

Title	個人選好による情報アクセスに適したデータモデルについて
Author(s)	林, 幸雄
Citation	情報処理学会研究報告 : データベース・システム研究会報告, 98(58): 381-388
Issue Date	1998-07
Type	Journal Article
Text version	publisher
URL	http://hdl.handle.net/10119/3415
Rights	<p>社団法人 情報処理学会, 林幸雄, 情報処理学会研究報告 : データベース・システム研究会報告, 1998(57), 1998, 381-388. ここに掲載した著作物の利用に関する注意: 本著作物の著作権は(社)情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。 The copyright of this material is retained by the Information Processing Society of Japan (IPSJ). This material is published on this web site with the agreement of the author (s) and the IPSJ. Please be complied with Copyright Law of Japan and the Code of Ethics of the IPSJ if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof. All Rights Reserved, Copyright (C) Information Processing Society of Japan.</p>
Description	

個人選好による情報アクセスに適した データモデルについて

林 幸雄 (yhayashi@jaist.ac.jp)

北陸先端科学技術大学院大学 知識科学研究科

概要. 知的な情報アクセスにおける人間的な主観の重要性を改めて指摘し、それに適したデータモデルを紹介する。本モデルは、従来のデータモデルの代数的な操作とは異なる幾何学的な操作を持ち、それは主観的な視点に応じた情報の関連性の強弱で張られる関数空間上で行なわれる。また、情報幾何学の α -分布族によるモデルの具現化を考え、ファジィ論理や確率的情報検索モデル、さらにはマーケティングサイエンスにおける個人選好モデルとの接点を示す。嗜好性を表す関数空間上で、これらは統一的に議論できるかも知れない。

On a New Data Model suitable for Intellectual Accesses by Personal Preference

Yukio Hayashi (yhayashi@jaist.ac.jp)

The School of Knowledge Science,
Japan Advanced Institute of Science and Technology, Hokuriku.

Abstract. We will point out the importance of subjectivity on intellectual information accesses, because we may require human knowledge rather than machine data in the processes of comprehension and creation. On this point, we introduce a new data model suitable for such accesses. It has geometrical data operations on a function space which consists of the degree of relations for information fragments (e.g. Home Pages) from subjective viewpoints, comparing with the conventional data models with algebraic operations. It is shown that our data model defined by the α -family of distributions in information geometry has a relation with Fuzzy logic and a probabilistic model of information retrieval. Further, we show another relation with mathematical models of personal preference in marketing science. We may totally consider these different types of models in the function space of preference.

1 はじめに

近年のインターネットの急成長によって、情報過多によるアクセスの困難さが知的生産性の低下をもたらすこと [2] が、より現実的で一層深刻な問題になりつつある。さまざまな知的資産が氾濫する情報（知識）社会において、我々は何を求めて情報や知識を得ようとしているのだろうか？特に、ネットワーク環境に分散した情報や知識に対しては、「偶然かも知れない知的な出会いが期待はずれでない（新しいメッセージが感じとれる）」ことを求めているのではないだろうか。それは、情報や知識を利用する側の解釈の多様性¹によって感じとれるものと考えられる。例えば、科学やビジネス分野などの研究活動においては、原著者の観点による内容的な関連性を示す引用文献を参考にしながらも、読者自らの観点でさまざまな関連性を再構築していく過程で、発見的な探索活動を行なっている。このような研究や新製品/市場の創出などにおける知的活動には、単に（電子化された）内容をキーワードなどによる索引付きデータとして管理するだけでは不十分で、利用者の活きた視点や観点から情報にさまざまな解釈が与えられ、情報間の関連性を吟味できるような、発見的な知の探索を支援するシステムが要求される。

2 個々人に宿る人間的な知（主観と客観の再考）

一般に、「知識」というと体系化された客観的存在をイメージしがちである。しかしながら、そもそも「客観」自体が合理的な主観とも言えるので、本質的に外化/形式化できない人間の認識（理知的な深層理解）に深くかかわっており、特に知の探索においては、常識的な客観から外れた主観的な知の重要性を改めて再考する必要がある。

ファジィ理論では、近代合理主義における主客二原論や分析主義（近代人の知の根拠）に対して、2値論理に還元不能な主観そのものを科学の場に持ち込む＜科学の主観化＞や現実的人間（日常的文脈）における主観の重要性が指摘されている [18]。すなわち、デカルトの（機械的/記号处理的な）客観に対して、パスカルの（心情認識的/直観的な）主観が擁護される。その主観は、

- 個人的主観、
- （根源的な認識にかかわる人間に共通の）普遍的主観、
- （過去の宗教観/価値観などに対する）反省された主観

¹個々の情報に対する解釈の多様性は、情報集合全体でみると情報を一意に分類できないこと [6] に相当する。例えば、文献 [12] では、“分類は検索に対する事物の局在化（検索範囲を限定可能）である。単純で確実なヒットの為には量が多いほど多次元分類が必要。分類は整理であり整頓とは異なる。つまり、狭義にはアクセスの為に、広義には前代からの秩序の洗練・厳密化がその本質である。情報は無限で新しい関係の発見が可能（知的万華鏡）。”等、文献 [22] では、“分類による体系化は対象の1つの解釈で、物の見方や捉え方が時代によって変われば分類法も変わり得る。時間がたてば、個人の分類基準も他人のと同様に使いづらくなる。量が増えると固い分類では曖昧になりやすい。一方、分類を階層的にするほど複雑で使いにくくなり、分類自体が無意味になる。1つの体系をあきらめて種々の視点や観点で関係付けたネットワーク分類が有効となってくるかもしれない。”等と述べられている。

などに区別されている。ファジィ理論は、曖昧な意図に適合した個々のシステムや論理を構築するために特に個人的主観に注目したが、我々は、開放ネットワーク系における知の流通、活用、創造のプロセスを対象とするために、相補的な主観と客観の相互作用や、知のスパイラルアップ [16] まで含めて、主観を重視する（図 1 参照）。

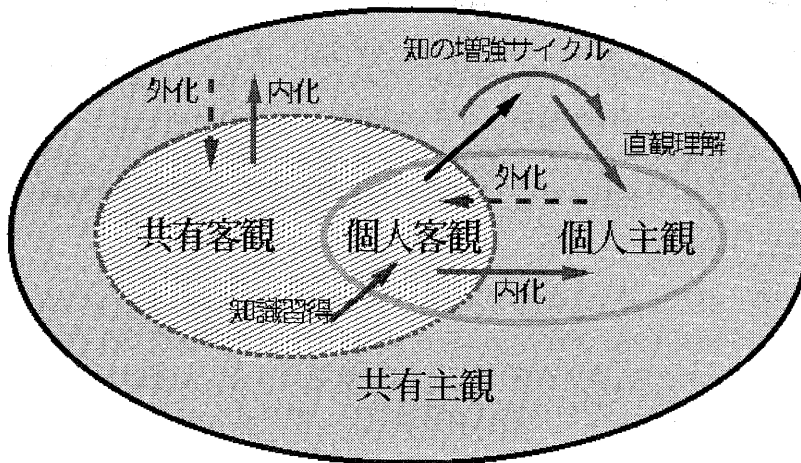


図 1: 個人知と共有知における知の流通、活用、創造: 共有主観や共有客観からインスパイアされた各個人の直観を介して、新しい概念を形成しつつ理解を促進/共有化していく

3 動的に知の空間を構築するモデル

動的に知の空間を構築するための枠組として提案された関係分布モデル [10] の概要を述べた後、従来のデータモデルとの相違点や情報検索モデル（ファジィ論理演算）との接点 [9] を概説する。ここで我々は、機械的なデータよりも人間的な情報や知識を対象とすることを強調しておきたい。

3.1 関係分布モデルの提案

まず、モデルのデータ構造を、 m 個の情報 $x \in \{1, 2, \dots, m\} = \mathcal{X}$ と、 n 個の異なる視点（属性）名、 n 個の各視点 i に対する情報 x の関連度: $p_i(x)$ 、利用者が情報探索を行なう際に指示する各視点 i の重要度パラメータ: θ^i とで規定する²。

次に、関係分布モデルの最も基本的なデータ操作を、予めモデルに規定された関連度 $p_i(x)$ を、視点の重要度パラメータ θ の値に従って重み付けて組合せて、新たな関連度 $p(x; \theta)$ を生成することで定義する。具体的には、情報幾何学の α -分布族 [1] などの統計的多様体

²関係分布モデルでは、視点（属性）名と情報 x の関連度 $p_i(x)$ の規定に関して、キーワードなどによる統語的で画一的な情報の関連付けではなく、視点に応じた状況に則して情報が関連付けられた空間を構築できるように、個々人の情報の選択眼や嗜好性などの主観的なデータを基にしながらも、具体的なデータ操作が可能なモデルを考えて、知的な情報探索の数理的性質を捉えることを狙っている。

$S = \{p(x; \theta)\}$ 上で視点の重要度パラメータ θ の値を変化させながら、関連度 $p(x; \theta)$ を動的に生成する。このとき、関連度の値が大きい情報 x を選択的に参照することで、(重要度の重み付けに従った)複合的な視点からの情報探索ができる。

3.2 従来のデータモデルとの比較

従来のデータモデルと関係分布モデルとの基本的な考え方の違い [10] を述べよう。

従来モデル: ハイパーテキスト等のように直接張られたリンクはもちろん、属性値の条件一致などで関係付けられる (関係表などにおける) 間接的なリンクも含めた、情報間のリンクを介した操作から、選択的に情報 (集合) を抽出する。情報とリンク集合中の個々の要素が操作対象。

関係分布モデル: 視点に対する情報の関連性の強弱を表すリンクを、(重み付けて) 組み合わせた全体的なリンク強度の分布空間を考えて、リンク強度の分布空間に対する幾何学な操作を介して、選択的に情報 (集合) を抽出する。関連の強弱の付いたリンク強度の集まりである、確率分布の関数空間が直接の操作対象。

これらの違いをアナロジー的に言えば、前者は構造化されたネットワークの関係リンクをたぐるような (代数的な) 操作モデルで、後者は拡散的に関係を持つ場の空間に対する (微分幾何学的な) 操作モデルと考えられる。

まず、確率的なモデルを採用した根拠として、利用者の質問に応じた動的な構造化や適当な階層化を可能とするように提言された、次世代ハイパーメディアの7つの課題のうちの1つである、リンク専制時代の終焉³に対処できそうなモデルであることや、人間の主観に基づく曖昧な情報処理を実用的に扱う一方で、統計理論などに基づく理論的な考察も可能と考えられること等がある。これらは、データベースの研究として今後取り組むべき課題 [15] として提起され項目⁴にもほぼ合致する。

次に、モデル構成上の利点として、統語的な体系化 [5][17] などにおける情報間のリンク属性を規定する場合に比べて労力が大幅に削減できること、視点という大上段から情報の関連性を比較的明示的にながめられること、視点と情報との関連の強さは規定するが、情報間の直接の関連性を縛るものではない (情報間の関連性や解釈の自由度が高い) こと等がある。つまり、動的な関連度 $p(x; \theta)$ の生成は、固い関係構造に縛られずに発見的な探索が行なえるよう、構造破壊と再構造化を積極的に行なっていると言える。

³文献 [7][8] では、“実際にネットワーク構造を持たないモデルを含めた、従来のリンクにとらわれない新しいハイパーメディアの概念が必要である。システムがナビゲーション空間を自動生成するような研究も行う必要がある。”と提言されている。このことは、ハイパーメディアに限らず、情報の多様な解釈や非一意分類性を扱う全てのモデルに当てはまる課題でもある。

⁴それらの項目は、応用分野に独立なマシン指向のこれまでのデータモデルの枠組を大きく変えて (人間の表現に合った) 人間指向の知識処理に適したモデルを考えること、環境に応じてアクセス構造やデータを能動的に再設計できる自己組織データベースを構築すること、人的な知識を共有化/再利用 (多様な解釈や人間の主観を活用) すること、言語 (論理的な記号処理) で普遍的な知識が表せるはずだと考えた AI の誤りを教訓にすること、ユーザの立場やシステムとの対話性をより重視すべきこと、などである。

3.3 α -分布族のファジィ平均化作用

関係分布モデルを具現化する α -分布族について述べる. パラメトリックなモデル全体の(微分幾何学的な)性質を研究する方法論として提唱された情報幾何学 [1] では, 統計的多様体の接続パラメータ $\alpha \in R$: 実数($\alpha \neq 1$)を持つ α -分布族 $S_\alpha = \{p_\alpha(x; \theta)\}$,

$$p_\alpha(x; \theta) \stackrel{\text{def}}{=} \left[\frac{1-\alpha}{2} \sum_{i=1}^{n+1} \theta^i F_i(x) \right]^{\frac{2}{1-\alpha}}, \quad (1)$$

が扱われている. 但し, $F_i(x) = [p_i(x)]^{\frac{1-\alpha}{2}}$, 各 $p_i(x) > 0$ は確率分布で, $\frac{\partial \log p_\alpha}{\partial \theta^i}$ が一次独立になるものとする. また, θ^{n+1} は規格化 $\sum_{x \in X} p_\alpha(x; \theta) = 1$ の為のものである. 特に, $\alpha \rightarrow 1$ ($\alpha = 1$) が指数型分布族 (Exp), $\alpha = -1$ が混合型分布族 (Mix) として知られている.

$$\text{Exp}: p(x; \theta) = \frac{\prod_{i=1}^n [p_i(x)]^{\theta^i}}{\sum_{x \in X} \left\{ \prod_{i=1}^n [p_i(x)]^{\theta^i} \right\}}, \quad (\theta^i \in R) \quad (2)$$

$$\text{Mix}: p(x; \theta) = \frac{\sum_{i=1}^n \theta^i p_i(x)}{\sum_{i=1}^n \theta^i}. \quad (\theta^i \geq 0, \sum_i \theta^i = 1) \quad (3)$$

式 (1) は規格化条件を緩めた, α -アファイン多様体 $\bar{S} = \{\bar{p}(x; \xi) \mid \bar{p}(x; \xi) > 0, \sum_{x \in X} \bar{p}(x; \xi) < \infty\}$,

$$\bar{p}(x; \xi) \stackrel{\text{def}}{=} \left[\frac{1}{n} \sum_{i=1}^n \xi^i F_i(x) \right]^{\frac{2}{1-\alpha}} = \left[\frac{1}{n} \sum_{i=1}^n u_i(x) \right]^{\frac{1}{\beta}}, \quad (4)$$

に拡張できる ($\alpha \neq 1 \Leftrightarrow \beta \neq 0$). 但し, $\beta \stackrel{\text{def}}{=} \frac{1-\alpha}{2}$, $\xi^i \stackrel{\text{def}}{=} n\beta\theta^i$, $u_i(x) \stackrel{\text{def}}{=} (\xi^i)^{\frac{1}{\beta}} p_i(x)$ とする.

式 (4) の中乗関数は, 最小値関数と最大値関数の間の値をとり, 平均値関数 [21] やファジィ平均演算子 [13] の関数クラスに含まれる. すなわち, α -分布族は $0 \leq u_i \leq 1$ の広い関数クラスにおいて, ファジィ平均化に規格化を施した作用を持つ. これによって, α -分布族で具現化した関係分布モデルは, シソーラスのないファジィ情報検索モデル [20] と見なすことができ, さらに指数型分布族は(統語的なデータ定義をあえてした形式的な対応の意味で), Croft の確率的情報検索モデル [4] の一種に相当すること [10] が知られている.

4 個人選好に基づく市場分析のモデル

本章では, マーケティングサイエンスにおける消費者需要と市場シェアモデルを概説し, 情報幾何学の α -分布族との密接な関連性を示す. ここで, 商品やサービスに対する消費者の効用は, 複数のニーズや欲求(属性⁵)の充足度を加重平均した形で得られ, 効用 \approx 選好は, 消費者の感情的反応と合理的思考の両方を反映する主観的評価とする [14].

4.1 消費者需要モデル(特性 \rightarrow 選好)

消費者需要モデルでは, 製品市場の特性 i (例えば, 食品栄養素などの客観的な特性) の消費量 q_i に対する効用 $U(q)$ が,

⁵属性は, ニーズや欲求充足に関する主観的判断を表す定性的な概念(客観的な特性から知覚符合化される)であるが, ある性質を持つブランド集合に属するかどうかの度合などで計られる. このような主観的な性質と客観的な特性を合わせて属性と呼ぶこともある.

Cobb-Douglas 型関数:

$$U(q) = \prod_{i=1}^n q_i^{\theta^i}, \quad (\beta \rightarrow 0) \quad (5)$$

CES 型関数:

$$U(q) = \left[\sum_i \theta^i q_i^\beta \right]^{\frac{1}{\beta}}, \quad (\beta < 1) \quad (6)$$

でそれぞれ規定される [11] [19]. ここで, θ^i ($\theta^i \geq 0$, $\sum_{i=1}^n \theta^i = 1$) は特性間のトレードオフを規定する個人の嗜好性を表すパラメータである.

4.2 市場シェアモデル (属性→選好→選択)

市場シェアモデルでは, 対象 (ブランド) $k \in \mathcal{X}$ が選択される確率としての確率的効用 $Prob(k) \approx A_k$ が,

MNL モデル:

$$Prob(k) = \frac{\exp(\sum_{i=1}^n \theta^i X_{ki})}{\sum_{j=1}^m \exp(\sum_{i=1}^n \theta^i X_{ji})}, \quad (7)$$

MCI モデル:

$$Prob(k) = \frac{\prod_{i=1}^n X_{ki}^{\theta^i}}{\sum_{j=1}^m \left\{ \prod_{i=1}^n X_{ji}^{\theta^i} \right\}}, \quad (8)$$

一般化 MCI モデル:

$$\text{Market - Share : } S_k = \frac{A_k}{\sum_{j=1}^m A_j}, \quad \text{Attraction : } A_k = \prod_{i=1}^n \Phi(X_{ki})^{\theta^i}, \quad (9)$$

でそれぞれ規定される [3] [11] [14]. ここで, X_{ki} は属性 (製品やコンセプトの性質よりも, 広告, 価格, 流通などのマーケティング変数で定義), θ^i は各個人 i の反応パラメータである. 特に, MNL モデルと MCI モデルは, 選択確率を選好 (確定的効用) と気まぐれ因子とで規定した時に, 後者の確率を第 1 種極値分布と第 2 種極値分布にした場合に相当する.

以上のモデルには, 有効性を支持する多くの経験的研究上の成果があるとともに, 対数線形回帰によるパラメータ推定が可能という利点がある [14]. 但し, どのモデルが妥当かについては, OR における平均値関数 [21] と同様に応用対象ごとの属性変数に依存するとともに, 事前に属性間の依存関係が明らかな場合はその構造を取り入れた方が良い.

4.3 α -分布族との関連性

まず, $\beta < 1$ の式 (4) は式 (6) の CES 型関数に他ならず, 式 (2) の指数型分布は分母の規格化成分を除いて式 (5) の Cobb-Douglas 型関数に一致する. 次に, 式 (8) の MCI モデルの $Prob(k)$ は式 (2) の指数型分布そのものであり, 式 (7) の MNL モデルの $\log Prob(k)$ は分母の規格化成分を除いて式 (3) の混合型分布に相当する. しかし, これら以外の一般化 MCI モデルと α -分布族の関数型は一致していない.

さらに、これらの個人選好モデルの特性や属性に対する効用 \approx 選好は、より主観性を重視した上で、関係分布モデルの視点に対する関連度に対応付けられる（表1）。まとめると、情報幾何学の指数型分布族と混合型分布族を2接点として、ファジィ平均演算子や個人選好モデルが、関数クラス（関数空間の土壌）において統一的に捉えられる（表2）。

一方、 α -分布族による関係分布モデルの妥当性（妥当な α の値）に関しては、上記のモデルと同様に応用対象（視点や情報集合の選び方）に依存すると考えられる。そればかりか、知的な探索を触発するような未知なる関係構造を生成する方が望ましく、そもそも事前知識による構造化（例えば、ロジットモデルにおける因果関係や高次相関の考慮 [11]）は極めて困難と考えられる。これらの点は具体的な事例を通じて検討していきたい。

RPM (α -分布族)	一般化 MCI モデル	CES 型関数
θ に対する関連度 (の対数) $\log p(x; \theta)$ 視点 i に対する情報 x の関連度 $p_i(x)$ 視点 i の重要度パラメータ θ^i	ブランドの吸引力/魅力 A_k 対象 j の k 番目の属性 X_{kj} 個人の反応パラメータ θ^i	効用 $U(q)$ 需要/数量 (製品の物理特性) q_i 個人の嗜好パラメータ θ^i

表 1: 関係分布モデル (RPM) と個人選好モデルとの各変数の対応

β	α	α -接続	平均 (論理演算)	消費者需要	市場シェア
$-\infty$	∞	-	最小 (AND)	CES 型	
0	1	指数型	幾何平均 (AND 的)	Cobb-Douglas 型	MCI
1/2	0	リーマン	平方根平均 (AND 的)	CES 型	
1	-1	混合型	算術平均 (中庸)		MNL
2	-3	-	遠心力平均 (OR 的)		
∞	$-\infty$	-	最大 (OR)		

表 2: α -分布族と平均演算子や個人選好モデルの関数クラスの対応 (空欄は対応なし)

謝辞. 本研究の一部は、(財)人工知能研究振興財団、及び、文部省科学研究費 10878045 の援助を受けている。

参考文献

- [1] 甘利 俊一, 長岡 浩司. 情報幾何の方法. 岩波講座 応用数学 [対象 12], 岩波書店, (1994).
- [2] Bush, V. As We May Think. *Reprint of The Atlantic Monthly*, pp. 101-108, (1945).
- [3] Cooper, L.G., Nakanishi, M. *Market-Share Analysis*. Klumer Academic Publishers, (1988).
- [4] Croft, W.B. Document Representation in Probabilistic Models of Information Retrieval. *Journal of the American Society for Information Science*, pp. 451-457, (1981).

- [5] 藤澤 浩道, 山崎 直子, 橋本 哲也. 概念ブラウザと個人情報環境. 情報処理学会研究会報告, IM7-6, pp. 41-48, (1992).
- [6] 藤原 讓. 情報知識学のフロンティア. 情報知識学会誌, Vol. 3, No. 1, pp. 3-13, (1993).
- [7] Halasz, F. Reflections on Note Cards: Seven Issues for the Next Generation of Hypermedia Systems. *Communication of the ACM*, Vol. 31, No. 7, pp. 836-852, (1988).
- [8] Halasz, F. Seven Issues: Revisited. *ACM HYPERTEXT '91*, Slides from Keynote Talk, (1991).
- [9] 林 幸雄. 個々人の嗜好を重視した魅力的な情報空間の構築に向けて-情報幾何学と平均演算子-. 第2回 曖昧な気持ちに挑むワークショップ, pp. 69-72, (1997).
- [10] 林 幸雄. データベースの数理モデルの視点パラメータの推定と関数空間におけるモデル拡張. 情報処理学会論文誌, Vol. 39, No. 6, pp. 1-11, (1998).
- [11] 片平 秀貴. マーケティングサイエンス. 東大出版会, (1987).
- [12] 加藤 秀俊. 整理学 - 忙しさからの解放 -, 中公新書 13, 中央公論社, (1963).
- [13] M. Mizumoto. Pictorial Representations of Fuzzy Connectives, Part I: Cases of t-norms, t-conorms and Averaging Operators. *Fuzzy Sets and Systems*, Vol. 31, pp. 217-242, (1989).
- [14] 中西 正雄 編著. 消費者行動のニューフロンティア. 誠文堂新光社, (1984).
- [15] 西尾 章治郎, 大須賀 節雄, 桐山 孝司, 西田 豊明, 藤原 讓, 松本 裕治, 溝口 理一郎. 座談会「知識処理応用とデータベース」. 人工知能学会誌, 特集「知識処理応用とデータベース」, Vol. 10, No.1, pp. 3-16, (1995).
- [16] 野中 郁次郎, 竹内 弘高 (梅本 勝博 訳). 知識創造企業. 東洋経済新報社, (1996).
- [17] 沢田 裕司, 大川 剛直, 馬場口 登. 観点を考慮した連想機構の実現. 情報処理学会論文誌, Vol. 35, No. 5, pp. 714-724, (1994).
- [18] 菅野 道夫. ファジィ理論の展開 -科学における主観性の回復-. サイエンス社, (1990).
- [19] 竹内 啓 編. 統計学辞典. 東洋経済新報社, (1989).
- [20] 馬野 元秀 編. ファジィデータベースと情報検索. 講座 ファジィ 9, 日刊工業新聞社, (1993).
- [21] 柳井 浩. 平均値の意味と構造 I, II. オペレーションズリサーチ, 3月号 pp. 155-160, 4月号 pp. 222-226, (1997).
- [22] 吉田 政幸. 分類学からの出発 -プラトンからコンピュータへ-, 中公新書 1148, 中央公論社, (1993).