〈シンポジウム〉

組合せ理論

日	時	昭和49年4月5日（金） 13:00～16:30
会	場	東京工業大学第4新館223講義室
司	会	高橋磐郎（早大）

ま と め

昨年春行なわれた第1回のシンポジウムでは，ORの中心課題とも典型ともいうべきMPについての総合的報告が行なわれたが，今回はORからみればやや特殊なあるいは新鮮な“組合せ理論”についての最近の話題が以下にのべる3人の方々によって紹介された．特殊な話題であるにもかかわらず会場は超満員で，この種の問題への関心が高いことが示された．

講演者の方々はいずれもその方面で第一線の研究成果をあげておられる方々で，最先端の話題を紹介されたため，内容は高度であったが，解説はわかりやすく問題の本質は十分伝えられたと思われる．

第1は，「組合せ理論の情報検索への応用」と題して広島大学の山本純恭氏が講演された．情報検索におけるファイル構成に組合せ数学的理論を導入したのは実験計画法で有名なR.C. Boseらが最初と思われる．Bose氏は，実験計画法で生まれたBIB (balanced incomplete block) や直交表およびこれらの基礎になる有限幾何を用いて効率のよいファイルを構成することを提案されたが，山本氏はファイルの問題を冗長度の最小化という一種の最適化問題としてとらえ，それが完全グラフのクロー (claw) 分割というグラフ理論上の問題になることを示し，その分割の実際のアルゴリズムを開発され，（ある条件の下での）最良のファイルを構成された．このご自身の業績もふくめ，組合せ理論を応用したファイル構成問題の一連の話題を紹介された．

第2は，「組合せ理論におけるグラフとマトロイド」と題して東京大学の伊理正夫氏が話された．周知のように氏はネットワーク理論，グラフ理論の第一人者で，最近はマトロイド理論に手をそめ，その方面の研究に成果を上げておられる．マトロイド (matroid) とは matrix と oid（のようなもの）との合成で，“マトリックスのようなもの”とでもいうべきもので，純粋数学の分野で線形代数におけるベクトルなどの一次独立性の概念を抽象化し公理的に構成された理論であるが，最近 W.T. Tutte などによってグラフ理論に応用されはじめたものである．伊理氏はこのマトロイドの公理論を概説されたあと，最近のご自身のおもな研究成果を二つ紹介された．その一つは bipartite グラフの点をマトロイドの元とみて，そこにいわゆるマッチングの問題を考えることにより，グラフ理論やネットワーク理論でのいくつかの重要な問題を統一的に論ずることができ，またその問題の解法アルゴリズムを与えることができること，第二は0,1変数を含む線形計画の双対問題をマトロイド的に論ずることができることを示された．

第3は，埼玉大の古林　隆氏の「クラスター分析」で，組合せ理論としてはもっとも実用的価値の高いものと見られている．これはいくつかの特性をもつ個体の集団があるとき，その任意の二つの個体間に，その特性に基づく，近さ（一種の距離）の概念を導入し，その集団を近いものどうしのグループに分割する手法である．はじめは生物学などでの階層分類構造に用いられていたようであるが，最近はOR手法としてかかせないものとなってきた．古林氏は，従来無批判に用いられていた距離の概念を，分割の目的に応じた目的関数の最適化との関連において，定義すべきことを，強調され，その観点から従来の手法を総合しまとめて論じられた．

以下に3氏のシンポジウム予稿を掲載することにする．ただし，古林氏の分はあとから追加されたもの，伊理氏の分には筆者が若干の注釈をつけ加えた．

（高橋磐郎）

プログラム

13:00～14:00 組合せ理論の情報検索への応用
山本 純恭(広島大学)
14:10～15:10 組合せ理論におけるグラフとマトロイド
伊理 正夫(東京大学)
15:30～16:30 クラスター分析
古林 隆(埼玉大学)

組合せ理論の情報検索への応用

山本 純恭(広島大学理学部)

1. 序

大量の情報の蓄積および検索に電子計算機が活用されるようになって久しいが，レコードそのものの特性を検索目的に従って抽出し，フォーマット化する問題，検索目的の分析すなわち質問の集合の構造解明の問題，フォーマット化されたレコードのファイル方式の問題，検索方式の問題など，たがいに別個に考察することの不可能な問題であり，今後の研究の進展にまつべきものが多い．

ここではフォーマット化された大量のレコードを収納し，一定の質問の集合に対して効率よく応答するファイルを構成する問題に，組合せ数学を応用する最近の研究について，われわれの最新の結果を含めて紹介したい．

2. 転置ファイル

最も原始的なファイル方式は，検索要求とは無関係にレコードの発生順に一連にファイルするいわば**積上げ法**とでもいうべきものであろう．質問の発生ごとにファイル全体を通した検索が必要であるから，収納時間，ファイル容量ともに最小であるが，検索時間最大であり実時間的でない．

これに対して，主要な質問に対して1：1に**バケツ**を用意し，該当するレコード(固有番号等)を収納する方式は**転置ファイル**とよばれ，広く用いられている．この方式によれば，検索時間は最小であるが，予想される質問のすべてに対してバケツを用意することになると，同一のレコードをくり返していくつかのバケツに収納することになり，収納に多大の時間と手数を必要とし，**冗長度**の大きい大容量のファイルとなる．基本的な特性項目に限定して転置ファイルを構成すればこれらの欠陥はいくらか緩和されるが，合成質問に対して**マッチング**の手数を必要とし，時間的にも無視できなくなる．

これらの収納時間，検索時間，ファイル容量のたがいに競合する関係を調和するファイル方式の追及に組合せ理論を応用する研究が開始されたのは1960年代の後半である．IBMのAbraham, Ghosh, Ray-Chaudhuri等とノースカロライナ大学のBose教授等の共同研究がその最初であろう．

3. BFS_2 と $BMFS_2$

l 個の項目のそれぞれについて該当するかどうかで特性づけられるいわゆる2値のレコードをファイルする方式として，Abraham等[A1]は有限射影幾何を用いて BFS_2 (Balanced filing schemes) を構成した．項目を点に，バケツを直線に対応させる方式である．実験計画法でよく研究されている会合数1の不完備ブロック計画 (BIBD) を用い，l 個の処理を項目に，b 個のブロックをバケツに対応させることにより，ブロックの大きさ k のとき，各バケツは $\binom{k}{2}$ 個の2項目の質問を含み，$\lambda=1$ より2項目の質問はすべてただ一つのバケツに属している．$\binom{l}{2}$ 個の2項目質問を $\binom{k}{2}$ 個ずつまとめてバケツを構成することになるから，バケツ数は転置ファイルにくらべ $1/\binom{k}{2}$ となるのみならず，二つ以上の2項目質問に該当するレコードの重複収納を相当程度回避できる特色がある．

各項目が多水準の場合については，Ghosh等[A5]の有限幾何を用いる2項目質問に対する $BMFS_2$ がある．われわれは有限射影幾何の構造定理を積極的に用いる $BMFS_2$ を構成した[A 8][A 9]．

この種の研究の詳細は[A 1]～[A 11]を参照されたい．

4. CRFS

Ghosh [B 1] [B 2] [B 3] は，最近一連検索可能性 (consecutive retrieval property, CR性) をもつファイル方式について考察している．たとえば二つの質問 A, B に該当するレコードをファイルする場合，レコードを $A\cap B^c$, $A\cap B$, $A^c\cap B$ の順に一連にすれば，質問 A, B のみならず $A\cap B$ に対しても一連に検索することができる．重複して収納する手順が省略できるのみならず，$A\cap B$, $A\cup B$ に対してマッチングの必要がない点ですぐれているという発想である．しかし，わずか三つの質問に対してすら，CR

性をもたせることは必ずしも可能でない．

Ghosh は質問の集合が CR 性をもつための種々の十分条件を与えると同時に，質問の全体を部分集合すなわちバケツの和に分割し，各バケツがレコードの集合に対して CR 性をもつ例を示している．

われわれは l 項目について 2 値のレコードの全体 $\boldsymbol{R}^{(l)}$ に対する 2 値の質問の全体 $\boldsymbol{Q}^{(l)}$ を，それぞれが CR 性をもつバケツに分割する問題（CR 分割とよぶ）について(1) CR 性をもつバケツに含まれる質問の最大数は $2l-1$ であること，(2) CR 分割数の下限は $t_l=\left[\left\{\binom{l}{[l/2]}\right\}+1\right\}/2\right]$ であること，および(3)最小 CR 分割を求めるアルゴリズムへの手がかり（$l=3,4,5,6$ については下限）を与えることに成功した．詳細は [B1〜6] を参照されたい．

5. NBFS$_2$ から HUBFS$_2$ へ

BFS$_2$ は l 項目について 2 値のレコードの全体に対する $\binom{l}{2}$ 個の 2 項目質問を $\binom{k}{2}$ 個ずつまとめてバケツを構成する方式であるが，Chow [C 1] はBFS$_2$ は有限幾何を用いるからそのパラメータに制約があること，有限体上の計算を必要とすることが欠点であるとし，$\binom{l}{2}$ 個の質問に具体的な順序をつけ，その順序に従って C 個ずつまとめてバケツを構成する方式を提案し NBFS$_2$ (New BFS$_2$) と名づけた．また BFS$_2$ とパラメータの等しい NBFS$_2$ をいく通りか比較し，冗長度も BFS$_2$ より小さいことを示している．

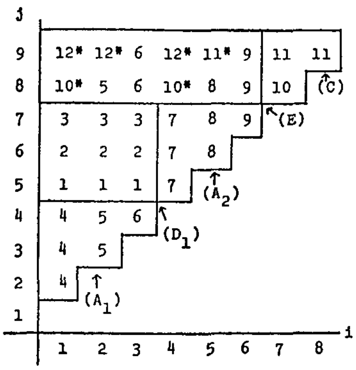
われわれは，2 項目の質問をたとえば C 個集めて一つのバケツを構成し，サブバケツを適当に用意することにより，これらの質問の少なくとも一つに該当するレコード（のアクセッション・ナンバー）をただ 1 回だけ収納し，いずれにも該当しないレコードは収納しない方式をとる場合，一つのレコードがそのバケツに収納される確率すなわちバケツの冗長率を問題にした．このとき，問題の一つはレコードの確率分布をどう考えるかということと，いま一つは，項目の点に，2 項目の質問を線に対応させるとき，C 個の質問からなるバケツは C 本の線の集合に対応するが，この集合のグラフ的構造と冗長率の関係である．

前者については，従来一様分布が考えられてきたが，これはかなり非現実的であると思われる．われわれは，項目の置換に対して不変な分布のクラスを考えた．これはレコード ω の 2 値の特性値ベクトルが $(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_l)$ $(\varepsilon_i=0 \text{ or } 1)$ であるとき，確率 $P(\omega)$ がウエイト $w=\Sigma\varepsilon_i$ のみの関数である場合であって，従来の一様分布はもちろん，各項目が独立でかつ該当する確率が一定値 p である場合を含む広い分布のクラスである．

このような分布のもとで，冗長率を考えるとき，バケツの冗長率は，そのバケツのグラフ構造が claw

表 1 HUBFS$_2$ のバケツの構成例 ($l=9$, $C=3$)

K_9 の claw 分解



バケツと対応する質問

バケツ	質　問
B_1	(1,5)(2,5)(3,5)
B_2	(1,6)(2,6)(3,6)
B_3	(1,7)(2,7)(3,7)
B_4	(1,2)(1,3)(1,4)
B_5	(2,3)(2,4)(2,8)
B_6	(3,4)(3,8)(3,9)
B_7	(4,5)(4,6)(4,7)
B_8	(5,6)(5,7)(5,8)
B_9	(6,7)(6,8)(6,9)
B_{10}	(1,8)(4,8)(7,8)
B_{11}	(5,9)(7,9)(8,9)
B_{12}	(1,9)(2,9)(4,9)

表 2 BFS$_2$ のバケツの構成例 ($l=9$, $C=3$)

EG (2, 3) の点と直線の結合行列

線	点（項　目）								
（バケツ）	1	2	3	4	5	6	7	8	9
1	1				1				1
2		1				1			1
3			1				1		1
4				1				1	1
5	1		1					1	
6	1	1		1					
7		1	1		1				
8			1	1		1			
9				1	1		1		
10					1	1		1	
11	1					1	1		
12		1					1	1	

バケツと対応する質問

バケツ	質　問
B_1	(1,5)(1,9)(5,9)
B_2	(2,6)(2,9)(6,9)
B_3	(3,7)(3,9)(7,9)
B_4	(4,8)(4,9)(8,9)
B_5	(1,3)(1,8)(3,8)
B_6	(1,2)(1,4)(2,4)
B_7	(2,3)(2,5)(3,5)
B_8	(3,4)(3,6)(4,6)
B_9	(4,5)(4,7)(5,7)
B_{10}	(5,6)(5,8)(6,8)
B_{11}	(6,7)(6,1)(7,1)
B_{12}	(7,8)(7,1)(8,1)

表3 $NBFS_2$ のバケツ構成例 ($l=9$, $C=3$)

質問の質号

j \ i	1	2	3	4	5	6	7	8
9	8	15	21	26	30	33	35	36
8	7	14	20	25	29	32	34	
7	6	13	19	24	28	31		
6	5	12	18	23	27			
5	4	11	17	22				
4	3	10	16					
3	2	9						
2	1							
1								

バケツと対応する質問

バケツ	質問
B_1	(1,2)(1,3)(1,4)
B_2	(1,5)(1,6)(1,7)
B_3	(1,8)(1,9)(2,3)
B_4	(2,4)(2,5)(2,6)
B_5	(2,7)(2,8)(2,9)
B_6	(3,4)(3,5)(3,6)
B_7	(3,7)(3,8)(3,9)
B_8	(4,5)(4,6)(4,7)
B_9	(4,8)(4,9)(5,6)
B_{10}	(5,7)(5,8)(5,9)
B_{11}	(6,7)(6,8)(6,9)
B_{12}	(7,8)(7,9)(8,9)

表4 冗長度の比較 ($l=9$, $C=3$)
($HUBFS_2$, BFS_2, $NBFS_2$, IFS_2)

レコードのウエイト	確率分布 一様	独立	その他	
	$\binom{l}{w}\left(\frac{1}{2}\right)^l$	$\binom{l}{w}\left(\frac{1}{3}\right)^w\left(\frac{2}{3}\right)^{l-w}$	$\binom{l}{w}p_w^{(1)}$	$\binom{l}{w}p_w^{(2)}$
0	.00195	.02601	—	—
1	.01758	.11706	.1	—
2	.07031	.23411	.2	.25
3	.16401	.27313	.4	.50
4	.24609	.20484	.2	.25
5	.24609	.10242	.1	—
6	.16406	.03414	—	—
7	.07031	.00732	—	—
8	.01758	.00091	—	—
9	.00195	.00005	—	—
冗長度				
IFS_2	9.000	4.000	3.600	3.250
BFS_2	6.000	3.111	2.971	2.821
$NBFS_2$	5.500	2.922	2.817	2.706
$HUBFS_2$	5.250	2.815	2.724	2.631

型（完全バイグラフ $K_{1,c}$）であるとき，最小となることが証明された．

したがって，$\binom{l}{2}$ 個の2項目質問を C 個ずつ集めてバケツを構成し，ファイル方式を作る場合，すべてのバケツを claw 型にすることが可能であれば，ファイル方式として最小の冗長度をもつものが構成されることになる．

この問題に対する解答は，完全グラフの claw 分解定理という形で与えられ，その必要十分条件は，$C\left|\binom{l}{2}\right.$, $l\geq 2C$ である．

証明および分解のアルゴリズムは省略するが，$l=9$, $D=3$ の場合について表1に例示した．なお，この方式を $HUBFS_2$ (Hiroshima Univ. BFS_2) と名づけた．

表2,3は $l=9$, $C=3$ の場合の BFS_2, $NBFS_2$ のバケツ構成例を示し，表4は4種のレコードの分布について，これらの方式および転置ファイル (IFS_2) の冗長度を比較したものである．

詳細は文献 [C 1]～[C 10] を参照されたい．

参考文献

[A]

[1] Abraham, C. T., S. P. Ghosh and D. K. Ray-Chaudhuri, *Information and Control*, **12** (1968).

[2] Bose, R. C., C. T. Abraham and S. P. Ghosh, *Proc. Symp. on Combinat. Math.*, 1969.

[3] Bose, R. C. and G. G. Koch, *SIAM J. Appl. Math.*, **17** (1969).

[4] Ghosh, S. P., *Information Sci.*, **1** (1969).

[5] Ghosh, S. P. and C. T. Abraham, *IBM J. Res. Develop*, **12** (1968).

[6] Koch, G. G., *Univ. of North Carolina Mimeo Ser.*, 552 (1967).

[7] Ray-Chaudhuri, D. K., *SIAM J. Appl. Math.*, **16** (1968).

[8] 山本純恭，寺本武昭，二神かほる，情報処理学会 第12回，1971.

[9] Yamamoto, S., T. Teramoto and K. Futagami, *Information and Control*, **12** (1972).

[10] 山本純恭，オペレーションズ・リサーチ，12月号 (1972).

[11] 高橋磐郎，生産研究所紀要，No. 4 (1973).

[B]

[1] Ghosh, S. P., *IBM Research Report*, No. RJ 708 (1970).

[2] Ghosh, S. P., *Comm. ACM*, **15** (1972). (*IBM Res. Rep.*, RJ 765)

[3] Ghosh, S. P., *IBM Research Report*, No. RJ 895 (1971).

[4] 山本純恭，潮 和彦，"対話型情報処理に関する研究" (1973).

[5] 山本純恭，潮 和彦，池田秀人，玉利文和，"対話型情報処理に関する研究" (1973).

[6] 山本純恭，潮 和彦，池田秀人，玉利文和，情報処理学会第14回大会，1973.

[C]
[1] Chow, D. K., *Information and Control*, **15** (1969).
[2] 池田秀人，垂水共之，山下有二，山本純恭，情報処理学会第14回大会，1973.
[3] 中野博信，鶴本良夫，古川博之，CFS研究会資料 第2回 (1973).
[4] 山本純恭，潮 和彦，重枝新成，特定研究集会（8月，東北大学），1973.
[5] 山本純恭，重枝新成，潮 和彦，池田秀人，特定研究集会（12月，大阪大学），1973.
[6] 山本純恭，重枝新成，潮 和彦，池田秀人，浜田 昇，情報処理学会第14回大会，1973.
[7] 山本純恭，潮 和彦，重枝新成，池田秀人，浜田 昇，情報処理学会第14回大会，1973.
[8] 山本純恭，池田秀人，潮 和彦，重枝新成，日本数学会（応数），1974.
[9] 山本純恭，池田秀人，潮 和彦，重枝新成，日本数学会（応数），1974.
[10] 山本純恭，池田秀人，重枝新成，潮 和彦，浜田 昇，CFS研究会資料第4回(1974).

組合せ理論におけるグラフとマトロイド

伊理 正夫（東京大学工学部）

1. マトロイド (matroid) [1]～[4]

matroid $M\ (E,\ *)$ (E：有限集合)

(イ) $*=\boldsymbol{B}(\subseteq 2^E)$, $\boldsymbol{B}\ni B$：基

$\langle \boldsymbol{B}1\rangle$ $B_1, B_2\in\boldsymbol{B}$ ⇨「$B_1\subsetneqq B_2$ ということはない」

$\langle \boldsymbol{B}2\rangle$ $B_1, B_2\in\boldsymbol{B},\ x\in B_1$ ⇨ $\exists\, y\in B_2:(B_1-\{x\})\cup\{y\}\in\boldsymbol{B}$

(ロ) $*=\boldsymbol{I}(\subseteq 2^E)$, $\boldsymbol{I}\ni I$：独立集合

$2^E-\boldsymbol{I}\ni D$：従属集合

$\langle \boldsymbol{I}1\rangle$ $I_1\in\boldsymbol{I},\ I_2\subseteq I_1$ ⇨ $I_2\in\boldsymbol{I}$

$\langle \boldsymbol{I}2\rangle$ $I_1, I_2\in\boldsymbol{I},\ |I_2|>|I_1|$ ⇨ $\exists\, x\in I_2-I_1: I_1\cup\{x\}\in\boldsymbol{I}$

(ハ) $*=\rho\ (:2^E\to N^+)$, $\rho(A)$：Aの階数

$\langle\rho1\rangle$ $0\leqq\rho(A)\leqq|A|\ (A\subseteq E)$

$\langle\rho2\rangle$ $A\subseteq B\ (\subseteq E)$ ⇨ $\rho(A)\leqq\rho(B)$

$\langle\rho3\rangle$ $\rho(A\cup B)+\rho(A\cap B)\leqq\rho(A)+\rho(B)$

(ハ′) $*=\rho$

$\langle\rho'1\rangle$ $\rho(\phi)=0$

$\langle\rho'2\rangle$ $\rho(A\cup\{x\})=\rho(A)$または$\rho(A)+1$

$\langle\rho'3\rangle$ $\rho(A\cup\{x\})=\rho(A\cup\{y\})=\rho(A)$ ⇨ $\rho(A\cup\{x,y\})=\rho(A)$

(ニ) $*=\boldsymbol{C}\ (\subseteq 2^E)$, $\boldsymbol{C}\ni C$：サーキット

$\langle \boldsymbol{C}1\rangle$ $C_1, C_2\in\boldsymbol{C}$ ⇨「$C_1\subsetneqq C_2$ ということはない」

$\langle \boldsymbol{C}2\rangle$ $C_1, C_2\in\boldsymbol{C},\ C_1\neq C_2,\ x\in C_1\cap C_2$ ⇨ $\exists\, C\in\boldsymbol{C}: C\subseteq(C_1\cup C_2)-\{x\}$

(ホ) $*=\mathrm{cl}\ (:2^E\to 2^E)$, $\mathrm{cl}(A)$：Aの閉包

$\langle \mathrm{cl}\,1\rangle$ $A\subseteq\mathrm{cl}(A)=\mathrm{cl}(\mathrm{cl}(A))(A\subseteq E)$

$\langle \mathrm{cl}\,2\rangle$ $A\subseteq B(\subseteq E)$ ⇨ $\mathrm{cl}(A)\subseteq\mathrm{cl}(B)$

$\langle \mathrm{cl}\,3\rangle$ $x\in\mathrm{cl}(A\cup\{y\})-\mathrm{cl}(A)$ ⇨ $y\in\mathrm{cl}(A\cup\{x\})-\mathrm{cl}(A)$

(イ⇄ロ) 基＝極大独立集合
独立集合＝基の部分集合

(ロ⇄ハ) $\rho(A)=\max\{|I|\,|\,I\subseteq A,\ I\in\boldsymbol{I}\}$
$\boldsymbol{I}=\{I|\rho(I)=|I|\}$

(ロ⇄ニ) サーキット＝極小従属集合
独立集合＝サーキットを含まない集合

(ハ⇄ホ) $\mathrm{cl}(A)=\{x|\rho(A\cup\{x\})=\rho(A)\}$
$\rho(A)=\min\{|I|\,|\,I\subseteq A,\ \mathrm{cl}(I)\supseteqq A\}$

∘ $\mathrm{cl}(A)=A\cup\{x|\exists\, C\in\boldsymbol{C}: x\in C\subseteq A\cup\{x\}\}$

∘ $B_1, B_2\in\boldsymbol{B}$ ⇨ $|B_1|=|B_2|$

∘ $I\in\boldsymbol{I},\ x\in\mathrm{cl}(I)-I$ ならば，$I\cup\{x\}$に含まれるサーキットがただ一つ存在し，それは$\{y|(I\cup\{x\})-\{y\}\in\boldsymbol{I}\}$と表わせる．

双対マトロイド $M(E,\ \boldsymbol{B})$, $M^*(E,\ \boldsymbol{B}^*)$

$\boldsymbol{B}^*=\{B^*|E-B^*\in\boldsymbol{B}\}$

直和マトロイド

$M_1(E_1,\ \boldsymbol{I}_1)+M_2(E_2,\ \boldsymbol{I}_2)=M(E,\ \boldsymbol{I})$

$E=E_1\cup E_2\quad(E_1\cup E_2=\phi)$

$\boldsymbol{I}=\{I_1\cup I_2|I_1\in\boldsymbol{I}_1, I_2\in\boldsymbol{I}_2\}$

写像 $f:E\to E^*$

$f(M(E,\ \boldsymbol{I}))=M^*(E^*,\ \boldsymbol{I}^*)$

$\boldsymbol{I}^*=\{f(I)|I\in\boldsymbol{I}\}$

合併マトロイド

$M_1(E_1,\ \boldsymbol{I}_2)\cup M_2(E_2,\ \boldsymbol{I}_2)=M(E,\ \boldsymbol{I})$

$E=E_1\cup E_2$

$\boldsymbol{I}=\{I_1\cup I_2|I_1\in\boldsymbol{I}_1,\ I_2\in\boldsymbol{I}_2\}$

独立マッチング $\boldsymbol{M}$

$M_1(E_1,\ \boldsymbol{I}_1)$, $M_2(E_2,\ \boldsymbol{I}_2)$：マトロイド

$G(E_1, E_2,\ U)$：E_1, E_2を両側の点集合とする2部グラフ

$\boldsymbol{M}$：Gの上のマッチング($\subseteq U$)

$\partial^+\boldsymbol{M}\equiv\boldsymbol{M}$の端点になっている$E_1$の部分$\in\boldsymbol{I}$

$^-\partial\boldsymbol{M}\equiv\boldsymbol{M}$の端点になっている$E_2$の部分$\in\boldsymbol{I}_1$

∘ $G_{\boldsymbol{M}}$：$G(E_1, E_2,\ U)$から次のようにして作られ

る有向グラフ

$U-\boldsymbol{M}$ の枝は E_1 側から E_2 側へのみ通れるとする.

$\boldsymbol{M}$ の枝は E_2 側から E_1 側へ通れるとする.

$\partial^+\boldsymbol{M}$ の点 x から $\mathrm{cl}(\partial^+\boldsymbol{M})-\mathrm{cl}(\partial^+\boldsymbol{M}-\{x\})-\{x\}$ の各点へ向かう枝を作る (cl は M_1 におけるもの).

$\mathrm{cl}(\partial^-\boldsymbol{M})-\partial^-\boldsymbol{M}$ の点 y から $\partial^-\boldsymbol{M}\cup\{y\}$ に含まれるサーキット (M_2 における) の y 以外の各点へ向かう枝を作る.

◦最大独立マッチングの作り方 [5]

独立マッチング $\boldsymbol{M}$ に対して $G_{\boldsymbol{M}}$ を作り, $G_{\boldsymbol{M}}$ の上で $E_1-\mathrm{cl}(\partial^+\boldsymbol{M})$ の点から $E_2-\mathrm{cl}(\partial^-\boldsymbol{M})$ の点への最短路 P を求める. P の上では $\boldsymbol{M}$ の枝と $U-\boldsymbol{M}$ の枝は交互に並んでいるから, $\boldsymbol{M}'=(\boldsymbol{M}-P\cap\boldsymbol{M})\cup(P\cap(U-\boldsymbol{M}))$ も独立マッチングで $|\boldsymbol{M}'|=|\boldsymbol{M}|+1$ である. そのような P が存在しなければ $|\boldsymbol{M}|$ は最大である.

独立割当問題 [6]

$G(E_1, E_2, U)$ の U の各枝に重みが与えられているとき, 最大独立マッチングで重みの和が最小のものを求める問題.

合併マトロイドの基と最大独立マッチング

$M(E, \boldsymbol{I})=M_1(E_1, \boldsymbol{I}_2)\cup M_2(E_2, \boldsymbol{I}_2)$ の基 B は, $B=B_1\cup B_2(B_i: M_i$ の基) という形にかかれる. Bを求めるには $|B_1\cap B_2|$ が最小であるような B_1 と B_2 を求めればよい. $|B_1\cap B_2|=|B_2|-|B_2\cap B_1{}^*|$ ($B_1{}^*$: M_1 の双対マトロイド $M_1{}^*$ の基) であるから, $M_1{}^*$ と M_2 で共通部分最大の基 $B_1{}^*B_2$ を求めればよい. $(G|E_1, E_2, U)$ の U は E_1 と E_2 の対応する元を結ぶ枝を採ればよい.

2. 行列, グラフとマトロイド

$\boldsymbol{M}$: 行列, $\boldsymbol{E}$: $\boldsymbol{M}$ の列ベクトルの集合

$\boldsymbol{I}$= {一次独立な列ベクトルの集合}

$\boldsymbol{G}$: グラフ, $\boldsymbol{E}$: $\boldsymbol{G}$ の枝の集合

① M
$\boldsymbol{I}$ = {無ループ集合}, $\boldsymbol{B}$= {木},
$\boldsymbol{C}$ = {初等的なループ}, $\rho(A)=A$ よりなる G の部分グラフの階数, $\mathrm{cl}(A)=A$ の枝とともにループをなすような枝の集合

② M^*
$\boldsymbol{I}^*$= {無カットセット集合}, $\boldsymbol{B}^*$ = {補木},
$\boldsymbol{C}^*$= {初等的なカットセット}, $\rho^*(A)=E-A$ の枝を短絡したグラフの零度

Lehman の問題 [7], [8], [9]

グラフ G の上に二つの木 B_1, B_2 を選び, $|B_1\cup B_2|$ をできるだけ大きくすること.

共通木の問題 [10], [11]

同一の枝の集合 E からなる二つのグラフ G_1, G_2 の上に, それぞれ, 木 B_1, B_2 を選び, $|B_1\cap B_2|$ をできるだけ大きくすること.

枢軸変換によって行列の項別階数を最小にする(階数を最大にする) 問題 [12]

大野の問題 [13]

階数 n の $n\times 2n$ 行列 M_1 の列ベクトルの集合 E_1 の中から, "i 番目の列と $n+i$ 番目の列とは同時に選ばないようにして", なるべく多くの一次独立なものを選ぶこと (M_2 として $n\left\{\overbrace{\begin{bmatrix} \begin{matrix}1 & & O\\ & \ddots & \\ O & & 1\end{matrix} & \begin{matrix}1 & & O\\ & \ddots & \\ O & & 1\end{matrix}\end{bmatrix}}^{2n}\right.$ という形の行列をとり, 最大独立マッチングの問題に帰着させる).

3. {0, 1} − 変数を含む線形計画問題の双対問題

添字集合 I, J, E

マトロイド $M(E, \boldsymbol{I})$

主問題:

◦変数 $x_j\geqq 0\ (j\in J),\ y_k=0 \text{ or } 1(k\in E)$

◦条件 $\sum_{j\in J} a_{ij}x_j-\sum_{k\in E} b_{ik}y_k=c_i(i\in I)$,

$\{k\,|\,y_k=1\}\in\boldsymbol{I}$

◦目的関数 $f(x)=\sum_{j\in J} e_jx_j\rightarrow \min.$

双対問題:

◦変数 $u_i(i\in I)$

◦条件 $\sum_{i\in I} a_{ij}u_i\leqq e_j\ (j\in J)$

◦目的関数 $g(u)=\sum_{i\in I} c_iu_i+\underbrace{\min_{K\in I}\Big[\sum_{k\in K}\Big(\sum_{i\in I} b_{ik}u_i\Big)\Big]}_{\text{マトロイドにおける minimum spanning tree の問題 (→Kruskal の方法)}}$

双対定理: 主問題の実行可能解 x, y と双対問題の実行可能解 u に対して,

$f(x)\geqq g(u)$

(最適解に対して=とは必ずしもならないが, =になれば両者はともに最適解.)

参 考 文 献

[1] Crapo, H. H. and G. C. Rcta, *Combinatorial Geometries*, MIT Press, 1971.

[2] Tutte, W. T., "Lectures on Matroids," *J. of Research of the NBS*, **69 B** (1965), 1–47.

[3] Wilson, R. J., "An Introduction to Matroid

Theory," *Amer. Math. Monthly*, **80** (1973), 500–525.
[4] Whitney, H., "On the Abstract Properties of Linear Dependence," *Amer. J. of Math.*, **57** (1935), 509–533.
[5] 富沢, 伊理, 電子通信学会回路とシステム研資料, CST 73-78 (1974年1月).
[6] 富沢, 伊理, 独立割当問題の解法と回路の複雑度の問題への応用（電子通信学会論文誌投稿中).
[7] Lehman, A., "A Solution of the Shannan Switching Game," *SIAM J.*, **12** (1964), 687–725.
[8] Bruno, J. and L. Weinberg, *Linear Algebra & Its Appl.*, **4** (1971), 17–54.
[9] 岸, 梶谷, 電子通信学会論文誌, **51-A** (1968), 196–203.
[10] 小沢, 電子通信学会回路とシステム研資料, CST 73-47.
[11] 伊理, 冨沢, 同上, CST 73-48.
[12] Iri, M., *Linear Algebra & Its Appl.*, **2** (1969), 427–466.
[13] Oono, Y., *Proc. Symp. on Active Networks & Feedback Systems*, PIB, 1960, 475–486.

(注1) マトロイドの定義について

有限集合 E の部分集合の族 $\boldsymbol{B}$ ($\subseteq 2^E$) が公理〈$\boldsymbol{B}$1〉, 〈$\boldsymbol{B}$2〉(→本文) をみたすとき, $(E, \boldsymbol{B})$ をマトロイドといい $M(E, \boldsymbol{B})$ とかく. また $\boldsymbol{B}$の元 B を基という (2^E はE の部分集合全体).

E をベクトルの有限集合とみたとき, 上記 B ($\in \boldsymbol{B}$) の定義は, E の中での基底の概念を抽象化したものであることは容易にうなずけるであろう. またあるグラフ, たとえば図1の枝全体を E とし, 木 (すべての点をつなぐ極小の枝の集合) を基とみれば, 基全体$\boldsymbol{B}$は上の公理をみたす. 図1の例では $E=\{1,2,3,4,5\}$, $\boldsymbol{B}=\{\{2,3,4\}, \{1,4,5\}, \{1,2,5\}, \{1,2,4\}, \{2,3,5\}, \{1,3,4\}, \{2,4,5\}, \{1,3,5\}\}$.

しかしこのような直観的イメージから離れてみれば, マトロイド $M(E, \boldsymbol{B})$ とは E の上に〈$\boldsymbol{B}$1〉, 〈$\boldsymbol{B}$2〉なる公理をみたす, $\boldsymbol{B}$ ($\subseteq 2^E$) を与えたものにほかならない.

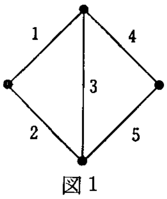


図1

$\boldsymbol{B}$の元つまり基の部分集合をすべて独立集合と呼び, その全体を$\boldsymbol{T}$とする, つまり

$$\boldsymbol{I} = \{I \mid I \subseteq B, \ B \in \boldsymbol{B}\}$$

とすると, $\boldsymbol{I}$は〈$\boldsymbol{I}$1〉, 〈$\boldsymbol{I}$2〉(→本文) の関係をみたす. 逆に〈$\boldsymbol{I}$1〉, 〈$\boldsymbol{I}$2〉をみたすような $\boldsymbol{I}$ ($\subseteq 2^E$) を勝手に与え, $\boldsymbol{I}$の, 極大の, 元全体を$\boldsymbol{B}$とすれば, $\boldsymbol{B}$は〈$\boldsymbol{B}$1〉, 〈$\boldsymbol{B}$2〉をみたすことがわかる. したがって〈$\boldsymbol{I}$1〉, 〈$\boldsymbol{I}$2〉をみたす$\boldsymbol{I}$を与えることによってマトロイドを定義することもできる. このときマトロイドを $M(E, \boldsymbol{I})$ とかく.

このほかサーキットの族やランクを与えてもマトロイドを定義することができる. いずれにしても E 上になんらかの関係を与えることであるから, これを $M(E, *)$ と書くのである. 本文ではいくつかの定義の与え方とそれらの関連を示してある.

(注2) 最大独立マッチングについて

本文で与えている2部グラフ $G(E_1, E_2, U)$ で Uは問題に応じて適宜選ぶべきものである. 例として図2のグラフで木を基とするマトロイド $M(E, \boldsymbol{B})$ とその双対マトロイド $\boldsymbol{M}^*(E, \boldsymbol{B}^*)$ を考え, M の基 (つまり木) と M^* の基 (つまり補木) との共通部分で最大となるようなものを求めてみよう.

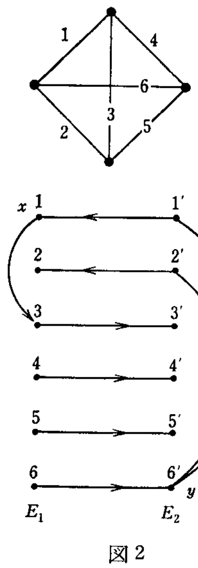


図2

この場合, E_1, E_2 は E のコピーにほかならない. Uとしては対応する元を単に結んだもの (→図2). このとき

$$\boldsymbol{M} = \{(1,1'), (2,2')\}$$

は独立マッチングであり,

$$\partial^+\boldsymbol{M} = \{1,2\}, \ \mathrm{cl}(\partial^+\boldsymbol{M}) = \{1,2,3\}$$

$$\partial^{-}\boldsymbol{M}=\{1',2'\},\ \mathrm{cl}(\partial^{-}\boldsymbol{M})=\{1',2',6'\}$$

となるから，$G_{\boldsymbol{M}}$ を作ると図2のようになり，最短パスは

$$P=(6,6',\ 1',1,\ 3,3')$$

であって，

$$\boldsymbol{M}'=\{(2,2'),\ (3,3'),\ (6,6')\}$$

となる．

同一の操作をくり返すと $\boldsymbol{M}'$ が最大独立マッチングであることがわかる．

(注3) $\{0,1\}$ 変数を含む線形計画の双対問題について

ここで定式化された問題は，もし $\boldsymbol{I}$ として 2^E をとれば，つまり E のすべての部分集合を独立集合とみれば，いわば自明な結果であるが，$\boldsymbol{I}$ にある構造をもたせると，双対問題の目的関数の第2項が，いわゆる minimum spanning tree などの効率のよいアルゴリズムによって求まるため，有効な結果を生む．とくにグラフやネットワーク上の線形計画に有用であろう．

クラスター分析

古林　隆（埼玉大学）

1. クラスター分析の目的

(1) 与えられた集合の階層的構造（階層的分類）を求めること．

(2) 与えられた集合の分割の中で，ある規準に従って最適なものを求めること．

クラスター分析の起こりは(1)の場合であるが，最近各方面でクラスター分析が注目されるようになったのは，(2)の場合に使われるようになったからである．

2. 階層的手法の一般手順

n：個体数（個体は番号 $1,2,\cdots,n$ で表わす）

$\delta(A,B)$：クラスター A,B 間の距離

（クラスターは $\{1,2,\cdots,n\}$ の部分集合で表わす）

(Ⅰ) $\Gamma=\{\{1\},\{2\},\cdots,\{n\}\}$ とおく．

$\delta(\{i\},\{j\})$ を計算する（かまたは読む）．

(Ⅱ) $\min_{A,B\in\Gamma}\delta(A,B)=\delta(G,\ H)$

となる $G,\ H(\in\Gamma)$ を求める．

(Ⅲ) G と H を結合する．すなわち，

$\Gamma-\{G,\ H\}+\{G\cup H\}\ \rightarrow\ \Gamma$

とし，Γ を書きだす．$|\Gamma|=1$ ならば終わり．

(Ⅳ) すべての $A(\in\Gamma,\ \neq G\cup H)$ に対して $\delta(A,G\cup H)$ を計算して，(Ⅱ) にもどる．

階層的手法は，元来，前項(1)の場合のために考えだされたものであるが，(2)の場合にも簡便法として使われている．

3. 階層的手法の評価規準

(1) $\delta(A,B)$ はどのような実際的な意味をもっているか．

(2) $\delta(A,\ G\cup H)$ の計算が簡単であるか．定義によらずに，$\delta(A,\ G)$, $\delta(A,\ H)$, $\delta(G,H)$, $|A|$, $|G|$, $|H|$ などで計算できることが望ましい．

(3) どのような規準に関して最適な（または最適に近い）分割を与えるか．

規準としては，分割 Γ の関数 $f(\Gamma)$ を考えて，それを最大または最小にするものを最適と考えることにする．

従来，(1)，(2)に重きがおかれていたが，今後は規準（目的関数）をはっきりさせ，最適な分割あるいはそれに近いものを作りだす「よい手法」を開発することが望まれる．

4. おもな階層的手法

D型（d_{ij} は個体 i と個体 j との間の距離を表わす）

名　称	$\delta(A,B)$	$\delta(A,G\cup H)$ の計算	目的関数 ($\Gamma=\{C_1,C_2,\cdots,C_v\}$
最短距離法	$\min_{i\in A,j\in B} d_{ij}$	$\min\{\delta(A,G),\ \delta(A,H)\}$	$\min_{\alpha,\beta}\min_{i\in C_\alpha,j\in C_\beta} d_{ij}\ \rightarrow\ \max$
最長距離法	$\max_{i\in A,j\in B} d_{ij}$	$\max\{\delta(A,G),\ \delta(A,H)\}$	$\max_{\alpha}\max_{i,j\in C_\alpha} d_{ij}\ \rightarrow\ \min$
群間平均距離法	$P(A,B)$	$\frac{1}{\lvert G\cup H\rvert}\{\lvert G\rvert\delta(A,G)+\lvert H\rvert\delta(A,H)\}$	$\min_{\alpha,\beta} P(C_\alpha,C_\beta)\ \rightarrow\ \max$

群内平均距離法	$Q(A \cup B)$	$\frac{1}{w(A \cup G \cup H)}\{w(A \cup G)\delta(A,G)$ $+w(A \cup H)\delta(A,H)+w(G \cup H)\delta(G,H)$ $-w(A)Q(A)-w(G)Q(G)-w(H)Q(H)\}$	$\max_{\alpha} Q(C_\alpha) \to \min$

ただし，$P(A,B)=\frac{1}{|A||B|}\sum_{i\in A, j\in B} d_{ij}$，$Q(A)=\frac{1}{\binom{|A|}{2}}\sum_{\substack{i,j\in A\\ i<j}} d_{ij}$，$w(A)=\binom{|A|}{2}$とする．

X型（x_{ik}は個体iの特性kについての観測値を表わす）

名　称	$\delta(\{i\},\{j\})$	$\delta(A,B)$	$\delta(A,G \cup H)$の計算	目的関数($\Gamma=\{C_1,C_2,\cdots,C_v\}$)
重心法	$\sum_k (x_{ik}-x_{jk})^2$	$\sum_k (\bar{x}_k{}^A-\bar{x}_k{}^B)^2$	$\frac{1}{\lvert G \cup H\rvert^2}\{\lvert G\rvert\lvert G \cup H\rvert\delta(A,G)$ $+\lvert H\rvert\lvert G \cup H\rvert\delta(A,H)$ $-\lvert G\rvert\lvert H\rvert\delta(G,H)\}$	$\min_{\alpha,\beta}\sum_k(\bar{x}_k{}^{C\alpha}-\bar{x}_k{}^{C\beta})^2 \to \max$
Ward 法	$\frac{1}{2}\sum_k (x_{ik}-x_{jk})^2$	$S(A \cup B)$ $-S(A)-S(B)$	$\frac{1}{\lvert A \cup G \cup H\rvert}\{\lvert A \cup G\rvert\delta(A,G)$ $+\lvert A \cup H\rvert\delta(A,H)$ $-\lvert A\rvert\delta(G,H)\}$	$\sum_\alpha S(C_\alpha) \to \min$
群内平均距離法	$\frac{1}{4}\sum_k (x_{ik}-x_{jk})^2$	$V(A \cup B)$	$\frac{1}{w(A \cup G \cup H)}\{w(A \cup G)$ $\delta(A,G)+w(A \cup H)\delta(A,H)$ $+w(G \cup H)\delta(G,H)$ $-w(A)V(A)-w(G)V(G)$ $-w(H)V(H)\}$	$\max_\alpha V(C_\alpha) \to \min$

ただし，$\bar{x}_k{}^A=\frac{1}{|A|}\sum_{i\in A} x_{ik}$，$S(A)=\sum_k\sum_{i\in A}(x_{ik}-\bar{x}_k{}^A)^2$，$V(A)=\frac{1}{|A|}S(A)$，$w(A)=|A|^2$ とする．

参　考　文　献

古林　隆，“クラスター分析における階層的手法”，経営科学，**18**，1 (1974)，3–10.

上の論文の参考文献にあげたもの以外に，次の文献（単行本）がある．

Anderberg, M. R., *Cluster Analysis for Applications*, Academic Press, 1973.

Jardine, N. and R. Sibson, *Mathematical Taxonomy*, Wiley, 1970.