

多次元尺度法

点 i と点 j の p 次元空間における座標 $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ と $\mathbf{x}_j = (x_{j1}, \dots, x_{jp})$ が与えられるとその距離 d_{ij} を算出することができる。距離が r-metric で与えられるとき次のようになる。

$$d_{ij} = \left(\sum_r |x_{ir} - x_{jr}|^r \right)^{1/r}$$

逆に、 n 個の点の間の距離が与えられたとき、それらの距離に対応する点の布置を求めるのが多次元尺度法である。プログラム PMDSKruskal.dpr は、非類似度データ δ_{ij} が与えられたとき、非類似度と単調関係にある r-metric による距離を与える布置を求めるものである。

このプログラムを実行する¹と次のフォームが表示される。

図1 起動時のフォーム

行と列は「追加」ボタンのクリックでデータの設定に必要なだけ増やすことができる。「追加」ボタンのクリックで、アクティブなセルの下に空白の行が挿入され、行に応じて列も挿入される。セルはクリックによりアクティブになる。「削除」ボタンをクリックすると、

¹ PMDSKruskal.dpr は Delphi 6 から開いて実行するものであるが、コンパイル済みの実行形式のファイル PMDSKruskal.exe はダブルクリックにより直接実行することができる。

アクティブなセルを含む行が削除される。行が削除されると対応する列も削除される。必要なだけの行および列を用意してデータを図2のように設定する。

Figure 2 shows a software window titled "Form1" with a data table and control buttons. The table has columns for labels and objects, and rows for objects. Callouts point to the "r" value field and the "No. of Dims" field.

	ラベル	対象1	対象2	対象3	対象4	対象5
ラベル		A	B	C	D	E
対象1	A	0				
対象2	B	1.4				
対象3	C	2	1.4			
対象4	D	1.4	2	1.4		
対象5	E	1	1	1	1	0

Buttons on the right: 閉じる, 追加(行), 削除(行), 保存(CSV), 読み込(CSV), 計算, 散布図, 布置.

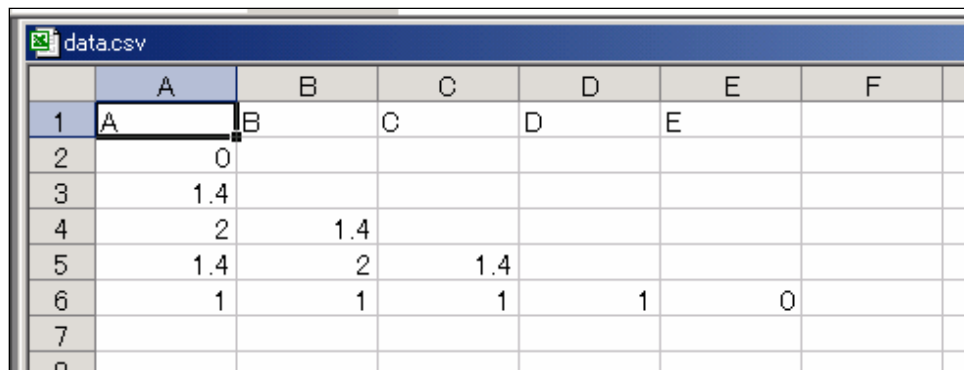
Callouts:

- r-metric の r の値 (points to r = 2.0)
- 次元数 (points to No. of Dims = 2)

図2 データの設定

ラベルの欄の行と列に対象のラベル（文字列）を設定する。このラベルの先頭2バイト分の文字列が布置の描画において用いられる。距離（非類似度）データは対称な位置のいずれかに設定されておればよい。すなわち、図2ではデータはすべて下三角行列の位置に設定されているが、上三角行列の位置に設定されていてもよい。もちろん、すべてのセルにデータを設定してもよい。対角線の位置に設定される自分自身との非類似度データは、設定しても無視される。

設定したデータは「保存」ボタンのクリックで保存することができる。データはCSV形式で保存されるのでExcelなどで開くことができる。図3は図2のデータをCSV形式で保存したものをExcelで開いたものである。



	A	B	C	D	E	F
1	A	B	C	D	E	
2	0					
3	1.4					
4	2	1.4				
5	1.4	2	1.4			
6	1	1	1	1	0	
7						

図3 Excelでのデータ

逆に Excel において図3の形式で用意されたデータは、CSV形式で保存すると図2の「読み込」ボタンのクリックで図2のように読み込むことができる。

図2のデータは、図4の布置に対してユークリッド距離($r=2$)を算出したものである。

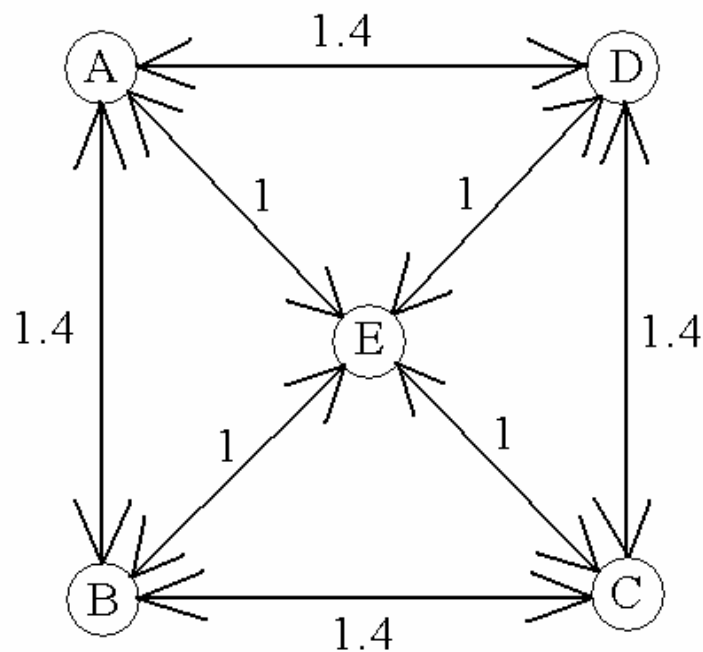


図4 データ(図2)の算出された布置

データの設定後、r-metricの r の値と次元数の値を適当な値に設定した後、「計算」ボタンをクリックして計算を始める。 r の値は1以上($r \geq 1$)である。 $r=1$ のときは city block distance

$$d_{ij} = \sum_t |x_{it} - x_{jt}|$$

を表す。 $r = 2$ のときはユークリッド距離

$$d_{ij} = \sum_t (x_{it} - x_{jt})^2$$

である。 r が大きくなると sup metric

$$d_{ij} = \max_t |x_{it} - x_{jt}|$$

に近づく。 r を大きくし過ぎると実行時に計算エラーが出易くなる。 $r = 50$ ぐらいの値で十分 sup metric とみなせる。 r の値の変化によって距離関数の様子がどのように変わるかはプログラム PUnitCircle.dpr によって視覚的に確かめることができる。

「計算」ボタンをクリックすると、計算結果などを書き出すテキストファイルの名前の設定を求めるダイアログボックスが表示される（図 5）。

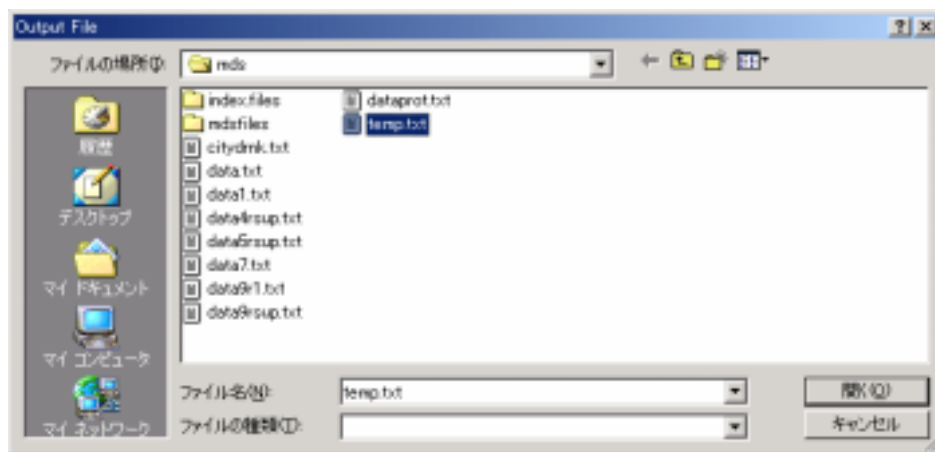


図 5 出力用ファイル名の設定

名前の設定後、図 5 の「開く」ボタンをクリックすると計算が始まる。図 2 のデータの場合は瞬時に計算が終了するが、計算時間が長いときは計算の途中結果が図 6 のように表示されるのを見ることが出来る。

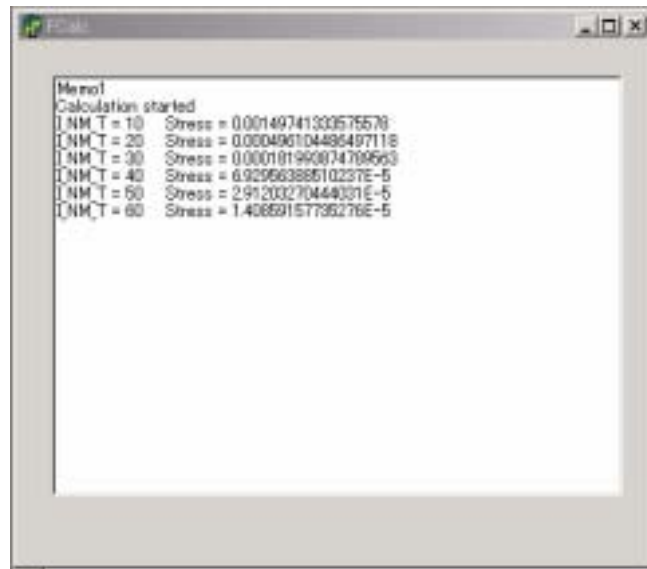


図6 計算の途中結果の表示例（図2のデータとは異なるものを用いた例）

図6における Stress の値は次式によるものである。

$$Stress = \sqrt{\frac{\sum (\hat{d}_{ij} - d_{ij})^2}{\sum d_{ij}^2}}$$

Stress の値が小さいほどモデルから計算された距離 d_{ij} はデータ δ_{ij} に順序関係がよく対応

していることになる。 \hat{d}_{ij} は、 d_{ij} を δ_{ij} との単調関係が保たれるように変換（isotonic transformation, nonmetric transformation）したものである。

計算が終了すると図7のフォームになる。

	ラベル	対象1	対象2	対象3	対象4	対象5
ラベル		A	B	C	D	E
対象1	A	0				
対象2	B	1.4				
対象3	C	2	1.4			
対象4	D	1.4	2	1.4		
対象5	E	1	1	1	1	0

Enabled になっている

計算

散布図

布置

r = 20 No. of Dims = 2

図 7 計算終了時のフォーム

計算終了後には、「散布図」ボタンと「布置」ボタンがイネーブルになっている。「散布図」ボタンをクリックすると図 8 の画面になる。

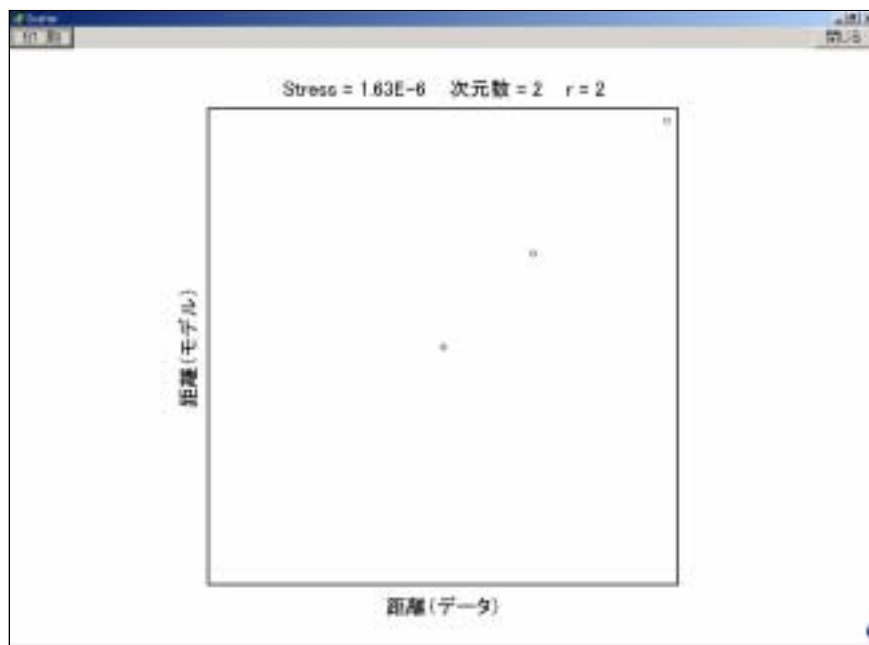


図 8 散布図

図 8 の散布図は、横軸にデータ δ_{ij} 、縦軸に求めた座標値（布置）から算出される距離 d_{ij} を

として、点 (δ_{ij}, d_{ij}) をプロットしたものである。距離 d_{ij} とデータ δ_{ij} との単調関係を視覚的に判断することができる。図の上部には Stress の値が表示されている。図2のデータは実際の点の布置(図4)から与えられているので、Stress の値は実質的には0であり、散布図における距離 d_{ij} とデータ δ_{ij} との単調関係もきれいに現れている。非類似度として集められた実際の生のデータでは、このようなきれいな関係はなかなか得られない。図8の画面における「印刷」ボタンをクリックすると、散布図がプリンタに出力される。

「布置」ボタンをクリックすると図9の画面になる。

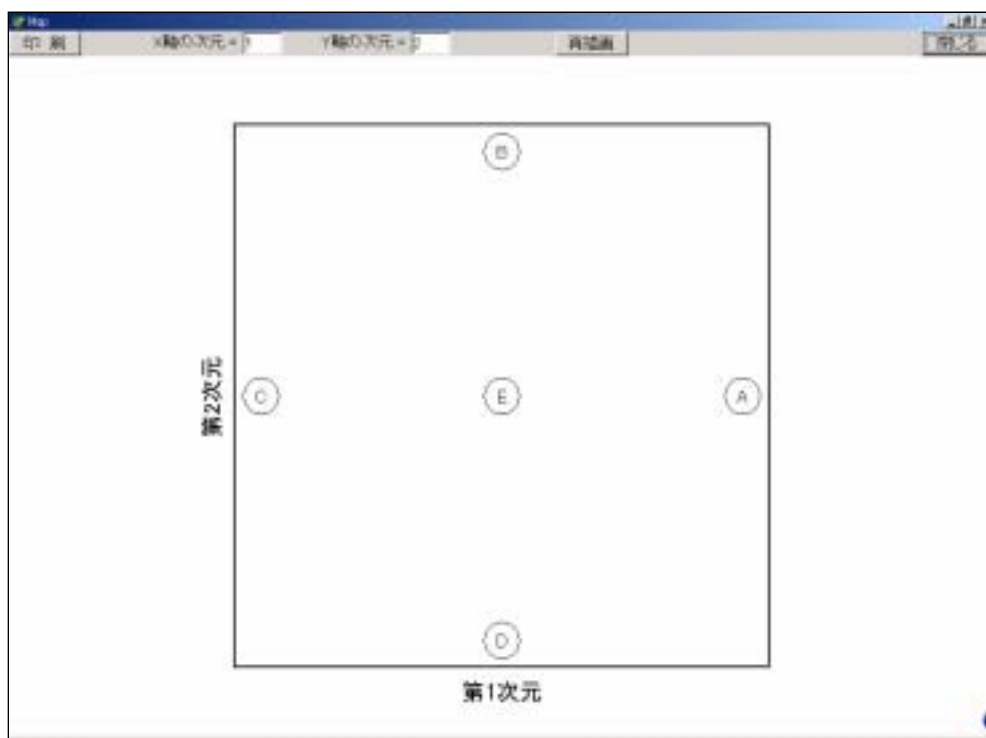


図9 解の布置

図9の解の布置は、対象 i の座標値 (x_{i1}, x_{i2}) をプロットしたものである。各対象の位置がそれぞれのラベルとして設定された文字列(図2)の先頭2バイトまでの文字列で表されている。図2のデータの場合は、ラベルが1バイトの文字列であるので、設定されている文字がそのまま表示されている。図9の解の布置は、データの値が算出された図4の布置によく対応している。画面上部にはX軸の次元とY軸の次元が設定できるようになっている。各次元の設定後、「再描画」ボタンをクリックすると設定した次元での布置が描画される(図10)。

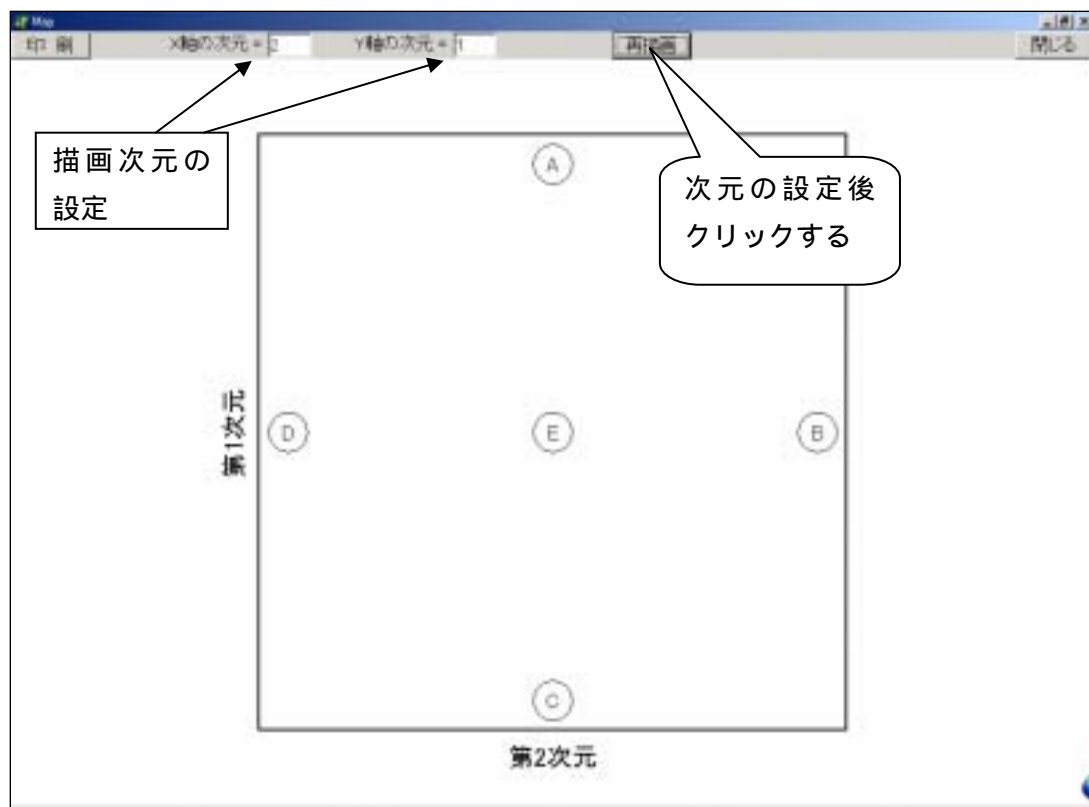


図 1 0 描画次元の設定と再描画

図 1 0 の画面上部の「印刷」ボタンをクリックすると、画面に描かれている布置がプリンタに出力される。

図 1 0 の画面上部の「閉じる」ボタンをクリックすると、図 7 の計算終了時の画面に戻る。この画面の「閉じる」ボタンをクリックすると、プログラムの実行が終了する。計算の終了後、図 5 で設定した名前のテキストファイルを開くと、リスト 1 のように座標値などが書き出されていることがわかる。

リスト 1 計算結果の出力ファイルの一部

座標値 =		
対象(A)	ラベル -> A	
	0.50000	-0.00000
対象(B)	ラベル -> B	
	0.00000	0.50000
対象(C)	ラベル -> C	
	-0.50000	-0.00000
対象(D)	ラベル -> D	
	0.00000	-0.50000
対象(E)	ラベル -> E	
	-0.00000	0.00000

参 考 文 献

- (1) Kruskal,J.B. (1964a) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, 1-27.
- (2) Kruskal,J.B. (1964b) Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29, 115-129.
- (3) Barlow,R.E., Bartholomew,D.J., Bremner,J.M. and Brunk,M.D. (1972) *Statistical inference under order restrictions: The theory and application of isotonic regression*. John Wiley & Sons Ltd.