

数理計画法による継次カテゴリデータの解析法

上田 徹
成蹊大学

(受理 2005年 12月 2日; 最受理 2007年 2月 8日)

和文概要 西里は双対尺度法の(カテゴリに順序のある)継次カテゴリデータへの適用法を工夫しているが,その定式化は面倒である.しかし,継次性は数理計画法の制約として捉えればその表現は容易なので,まず MS Excel のソルバーで解ける程度の簡易な数理計画法としての定式化について述べている.数理計画法の利点は目的関数や制約の修正,追加が極めて容易なことであり,その特徴を生かして様々な目的関数について検討している.ところで,回答者に評価対象を主観的に評価してもらう場合はその評価結果は確定値ではなくこの程度といった曖昧さを持っているであろう.そこで,双対尺度法において曖昧さをモデルに持ち込むためにファジィ数を用いることも提案している.三つの回答例を用いて本論文で提案するモデルの有効性を確認している.

キーワード: 数理計画、双対尺度法、ファジィ数、カテゴリデータ

1. はじめに

双対尺度法(西里 [1])は数量化理論第3類や符合解析法などと呼ばれている手法を拡張し,様々なデータに使えるように工夫されている.しかし,継次カテゴリデータ(「優」,「良」,「可」,「不可」のように反応カテゴリに順位のあるもの)に対する双対尺度法におけるモデル行列の作成は継次性を表現するための工夫がゆえに直感的理解が難しく,その作成が容易でない.しかし,継次性は数理計画法の制約として捉えればその表現は容易である.そこで,第1の目的として,数理計画法としての定式化を検討する.数理計画法の利点は目的関数や制約の修正,追加が極めて容易なことであり,その特徴を生かして様々な目的関数について検討する.

ところで,回答者に評価対象を主観的に評価してもらう場合はその評価結果は確定値ではなくこの程度といった曖昧さを持っていることがある.10点満点のうちの何点かを答えるとした場合に6点をつけるか7点をつけるか迷うであろう.また「優」,「良」,「可」のうちの「優」にするか「可」にするか迷う場合もあると思われる.そのような場合に,回答の中に自信のなさを告げてもらう(たとえば5点から8点の間のように区間で答えてもらうとか「優」が60%,「良」が40%のように隣り合うカテゴリに重みを付けさせる)ことが考えられる.たとえば,数量化理論II類や数量化理論I類における説明変数の曖昧さをファジィ数として捉えた和多田・田中・浅居 [4], [5]の研究がある.しかし,なぜ5点は含まれて4点は含まれないのか,なぜ8点は含まれて9点は含まれないのか「優」の重みは55%ではなくて60%なのかと言った具合に,詳しく述べてもらえばもらうほど曖昧さを含んでいる数字やカテゴリが新たな曖昧さを生み出してしまいかもしれない.そこで詳しい回答を求めるのではなく分析側で対処することを考える.先にコンジョイント分析法ではカテゴリ

に与える数値を三角型ファジィ数 (下限値 a_1 , 中央値 a_2 , 上限値 a_3) として取り込み, 和多田ら (1982,1983) とは違って a_1, a_2, a_3 も推定対象とすることによりその目的がほぼ実現できることを示した (上田 [3]). ここでは第 2 の目的として, 曖昧さを取り扱うために上田 [3] と同様のファジィ数を用いたモデルの検討も行う.

2. 扱うデータと, 対応する双対尺度法

ここでは既存の手法 (西里 [1]) を紹介する. N 人の回答者の各々が M 個の対象を評価する. ただし, 評価値は K 個のカテゴリの一つを選ぶこととし, 番号の大きいカテゴリほどよい評価値とする. カテゴリ k とカテゴリ $(k+1)$ の境界に与える値を τ_k とすると, $\tau_k \leq \tau_{k+1}$ を満たしてほしい ($1 \leq k \leq K-1$). また n 番目の対象 O_n に与える値を μ_n とすると, $\tau_k < \mu_n \leq \tau_{k+1}$ ならば O_n はカテゴリ $k+1$ に属するものと捉える. M 個の μ_n の値は回答を反映してほしい. そこで西里 (1984) は特別なデータ行列 $F = (f_{ij})$ の作り方を提案している. F は N 行 $\{M(K-1)\}$ 列で, その各行は対象ごとに $(K-1)$ 個の要素を与えられ, カテゴリ k に属すると答えられた場合には $(k-1)$ 以下のカテゴリの値は (-1) とし, k 以上のカテゴリの値は 1 とする. ただし, $(K-1)$ 個の要素がすべて (-1) のときにはカテゴリ K であったものとする. これにより,

$$f_{i,j(K-1)+h} = 1 : \text{人 } i \text{ は } \mu_j < \tau_h \text{ と評価}$$

$$-1 : \text{人 } i \text{ は } \mu_j > \tau_h \text{ と評価}$$

と考える. 3 個の対象 A,B,C があり, カテゴリ数 $K=3$ とする. i 番目の人が, A はカテゴリ 1, B はカテゴリ 2, C はカテゴリ 3 と評価したとき, F の第 i 行は

$$1 \quad 1 \quad | \quad -1 \quad 1 \quad | \quad -1 \quad -1$$

となる.

模型行列 A は $\{M(K-1)\}$ 行 $\{M+(K-1)\}$ 列であり,

$$A = \begin{bmatrix} I_m & -\mathbf{1}_m & \mathbf{0}_m & \cdots & \mathbf{0}_m \\ I_m & \mathbf{0}_m & -\mathbf{1}_m & \cdots & \mathbf{0}_m \\ I_m & \mathbf{0}_m & \mathbf{0}_m & \cdots & \mathbf{0}_m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ I_m & \mathbf{0}_m & \mathbf{0}_m & \cdots & -\mathbf{1}_m \end{bmatrix} \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ M \end{matrix} \tag{1}$$

$$m \quad m+1 \quad m+2 \quad \cdots \quad m+M$$

である ($m = K-1$). I_m は $m \times m$ の単位行列, $-\mathbf{1}_m, \mathbf{0}_m$ は m 個の要素がすべて (-1) , 0 の列ベクトルである.

$$\mathbf{x} = (\tau_1, \tau_2, \dots, \tau_{K-1}, \mu_1, \mu_2, \dots, \mu_M)^t$$

とすると, $A\mathbf{x}$ の最初の $m (=K-1)$ 行からなるブロックは

$$(\tau_1 - \mu_1, \tau_2 - \mu_1, \dots, \tau_{K-1} - \mu_1)^t$$

である. 同様に $A\mathbf{x}$ の h 番目のブロックは

$$(\tau_1 - \mu_h, \tau_2 - \mu_h, \dots, \tau_{K-1} - \mu_h)^t$$

である．すなわち， Ax の $(hm+i)$ 番目の要素は $(\tau_i - \mu_h)$ である．

これだけでも結構理解が難しいが， τ_k 間の比較， μ_n 間の比較結果を反映するために $N \times \{m-1 + M(M-1)/2\}$ の行列 F_a と $\{m-1 + M(M-1)/2\} \times (m+M)$ の行列 A_a を導入している． F_a の最初の $(m-1)$ 列は $\tau_k \leq \tau_{k+1}$ を反映して固定的にすべて (-1) の値を与えられ，残りの $\{M(M-1)/2\}$ 列の第 i 行は人 i の

$$(\mu_1 - \mu_2, \mu_1 - \mu_3, \dots, \mu_1 - \mu_M; \mu_2 - \mu_3, \mu_2 - \mu_4, \dots, \mu_M; \dots; \mu_{M-1} - \mu_M)$$

に関する評価結果を反映している．すなわち， F_a の第 i 行の $\{m-1 + h(2M-h-1)/2+j\}$ 番目 ($0 \leq h \leq M-2, 1 \leq j \leq M-h-1$) の値は

$$\mu_{h+1} < \mu_{h+1+j} \text{ ならば } -1$$

$$\mu_{h+1} = \mu_{h+1+j} \text{ ならば } 0$$

$$\mu_{h+1} > \mu_{h+1+j} \text{ ならば } 1$$

である． A_a の第 j ($\leq m-1$) 行の第 j 列は 1 ，第 $(j+1)$ 列は (-1) ，他の列はゼロであり， A_a の第 $\{m-1 + h(2M-h-1)/2+j\}$ 行の第 $(m+h+1)$ 列 ($0 \leq h \leq M-2, 1 \leq j \leq M-h-1$) は 1 ，第 $(m-1+h+1+j)$ 列は (-1) ，他の列はゼロである．たとえば

$$\mathbf{x} = (\tau_1, \tau_2, \tau_3, \mu_1, \mu_2, \mu_3)^t$$

で，人 i が $\mu_1 = \mu_3 < \mu_2$ と判断したとすると， F_a の第 i 行は $(-1, -1 \mid -1, 0, 1)$ であり，

$$A_a = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

である．

結局，

$$E = [F, F_a] \begin{bmatrix} A \\ A_a \end{bmatrix} \quad (2)$$

をつくり，

$$\mathbf{x} = (\tau_1, \tau_2, \dots, \tau_{K-1}, \mu_1, \mu_2, \dots, \mu_M)^t$$

を $E^t E$ の最大固有値に対応する固有ベクトル，回答者からなるベクトル

$$\mathbf{y} = (y_1, y_2, \dots, y_N)^t$$

を EE^t の最大固有値に対応する固有ベクトルとして求める． \mathbf{x} と \mathbf{y} の間には

$$\mathbf{x} = aE^t \mathbf{y}; \mathbf{y} = aE \mathbf{x} \quad (a: \text{スカラー}) \quad (3)$$

の関係があり、一方が求めれば他方も求まる。\$E^t E\$ の最大固有値に対応するベクトルを求めることは全分散一定*のもとで層(回答者)間分散に比例する \$x^t E^t E x\$ を最大化することに相当する。すなわち、\$N\$ 人の \$x\$ で重み付けられた評価データ \$E x\$ を使って \$N\$ 人を最もうまく判別できる \$x\$ を求めていることになる。

* ここでは \$x\$ のノルムを一定とすることと同じである。

3. 数理計画法としての定式化

表1(西里[1]の表4.4)のような簡単な例を用いて説明する。回答者 \$i\$ に \$y_i\$、対象 \$j\$ に \$x_j\$、カテゴリ \$k\$ (\$=1:\times, 2:\triangle, 3:\circ\$) に評価値 \$t_k\$ を与えることとし、回答者 \$i\$ の対象 \$j\$ に与えた評価値を \$e_{ij}\$ とする(節2では \$t_k\$ と \$t_{k+1}\$ の境界値 \$\tau_k\$ を求めていることに注意)。たとえば表1では、回答者1は対象A(=1)をカテゴリ2としているので \$e_{11} = t_2\$、対象B(=2)をカテゴリ3としているので \$e_{12} = t_3\$ となる。

回答者間の分散が最大になるよう、対象間の分散が最大になるよう考えると、以下のような定式化が考えられる。

$$\text{【定式化1】目的: } \max \sum_{i=1}^N (y_i - \mu)^2; \mu: \text{全体平均} \tag{4}$$

$$\text{s.t. } \sum_{i=1}^N y_i / N = \mu = 0; y_i = \sum_{j=1}^M e_{ij} / M \tag{5}$$

$$\sum_{i=1}^N \sum_{j=1}^M (e_{ij} - \mu)^2 / (NM) = 1 \tag{6}$$

$$t_1 \leq t_2 \leq \dots \leq t_K \tag{7}$$

ここで、回答者 \$i\$ がカテゴリ \$k\$ を選んだ対象数を \$h_{ik}\$ とすると、

$$y_i = \sum_{k=1}^K h_{ik} t_k / M \tag{8}$$

である。全分散 = 1 (式(6))のもとで \$y\$ の分散最大(式(4))を目的としている。制約は式(7)に凝縮されている。

$$\text{【定式化2】目的: } \max \sum_{j=1}^M (x_j - \mu)^2; \mu: \text{全体平均} \tag{9}$$

$$\text{s.t. } \sum_{j=1}^M x_j / M = \mu = 0; x_j = \sum_{i=1}^N e_{ij} / N \tag{10}$$

$$\sum_{i=1}^N \sum_{j=1}^M (e_{ij} - \mu)^2 / (NM) = 1 \tag{11}$$

$$t_1 \leq t_2 \leq \dots \leq t_K \tag{12}$$

全分散 = 1 (式(11))のもとで \$x\$ の分散最大(式(9))を目的としている。

どちらの定式化でも隣り合う \$t_k\$ の値が等しくなってしまうこともあり、

表 1: 回答例 1

人	対象				
	A	B	C	D	E
1	2	3	3	3	3
2	1	2	3	1	1
3	2	3	2	2	1
4	1	1	1	1	1
5	3	1	1	2	3
6	3	3	3	3	3
7	2	2	2	2	1
8	1	3	3	2	1
9	3	1	1	3	3
10	1	2	2	1	1
和	19	21	21	20	18

1 : × , 2 : △ , 3 : ○

$$t_k - t_{k-1} \geq C : \text{一定値 } (k \geq 2) \quad (13)$$

の制約を加えることも必要となる。

定式化 1 で得られた t_k を式 (10) の第 2 式に代入すれば x が得られ, 定式化 2 で得られた t_k を式 (5) の第 2 式に代入すれば y が得られる。ただし, 定式化 1 と定式化 2 で得られる t_k の値は一般には異なる。すなわち目的関数, 制約の作り方で様々な解が得られる可能性がある。次節では様々な目的関数を取り上げる。

4. 様々な目的関数

次のような様々な目的関数が考えられる。

前節では x に着目する場合と y に着目する場合とで異なる t_k の値が得られるため, それらを統合する目的関数として

$$(i) \max \sum_{j=1}^M (x_j - \mu)^2 / M + \sum_{i=1}^N (y_i - \mu)^2 / N \quad (14)$$

が考えられる。さらに, どちらかをより強調したい場合には

$$(ii) \max w_1 \sum_{j=1}^M (x_j - \mu)^2 / M + w_2 \sum_{i=1}^N (y_i - \mu)^2 / N \quad (15)$$

のように重み w_i を考慮してもよい。また,

$$(iii) \max \sum_{j=1}^M (x_j - \mu)^2 / M; \quad \sum_{i=1}^N (y_i - \mu)^2 / N \geq C_1 \quad (16)$$

$$(iv) \max \sum_{i=1}^N (y_i - \mu)^2 / N; \quad \sum_{j=1}^M (x_j - \mu)^2 / M \geq C_2 \quad (17)$$

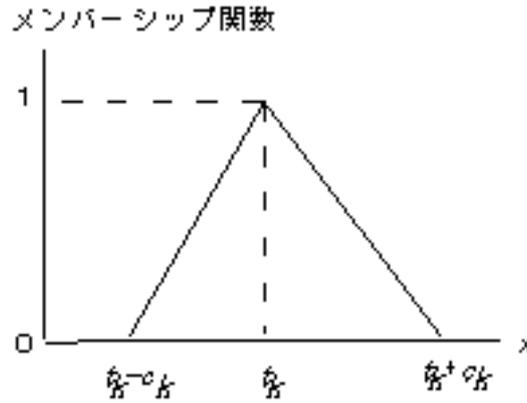


図 1: 三角型ファジィ数のイメージ図

のように，一方に確保すべき最小値 C_i を制約として，他方を最大化することも考えられる．

以上では人の区別をしようとするのに対し，人の評価になるべく差が出ないような評価をねらった

$$(v) \min \sum_{i=1}^N (y_i - \mu)^2 / N; \quad \sum_{j=1}^M (x_j - \mu)^2 / M \geq C_2 \quad (18)$$

も考えられる．

以上は x や y の分散に着目していたのに対し，評価値の差に着目した

$$(vi) \max \sum_k (t_{k+1} - t_k)^2 \quad (19)$$

も考えられる．

5. ファジィ数の利用

回答者を通じて確定的な値 t_k があるのではなく，それはもう少し曖昧なものと考えて， t_k を下限値 ($t_k - c_k$)，モード t_k ，上限値 ($t_k + d_k$) の三角型ファジィ数とする（メンバーシップ関数が三角形で表される．図 1 参照）．このとき，ファジィ数の比較に使われる代表通常数は図 2 から分かるように二つの台形の面積の平均で与えられ，

$$T_k = \{(t_k - c_k) + 2t_k + (t_k + d_k)\} / 4 \quad (20)$$

である（田中・松岡 [2]）．定式化 1 に対応して次のような定式化が考えられる．

$$\text{【定式化 F 1】} \max \left\{ \sum_{i=1}^N (y_i - \mu)^2 - \sum_{k=1}^K (c_k^2 + d_k^2) \right\}; \mu : \text{全体平均} \quad (21)$$

$$\text{s.t.} \quad \sum_{i=1}^N y_i / N = \mu = 0; y_i = \sum_{k=1}^K h_{ik} T_k / M \quad (22)$$

$$\sum_{k=1}^K H_k T_k^2 / (NM) = 1; H_k = \sum_{i=1}^N h_{ik} \quad (23)$$

$$T_1 \leq T_2 \leq \dots \leq T_K \quad (24)$$

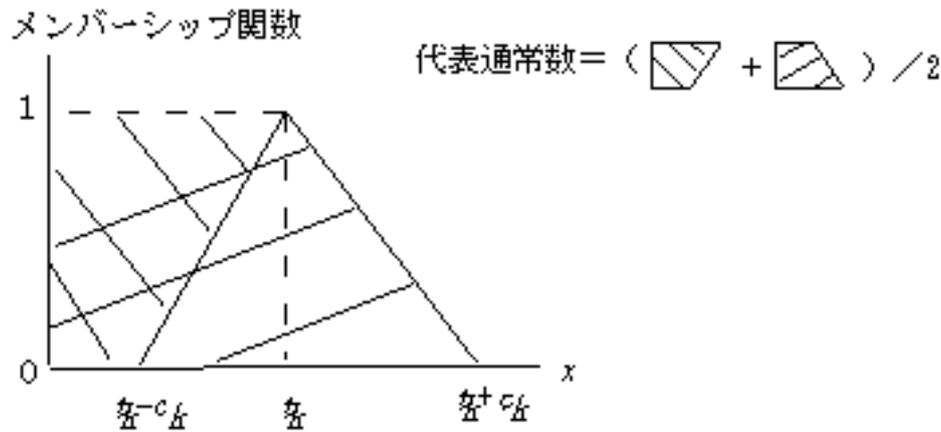


図 2: 代表通常数のイメージ図

ここで、三角型ファジィ数の二乗は三角型ファジィ数にならないことを考慮して T_k^2 の代りに

$$\begin{aligned} & [\int_0^1 \{t_k - c_k(1 - \alpha)\}^2 d\alpha + \int_0^1 \{t_k + d_k(1 - \alpha)\}^2 d\alpha] / 2 \\ & = t_k^2 - t_k(c_k - d_k) / 2 + (c_k^2 + d_k^2) / 6 \end{aligned} \quad (25)$$

を使うことも考えられる。

ここでも、隣り合う t_k の値が等しくなってしまうこともあり、

$$t_k - t_{k-1} \geq C : \text{一定値 } (k \geq 2) \quad (26)$$

の制約を加えることが必要となる。

式 (21) の第 2 項がゼロとなるような場合は、定式化 1 と一致する。なお、定式化 2 に対応する定式化 F2 も同様に得られるが、容易に類推できるのでここでは省略する。

6. 分析例

ここでの分析には MS Excel のアドインソフトであるソルバーを用いた。

(1) 回答例 1 の分析

表 1 に示した回答例 1 における y の分散 $V(y)$ 最大化基準のもとでの x の値と大きさの順位を表 2、 y の値と大きさの順位を表 3 に示す。また、 x の分散 $V(x)$ 最大化基準のもとでの x の値と大きさの順位を表 4、 y の値と大きさの順位を表 5 に示す。

表 2: 回答例 1 [定式化 1 : $V(y)$ 最大] における x の値と大きさの順位 : () 内の数字

	A	B	C	D	E	目的関数値
西里	-0.402(4)	0.440(1)	0.426(2)	-0.210(3)	-0.546(5)	
$C=0$	-0.125(4)	0.083(1)	0.083(1)	-0.125(4)	0.083(1)	4.792
$V(y)$ $C=0.2$	-0.121(5)	0.094(1)	0.094(1)	-0.101(4)	0.034(3)	4.774
最大 $C=0.4$	-0.117(5)	0.105(1)	0.105(1)	-0.077(4)	-0.015(3)	4.734

表 3: 回答例 1 [$V(y)$ 最大] における y の値と大きさの順位 : () 内の数字

	西里	$C=0$	$C=0.2$	$C=0.4$	カテゴリ値計
1	-0.070(5)	0.917(2)	0.938(2)	0.953(2)	14(2)
2	-0.413(10)	-0.333(6)	-0.357(7)	-0.379(7)	8(8)
3	-0.329(8)	-0.333(6)	-0.277(6)	-0.219(6)	10(4)
4	-0.106(6)	-0.750 (8)	-0.829(10)	-0.903(10)	5(10)
5	0.362(3)	0.083 (4)	0.074(4)	0.065(4)	10(4)
6	0.071(4)	1.333 (1)	1.329(1)	1.317(1)	15(1)
7	-0.255(7)	-0.750 (8)	-0.669(8)	-0.583(8)	9(7)
8	0.433(1)	0.083 (4)	0.074(4)	0.065(4)	10(4)
9	-0.387(9)	0.500 (3)	0.466(3)	0.429(3)	11(3)
10	0.407(2)	-0.750 (8)	-0.749(9)	-0.743(9)	7(9)
	$V(y)$	0.479	0.477	0.473	
	$V(x)$	0.010	0.009	0.008	
	$V(y)+V(x)$	0.490	0.486	0.482	

表 4: 回答例 1 [$V(x)$ 最大] における x の値と大きさの順位 : () 内の数字

	A	B	C	C	E	目的関数の値	
西里	-0.402(4)	0.440(1)	0.426(2)	-0.210(3)	-0.546(5)		
$V(x)$	$C=0$	-0.041(4)	0.165(1)	0.165(1)	0.165(1)	-0.453(5)	0.289
	$C=0.2$	-0.051(4)	0.163(1)	0.163(1)	0.143(3)	-0.418(5)	0.251
最大	$C=0.4$	-0.060(4)	0.160(1)	0.160(1)	0.120(3)	-0.381(5)	0.214

表 5: 回答例 1 [$V(x)$ 最大] における y の値と大きさの順位 : () 内の数字

西里	$C=0$	$C=0.2$	$C=0.4$	カテゴリー値計	
1	-0.070(5)	0.783(1)	0.824(2)	0.861(2)	14(2)
2	-0.413(10)	-0.453(8)	-0.458(8)	-0.461(8)	8(8)
3	-0.329(8)	0.371(3)	0.317(3)	0.261(3)	10(4)
4	-0.106(6)	-1.277(10)	-1.273(10)	-1.262(10)	5(10)
5	0.362(3)	-0.041(5)	-0.031(6)	-0.020(6)	10(4)
6	0.071(4)	0.783(1)	0.864(1)	0.941(1)	15(1)
7	-0.255(7)	0.371(3)	0.277(4)	0.181(4)	9(7)
8	0.433(1)	-0.041(5)	-0.031(6)	-0.020(6)	10(4)
9	-0.387(9)	-0.041(5)	0.009(5)	0.060(5)	11(3)
10	0.407(2)	-0.453(8)	-0.498(9)	-0.541(9)	7(9)
$V(y)$	0.355	0.369	0.383		
$V(x)$	0.058	0.050	0.043		
$V(y)+V(x)$	0.413	0.419	0.426		

表 3, 表 5 から分かるように $V(y)$ は $V(x)$ の約 50 倍であるため $\{V(y)+V(x)\}$ の最大化は $V(y)$ の最大化で実現された。すなわち, 本例では $V(y)$ 最大化と $\{V(y)+V(x)\}$ 最大化とは一致した。なお, 定式化 F1 においては $c_k = d_k = 0$ であったので定式化 1 と定式化 F1 とは結果が一致した。これは式 (21) の第 2 項が効きすぎているためと考えられる。そこで式 (21) の第 2 項を削除し, $c_k \leq |t_k|, d_k \leq |t_k|$ のような制約を設けることが考えられるが, その場合には制約が等号で成り立つ場合が多く, 制約に縛られた結果となり, 望ましくない。

本論文で提案したものは西里よりも表 1 の行和, 列和に近い。行和, 列和に近いことがよいのなら双対尺度法などといわず行和, 列和を人, モノの評価値とすればよいので本論文で提案したもののほうがよいことを意味しないが, 西里よりも「どのカテゴリーとしたか」をより重視していることになっていると言える。

表 6: 回答例 2: テレビ局の視聴調査結果

	NHK	教育	日本	TBS	フジ	朝日	東京
1	3	2	4	4	4	3	3
2	3	2	4	4	4	4	3
3	3	1	4	4	4	3	2
4	2	2	4	4	4	4	3
5	1	1	4	4	4	3	3
6	1	1	3	3	4	4	2
7	3	2	4	4	4	4	3
8	3	2	2	2	4	3	2
9	2	1	2	3	4	3	4
10	2	2	4	3	4	3	2

4: よく見る, 3: 見る, 2: あまり見ない, 1: 見ない

(2) 回答例 2 の分析

表 6 にテレビ局の視聴調査結果 (回答例 2) を示す。表 7 に x の値を示す。 $V(x)$ および $\{V(y)+V(x)\}$ の最大化では $C=0$ のときに ($t_{i+1} - t_i > 0.4$) だったので $C=0.2, 0.4$ のときも同じ結果だった。 x の値の順位は西里とほぼ同様の結果だが、本論文で示した方法は同じ値を取りやすいことが分かる。表 8 に 3 種の目的関数に対応した分散の値を示す。ここでは $V(x)$ が $V(y)$ より大きいとその差は回答例 1 より小さいため $\{V(y)+V(x)\}$ の最大化は $V(x)$ の最大化の結果と近かったが若干異なる結果となった。この例でも定式化 F 1 は有効ではなかった。

表 7: 回答例 2 における x の値

	NHK	教育	日本	TBS	フジ	朝日	東京
西里	-0.698	-1.421	0.947	0.913	1.424	0.665	-0.332
$C=0$	-0.408	-1.123	0.306	0.306	0.306	0.306	0.306
$V(y)$ $C=0.2$	-0.493	-1.252	0.365	0.365	0.465	0.345	0.205
最大 $C=0.4$	-0.568	-1.353	0.416	0.416	0.616	0.376	0.096
$V(x)$ 最大	-0.734	-1.370	0.527	0.505	1.063	0.366	-0.357
$V(x)+V(y)$ 最大	-0.729	-1.383	0.523	0.506	1.051	0.375	-0.343

表 8: 回答例 2 における分散の値

	max $V(y)$			max $V(x)$		
	$C=0$	$C=0.2$	$C=0.4$	$C=0$	$C=0.2$	$C=0.4$
$V(y)$	0.1667	0.1542	0.1406	0.1038	0.1038	0.1038
$V(x)$	0.2708	0.3506	0.4328	0.6199	0.6199	0.6199
$V(x)+V(y)$	0.4375	0.5047	0.5735	0.7238	0.7238	0.7238
	max $V(x)+V(y)$					
	$C=0$	$C=0.2$	$C=0.4$			
$V(y)$	0.1042	0.1042	0.1042			
$V(x)$	0.6197	0.6197	0.6197			
$V(x)+V(y)$	0.7240	0.7240	0.7240			

(3) 回答例 3 の分析

表 9 に自動車メーカー選好度調査結果 (回答例 3) を示す (大変興味がある「4」、興味がある「3」、あまり興味がない「2」、興味がない「1」)。定式化 F2 においては $c_k = d_k = 0$ であったので定式化 2 と定式化 F2 とは結果が一致した。しかし、定式化 F1 では、表 10 に示すように c_k, d_k にゼロでないものがあり、定式化 1 と一致しない。これは $N=25$ に対して $M=6$ と少ないことが原因の一つと考えられる。また、定式化 F1 では定式化 1 よりも変数が多いため $C=0.08$ でも解があったが、定式化 1 では $C=0.07$ で解がなかった。

表 9: 自動車メーカーの選好度

被験者	日産	トヨタ	マツダ	ホンダ	スバル	ダイハツ
1	4	4	3	3	1	1
2	2	4	2	4	4	2
3	4	4	4	4	4	1
4	1	4	3	3	4	3
5	2	4	2	4	2	1
6	3	4	1	4	1	1
7	4	4	4	4	2	1
8	1	3	3	2	2	2
9	3	4	1	2	1	1
10	4	4	3	4	1	1
11	4	3	1	4	1	1
12	2	4	1	4	1	1
13	2	3	4	3	2	1
14	2	1	1	1	1	1
15	2	3	2	3	2	2
16	3	3	1	4	2	2
17	3	3	1	2	2	2
18	3	4	2	4	2	1
19	4	4	2	3	2	2
20	4	4	2	4	1	1
21	1	1	1	1	1	1
22	4	4	3	4	1	1
23	4	4	1	4	1	1
24	3	4	2	3	4	2
25	1	4	1	1	1	1

4 : 大変興味がある , 3 : 興味がある , 2 : あまり興味がない , 1 : 興味がない

表 10: 変数の値 (回答例 3)

	$C=0.01$	$C=0.05$	$C=0.08$		$C=0.01$	$C=0.05$	$C=0.08$
t_1	0.063	0.096	0.124	c_3	0	0	0
t_2	0.053	0.046	0.044	c_4	0	0	0
t_3	0.043	-0.004	-0.036	d_1	0	0	0
t_4	-0.116	-0.112	-0.116	d_2	0.018	0.007	0
c_1	0.062	0.072	0.122	d_3	0.076	0.099	0.074
c_2	0	0	0.038	d_4	0	0	0.086

7. むすび

継次カテゴリデータの数理計画法に基づく分析法を提案し、回答例 1, 2 を通して西里との相違を明らかにした。回答例 3 では三角型ファジィ数が解になる場合を示した。西里の方法では x と y に双対性（一方が得られれば他方が得られる）があり、1 種類の解しか得られないが、ここで提案した方法では評価基準ごとに異なる結果となり、どの基準を使えばよいのか迷ってしまうという欠点がある。 $V(x)$ と $V(y)$ に差があるときには、 $\{V(y)+V(x)\}$ 最大化の結果はその大きい方にほぼ等しくなったので、 $\{V(y)+V(x)\}$ 最大化から得られた x, y を用いることにすれば統一した値が得られることとなる。

本論文の定式化は MS Excel のソルバーで解けるように、非常に簡易である。ただし、その初期値によっては局所解になることがあるので初期値を取り替えた解の導出も必要である。

条件 $\{t_k - t_{k-1} \geq C\}$ における C の値をどのように決めるのか、多次元の解は求められるのかは今後の課題である。

参考文献

- [1] 西里静彦：質的データの数量化（朝倉，1984），ただし，1982 年発行のものには掲載されていないことに注意。
- [2] A. Kaufmann and M.M. Gupta : *Fuzzy Mathematical Models in Engineering and Management Science* (Elsevier, 1988). 田中英夫 監訳・松岡浩訳：ファジィ数学モデル（オーム社，1992）.
- [3] 上田徹：コンジョイント分析における曖昧な回答の扱い方. オペレーションズ・リサーチ, 44 (1999), 496-502.
- [4] 和多田淳三，田中英夫，浅居喜代治：ファジィ数量化理論 II 類. 行動計量学, 9 (1982), 24-32.
- [5] 和多田淳三，田中英夫，浅居喜代治：ファジィ数量化理論 I 類. 行動計量学, 11 (1983), 66-73.

付録 μ_j と τ_k の両方を用いた数理計画法による定式化例

西里が設けた枠組みに合わせてカテゴリ k とカテゴリ $(k+1)$ の境界に与える値を τ_k とし、 n 番目の対象 O_n に与える値を μ_n とし、数理計画法による定式化を考える。人 i が対象 O_n に持っている評価値を v_{in} とする。当然、 μ_n と v_{in} とは一致せず、誤差

$$e_{in} = |\mu_n - v_{in}| \tag{A.1}$$

が存在する。人 i が、「対象 O_n はカテゴリ k に属している」と評価すると

$$\tau_{k-1} \leq e_{in} \leq \tau_k \tag{A.2}$$

でなければならない。ただし、 $\tau_0 = 0$ とする。

このとき、次のような定式化が考えられる。

$$\text{【定式化 A1】 目的：} \min \sum_{i,n} e_{in} \quad (\text{A.3})$$

$$\text{s.t. } e_{in} = |\mu_n - v_{in}| \quad i=1,2,\dots,N; n=1,2,\dots,M \quad (\text{A.4})$$

$$\tau_{k-1} \leq e_{in} \leq \tau_k \quad i=1,2,\dots,N; n=1,2,\dots,M \quad (\text{A.5})$$

$$\tau_0 = 0, \tau_k - \tau_{k-1} = c > 0 \quad k=1,2,\dots,K-1 \quad (\text{A.6})$$

$$\sum_{j=1}^M \mu_j = d > 0 \quad (\text{A.7})$$

$$\mu_j, \tau_k, v_{in}, e_{in} \geq 0 \quad (\text{A.8})$$

$$\text{【定式化 A2】 目的：} \max \sum_{j=1}^M (\mu_j - d/5)^2$$

$$\text{s.t. 式 (A.4) ~ (A.8) および}$$

$$\sum_{i,n} e_{in} \leq T \quad (\text{A.9})$$

ここで, c, d, T は適当に設定された定数である.

定式化 A1 は誤差をなるべく小さくしようとするもので, 線形計画法で解くことができる. 定式化 A2 は誤差和を一定値以下に抑えた上で μ_j のばらつきを最大にしようとするもので, 2 次計画法で解くことができる. しかし, 式 (A.5) は結構面倒である. たとえば表 1 の人 2 の回答したカテゴリは 1,2,3,1,1 となっているので

$$0 \leq e_{21} \leq \tau_1, \tau_1 \leq e_{22} \leq \tau_2, \tau_2 \leq e_{23} \leq \tau_3, 0 \leq e_{24} \leq \tau_1, 0 \leq e_{25} \leq \tau_1$$

のように回答に合わせた制約を立てる必要がある.

このように制約を立てる面倒さにも拘わらず, $c=0.5, d=10, T=8$ のときの定式化 A1 による解は

$$\mu_1 = \mu_5 = 1.7, \mu_2 = \mu_3 = \mu_4 = 2.2,$$

定式化 A2 による解は

$$\mu_1 = \mu_4 = \mu_5 = 1.8, \mu_2 = \mu_3 = 2.3$$

となり, 評価対象の区別は節 3 で示したものよりも劣る.

上田 徹

成蹊大学 理工学部 情報科学科

〒 180-8633 武蔵野市吉祥寺北町 3-3-1

ABSTRACT

ANALYSIS OF SUCCESSIVE CATEGORIES DATA
USING MATHEMATICAL PROGRAMMING

Tohru Ueda

Seikei University

Complex procedures are used in the formulation of dual scaling for successive categories data proposed by Nishisato. However, in a mathematical programming approach, successiveness is easily represented as constraints. Moreover, in a mathematical programming approach it is easy to add and modify objective functions and constraints. Thus mathematical programming models with various objective functions are proposed. Also a model that treats fuzzy numbers is proposed for the purpose of representing lack of assurance or vagueness in answers. The usefulness of these methods is shown by applying them to three data sets.