北海道大学工学部研究報告
第78号（昭和51年）

Bulletin of the Faculty of Engineering,
Hokkaido University, No. 78 (1976)

多次元尺度構成の一手法

佐藤義治* 持田泰昭* 河口至商*

（昭和50年6月30日受理）

A Certain Technique of Multidimensional Scaling

Yoshiharu SATO Yasuaki MOCHIDA Michiaki KAWAGUCHI

(Received June 30, 1975)

Abstract

The purpose of multidimensional scaling is to assign a certain scaling value to each object which could not be measured previously with any numerical scale, but however showed the interrelation among them.

In this paper, we have employed the ordinary real value as the scale. But since each object can not always be arranged on a real line, we generally need multidimensional Euclidean space to represent each object as the point of its space, where the number of dimensions of this space is minimum for some criterion. And we have proposed an iterative method for determination of the minimum dimensional coordinates of each object. By this iterative method, the number of unknown variables with respect to be solved equation is reduced. Then we can apply this technique to practical problem without relying on a large memory of the computer as compared with other corresponding techniques.

1. 序論

多次元尺度構成とは，対象となる個々の事象，あるいは個体の，各相互間に，何らかの基準によって定められた類似性 (similarity)，又は距離等の量が与えられたとき，これらの量に基づき，個々の事象を何次元かのユークリッド空間（一般には，他の距離空間であってもよいし，またその方がより妥当な距離関数を用いることができる場合があるが，直観的に，ユークリッド空間の方が理解しやすいので，ここではユークリッド空間のみを対象とする）に配置することにより，その構造を解明すると共に，全体の関連を個々に与えられる尺度（座標）によって見ようとするものである。従ってこの座標をいかにして決定するのかということが，手法の中心をなすものである。このときユークリッド空間の次元はできるだけ小さい方がより望ましいのである。

従来，多次元尺度構成に関する研究は，R. N. Shepard[1),2)], J. B. Kruskal[3),4)], L. Guttman[5)] および W. S. Torgerson[6)] 等によって研究されているものであるが，これらにある考え方の1つは，データとして得られる類似性のランクオーダーのみを問題としているものである。この場合には各個々の相互関連性を全体的に握把するには有効であるが，個々の対象の関連を詳細に議論

* 工学部情報数理工学第一講座

するには不適当であるように思われる。他方 Torgerson の考え方は類似性（あるいは非類似性）として得られる量をそのまま用いているのではあるが，一般的に対象となる個数が多くなる場合，あるいはその配置を可能にするユークリッド空間の次元が大きくなる場合には，理論的にその解を得ることは可能であろうが，実際には現在の計算処理能力からすると不可能であるように思われる（非常に多くの時間と空間の使用が許されるならば，可能性はあると考えられる）。

従って本論文においては，Torgerson の立場に立って，より実用的な手法を提案すること，及びその応用可能性を示すことを目的とするものである。

2. 手法の概説

対象となる N 個の個体（対象とする個々のものを以下個体と呼ぶ）$O_1, O_2, \cdots, O_N$ に関して，すべての２つの個体間に，ある基準によって定められた類似度あるいは非類似度（類似度とは逆の関係にあるもの）が与えられているものとする。ここに類似度とは，互いに類似していると考えられるものは大きな値，似ていないと考えられるものは小さな値をもつものである。ここではこれら N 個の個体をユークリッド空間に配置しようとするのであるから，その関係をユークリッド距離として表現するために，類似度とは逆の非類似度を用いることにする。さらに各個体間の関係は対称性をもつものと仮定する。従って $N(N+1)/2$ 個の非類似度を $d_{ij}(i, j=1, 2, \cdots, N, d_{ij}=d_{ji})$ によって表わす。

この d_{ij} を用いて各個体 O_i をユークリッド空間に配置するのであるが，このとき何次元のユークリッド空間を必要とするのかは未知である。ここで述べる手法は，まず１次元に配置可能であると仮定して，１次元の配置を求め，それで不十分（以下に述べる基準に従う）であるならば，１次元で説明できる部分を除外して，その残りのものを直交する他の１次元（全体として２次元）で説明させ，さらにそれでも不十分ならば，以下順次次元を増加させ，最終的な空間配置を得ようとするものである。

各個体 O_i に与えられる１次元の座標を x_i とするとき，

$$W^2 = \sum_{i \neq j}^{N}\sum \left\{d_{ij}^2-(x_i-x_j)^2-L\right\}^2 \tag{1}$$

なる W^2 の値が最小となるように $x_i(i=1, \cdots, N)$ を求めるのである。ここに L は定数であり，非類似度 d_{ij} は真にユークリッド距離の関係を満たしているとは限らないため，この L によってそれを修正しようとするものである。さらにこの W^2 から得られる誤差分散

$$J^2 = W^2/N(N-1)$$

と，d_{ij} の分散 V^2 との比 J^2/V^2 の値によって配置の良さを表わすものとする。すなわち

$$P = \{1-J^2/V^2\}^2$$

の値で，得られたユークリッド空間の配置によって d_{ij} を何パーセント説明できるかという基準を与えることができる。

式(1)で与えられる W^2 の値を最小にする１次元の配置（座標）$x_i(i=1, \cdots, N)$ を求めるため，W^2 を x_k で偏微分して０とおいた連立方程式

$$\sum_{i=1}^{N}\left\{d_{ik}^2-(x_i-x_k)^2\right\}(x_i-x_k) = 0 \qquad (k=1, 2, \cdots, N) \tag{2}$$

を解けばよいのであるが，この式をそのまま解くことは困難であるため，つぎのような逐次近似により各々の x_i を求める。

各 x_i に対するある初期値 x_i^0 を与え，

$$x_i = x_i^0(1+\varDelta x_i) \qquad (i=1,2,\cdots,N) \tag{3}$$

なる関係によって x_i が得られるものとする。

このときの $\varDelta x_i$ を式(3)を式(2)の連立方程式に代入し，$(\varDelta x_i)^2$ 以上の項を無視して，$\varDelta x_i$ の連立線形方程式

$$\left\{-\sum_{i\neq k}^{N} u_{ki}\right\}x_k^0\varDelta x_k+\sum_{i\neq k}^{N} u_{ki}x_i^0\varDelta x_i = \sum_{i\neq k}^{N}(v_{ki}-w_{ki}) \qquad (k=1,2,\cdots,N) \tag{4}$$

により求める。ここに

$$u_{ki} = d_{ik}^2-3(x_i^0-x_k^0)^2$$
$$v_{ik} = d_{ik}^2(x_i^0-x_k^0)$$
$$w_{ik} = (x_i^0-x_k^0)^3$$

とおく。このとき，連立方程式(4)の階数が一般に $N-1$ となることから，(4)のみの関係では N 個の $\varDelta x_i$ が一意に定まらない。従って，各個体 O_i を1次元に配置するとき，その原点は任意であるから，$\sum_{i=1}^N x_i/N=0$ なる条件を付けても一般性は失なわれない。この条件と(3)の関係から

$$\sum_{i=1}^{N} x_i^0\varDelta x_i = 0$$

なる関係式が得られ，この式と(4)で与えられている連立方程式の任意の$(N-1)$個の式を組み合わせることにより各 $\varDelta x_i$ を求める。

初期値 x_i^0 の設定は以下の通り行なう*。その基本的な考え方は最初に d_{ij}^2 と $(x_i-x_j)^2$ との相関係数が最大となるように x_i を決定することである。すなわち，

$$R = \frac{\sum\sum_{i\neq j}^{N} d_{ij}^2(x_i-x_j)^2}{\frac{1}{N}\sum_{i=1}^{N}x_i^2-\bar{x}^2}$$

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N}x_i$$

なる R の値を最大にする x_i を求める。ただしここでは条件

$$\bar{x} = 0$$
$$\frac{1}{N}\sum_{i=1}^{N}x_i^2-\bar{x}^2 = \text{const}.$$

を満たすものとする。(一般性は失なわれない) 従ってこの条件を満たし R を最大にする x_i はつぎのような固有方程式における最大の固有値に対応する固有ベクトルとして求められる。

$$\left\{-\sum_{\substack{i=1\\ i\neq k}}^{N} p_{ki}\right\}x_k+\sum_{\substack{i=1\\ i\neq k}}^{N} p_{ki}x_i = \lambda x_k \tag{5}$$

ただし

$$p_{ki} = -(d_{ki}^2+d_{ik}^2) = -2d_{ki}^2$$

とする。ここで得られる x_i に対して，評価関数(1)が適用できるように x_i のスケールを以下の式によって修正する。それは，

$$Q^2 = \sum\sum_{i\neq j}^{N}\left\{d_{ij}^2-K(x_i-x_j)^2-L\right\}^2$$

---

* 林知己夫氏の「e_{ij} 型の数量化」に相当するものである。

が最小となるような K, L を求める。すなわち K, L は連立方程式

$$\begin{cases} \sum\sum_{i \neq j}^{N} (x_i - x_j)^2 K + N(N-1) L = \sum\sum_{i \neq j}^{N} d_{ij}^2 \\ \sum\sum_{i \neq j}^{N} (x_i - x_j)^4 K + \sum\sum_{i \neq j}^{N} (x_i - x_j)^2 L = \sum\sum_{i \neq j}^{N} d_{ij}^2 (x_i - x_j)^2 \end{cases} \tag{6}$$

の解として得られる。この K と (5) より求められる固有ベクトル x_i を用いて初期値 x_i^0 を

$$x_i^0 = \sqrt{K} x_i$$

によって決定する。

この初期値を用いて上述の連立方程式 (4) より $\varDelta x_i$ を求め $x_i = x_i^0(1 + \varDelta x_i)$ として得られる x_i により J^2, あるいは P の値を計算し，それが設定された基準 (例えば，$J^2 \leqq 0.05$, $P \geqq 0.9$) を満たすならば，そのときの x_i を求める座標とする。もしこの基準が満たされないならば，この x_i を用いさらに (6) より K, L を求め，$\sqrt{K} x_i$ を初期値として $\varDelta x_i$ を計算し以下同様の操作をくり返す。このとき，くり返しは，J^2, 又は P の値が基準を満たすか又は $\varDelta x_i$ が 0 に収束* した場合に中止する。もし $\varDelta x_i$ が 0 に収束したとき J^2, P の値が基準を満たさないならば，N 個の個体 O_i は 1 次元に配置することはできないことになる。従ってつぎに d_{ij}^2 から 1 次元で説明できる部分を除いたもの，すなわち

$$d_{ij}^{2\prime} = d_{ij}^2 - (x_i - x_j)^2$$

をいま求めた 1 次元とは直交する他の 1 次元 (全体として 2 次元) で説明しようとするところが，本論文で述べる手法の中心をなすものである。計算はこの $d_{ij}^{2\prime}$ を上述の d_{ij}^2 としていままでと同様の操作を行なうことになる。ここで求まる座標の値を $y_i (i = 1, \cdots, N)$ とすると，結局 d_{ij}^2 を

$$d_{ij}^2 - (x_i - x_j)^2 - (y_i - y_j)^2 - L$$

なる値が最小となるような 2 次元 (x_i, y_i) の距離関係で説明しようとするものであり，これはユークリッド空間の距離関係としては極く自然なものであろう。もし 2 次元でも，J^2, P の値が満足できるものでなければ，さらに

$$d_{ij}^{2\prime\prime} = d_{ij}^{2\prime} - (x_i - x_j)^2 - (y_i - y_j)^2$$

とおいて全体として 3 次元の空間への配置を求め，以下順次次元を増加させて行くことができる。

3. 応 用 例

G. Ekman[8] の行なった実験のデータを用いて 14 色の色の空間配置を上述の手法によって求める。Ekman の実験はつぎのような方法で行ったものである。

「被験者は，スクリーンから 2.5 m 離れたところにいて，スクリーン上の 1 cm 離れた直径 1.5 cm の 2 つの円形の窓に示される波長が 434 mμ から 674 mμ までの 14 色のうち 2 つの色を見て，その 2 つの色の似ている程度を判断し，全く似ていないと判断したものには 0，全く同じ色と判断したものには 4 というように 0 から 4 まで 5 段階の score を与える。」

この score は 2 つの色の類似度をあらわすものであり，ここで述べた手法を適用するためにこれを逆 (大きさの程度を逆にする) にし，それを [0, 1] 区間に変換して得られたものが次表である。この非類似度 (距離) を用いて，ここで述べた手法により得られる 14 色の色の空間配置を示したものが次図である。

* 実際の計算では，必要な精度に応じて $|\varDelta x_i| \leqq 10^{-3}$ 等の適当な値を決定する。

表　dissimilarity matrix

波長 mμ	434	445	465	472	490	504	537	555	584	600	610	628	651	674
434		.14	.58	.58	.82	.94	.93	.96	.98	.93	.91	.88	.87	.84
445	.14		.50	.56	.78	.91	.93	.93	.98	.96	.93	.89	.87	.86
465	.58	.50		.19	.53	.83	.90	.92	.98	.99	.98	.99	.95	.97
472	.58	.56	.19		.46	.75	.90	.91	.98	.99	1.00	.99	.98	.96
490	.82	.78	.53	.46		.39	.69	.74	.93	.98	.98	.99	.98	1.00
504	.94	.91	.83	.75	.39		.38	.55	.86	.92	.98	.98	.98	.99
537	.93	.93	.90	.90	.69	.38		.27	.78	.86	.95	.98	.98	1.00
555	.96	.93	.92	.91	.74	.55	.27		.67	.81	.96	.97	.98	.99
584	.98	.98	.98	.98	.93	.86	.78	.67		.42	.63	.73	.80	.77
600	.93	.96	.99	.99	.98	.92	.86	.81	.42		.26	.50	.59	.72
610	.91	.93	.98	1.00	.98	.98	.95	.96	.63	.26		.24	.38	.45
628	.88	.89	.99	.99	.99	.98	.98	.97	.73	.50	.24		.15	.32
651	.87	.87	.95	.98	.98	.98	.98	.98	.80	.59	.38	.15		.24
674	.84	.86	.97	.96	1.00	.99	1.00	.98	.77	.72	.45	.32	.24	

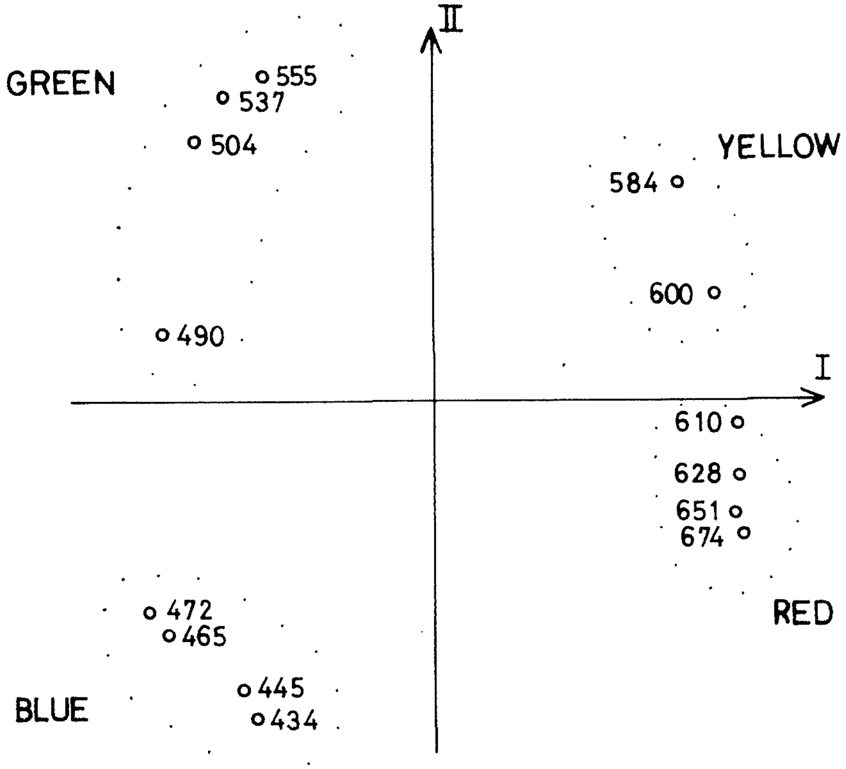

図　色の類似性による空間配置

この結果，視覚にもとづいた色の関連性を示す空間配置が得られたわけである。全体が配置される空間の次元が2次元で十分であること，さらに3原色の混合から得られる色度図の色の配置とほぼ一致することは興味ある結果である。

4. 結　論

多次元（多変量）データの構造を，より低次元の空間に要約して，それをより単純な構造としてとらえようとすることは多変量解析の主な目的の1つである。しかし，そのデータが，必ずしも連続量，すなわち普通の多変量統計学の対象となる量であるとは限らない。例えば本論文の例にもあるように，人の感覚（好きか嫌いかの程度，どれ程似ているか，等）は，ある discrete な変量で表現せざるを得ないこともある。このようなデータに対して，個々の相互間の関係から，各個体を配置可能な空間を決定することが，本論文で述べた多次元尺度構成（multidimensional scaling）の目的である。本論文の基本的な考え方は，W. S. Torgerson に負うところが大きいが，それとの根本的な相異は，その手法の中で次元を増加させていく際に，Torgerson は，例えば2次元への展開をする場合には $(x_i, y_i)\ (i=1, \cdots, N)$ なる $2N$ 個の値を一度に決定しようとするものであるが，本論文では，これを x_i を求めてから，さらに必要であれば，y_i を求める。すなわち，一次元ずつ逐次的に増加させていくものである。従って，本論文の手法の利点は，未知数の数が各次元で N 個ですむことである。Torgerson の手法では，k 次元の展開を考える場合には未知数は kN 個となり，N が大きいか又は k が大きいときには実際には計算が不可能になることがある。

参 考 文 献

1) Shepard R. N.: Psychometrika, **27** (1962), pp. 125–140.
2) Shepard R. N.: Psychometrika, **27** (1962), pp. 219–246.
3) Kruskal J. B.: Psychometrika, **29** (1964), pp. 1–26.
4) Kruskal J. B.: Psychometrika, **29** (1964), pp. 115–129.
5) Guttman L.: Psychometrika, **33** (1968), pp. 469–506.
6) Torgerson W. S.: Psychometrika, **30** (1965), pp. 379–393.
7) 林知己夫，樋口伊佐夫，駒沢　勉：情報処理と統計数理，(1970)，産業図書.
8) Ekman G.: The Jour. of Psychology, **38** (1954), p. 467–474.