

論文

確率推移検索システム*

磯本征雄** 山県敬一***

Abstract

Recently, a graph theory and a probability theory have been applied to various fields of researches in a computer science. It is convenient to discuss transition phenomena with the use of a directed graph and a concept of probability. And also, we are interested in developing a question-answering system by the aid of the graph theory and the probability theory. One of the purposes of this paper is to make a theoretical foundation of the system on these theories.

This paper shows a formulation with which Markov chain graphs having transition probabilities are efficiently analyzed by a computer. Using the formulation, we have realized the probable transition searching system in a FORTRAN program. The system analyzes just only the structure of directed graphs, apart from the concrete meanings of the graphs. Then, we consider that the system has many possibilities of applications.

1. ま え が き

相互の関係が重要な意味を持つ事象集合は多い。そして、それら事象間の関係の構造に関する探索の問題は、情報検索システム¹⁾や問題解決システム²⁾などに、しばしば現れる。この場合、探索の基準となる関係の構造には、その具体的対象に応じていろいろのものがああり、プログラムのパッケージとしても多くのものが作られている。

ところで、関係の構造の探索手法について、性質の面から、決定論的なものと、確率論的なものがある。

前者については、多くの場合、組み合わせ論的な探索手順がとられ、その手順をなるべく簡略化するように工夫がなされる。他方、後者の場合は、ある程度、確率論における帰結を利用することにより、組み合わせ論的な探索手順を用いなくても、ある種の特徴抽出、あるいは、それに基づく事象探索が可能である。

この論文では、後者の確率論的モデルについて、一般的な手法を明らかにし、さらに具体的にプログラム

でどのように実行するかをのべる。これには、グラフ理論と、確率論的モデルの両面の手法を利用する。

初めに与えられる事象集合の関係は、具体的対象に応じて異なったものとなるが、それは関係を表わすグラフが変わるだけで、検索機能は同じものが使える場合が多くあると思われる。そのため、この論文では、検索対象となるグラフとしてはあまり強い制約を与えず、なるべく一般的な形で取り扱えるように考慮してある。したがって、ここでの検索機能は比較的一般的な性質を持っており、適用できる応用範囲はかなり広いと考えられる。

2. 記号の定義とシステムの定式化

有限事象集合 $V = \{v_1, v_2, \dots, v_m\}$ と、その上での2項関係 R を考える。すべての事象 v_i を頂点とし、関係 $v_i R v_j$ を満たす v_j に対して、 v_i から v_j へ向う有向辺で頂点を結べば、有向グラフ G を得る。ただし、ここに $v_i, v_j \in V$ である。

ある事象 v_i について、 $v_i R v_j$ を満たす v_j が複数個あるときは、各々の有向辺に、関係の強さを表わす確率的重み p_{ij} をつける。すなわち、 p_{ij} は次式を満たす。

$$\sum_j p_{ij} = 1 \quad (v_i, v_j \in V, v_i R v_j) \quad (2.1)$$

* Probable Transition Searching System by Yukuo ISOMOTO (Osaka University Computation Center) and Keiichi YAMAGATA (Department of Precision Engineering, Osaka University)

** 大阪大学大型計算機センター

*** 大阪大学工学部精密工学科

$v_i R v_j$ を満たす v_j が唯一つであるときは、 $p_{ij}=1$ とする。辺につけられた重みが、2項関係 R だけに依存するとすれば、この有向グラフは、時間的に一様なマルコフ連鎖グラフとなる(付録 1 (701 頁参照))。

グラフ理論的考察を行うため、 V から V への対応 Γ を導入する³⁾。任意の頂点 $v_i \in V$ について、 $v_i R v_j$ を満たす v_j の集合を $\Gamma(v_i)$ と表わし、また、任意の頂点 v_j について $v_i R v_j$ を満たす v_i の集合を $\Gamma^{-1}(v_j)$ と書く。さらに、任意の部分集合 $U \subset V$ に対して $\Gamma(U)$ と $\Gamma^{-1}(U)$ を次式で定義する。

$$\Gamma(U) = \bigcup_{v_i \in U} \Gamma(v_i), \quad \Gamma^{-1}(U) = \bigcup_{v_i \in U} \Gamma^{-1}(v_i) \quad (2.2)$$

いま、有向辺にそって、結びつけられた事象をたどることを(確率)推移と呼ぶことにし、 n 段階の推移を表わすために Γ^n を次式で定義する。

$$\Gamma^0(v_i) = \{v_i\}, \quad \Gamma^1(v_i) = \Gamma(v_i) \quad (2.3)$$

$$\Gamma^n(v_i) = \Gamma(\Gamma^{n-1}(v_i)) \quad (n \geq 2)$$

以上の表現を用いて、

$$\hat{\Gamma}(v_i) = \bigcup_{n=0}^{\infty} \Gamma^n(v_i) \quad (2.4)$$

を定義すれば、 $\hat{\Gamma}(v_i)$ は事象 v_i からの到達可能集合を表わす。

任意の $v_i, v_j \in V$ に対して、 $v_j \in \hat{\Gamma}(v_i)$ が成り立つとき、かつその時に限り $v_i \leq v_j$ であるものとして、関係 \leq を定義すれば、この関係は $V(G)$ における擬順序関係である³⁾。

有向辺につけられた重み p_{ij} について、これを第 i 行第 j 列の成分とする行列 P を考えれば、これは推移確率行列⁴⁾となる。初期状態における事象 v_s の出現確率を $\alpha_s(0)$ として、これを第 k 成分とする行ベクトル $\alpha(0)$ で、事象の初期分布を表わせば、確率論の帰結を用いて、第 n 段階推移後の分布は、

$$\alpha(n) = \alpha(0)P^n \quad (2.5)$$

で与えられる。とくに、初期分布として、特定の事象 v_s からのみ推移が始まる場合には、 $\alpha(0)$ における第 s 成分のみが 1 で、他の成分は 0 である。

上に述べたようなグラフの構造は、具体的な検索対象に依存して決まる。例えば、事象集合 V の全体が強連結であるような、既約な有向グラフの場合、確率モデルとしては興味のあるものであるが、因果関係に基づく検索を考える場合等は、むしろ擬順序関係に注目する必要がある。当面の問題として、この論文でもグラフの全体が既約である場合は、考慮の対象外とし、擬順序関係に注目したモデルを考える。

ただし、確率論的に重みづけられた関係を取り扱う

以上、部分グラフとして強連結なものを含むことは十分あり得る。有向グラフ G に含まれる強連結なセクション部分グラフ中の事象の集合を、ここではクラスター $C_{(i)}$ と呼ぶ。ただし、添字 i は、擬順序関係 \leq における同値類の番号づけである。

有限なマルコフ連鎖では、グラフ G 全体が強連結である場合を除いて、過渡的事象から出発して推移を重ねる内に、やがてある閉じた出口のない集合に吸収される。このような吸収事象としては、もともと有向グラフ G において、単一事象が閉じているものと、複数個の事象を持つクラスターとして閉じているものがある。前者については、

$$\Gamma(v_i) = \{v_i\}, \quad p_{ii} = 1 \quad (2.6)$$

である。

後者については、

$$\Gamma(C_{(i)}) = C_{(i)} \quad (2.7)$$

である。この $C_{(i)}$ を一つの事象 v_c でおきかえ、このクラスターに入って来るすべての有向辺を v_c につなぎ、さらに $p_{cc} = 1$ となるように、 v_c から v_c への一つの有向辺をつける。このようにして作られた有向グラフ G' は、吸収マルコフ連鎖を表わし、その事象集合 V' は、過渡的事象集合 $V_{T'}$ と吸収事象集合 $V_{E'}$ に分けることができる。

検索システムにおいては、 $V_{E'}$ の要素は、関係 R の推移における究極の到達点を表わす。その意味で、 $V_{E'}$ の要素を特に端末事象と呼ぶ。

事象につけられた番号をつけかえて、 $V_{T'} = \{v_1, v_2, \dots, v_r\}$ 、 $V_{E'} = \{v_{r+1}, v_{r+2}, \dots, v_m\}$ とすれば、推移確率行列 P' 、ならびに事象の分布を示す確率ベクトル $\alpha'(n)$ は、以下のように書ける。

$$P' = \begin{bmatrix} P_{T'} & P_{E'} \\ 0 & I \end{bmatrix}, \quad \alpha'(n) = [\alpha_{T'}(n), \alpha_{E'}(n)] \quad (2.8)$$

ここに、 I は $(m'-r) \times (m'-r)$ の単位行列である。なお、 $V_{T'}$ は空ではないと仮定しておく。

いま初期事象 $v_s \in V_{T'}$ から出発したとき、端末事象 $v_e \in V_{E'}$ に吸収される確率 b_{se} を行列 B の第 s 行 e 列の要素とすれば、

$$B = [b_{se}] = (I - P_{T'})^{-1} P_{E'} \quad (2.9)$$

が成り立つ⁴⁾。これを用いれば、有向グラフの個々の事象をたどることなしに、初期事象と端末事象との関係の強さを調べることが可能である。式(2.7)のクラスター $C_{(i)}$ の推移確率行列を $P_{C_{(i)}}$ とする。 $C_{(i)}$ の定常分布における事象 $v_i \in C_{(i)}$ の相対的な分布の確率は、次式を満たすベクトル $\alpha_{C_{(i)}}$ の第 i 番目の成分

で与えられる。

$$q_{C(v)} P_{C(v)} = q_{C(v)} \quad (2.10)$$

すなわち、式 (2.1) より $q_{C(v)}$ は行列 $P_{C(v)}$ の最大固有値 1 の固有ベクトルである⁵⁾。 $q_{C(v)}$ は、 $C(v)$ 内の事象の相対的重要度を評価する上で、指標として使うことができる。

つぎに、初期事象 v_1 と末端事象 $v_n \in \hat{\Gamma}(v_1)$ が定められたとき、両者の間を結ぶ順路は、一般に複数個存在する。それらの順路のそれぞれについて、順路に沿う推移確率の積を考え、それが最大であるような順路を最大推移確率順路 $F_S(v_1)$ と呼び、その順路における確率の積を、 v_1 と v_n を結ぶ最大推移確率 $\pi_1(v_1)$ と呼ぶものとする。これらは、 v_1 と v_n を結ぶ順路の選択に使われる。

3. 検索手順の定式化

確率推移検索システムとは、与えられたマルコフ連鎖グラフ G において、初期事象 v_1 が指定されたとき、

- i) 初期事象 v_1 からの到達可能集合 $\hat{\Gamma}(v_1)$ 、
- ii) $\hat{\Gamma}(v_1)$ 内のクラスター $C(v)$ 、
- iii) 末端事象 $v_n \in V_2'$ への吸収される確率 b_{v_n} 、
- iv) 最大推移確率順路 $F_S(v_1)$ とその最大推移確率 $\pi_1(v_1)$

などを、計算機により求めるシステムである。

以下に、この検索の具体的な手順を定式化する。

3.1 到達可能集合 $\hat{\Gamma}(v_1) \subset V$ の抽出

$\hat{\Gamma}(v_1)$ を直接に求める方法は、定義式 (2.3)、(2.4) を忠実に実行することであるが、この方法を一般的な有向グラフに適用すると、同一事象について、何回も重複して隣接事象を調べることになる。とくにクラスターを含む場合は、同一事象が巡回して何回も現れる。そこで、探索された事象の重複を即座に検出して、探索効率の低下を防ぐアルゴリズムが必要となる。

この目的のために、以下の漸化式で $\lambda_n(v_1)$ 、 $A_n(v_1)$ を定義する。

$$A_0(v_1) = \{v_1\},$$

$$\lambda_1(v_1) = \{v_1\} \cap \Gamma(v_1), \quad A_1(v_1) = \{v_1\} \cup \Gamma(v_1). \quad (3.1)$$

$$\lambda_n(v_1) = A_{n-1}(v_1) \cap \Gamma(\lambda_{n-1}(v_1)),$$

$$A_n(v_1) = A_{n-1}(v_1) \cup \lambda_n(v_1), \quad n \geq 2 \quad (3.2)$$

ここで、 $A_{n-1}(v_1)$ と $\{v_1\}$ はそれぞれ A_{n-1} と $\{v_1\}$ の V に対する補集合である。これらの表現を用いると、次式が成立する (付録 2 (702 頁参照))。

$$A_n(v_1) = \bigcup_{k=0}^n \Gamma^k(v_1), \quad n \geq 1 \quad (3.3)$$

式 (3.2) によって、新しい集合 $A_n(v_1)$ を作る時、あらたに加えられる部分集合 $\lambda_n(v_1)$ と、もとの集合 $A_{n-1}(v_1)$ との間には、両者に共通に含まれる事象はなく、これによって、検索の重複が避けられる。

ところで、式 (3.2) の定義から明らかに、 $n=1$ から集合演算を開始して、 n が大きくなるに従って、 $A_n(v_1)$ に属する事象の数は単調に増大する。他方、集合 V は有限であることから、ある有限の数 N が存在して、 $\lambda_N(v_1) = \phi$ となる。また、 $\Gamma(\phi) = \phi$ であるから、 $n \geq N$ なる任意の n について、 $\lambda_n(v_1) = \phi$ であり、 $A_n(v_1) = A_N(v_1)$ が成り立つ。したがって、式 (3.2) による有限回の反復操作で $\hat{\Gamma}(v_1)$ が求められる。すなわち、

$$\hat{\Gamma}(v_1) = \lim_{n \rightarrow \infty} A_n(v_1) = A_N(v_1), \quad \lambda_N(v_1) = \phi. \quad (3.4)$$

初期事象 v_1 が指定されて、 $\hat{\Gamma}(v_1)$ が抽出されると、後は $v_i \in \hat{\Gamma}(v_1)$ を満たす事象だけを持つセクション部分グラフだけを考えればよい。この段階でこのグラフの隣接行列が作られる。2. で述べたグラフの性質から、隣接行列の要素は 0 または 1 の値しかとらない。したがって、ビット単位の行列で表現できる。

この隣接行列の第 i 行のビット列は $\Gamma(v_i)$ を表わし、第 i 列のビット列は $\Gamma^{-1}(v_i)$ を表わす。これらを用いると、事象集合における集合演算はビット単位の論理演算で処理できることになり、効率がよい。

この手法は、ここで述べた $\hat{\Gamma}(v_1)$ の抽出手順にも、また次節のクラスターの抽出手順にも応用される。

3.2 クラスターの抽出

混乱は起らないと思われるので、前節で抽出されたセクション部分グラフだけを考慮して、 $\hat{\Gamma}(v_1)$ を改めて事象集合 V とする。また、隣接行列 D も前節で述べたビット単位の行列として与えられているものとする。

ここで述べるクラスター抽出の手法は、事象の部分集合をビット列で表現し、集合演算を論理演算で効率良く実行することに主眼をおくものである。したがって、有向グラフを直接有向辺に沿ってたどるというアルゴリズムは、表面上に現れない。

実行能率の向上を目的として、クラスター抽出手順を二段階に分ける。第一段階で、クラスターに含まれないことが容易にわかる事象について、それらをまず取り除き、しかる後に強連結成分に分解する。

明らかに、前節の手順で抽出された有向グラフについては、初期事象 v_0 と、1つの事象だけで閉じている事象 v_i (この集合を V_E と書く) は、クラスターに含まれない。次式で V の部分集合 S_0 を定義する。

$$S_0 = V \cap V_E^c \cap \{v_i\}^c \quad (3.5)$$

さらに、次の漸化式で V の部分集合 S_m を定義する。

$$S_m = S_{m-1} \cap \{v_i\}^c, \quad v_i \in S_{m-1}, \quad m \geq 1 \quad (3.6)$$

ただし、 v_i は次の関係を満たす事象である。

$$\Gamma(v_i) \cap \{v_i\}^c \cap S_{m-1} = \phi,$$

または

$$\Gamma^{-1}(v_i) \cap \{v_i\}^c \cap S_{m-1} = \phi. \quad (3.7)$$

式 (3.6) の漸化式に従って、順に S_m を求めれば、クラスターに含まれ得ない事象が次々に取り除かれる。ある $m=M$ において、 S_M の中に式 (3.7) を満たす事象が見出せなくなったとき、 S_M はクラスター及びクラスター間を結ぶ事象よりなる V の部分集合である。もともとグラフにクラスターが含まれていなければ、この段階で $S_M = \phi$ となる。

$S_M \neq \phi$ であれば、強連結成分の抽出の第2段階に移る。いま、隣接行列 D はすでにわかれているので、任意の $v_i \in S_M$ について $\Gamma(v_i)$ 、 $\Gamma^{-1}(v_i)$ は共にわかっている。 $\Gamma(v_i)$ を用いて式 (3.1)、(3.2) の漸化式を用いれば、 $\hat{\Gamma}(v_i) \subseteq S_M$ を得る。

$\hat{\Gamma}(v_i)$ は v_i より到達可能な事象集合であるが、強連結成分を検出するには、 v_i へ到達可能な事象集合が必要である。これを $\tilde{\Gamma}(v_i) \subseteq S_M$ とする。有向辺の向きをすべて逆にした有向グラフを想定すれば、式 (3.1)、(3.2) の漸化式に $\Gamma^{-1}(v_i)$ を適用することにより、全く同様の手順で $\tilde{\Gamma}(v_i)$ が得られる。詳しい定式化は省略する。

$v_i \in S_M$ について、次の集合を定義する。

$$C = \hat{\Gamma}(v_i) \cap \tilde{\Gamma}(v_i). \quad (3.8)$$

この時、任意の事象 $v_j \in C$ について、 v_i から v_j への順路、および v_j から v_i への順路が共に存在する。故に C はクラスターである。

クラスターが見つければ、これに属する事象を S_M から有向辺と共に取り除き、同様の操作を反復することによって、総てのクラスターを抽出する。

3.3 端末事象へ吸収される確率 b_{ii}

前節で述べた手順に従って、クラスターを抽出し、閉じていて出口のないクラスターを1つの端末事象として、 V_E につけ加えれば、式 (2.9) の計算は、行列に関する計算のみに帰着される。この場合も、いま考えているセクション部分グラフについてのみ、計算す

れば良い。

3.4 端末事象への最大推移確率とその順路

初期事象 v_0 から端末事象 v_e への順路 (一般には複数個ある) について、順路に沿う確率の積を考える。任意の事象 $v_i \in V$ について、いろいろな順路に沿う、 v_e への推移確率の中で最大のものを $f_i(v_e)$ と表わすと、 $f_i(v_e)$ は次式を満たす⁶⁾。

$$f_i(v_e) = \max_{v_j \in \Gamma(v_i)} [p_{ij} f_j(v_e)], \quad v_i \in V. \quad (3.9)$$

ただし、 $f_e(v_e) = 1$ とする。

逐次近似法で $f_i(v_e)$ を求めるために、第1近似として、次の値を付与する。

$$\left. \begin{aligned} f_i^{(1)}(v_e) &= p_{ie}, \quad v_i \in \Gamma^{-1}(v_e). \\ f_i^{(1)}(v_e) &= 0, \quad v_i \notin \Gamma^{-1}(v_e). \end{aligned} \right\} \quad (3.10)$$

この値は、ただ1段階だけの推移で v_e に到達できる順路についての評価である。1段階の推移で到達する順路が存在しないときは、重み0を与える。

つぎに、高々 k 段階の推移で到達できる順路についての評価は、以下の漸化式で与えられる。

$$f_i^{(k)}(v_e) = \max_{v_j \in \Gamma(v_i)} [p_{ij} f_j^{(k-1)}(v_e)]. \quad (3.11)$$

これを反復適用して、総ての $v_i \in V$ について

$f_i^{(k)}(v_e) = f_i^{(k-1)}(v_e)$ となったならば、 $\pi_i(v_e)$ は次式で得られる⁶⁾。

$$\pi_i(v_e) = f_i^{(k)}(v_e). \quad (3.12)$$

最大推移確率を与える順路は、式 (3.11) において、最大値を与える事象番号 j を記憶しておけば求められる。

4. 計算機によるデータの取り扱い

確率推移検索システムは、FORTRAN-700⁷⁾ (JIS-FORTRAN 水準 7000 相当) によりプログラムされている。以下、FORTRAN-700 によるプログラミングの説明を行うが、このために一般性を失うことはないであろう。前章までの定式における P 、 v_i 、 $\Gamma(v_i)$ は、配列により定義される。集合 $\Gamma(v_i)$ 、 $\lambda_n(v_i)$ 及び $\Lambda_n(v_i)$ 等はビット形式で定義される変数であり、集合要素は変数の各ビットに対応させる。プログラムの平易さ、処理の高速化及び記憶領域の縮小化のために集合算をマスキング式によるビット単位での OR、AND 及び NOT の演算子で処理する。以下、プログラミング上、特に工夫の必要な事項について説明する。

4.1 ファイルにおけるデータ構造

一次情報 v_i 、 $\Gamma(v_i)$ 、 $p_{ij}(v_j \in \Gamma(v_i))$ は、予めファイルに Fig. 4.1 (次頁参照) に示す書式で記録されてい

る。事象 v_i は、その内容を示す文章 (MMM... MMM) と共にコード番号 (IIII) で記録されている。同時に、 $\Gamma(v_i)$ の中の要素の数 (J) が与えられている。次の行には、 $\Gamma(v_i)$ の要素がコード番号 (KKK 1, KKK 2 等) により与えられ、それらへの推移確率 (XXX...XX, YYY...YY 等) が各々のコード番号の直後に記録される。

コード番号は、事象 v_i をファイルより読み出す時の事象の認識に使われる。添字 i は、説明の便宜上付けられたものであり、ファイル上では特別の意味はない。 J は、 $\Gamma(v_i)$ 、 p_{ij} の読み出しの際に使われるパラメータである。

4.2 データの主記憶への転送

事象 $v_i \in V$ 及び推移 $\Gamma(v_i)$ より成るグラフ全体は、ファイルの中に記録されている。初期事象 v_1 の入力直後、ファイル中の v_1 が確認され、漸化式 (3.1), (3.2) により、ファイル中の事象を探索しながら $\hat{\Gamma}(v_1)$ が主記憶上に取り出される。 v_i , $\Gamma(v_i)$, $p_{ij}(v_j \in \Gamma(v_i))$ は、予め主記憶上に定義されている配列に記憶される。事象 v_i の添字 i は、主記憶に記憶される順序にしたがって、改めて付けられる番号である。番号付けは、配列添字の小さいものから大きなものへほぼ推移の順番に従う。以下、簡単のため Fig. 4.2 の具体例をもって説明する。Fig. 4.2 の各事象に付けられた添字の値は、主記憶に記憶される時点で与えられるものであるが、誤解の恐れはないと思われるので、この点での注釈は以後省略する。

v_1 がコード番号により入力された直後に、 $v_1, \Gamma(v_1) = \{v_2, v_3\}$, p_{12}, p_{13} がファイルより読み出され、同時に式 (3.1) より次の値が定められる。

$$\begin{aligned} \lambda_1(v_1) &: 01100000\dots, \\ A_1(v_1) &: 11100000\dots, \\ \Gamma(v_1) &: 01100000\dots, \end{aligned}$$

配列要素 v_1, v_2, v_3 には、コード番号が記憶されている (Fig. 4.2 参照)。集合 $\lambda_1(v_1)$ の第 2, 3 ビットが 1 であることより、 $\{v_1\}$ に新たに v_2 と v_3 が追加されることがわかる。 $A_1(v_1)$ より、 v_1, v_2, v_3 が、この時点で到達可能集合の要素であることがわかる。集合を表す配列要素 $\Gamma(v_1)$ より、 v_1 から v_2 と v_3 への推移が可能であることがわかる。

つぎに、 $\lambda_1(v_1)$ の示すところから従って、コード番号を用いて v_2 と v_3 をファイルより探索して $\Gamma(v_2) = \{v_5, v_4, v_6\}$, $\Gamma(v_3) = \{v_5, v_6\}$ 及び $p_{23}, p_{24}, p_{25}, p_{35}, p_{36}$ を読み出す。式 (3.2) より、次の値が与え

```
(IIII)J/MMMMMMMMMMMMMMMMMMMM...
/KKK1/XXXXXXXXXX//KKK2/YYYYYYYYYY//...
```

- IIII ; I4, a code number for a phenomenon.
- J ; I1, the number of transitions from IIII.
- MM...MM ; 8A8, a message for the phenomenon IIII.
- KKK1 ; I4, a code number transferred from IIII.
- XX...XX ; F10.5, a transition probability from IIII to KKK1.
- KKK2 ; I4, a code number transferred from IIII.
- YY...YY ; F10.5, a transition probability from IIII to KKK2.

Fig. 4.1 Data structure in the file. The data are stored in the above mentioned format.

i	v_i	J	$\Gamma(v_i)$
1	0001	2	0110000000.....
2	0020	3	0011100000.....
3	0405	2	0000110000.....
4	0620	1	0000001000.....
5	1432	3	0000000111.....
6	0580	1	0000001000.....
7	3400	⋮	⋮
8	2340	⋮	⋮
9	3810	⋮	⋮
10	1001	⋮	⋮
⋮	⋮	⋮	⋮

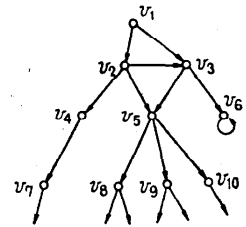


Fig. 4.2 A directed graph and its data structure in the main memory.

られる。

$$\begin{aligned} \lambda_2(v_1) &: 0001110000\dots, \\ A_2(v_1) &: 1111110000\dots, \\ \Gamma(v_2) &: 0011100000\dots, \\ \Gamma(v_3) &: 0000110000\dots, \end{aligned}$$

$\lambda_2(v_1)$ の第 4, 5, 6 ビットが 1 であることから、 v_4, v_5, v_6 が新たに追加される事象であることがわかる。 $A_2(v_1)$ の第 1~6 ビットが 1 であることから、この時点で確認された $\hat{\Gamma}(v_1)$ の要素は v_1, v_2, \dots, v_6 であることがわかる。 $\Gamma(v_2)$ は、 v_2 より v_3, v_4 と v_5 への推移が可能であることを示し、 $\Gamma(v_3)$ は、 v_3 より v_5 と v_6 への推移が可能であることを示す。同様の操作は、 $\lambda_N(v_1) = \emptyset$ となるまで繰り返される。

$\hat{\Gamma}(v_1)$ が得られた後に、 V_T と V_E の分離を行う。 V_E の要素は、配列添字の大きい側に移してひとまとめにする。また、 $\Gamma(C) = C$ なるクラスターは、その振舞いが端末事象に類似しているため V_E の部分集合とみなす。上記説明中では省略したが、各事象に関するメッセージも事象のコード番号と同様に主記憶上に転

文で示された機能は、一般的又は抽象的であるかも知れない。しかし、確率推移検索システムが取り扱う具体的対象とその意味・内容は、総て一次情報である v_i , $\Gamma(v_i)$ 及び $p_{ij}(v_j \in \Gamma(v_i))$ により与えられ、システムの機能とは分離して考える。

このことをわかり易くするために、プログラム・デバッグの際の、バッグの類推と追跡の過程を一例に考える。初期事象 v_i をプログラム誤動作の表面的な現象であるとする。 $v_i \in V_T$ をバッグ及びバッグの示す現象、さらにはバッグ発見のための手法・手順についての説明であるとする。 $v_e \in V_E$ を最終的デバッグの手続きであるとする。この時、 $\Gamma(v_i)$ によって作られる有向グラフとは、プログラム誤動作の表面的な現象から始めて、バッグの示す諸々の現象についての因果関係を経て、デバッグ手続きに至る説明のための様々の可能な順路を示すグラフである。有向グラフ上にある多くの分岐路は、これら順路の様々の可能性を示すものであり、推移確率とはその可能性の大きさを与えるものである。

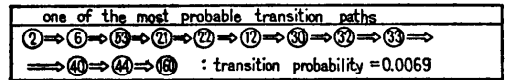
確率推移検索システムは、データの構造のみを解析する。解析された情報の具体的な意味・内容は、このシステムを離れた別の次元で解釈される。この点は、システムの応用範囲をより広く、一般的なものにするために行った工夫でもある。

Fig. 5.1 (前頁参照)及び Table 5.1 は、確率推移検索システムのテストのために用いた有向グラフとその処理結果の一部分である。テストは、様々の事象を初期事象として選んで繰り返した。この結果、Fig. 5.1 の有向グラフに対するテスト例では、最大処理時間は約6秒であった。この点は、本システムが実用的なものになり得るとの確信を与える。

本システムは、確率推移の解析システムとしてのみ開発されたものではない。確率の付随した有向グラフにより表わせられる諸々の過程に対して、多少の修正を行うのみで、容易に広い応用範囲が得られるであろう。また、本システムに、有向グラフを積極的に書き替える機能を追加することにより、質問応答システムへの拡張・応用が容易に行える。現在、我々の所では、本システムを用いて、誤動作を起したプログラムの誤り追跡の過程における可能性の類推への援助システムを開発し、試用段階にある。

この論文に関しては、 $\Gamma(C) \ni C$ なるクラスターの取り扱い方法は余り重要でなかった。しかし、有向グラフの取り扱い方法を変えた場合には、 $\Gamma(C) \ni C$ なるク

Table 5.1 Results of searching the graph in fig. 5.1.

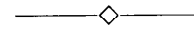


	final actualizing probability of terminal phenomena
①	0.049
②	0.104
③	0.055
④	0.068
⑤	0.168
⑥	0.056
⑦	0.167
⑧	0.333
Total	1.000

ラスターの取り扱い方法が問題になる。この点に関する議論は、別の機会に行う予定である。なお、このシステムには、大阪大学大型計算機センター・ライブラリ・プログラム⁸⁾、INVERS と UPOWRS が使われている。

参考文献

- 1) Gerard Salton: Automatic Information Organization and Retrieval (Computer Science Series), McGraw-Hill Book Company (1968).
- 2) G.W. Ernst and A. Newell: ACM Monograph Series GPS: A Case Study in Generality and Problem Solving, Academic press (New York and London) (1969).
- 3) 尾崎弘, 白川功: グラフとネットワークの理論, コロナ社, (1973).
Clande Berge: The Theory of Graphs and its applications, Mathuen & Co. Ltd., London (1966).
- 4) S. カーン著, 佐藤健一・佐藤由身子訳: 数理解析とその周辺3, 確率過程講義, 産業図書 (1974).
北川敏男: 情報科学講座, A. 5. 1, マルコフ過程, 共立出版, (1967).
- 5) R. S. Varga: Matrix Iteration Analysis, (Prentice-Hall, Inc. Maruzen Co. Ltd.), (1964).
- 6) R. Bellman, K. L. Cooke and J. Lockett: Algorithm, Graph and Computers, Academic Press, (New York and London) (1970).
- 7) FORTRAN-700 文法説明書, NEC 日本電気(株).
- 8) 大阪大学大型計算機センター・ニュース, No. 8, 大阪大学大型計算機センター (1972).



付録1 2項関係とマルコフ連鎖グラフ

2項関係を表す有向グラフ一般に対して、これに確率論的重みをつける場合、ある事象 v_i について、 $v_i R^* v_j$ を満たす v_j が1つも存在しない場合の扱いが

問題となる。 $p_{ij}=0$ とすると、式 (2.1) が満たされない。

この論文では、2. に述べる検索機能の内容を考慮して、出て行く有向辺が1つも存在しない事象については、自分自身に向う有向辺を付加して $p_{ii}=1$ として扱う。与えられた2項関係に対して、ここでの検索システムを適用する場合、 $v_i R v_j$ を満たす v_j が空である場合と、それが唯1つ v_i 自身である場合との区別が必要になることがある。ただ、この論文で述べる検索手順は、その区別の必要性の有無には無関係であるので、この問題には立ち入らないこととする。

付録2 式 (3.3) の証明

数学的帰納法により証明する。式 (3.1) より明らかに $n=0, 1$ について、次式が成り立つ。

$$A_0(v_i) = \{v_i\}, \quad A_1(v_i) = \bigcup_{k=0}^1 \Gamma^k(v_i). \quad (A1.1)$$

今、 $n = \nu - 1$ 、 ν について次式が成り立つと仮定する。

$$A_n(v_i) = \bigcup_{k=0}^n \Gamma^k(v_i). \quad (A1.2)$$

つきに $n = \nu + 1$ のとき Γ の定義から、次式を得る。

$$\bigcup_{k=0}^{\nu+1} \Gamma^k(v_i) = \{v_i\} \cup \Gamma\left(\bigcup_{k=0}^{\nu} \Gamma^k(v_i)\right).$$

式 (A1.2) の仮定及び式 (3.2) より上式は次式となる。

$$\bigcup_{k=0}^{\nu+1} \Gamma^k(v_i) = \{v_i\} \cup \Gamma(A_{\nu-1}(v_i)) \cup \Gamma(\lambda_\nu(v_i)).$$

再び仮定 (A1.2) と Γ の定義より、次の関係を得る。

$$\bigcup_{k=0}^{\nu+1} \Gamma^k(v_i) = A_\nu(v_i) \cup \Gamma(\lambda_\nu(v_i)). \quad (A1.3)$$

また一方、式 (3.2) より次の関係を得る。

$$A_{\nu+1}(v_i) = A_\nu(v_i) \cup \Gamma(\lambda_\nu(v_i)) \quad (A1.4)$$

式 (A1.3) と (A1.4) より、次式を得る。

$$A_{\nu+1}(v_i) = \bigcup_{k=0}^{\nu+1} \Gamma^k(v_i) \quad (A1.5)$$

よって、任意の自然数 n について

$$A_n(v_i) = \bigcup_{k=0}^n \Gamma^k(v_i)$$

が成り立つ。

(昭和49年10月26日受付)

(昭和50年11月4日再受付)