数理解析研究所講究録
第 421 巻 1981 年 149-162

記憶階層のもとでの結合操作について

*上林 弥彦　*安浦 寛人
†岩間 一雄　*矢島 脩三
*京都大学 工学部
†京都産業大学 理学部

1. まえがき

計算機のデータベースへの利用が拡がるにつれ，データベースの諸演算を効率良く処理することが重要な問題となってきている。本稿は，記憶階層のある場合について，関係代数演算の中で最も時間のかかる操作の一つである結合操作の最適化について考察したものである。

最近のVLSI技術の進歩により主記憶容量は大幅に拡がってきているが，同時に高密度磁気記録の技術の進歩により二次記憶の容量もまた大きく拡がりつつあり，ランダムアクセスを主体とした主記憶と順アクセスを主体とした二次記憶からなる記憶の階層性は当分変化しないと考えられる。この意味で，記憶の階層性から生じるページ入れ換えの問題を考えたデータ処理のための効率の良いアルゴリズムの開発は非常

に重要であろう。記憶の階層性を考慮した場合，主記憶内の処理コストは無視して，ページ入れ換え回数のみで近似的にコストの評価ができる。

データベースの関係代数演算の中には，射影，結合，制約や各種集合演算等が含まれるが，この中では結合演算と直積が最も複雑で記憶容量も大きい。Kim [1] は，すべてのページの組合せを調べる場合について，これらの操作を行なう場合の，ページ入れ換え回数の最小化を行なう問題について検討した。実際には，これらの操作は他の操作と組合されることも多く，その場合にはすべてのページの組合せを調べなくても良いことが多い。本稿では，システムが各ページの内容に関するディレクトリを持っていて，この情報を利用することによりすべての組合せを調べなくて良いという場合について，ページ入れ換え回数を最小化する方法について考察する。

この最適解を求める問題は，特殊なグラフ上でのハミルトニアンパスを求める問題と同等になり，容易に解の存在判定がＮＰ－完全となることが示せる。このため，解が容易に求まるための条件や近似解の導出法について考察することが重要であり，いくつかの十分条件による解法等を示した。

2、問題のモデル

RおよびSを関係とし，それぞれ，$r_1, r_2, \ldots, r_p$ および $s_1, s_2, \ldots, s_q$ という部分関係に分けられているとする。ここで，各 r_i または s_j は1つのページに納められているものとする。2つのページ r_i と s_j よりなるページ対を u_{ij} で表わす。2つのページ対 u_{ij} と u_{kh} との間の距離 d は，u_{ij} から u_{kh} へバッファの内容を変化させるのに要するページ入れ換え回数で定義される。すなわち，

$$d(u_{ij}, u_{kh}) \begin{cases} = 0 & \text{if } i=k \text{ and } j=h, \\ = 1 & \text{if } i=k \text{ or } j=h \text{ but not both}, \\ = 2 & \text{otherwise.} \end{cases} \quad (1)$$

計算すべきページ対集合が与えられた場合に，それぞれのページ対が必ず一回現われるようなページ対の系列を，このページ対集合に対するスケジュールと呼ぶ。スケジュール $\omega = u_1 u_2 \ldots u_t$ のコストは，このスケジュールに従う順序でバッファ内のページ対を変化させてゆく場合に必要なページ入れ換え回数で定義される。すなわち，

$$c(u_1 u_2 \cdots u_t) = \sum_{i=1}^{t-1} d(u_i, u_{i+1}). \quad (2)$$

正確には最初のページ対をバッファに入れるため2だけのコストが必要となるが，この値はいつも一定なので本稿では

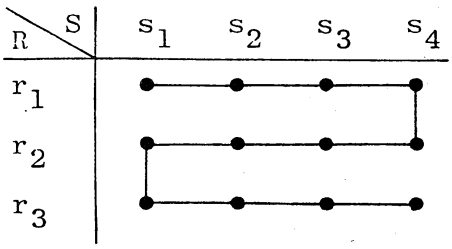
考えないこととする。つぎの性質は自明である。

$$c\ (\ u_1u_2\cdots u_t\ ) \leqq 2\ (t-1), \tag{3}$$

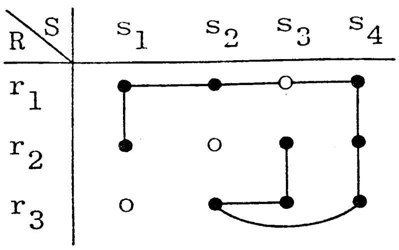
$$c\ (\ u_1u_2\cdots u_t\ ) \geqq t-1. \tag{4}$$

与えられたページ対集合に対し c の値を最小にするものを最適解という。最適解のコストが t－1 であれば最小解という。

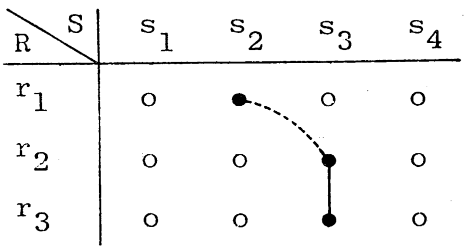
図1には，本稿の動機となった問題がまとめられている。もしすべてのページ対を計算する必要のある場合は，t=p をとなる。このような場合には必ずコストが t－1 の最小解が存在する[1]（図1 (a) 参照）。



(a) Cost = 11



(b) Cost = 8



(c) Cost = 3

図1　すべてのページ対を必要としない場合の最適解

Rは属性AとC上の関係で，Sは属性A，B，D上の関係とし，RはAの値でソートされ，SはまずBで次にAの値でソートされているとする。r_iとs_jは次の値の範囲のタプルを含むものとする。

r_1 : $1 \leqq A < 2$,

r_2 : $2 \leqq A < 4$,

r_3 : $4 \leqq A \leqq 6$,

s_1 : $B = 1,\ 1 \leqq A < 4$,

s_2 : $\begin{cases} B = 1,\ 4 \leqq A \leqq 6 \text{ and} \\ B = 2,\ 1 \leqq A < 2, \end{cases}$

s_3 : $B = 2,\ 2 \leqq A \leqq 6$,

s_4 : $B = 3$.

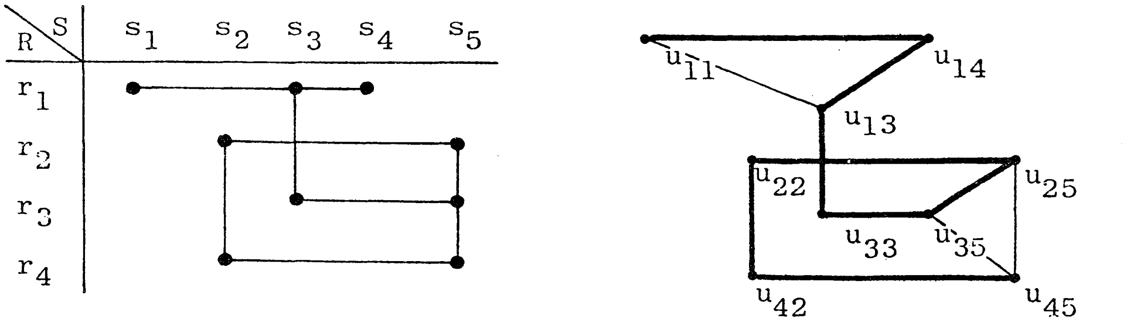
たとえば，r_2とs_2の組合せでは共通のタプルが存在しないのでu_{22}は計算しなくてよい。計算の必要なページ対は図1(b)に●で示されている9個であり，図の順で処理すればコストは8となる。

さらに$B = 2$という制約が関係Sに課されているとすると，図1(c)の3つのページ対のみを計算すれば良いことになる。この場合，最適解のコストは3で，最小解は存在しない。

この例のように，現実にすべてのページ対を計算しなくて良いことが多く，本稿ではこのような場合の最小解や最適解について考察する。

3、問題の複雑さ

図1に示すようなページ対グラフでは距離1のページ対どおしも直接つながれていないことが多いので，これをつないだページ距離グラフを考え，それによって問題の複雑さを検討する。



(a) A page-pair graph

(b) A page-pair distance graph

図2　ページ対グラフとページ距離グラフ

図2(a)にはページ対グラフの例が示されており，図2(b)は対応するページ距離グラフが示されている。ページ距離グラフでは，1つのページを共有するページ対どおしが部分完全グラフを作ることになる。つぎの命題は自明である。

<u>命題1</u>：与えられた問題に対し最小解の存在するための必要十分条件は，対応するページ距離グラフがハミルトニアンパスを持つことである。

図2の例では，そのようなハミルトニアンパスが太線で示

されており，最小解が存在することが分かる。

ハミルトニアンパスが存在するかどうかの判定問題は，よく知られているようにNP-完全である。命題1では，最小解の存在判定問題がハミルトニアンパスの存在判定問題に帰着できることが示されているので，すでにNP-完全であると知られている問題がこの最小解の存在判定問題に帰着できることを示せば，最小解の存在判定問題はNP-完全となる。このような問題として，立体グラフのハミルトニアンパスの存在判定を用いる。

〃立体グラフ（すべての節点の次数が3である）のハミルトニアンパスの存在判定問題はNP-完全である。〃[2]

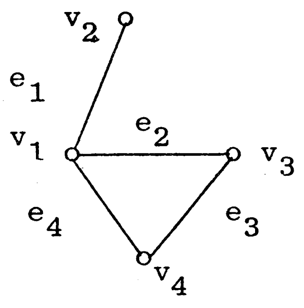
グラフGに対してH(G)は，Gの次数iの節点を節点数iの完全グラフとし，もとのi本の枝をそれぞれの節点に1本づつつないだものとする。

<u>命題2</u>：グラフGが与えられた場合，次の操作によりH(G)を得ることができる。

(1) Gの節点枝行列（incidence matrix）を求める。

(2) 各節点と各枝をそれぞれ1つの関係の中のページであるとみなして，ページ対グラフを作る。

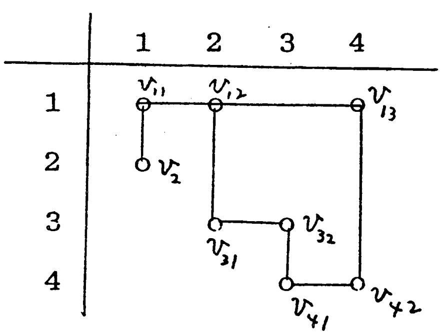
(3) このページ対グラフに対応するページ距離グラフが求めるH(G)である。



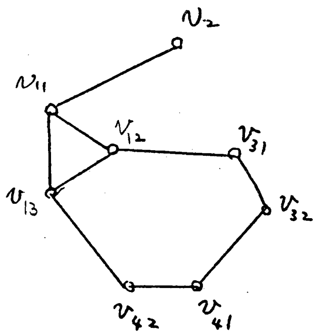
(a)

	e_1	e_2	e_3	e_4
v_1	1	1		1
v_2	1			
v_3		1	1	
v_4			1	1

(b)



(c)



(d)

図3　命題2の例

図3には，この操作の例を示している。図3(a)のグラフGの節点枝行列を図3(b)に示す。これをこのままページ対グラフとしたのが図3(c)である。このページ対グラフに対応するページ距離グラフは図3(d)に示される。このグラフでは，Gのv_1はv_{11}, v_{12}, v_{13}に，v_3はv_{31}, v_{32}に，v_4はv_{41}, v_{42}に対応しており，H(G)となっていることが分かる。

命題3：立体グラフのハミルトニアンパスの存在判定問題は，最小解の存在判定問題に帰着できる。

証明：Gを立体グラフとする。命題2の操作でH(G)を得ることができる。最小解の存在判定問題はH(G)の上でのハミルトニアンパスの存在判定問題となる。H(G)の上でのハミルトニアンパスが存在すればGのハミルトニアンパスも存在することを示せば証明は完了する。次の2つの場合がある。

(1) H(G)のハミルトニアンパスが，Gの1つの節点に対応する3節点を順次通っている場合（図4(a))：この場合はGの対応するハミルトニアンパスは，対応する枝より入り対応する枝より出るものとなる(図4(b))。

(2) H(G)のハミルトニアンパスが，Gの1つの節点に対応する3節点のうち2節点と1節点を別々に通っている場合（図4(c))：この場合，1節点はハミルトニアンパスの端点となっているので，対応する端点をずらすとよい（図4(d))．

命題4：最小解の存在判定問題はNP-完全である。

このことから近似的手法の開発が重要となる．

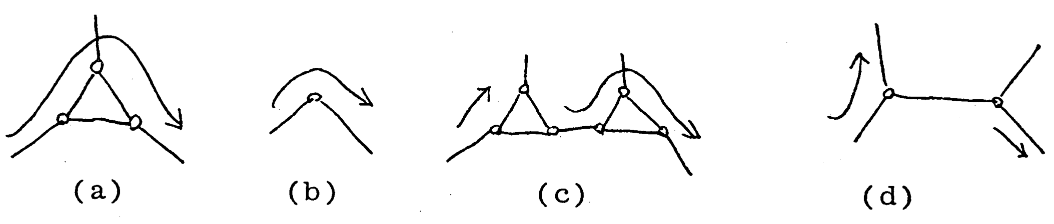


図4　H(G)のハミルトニアンパスからGのハミルトニアンパスへの帰着

4. 二部グラフによる解法

ハミルトニアンパスを求める問題は複雑であるがオイラーパスを求めるのは容易である。ページ対を枝に対応させることにより，最小解を求める問題をオイラーパスの問題に帰着させることを考える。

ページ対二部グラフは，2つの節点集合が，それぞれ，1つの関係のページに対応するもので，計算の必要なページ対 u_{ij}（r_i と s_j のページ対）があれば，r_i と s_j に対応する節点を結ぶ枝 e_{ij} が存在する。図5にその一例を示す。

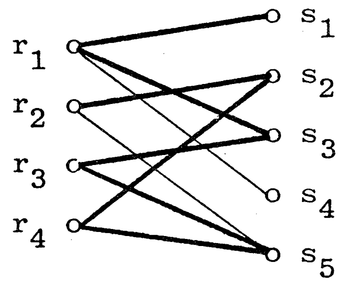


図5　ページ対二部グラフ

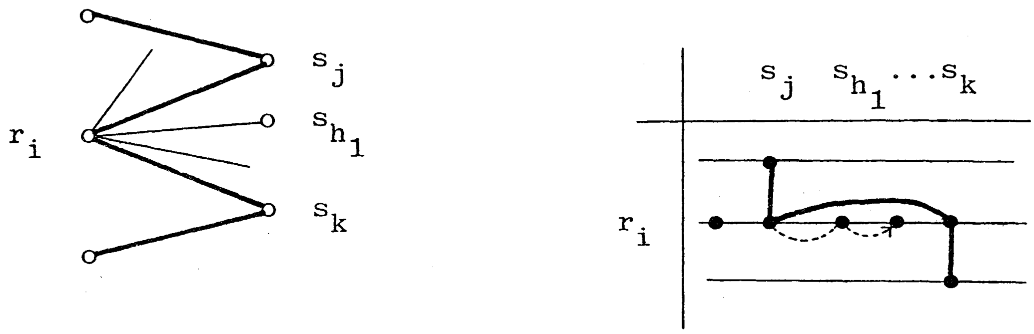
2つのページ対間の距離が1である場合に限って，それらに対応する枝は端節点を共有することとなるため，次の命題が得られる。

命題5：与えられた問題が最小解を持つための十分条件は，対応するページ対二部グラフがオイラーパスを有することである。すなわち，ページ対二部グラフが連結していて，奇次数の頂点数が0か2であることである。

図5のグラフは，オイラーパスを持たないが最小解の存在する例となっている。奇数次の節点数は3であるが最小解は次のようなものである。

$u_{11}u_{14}u_{13}u_{33}u_{35}u_{45}u_{42}u_{22}u_{25}$

図6は，この問題を説明するためのものである。



(a)　　　　　　(b)

図6　命題5が十分条件である理由

二部グラフ上であるパス（図6(a)の太線）を考える。これに対応して，ページ対グラフ上でのスケジュールを作ることができる（図6(b)）。もとのスケジュールには r_i と s_{h_1} よりなるページ対は含まれていないが，(r_i, s_j) (r_i, s_{h_1}) (r_i, s_k) の順で処理することによりこのページ対を含ませることができる。すなわち，ページ対二部グラフ上のパスが通る節点を端節点とする枝に対応するページ対はすべてスケジュールに含めることができる。

図5のグラフでは，太線で示したパスにより，最小解を求めることができる。

命題6：与えられた問題が最小解を持つための必要十分条件は，対応するページ対二部グラフに次の条件を満足するパスの存在することである。"任意の枝に対して，その端節点のうち少なくとも一方がパスに含まれる。"

図5の太線に対応するスケジュールは次の通りである。

$$u_{11}u_{13}u_{33}u_{35}u_{45}u_{42}u_{22}$$

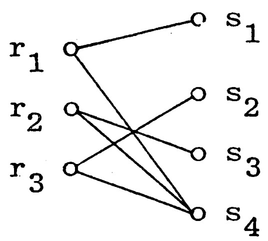
図6に示す方法により u_{14} と u_{25} を追加すれば最小解が得られる。

$$u_{11}u_{\underline{14}}u_{13}u_{33}u_{35}u_{\underline{25}}u_{45}u_{42}u_{22}$$

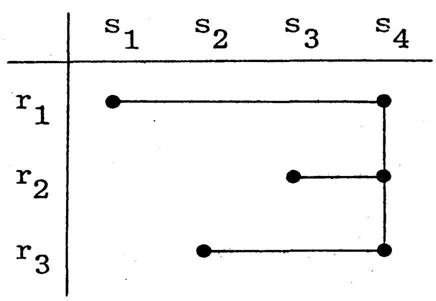
系1：RまたはSに対応する節点をすべて含むパスがページ対二部グラフに存在すれば，最小解が存在する。

系2：ページ対二部グラフの枝数が5以下の場合，最小解の存在する必要十分条件は，そのグラフが連結であることである。

系2はすべての可能性を調べて証明される。枝数が6で連結ではあるが最小解の存在しない例を図7に示す。



(a)



(b)

図7 連結で最小解のない場合

ページ対二部グラフにより，次のような方法が得られる。

[ページ対二部グラフによる方法]

(1) ページ対二部グラフを求める。もしグラフが連結していなければ，以下の方法でそれぞれの連結成分に対しスケジュールを決定した後，すべての成分を連結にするため枝を付け加える。

(2) 奇数次の節点数が4個以上の場合，次の操作でその数を減らす。

(2-1) 次の条件を満足するパスがあれば，そのパスに含まれる全ての枝を除く。

始点と終点は奇数次の節点である。これら以外のパス上の節点の次数は4以上である。

(2-2) 奇数次の節点がk個の奇数次の節点（次数3以上）とh個の偶数次の節点（次数4以上）とつながっていて，$k \geq h$なら，これらの枝を取り去る。命題6の条件を満足するパスを求めるとき，最初の奇数次の節点以外のすべての節点を通る必要がある。

(3) 上記の操作で奇数次の節点数を減らすことができなくなった場合，枝を加えることによって奇数次の節点を減らす。二部グラフの同じ節点集合に属さない節点どおしをつなぐ枝のコストは1であるが，同じ節点集合に属する節点どおしを

つなぐ枝のコストは2である。

(4) 奇数次の節点数が0か2になれば，オイラーパスを求めることができる。(2)(3)による変形を対応させることによりスケジュールを求めることができる。

5．むすび

本稿で与えた問題は，McGill大学のMerrett教授の問題に著者の一人上林が手を加えたものである。問題のNP-完全性については，McGill大学のAvis助教授も考察しているが，本稿では岩間によるものを示している。本稿は，矢島研究室で上林が示した問題に対して研究室内より出たアイデアをまとめたものであり，Merrettらの結果も含めた形での論文を作ることも考慮中である。

文献

〔1〕 W.Kim, "A new way to compute the product and join of relations," ACM SIGMOD '80, pp.179-187.

〔2〕 M.R.Gray, D.S.Johnson, R.E.Tarjan, "The planar hamiltonian circuit problem is NP-complete," SIAM J.Compting, 1976.

〔3〕 F.Harary, "Graph Theory", Addison-Wesley, Reading Mass., 1969.

〔4〕 A.V.Aho, J.E.Hopcroft and J.D.Ullman,"The Design and Analysis of Computer Algorithms," Addison-Wesley, Reading Mass., 1974.