

メトリックス・統計調査による ベクトル処理用プログラムの特徴の把握

5D-10

前富 博† 島崎 真昭‡

†九州大学工学部 ‡九州大学大型計算機センター

1. はじめに

スーパーコンピュータの自動ベクトル化コンパイラの発達により、ベクトルスーパーコンピュータのプログラム開発はかなり容易になった。しかし、自動ベクトル化コンパイラの能力は近年大変に向上したとは言え、ベクトルスーパーコンピュータの能力を最大限に引き出すためには、利用者がプログラムをベクトル演算向きに修正することが重要である。このプログラムのチューニングはシステムに備え付けの支援ツールを利用しながら行うのが一般的であるが、計算機毎のノウハウも必要とし、利用者の負担は相当なものとなっている。

このようなベクトル処理用のプログラムのソーステキスト上の特徴を把握することを目的として、相当規模のプログラムの統計的調査を行った。その結果および考察について報告する。

2. 調査対象プログラムと調査方法

2.1 調査対象プログラム

今回の調査対象となったのは、下記の FORTRAN のプログラムグループである。

- i) LAPACK^[1]
- ii) Perfect ベンチマークプログラム^[2]
- iii) NUMPAC (スカラー版)^[3]
- iv) NUMPAC (ベクトル版)
- v) 九州大学のスーパーコンピュータ用ベンチマークプログラム
- vi) 行列計算ライブラリー^[4]

(1)、(3)、(4)、(6) はライブラリーレベルのプログラムであり、(2) は応用レベルのプログラムである。(5) は核ループレベルのプログラム、ライブラリーレベルのプログラム、応用レベルのプログラムを含んでいる。(6) のプログラムの著者には(1) のプログラムの著者の一人である DR. Dongarra が含まれている。

2.2 調査項目と調査方法

スーパーコンピュータの性能評価、特に実行時の性能を問題とすると、ソフトウェアだけでなくハードウェアを含んだシステム全体の評価が中心となってしまう。今回はソフトウェアを重点とし、プログラムのテキストレベルの実態調査を実施した。

本稿では、Knuth^[5] の FORTRAN プログラムの実態調査や McCabe^[6] の V(G) を参考にして、ベクトル処理プログラムのメトリックスで有効なものが定義できないかを調べる基礎調査の目的で、FORTRAN の文の頻度調査、V(G) の計測を行なった結果を述べる。

米国 Bell 研究所で作成されたプログラムに f2c という FORTRAN から C への変換プログラムがある。これは LEX と YACC を使い作成されているため、今回はこれをベースにして、主に YACC の部分を変更して、計測用プログラムを作成した。統計整理には九州大学大型計算機センターの UXP の S システムを利用した。V(G) はプログラムのフローグラフについて以下の式に定義される。

$$V(G) = e - n + 2p$$

e はエッジの数、n はノードの数、p は独立なコンポーネントの数である。本稿では、この目的で以下の式を計算した。

$$V(G) = 1 + (\text{DO 文の数}) + (\text{論理 IF の数}) + (\text{算術 IF の場合、分岐の数に応じて 2 又は 1})$$

3. 調査の結果と考察

3.1 FORTRAN の文の頻度調査結果

今回の調査では、実行文と非実行文の割合、文の種類別の頻度の割合、代入文の右辺の項数の3種類の項目について計測を行なった。

結果の考察として、FORTRAN プログラムは代入文と DO 文とで約 50~60% を占めることが一般的といえ、ループを主体とする計算のプログラムの特徴といえる。特に NUMPAC ベクトル版、九大ベンチマーク、行列計算ライブラリーといったベクトル計算向きのプログラムではその特徴が明確にでている。

次に代入文の右辺の項数であるが、他のソフトウェアと比較して LAPACK および Perfect ベンチマークで

Metrics and Statistical Analysis of Programs for Vector Processing

Hiroshi Maetomi, Masaaki Simasaki
Faculty of Engineering, Kyushu Univ.
Computer Center, Kyushu Univ.

は1、2項だけの単純な代入文が多く、一方 NUMPAC ベクトル版や九大ベンチマークでは1項だけの代入文の割合よりも3項以上の割合が多い。1CPU内に多くの並列演算パイプラインを持つベクトル計算機特に日本のベクトル計算機の場合、代入文の右辺の項の数が多いほど演算器利用率を高め保てる可能性が高く、NUMPACベクトル版や九大ベンチマークの結果はプログラミング上推奨されていることの結果の一つといえる。米国のスーパーコンピュータの場合、1CPUの演算パイプラインよりむしろCPUレベルの並列方式が早くから採用されてきた。LAPACKやPerfectベンチマークはそのことが反映されているといえる。

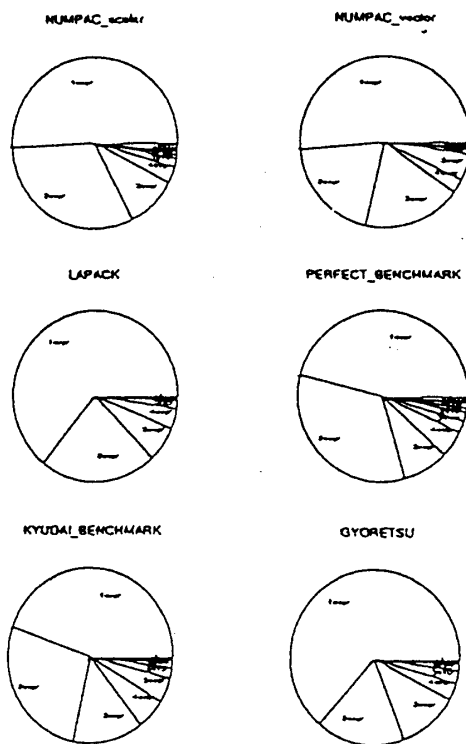


図1: 代入文の右辺の項数調査

3.2 V(G)に関する調査結果

結果から文の数とV(G)には線形の回帰関係があり、その係数である文当りのV(G) (表1参照)はソフトウェアにより特徴的な値を持っているといえる。例えば、NUMPACベクトル版や九大ベンチマークのその値は他のプログラムグループの値より大きな値を持っている。これはベクトル計算機の場合、ベクトル化率向上のためのDO文の比率向上が推奨されており、その影響と思われる。

次にNUMPACのスカラ版、ベクトル版の両方にある共通のサブルーチンを取り出して比較してみた。共通のサブルーチンについては表1の場合と異なり、スカラ版、ベクトル版のV(G)/文の数は0.24,0.25では

と同じであった。ベクトル版はプログラムの書換えにより、ループアンローリングが多用されている。この書換えにより、ベクトル版はスカラ版よりDO文の数が増加しているがプログラム全体の行数も同様に増加しているため、V(G)/文の数に変化がなかったのである。一方代入文の右辺の項数はベクトル版ではスカラ版より増加していることがわかった。

プログラム名	V(G)/文の数
NUMPAC(scalar)	0.132
NUMPAC(vector)	0.260
LAPACK	0.158
PERFECT	0.143
KYU-BENCH	0.262
GYORETSU	0.191

表1: 各プログラムの文当りのV(G)の割合数値データ

4. まとめ

実用されている相当規模のプログラムを対象にして、ベクトル処理用プログラムのメトリクス・統計調査を行なった結果について報告した。文の種類の種類統計などに興味深い結果が見られた他、V(G)/文の数がプログラムグループに特徴的であることを示した。プログラムをベクトル計算機向きに書き換える手法として良く知られるループアンローリングでは文の数、DOループの数が増えるが、V(G)/文の数は変化しないこと、代入文の右辺の項数を増加するのに効果のあることがデータとして示された。

参考文献

- [1] Anderson, E., Dongarra, J.J. "LAPACK Working Note 18.", Univ.Tennessee, 1990.
- [2] Cybenko, G., Bruner, J., Ho, S. Sharma, S. "Parallel computing and the Perfect Benchmark", Proceedings of the International Symposium on Supercomputing, Fukuoka, pp.99-110, Nov.6-8, 1991,
- [3] 二宮市三監修、「ライブラリー・プログラム利用の手引き(数値計算編)」、NUMPAC vol.1-3、名古屋大学大型計算機センター、1991.
- [4] 小国力編著、村田健郎、三好俊郎、ドンガラ,J.J.、長谷川秀彦 著、「行列計算ソフトウェア」、丸善株式会社、1991.
- [5] Knuth, D. E. "An empirical study of FORTRAN programs", Software Practice and Experience, vol. 1, pp. 105-133, 1971.
- [6] McCabe, T. J. "A complexity measure", IEEE Trans, on Software Engineering SE-2, 4, pp. 308-320, 1976.