

フローグラフや状態遷移図の特性のパス数による分析
若杉忠男

若杉情報技術コンサルタントオフィス
〒251 神奈川県藤沢市片瀬山3-11-1 電話 0466-23-4832

フローグラフや状態遷移図のL個の連続したリンクを長さLのパスと定義し、長さLのパスの個数を P_L 、それをベクトル状に並べた $\{P_L\}$ をパスベクトル、 P_{L+1}/P_L を P_L の増加率と呼ぶ。パスベクトル $\{P_L\}$ は連結行列によって求められる。この $\{P_L\}$ からつぎのようなことが分かる。Lを大にしたとき、フローグラフが(1)ループを含まない場合には P_L は0となり、(2)ループを含むがそのループが互いに共通のノードをもたない場合には P_L はLのべき乗のオーダーで増加し、(3)共通のノードをもつループを含む場合はLの指数関数のオーダーで増加する。

On analysis of characteristics of
flowgraphs and state transition diagrams by numbers of paths
Tadao WAKASUGI, Member

3-11-1 Kataseyama Fujisawa-city, 251 Japan

Concerning the flowgraphs and state transition diagrams, a sequence of consecutive links is called path length L, the number of paths of length L is described as P_L , their sequence is called path vector $\{P_L\}$, and P_{L+1}/P_L is defined as increasing rate of P_L . From the $\{P_L\}$, we can obtain the following results; Flowgraphs and state transition diagrams are classified into 3 classes depending on the characteristics of loops they contain. When L increases, (1) if no loops are contained then P_L becomes 0, (2) if loops have no common nodes then P_L increases at a rate proportional to a power of L, and (3) if loops have common nodes then P_L increases at a rate proportional to an exponential function of L.

1. はじめに

プログラムの複雑度の指標については多くの論文がある。最も単純な指標としてステートメント数があるが批判は多い。またプログラムの設計の複雑さを示す尺度として、McCabeのサイクロマチック数がある。これは(1)式のようにリンクとノードの差だけが評価の対象となり、プログラムの規模は示さない。

$$\text{サイクロマチック数} = (\text{リンク数}) - (\text{ノード数}) + 2 \quad (1)$$

これらの他にもプログラムの性質を表す指標は多く、31種もの指標を列挙し比較している文献もあるが^[1]、ここでは、プログラム(またはプロトコル)の特性をフローグラフ(または状態遷移図)のパスの数で表すことを考える。

2. パスとフローグラフ

フローグラフは、ノードとそれをつなぐリンクからなる。リンクはプログラムフローチャートでは一連のステートメントに対応し、状態遷移図では状態の遷移に対応する。まず前提として次の定義をする。

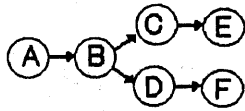
前提：フローグラフのパスとは、任意のノードから出発し任意のノードで終わる連続したリンクのシーケンスをいう。パスを構成するリンクの個数をパスの長さといい、また長さLのパスの個数を P_L とする。 P_{L+1}/P_L を P_L の増加率という。

図1のフローグラフの数 P_L は、ノードAからノードBに行くパスをA→Bと表すと、
長さ1のパス：A→B, B→C, B→D, C→E, D→Fの5本
長さ2のパス：A→B→C, A→B→D, B→C→E, B→D→Fの4本
長さ3のパス：A→B→C→E, A→B→D→Fの2本

となる。上記のフローグラフのパスの個数をベクトルの形で $\{P_L\} = (5, 4, 2)$ と記述し、パスベクトルという。

フローグラフはまた連結行列でも表される。連結行列とは図1に示すように、フローグラフのノードの数がN個の場合、 $N \times N$ のマトリックスとなり、各i, j要素はノードiからノードjへのリンクの数を表す。i, j要素が0ということは、

Flowgraph



Path vector

$$\{P_L\} = \{5, 4, 2\}$$

Increasing rate vector

$$\{P_{L+1}/P_L\} = \{4/5, 1/2\}$$

Connection matrix

	A	B	C	D	E	F
A	0	1	0	0	0	0
B	0	0	1	1	0	0
C	0	0	0	0	1	0
D	0	0	0	0	0	1
E	0	0	0	0	0	0
F	0	0	0	0	0	0

図1 ループをもたないフローグラフとその連結行列, $\{P_L\}$, $\{P_{L+1}/P_L\}$ の一例

i と j がリンクで直接はつながっていないことを示す。こうして作った行列を M とし、このマトリックスの i, j 要素を m_{ij} と表すと、行列 $M \times M$ の各要素は、

$$\sum_{k=1}^N m_{ik} \times m_{kj} \quad (2)$$

となり、これはノード i からノード j に長さ 2 で遷移するパスの個数を表す。同様にして行列 M^L の要素の合計は長さ L のパスの総数を表す。

P_L については次の定理が成立する。

定理 1: フローグラフについて、ノードの数を N, 一つのノードから出るリンクの最大本数を K とすると、次の式が成り立つ。

$$P_L \leq N \times K^L \quad (3)$$

証明

一つのノードから出るリンクの数の最大を K とすると、一つのノードから出る長さ L のパスは K^L 個以下である。したがって長さ L のパスの全体は $N \times K^L$ 以下である。

Q E D

3 フローグラフの特性とパス

パスベクトル $\{P_L\}$ の第一項 P_1 はリンクの総数であるから、ステートメント数にほぼ比例するのでフローグラフのボリュームを表し、また P_2 はブランチ数にほぼ比例するので、プログラムの複雑さを表す。また P_1 と P_2 はそれぞれパスカバレッジ試験というステートメントカバレッジとブランチカバレッジ試験に必要なチェック項目数の指標と考えられる。

実例として交換ソート (バブルソートともいう) を考える。そのフローグラフは図 2 の実線部分である。パスベクトルは、定理 1 から $P_L \leq 7 \times 2^L$ となるがこの見積りは大きすぎる。より正確に求めるには P_{L+1}/P_L を計算し、 $L \rightarrow \infty$ としたとき

の値を求めると、図 2 のグラフの曲線 a に示すようにサイクル 3 で振動する。その 3 つの値の幾何平均は 1.44 となる。よって $P_L = P_1 \times 1.44^{L-1} = 9 \times 1.44^{L-1} = 6.25 \times 1.44^L$ として連結行列で求めた数値と比較すると、 $L = 1 \sim 10$ の範囲ではその誤差は 10% 以内である。

図 2 のフローグラフは GOTO レスプログラミングの条件を満たしているが、ここで図 2 の点線に示す $D \rightarrow B$ というリンクを入れると GOTO レスプログラミングでなくなる。そうすると図 2 のグラフの曲線 b に示すように P_L も P_{L+1}/P_L の値も増加する。先と同じように連続する 3 つの P_{L+1}/P_L の幾何平均を求めると 1.59 となる。これによりプログラムに GOTO 文を 1 本挿入するだけでパス数が変わることが分かる。増加率の幾何平均は、これが大きいと試験しなければならない条件の組合せの数が増え、したがってプログラムの試験のしにくさを示す数値であると考えられる。すなわち、

定義: P_L の増加率 P_{L+1}/P_L の極限値を " P_L の増加率の極限" と呼ぶ。誤解を招かない場合には単に極限と呼ぶ。増加率がサイクリックな値をとる場合はその一連の極限値の幾何平均値とする。

先に述べたサイクロマチック数の式を連結行列で表現すると、

$$\text{サイクロマチック数} = \frac{\text{連結行列の要素の和}}{\text{連結行列の行数} + 2} \quad (4)$$

となる。この値が小さいということは、行列の 0 要素が多くなることである。0 要素の多い行列はべき乗してもその要素の数の増加が少ないから、 P_L の増加率の極限の値は小さくなる。図 2 の例で P_L の増加率の極限とサイクロマチック数を比較すると、 $D \rightarrow B$ のパスのない場合にはそれぞれ 1.44 と 4 となり、 $D \rightarrow B$ のある場合には 1.59 と 5 になる。

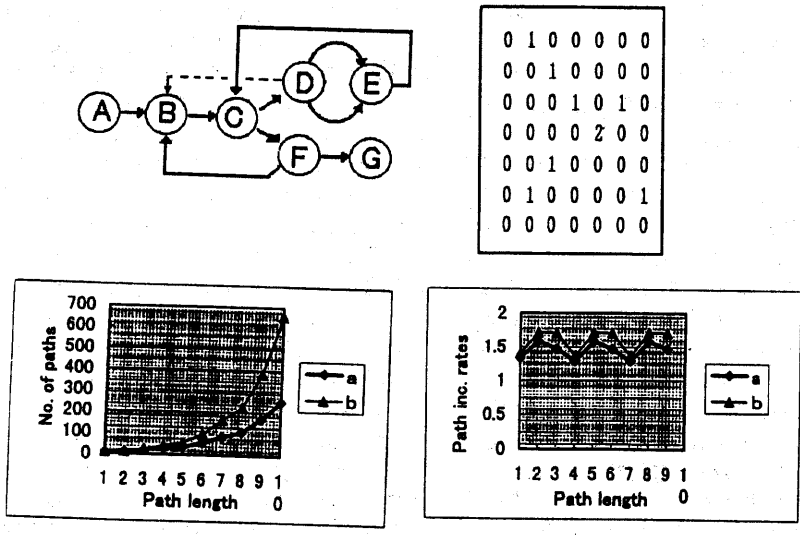


図2 交換ソートのフローグラフとその連結行列, $\{P_L\}$, $\{P_{L+1}/P_L\}$ の例

4 パスの個数

フローグラフについて P_L の評価を試みる。考察の対象とするフローグラフは、構造化プログラムの基本パターンであるシーケンスとブランチとループと、それらを合成したものであり、またスタートノード一つ、エンドノード一つで、ループからのとび出しやループへのとび込みのないものとする。またループとは無限に反復可能と考える。

4.1 ループと P_L の関係

定理2: フローグラフがループを含まない場合には、 N をノード数として $L \geq N$ のとき $P_L = 0$ となる。逆に $L < N$ となる長さ L のパスはループを含む。

証明: フローグラフのノードの個数を N とすると、長さ N 以上のパスは必ず同じノードを2度通ることになり、そこにループができる。したがって、ループがなければ $L \geq N$ となるパスはない。
Q E D.

図1がループのないフローグラフの例である。連結行列はスタートノードからエンドノードへと番号付けすると、図1のように左下半分がすべて0要素である三角行列にすることができる。三角行列は少なくとも N 回べき乗すると0行列になる。

以降ではノードAに始まりBに到達する長さ L のパスの個数を $P_{AB}(L)$ などと表す。

4.2 共通のノードをもたないループの場合

定理3: フローグラフがループを含みそのループが互いに共通のノードをもたない場合は、全体のノード数を N 、一つのパスでつながっているループの最大数を I 、 $L \geq N$ とすると、長さ L のパス数 P_L は L^{I-1} 次の多項式 ($I=1$ のときは振動または一定値) になる。

また、その各ループに含まれるノードの数をそれぞれ J_1, J_2, \dots, J_I 、その最小公倍数を J とすると、 J 個の連続する増加率の幾何平均は1に近づく。すなわち、

$$\left\{ \left(\frac{P_{L+1}}{P_L} \right) \times \left(\frac{P_{L+2}}{P_{L+1}} \right) \times \dots \times \left(\frac{P_{L+J}}{P_{L+J-1}} \right) \right\}^{1/J} \rightarrow 1 \quad (6)$$

証明: 図3のフローグラフを使い4段階に分けて証明する。

① まず図3 aのように各ループに一つしかノードがない場合を考える。これは $I=3, J=1$ の例である。最も多くのループを含むパス、 $A \rightarrow B \rightarrow C$ という長さ2のパス $P_{AC}(2)$ がループを一つまわると長さ3のパスになる。ループは3個あるから長さ3のパスは、 $P_{AC}(3) = P_{AC}(2) \times 3$ である。パスの長さが K でループが I 個ある場合、 $P_{AC}(K+1) = P_{AC}(K) \times I$ となる。同様にして

$P_{AC}(K+2) = P_{AC}(K) \times (I \text{ 個の中から } 2 \text{ 個を順序を考慮せずに選んだ場合の数} = {}_I H_2)$ となる。したがって I 個のループを含む長さ $L = K + (L - K)$ のパス数は、 K を固定して L を大としたとき、

$$P_{AC}(L) = P_{AC}(K) \times {}_I H_{L-K}$$

$$= P_{AC}(K) \times (L-K+1)(L-K+2) \cdots (L-K+I) / (I-1)!$$

$$= P_{AC}(K) \times \{ (L-K) \text{ の } I-1 \text{ 次多項式} \}$$

→ L の $I-1$ 次式

となり、 $P_{AC}(K)$ は $L \geq N$ の場合 L の $I-1$ 乗で増加する。つながっているループの数が I より少ないパスの個数は、 $I-2$ 次以下の多項式になるから無視でき、結局 P_L は L の $I-1$ 乗で増加する。

② 各ループがノードを二つ以上もつ場合、すなわち図3 bの $P_{AC}(L)$ について考える。この図では $J=2$ であるが、ループA, B, Cのノード数を J_A, J_B, J_C 、その最小公倍数を J とし、パス $P_{AC}(L)$ を長さ J だけ伸ばすと、 I 個のループのどれでも選択できるから、 $P_{AC}(L+J) \geq P_{AC}(L) \times I$ となる。したがって①と同じ論法によって、 $P_{AC}(L)$ は $L \geq N$ で L が J の倍数のとき L/J の $I-1$ 次多項式より大きい。

図3 aとbの $P_{AC}(L)$ を比較すると、図3 aの $P_{AC}(L)$ のうち各ループを0または偶数回まわるパスの数と図3 bの $P_{AC}(L)$ のパスの数とが等しい。しかし図3 aのパスのうち各ループを奇数回まわるパスは図bには対応するパスがなく、それだけ図bのパスの方が数は少ないので、

図3 aの $P_{AC}(L) >$ 図3 bの $P_{AC}(L)$ となる。したがって一般に図3 bの $P_{AC}(L)$ は、 L が J の倍数のとき、 L の二つの $I-1$ 次式にはさまれる。

③ L が J の倍数でない場合は、図3 bの両端のループA, Cにそれぞれ J_A, J_C 個のノードがあるとすると、パス $A \rightarrow B \rightarrow C$ を含むパスは $J_A \times J_C$ 組だけある。それ以外のパスは含むループの数が少ないので無視できる。図3 bの例では $J_A = J_C = 2$ だから、ループを3個含むパスは、 $P_{AC}, P_{DC}, P_{AF}, P_{DF}$ の4組である。 P_{DF} を考えると、これは P_{AC} から $A \rightarrow D$ と $F \rightarrow C$ という二つのパスを除いたものであるから、 $P_{DF}(L) = P_{AC}(L-2)$ という関係がある。同様に $P_{DC}(L) = P_{AF}(L) = P_{AC}(L-1)$ であり、図3 bの P_L は、 $A \rightarrow C$ を含まないパスの個数は無視することを \cong で表すと

$$L \text{ が偶数ならば, } P_L \cong P_{AC}(L) + P_{DF}(L)$$

$$\cong P_{AC}(L) + P_{AC}(L-2)$$

$$L \text{ が奇数ならば, } P_L \cong P_{DC}(L) + P_{AF}(L)$$

$$\cong 2 \times P_{AC}(L-1)$$

となり、いずれにしても L の $I-1$ 次式となる。

$J_A = J_C = 2$ でない場合についても同様にして、 $L > N$ なるすべての L について、 L が J の倍数のとき P_L は $J-1$ 次式となることがいえる。

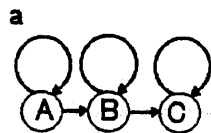
④ (6)式は、 P_L がサイクル J で (L の $I-1$ 次式) となるので、それを $\alpha_L L^{I-1}, \alpha_{L+1} L^{I-1}, \dots, \alpha_{L+J-1} L^{I-1}$ と表すと、 $\alpha_L = \alpha_{L+J}$ で、任意の L から $L+J-1$ について、(6)式は

$$\frac{\{\alpha_{L+1} (L+1)^{I-1} \times \dots \times \alpha_{L+J-1} (L+J-1)^{I-1} \times \alpha_{L+J} (L+J)^{I-1}\}^{1/J}}{\{\alpha_L L^{I-1} \times \alpha_{L+1} (L+1)^{I-1} \times \dots \times \alpha_{L+J} (L+J)^{I-1}\}^{1/J}}$$

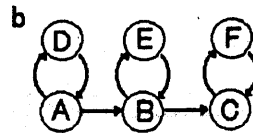
$$= \left\{ \frac{(L+J)^{I-1}}{L^{I-1}} \right\}^{1/J}$$

$$= \left(1 + \frac{J}{L} \right)^{(I-1)/J} \rightarrow 1$$

となる。 Q.E.D.



{5, 8, 12, 17, 23, 30...}



{8, 11, 14, 18, 22, 27...}

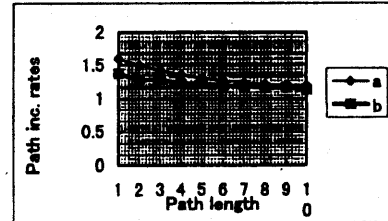
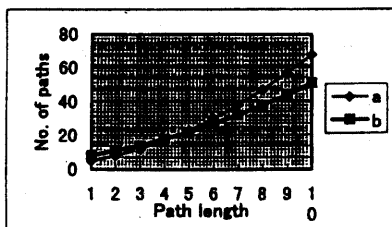


図3 共通のノードのないループを含むフローグラフとそのパス数 $\{P_L\}$ の例

P_L は連結行列をべき乗してその要素を合計すれば数値として得られる。 $L \geq N$ なるパスの P_L を式の形で求めるには、すべてのノードの組合せについてパス数を L の関数として求め合計すればよい。このような求め方は面倒で実際的ではないが、図3 aのフローグラフについて実例を示す。

$$P_{AC}(L) = {}_3H_{L-2} = L \times (L-1) / 2$$

(${}_2H_{L-2}$ の-2は、A→BとB→Cのリンク)

$$P_{AB}(L) = P_{BC}(L) = {}_2H_{L-1} = L$$

$$P_{AA}(L) = P_{BB}(L) = P_{CC}(L) = 1$$

したがって、これらを合計して

$$P_L = P_{AC}(L) + P_{AB}(L) + P_{BC}(L) + P_{AA}(L) + P_{BB}(L) + P_{CC}(L)$$

$$= L \times (L-1) / 2 + L + L + 1 + 1 + 1$$

$$= 0.5L^2 + 1.5L + 3$$

となる。

4.3 共通ノードをもつループがある場合

定理4：フローグラフに同じノードを共有しているループのグループがある場合、ノードを共有するループの数を $I > 1$ 、 N をノードの総数とすれば、 $L \geq N$ なるときに、 P_L は I^L のオーダーで増加する。

また、各グループに含まれるノードの数をそれぞれ $J_1, J_2 \dots J_I$ とし、その最小公倍数を J とすると、 $L \rightarrow$ 大のとき連続する J 個の P_L の増加率の幾何平均は、1より大なる数値に収束する。すなわち字式が成立する。

$$\left\{ (P_{L+1} \times P_{L+2} \times \dots \times P_{L+J}) / (P_L \times P_{L+1} \times \dots \times P_{L+J-1}) \right\}^{1/J} \rightarrow C > 1 \quad (7)$$

証明：図4を使う。aの例では $I = 2, J = 1,$

bでは $I = 2, J = 2$ である。定理3と同じく4段階に分けて証明する。

① 定理1によって、 $L \geq N$ なる L について、 $(J_1 + J_2 + \dots + J_I - I + 1) \times I^L \geq P_L$ である。

② 図4 bのようにループが二つ以上のノードをもつ場合、ループのもつノード数の最小公倍数を J とすれば、 $P_{AA}(J) \geq I, P_{AA}(2J) \geq I^2$ 、同様に $P_{AA}(N \times J) \geq I^N$ で、 L が J の倍数のときに $P_L \geq P_{AA}(L) \geq I^{L/J}$ となる。したがって①と併せて、 $L \geq N$ でかつサイクル J のとき P_L は L の二つの指数関数に挟まれる。

③ $L \geq N$ なる任意の L について長さ L のパスはサイクル J でノードAを通り、したがって $P_{AA}(L)$ を含むパスを含む。 $P_{AA}(L)$ は L の指数関数で増加するので、 P_L は L の指数関数で増加する。

④ (7)式については、 P_L はサイクル J で指数関数で増加するので、それらを $\beta_L \alpha_L^L, \beta_{L+1} \alpha_{L+1}^{L+1}, \dots, \beta_{L+J-1} \alpha_{L+J-1}^{L+J-1}$ と表す。 $\alpha_{L+J} = \alpha_L, \beta_{L+J} = \beta_L$ で、また $I > 1$ であるから $\alpha_L > 1$ である。したがって任意の L から $L+J-1$ について

$$(7)式 \rightarrow (\beta_{L+1} \alpha_{L+1}^{L+1} \times \beta_{L+2} \alpha_{L+2}^{L+2} \times \dots \times \beta_{L+J-1} \alpha_{L+J-1}^{L+J-1})^{1/J} / (\beta_L \alpha_L^L \times \beta_{L+1} \alpha_{L+1}^{L+1} \times \dots \times \beta_{L+J-1} \alpha_{L+J-1}^{L+J-1})^{1/J} \rightarrow \alpha_L > 1$$

となる。

Q.E.D.

図4のa, bについて P_{L+1}/P_L を計算すると、aでは $\alpha = 2$ でありループの数と一致する。bについてはサイクルが2で $\alpha_1 = 1.333$ と $\alpha_2 = 1.5$ という値に収束し、その積は2となり、サイクル2で

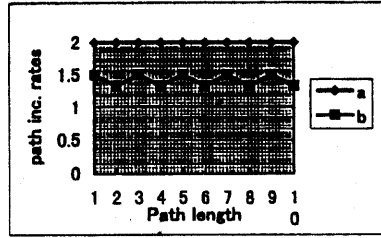
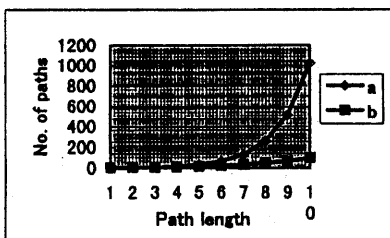
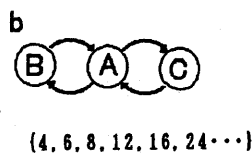
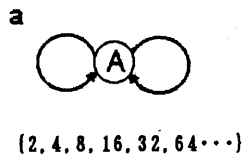


図4 共通のノードをもつループを含むフローグラフの例

2倍ずつ増加していることを示す。また図2の例ではノードCを共有している二つのループC→D→EとB→C→Fがそれぞれ3つのノードをもっているため最小公倍数Jは3となり、 P_L の増加率がサイクル3で振動している。

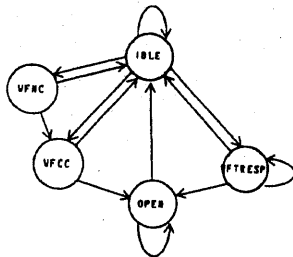
5 状態遷移図の評価例

状態遷移図のアイドルノードを二つに分割し、一方をスタートノード、他方をエンドノードとするとフローグラフになる。図5にトランスポートプロトコルのクラス0のフローグラフと、連結行列から得た P_L の値を示す。このフローグラフのループは共通のノードをもたず、もっとも多くのループをもつパスはSTART→WFTRESP→OPEN→ENDで、ループの数は2なので、定理3によりLが大のとき P_L はLの1次多項式で表される。その多項式は連結行列から得た数値と一致するように係数を定めれば求められる。同様に、クラス0, 2,

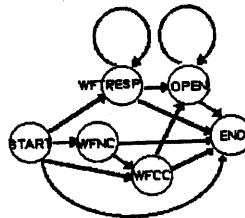
4の式はそれぞれ $P_L = 10 + 4L$, $6 + 8L + 2L^2$, $(54 + 62L + 23L^2 + L^4)/6$ となる^[2]。トランスポートプロトコルはクラスが上がるほど複雑になることが P_L の式の次数が増えることで示されている。このことはすでに発表済みではあるが、本論文によって一般化と完全な証明とを行った。

一般に近似式は独立変数の小さいところで近似がよく、大きくなると誤差が増えるが、定理3の場合には逆に $L \geq$ ノード数で正確な値となり、Lが小さいところでは正しさは保証されない。トランスポートの場合を見るとノード数は6であるが、 $L \geq 3$ で正しく連結行列と一致している。この3という値はさきに示したSTART→WFTRESP→OPEN→ENDというパスの長さに対応している。

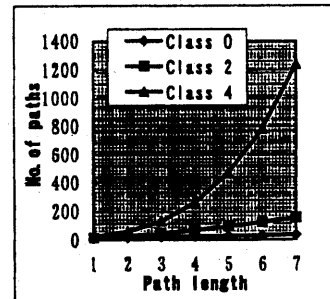
一般のフローグラフでは指数関数と多項式が混在するので式の形で求めることは簡単ではない。



State transition diagram of class 0



Flowgraph of class 0



P_L of Transport protocols

図5 トランスポートプロトコルのパス数 P_L

6 まとめと展望

パスベクトル (P_L) について分析し、次のことを示した。

パスベクトルにはフローグラフの性質が現れ、構造化プログラミングやサイクロマチック数などが反映される。またそれに含まれるループによって、L→大のときに、

- ①ループのない場合には、 $P_L \rightarrow 0$
- ②ループを含みそのループが共通のノードをもたない場合には、パス数はLの多項式で表現され、その増加率の幾何平均→1
- ③二つ以上のループが同じノードを共有する場合には、パス数はLのべき乗で増加し、その増加率の幾何平均→C>1

となるという3つのタイプに分類することができる。

今後は、パスカバレッジ試験と結び付けてエラー数の推定に展開してゆく予定である^[3]。

参考文献

- [1] Li, H. L. & Chung, W. K.: An Empirical Study of Software Metrics, IEEE Tr. on Soft. Eng. Vol. SE-13, No. 6 (1987).
- [2] 若杉忠男: ISOで開発したトランスポートプロトコルの適合性試験スイートの質の評価, 情報処理学会論文誌, Vol. 37, No. 3, pp451-459 (1996-3).
- [3] 若杉忠男: 残存エラー数の推定が可能なソフトウェア試験法について, 情報処理学会ソフトウェア工学研究会報告, 96-SE-112, PP9-16 (1996)