数理解析研究所講究録
第458巻 1982年 276-285

シャフルされた記号列の復元問題

京産大 理　　岩間一雄

1. まえがき

分散処理で必要となる多対多通信を簡便にかつ低コストで実現する方法としては，すべての局が単一の受動通信媒体（バスやループの形をとる）へ共通にアクセスする方法が最も有望であると考えられている[1]。本稿では分散処理系のこのような通信に焦点を当てたモデルを提案し，情報交換が混乱なく進行するための条件を数学的観点から議論する。

図1に示される様に，モデルは $n+m$ 個の並行的に動作するオートマトンと単一の通信チャネルより成る．$T_1,\cdots,T_n$ は送信局と呼ばれ，$\{0,1\}$ 上の列 $u_1,\cdots,u_n$ を送信する。各 u_i はあらかじめ与えられた言語 L_i に属すことが保障されている。他の m 個のオートマトン $R_1,\cdots,R_m$ は受信局と呼ばれ，それぞれ列 $v_1,\cdots,v_m$ を受信する。$T_1,\cdots,T_n$ が完全に非同期的な並行動作を行うなら R_i によって受信される列 v_i は $u_1\odot u_2\odot\cdots\odot u_n$ の中からランダムに選ばれる（$\odot$ はシャフル演算を表わす[2]-[4]）。

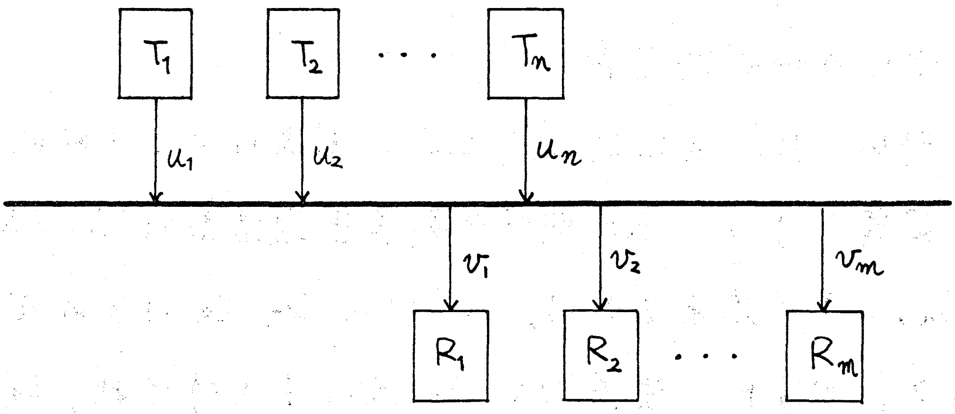

図1. モデル

"情報交換が混乱なく進行する"ために必要な最も基本的条件は，各R_iが受信列v_iから送信列$u_1, \cdots, u_n$を一意に復元できることであろう。本稿の目的は，このような一意復元可能性が常に保障されるためには送信局が生成する言語$L_1, \cdots, L_n$にどのような条件が要求されるかを議論することである。分散並行システムに関しては多くの形式的研究が知られているが[7]，その通信のメカニズムに焦点を当てたものはほとんど存在しなかったと思われる。

〔例題1〕 $L_1 = (10)^*$, $L_2 = (11+00)^*$ とする。列$v_i = 11100010$が受信されたとき，$v_i \in u_1 \odot u_2$ $(u_1 \in L_1, u_2 \in L_2)$であるような$u_1$と$u_2$は，$u_1 = 1010$と$u_2 = 1100$以外にはありえず，$v_i$は一意に復元可能である。しかし，$v_i' = 11010110$の場合は，$u_1 = 10$と$u_2 = 110011$，$u_1 = 101010$と$u_2 = 11$の（少くとも）2通りの復元が可能であり不都合が生じる。

2. モデルの形式的定義

送信列$u_1, \cdots, u_n$と受信列$v_1, \cdots, v_m$の関係はより一般的なものを許す定義とする．非同期時分割多重通信系は$C=(\Sigma_T, \Sigma_R; L_1, \cdots, L_n; h)$で表わされる．ここで，$\Sigma_T, \Sigma_R$はそれぞれ，送信アルファベット，受信アルファベットと呼ばれ，$L_1, \cdots, L_n$はΣ_T上の言語である．hは$L_1\times\cdots\times L_n$から$\Sigma_R^*$の空でない部分集合の族の中への写像である．$\bigcup_{u_1\in L_1, \cdots, u_n\in L_n} h(u_1, \cdots, u_n)$を$h(L_1, \cdots, L_n)$で表わす。

$v\in\Sigma_R^*$に対し，$v\in h(u_1, \cdots, u_n)$であるような列$u_1\in L_1, \cdots, u_n\in L_n$が存在するとき，$v$は$(u_1, \cdots, u_n)$へ復元可能であるといい，$v \overrightarrow{C} (u_1, \cdots, u_n)$とかく．$v \overrightarrow{C} (u_1, \cdots, u_n)$かつ$v \overrightarrow{C} (u_1', \cdots, u_n')$なら$u_1=u_1', \cdots, u_n=u_n'$が成立するとき，$v$は$(u_1, \cdots, u_n)$へ一意復元可能であるという。任意の$w\in h(L_1, \cdots, L_n)$が一意復元可能であるとき$C$は一意復元可能であるという。

本稿では，$\Sigma_T=\Sigma_R=\{0,1\}$，$h(u_1, \cdots, u_n)=u_1\odot\cdots\odot u_n$の場合を主に議論する．なお，$\odot$はシャフル演算[2]を表わし，

$$u_1\odot u_2=\{y_1z_1y_2z_2\cdots y_\ell z_\ell \mid \ell\geq 1,\ y_i, z_i\in\Sigma_T^*,\ y_1y_2\cdots y_\ell=u_1 \text{ かつ } z_1z_2\cdots z_\ell=u_2\}.$$

で定義される。$u_1\odot\cdots\odot u_n$ $(n\geq 3)$の場合も同様であり，又言語へ自然に拡張される（$L_1\odot L_2=\bigcup_{u_1\in L_1, u_2\in L_2} u_1\odot u_2$）。以下では

特にことわらない限り，Σ_T, Σ_R, h の記述は省略し，$C=(L_1, \cdots, L_n)$ で表わす。$L_1, \cdots, L_n$ が前後の文脈から自明のときは，$\overset{C}{\to}$ は単に $\to$ で置き換えてよい。例題1の L_1 と L_2 に関していえば，(L_1, L_2) は一意復元可能ではない。

$h(u_1, \cdots, u_n)=u_1 \odot \cdots \odot u_n$ は，1.で述べた様に各 T_i が完全にランダムな並行動作を行う場合を想定している。このような並行性に一定の制限を付すなら問題は一般に簡単になる。例えば，$h(u_1, \cdots, u_n)=\{u_{k_1}u_{k_2}\cdots u_{k_n} \mid k_i \neq k_j \ (i \neq j) \ 1 \le k_i, k_j \le n\}$ と決めることは，列 u_i を送信する間はチャネルが送信局 T_i によって排他的に専有されることを意味している。この制限を実現する実際的手法も考案されているが，[(5),(6)] 制御信号等の実時間通信には不向きであろう。また，信号がしょう突した場合に別の信号に変わる（例えば0と1のしょう突によって第3の記号2が生じる）という設定も現実性がある。このような場合も考慮に入れ一般に $\Sigma_T \neq \Sigma_R$ でよいことにしている。

モデルをさらに一般化することも可能である。興味深い一つの拡張は，送信局 T_i はチャネル上の信号の流れを観測でき，それによって，適応的に自らの送信列 u_i やその送信タイミングを変更できる能力を与えることである。一意復元のより容易な系が構成できると予想されるが，議論は別の機会にゆずることにする。

3. 一意復元可能性の判定

$L_1=(10)^*$, $L_2=(111+000)^*$ とする．このとき (L_1,L_2) は一意復元可能であることが証明できる．この場合も含め，$h(L_1,\cdots,L_n)$ が無限集合の場合は，$(L_1,\cdots,L_n)$ が一意復元可能であるかどうかの判定は一般には易しくない．

もし，$L_1,\cdots,L_n$ がすべて正規集合であれば，入力テープ1本，出力テープ n 本の非決定性有限変換器Mで，関係

$$R(M)=\{(z,x_1,\cdots,x_n)\mid z\xrightarrow[(L_1,\cdots,L_n)]{}(x_1,\cdots,x_n)\}$$

を限定するものが構成できる．(図2に，上記 $L_1=(10)^*$, $L_2=(111+000)^*$ の場合の変換器 M_1 の状態図を示す。) このような変換器Mに対し，それがある意味で決定性の動作を行うかどうか，即ち，任意の $z\in\{0,1\}^*$ に対し

$$|\{(z,x_1,\cdots,x_n)\mid(z,x_1,\cdots,x_n)\in R(M)\}|\leq 1$$

が成立するかどうかを判定するアルゴリズムの存在が証明できる ($(L_1,\cdots,L_n)$ が一意復元可能であるための必要十分条件が上記条件の成立であることは自明)．図2からも判るように，Mは ε (空列) 出力が許され，又過渡的な非最終状態の使用も許されているので，アルゴリズムの構築は注意深く行う必要がある (詳細略[8])。

［定理1］ 正規集合 $L_1,\cdots,L_n$ に対し，$(L_1,\cdots,L_n)$ が一意復元可能であるかどうか判定するアルゴリズムが存在する。

未確認ではあるが，効率の良いアルゴリズムは存在しないだろうと予想される。上記手法に従うなら，変換器Mの状態数rに対し，必要なステップ数はおよそ8^{r^2}になる。

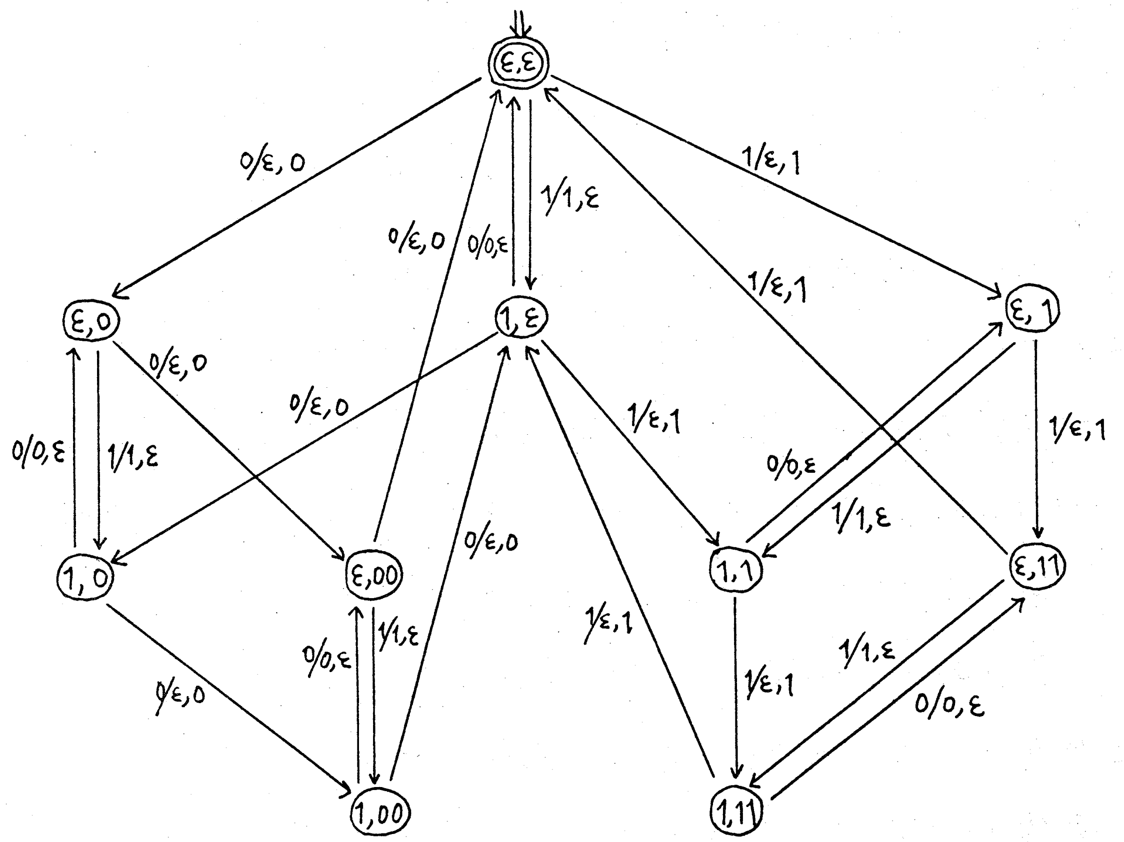

図2. 有限変換器M_1の状態図

4. 耐シャフル性のある符号化

現在までの議論より，一意復元可能性を実現するためには，送信局T_iが送信する可能性のある列L_iは相当多くの冗長性を有する（L_iが$\{0,1\}^*$に比べて十分疎である）ことが必要になると推察される。このことをより厳密に議論するため，各言

語L_iが，$L_i = g_i(\{0,1\}^*)$で与えられる場合を考察する。ここでg_iは$\{0,1\}^*$からそれ自身の中への一対一準同型（一意復号可能な符号化関数）である。以下，$(g_1(\{0,1\}^*), \cdots, g_n(\{0,1\}^*))$を簡単に$(g_1, \cdots, g_n)$と書く。

〔例題2〕　g_1とg_2が，$g_1(0)=100100$，$g_1(1)=011011$，$g_2(0)=110110$，$g_2(1)=001001$で決まる準同型とする．このとき(g_1, g_2)は一意復元可能ではない．なぜなら，$(011)^4 \to ((011)^4, \varepsilon)$かつ$(011)^4 \to ((011)^2, (110)^2)$であるから．

一意復元可能なg_1とg_2を構成しようと考え，仮に，$g_1(0)=11110000$，$g_1(1)=1010101010$と決めたとする．すると，

$$g_1(01) = 1111\,0000\,1010101010$$

$$g_1(10) = 10101010\ \ 10\,1111\,0000$$

とかける．2つの列の違いを観察することにより，(g_1, g_2)を一意復元可能にしようとするなら，$g_2(0)$，$g_2(1)$，$g_2(00)$，$\cdots$のいずれの列も部分列000を含んではならないことが判る。そうでないなら，例えば$g_2(0)=100001$であるなら，$g_1(01) \odot g_2(00)$と$g_1(10) \odot g_2(00)$は共通の列

$$10101010100001110101010100001$$

を含み，従って(g_1, g_2)は一意復元可能ではない。

この考え方を厳密化して我々は$(g_1, \cdots, g_n)$が一意復元可能であるための（線型時間で判定できる）一つの必要条件を得

ることができる。詳細は省くが[8]，本条件は次のことを主張している。一意復元可能な$g_1, \cdots, g_n$を構成しようとするなら，(i) $g_i(0)$と$g_i(1)$における0あるいは1の連続はできるだけ短くすべきであり，かつ(ii) $g_i(01)$と$g_i(10)$の上で述べた"相異度"をできるだけ大きくすべきである。直観的にいって，(i)と(ii)はトレードオフの関係にあり，例えば以下の事実が導ける。

［性質1］ $g_1(0)$，$g_1(1)$，$g_2(0)$，$g_2(1)$の長さがすべて5以下なら(g_1, g_2)は一意復元可能ではない。

このように，一意復元可能な$g_1, \cdots, g_n$を得ようとするなら$g_i(0)$や$g_i(1)$の長さはかなり長くなることが予想される。ところが，$(g_1, \cdots, g_n)$の一意復元可能性を検証する効率のよい方法は知られておらず，前節のアルゴリズムを用いるなら，$g_i(0)$と$g_i(1)$の長さの和が10程度になるとたとえ$n=2$の場合でもステップ数が2^{100}にもなってしまう．一意復元可能な$g_1, \cdots, g_n$が構成できるかは，それらが満たすべき（判定の容易な）十分条件を発見できるかにかかっている。

［定理2］ (g_1, g_2, g_3)が一意復元可能であるような一対一準同型g_1, g_2, g_3が存在する．

（証明） 方針だけ簡単に述べ詳細は文献[8]にゆずる。先ず，

$f_1(0) = 1^8 0^8$，$f_1(1) = 0^8 1^8$

$f_2(0) = (110)^{37}$，$f_2(1) = (001)^{37}$

で決まる (f_1, f_2) が一意復元可能であることを証明する．列 x のプレフィクスの全体を $pr(x)$ で表わし，$x_1 \in pr(x_2)$ 又は $x_2 \in pr(x_1)$ のとき $x_1 \approx x_2$ とかく。以下の4条件を満たす列 γ, $\gamma_1, u, v, w, x, u_1, v_1, w_1, x_1$ を考える。

(1) $\gamma \xrightarrow[(f_1,f_2)]{} (u, v)$ かつ $\gamma \xrightarrow[(f_1,f_2)]{} (w, x)$．

(2) $\gamma_1 \in pr(\gamma)$, $u_1 \in pr(u)$, $v_1 \in pr(v)$, $w_1 \in pr(w)$ かつ $x_1 \in pr(x)$.

(3) $\gamma_1 \in u_1 \odot v_1$ かつ $\gamma_1 \in w_1 \odot x_1$

(4) $u_1 \approx w_1$ かつ $v_1 \approx x_1$

つまり，(3)は(1)の復元の途中経過を示している。f_1 と f_2 の特色は，これ了の条件のもとでは $|u_1|$ と $|w_1|$ （$|v_1|$ と $|x_1|$）の差がそれ程大きくなうないことで，せいぜい8でおさえられることが示せる。

$u \neq w$ 又は $v \neq x$ （(f_1, f_2) が一意復元可能でない）と仮定しよう．上記性質より $u \approx w$ かつ $v \approx x$ ということはありえない．よって，ある列 $u_2, u_3, u_4, v_2, v_3, v_4$ に対し，(i) $u = u_2 f_1(0) u_3$ かつ $w = u_2 f_1(1) u_4$ 又は (ii) $v = v_2 f_2(0) v_3$ かつ $x = v_2 f_2(1) v_4$ のいずれかにかける．再び上記性質を利用し，このような u, v, w, x に対しては，$\gamma \to (u, v)$ かつ $\gamma \to (w, x)$ とはなりえないことが示せる。このとき，列 $f_1(0)$ と $f_1(1)$ においては0と1の個数がバランスしているが，$f_2(0)$ と $f_2(1)$ においてはかなり異っていることを利用する。

同様の手法により，以下の(g_1, g_2, g_3)が一意復元可能になるような整数p, q, rが存在することが示せる。

$g_1(0) = 1^p 0^p$, $g_1(1) = 0^p 1^p$

$g_2(0) = (1^{16} 0^8)^q (0^{16} 1^8)^q$, $g_2(1) = (0^{16} 1^8)^q (1^{16} 0^8)^q$

$g_3(0) = (110)^r$, $g_3(1) = (001)^r$

5. むすび

一意復元可能な$(g_1, \cdots, g_n)$は任意のnに対して存在すると予想されるが現在のところ証明されていない。ランダムな並行動作を制限した場合の一意復元可能性についても実用上有用と思われるいくつかの結果が得られているが別稿にゆずる。

謝辞　日頃御指導いただく京都大学情報工学教室矢島脩三教授，また貴重な御意見をいただいた同上林弥彦助教授に深謝します．本研究は一部文部省科学研究費による。

文献　(1) M.V. Wilkes, Computer 13 (1980). (2) S. Ginsburg, McGraw-Hill (1966). (3) A.C. Shaw, IEEE Trans. Software Eng. SE-4 (1978). (4) M. Jantzen, Theor. Comput. Sci 14 (1981) (5) R.M. Metcalfe and D.R. Boggs, CACM 19 (1976). (6) D.L. Nelson and R.L. Gordon, Proc. COMPCON Fall 78 (1978). (7) 福村，稲垣，AL81-81 (1981), (8) 岩間，AL81-111 (1982).