

研究論文

特徴間類似性モデルとその最小二乗解

足立 浩平¹⁾

Inter-feature similarity models and their least squares solutions

ADACHI Kohei

In this paper, a model is presented that represents the similarity of a stimulus pair as either the sum or the average of inter-feature similarity parameters. Those parameters stand for the similarities between the features of a stimulus of the pair and those of its counterpart. An extended version of the model is also presented which includes the parameters representing the weights of features. Least squares algorithms are developed to obtain the parameters with a data set of inter-stimulus similarities and stimulus features. A numerical example is given to illustrate the presented models and algorithms, and mathematical properties of the models are discussed.

Key words : inter-stimulus similarity, inter-feature similarity, least squares methods

キーワード : 刺激間類似性, 特徴間類似性, 最小二乗法

1. 序 論

「AさんとBさんの服装は似ている」, 「C町とD町は雰囲気が異なる」といった表現に例示されるように, 人は二つの対象(刺激)間の類似・非類似を知覚し, それに言及することが多い。こうした類似性について, 「個々の刺激が有する諸特徴(属性)の中で, どのような特徴が, 刺激間の類似性に深く関与しているか」といった問いが, しばしば興味の対象となる。ここで, 関与の程度をウェイトと言い換えれば, 上記の問いは, 「刺激間の類似性に対して, 刺激の諸特徴がどの程度のウェイトを持つか」と言い換えられ, ウェイトと類似性を関係づけるものとして, 特徴ウェイトモデルと総称できるモデルの群がある。

こうした特徴ウェイトモデルの中で最も簡潔

なもの共通特徴モデルである。共通特徴モデルでは, 刺激*i*の特徴集合を O_i で表すと, 刺激*i, j*間の類似性が, O_i と O_j の共通集合の関数 $f(O_i \cap O_j)$ によって表せると仮定する。ここで, 関数*f*は, 集合内の特徴のウェイトを合計するものである(例えば, Shepard & Arable, 1979)。例えば, 人物*i*および人物*j*の性格(特徴)の集合をそれぞれ $O_i = \{\text{内向的, 神経質, 努力家}\}$ および $O_j = \{\text{神経質, 努力家, 知的, 社会的}\}$ とすると, 両者の性格の類似性は, $O_i \cap O_j = \{\text{神経質, 努力家}\}$ の要素のウェイトの合計と考えるわけである。この方法では, 考えられる全ての特徴の集合(特徴の全集合)を $O = \{k | k = 1, 2, \dots, m\}$ とおいた上で, 特徴データ x_{ik} を, 刺激*i*が特徴*k*を持つときは $x_{ik} = 1$, 持たないときは $x_{ik} = 0$ と定義し, 特徴*k*にウェイト w_k を対応させると, 刺激*i, j*間の類似性は, $\phi_{ij} = f(O_i \cap O_j) = \sum_{k=1}^m w_k \min(x_{ik}, x_{jk})$ と表せる。以上の共通特徴モデルとは反対に, 刺

1) 立命館大学文学部

刺激 i, j 間の示差特徴集合 $(O_i - O_j) \cup (O_i - O_j) = \sum_{k=1}^m w_k |x_{ik} - x_{jk}|$ を非類似性（つまり類似性の小ささ）に対応させるモデルもある（例えば、北條，1982）。さらに、特徴の共通集合と同時に示差集合を考慮したものが、Tversky（1977）のコントラストモデルである。足立（1998）は、以上の諸モデルに基づき、類似性および特徴のデータから、特徴のウェイトを求める統計解析法、すなわち、類似性データを従属変数、特徴データを説明変数とした重回帰分析を提案している。

さて、特徴ウェイトモデルでは、個々の特徴のウェイトを考慮するが、特徴と特徴間の関係を考慮していないのが欠点といえる。例えば、特徴データ $O_i = \{\text{太った, 色黒}\}$ および $O_j = \{\text{丸顔, スポーツマン風}\}$ によって記述される2人の人物間の類似性は、両特徴集合間に文字通りの共通特徴がないため、各特徴のウェイトがいかなる値であっても、共通特徴モデルから導かれる類似性が0、つまり、両者は全く似ていないことになる。同様に、示差特徴モデルやコントラストモデルでも、両者は非常に非類似であることが予測される。しかしながら、特徴「太った」と「丸顔」、および、「色黒」と「スポーツマン

風」は類似するため、両者はむしろ似ているものと思われる。このように、特徴間関係を考慮しない特徴ウェイトモデルでは低い類似性が予測されるが、個々の特徴どうしが類似するために、似ていると判断される刺激対の例は他にも数多くあろう。

そこで、本研究では、特徴どうしの類似性がパラメータとして組み込まれた刺激間の類似性のモデルを提案する。すなわち、刺激の対が互いにもつ特徴間の類似性の高さに応じて、刺激間の類似性も高くなるモデルを考える。提案するモデルには、幾つかの下位モデルに分かれるが、特徴間類似性モデルと総称する。モデル構成の後、モデルから予測される刺激間類似性と、実際に観測される刺激間類似性の二乗誤差を最小にするパラメータを求めるための、アルゴリズムを考えることが研究の第2の目的となる。提案したアルゴリズムのデータへの適用、および、モデルとアルゴリズムの数理的性質の考察も行う。

2. 単純総和モデル

計 m 個の特徴のそれぞれを k または l で表し、

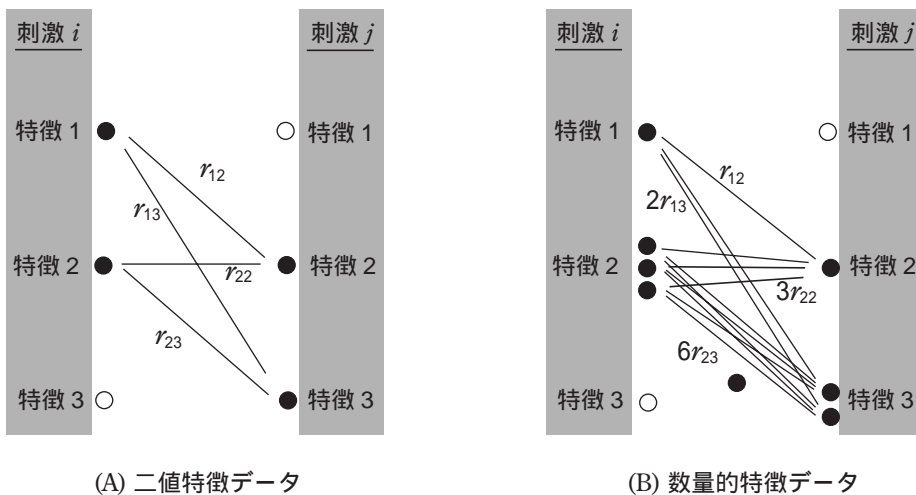


図1. 特徴間類似性モデル（○は特徴がないこと、●の数は特徴量を表す）

特徴 k と特徴 l の間の類似性を表すパラメータを r_{kl} と表記する。ただし、特徴間類似性は対称、すなわち

$$r_{kl} = r_{lk} \quad (k = 1, \dots, m; l = 1, \dots, m) \quad (1)$$

とする。全ての特徴対に関する特徴間類似性のパラメータをまとめると、 m 次の対称行列

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{12} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1m} & r_{2m} & \cdots & r_{mm} \end{bmatrix} \quad (2)$$

によって表せる。これらのパラメータから刺激間のモデル類似性を定義する。定義法として最も簡潔なものは、2つの刺激が有する特徴間の類似性を加算するものである。この定義法を、特徴の総数 m を3として例示したのが図1である。特徴データ x_{ik} が二値、つまり、特徴の有無で記述される場合を想定し、例えば、刺激 i が特徴1と特徴2を持ち、刺激 j が特徴2と特徴3を持つケースを例示したのが、図の(A)である。(A)に示すように、両刺激の所有特徴を結び付ける結合を考え、各結合に対応する類似性を足しあげた値 $r_{12} + r_{13} + r_{22} + r_{23}$ をモデル類似性とする。特徴データ x_{ik} が1,0以外の数量である場合には、図の(B)に示すように、刺激 i は特徴 k を「 x_{ik} 個持っている」とみなした上で、この個数に応じた分だけの結合を考え、結合に対応する類似性の総和をモデル類似性とする。このことは、 x_{ik} が整数に限らず、1.5や3.6といった実数の場合一般に拡張できる。

以上の定義法を式で表すと、特徴が二値および数量のいずれの場合も、モデル類似性は、

$$\phi_{ij} = \phi(\mathbf{x}_i, \mathbf{x}_j; \mathbf{R}) = \sum_{k=1}^m \sum_{l=1}^m r_{kl} x_{ik} x_{jl} = \mathbf{x}_i' \mathbf{R} \mathbf{x}_j \quad (3)$$

と表現できる。これを特徴間類似性モデルと呼ぶことにする。モデル類似性に加算定数 c を加えた

$$\hat{s}_{ij} = \sum_{k=1}^m \sum_{l=1}^m r_{kl} x_{ik} x_{jl} + c \quad (4)$$

を予測類似性として、これとデータとの二乗誤

差の総和

$$\eta = \sum_{(i,j)} (s_{ij} - \hat{s}_{ij})^2 = \sum_{(i,j)} \left[s_{ij} - \left(\sum_{k=1}^m \sum_{l=1}^m r_{kl} x_{ik} x_{jl} + c \right) \right]^2 \quad (5)$$

を最小にする $\mathbf{R} = (r_{kl})$ を求めることが、特徴間類似性モデルによる解析の目的になる。

この問題を解くため、幾つかの行列とベクトルを定義する。定義は幾分煩雑であるが、定義を終えると解法はきわめて簡潔である。まず、

$$y_{ijkl} = \begin{cases} x_{ik} x_{jl} & (k = l) \\ x_{ik} x_{jl} + x_{il} x_{jk} & (k \neq l) \end{cases} \quad (6)$$

とした上で、刺激 i, j について、全ての特徴対の組み合わせに関する $y_{ijkl} (k \leq l)$ を並べた後に要素1を加えた $[m(m+1)/2+1]$ 次元のデータ・ベクトル

$$\mathbf{y}_{ij} = [y_{ij11}, y_{ij12}, \dots, y_{ij1m}, y_{ij22}, y_{ij23}, \dots, y_{ij2m}, \dots, y_{ij,m-1,m-1}, y_{ij,m-1,m}, y_{ijmm}, 1] \quad (7)$$

を定義する。また、全ての特徴対の組み合わせに関する $r_{kl} (k \leq l)$ を上記の \mathbf{y}_{ij} の要素と同様に並べた後に、加算定数 c を加えた $[m(m+1)/2+1]$ 次元のパラメータ・ベクトル

$$\mathbf{r} = [r_{11}, r_{12}, \dots, r_{1m}, r_{22}, r_{23}, \dots, r_{m-1,m-1}, r_{m-1,m}, r_{mm}, c] \quad (8)$$

(一連の特徴間類似性を、それらの対称性を考慮して、行列 \mathbf{R} でなく、一続きに並べた後に c を加えたベクトル) を定義する。さらに、一連の類似性データを要素とする N 次元ベクトル \mathbf{s} の要素の並びに対応させて、各行に行ベクトル \mathbf{y}_{ij}' を配置した $N \times [m(m+1)/2+1]$ の行列を \mathbf{Y} と表す。自己類似性を考慮しない場合には、

$$\mathbf{s} = \begin{bmatrix} s_{12} \\ s_{13} \\ \vdots \\ s_{n-1,n} \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} \mathbf{y}'_{12} \\ \mathbf{y}'_{13} \\ \vdots \\ \mathbf{y}'_{n-1,n} \end{bmatrix} \quad (9)$$

となる。以上の表現のもとに、一連の予測類似性 \hat{s}_{ij} (4) を \mathbf{s} と同様の順に並べた N 次元ベクトルは、

$$\hat{\mathbf{s}} = \mathbf{Y} \mathbf{r} \quad (10)$$

と表せる。従って、最小二乗基準(5)は、

$$\phi(\mathbf{r}) = \|\mathbf{s} - \mathbf{Y}\mathbf{r}\|^2 \quad (11)$$

と表せる。これは線形重回帰分析の問題に他ならず、 \mathbf{r} の最小二乗解は、

$$\hat{\mathbf{r}} = (\mathbf{Y}\mathbf{Y})^{-1}\mathbf{Y}\mathbf{s} \quad (12)$$

で与えられる。つまり、一連のパラメータの最小二乗解は解析的に求められる。

3. 非負対角モデル

特徴間類似性モデル(3)における同じ特徴どうしの類似性パラメータ r_{kk} ($k=1, \dots, m$)、つまり、特徴の自己類似性の意味を考えてみよう。2つの刺激が共通して特徴 k を持つとすると、これらの刺激間のモデル類似性には r_{kk} が加算されることになり、この点で特徴の自己類似性は、特徴ウェイトモデルの共通特徴モデルにおけるウェイトと同様の意味を持つと考えられる。このように考えると、特徴の自己類似性 r_{kk} ($k=1, \dots, m$) は0以上つまり非負であると仮定するのが自然であるといえる。そこで、特徴間類似性モデル(3)において、特徴の自己類似性パラメータに非負制約

$$r_{kk} \geq 0 \quad (k=1, \dots, m) \quad (13)$$

を導入する。自己類似性パラメータは(2)の行列 \mathbf{R} の対角要素をなすので、上記の非負制約を課したモデル(3)を特に非負対角モデルと呼ぶことにする。この非負対角モデルの解法の提示のために、次に幾つかの行列とベクトルを定義する。ここでも定義は煩雑であるが、解法は比較的簡潔である。

まず、非負制約を課す自己類似性パラメータと、無制約のその他の特徴間類似性パラメータに関するベクトルを別々に定義する。すなわち、(8)のパラメータ・ベクトル \mathbf{r} の要素を2つに分けて、自己類似性に関する m 次元ベクトル、および、他の特徴間類似性に加算定数を加えた $[m(m-1)/2+1]$ 次元ベクトルを、それぞれ、

$$\mathbf{r}_D = [r_{11}, r_{22}, \dots, r_{mm}]' \quad (14)$$

$$\mathbf{r}_O = [r_{12}, r_{13}, \dots, r_{1m}, r_{23}, r_{24}, \dots, r_{2m}, \dots, r_{m-1,m}, c]' \quad (15)$$

と表す。なお、 \mathbf{r}_D および \mathbf{r}_O に含まれる特徴間類似性パラメータはそれぞれ行列 \mathbf{R} の対角要素および非対角要素であるので、Diagonal (対角) および Off-diagonal (非対角) の頭文字を $\mathbf{r}_D, \mathbf{r}_O$ の添え字としている。これらのパラメータ・ベクトルと同様に、(7)に記す y_{ij} の要素を \mathbf{r}_D への対応部と \mathbf{r}_O への対応部に分けて、それぞれ、 m 次元および $[m(m-1)/2+1]$ 次元のベクトル

$$\mathbf{y}_{ij}^{(D)} = [y_{ij11}, y_{ij22}, \dots, y_{ijmm}]' \quad (16)$$

$$\mathbf{y}_{ij}^{(O)} = [y_{ij12}, y_{ij13}, \dots, y_{ij1m}, y_{ij23}, y_{ij24}, \dots, y_{ij2m}, \dots, y_{ij,m-1,m}, 1]' \quad (17)$$

によって表す。さらに、行ベクトル ($\mathbf{y}_{ij}^{(D)}$) および ($\mathbf{y}_{ij}^{(O)}$) を、 \mathbf{s} の要素の並びに対応させて各行に配置した $N \times m$ の行列および $N \times [m(m-1)/2+1]$ の行列を、それぞれ、 \mathbf{Y}_D および \mathbf{Y}_O と表す。自己類似性を考慮しない場合には、

$$\mathbf{s} = \begin{bmatrix} s_{12} \\ s_{13} \\ \vdots \\ s_{n-1,n} \end{bmatrix}, \mathbf{Y}_D = \begin{bmatrix} (\mathbf{y}_{12}^{(D)})' \\ (\mathbf{y}_{13}^{(D)})' \\ \vdots \\ (\mathbf{y}_{n-1,n}^{(D)})' \end{bmatrix}, \mathbf{Y}_O = \begin{bmatrix} (\mathbf{y}_{12}^{(O)})' \\ (\mathbf{y}_{13}^{(O)})' \\ \vdots \\ (\mathbf{y}_{n-1,n}^{(O)})' \end{bmatrix} \quad (18)$$

と表せる。以上の定義のもとに予測類似性ベクトルは、

$$\hat{\mathbf{s}} = \mathbf{Y}_D \mathbf{r}_D + \mathbf{Y}_O \mathbf{r}_O \quad (19)$$

と表せるので、最小化すべき基準は、

$$\eta(\mathbf{r}_D | \mathbf{r}_O) = \|\mathbf{s} - (\mathbf{Y}_D \mathbf{r}_D + \mathbf{Y}_O \mathbf{r}_O)\|^2 \quad (20)$$

となる。これを最小にする $\mathbf{r}_D, \mathbf{r}_O$ は一度に解析的には得られないので、交互最小二乗法を用いる。

まず、 \mathbf{r}_D を一定としたときの基準 $\eta(\mathbf{r}_O | \mathbf{r}_D)$ を最小にする \mathbf{r}_O は、

$$\eta(\mathbf{r}_O | \mathbf{r}_D) = \|(\mathbf{s} - \mathbf{Y}_D \mathbf{r}_D) - \mathbf{Y}_O \mathbf{r}_O\|^2 \quad (21)$$

と表せることから (線形重回帰分析に他ならず)、

$$\mathbf{r}_O = (\mathbf{Y}_O' \mathbf{Y}_O)^{-1} \mathbf{Y}_O' (\mathbf{s} - \mathbf{Y}_D \mathbf{r}_D) \quad (22)$$

で与えられる。

\mathbf{r}_O が所与のときの \mathbf{r}_D の推定は、 \mathbf{r}_D の各要素に非負制約(13)があるため、個々の要素ごとに行

う。すなわち r_D からいずれか 1 つの要素 r_{KK} を除いた $(m-1)$ 次元ベクトルを $r_D^{(KK)}$ と表すと、 r_0 , $r_D^{(KK)}$ を一定としたときの基準 $\eta(r_{KK}|r_0, r_D^{(KK)})$ を最小にする r_{KK} は、以下のようにして求められる。まず、

$$b_{ij}^{(KK)} = s_{ij} - r_0' y_{ij}^{(O)} - \sum_{k \neq K} r_{kk} y_{ijkk} \quad (23)$$

とおくと、(5)より、最小化すべき基準は

$$\eta(r_{KK}|r_0, r_D^{(KK)}) = \sum_{(i,j)} (b_{ij}^{(KK)} - r_{KK} y_{ijkk})^2 \quad (24)$$

のように r_{KK} の二次関数となり、無制約の解は、 $d\eta(r_{KK}|r_0, r_D^{(KK)})/dr_{KK} = 0$ より

$$\tilde{r}_{KK} = \sum_{(i,j)} b_{ij}^{(KK)} y_{ijkk} / \sum_{(i,j)} y_{ijkk}^2 \quad (25)$$

で与えられる。従って、非負制約を課した解は、

$$\hat{r}_{KK} = \begin{cases} \tilde{r}_{KK} & (\tilde{r}_{KK} \geq 0) \\ 0 & (\tilde{r}_{KK} < 0) \end{cases} \quad (26)$$

で与えられる。(26)によって \hat{r}_{KK} が得られれば、 r_0 の第 K 要素をこのに置き換える。以上のステップを K を 1 から m まで変えて順次行う。

r_D を適当な初期値に設定した上で、ここまで記した r_0 および r_D の推定を、収束したと判断されるまで、交互に繰り返せばよい。

4. ウェイトつきモデル

特徴の自己類似性がウェイトに対応するという観点のもとに、前項の非負対角モデルでは自己類似性パラメータに非負制約を課したが、このアプローチは、特徴のウェイトと特徴間類似性を分離していない点で、両者を同時に考慮する目的に照らせば部分的解決に過ぎない。すなわち、ウェイトを各特徴に付与される顕著性であると考えると、こうしたウェイトと特徴間関係である類似性とを区別して考える方が自然であるといえよう。そこで、特徴 k のウェイトを $w_k (k=1, \dots, m)$ で表した上で、 w_k と同時に特徴 k

と他の特徴 l の類似性 r_{kl} も考慮するモデルを考え、これをウェイトつき特徴間類似性モデルと呼ぶことにする。このモデルでは、刺激 i, j の対の特徴 k と l の結合の強さは、前項までのモデルで考慮した $r_{kl} x_k x_l$ に、さらに、各特徴のウェイト w_k と w_l を乗じたものに等しいと仮定する。これを x_{ik} が二値の場合について例示したのが図 2 である。この図では、刺激 i, j がそれぞれ特徴 1 および 2 を有するが、両特徴間の類似性 r_{12} に、各特徴のウェイト w_1, w_2 を乗じたものが結合の強さになることを示している。そして、こうした結合の強さの合計が刺激間類似性に等しいと考える。

以上のウェイトつき特徴間類似性モデルを式で表現すると、

$$\phi_{ij} = \sum_{k=1}^m \sum_{l=1}^m w_k w_l r_{kl} x_{ik} x_{jl} \quad (27)$$

と表せる。ただし、ウェイトは非負

$$w_k \geq 0 \quad (k=1, \dots, m) \quad (28)$$

とする。また、ウェイト・パラメータと類似性パラメータの分離した推定を可能にするため、全ての特徴の自己類似性について

$$r_{kk} = 1 \quad (k=1, \dots, m) \quad (29)$$

と制約する。(14),(15)に定義する特徴間類似性のベクトル r_D および r_0 に基づくと、この制約は r_D の要素は全て 1 と固定することであり、 r_0

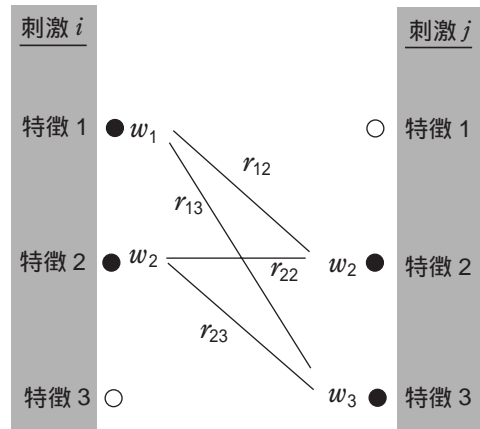


図 2 ウェイトつき特徴間類似性モデル

および w_1, \dots, w_m が推定対象となる。(29)を考慮し(6)で定義する y_{ijkl} を用いると、モデル類似性(3)は、

$$\phi_{ij} = \sum_{k,l>k} w_k w_l r_{kl} y_{ijkl} + \sum_{k=1}^m w_k^2 y_{ijkk} \quad (27)$$

と書き換えられる。これに加算定数 c を加えたものを予測類似性 $\hat{s}_{ij} = \phi_{ij} + c$ として、二乗誤差の総和

$$\eta = \sum_{(i,j)} (s_{ij} - \hat{s}_{ij})^2 = \sum_{(i,j)} \left[s_{ij} - \left(\sum_{k,l>k} w_k w_l r_{kl} y_{ijkl} + \sum_{k=1}^m w_k^2 y_{ijkk} + c \right) \right]^2 \quad (28)$$

を最小にすることが解析の目的となる。すなわち、(15)で定義した r_0 と一連のウェイト $\mathbf{w} = [w_1, \dots, w_m]'$ の最小二乗解を求めることが目的となるが、このために交互最小二乗法を用いる。

まず、一連のウェイト \mathbf{w} を一定にしたときの基準 $\eta(\mathbf{r}_0 | \mathbf{w})$ を最小にする \mathbf{r}_0 は、以下のようにして解析的に求められる。まず、互いに異なる2つの特徴のウェイトの積 $w_1 w_2, w_1 w_3, \dots, w_1 w_m, w_2 w_3, w_2 w_4, \dots, w_2 w_m, \dots, w_{m-1} w_m$ を対角要素に配置した後に、最後の対角要素を1とした $[m(m-1)/2+1]$ 次の対角行列を

$$D_{ww} = \begin{bmatrix} w_1 w_2 & 0 & \dots & 0 \\ 0 & w_1 w_3 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ \vdots & \vdots & 0 & w_{m-1} w_m \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \quad (29)$$

と表し、また、各特徴のウェイトの二乗を要素とする m 次元ベクトルを

$$\mathbf{w}^{(2)} = [w_1^2, w_2^2, \dots, w_m^2]' \quad (30)$$

と表記すると、(18)に示す行列 Y_0 および Y_D を用いて、一連の予測類似性を並べたベクトルは、

$$\hat{\mathbf{s}} = Y_0 D_{ww} \mathbf{r}_0 + Y_D \mathbf{w}^{(2)} \quad (31)$$

と表せる。しかるに、最小化すべき基準は、

$$\eta(\mathbf{r}_0 | \mathbf{w}) = \|(\mathbf{s} - Y_D \mathbf{w}^{(2)}) - Y_0 D_{ww} \mathbf{r}_0\|^2 \quad (32)$$

と表せ(これも線形重回帰分析の問題であり)、

求める解は、

$$\hat{\mathbf{r}}_0 = (D_{ww} Y_0' Y_0 D_{ww})^{-1} D_{ww} Y_0' (\mathbf{s} - Y_D \mathbf{w}^{(2)}) \quad (33)$$

と与えられる。

次に、 \mathbf{r}_0 を一定としたときの \mathbf{w} の推定は非負制約(28)があるので、ウェイト $w_k (k=1, \dots, m)$ の1つずつについて推定を繰り返す。すなわち、 \mathbf{r}_0 、および、 \mathbf{w} から1つの要素 w_K を除いた $(m-1)$ 次元ベクトル $\mathbf{w}_{(K)}$ を一定にしたときの基準 $\eta(w_K | \mathbf{r}_0, \mathbf{w}_{(K)})$ を最小にする w_K を求めるステップを、 K を1から m まで順次変えながら繰り返す。このために、基準を

$$\begin{aligned} \eta(w_K | \mathbf{r}_0, \mathbf{w}_{(K)}) &= \sum_{(i,j)} \left(s_{ij} - \sum_{k,l>k} w_k w_l r_{kl} y_{ijkl} - \sum_{k=1}^m w_k^2 y_{ijkk} - c \right)^2 \\ &= \sum_{(i,j)} \left(s_{ij} - \sum_{\substack{k,l>k \\ k \neq K, l \neq K}} w_k w_l r_{kl} y_{ijkl} - \sum_{k=K} w_K w_k r_{kK} y_{ijkK} - \sum_{l=K} w_K w_l r_{Kl} y_{ijKl} \right. \\ &\quad \left. - w_K^2 y_{ijKK} - \sum_{k \neq K} w_k^2 y_{ijkk} - c \right) = \sum_{(i,j)} (d_{ij}^{(K)} - q_{ij}^{(K)} w_K - y_{ijKK} w_K^2)^2 \end{aligned} \quad (34)$$

と表現する。ここで、

$$d_{ij}^{(K)} = s_{ij} - \sum_{\substack{k,l>k \\ k \neq K, l \neq K}} w_k w_l r_{kl} y_{ijkl} - \sum_{k \neq K} w_k^2 y_{ijkk} - c \quad (35)$$

$$q_{ij}^{(K)} = \sum_{k>K} w_k r_{kK} y_{ijkK} + \sum_{l>K} w_l r_{Kl} y_{ijKl} \quad (36)$$

である。(34)は w_K の四次関数であり、これを展開すると、

$$\eta(w_K | \mathbf{r}_0, \mathbf{w}_{(K)}) = b_1 w_K^4 + b_2 w_K^3 + b_3 w_K^2 + b_4 w_K + \sum_{(i,j)} d_{ij}^{(K)2} \quad (37)$$

と表せる。ここで、各係数は次のとおりである。

$$b_1 = \sum_{(i,j)} y_{ijKK}^2, \quad b_2 = 2 \sum_{(i,j)} q_{ij}^{(K)} y_{ijKK},$$

$$b_3 = \sum_{(i,j)} (q_{ijR}^{(K)2} - 2d_{ij}^{(K)} y_{ijKK}), \quad b_4 = -2 \sum_{(i,j)} d_{ij}^{(K)} q_{ij}^{(K)}. \quad (38)$$

一般に、(34)の四次関数は、1つの極小値を持つ場合と、2つの極小値を持つ場合に分けられる。前者の場合に1つの極小値を \tilde{w}_K で表すと、非負制約のもとでの解は、

$$\hat{w}_K = \begin{cases} \tilde{w}_K & (\tilde{w}_K \geq 0) \\ 0 & (\tilde{w}_K < 0) \end{cases} \quad (39)$$

で与えられる。後者の場合に2つの極小値を $\hat{w}_K^{(1)}, \hat{w}_K^{(2)}$ で表し、 $\hat{w}_K^{(1)} < \hat{w}_K^{(2)}$ とすると幾つかのケースが考えられ、非負制約のもとでの解は、

$$\hat{w}_K = \begin{cases} 0 & (\tilde{w}_K^{(2)} \leq 0) \\ 0 & (\tilde{w}_K^{(1)} \leq 0 < \tilde{w}_K^{(2)} \text{ かつ } \eta(0 | r_0, w_{(k)}) \leq \eta(\tilde{w}_K^{(2)} | r_0, w_{(k)}) \\ \tilde{w}_K^{(2)} & (\tilde{w}_K^{(2)} \leq 0 < \tilde{w}_K^{(1)} \text{ かつ } \eta(0 | r_0, w_{(k)}) > \eta(\tilde{w}_K^{(2)} | r_0, w_{(k)}) \\ \tilde{w}_K^{(1)} & (\tilde{w}_K^{(1)} > 0 \text{ かつ } \eta(\tilde{w}_K^{(1)} | r_0, w_{(k)}) \leq \eta(\tilde{w}_K^{(1)} | r_0, w_{(k)}) \\ \tilde{w}_K^{(1)} & (\text{上記以外}) \end{cases} \quad (40)$$

で与えられる。さて、上記の極小値あるいは $\hat{w}_K^{(1)}, \hat{w}_K^{(2)}$ は、(37)を w_K で微分したものを0とおいて得られる三次方程式

$$\frac{d\eta(w_K | r_0, w_{(k)})}{dw_K} = 4b_1 w_K^3 + 3b_2 w_K^2 + 2b_3 w_K + b_4 = 0 \quad (41)$$

を解くことによって得られる。この解はカルダノ (Cardano) の方法 (例えば、森, 1990) によって求められる²⁾。

w の初期値を適当な値に設定した上で、ここまで記した r_0 および w の推定を収束したと判断されるまで、交互に繰り返せばよい。

5. モデル間の関係

ここまで特徴間類似性モデルとして、単純総和モデル(3)、これに制約(13)を課した非負対角モデル、そして、ウェイトつきモデル(27)を考察したが、これらのモデルはある種の条件のもとで同等になる。まず、単純総和モデルにおける特徴間の自己類似性が全て非負である場合には、これと非負対角モデルが同等になることは自明である。さらに、単純総和モデルの自己特徴類似性が全て正であるとき、これとウェイトつきモデルが同等になる。このことは、次のように、各モデルを行列表現すると容易に示すことができる。

全ての刺激対に関するモデル類似性をまとめ

て、 n 次の対称行列

$$\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1n} \\ \phi_{12} & \phi_{22} & \cdots & \phi_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{1n} & \phi_{2n} & \cdots & \phi_{nn} \end{bmatrix} \quad (42)$$

で表すと、特徴間類似性の単純総和モデルは

$$\Phi = XRX^T \quad (43)$$

と表せる。ここで、 X は x_{ij} を要素とする n (刺激) \times m (特徴) のデータ行列で、 R は(2)で定義した特徴間類似性行列である。一方、ウェイトつきモデルは、

$$\Phi = XD_w R_1 D_w X^T \quad (44)$$

と表せる。ここで、

$$D_w = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & w_m \end{bmatrix}, \quad R_1 = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1m} \\ r_{12} & 1 & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1m} & r_{2m} & \cdots & 1 \end{bmatrix} \quad (45)$$

である。

さて、特徴の自己類似性つまり R の対角要素 r_{kk} ($k=1, \dots, m$) が全て正であるとき、 r_{kk} の平方根を対角要素とする対角行列 $D_R^{1/2}$ およびその逆行列 $D_R^{-1/2}$ 、

$$D_R^{1/2} = \begin{bmatrix} r_{11}^{1/2} & 0 & \cdots & 0 \\ 0 & r_{22}^{1/2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & r_{nn}^{1/2} \end{bmatrix}, \quad D_R^{-1/2} = \begin{bmatrix} r_{11}^{-1/2} & 0 & \cdots & 0 \\ 0 & r_{22}^{-1/2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & r_{nn}^{-1/2} \end{bmatrix} \quad (46)$$

が定義できる。もちろん、両行列の積 $D_R^{1/2} D_R^{-1/2}$ および $D_R^{-1/2} D_R^{1/2}$ はともに単位行列 I_m に等しいので、

$$R = D_R^{1/2} D_R^{-1/2} R D_R^{-1/2} D_R^{1/2} \quad (47)$$

と表せる。ここで、右辺の真ん中の $(D_R^{-1/2} R D_R^{-1/2})$ の対角要素が全て1となることは容易に確認できる。さらに、 $D_R^{1/2}$ の対角要素は全て正であるので、

$$R_1 = D_R^{-1/2} R D_R^{-1/2}, \quad D_w = D_R^{1/2} \quad (48)$$

とおくことができる。すなわち、 R の対角要素が全て正であれば、

$$R = D_w R_1 D_w \quad (49)$$

のように、 R を D_w (ウェイト) と R_1 (全対角要素が1の特徴間類似性行列) に分解でき、単純総和モデルおよび非負対角モデルとウェイトつき

モデルは同等になる。

上記の性質より、単純総和モデルによる解析の結果、自己類似性つまり R の対角要素の推定値が全て正であれば、非負対角モデルはもちろん、ウェイトつきモデルの解析も改めて行う必要はなく、(48)によって特徴のウェイトと特徴間類似性パラメータが得られる。

ここで注意すべきことは、3つのモデルが同等になるのが、上記のケース (R の全対角要素が正のとき) だけに限られないことである。例えば、 $m=3$ として、

$$R = \begin{bmatrix} 4 & 0 & 3 \\ 0 & 0 & 0 \\ 3 & 0 & 9 \end{bmatrix}, D_w = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}, R_1 = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix} \quad (50)$$

のとき、 $R = D_w R_1 D_w$ つまり単純総和、非負対角およびウェイトつきモデルは同等になる (この場合には R_1 は一意に定まらず、上記以外の行列でも等号が成り立つ)。このように他にも同等関係が成り立つ条件はある。ただし、行列 ($D_w R_1 D_w$) の対角要素は $w_i^2 (i = 1, \dots, m)$ であるので、少なくとも R の対角要素が全て非負でなければ、上記の同等関係が得られないことは明らかである。さらに、この R の対角要素の非負条件から、3つの特徴間類似性モデルについて次の階層関係が導ける。まず、上述したように、単純総和モデル(43)において、 R の対角要素が非負であれば非負対角モデルと同等になり、非負対角モデルの中でも特に対角要素が全て正の場合や(50)に例示するような場合のときに、ウェイトつきモデルが導かれる。このことから、最も拘束の強い(パラメータの存在範囲が狭い)モデルがウェイトつきモデルであるのに対して、単純総和モデルは最も拘束が弱く、非負対角モデルは両モデルの中間に位置するといえる。

6. 平均モデル

前項まで述べてきた一連の特徴間類似性モデルに共通する点は、図1および図2に示すように特徴間の結合度を合計する点である。こうした合計によるモデルは、特徴ウェイトモデルにおける素データ・モデルと同様に、刺激が有する特徴数の影響を受けるといえる。例えば、 x_{ik} が二値つまり特徴が存在の有無によって定義される場合、それぞれ2個および3個の特徴を持つ2つの刺激間のモデル類似性は計6回の加算に基づくが、5個および6個の特徴を持つ2つの刺激間のモデル類似性は計30回の加算に基づく。こうした加算回数つまり刺激が有する特徴数がモデル類似性に影響することが望ましくないという考えに立てば、何らかの方法で、特徴数の影響を除去する必要がある。こうした影響を除去するため、結合の合計ではなく平均を利用し、単純総和モデル(3)の右辺を $x_i = \sum_{k=1}^m x_{ik}$ と $x_j = \sum_{k=1}^m x_{jk}$ の積つまり加算回数で除して、

$$\phi_{ij} = \frac{1}{x_i x_j} \sum_{k=1}^m \sum_{l=1}^m r_{kl} x_{ik} x_{jl} = \sum_{k=1}^m \sum_{l=1}^m r_{kl} z_{ik} z_{jl} \quad (51)$$

とするモデルが考えられる。ここで、

$$z_{ik} = \frac{1}{x_i} \sum_{k=1}^m x_{ik} \quad (52)$$

である。すなわち、ここまでの特徴間類似性モデルの一連の x_{ik} を z_{ik} に換えるわけである。(51)に加算定数 c を加えると、予測類似性は

$$\hat{s}_{ij} = \sum_{k=1}^m \sum_{l=1}^m r_{kl} z_{ik} z_{jl} + c \quad (53)$$

と表される。以上のモデルを特徴間類似性の平均モデルと呼び、前項までの一連のモデルを合計モデルと呼んで区別する。

さて、平均モデルは、パラメータが一意に定まらないという性質を持つ。すなわち、全ての特徴間類似性パラメータに一定の定数を加えても、加算定数からこの定数を減じれば予測類似性(52)は不変である。この性質は次のようにし

て示される。任意の定数 λ を用いて

$$r_{kl}^* = r_{kl} + \lambda \quad (k=1, \dots, m; l=1, \dots, m), \quad c^* = c - \lambda \quad (54)$$

とおき、これらを新たにパラメータとすると、

$$\hat{s}_{ij}^* = \sum_{k=1}^m \sum_{l=1}^m z_{ik} z_{jl} r_{kl}^* + c^* = \sum_{k=1}^m \sum_{l=1}^m z_{ik} z_{jl} r_{kl} + \lambda \sum_{k=1}^m \sum_{l=1}^m z_{ik} z_{jl} + c - \lambda \quad (55)$$

となるが、ここで、

$$\sum_{k=1}^m \sum_{l=1}^m z_{ik} z_{jl} = \frac{1}{x_i x_j} \sum_{k=1}^m \sum_{l=1}^m x_{ik} x_{jl} = \frac{1}{x_i x_j} \sum_{k=1}^m x_{ik} \sum_{l=1}^m x_{jl} = 1 \quad (56)$$

より、

$$\hat{s}_{ij}^* = \sum_{k=1}^m \sum_{l=1}^m z_{ik} z_{jl} r_{kl} + \lambda + c - \lambda = \hat{s}_{ij} \quad (57)$$

となり、予測類似性は不変である。

以上の性質より、単純総和モデルと非負対角モデルおよびウェイトつきモデルの区別が意味をなさないと言えよう。すなわち、単純総和モデルで推定された特徴の自己類似性の中に負のものがあったとしても、十分に大きな定数1を加えれば（加算定数から1を減じることによって）、予測類似性を変えることなく自己類似性を非負にでき、単純総和モデルを非負対角モデルに変えることができる。同様にして、自己類似性を正にして、前項に記したように、ウェイトつきモデルが導ける。

平均モデルをそのまま利用して何らかの方法で最小二乗解を求めても、上記のように、それらに定数を加えたものも同様に最小二乗解であるので、解析結果の明瞭な解釈ができない。こうした平均モデルにおけるパラメータの不定性は、一連のパラメータの中のいずれか1つを一定の数値に固定することによって除去できる。ここで、どのパラメータを固定するかが問題となるが、加算定数を固定するのが1つの方法であろう。例えば、 $c = \bar{s}$ （類似性データの平均値）として、予測類似性を

$$\hat{s}_{ij} = \phi_{ij} + \bar{s} \quad (58)$$

とする方法が考えられる。これに基づく最小二乗基準は、 $s_{ij}^* = s_{ij} - \bar{s}$ とおくと、

$$\eta = \sum_{(i,j)} (s_{ij}^* - \phi_{ij})^2 \quad (59)$$

と表せ、平均からの偏差値に変換された類似性データ s_{ij}^* にモデル類似性 ϕ_{ij} を対応させるアプローチといえる。ここで、 s_{ij}^* が0を中心に負から正の値をとることから、「 s_{ij}^* は、類似・非類似のいずれでもない中庸を表す0を境に、負の値はむしろ非類似、正の値はむしろ類似を表す」と意味付ければ、こうしたデータから推定される r_{kl} にも同様の意味付けを与えられることが期待でき、解釈が促されることもありえよう。ただし、 $s_{ij}^* = 0$ が必ずしも中庸を表している根拠はなく、上記のことはあくまで期待である。

平均モデルのもとで(59)を最小にするパラメータの推定には、2～4項に述べた方法に若干の修正を加えれば、本質的に同じ方法が利用できる。修正点の1つは、類似性データ s_{ij} および特徴データ x_{ik} をそれぞれ s_{ij}^* および z_{ik} に置き換えることである。もう1つの修正点は一連の式の定義から加算定数 c および c に対応する部分を除くことである。以上の置換および削除を行えば、ここまでと同じアルゴリズムで最小二乗解が得られる。

7. 数値例

71種の文献（刺激）間の類似性および特徴データ（足立・田中，1995）に、一連の特徴間類似性モデルを適用し、最小二乗解を求めた。ここで、特徴データは、各文献の13種のキーワード（特徴；表2参照³⁾）を1か0で表す2値変数である。平均モデルを適用する際には、前節に記したように、加算定数を類似性データの平均値に固定した。

一連の類似性データ s_{ij} が、互いに独立に、平均が \hat{s}_{ij} 、分散が未知の定数 σ^2 である正規分布に従うと仮定して、解析結果の各モデルのAkaike（1974）の情報量規準（AIC）を求めた。

表1. 特徴間類似性モデルのAIC (括弧内は有効パラメータ数)

モデル	合計 ($q=93$)			平均 ($q=92$)		
	単純総和	非負対角	ウェイトつき	単純総和	非負対角	ウェイトつき
AIC	2048.5	2048.5	2048.5	1911.0	1911.1	1921.7

その結果を表1に示す。合計モデルでは、単純総和および非負対角モデルともに、特徴間類似性パラメータ $m(m+1)/2$ 個と加算定数および分散が推定対象となり、AICの算出に必要となる有効パラメータ数は、

$$\varphi = \frac{m(m+1)}{2} + 2 \quad (60)$$

となる。ウェイトつきモデルについても、 $m(m-1)/2$ 個の特徴間類似性、 m 個のウェイト、加算定数および分散がパラメータとなり、有効パラメータ数は(60)に等しくなる。平均モデルでは加算定数を事前に固定するので、有効パラメータ数は(60)より1つだけ少ない。なお、合計および平均モデルのそれぞれについて、単純総和、非負対角およびウェイトつきモデルの有効パラメータ数が等しいのに対して、パラメータの存在範囲が上記の順で狭くなっていく。従って、各モデルのAICの値には常に、「単純総和モデルのAIC ≤ 非負対角モデルのAIC ≤ ウェイトつきモデルのAIC」の関係が成り立つ。

表1から、合計モデルより平均モデルの方が妥当性が高く、刺激の特徴数の影響を抑えることの有効性が窺われる。同じデータに、特徴ウェイトモデルに基づく足立(1998)の回帰分析を適用すると、共通特徴、示差特徴、および、コントラストモデルのAICは、それぞれ、2801.4($q=15$)、2808.8($q=15$)、2631.5($q=16$)であった。これらより、特徴間類似性のいずれのモデルのAICも遥かに値が小さい。このことから、特徴間類似性を考慮することの重要性がわかる。

平均モデルの中でも、単純総和モデルの解析

結果では、特徴間の自己類似性に負の推定値が見られたので、これに非負制約を課した非負対角モデルのパラメータの推定値を表2に示す。表2で正の高い値を示す特徴どうしは互いに似ており、また、負の値を示す特徴どうしは、むしろ非類似の関係にあると見なされる。表を見渡すと比較的大きい値が散見され、例えば、特徴「EMアルゴリズム」を取り上げると自己類似性「3.63」よりも、他の特徴「潜在クラスモデル」との類似性の方が高い値「5.08」を示している。こうした点からも、特徴間類似性が刺激間類似性に大きな影響を与えることが窺われる。

8. 考察

本研究の目的は、1) 特徴間関係が刺激間の類似性に及ぼす効果を考慮して、特徴間類似性をパラメータとしたモデルを考案すること、および、2) データからモデルのパラメータを推定するためのアルゴリズムを考案することであった。1) については、3種の特徴間類似性モデルを考えた。すなわち、パラメータの単純加算によって類似性を表現する単純総和モデル、特徴の自己類似性のパラメータを0以上に制約する非負対角モデル、および、各特徴のウェイトを考慮した特徴ウェイトモデルを提案した。これらのモデルは相互に関連し、5節で論じたように、単純総和モデル、非負対角モデル、特徴ウェイトモデルの順にパラメータへの拘束が強くなるという階層関係がある。2) については、各モデルについて、最小二乗法によって

表 2 . 非負対角（平均）モデルにおけるパラメータの推定値（ \hat{R} ）

特 徴	数量化	MDS	クラス ター	類似 性	因子 分析	共分散 構造	IRT	潜在 クラス	最小 二乗法	最尤法	ベイズ 法	EM	質的 データ
数量化法	14.80	3.88	-3.97	0.82	-2.07	-5.01	-0.49	-5.47	-1.58	-3.41	-5.94	-3.62	0.37
多次元尺度法 (MDS)	3.88	9.82	1.62	3.01	-4.09	-6.12	-2.50	-1.36	-0.02	-0.98	-4.76	-3.30	-0.48
クラスター分析	-3.97	1.62	11.62	-1.63	-5.51	-2.54	-4.47	-1.01	-4.07	-3.06	-3.72	-1.07	-3.82
類似性	0.82	3.01	-1.63	3.70	-0.02	-2.85	-4.96	-3.02	0.80	-0.92	-1.42	-1.90	-0.86
因子分析	-2.07	-4.09	-5.51	-0.02	9.63	-2.03	-2.66	-1.79	2.54	0.09	1.66	-0.66	0.27
共分散構造分析	-5.01	-6.12	-2.54	-2.85	-2.03	16.56	1.67	4.55	-0.92	1.55	1.98	4.51	-1.57
項目反応理論 (IRT)	-0.49	-2.50	-4.47	-4.96	-2.66	1.67	11.04	4.92	-2.27	-0.12	3.99	2.55	-0.20
潜在クラスモデル	-5.47	-1.36	-1.01	-3.02	-1.79	4.55	4.92	1.41	-1.85	1.84	2.63	5.08	1.73
最小二乗法	-1.58	-0.02	-4.07	0.80	2.54	-0.92	-2.27	-1.85	0.71	-0.27	-1.49	-1.72	0.88
最尤法	-3.41	-0.98	-3.06	-0.92	0.09	1.55	-0.12	1.84	-0.27	2.87	2.32	2.71	0.81
ベイズ法	-5.94	-4.76	-3.72	-1.42	1.66	1.98	3.99	2.63	-1.49	2.32	7.94	3.65	-0.18
EM アルゴリズム	-3.62	-3.30	-1.07	-1.90	-0.66	4.51	2.55	5.08	-1.72	2.71	3.65	3.63	0.07
質的データ	0.37	-0.48	-3.82	-0.86	0.27	-1.57	-0.20	1.73	0.88	0.81	-0.18	0.07	0.00

パラメータの推定値を求めるアルゴリズムを考案した。特徴間類似性モデルは、特徴ウェイトモデルに比べて、特徴と特徴の間に定義される類似性を考慮する分だけ推定すべきパラメータ数は多くなるが、アルゴリズムはさほど複雑なものではない。すなわち、無制約の単純総和モデルのパラメータは解析的に求めることができ、また、非負対角モデルや特徴ウェイトモデルのアルゴリズムは反復計算に基づくものの、反復の各ステップの演算は比較的簡潔であり、実用に耐えうるものと考えられる。数値例では、特徴間類似性モデルが特徴ウェイトモデルより高い適合度を示し、特徴間関係を考慮することの必要性が示唆されたといえる。

特徴間類似性の平均モデルについては、パラメータの不定性を除去するため、加算定数を類似性データの平均値に固定する方法を取り上げ、適用例でもこの方法を用いた。しかしながら、この方法は一例に過ぎず、他に合理的な数値があれば、加算定数をそれに等しいと固定することも考えられる。さらに、そもそも不定性

を除去せずに、全てのパラメータを未知として解析を行うことも可能である。この場合、解析結果の一義的な解釈は難しいが、 x_{ik} を z_{ik} に換えさえすれば、3 および 4 項に記したアルゴリズムをそのまま利用して最小二乗解の 1 つを求めることができる。ただし、2 項に記す単純総和モデルの解法は逆行列を要するものであるが、平均モデルではこの逆行列が定義できず、(12) 式では解が得られない。以下に、単純総和の平均モデルで全パラメータを未知とした場合の解法を補足しておく。

まず、(6) の x_{ik} を z_{ik} に換えて、

$$\tilde{y}_{ijkl} = \begin{cases} z_{ik} z_{jl} & (k = l) \\ z_{ik} z_{jl} + z_{il} z_{jk} & (k \neq l) \end{cases} \quad (61)$$

と定義し直せば、最小二乗基準を導くまでの議論は 3.5.1 項と同じである。すなわち、 $[m(m+1)/2+1]$ 次元のデータ・ベクトル $\tilde{y}_{ij} = [\tilde{y}_{ij11}, \tilde{y}_{ij12}, \dots, \tilde{y}_{ij1m}, \tilde{y}_{ij22}, \tilde{y}_{ij23}, \dots, \tilde{y}_{ij2m}, \dots, \tilde{y}_{ij,m-1}, \tilde{y}_{ijmm}, 1]'$ を、類似性データ・ベクトル s の要素の並びに対応させて、各行に配置した $N \times [m(m+1)/2+1]$ の行列を \tilde{Y} とおけば、予測類似性ベクトルは s

$= \tilde{Y}r$ と表せる。ここで $r=[r_{11}, r_{12}, \dots, r_{1m}, r_{22}, r_{23}, \dots, r_{2m}, \dots, r_{m-1,m-1}, r_{m-1}, r_{mm}, c]$ である。従って、最小二乗基準は

$$\phi(r) = \|s - \tilde{Y}r\|^2 \tag{62}$$

と表せる。ここで、 \tilde{Y} の階数が $[m(m+1)/2+1]$ 未満であることに注意しなければならない。すなわち、各行の最終列の要素を除く総和は、

$$\sum_{k,l \neq k} \tilde{y}_{ijkl} = \sum_{k=1}^m \sum_{l=1}^m z_{ik} z_{il} = \sum_{k=1}^m z_{ik} \sum_{l=1}^m z_{jl} = 1 \tag{63}$$

となって最終列の1に等しくなり、 \tilde{Y} の一連の列ベクトルは一次独立ではなく、 $\tilde{Y}\tilde{Y}$ は特異行列となり、逆行列が定義できないため、 $\tilde{Y}\tilde{Y}$ のムーアペンローズ逆行列 $(\tilde{Y}\tilde{Y})^+$ を利用しなければならない(例えば、足立, 2001)。すなわち、(62)の最小二乗解は

$$\hat{r} = (\tilde{Y}\tilde{Y})^+ \tilde{Y}s + [I_{M+1} - (\tilde{Y}\tilde{Y})^+ \tilde{Y}\tilde{Y}]b \tag{64}$$

で与えられる。ここで、 $M=m(m+1)/2$ 、 b は $(M+1)$ 次の任意のベクトルであり、解は複数個ある。ただし、 \tilde{Y} の階数が $(M+1)$ ではないが M であるとして、1を任意の定数とすると、(64)の右辺の第2項は、

$$[I_{M+1} - (\tilde{Y}\tilde{Y})^+ \tilde{Y}\tilde{Y}]b = \lambda \begin{bmatrix} 1_M \\ -1 \end{bmatrix} \tag{65}$$

となることが証明される4,5。従って、

$$\hat{r} = (\tilde{Y}\tilde{Y})^+ \tilde{Y}s \tag{66}$$

とおいた上で(65)を利用すると、最小二乗解(64)は

$$\hat{r} = \tilde{r} + \lambda \begin{bmatrix} 1_M \\ -1 \end{bmatrix} \tag{67}$$

と表せる。

さて、上記の最小二乗解の含意をみるため、 \hat{r} および \tilde{r} の要素をそれぞれ \hat{r}_{kl} 、 \tilde{c} 、および、 \tilde{r}_{kl} 、 \tilde{c} と表すと、(67)は、

$$\hat{r}_{kl} = \tilde{r}_{kl} + \lambda (k, l=1, \dots, m; k \leq l), \quad \hat{c} = \tilde{c} - \lambda \tag{68}$$

と表せ、(54)と同等の表現が得られる。すなわち、最小二乗解の1つは(66)で与えられるが、(68)によって変換したのもも解となるわけである。ここで特徴間類似性パラメータに着目する

と、ここまで繰り返したように、一意には定まらないが、その差は一意的に定まるといえる。すなわち、2種の特徴の対 (k,l) および (K,L) に関して、上記の議論より、

$$\hat{r}_{kl} - \hat{r}_{KL} = (\tilde{r}_{kl} + \lambda) - (\tilde{r}_{KL} + \lambda) = \tilde{r}_{kl} - \tilde{r}_{KL} \tag{69}$$

となり、パラメータの差については一定に保たれる。例えば、 $\hat{r}_{kl}=1$ 、 $\hat{r}_{KL}=-2$ という推定値が得られたとすると、この1や-2という数値そのものは意味を持たないが、両者の差をとり、 r_{kl} は r_{KL} より3だけ類似性が大きいということが出来る。すなわち、上記の特徴間類似性パラメータの推定値は、比率尺度ではなく、間隔尺度上の値であるといえよう。

特徴間類似性モデルの短所は、各特徴間の類似性パラメータの合計に基づくため、数多くの未知パラメータを持つことであろう。一般に、データ数に比較してパラメータが過多であると、その推定精度が劣化するといえる。さらに、パラメータ数の過多のため、明瞭な解析結果の解釈ができない点も短所である。すなわち、表2のように、特徴間類似性パラメータの推定値が m (特徴) \times m (特徴)のクロス表(行列)で表現され、こうした表から特徴間関係を把握することは容易ではない。こうした短所を改善するためには、例えば、特徴を点やベクトルによって表すといった、特徴間類似性モデルに空間的な構造を導入することが、一つの方策であると考えられる。

注

2) 一般に三次方程式の解は以下ようになる(ここで用いる記号は本文中の記号と無関係である)。 $x^3 + ax^2 + bx + c = 0$ において、 $x = y - a/3$ とおくと $y^3 + py + q = 0$ となる。ここで、 $p = b - \frac{a^2}{3}$ 、 $q = \frac{2a^3}{27} - \frac{ab}{3} + c$ である。 $\omega = \frac{-1+i\sqrt{3}}{2}$ 、 $\omega^2 = \frac{-1-i\sqrt{3}}{2}$ (i は虚数単位)とおくと、 $y^3 + py + q = 0$ を満たす y は $m+n$ 、 $\omega m + \omega^2 n$ および $\omega^2 m + \omega n$ で与えられる。ここで、 $m = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$ 、 $n = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$ である。

- 3) キーワードの意味の理解には、例えば、柳井（1994）や繁樹（1995）などが参考になる。
- 4) $\begin{bmatrix} \mathbf{1}_M \\ -1 \end{bmatrix}$ は第1,2,...,M要素が全て1で、第M+1要素が-1の(M+1)次元ベクトルである。
- 5) 一般に、Aがa次の対称べき等行列であり $\mathbf{AB} = \mathbf{0}$ の場合に、もし、 $\text{rank}\mathbf{A} + \text{rank}\mathbf{B} = a$ であれば、 $\mathbf{A} = \mathbf{I}_a - \mathbf{BB}^+$ になる(Magnus & Neudecker,1988, p35, theorem 9)。この定理を利用すればよい。すなわち、 $(\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}})^+(\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}})$ はM+1次の対称べき等行列であり、(63)より $(\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}})^+(\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}})\begin{bmatrix} \mathbf{1}_M \\ -1 \end{bmatrix} = \mathbf{0}$ である。さらに、 $\text{rank}(\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}})^+(\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}) + \text{rank}\begin{bmatrix} \mathbf{1}_M \\ -1 \end{bmatrix} = M+1$ より、 $(\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}})^+(\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}) = \mathbf{I}_{M+1} - \begin{bmatrix} \mathbf{1}_M \\ -1 \end{bmatrix} \begin{bmatrix} \mathbf{1}_M \\ -1 \end{bmatrix}^+$ となる。ここで、 $\begin{bmatrix} \mathbf{1}_M \\ -1 \end{bmatrix}^+ = \frac{1}{M+1} [\mathbf{1}'_{M+1} - 1]$ となることは容易に確認できる。しかるに、 $[\mathbf{I}_M - (\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}})^+(\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}})]\mathbf{b} = \frac{1}{M+1} \begin{bmatrix} \mathbf{1}_M \\ -1 \end{bmatrix} [\mathbf{1}'_{M+1} - 1]\mathbf{b}$ より、 $\lambda = [\mathbf{1}'_{M+1} - 1]\mathbf{b}/(M+1)$ とおけば(65)が与えられる。

文献

足立浩平（1998）. 特徴集合モデルに基づく類似性データの回帰分析．甲子園大学紀要 人間文化学部編 No. 2, 1-16 .

足立浩平（2001）. 文科系のための線形代数．甲子

園大学紀要 人間文化学部編 No. 5, 21-33 .

足立浩平・田中 博（1995）. 対象間の非類似性を外的基準としたカテゴリカル属性の尺度構成法. 行動計量学, 22, 110-125 .

Akaike (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716-723.

北條 弘（1982）. 日本語子音素の類似性 IND-SCALと数量化理論第 類による分析 . 心理学研究, 53, 72-79 .

Magnus, J.R., & Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics*. Chichester: Wiley.

森 克美（1990）. 非線形方程式．大野 豊・磯田和男（監修）数値計算ハンドブック（新版），pp. 589-688，オーム社 .

Shepard, R.N., & Arabie, P. (1979). Additive clustering: Representation of similarities as combination of discrete overlapping properties. *Psychological Review*, 86, 87-123.

繁樹算男（1995）. 意思決定の認知統計学．朝倉書店 .

Tversky, A (1977). Features of similarity. *Psychological Review*, 84, 327-352.

柳井晴夫（1994）. 多変量データ解析法 理論と応用 . 朝倉書店 .