展開法

Penalty 基準

展開法とは、対象の 2 組のグループ間の距離データから対象の布置を求めるものである。距離データが対象間の非類似度あるいは類似度データとして与えられたときは、それらの非類似度あるいは類似度データの大小関係が対象間の距離の大小関係と何らかの形で対応するように対象の布置を求める。例えば、色と象徴語との適合度評定データが表１のように与えられているとする[1]。

表１. 色と象徴語の適合度評定データ（女性)。「非常によく合う」から「全く合わない」までの 6 つの評定語に 6 から 1 までの整数値を順に対応させ、女性被験者 9 人の中央値を求めたもの。

色	象徴語									
	永遠	平静	郷愁	夢	家庭	幸福	愛	嫉妬	怒り	孤独
赤	4	1	3	4	5	5	6	6	6	2
黄	5	2	4	5	5	6	3	5	4	2
緑	4	4	4	4	4	4	3	2	3	3
シアン	5	6	4	4	4	3	3	3	2	5
青	5	5	4	4	3	3	2	4	4	6
マジェンタ	3	3	4	3	2	3	3	5	4	4

表１の適合度（類似度）データを Penalty 基準を用いた展開法で分析して得られた布置の例が図１である。

[1] 岡本安晴、1996、「色彩の心理学的意味空間の研究」、コスメトロジー研究報告、Vol. 4、136-144.

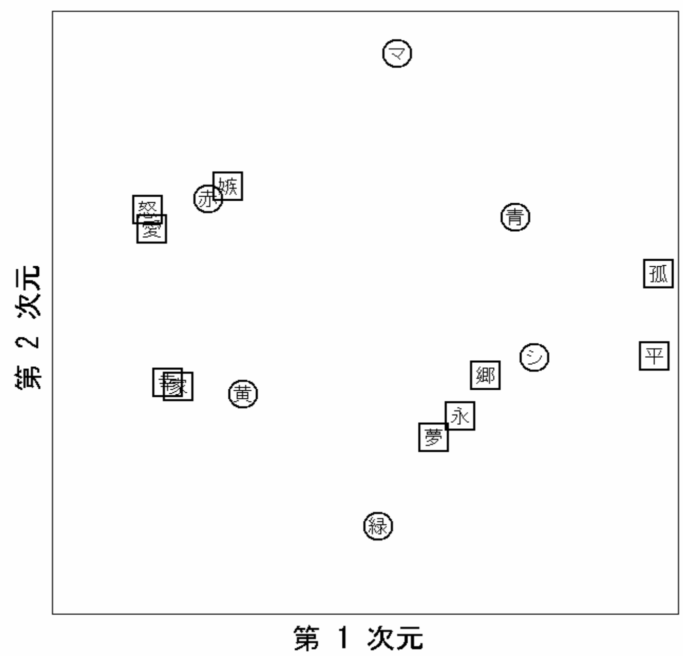

図１　表１のデータの分析例

図１の布置を見ると、色が赤、黄、緑、シアン、青、マジェンタの順に円環状に並んでおり、それぞれの色の近くにそれらの色と適合度が高いと評定された象徴語が配置されていることがわかる。

図１の分析はプログラム PUnfoldP.dpr によるものである。このプログラムは penalty 基準により展開法を行うものである。まず、その考え方について説明を行い、その後 PUnfoldP.dpr の使い方について解説する。

Penalty 基準による展開法[2]

M 個の対象 X_1、･･･、X_M の座標、$\left(x_{i,1} \quad \cdots \quad x_{i,p}\right)$、$i = 1, \cdots M$ 、および N 個の対象 Y_1、･･･、Y_N の座標、$\left(y_{j,1} \quad \cdots \quad y_{j,p}\right)$、$j = 1, \cdots N$ が与えられると、X_i と Y_j の距離 $d_{i,j}$ は次式で与えられる。

2　岡本安晴、1994、「色彩空間の展開法的分析」、日本心理学会第 58 回大会発表論文集、p.537.
　岡本安晴、1996、「色彩の心理学的意味空間の研究」、コスメトロジー研究報告、Vol. 4、136-144.

$$d_{i,j} = \sqrt{\sum_{t=1}^{p} \left(x_{it} - y_{jt} \right)^2}$$

逆に、$d_{i,j}$ が与えられたときに $d_{i,j}$ を与える座標 $\left(x_{i,1} \quad \cdots \quad x_{i,p} \right)$ および $\left(y_{j,1} \quad \cdots \quad y_{j,p} \right)$ を求めることが考えられる。この２組のグループ間の距離から布置を求める方法が展開法である。対象間の距離は直接与えられないが、距離と何らかの単調関係にあると考えられるデータ（類似度、非類似度など）が与えられたときは、そのデータと出来るだけ単調関係が成り立つ距離を与える布置を求める。対象 X_i と Y_j の類似度データ $s_{i,j}$ が与えられたときは距離 $d_{i,j}$ との間に次の単調関係が成り立つようにする。

$$s_{i1,j1} \geq s_{i2,j2} \quad \Leftrightarrow \quad d_{i1,j1} \leq d_{i2,j2} \qquad （1）$$

すなわち、X_{i1} と Y_{j1} の類似度 $s_{i1,j1}$ が X_{i2} と Y_{j2} の類似度 $s_{i2,j2}$ より高いとき、距離 $d_{i1,j1}$ は $d_{i2,j2}$ より小さくなるように対象の布置を求める。データが非類似度 $\delta_{i,j}$ のときは

$$\delta_{i1,j1} \geq \delta_{i2,j2} \quad \Leftrightarrow \quad d_{i1,j1} \geq d_{i2,j2}$$

が成り立つようにする。非類似度データ $\delta_{i,j}$ は

$$s_{i,j} = \max\left(\delta_{i,j}\right) - \delta_{i,j} + 1$$

あるいは

$$s_{i,j} = \delta_{i,j}^{-1}$$

などと簡単に類似度データに変換できる。

（１）式の関係が成り立つように推定された距離 $d_{i,j}$ から座標値を求める方法は nonmetric な方法として知られている。この nonmetric な方法は退化解の得られる危険があることが知られている。これは、次式

$$s_{i1,j1} \geq s_{i2,j2} \quad \Leftrightarrow \quad d_{i1,j1} = d_{i2,j2}$$

のように距離に等号関係が成り立っているときも、（１）式が成り立つからである。距離に等号関係が成り立つということは、例えば、複数の対象 X_i が１つの対象 Y_j を中心として円

環状に並んだり、あるいは複数の対象 X_i または複数の対象 Y_j が１点に重なったりする解が得られたりするということである。このような解が得られることを避けるために、 $s_{i1,j1}$ と $s_{i2,j2}$ の順序関係が $d_{i1,j1}$ と $d_{i2,j2}$ の順序関係に（１）式のような形で反映されるだけではなく、 $s_{i1,j1}$ と $s_{i2,j2}$ の間に差があれば $d_{i1,j1}$ と $d_{i2,j2}$ の間にも差があるようにモデルの適合度の基準を設定することを考える。ここでは、適合度の逆、すなわち penalty 基準 P を以下のように設定する。

$$P = \sum_{i1=1}^{M}\sum_{j1=1}^{N}\sum_{i2=1}^{M}\sum_{j2=1}^{N}\phi\left\{\left(s_{i1,j1}-s_{i2,j2}\right)\cdot\left(d_{i1,j1}^{2}-d_{i2,j2}^{2}\right)\Big/\sum_{i,j}d_{i,j}^{2}\right\} \qquad （２）$$

ここで、

$$\phi(z)=\frac{1}{1+e^{-c\cdot z}}\text{、}\ c>0$$

で、 c は適当な正定数である。

Penalty を与える関数 $\phi(z)$ は z が０より十分に小さい負の値のときは０に近い値をとり、０より大きい正の値のときは１に近い値をとる（図２）。

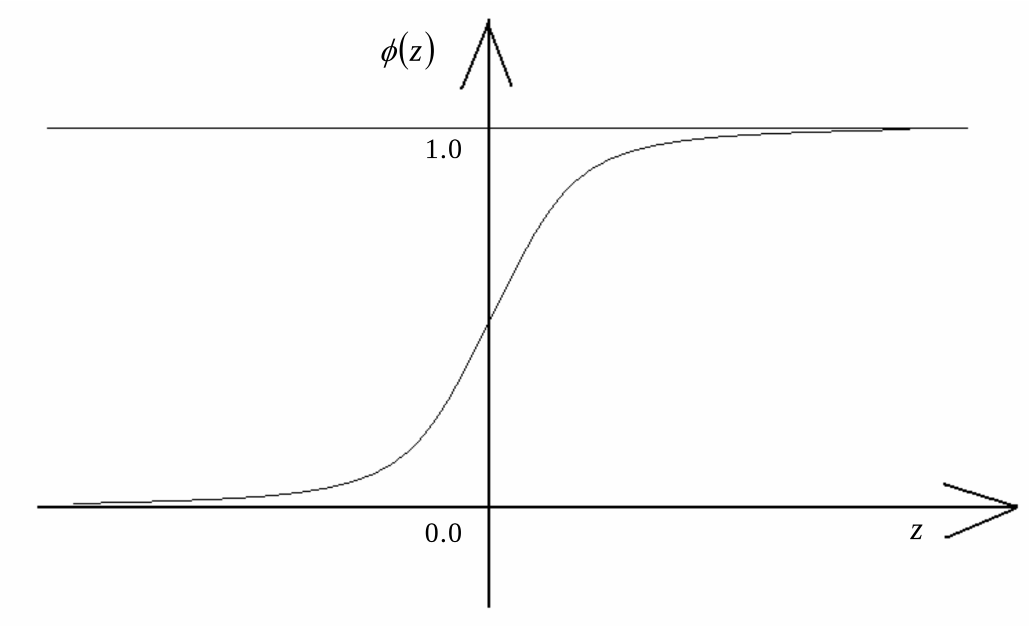

図２　Penalty 関数 $\phi(z)$

c の値は $z=0$ における関数の傾きを決めるもので、 c の値が大きいほど $z=0$ の近傍で関

数値は０から１に急激に変化する。

（２）式における関数$\phi(z)$の独立変数zの値を与える項

$$\frac{\left(s_{i1,j1}-s_{i2,j2}\right)\cdot\left(d_{i1,j1}^{2}-d_{i2,j2}^{2}\right)}{\sum_{i,j}d_{i,j}^{2}} \qquad (3)$$

の値は、$d_{i1,j1}$と$d_{i2,j2}$の順序関係が$s_{i1,j1}$と$s_{i2,j2}$の順序関係に対応しているとき、すなわち次式

$$s_{i1,j1} > s_{i2,j2} \quad \Leftrightarrow \quad d_{i1,j1} < d_{i2,j2} \qquad (4\mathrm{a})$$

あるいは

$$s_{i1,j1} < s_{i2,j2} \quad \Leftrightarrow \quad d_{i1,j1} > d_{i2,j2} \qquad (4\mathrm{b})$$

が成り立つとき、負の値となる。式（３）において、$d_{i,j}$ではなくその２乗$d_{i,j}^{2}$を用いているのは、計算が簡単になるからである。（４a）式あるいは（４b）式が成り立っているとき、$s_{i1,j1}$と$s_{i2,j2}$の間の差が大きいときに$d_{i1,j1}$と$d_{i2,j2}$の間の差も大きいならば、（３）式の値は絶対値の大きい負の値となり関数$\phi(z)$の値は０に近い値となる。逆に、（４a）式および（４b）式の両式とも成り立たないとき、$s_{i1,j1}$と$s_{i2,j2}$の間の差が大きいときに$d_{i1,j1}$と$d_{i2,j2}$の間の差も大きいならば（３）式の値は大きい正の値となり関数$\phi(z)$の値は１に近い値となる。以上のことから、（２）式の関数 P は、$d_{i1,j1}$と$d_{i2,j2}$の順序関係と$s_{i1,j1}$と$s_{i2,j2}$の順序関係の対応の悪さを表していると考えられる。したがって、P の値が出来るだけ小さくなるような距離$d_{i,j}$を与える座標$\left(x_{i,1} \quad \cdots \quad x_{i,p}\right)$および$\left(y_{j,1} \quad \cdots \quad y_{j,p}\right)$を求める。P の値を最小にする座標値を求めるための極値探索法としてプログラム PUnfoldP.dpr では Davidon-Fletcher-Powell（DFP）法[3]を用いている。DFP 法では１次の偏導関数が必要である。必要な偏導関数を以下に示す。

まず、$x_{i,h}$による偏導関数は次式[4]で与えられる。

---

[3] Davidon-Fletcher-Powell 法についての解説がホームページ
http://www.ikuta.jwu.ac.jp/~yokamoto/openwww/num/opt2/
にある。

[4] 添字の範囲を適当に設定することにより、項目の和の重複に対応するものとする。

$$\frac{\partial P}{\partial x_{i,h}} = \sum_{j1,i2,j2} \phi' \left\{ \left(s_{i,j1} - s_{i2,j2}\right)\left(d_{i,j1}^2 - d_{i2,j2}^2\right) \Big/ \sum_{a,b} d_{a,b}^2 \right\}$$

$$\times \frac{\partial}{\partial x_{i,h}} \left\{ \left(s_{i,j1} - s_{i2,j2}\right)\left(d_{i,j1}^2 - d_{i2,j2}^2\right) \Big/ \sum_{a,b} d_{a,b}^2 \right\}$$

$$+ \sum_{i1,j1,j2} \phi' \left\{ \left(s_{i1,j1} - s_{i,j2}\right)\left(d_{i1,j1}^2 - d_{i,j2}^2\right) \Big/ \sum_{a,b} d_{a,b}^2 \right\}$$

$$\times \frac{\partial}{\partial x_{i,h}} \left\{ \left(s_{i1,j1} - s_{i,j2}\right)\left(d_{i1,j1}^2 - d_{i,j2}^2\right) \Big/ \sum_{a,b} d_{a,b}^2 \right\}$$

ここで、

$$\phi'(z) = \frac{-1}{\left(1 + e^{-cz}\right)^2} \cdot \left(-c \cdot e^{-cz}\right)$$

$$= \frac{c \cdot e^{-cz}}{\left(1 + e^{-cz}\right)^2}$$

$$\frac{\partial}{\partial x_{i,h}} \left\{ \left(s_{i,j1} - s_{i2,j2}\right)\left(d_{i,j1}^2 - d_{i2,j2}^2\right) \Big/ \sum_{a,b} d_{a,b}^2 \right\}$$

$$= \frac{\left(s_{i,j1} - s_{i2,j2}\right)\left[\frac{\partial}{\partial x_{i,h}}\left(d_{i,j1}^2 - d_{i2,j2}^2\right) \cdot \left(\sum_{a,b} d_{a,b}^2 \right) - \left(d_{i,j1}^2 - d_{i2,j2}^2\right) \cdot \frac{\partial}{\partial x_{i,h}} \sum_{a,b} d_{a,b}^2 \right]}{\left(\sum_{a,b} d_{a,b}^2 \right)^2}$$

$$\frac{\partial}{\partial x_{i,h}}\left(d_{i,j1}^2 - d_{i2,j2}^2\right) = 2 \cdot \left(x_{i,h} - y_{j1,h}\right)$$

$$\frac{\partial}{\partial x_{i,h}}\left(d_{i,j1}^2 - d_{i,j2}^2\right) = 2 \cdot \left(x_{i,h} - y_{j1,h}\right) - 2 \cdot \left(x_{i,h} - y_{j2,h}\right)$$

$$= -2 \cdot \left(y_{j1,h} - y_{j2,h}\right)$$

$$\frac{\partial}{\partial x_{i,h}} \sum_{a,b} d_{a,b}^2 = \sum_b 2\left(x_{i,h} - y_{b,h}\right)$$

$$\frac{\partial}{\partial x_{i,h}}\left\{\left(s_{i1,j1}-s_{i,j2}\right)\left(d_{i1,j1}^{2}-d_{i,j2}^{2}\right)\Big/\sum_{a,b}d_{a,b}^{2}\right\}$$

$$=\frac{\left(s_{i1,j1}-s_{i,j2}\right)\left[\frac{\partial}{\partial x_{i,h}}\left(d_{i1,j1}^{2}-d_{i,j2}^{2}\right)\cdot\left(\sum_{a,b}d_{a,b}^{2}\right)-\left(d_{i1,j1}^{2}-d_{i,j2}^{2}\right)\cdot\frac{\partial}{\partial x_{i,h}}\sum_{a,b}d_{a,b}^{2}\right]}{\left(\sum_{a,b}d_{a,b}^{2}\right)^{2}}$$

$$\frac{\partial}{\partial x_{i,h}}\left(d_{i1,j1}^{2}-d_{i,j2}^{2}\right)=-2\cdot\left(x_{i,h}-y_{j2,h}\right)$$

である。

$y_{j,h}$ による偏導関数は次式で与えられる。

$$\frac{\partial P}{\partial y_{j,h}}=\sum_{i1,i2,j2}\phi'\left\{\left(s_{i1,j}-s_{i2,j2}\right)\left(d_{i1,j}^{2}-d_{i2,j2}^{2}\right)\Big/\sum_{a,b}d_{a,b}^{2}\right\}$$

$$\times\frac{\partial}{\partial y_{j,h}}\left\{\left(s_{i1,j}-s_{i2,j2}\right)\left(d_{i1,j}^{2}-d_{i2,j2}^{2}\right)\Big/\sum_{a,b}d_{a,b}^{2}\right\}$$

$$+\sum_{i1,j1,i2}\phi'\left\{\left(s_{i1,j1}-s_{i2,j}\right)\left(d_{i1,j1}^{2}-d_{i2,j}^{2}\right)\Big/\sum_{a,b}d_{a,b}^{2}\right\}$$

$$\times\frac{\partial}{\partial y_{j,h}}\left\{\left(s_{i1,j1}-s_{i2,j}\right)\left(d_{i1,j1}^{2}-d_{i2,j}^{2}\right)\Big/\sum_{a,b}d_{a,b}^{2}\right\}$$

ここで、

$$\frac{\partial}{\partial y_{j,h}}\left\{\left(s_{i1,j}-s_{i2,j2}\right)\left(d_{i1,j}^{2}-d_{i2,j2}^{2}\right)\Big/\sum_{a,b}d_{a,b}^{2}\right\}$$

$$=\frac{\left(s_{i1,j}-s_{i2,j2}\right)\left[\frac{\partial}{\partial y_{j,h}}\left(d_{i1,j}^{2}-d_{i2,j2}^{2}\right)\cdot\left(\sum_{a,b}d_{a,b}^{2}\right)-\left(d_{i1,j}^{2}-d_{i2,j2}^{2}\right)\cdot\frac{\partial}{\partial y_{j,h}}\sum_{a,b}d_{a,b}^{2}\right]}{\left(\sum_{a,b}d_{a,b}^{2}\right)^{2}}$$

$$\frac{\partial}{\partial y_{j,h}}\left(d_{i1,j}^{2}-d_{i2,j2}^{2}\right)=2\cdot\left(x_{i1,h}-y_{j,h}\right)\cdot(-1)=-2\cdot\left(x_{i1,h}-y_{j,h}\right)$$

$$\frac{\partial}{\partial y_{j,h}}\left(d_{i1,j}^2 - d_{i2,j}^2\right) = 2\cdot\left(x_{i1,h} - y_{j,h}\right)\cdot(-1) - 2\cdot\left(x_{i2,h} - y_{j,h}\right)\cdot(-1)$$

$$= -2\cdot\left(x_{i1,h} - x_{i2,h}\right)$$

$$\frac{\partial}{\partial y_{j,h}}\sum_{a,b} d_{a,b}^2 = \sum_a 2\left(x_{a,h} - y_{j,h}\right)\cdot(-1) = -2\cdot\sum_a\left(x_{a,h} - y_{j,h}\right)$$

$$\frac{\partial}{\partial y_{j,h}}\left\{\left(s_{i1,j1} - s_{i2,j}\right)\left(d_{i1,j1}^2 - d_{i2,j}^2\right)\Big/\sum_{a,b} d_{a,b}^2\right\}$$

$$= \frac{\left(s_{i1,j1} - s_{i2,j}\right)\left[\frac{\partial}{\partial y_{j,h}}\left(d_{i1,j1}^2 - d_{i2,j}^2\right)\cdot\left(\sum_{a,b} d_{a,b}^2\right) - \left(d_{i1,j1}^2 - d_{i2,j}^2\right)\cdot\frac{\partial}{\partial y_{j,h}}\sum_{a,b} d_{a,b}^2\right]}{\left(\sum_{a,b} d_{a,b}^2\right)^2}$$

$$\frac{\partial}{\partial y_{j,h}}\left(d_{i1,j1}^2 - d_{i2,j}^2\right) = -2\cdot\left(x_{i2,h} - y_{j,h}\right)\cdot(-1) = 2\cdot\left(x_{i2,h} - y_{j,h}\right)$$

プログラム PUnfoldP.dpr

上で説明した Penalty 基準（２）式による展開法を行うプログラムが PUnfoldP.dpr である。PUnfoldP.dpr を実行する[5]と図３のフォームが表示される。

[5] PUnfoldP.dpr は Delphi で開くプロジェクトファイルである。コンパイルされた実行形式のファイルは PUnfoldP.exe である。ここでは、「プロジェクトファイルを実行する」という言い方を用いる。

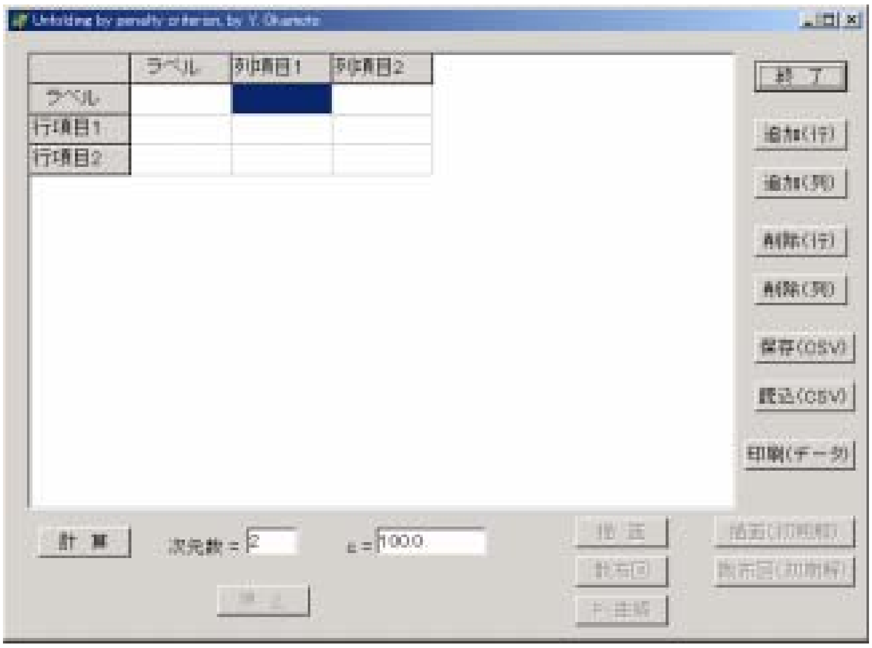

図３　実行開始時のフォーム

図３のフォーム内のグリッドの列数と行数をデータに合わせて設定する。図４は表１のデータを設定したものである。

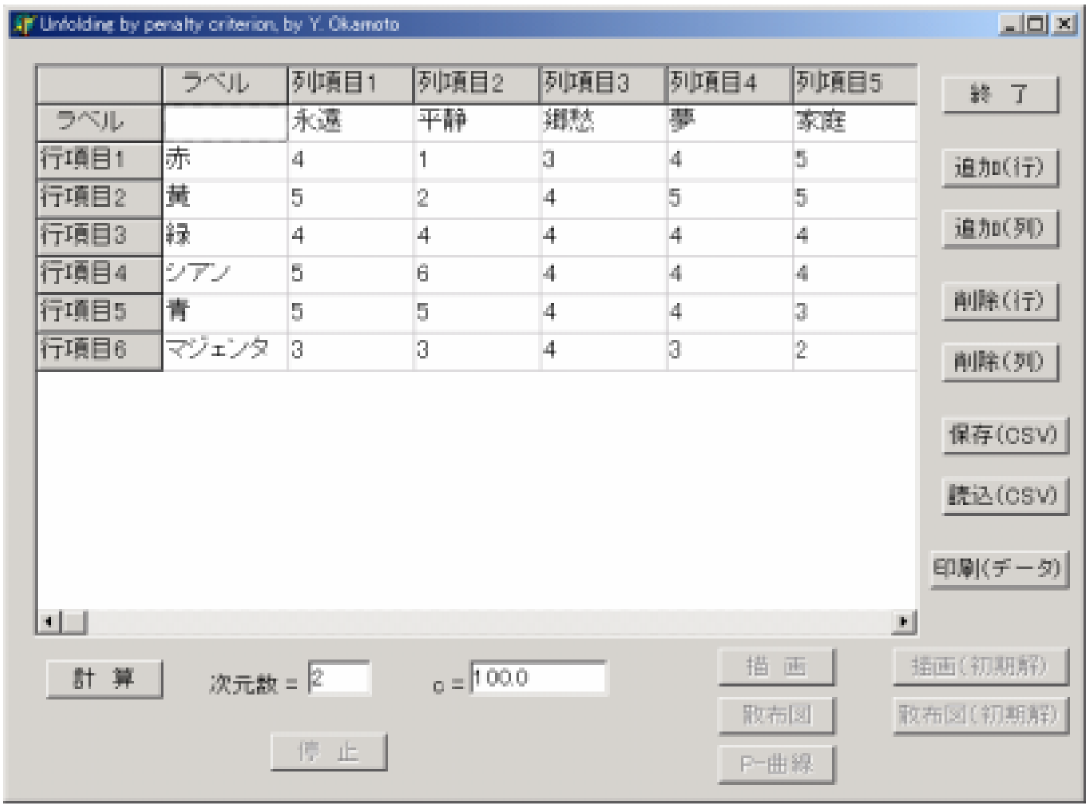

図４　表１のデータの設定

ラベルの行および列に対象のラベルを設定する。設定されたラベルは解の布置の描画において、それぞれの対象を表すのに用いられる。ラベルとして設定された文字列の先頭２バイトが用いられるので、英数字であれば先頭の２文字、漢字などの日本語文字列の場合は先頭の１文字が使用される。データのサイズに合わせてグリッドの行数と列数を調整するが、増やすときは「追加」ボタン、減らすときは「削除」ボタンをクリックする。「追加」ボタンのクリックでアクティブなセルを含む行あるいは列の後に行あるいは列の挿入が行われ、「削除」ボタンのクリックでアクティブなセルを含む行あるいは列が削除される。セルはクリックによりアクティブになる。

行項目ｉと列項目ｊの類似度を該当の行と列の交点に位置するセルに設定する。類似度 $s_{i,j}$ は非負の範囲の数値でなければならない。これは初期値を求める数量化４類の基準による展開法における制約である。

、設定されたデータは「保存」ボタンのクリックで CSV 形式のファイルとして保存される。図４のように設定されたデータを保存したものを Excel で開くと図５のように表示される。

DColorWord.csv

	A	B	C	D	E	F	G	H
1		永遠	平静	郷愁	夢	家庭	幸福	愛
2	赤	4	1	3	4	5	5	6
3	黄	5	2	4	5	5	6	3
4	緑	4	4	4	4	4	4	3
5	シアン	5	6	4	4	4	3	3
6	青	5	5	4	4	3	3	2
7	マジェンタ	3	3	4	3	2	3	3

図５　CSV 形式で保存したデータを Excel で開いたもの

逆に Excel で図５の形式に従って用意したデータを CSV 形式のファイルで保存（拡張子を.csv とする）したものは、図３のフォームにおいて「読込」ボタンをクリックすると図４のようにグリッド内に読み込むことができる。

設定されたデータは、「出力（データ）」ボタンのクリックでテキストファイルに出力することが出来る。「出力（データ）」ボタンをクリックすると図６のように出力先のファイル名の設定を求めるダイアログボックスが表示される。

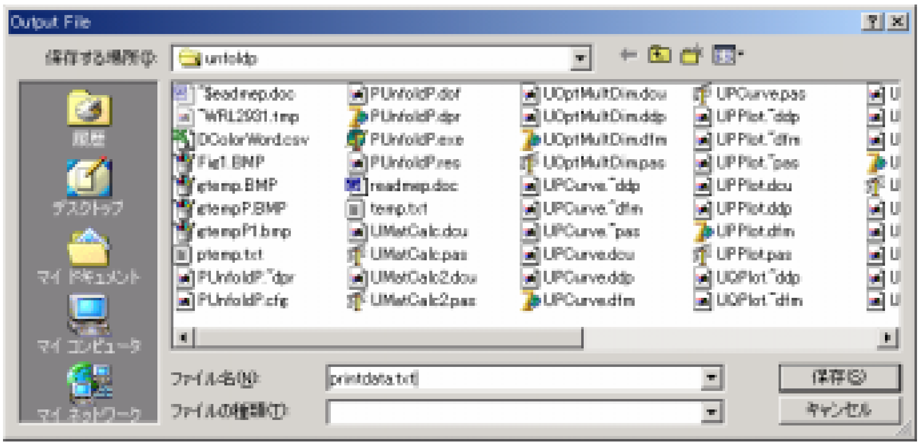

図６　出力先のファイル名の設定

出力先のファイル名を適当に設定して「保存」ボタンをクリックすると、設定したファイル名のテキストファイルにデータが書き出されるが、その前に図７のダイアログボックスが表示される。

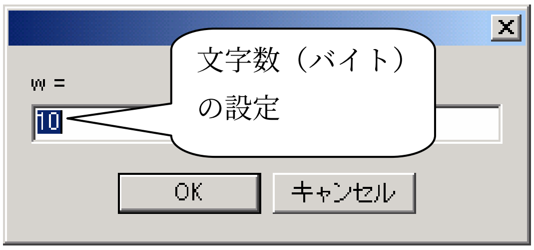

図７　文字数の設定

図７のダイアログボックスにおいて設定される「w =」の値は、データの書き出しにおけるセル当たりの文字数（バイト）である。ラベルなど日本語文字は２バイト単位になっているので、w の値としては偶数を設定するようにする。図７の「OK」ボタンのクリックで、セル当たりの文字数 w バイトでテキストファイルに書き出される。図８は書き出されたテキストファイルをエディタで開いたものである。

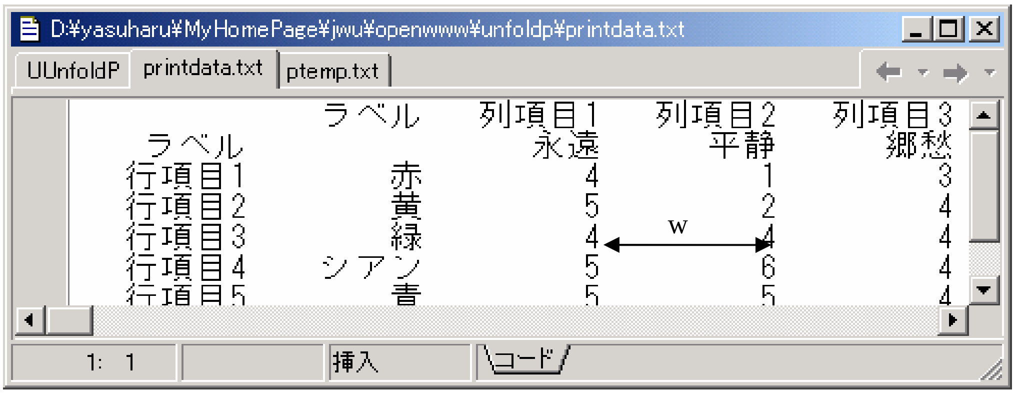

図８　データの出力例

データの設定後、「計算」ボタンをクリックすると計算が始まるが、その前に次元数とｃの値を適当な値に設定する（図９）。

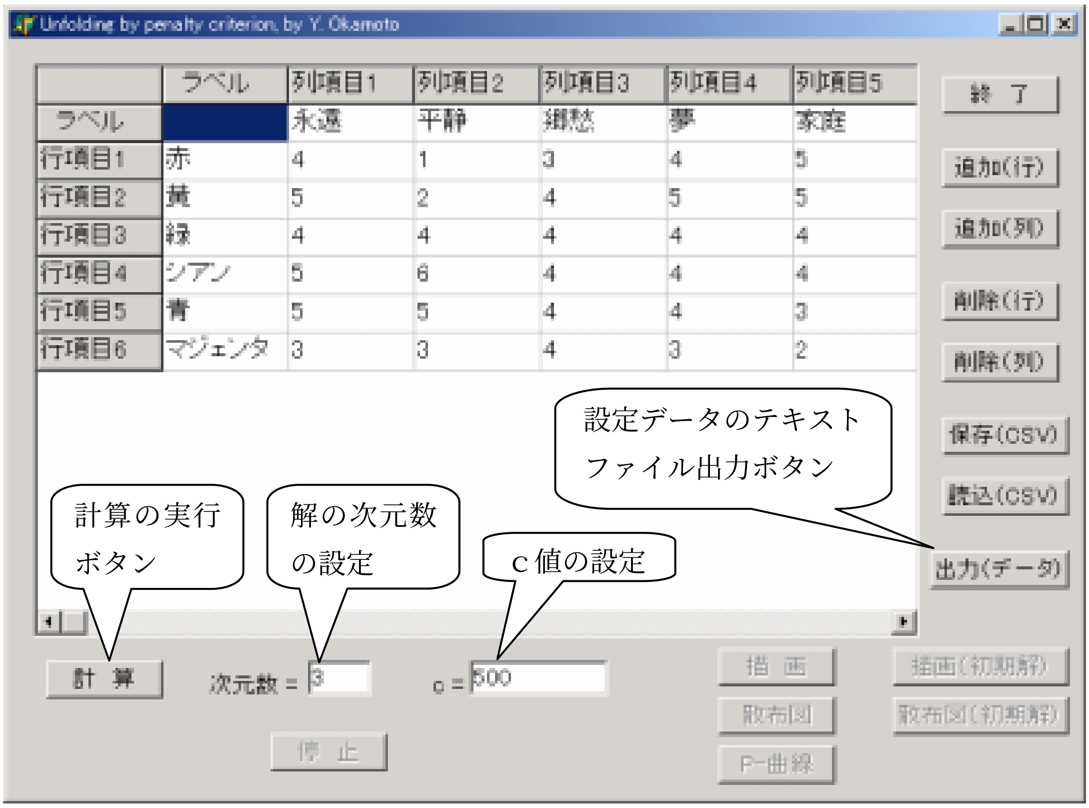

図９　計算の実行

ｃの値として設定された値が penalty 関数 $\phi(z)=1/\left(1+e^{-c\cdot z}\right)$ における z の係数 c の値として設定される。次元数は、解の次元ｐの値である。

「計算」ボタンをクリックすると、計算の実行の前に、計算結果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図１０のように表示される。

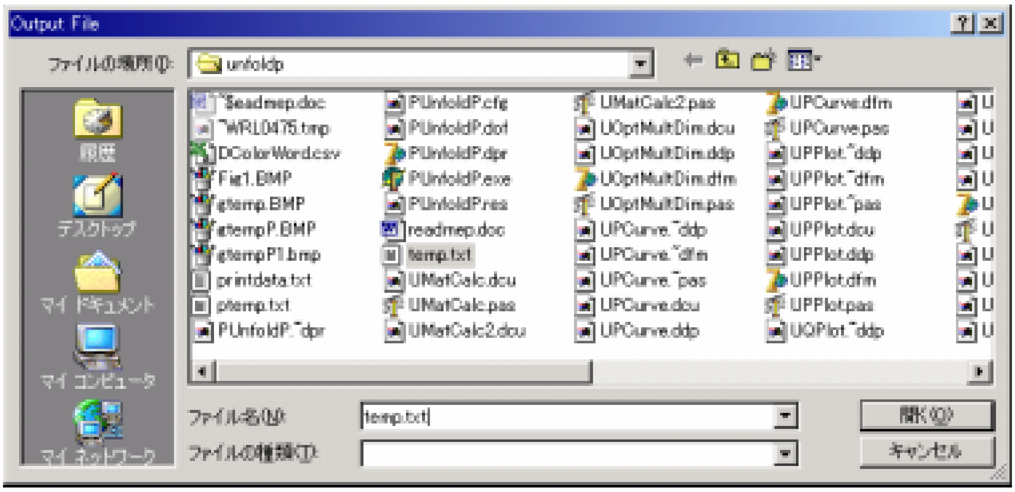

図１０　計算結果出力用のファイル名の設定

図１０のダイアログボックスで設定した名前のテキストファイルに計算結果が出力される。ファイル名の設定後、「開く」ボタンをクリックすると計算が始まる。解は（２）式で与えられる P の値を最小にするものとして極値探索法で求められる。極値探索法における初期値は、数量化４類の基準による展開法[6]による解を用いている。数量化４類の基準による展開法の解が求まると、その解を初期値とし極値探索が始まる。極値探索が行われている間は、その途中経過を表示するフォームが図１１のように提示される。

[6] 岡本安晴、1991、「数量化IV類の基準による展開法」、日本行動計量学会第 19 回大会発表論文抄録集、32-33.
岡本安晴、1993、「数量化４類の基準による展開法における制約条件について」、金沢大学文学部論集・行動科学科篇、第 13 号、35-54.
岡本安晴、1995、「数量化４類の基準による展開法」、行動計量学、第 22 巻、第２号、126-134.
岡本安晴、1998、「数量化４類の基準による展開法の試み」、第 66 回日本統計学会講演報告集、443-444.

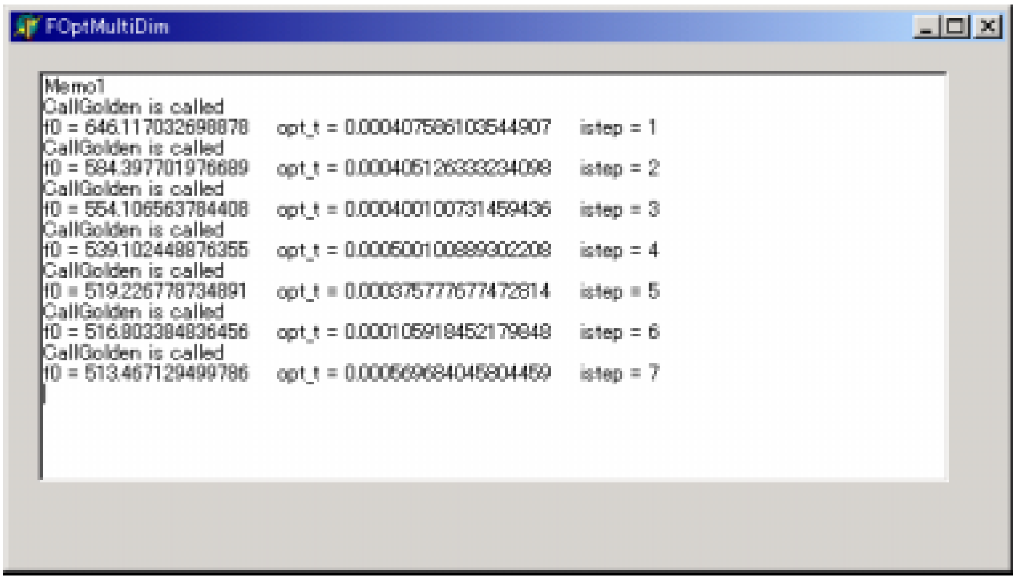

図１１　極値探索の途中経過の表示

　計算が終了すると、計算結果を書き出したテキストファイルの名前を表示するダイアログボックスが図１２のように提示される。

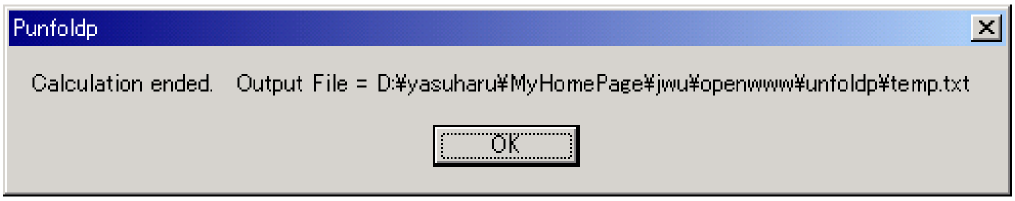

図１２　計算結果出力先のファイル名の表示

「OK」ボタンのクリックで、フォームは図１３のようになる。計算前は Enable でなかった「描画（初期解)」、「散布図（初期解)」、「描画」、「散布図」、「P-曲線」の各ボタンが、計算終了後は Enable になっている。

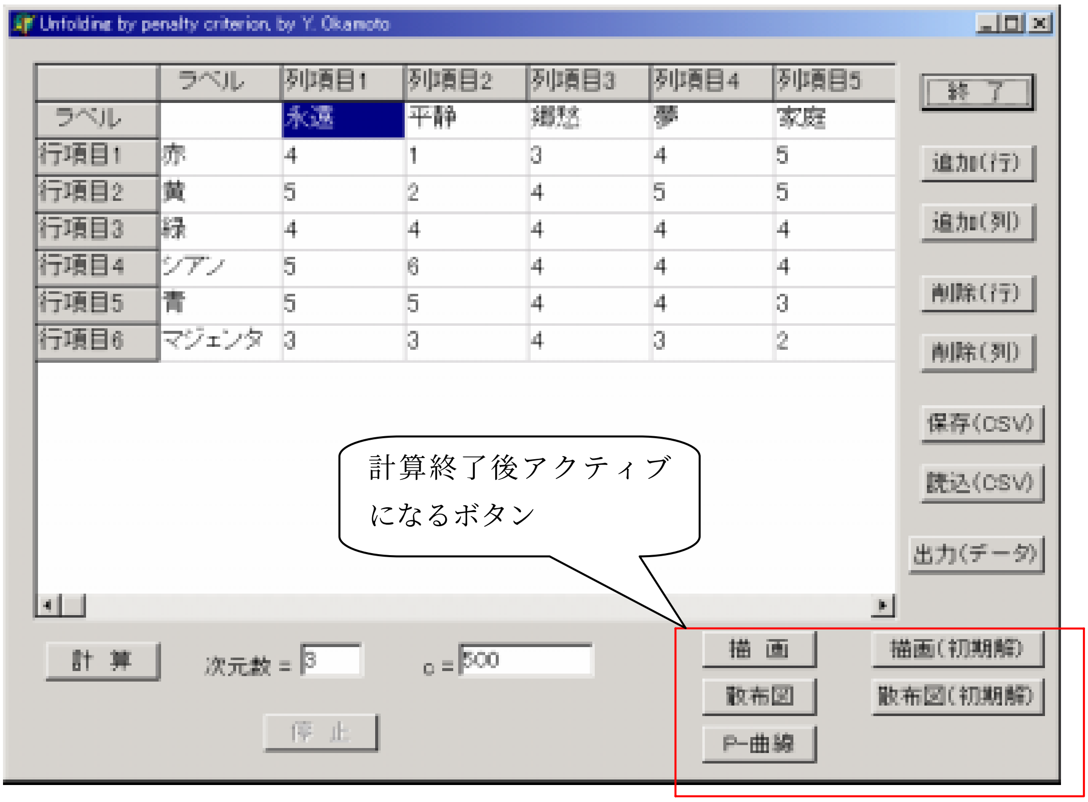

図１３　計算終了後のフォーム

「描画（初期解)」ボタンをクリックすると、初期値として求めた数量化４類の基準による展開法の解が描画される（図１４)。

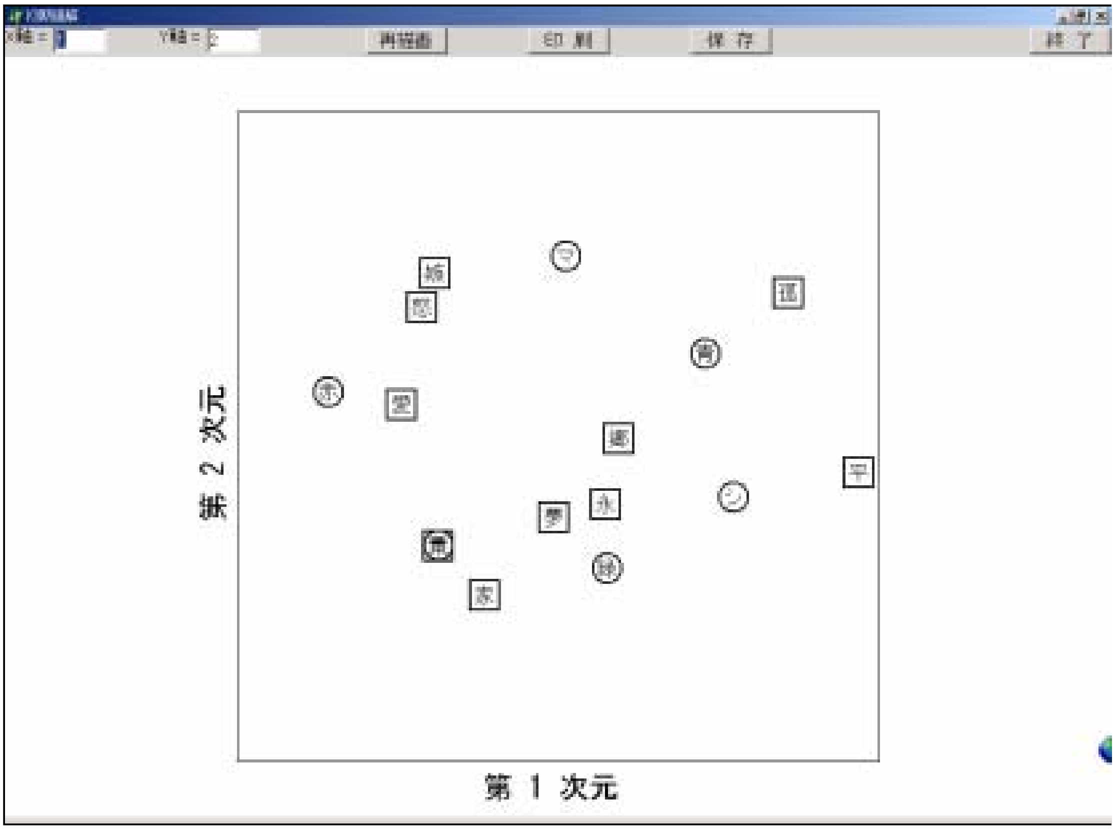

図１４　初期値（数量化４類の基準による解）

数量化４類の基準による展開法はデータのおおまかな情報をよくとらえるもので、図 14 に示される解も表１のデータの情報をよく表している。しかし、数量化４類の基準は類似度と距離の単調関係を厳しく求めるものではないので、類似度と解として与えられた布置における距離との関係を散布図として表すと単調関係はそれほどよくはない。数量化４類の基準による解には、２組の対象の布置においてそれぞれの組内での重み付き重心が原点にくること、原点から離れるに従って距離関係にバイアスが掛かることという特性がある[7]。類似度と距離の関係を示す散布図は「散布図（初期解）」ボタンのクリックで表示される（図１５）。

[7] 岡本安晴、1995、「数量化 4 類の基準による展開法」、行動計量学、第 22 巻、第２号、126-134.

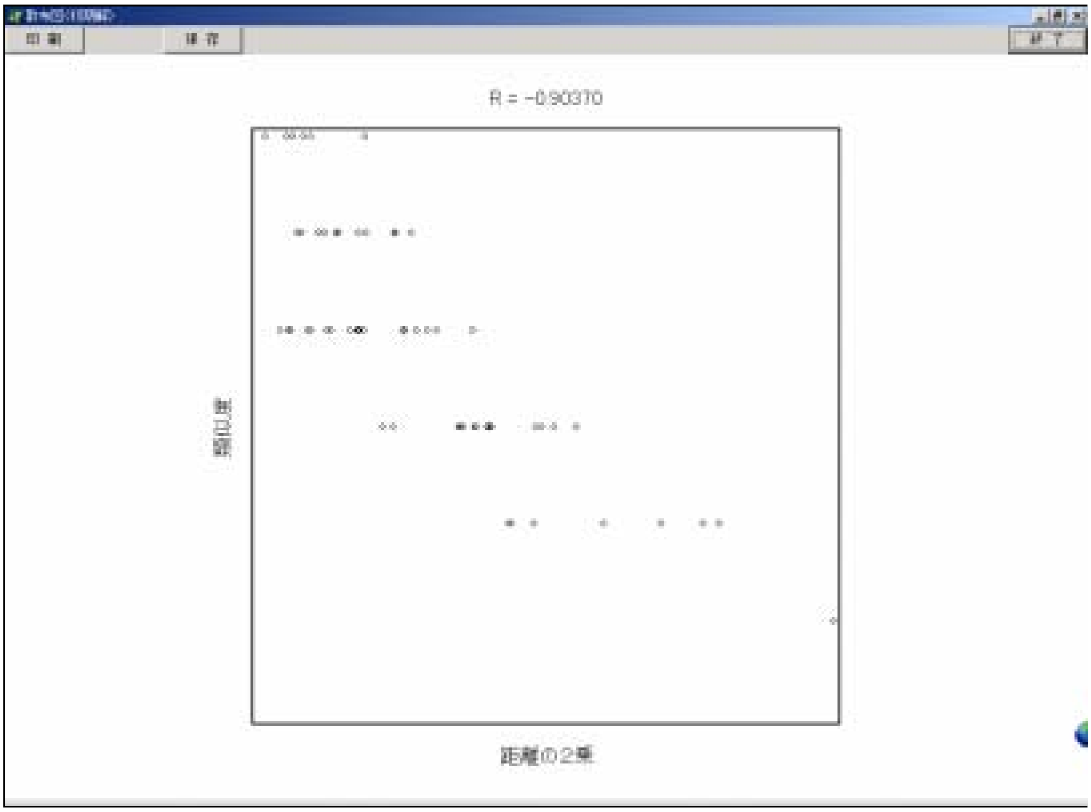

図１５　初期解に対する散布図

図１５は点$\left(d_{i,j}^2 \quad s_{i,j}\right)$をプロットしたものである。おおまかには距離の２乗$d_{i,j}^2$と類似度$s_{i,j}$の間に単調関係が認められる。図の上部に表示されているＲの値は、距離と類似度との間の順位相関係数である。ただし、類似度は６カテゴリの評定値における９人の評定者の中央値なので１から６までの整数値であり、タイ・スコアが多くあるが、タイ・スコアには順位の平均値を対応させるという方法はとっていない。評定値がタイ・スコアのときは、対応する距離の方の大小関係に基づいて評定値の方の順序付けを行っている。この方法による順位相関係数の方が、penalty 基準による解のデータに対する適合度がよく表されるからである。

Penalty 基準による解の布置は、「描画」ボタンのクリックにより表示される（図１６）。

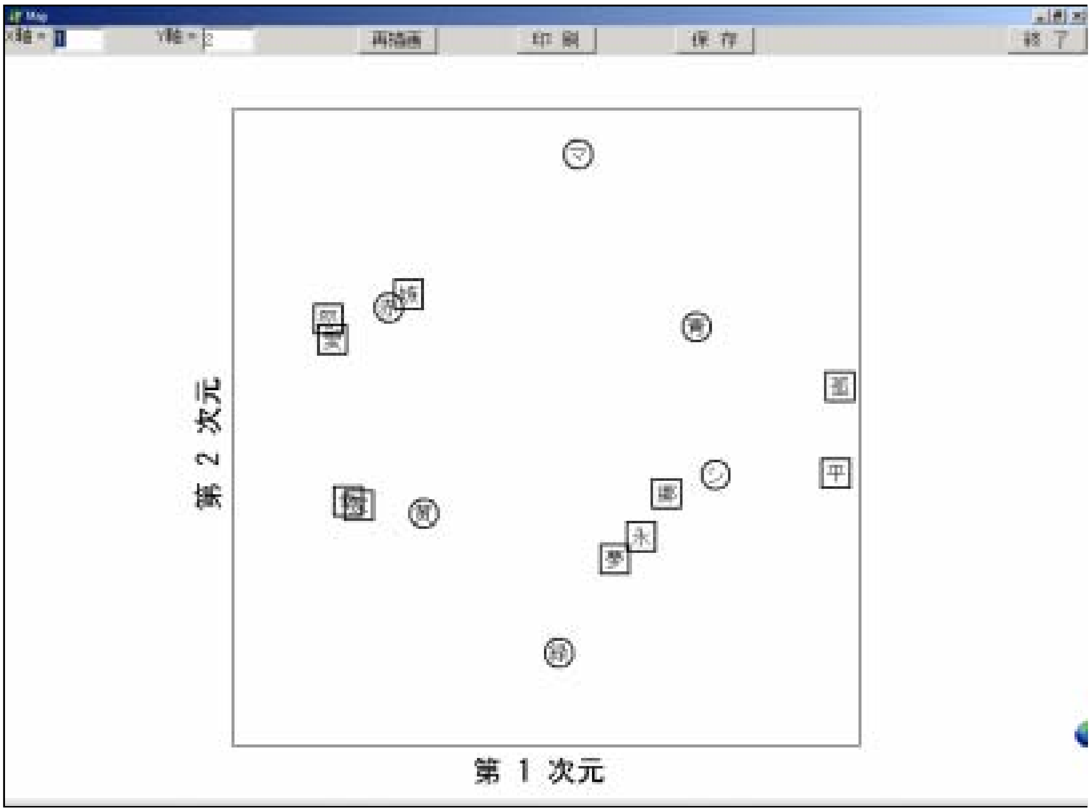

図１６　Penalty 基準による解

Penalty 基準による解は、座標値の絶対値の最大値が１になるように標準化されている。これは、（３）式の値

$$\frac{\left(s_{i1,j1}-s_{i2,j2}\right)\cdot\left(d_{i1,j1}^{2}-d_{i2,j2}^{2}\right)}{\sum_{i,j} d_{i,j}^{2}} \qquad (3)$$

が、座標値$\left(x_{i,t}\right)$および$\left(y_{j,t}\right)$をすべて定数α倍した座標値$\left(\alpha x_{i,t}\right)$および$\left(\alpha y_{j,t}\right)$に対しても変わらないからである。すなわち、$\alpha$倍した座標値$\left(\alpha x_{i,t}\right)$および$\left(\alpha y_{j,t}\right)$における距離を$d_{i,j}^{(\alpha)}$とおくと、

$$d_{i,j}^{(\alpha)}=\sqrt{\sum_{t}\left(\alpha x_{i,t}-\alpha y_{j,t}\right)^{2}}=\alpha\cdot d_{i,j}$$

であり、

$$\frac{\left(s_{i1,j1}-s_{i2,j2}\right)\cdot\left[\left(d_{i1,j1}^{(\alpha)}\right)^{2}-\left(d_{i2,j2}^{(\alpha)}\right)^{2}\right]}{\sum_{i,j}\left(d_{i,j}^{(\alpha)}\right)^{2}}=\frac{\left(s_{i1,j1}-s_{i2,j2}\right)\cdot\alpha^{2}\left(d_{i1,j1}^{2}-d_{i2,j2}^{2}\right)}{\alpha^{2}\sum_{i,j} d_{i,j}^{2}}$$

$$= \frac{\left(s_{i1,j1} - s_{i2,j2}\right)\cdot\left(d_{i1,j1}^{2} - d_{i2,j2}^{2}\right)}{\sum_{i,j} d_{i,j}^{2}}$$

が成り立ち、（３）式の値が座標値の定数倍によって変わらないからである。

図１６の画面上部左にあるＸ軸およびＹ軸の指定を変更後、「再描画」ボタンをクリックすると、再設定した次元による解の描画が行われる（図１７）。

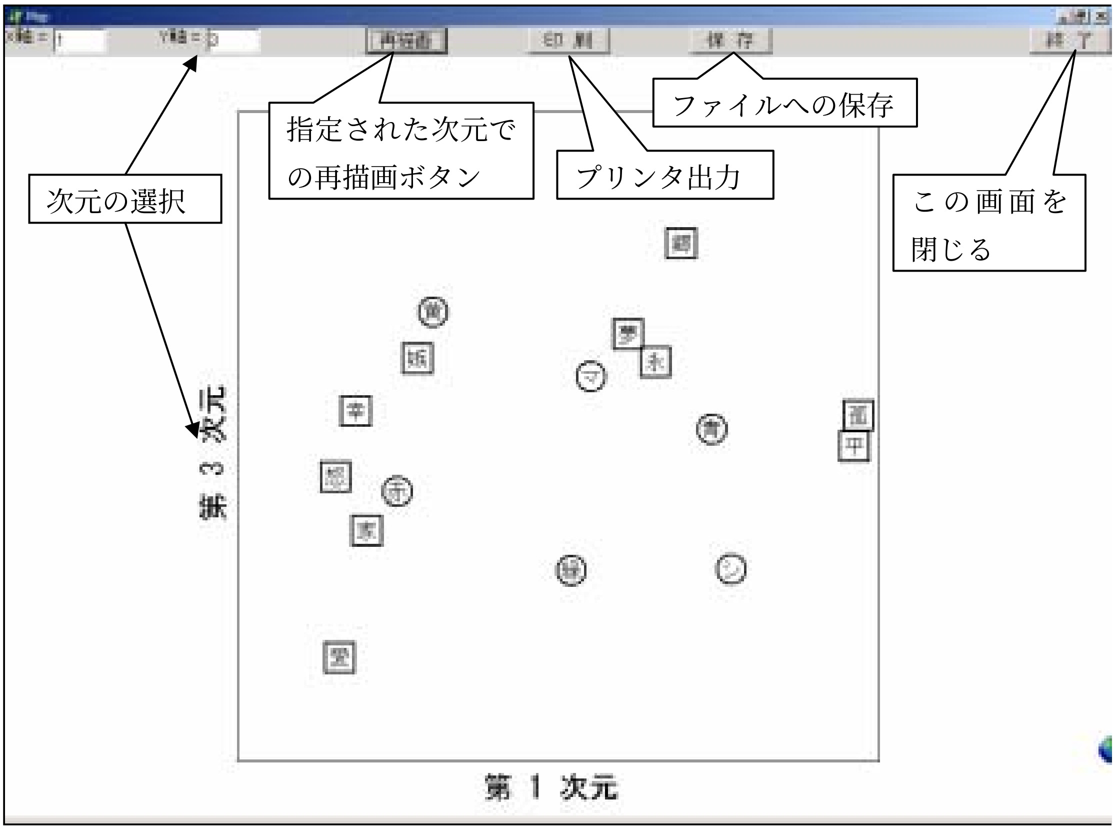

図１７　再描画・印刷・保存・終了ボタン

描画表示されている布置は、「印刷」ボタンのクリックでプリンタに出力される。「保存」ボタンをクリックすると、描画されている布置を画像ファイル（*.bmp ファイル）として保存することが出来る。

「終了」ボタンのクリックで描画画面が閉じられる。

「散布図」ボタンのクリックで penalty 基準による解に対する点$\left(d_{i,j}^{2}\quad s_{i,j}\right)$の散布図が表示される（図１８）。

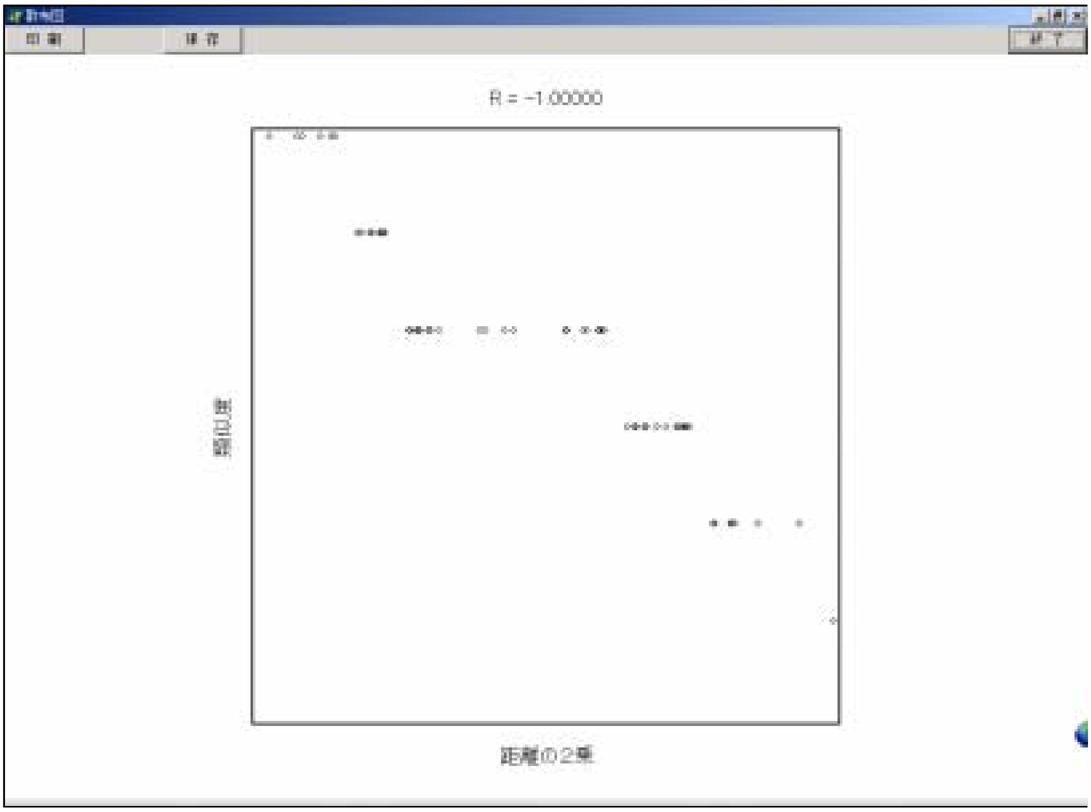

図１８　Penalty 基準による解における散布図

初期値（数量化４類の基準による解）に比べて距離と類似度の単調関係が非常によくなっていることがわかる。順位相関係数 R はー１と表示されている。この場合の順位相関係数の算出においても、初期値の場合（図１５）と同様、類似度がタイ・スコアのときは対応する距離の方の順序関係に合わせて順位付けを行っている。

「P-曲線」ボタンをクリックすると、Penalty 関数 $\phi(z)$ が表示される（図１９）。

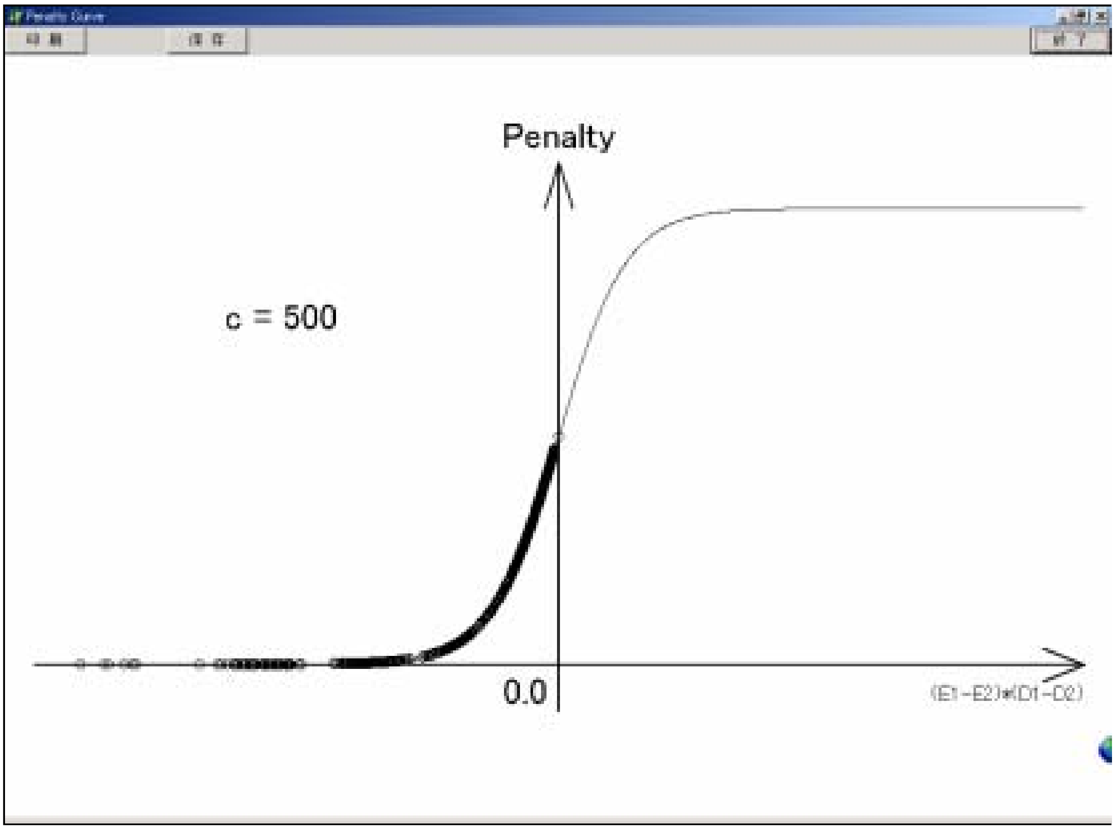

図１９　Penalty 関数

Penalty 関数を表す曲線上の小円は、点$\left(z_{i1,j1,i2,j2}\quad \phi\left(z_{i1,j1,i2,j2}\right)\right)$を表している。ただし、

$$z_{i1,j1,i2,j2}=\frac{\left(s_{i1,j1}-s_{i2,j2}\right)\left(d_{i1,j1}^{2}-d_{i2,j2}^{2}\right)}{\sum_{a,b}d_{a,b}^{2}}$$

である。図１９において、小円が原点の左側にあるときは$z_{i1,j1,i2,j2}<0$であるので、これは類似度と距離の順序関係が（４a）式あるいは（４b）式を満たし、うまく対応していることを表す。原点の上にある点は$z_{i1,j1,i2,j2}=0$であることを表す。表１の評定尺度データの場合、タイ・スコア（$s_{i1,j1}=s_{i2,j2}$）があるが、このときは常に$z_{i1,j1,i2,j2}=0$となる。図１９の penalty 関数の図は、図１８の散布図の図とともに、penalty 基準による解が表１のデータの順序関係をよく表していることを示すものである。

類似度と距離との単調関係は c の値が大きいほど対応がよくなるが、極値探索の計算時間も長くなる。プログラム PUnfoldP.dpr による分析においては、解の次元数の選択ととも

にｃの値の選択も重要である。次元数とｃの値との組み合わせをいろいろ試してみる必要がある。計算終了時のフォーム（図９）において、次元数とｃ値の再設定を行った後「計算」ボタンをクリックすると、再設定した値での計算が行われる。再設定した値での計算が始まるときに計算結果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が（図１０）、このときは前の計算において設定した名前とは別の名前を設定しておくと、後でそれぞれの計算結果を見ることができる。前に設定したものと同じ名前を設定すると、前の計算結果に上書きされて前の計算結果は消されてしまう。また、ｃの値が大きすぎて極値探索がなかなか終了しないときは、Enable になっている「停止」ボタンをクリックすると極値探索から途中で抜け出して計算が終了する。

図９の設定における計算結果の出力ファイルの内容は、リスト１のようになっている。

リスト１　テキストファイルへの出力例

```
行項目のラベル...
行項目-1 ==> 赤
行項目-2 ==> 黄
行項目-3 ==> 緑
行項目-4 ==> シアン
行項目-5 ==> 青
行項目-6 ==> マジェンタ

列項目のラベル...
列項目-1 ==> 永遠
列項目-2 ==> 平静
列項目-3 ==> 郷愁
列項目-4 ==> 夢
列項目-5 ==> 家庭
列項目-6 ==> 幸福
列項目-7 ==> 愛
列項目-8 ==> 嫉妬
列項目-9 ==> 怒り
列項目-10 ==> 孤独

類似度データ...
         4.00         1.00         3.00         4.00         5.00         5.00         6.00         6.00
6.00         2.00
         5.00         2.00         4.00         5.00         5.00         6.00         3.00         5.00
4.00         2.00
         4.00         4.00         4.00         4.00         4.00         4.00         3.00         2.00
3.00         3.00
         5.00         6.00         4.00         4.00         4.00         3.00         3.00         3.00
2.00         5.00
         5.00         5.00         4.00         4.00         3.00         3.00         2.00         4.00
4.00         6.00
         3.00         3.00         4.00         3.00         2.00         3.00         3.00         5.00
4.00         4.00

Standardized E =
 0.67 0.17 0.50 0.67 0.83 0.83 1.00 1.00 1.00 0.33
 0.83 0.33 0.67 0.83 0.83 1.00 0.50 0.83 0.67 0.33
 0.67 0.67 0.67 0.67 0.67 0.67 0.50 0.33 0.50 0.50
 0.83 1.00 0.67 0.67 0.67 0.50 0.50 0.50 0.33 0.83
```

```
 0.83 0.83 0.67 0.67 0.50 0.50 0.33 0.67 0.67 1.00
 0.50 0.50 0.67 0.50 0.33 0.50 0.50 0.83 0.67 0.67

---------------- Calculation of the Initial Configuration ----------------



得られた固有値の総数 = 6

eigen values =
  0.24197  0.12817  0.07841  0.04577  0.04005
最大固有値 1.0 は除きました



               contribution(%)     cum.contr(%)

dim 1                45.3              45.3
dim 2                24.0              69.3
dim 3                14.7              83.9
-------------------------------------------------
dim 4                 8.6              92.5
dim 5                 7.5             100.0

解の初期値（数量化展開法）

X =
       -0.05092        0.00986       -0.02157
       -0.02658       -0.02363        0.02944
        0.01069       -0.02864       -0.00993
        0.03864       -0.01304       -0.01764
        0.03233        0.01823        0.01226
        0.00159        0.03932        0.00718
Y =
        0.01046       -0.01483        0.00777
        0.06631       -0.00785       -0.01543
        0.01328       -0.00051        0.01138
       -0.00089       -0.01775        0.01127
       -0.01617       -0.03470       -0.01097
       -0.02645       -0.02391        0.01831
       -0.03468        0.00687       -0.04958
       -0.02708        0.03560        0.01526
       -0.02997        0.02809        0.00056
        0.05083        0.03119        0.00001

Rank correlation coeficient (Initial Configuration) = -0.90370


----------------- ペナルティ基準による解 -----------------

c = 500
Rank correlation coefficient = -1.00000

X =
       -0.53671        0.40354       -0.18025
       -0.41567       -0.28787        0.41032
        0.04397       -0.75679       -0.43763
        0.57643       -0.16001       -0.43489
        0.51056        0.33715        0.02691
```

```
        0.10731        0.91701        0.19615
Y =
        0.32299       -0.36717        0.24459
        0.98556       -0.15553       -0.03130
        0.41151       -0.22289        0.63353
        0.23398       -0.44087        0.33479
       -0.63569       -0.26259       -0.30834
       -0.67160       -0.24688        0.08379
       -0.72581        0.29454       -0.72749
       -0.46591        0.44568        0.25417
       -0.74092        0.36461       -0.13150
        1.00000        0.13807        0.06714
```

まず、行項目と列項目の対象のラベルが書き出されている。続いて、類似度データが書き出される。類似度データに続く「Standardize E」の値は、類似度データを最大値が１になるように標準化したものである。解を求める計算は、この最大値が１になるように標準化された類似度データに対して行われる。

標準化された類似度データの書き出しに続いて、数量化４類の基準による展開法の分析結果が書き出される。２組の対象の布置（座標値）の出力の後、この解における距離と類似度データとの順位相関係数が書き出されている。リスト１の場合、この順位相関係数は－0.90370 となっている。この値は図１５の画面で表示されているものである。

数量化４類の基準による展開法の解の書き出しに続いて、penalty 基準による解が出力されている。まず、用いられた c の値が書き出されている。リスト１の場合は「c = 500」となっている。続いて、順位相関係数が書き出されている。この場合の順位相関係数は 1.0 であり、この値は図１８の画面においても表示されている。最後に、２組の対象のそれぞれの座標値が書き出されている。