 ($P = M_i$) とクラスタ実行時 ($P > M_i$) では通信の過程が大きく異なるため，$P = M_i$ では N-T モデルをそのまま使用し，P-T モデルは $P > M_i$ の N-T モデルから構築することで精度の向上を図る．本研究では図 1 に示すように構成 $[P, M_i]$ に応じてモデルを切り替えて使用する．

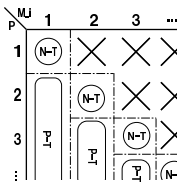

図 1: モデル切替

3 評価

表 1 に示す不均一クラスタを用いて HPL を実行し，実行時間予測モデルを構築する．本研究では，各グループ G_i 内で問題サイズ N，投入 PE 数 P_i，マルチプロセス数 M_i を変えながら測定を行なう．G_1, G_2 の性能比が 4:1 であったため，マルチプロセス数 M_1 は M_2 の 4 倍前後の値まで測定する．

表 1: HPL 実行環境

PE	Athlon 1.33 GHz×1 (G_1), Intel Pentium-II 400 MHz×8 (G_2)
OS, Network	RedHat Linux7.0J (kernel 2.4.2), 100base-TX (Intel Pro100+)
Compiler	gcc 2.96, -DHPL_DETAILED_TIMING -fomit-frame-pointer -O3 -funroll-loops -W -Wall, MPICH–1.2.5, ATLAS 3.2.1

表 2: クラスタ構成パラメータ

	Athlon		Pentium-II		構成数
	P_1	M_1	P_2	M_2	
構築時	1	1〜6	1〜8	1〜6	54
評価時	0〜1	1〜6	0〜8	1	62

代用モデル　G_1 の PE 数は 1 であり P-T モデルを構築することができない．そこで，G_2 の P-T モデルを N-T モデルの出力から求めた比で定数倍して代用する．今回は G_2 の P-T モデルを 0.307 倍し G_1 の P-T モデルとした．妥当性は実測値との比較から検証する．

テストケース測定　表 2(構築時) の各構成ごとに表 3 に示す N について測定を行なう．今回は測定セット数や測定時間の削減についても評価するため N の取り方が異なる 3 つの場合 (N9, N5, NS) について測定する．N9 は $N \leq 6400$ の 9 セット，N5 は同範囲で 5 セット, NS は $N \leq 1600$ の 5 セットとした．

表 3: N-T モデル構築時の N の測定ケース

	Measurement set of N	total	測定所要時間
N9	400,600,800,1200,1600,2400,3200,4800,6400	486	6.4 [hour]
N5	400,800,1600,3200,6400	270	4.2 [hour]
NS	400,600,800,1200,1600	270	0.37 [hour]

最適構成の予測　表 3 の各測定結果から予測モデルを構築した．次に表 2(評価時) の 62 構成に対して $N = 3200$〜9600 の実行時間を予測し，予測が最小となる予測最良構成を求めた．さらに同 62 構成の実際の実行時間 t も測定し，実測が最小となる実測最良構成も求めた．予測最良構成の予測実行時間 τ, 実測実行時間 $\hat{\tau}$ と実測最良構成の実行時間 $\hat{T}$ の間の誤差 ϵ, e を表 4 に示す．

この結果，N5 の ϵ は最大 14.6% と N9 よりもやや悪化したが，予測最良構成は一致し，誤差 e が 0%〜7.4%の実用的な (準) 最良構成が得られた．NS では $N = 9600$ で負の予測実行時間となる異常な結果が得られた．異常の原因は図 2 のように不適切な N-T モデルが構築されるためであった．NS の ϵ は 10.6%〜115%，e は 9.9%〜60.8%であった．

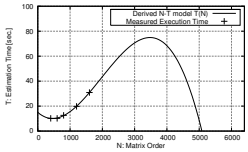

図 2: 破綻したモデル

この結果より N の測定範囲について，(1) 点数は 5 点程度でよいが，(2) 測定範囲が小さいとモデルが破綻するので，ある程度大きい N を測定する必要があることが分かった．

表 4: 各サイズ N における予測最良構成の誤差

Size	$\epsilon = (\tau - \hat{T})/\hat{T}$			$e = (\hat{\tau} - \hat{T})/\hat{T}$		
N	N9	N5	NS	N9	N5	NS
3200	-0.018	0.019	-0.106	0.000	0.000	0.608
4800	-0.099	-0.080	-0.559	0.074	0.074	0.238
6400	-0.096	-0.095	-0.787	0.022	0.022	0.134
8000	-0.124	-0.146	-0.983	0.015	0.015	0.100
9600	-0.093	-0.139	-1.146	0.000	0.000	0.099

4 おわりに

本研究では，不均一クラスタ上で既存の科学技術計算応用を負荷分散するため，マルチプロセス法について検討した．最適な PE 群およびマルチプロセス数を選択するため，HPL の実測値から実行時間予測モデルを構築して，実際に (準) 最良構成を予測することに成功した．

参考文献

[1] A. Petitet et al. “HPL – A Portable Implementation of the High-Performance Linpack Benchmark for Distributed-Memory Computers,” http://www.netlib.org/benchmark/hpl/.

[2] 岸本芳典, 市川周一: “不均一クラスタ上での実行時間予測モデルとその評価,” 情処研報 2003–HPC–95, pp. 161–166 (2003).