数理解析研究所講究録
第534巻 1984年 191-196

一連検索性をもつファイル構成について

近畿大・理工　田澤新成　（Shinsei Tazawa）

1. はじめに

情報収納検索システムにおいて質問に該当するレコードの検索に要する時間とレコードを収納するのに要する記憶容量との間には一般にトレードオフが存在する。しかしGhosh [2]は検索時間を短縮しかつ要する記憶容量の減少を実現するファイル方式を考案した。これはレコードを適当な順序に配列し各質問に該当するレコードを一連に検索することを可能にするような方式である。このようにレコードを一連に検索することを可能にする性質を一連検索可能性（CRP）といい、このときのレコードの配列をCR－配列とよぶ。

2値ベクトルの全体 $V_n=\{(v_1,v_2,\dots,v_n)\mid v_i=1,0;\ i=1,\dots,n\}$ を考える。ここでnは2値項目数，V_nのベクトルは2値レコードに対応する。S_iは第i番目について値1をもつ，すなわち，$v_i=1$なる2^{n-1}個のベクトルの列を表わし，$\overline{S_i}$は$v_i=0$なるベクトルの列を表わす。S_i（$\overline{S_i}$）はi番目の項目に該当

する（該当しない）レコードの検索を要求する質問に対応するということでS_iおよび$\bar{S}_i$は次数1の質問とよばれる。EhrichとLipski [1]は質問集合$\{S_1, \cdots, S_n\}$に対してCRPをもつV_nのベクトル（レコード）のCR－配列を構成するアルゴリズムを与えた。それの長さl_nは

$$(1) \qquad l_n = \left(\frac{2}{3}n + \frac{2}{9}\right)2^{n-1} - \frac{1}{9}(-1)^n$$

である。LuccioとPreparata [3]はl_nが質問集合$\{S_1, \cdots, S_n\}$に対するCR－配列について最小であることを示した。ここでは，次数1の質問集合$\{S_1, \cdots, S_n, \bar{S}_1, \cdots, \bar{S}_n\}$に対してCRPをもつ$V_n$のベクトルの配列の長さ$L_n$について

$$L_n \geqq \left(\frac{2}{3}n - \frac{1}{36}\right)2^n - \frac{2}{9}(-1)^n$$

が成り立ち，かつこの等号を満たすようなCR－配列を構成することができることを示す。

2. CR－配列の長さの下限

質問集合$\{S_1, \cdots, S_n, \bar{S}_1, \cdots, \bar{S}_n\}$に対して$V_n$のベクトルのCR－配列について長さが最小なものを$A_n$と記述する。$T_i = S_i$, $T_{i+n} = \bar{S}_i$ $(1 \leq i \leq n)$とおき，A_nにおいてT_iの最初の位置を$A_n[T_i]$とおく。ここで$A_n[T_{i_1}] < A_n[T_{i_2}] < \cdots < A_n[T_{i_{2n}}]$と仮定する。このとき$A_n$は次の性質をもつ。

性質1　T_iに属するベクトルが位置している連続した位置

の集合を$P[T_i]$と書くと，$|j-k|\geqq 2$なるj, kについて

(2) $$P[T_{i_j}] \cap P[T_{i_k}] = \phi$$

が成り立つ。

これは次のようにしてわかる。Q_kをT_k（$k=1, \cdots, 2n$）に属するベクトルの集合とする。T_{i_j}, $T_{i_{j+1}}$, $T_{i_{j+2}}$を考え，$P[T_{i_j}] \cap P[T_{i_{j+2}}] \neq \phi$とする。そのとき$P[T_{i_{j+1}}] \subseteq P[T_{i_j}] \cup P[T_{i_{j+2}}]$。その結果，$Q_{i_{j+1}} \subseteq Q_{i_j} \cup Q_{i_{j+2}}$である。このことから，ある$k$に対して$T_{i_j} = S_k$, $T_{i_{j+2}} = \bar{S}_k$　あるいは　$T_{i_j} = \bar{S}_k$, $T_{i_{j+2}} = S_k$である。だから$Q_{i_j} \cap Q_{i_{j+2}} = \phi$すなわち$P[T_{i_j}] \cap P[T_{i_{j+2}}] = \phi$となる。したがって(2)が成り立つ。

性質2　A_nの中で初めのn個の一連の空間$T_{i_1}, T_{i_2}, \cdots, T_{i_n}$をもつ部分列を$A_n^{(1)}$とし，残りの$n$個の一連の空間$T_{i_{n+1}}, T_{i_{n+2}}, \cdots, T_{i_{2n}}$をもつ部分列を$A_n^{(2)}$とする。このとき，$A_n^{(\alpha)}$の長さは$l_n$以上である（$\alpha = 1, 2$）。

$A_n^{(\alpha)}$が$\bar{S}_{j_1}, \bar{S}_{j_2}, \cdots, \bar{S}_{j_k}$を含むとする。$A_n^{(\alpha)}$について，各$t = 1, 2, \cdots, k$に対し$s_{j_t} = 1$のときは0に，$s_{j_t} = 0$のときは1に置きかえてえられる新しい列をBとかく。このときBは$\{s_1, \cdots, s_n\}$に対するCR－配列であり，その長さは$A_n^{(\alpha)}$のそれと同じである。Bの長さは[3]からl_n以上であるから，求める結果をうる。

定理　A_nの長さをL_nとかく。このとき

(3) $$L_n \geqq \left(\frac{2}{3}n-\frac{1}{36}\right)2^n-\frac{2}{9}(-1)^n$$

証明　異った2つの項目を指定した質問に該当するレコード（ベクトル）の個数は2^{n-2}であるから，$|P[T_{i_n}]\cap P[T_{i_{n+1}}]|\leqq 2^{n-2}$である。それ故，性質1，2から

$$L_n \geqq 2l_n-|P[T_{i_n}]\cap P[T_{i_{n+1}}]|\geqq 2l_n-2^{n-2}$$

が成り立つ。従って，(1)を適用して(3)をうる。

3．最小な長さをもつCR－配列の構成

この節で，(3)の等式に到達するCR－配列A_nを構成する手続きを述べる。この構成方法についてn=3の場合の構成例を図1に与える。与えられた質問集合に対するCR－配列Bにおいて，Bのj番目のベクトルをB[j]とかき，BにおけるセグメントTの最初の位置をB[T]とかく。

ステップ1．質問集合$\{S_1, S_2, \cdots, S_n\}$に対して，$A_n^{(1)}[S_{i+1}]<A_n^{(1)}[S_i]$，$i=1,2,\cdots,n-1$および$|P_1[S_1]-P_1[S_2]|=2^{n-2}$をみたすCR－配列$A_n^{(1)}$をEhrich-Lipskiアルゴリズム[1]を用いて構成する。ここで$P_1[T]$はセグメントTが位置している（$A_n^{(1)}$において）連続した位置の集合である。（図1では$P_1[S_1]-P_1[S_2]$の代りにS_1-S_2と示してある）

ステップ2．$A_n^{(1)}$の各ベクトルについて1を0に，0を1

に置きかえ，つぎに第1行と第2行を交換する。この新しい列を A^* とおく。ルール $A_n^{(2)}[j] \leftarrow A^*[l_n-j+1]$ によって $A_n^{(2)}$ を形作る。

注釈 $A_n^{(\alpha)}$ の長さは $\alpha=1,2$ について l_n である（[1], [3] を見よ）。$A_n^{(2)}$ は質問集合 $\{\bar{S}_1, \bar{S}_2, \cdots, \bar{S}_n\}$ に対するCR－配列である。さらに，$A_n^{(2)}[\bar{S}_2] < A_n^{(2)}[\bar{S}_1] < A_n^{(2)}[\bar{S}_3] < \cdots < A_n^{(2)}[\bar{S}_n]$ かつ $|P_2[\bar{S}_2] - P_2[\bar{S}_1]| = 2^{n-2}$ が成り立つ。ここで $P_2[T]$ はセグ

$$A_3^{(1)} = \begin{array}{ccccccccc} & & & \multicolumn{4}{c}{\overline{\quad S_2 \quad}} & \multicolumn{2}{c}{\overline{S_1-S_2}} \\ 1&1&0&0&0&1&1&1&1 \\ 1&0&0&1&1&1&1&0&0 \\ 1&1&1&1&0&1&0&1&0 \\ \multicolumn{4}{c}{\underline{\quad S_3 \quad}} & & \multicolumn{4}{c}{\underline{\quad S_1 \quad}} \end{array} \qquad A^* = \begin{array}{ccccccccc} 0&1&1&0&0&0&0&1&1 \\ 0&0&1&1&1&0&0&0&0 \\ 0&0&0&0&1&0&1&0&1 \end{array}$$

$$A_3^{(2)} = \begin{array}{ccccccccc} \multicolumn{4}{c}{\overline{\quad \bar{S}_2 \quad}} & & \multicolumn{4}{c}{\overline{\quad \bar{S}_3 \quad}} \\ 1&1&0&0&0&0&1&1&0 \\ 0&0&0&0&1&1&1&0&0 \\ 1&0&1&0&1&0&0&0&0 \\ \multicolumn{2}{c}{\underline{\bar{S}_2-\bar{S}_1}} & \multicolumn{4}{c}{\underline{\quad \bar{S}_1 \quad}} & & & \\ & & \multicolumn{7}{c}{\longleftarrow A^{**} \longrightarrow} \end{array} \qquad \text{in reverse order of } A^*$$

$$A_3 = \begin{array}{cccccccccccccccccc} & & & \multicolumn{4}{c}{\overline{\quad S_2 \quad}} & & & \multicolumn{4}{c}{\overline{\quad \bar{S}_2 \quad}} & & \multicolumn{4}{c}{\overline{\quad \bar{S}_3 \quad}} \\ 1&1&0&0&0&1&1&1&1&1&1&0&0&0&0&1&1&0 \\ 1&0&0&1&1&1&1&0&0&0&0&0&0&1&1&1&0&0 \\ 1&1&1&1&0&1&0&1&0&1&0&1&0&1&0&0&0&0 \\ \multicolumn{4}{c}{\underline{\quad S_3 \quad}} & & \multicolumn{4}{c}{\underline{\quad S_1 \quad}} & \multicolumn{4}{c}{\underline{\quad \bar{S}_1 \quad}} & & & & & \\ \multicolumn{9}{c}{\longleftarrow A_3^{(1)} \longrightarrow} & \multicolumn{9}{c}{\longleftarrow A^{**} \longrightarrow} \end{array}$$

図1　CR－配列 A_3 の構成

メントTが位置する位置の集合である。（図1では $P_2[\bar{S}_2]-P_2[\bar{S}_1]$ の代りに $\bar{S}_2-\bar{S}_1$ と記してある。）

ステップ3. $P_2[\bar{S}_2]-P_2[\bar{S}_1]$ に属する位置にあるすべてのベクトルを $A_n^{(2)}$ から削除する。そしてこの新しい列を A^{**} と書く。

ステップ4. $A_n^{(1)}$ と A^{**} を連接して $A_n=A_n^{(1)}A^{**}$ を得る。

注釈 $A_n^{(1)}$ の長さは ℓ_n, A^{**} の長さは ℓ_n-2^{n-2} であるから A_n の長さは $2\ell_n-2^{n-2}$, すなわち(3)の右辺に等しい長さである

参考文献

[1] H.D.Ehrich and W.Lipski,Jr., On the storage space requirement of consecutive retrieval with redundancy, Inform. Proc. Lett. 4(1976)101-104.

[2] S.P.Ghosh, File organization: The consecutive retrieval property, Comm. ACM 15(1972)802-808.

[3] F.Luccio and F.P.Preparata, Storage for consecutive retrieval, Inform. Proc. Lett. 5(1976)68-71.