

手続き間情報の解析と整理のツール AUDIE について†

牛島和夫†† 田町典子††

プログラムは読んで理解しやすいことが重要である。しかしソースプログラム自身がわかりやすく書かれていても、手続き間関係あるいは処理対象となるデータの役割を大域的な視野で把握することは容易でない。プログラムが大きくなるとこれらの関係がますます複雑になるため、何らかの支援ツールが必要となる。

既存の静的解析ツールは出力結果が大量になりがちで必ずしも利用しやすいとはいえない場合が多いので、必要な情報が探しやすいよう整理されていることが求められる。また動的解析結果を静的解析結果に加えて出力することによりさらに有効な情報を得ることが期待できる。

以上の考察に基づき手続き間情報の解析と整理のツール AUDIE を作成した。AUDIE は FORTRAN プログラムを静的に解析してその情報を蓄積し、さらに手元にある実行回数計数ツール FORDAP による動的解析情報を組み合わせ、整理して提示する。このシステムは 1980 年 12 月から九大大型計算機センターで一般の使用に供している。

本論文ではまず AUDIE の機能を概説し、その実現と AUDIE の作成するプログラムデータベースについて述べ、AUDIE の使用例を通じてその効用を説明する。なお AUDIE は他環境への移し換えを考慮して Portable FORTRAN で記述している。

1. はじめに

ソフトウェアの保守におけるコスト削減が重視されるに伴って、わかりやすくエラーの入りにくいアルゴリズム、データ構造、あるいは表記法が奨励されるようになってきた。しかしソースプログラム自身がわかりやすく書かれていても、手続き間関係あるいは処理対象となるデータの役割を大域的な視野で把握することは容易でない。プログラムが大きくなるとこれらの関係がますます複雑になるため、何らかの支援ツールが必要となる。

手続き間情報を得るためには FORTRAN やそのサブセットに対して静的解析を行う DAVE¹⁾ や PFORT verifier²⁾ 等の解析結果を使用することができる。しかしこれらの出力結果は大量になりがちで必ずしも利用しやすいとはいえない。このため出力が多すぎないこと、必要な情報が探しやすいよう整理されていることが求められる。また動的解析結果を静的解析結果に加えて出力することによりさらに有効な情報が得られる³⁾。

以上の考察に基づき手続き間情報の解析と整理のツール AUDIE (a tool for analyzing and docu-

menting interprocedural information including execution counts)⁴⁾ を作成した。AUDIE は、FORTRAN で書かれたプログラムを静的に解析してその情報を蓄積し、さらに我々の研究室で作成した実行回数計数ツール FORDAP⁵⁾ による動的解析情報を組み合わせ、整理して提示する。なおこのシステムは 1980 年 12 月から九大大型計算機センター FACOM-M 200 OSIV の下で一般の使用に供している⁶⁾。

以下 2 章で AUDIE の機能を概説し、3 章で AUDIE の実現と AUDIE の作成するプログラムデータベースについて述べる。4 章では AUDIE の使用例を通じてその効用を説明する。なお本稿で“手続き”とは、FORTRAN ではプログラム単位が対応する。

2. AUDIE の機能

2.1 構成

AUDIE は静的解析を行って静的情報を蓄積する静的解析部、動的解析を行って動的情報を蓄積する動的解析部、および蓄積された情報から必要な情報を取り出して編集し出力する表出力部の 3 つの部分から成る (図 1 参照)。各部分は独立に実行することができ、解析結果はファイルに保存されるので次のような利点を有する。

- 実行不可能なプログラム (作成途中、テスト段階、あるいは環境による制限等で実行できないもの) に対しても静的解析結果は得られる。

† A Tool for Analyzing and Documenting Interprocedural Information Including Execution Counts by KAZUO USHIJIMA and NORIKO TAMACHI (Department of Computer Science and Communication Engineering, Kyushu University).

†† 九州大学工学部情報工学科

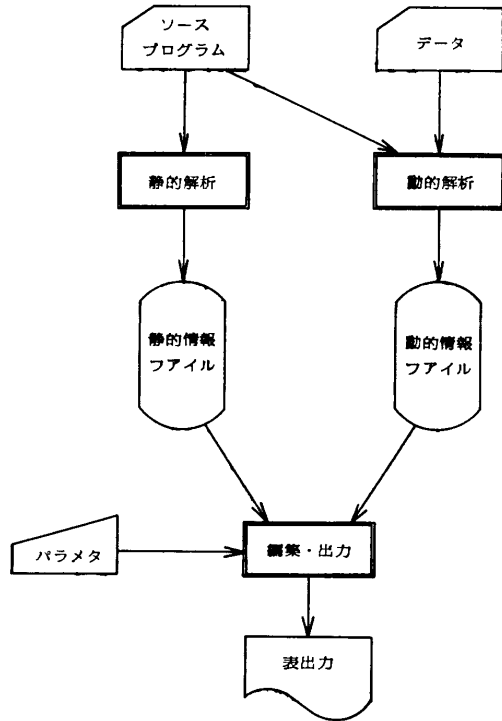


図 1 AUDIE の構成

Fig. 1 Overview of AUDIE.

● 幾通りかの実行状況（実行時のデータが異なるために起こる）を比較する場合に静的解析を繰り返す必要がない。

● 解析を繰り返すことなく何度でも出力を得られるので、必要な時に必要な表だけ出力させることができる。

2.2 文法エラーの検出

AUDIE の静的解析部は入力されるソースプログラムにコンパイラで検出されるエラー（手続き内文法エ

ラー）がないことを条件としているが、次のような手続き間のエラーを検出する。

- 手続きの再帰的呼出し・相互呼出しのテキスト上の可能性。
- 引数の受渡しにおける実引数と仮引数の数・型・使用方法の不一致。

2.3 出力情報

AUDIE の表出力部は以下の表またはグラフをユーザの選択により出力する。

(1) G表 (General descriptions of procedures in source program; 手続きの概略表, 図2参照): 各手続きに対する以下の情報を手続き名のアルファベット順に出力し、最後に合計値を出力する。

● ソースプログラム中何番目の手続きであるか (NO 欄)。

● 手続きに含まれるソースプログラムの行数 (CARD 欄)。

● 手続きに含まれる実行文の数。ただし、論理 IF 文を 2, CONTINUE 文を 0 と数える (STEP 欄)。

● 実行時に一度も実行されなかった実行文の数 (UN-EXECUTED 欄)。

● 実行時に他から呼ばれた回数の合計 (INVOKED 欄)。

● 実行時に各文が実行された回数の合計 (EXECUTION 欄)。

(2) P表 (Procedure invocation table; 手続き呼出し表): 手続き間の呼出し関係を行列の形で表わしたもの。

(3) C表 (Common block declaration table; 共通ブロック宣言表): 共通ブロックが手続きに組み込まれている状態を行列の形で表わしたもの。

(4) W表 (Within-procedure data usage table; 手続き内データ使用状況表, 図3参照): 各手続き中における仮引数および共通ブロック内の変数または配列（以下共通変数と呼ぶ）についての、手続き内だけからわかる使用状況を記号で示す (X: not used, R: referenced, M: modified, E: actual parameter)。この表では、手続き呼出しにおいて共通変数を介して行われる情報の受渡しは無視し、引数を介して行われるものは“E”で示す。

なお配列の使用状況については配列要素間の区別をしない。共通ブロックが各手続きにおいて同じ構造をもつよう宣言されていない場合（共通ブ

(G) GENERAL DESCRIPTIONS OF PROCEDURES IN SOURCE PROGRAM						
NAME	NO.	CARDS	EXECUTABLE	UN-EXECUTED	INVOKED	EXECUTION
BLACK	11	25	11	0	684	8892
BLUE	9	13	7	0	2109	18981
BROWN	10	16	7	1	5	30
GOLD	7	42	34	0	1	18009
GRAY	2	141	126	2	1	2349
GREEN	15	19	16	0	5	380
INDIGO	6	64	40	0	4	138934
LILAC	4	16	9	0	13	195
ORANGE	16	95	85	4	4	747918
PURPLE	13	12	8	0	60	1920
RED	12	22	11	0	95	2888095
SILVER	5	7	4	0	4	656
VIOLET	8	14	10	0	1430	5725
WHITE	3	22	16	0	24	624
YELLOW	14	37	32	5	60	1180
MAINPR	1	501	383	2	1	20000
TOTAL		1046	799	14	4500	3853888

図 2 G表: 手続きの概略表

Fig. 2 G-Table: General descriptions of procedures in source program.

(W) WITHIN-PROCEDURE DATA USAGE TABLE

PARAMETER	1	2	3	4	5	6	7	8
1	I I I	I R A I	R A I	I I I	I I I	I I I	I I I	I I I
2	I I I	R R A I	I I I	I I I	I I I	I I I	I I I	I I I
3	I I I	R R A I	I I I	I I I	I I I	I I I	I I I	I I I
4		I I I	R A I	I I I	I I I	I I I	I I I	I I I
5		I I I	R A I	I I I	I I I	I I I	I I I	I I I
6		I I I	R A I	I I I	I I I	I I I	I I I	I I I
7		I I I	R A I	I I I	I I I	I I I	I I I	I I I
8		I I I	R A I	I I I	I I I	I I I	I I I	I I I

COMMON VAR	TYPE
ALPHA ITALY	IA R R X . RM R . . . R
BETA USA	RA M R X
USSR	RA M R X
CHI CHINA	RA R R R R R RM
DELTA NORWAY	IA R M
SWEDEN	RA R M
GAMMA GREEK	RA RE R R RME
IOTA POLAND	RA X M . R R
KAPPA SPAIN	RA R RM
LAMBDA FRANCE	RA R RM
MU GERMAN	RA R RM
NU X	RA R R R . RM
PHI CANADA	RA M R
PI SUDAN	RA X X X E E RM
CONGO	RA R R R R M R
PSI CEYLON	RA M R
RHO SYRIA	RA R R R R M R
SIGMA IRAQ	I R X . X X RM
LIBYA	I X X . X X RM
JORDAN	I R R . R X RM
IRAN	I X R . X X RM
TAU EGYPT	RA R M R ME
THETA JAPAN	IA R M
XI BRAZIL	R R RM
PERU	R R RM
CHILI	R R RM
CUBA	R R RM

図 3 W表：手続き内データ使用状況表

Fig. 3 W-Table: Within-procedure data usage table.

ロック中の記憶領域が、ある手続きでは配列と宣言され他の手続きではそれに対応する領域が複数個の配列または変数として宣言されている場合等), その共通ブロックに対しては共通ブロック名のみが出力され、変数情報は出力されない。

また使用の便を考慮して、データの型を、仮引数は PARAMETER 欄の上段に、共通変数は変数名のす

(B) BETWEEN-PROCEDURE DATA USAGE TABLE

PARAMETER	1	2	3	4	5	6	7	8
1	I I I	I R A I	R A I	I I I	I I I	I I I	I I I	I I I
2	I I I	R R A I	I I I	I I I	I I I	I I I	I I I	I I I
3	I I I	R R A I	I I I	I I I	I I I	I I I	I I I	I I I
4		I I I	R A I	I I I	I I I	I I I	I I I	I I I
5		I I I	R A I	I I I	I I I	I I I	I I I	I I I
6		I I I	R A I	I I I	I I I	I I I	I I I	I I I
7		I I I	R A I	I I I	I I I	I I I	I I I	I I I
8		I I I	R A I	I I I	I I I	I I I	I I I	I I I

COMMON VAR	TYPE
ALPHA ITALY	IA R R X . RM R . . . RM
BETA USA	RA M R R RM
USSR	RA M R R RM
CHI CHINA	RA R R R R R RM
DELTA NORWAY	IA R R RM
SWEDEN	RA R R RM
GAMMA GREEK	RA R R R R RM
IOTA POLAND	RA X M . R RM
KAPPA SPAIN	RA R RM
LAMBDA FRANCE	RA R RM
MU GERMAN	RA R RM
NU X	RA R R R . RM
PHI CANADA	RA M RM
PI SUDAN	RA X X X R R X . RM
CONGO	RA R R R R RM R . RM
PSI CEYLON	RA M RM
RHO SYRIA	RA R R R R M RM
SIGMA IRAQ	I R X . X X RM
LIBYA	I X X . X X RM
JORDAN	I R R . R X RM
IRAN	I X R . X R RM
TAU EGYPT	RA R M R RM
THETA JAPAN	IA R RM
XI BRAZIL	R R RM
PERU	R R RM
CHILI	R R RM
CUBA	R R RM

図 4 B表：手続き間データ使用状況表

Fig. 4 B-Table: Between-procedure data usage table.

ぐ右側の TYPE 欄に、記号で表示する (I: integer, R: real, L: logical, C: complex, D: double precision, A: array, 仮引数の場合さらに, S: sub-program name). 後述の B表, O表でも同じ情報を同じ形式で示している。

(5) B表 (Between-procedure data usage table; 手続き間データ使用状況表, 図4参照): 仮引数およ

値で表示したものを。

(10) Hグラフ (Hierarchical invocation graph; 呼出しグラフ): 手続きの呼出し関係 (P表に示される) を木状のグラフに編集したもの。

2.4 使用上の制限

以下に該当する場合、静的解析部はメッセージを出力して適当な処理を行う。

(1) 共通ブロックの構造が手続き間で異なる場合、その共通ブロック中の変数は解析対象としない。ある手続きでは配列として宣言されている領域が他の手続きでは複数の配列または変数に対応する場合はこれにあたる。またこれと同じ機能を EQUIVALENCE 文を用いて実現している場合、COMMON 文で宣言された名前のみが対象となる。

(2) 同じ名前の手続き (主プログラム、ブロックデータを含む) が複数個あると最初のものだけを解析対象とする。

(3) 手続きの入口は唯一であること。ENTRY 文は無視し、これによって宣言される手続き名はソースプログラム中に本体がないものと見なす。

(4) 作業領域の大きさによる制限のため、約 1,000 ステップを越える手続きがあるとそれを無視して解析を続行する。作業領域を拡張すれば処理可能となるが、わかりやすいプログラムをつくるという観点からあまり大きな手続きを書くことは奨められない。

(5) 手続き名を引数とする場合、対応する実引数と仮引数とがともに手続き名であることしか検査しないので、実引数として渡された手続きの呼出しにより再帰的呼出しが起こる可能性があったとしてもエラーとして検出しない。ただしプログラムデータベース (3章参照) 内には、このような検査に対応できる情報を蓄えているので、必要に応じて取り出すことができる。

3. AUDIEの実現とプログラムデータベース

AUDIE の作成する解析情報ファイルはプログラムデータベースとして整備中である。これには静的情報を格納するプログラムデータベース S (static) と動的情報を格納するプログラムデータベース D (dynamic) がある。プログラムデータベース D は動的解析部により動的解析終了後直ちに作成される。 D の各レコードはソースプログラムの各行と 1対1に対応し、各実行文の開始行に対応するレコードにその文の実行数情報が格納される。

プログラムデータベース S は静的解析部により作成され、手続き内情報を格納する L (local) と手続き間情報を格納する G (global) に分かれる。 L は各手続きに対応して作成される小ファイルの集合であり、処理対象手続きに対応する小ファイルだけが主記憶中に置かれる。これは名前、基本ブロック、および基本ブロック中での名前の使用状況の3つに分かれ、それぞれの属性と相互を関連付けるポインタとを格納する。各小ファイルは対応する手続きの解析が終了することによって出力されて蓄えられる。

G は L の作成終了後プログラム全体に対して作成され、名前、その結合に伴う大域的データに関する情報、および L の管理のための情報の3つに分かれる。手続きの結合関係を考慮して解析を行うためには、解析対象手続きから呼ばれる手続きすべてがすでに解析済みでなければならない。すべての手続きをこのような条件下で解析するためには手続きの解析順序が問題になる。手続き内解析で作成された L の各小ファイルは、この条件を満たす順序で主記憶内に入力され手続き間解析が行われる。しかしこの小ファイルには解析中の手続きに関する情報しか含まれておらず、手続き呼出しに対応して他手続きに関する情報が必要な場合はこれを作成中の G から得なければならない。このとき使用する G 中の情報は、呼び出される手続き (解析順序により現在解析中の手続きに先立って解析される) の解析終了時にあらかじめこの G に追加されている。静的解析終了後 G はデータベースとして蓄えられる。

いったん静的解析の済んだプログラムの一部を書き替えた場合、その手続きのみを再解析して、プログラムデータベース L 中の対応する小ファイルだけを更新することができれば便利である。しかしその際でも変更に伴って G は全面的に作り直さねばならぬ。したがって、現在はプログラム全体を再解析して L と G とを更新している。

データフロー解析結果は、プログラムの誤り発見にとって有用である¹⁾。AUDIE はこれに対応できるような情報も静的解析部によって S の中に蓄えている。しかし、AUDIE が解析対象と想定しているそれほど小さくない規模のプログラム全体を一括してデータフロー解析するのはコストがかかりすぎる。まず全体を把握した後、蓄積された情報に基づき狭い範囲に限定して解析を行うようなシステムを実現するのが現実的ではないかと考えている。

4. AUDIE 使用の例

ここでは九大大型計算機センターのユーザプログラムのひとつを解析して得られた結果 (図 2~6) を用いて AUDIE 使用の一端を説明する。

G表 (図 2) は解析の対象となる各手続きに関するテキストとしての大きさと実行時の利用頻度とを手続きレベルでまとめたものである。UN-EXECUTED 欄の値からは実行時に実行されなかった文の存在を知ることができる。プログラム中に実行されていない文があれば完全にテストされたとはいえないので、プログラムテストに有用な情報である。

INVOKED 欄からは呼出し回数の集中を、また EXECUTION 欄からはその手続きに含まれる各文の実行回数の総和を知ることができ、プログラムの効率改善に役立つ。Knuth によれば I/O バウンドでないプログラム (入出力が計算の大部分を占めないようなプログラム) ではソーステキストの 3% 中に実行時間の 50% が集中しているという⁷⁾。手続き RED では実行文の数に対応する EXECUTABLE 値は総和の約 1.4% (11/799) にすぎないが、EXECUTION 値は総和の 75% (2888095/3853888) がこれに集中している (G表)。そこで RED 内での局所的な手直しによって全体の実行時間を 2/3 に短縮することができた。

手続きとその環境 (引数および共通変数の使用状況) を調べるには S・W・B の各表 (図 3~5) が役立つ。ここで再び RED を例に取り上げる。

RED は 95 回呼び出されるが (G表の INVOKED 欄) それは YELLOW 内 2カ所から (S表) に限られる。そこでまず RED を呼び出している YELLOW の呼出し環境を調べる。YELLOW は主プログラム内の 1カ所のみから呼ばれ、RED を 2カ所で、PURPLE と SQRT を各 1カ所で呼ぶ (S表)。SQRT は基本外部関数でありその呼出しによって外部の環境を変えない。そこで RED と PURPLE の呼出しによって YELLOW の環境に変化があるかどうか調べる。

このふたつの手続きはともに他の手続きを呼び出さない (S表)。W表によればこれら (共に関数副プログラム) においては引数も共通変数もその使用状況はすべて “R” または “X” である。したがってこれらの手続きを呼び出しても、その際に実引数として与える変数や、これらの手続きが参照する共通変数の値が変化することはない。

W表において YELLOW の 5個の引数は “RE”

“RE” “RE” “RM” “RM” と表示されている。ここで “E” の表示は YELLOW の第 1~第 3 引数が手続き呼出しの実引数として用いられることを示す。したがって YELLOW の第 1~第 3 引数の最終的な使用状況を知るためには、YELLOW から呼ばれる手続き内でその実引数がどのように使用されているかを知らねばならない。YELLOW が呼ぶ PURPLE と RED は引数を参照するだけだから、YELLOW の第 1~第 3 引数も参照されるだけのはずである。この解析結果が B表の “R” “R” “R” “RM” “RM” とし

て表示されている。YELLOW における共通変数はもう少し複雑である。共通ブロック SIGMA 中の 4つの変数は W表では “X” (宣言されているが使用されていない) と表示されるが B表ではそのうちの変数 IRAN に関して “R” に変わっている。これは RED 中で IRAN が参照され、RED を呼ぶ YELLOW にその使用状況が伝達されたためである。共通ブロック BETA・CHI・DELTA・LAMBDA・THETA は、W表では未宣言とされているが B表ではその中の変数が “R” になっている。これはこれらが主プログラムで宣言され、YELLOW が呼ぶ PURPLE あるいは RED で参照されるため、YELLOW にも暗黙に組み込まれているとみなすためである。なお B表で “X” (宣言されているが使用されない) と表示される共通ブロック (手続き BROWN・GOLD における ALPHA・IOTA 等) は、その宣言をそれぞれの手続きから取り除いても構わない。

D表と A表は実行回数情報をまとめたものであり、D表は手続きの呼出し回数、A表は引数や共通変数の引用回数を示している (紙数の関係で出力例は省略した)。D表と S表を並べてみるとプログラムの動的振舞いを大局的につかむことができ、A表と O表 (場合によっては W表、B表) とを並べてみるとデータの引用に関する動的振舞いをうかがうことができる。

主プログラムは YELLOW を 1カ所から 60 回、YELLOW は RED を 2カ所から 95 回呼出している。共通ブロック THETA 中の配列 JAPAN は主プログラム中 16カ所に現われ (O表) 使用状況は “M (代入)” (W表)、また RED 中では 4カ所に現われ使用状況は “R (参照)” である。すなわち配列 JAPAN (16個の要素を持つ) は主プログラムで値を設定され、もっぱら RED で参照される。アクセス回数は主プログラム中 16カ所で 16 回、RED 中 4カ所で 4,900 回である。RED の呼出し回数 95 回に対して 16 個しか要

素を持たない配列 JAPAN が 4,900 回も使用されていることは、添字付変数 JAPAN (I, J) を同じ添字値で何度も参照していることを意味し、計算過程でこの値の括り出しの可能性を示唆している。

このように A 表を O 表・W 表・B 表と並べてみると手続き間のデータの授受を知ることができる。プログラムが異常終了した場合など、エラーフラグの役割を果たす変数がプログラム中にあらかじめ用意してあれば、その変数がどの手続きで設定されたかを見出すことができる。またアクセス回数 0 の変数の存在が知ればテストの立場から有用な情報となる。

5. おわりに

AUDIE は静的解析結果・動的解析結果をプログラムデータベースという形で外部記憶に蓄積し、これに基づいて表出力を行う。しかし表中に現われるのは蓄積されている情報の一部にすぎず、実際にはもっと詳しい情報が格納されている。なるべくコンパクトでわかりやすい形で出力するという設計思想により、AUDIE ではこれらをすべて出力することはしないが、全体を把握したうえで部分的に詳細な情報を得たい場合もある。そのため AUDIE の作成するプログラムデータベースを会話的に検索して、ユーザのより詳細な要求にも対応できるシステムを作成中である。そのうち H グラフに対する視野指定機能、および名前によるソースプログラム検索機能はすでに単体として試作されている。このほか指示した表に対する視野指定機能、表中の値あるいはプログラムデータベース中の項目の検索機能、複数の動的解析結果の比較による未テスト部分抽出機能等を準備中である。

手続き間情報の授受は一般に引数と大域的データを介して行われる。引数の使用モードを仮引数宣言部で明記する手段を持った PASCAL などの言語の場合でも、大域的データによる手続き間情報の授受は理解するのが面倒である。大域的に定義された型や大域的に

宣言された変数と手続きとの関係を解析して整理する (AUDIE のような) ツールは、これらの言語に対しても有用性を発揮するものと考えられる。

AUDIE は実行回数計数部⁵⁾を除いてカード枚数約 5,400 枚・3,000 ステップから成る。すべて Portable FORTRAN²⁾ で記述され、ファイル形式は順編成なので、他環境への移換も容易である。九大大型計算機センターの FACOM-M 200 上での大きさは約 300 キロバイト、CPU 時間はカード約 8,000 枚、手続き 99 個、コモンブロック 23 個の入力プログラムに対して静的解析部約 32 秒、出力部約 4 秒である。

参 考 文 献

- 1) Osterweil, L. J. and Fosdick, L. D.: DAVE-A Validation Error Detection and Documentation System for FORTRAN Programs, Software-Practice and Experience, Vol. 6, pp. 473-486 (1976).
- 2) Ryder, B. G.: The PFORT Verifier, Software-Practice and Experience, Vol. 4, pp. 359-377 (1974).
- 3) Gannon, C.: A Debugging, Testing, and Documentation Tool for JOVIAL J 73, Proceedings COMPSAC 80, pp. 634-639 (1980).
- 4) 牛島, 田町: 実行回数を含む手続き間情報の解析と整理のツール, 電子通信学会技術研究報告, Vol. EC 79-57 (1980).
- 5) 牛島: FORTRAN プログラミングツール, 産業図書, p. 241 (1979).
- 6) 牛島, 田町: プログラム単位間情報の解析と整理のツール AUDIE の使用について, 九大大型計算機センター広報, Vol. 13, No. 4, pp. 469-488 (1980).
- 7) Knuth, D. E.: An Empirical Study of FORTRAN Programs, Software-Practice and Experience, Vol. 1, pp. 105-133 (1971).

(昭和 56 年 3 月 20 日受付)

(昭和 56 年 9 月 7 日採録)