

早稲田大学 和田 寛子  
 東京大学 後藤 正幸  
 早稲田大学 平澤 茂一

## 1. はじめに

マーケティング等の分野において、商品のイメージや顧客の嗜好を把握するため、アンケート調査を実施し、解析することは一つの有効な手段である。

従来、被験者と質問項目という二面性を持つアンケート調査対象においては、各質問項目について被験者の平均を求め、解析するのが一般的であった。しかし平均値による解析ではばらつき等が考慮されず、被験者と質問項目という二面性を持つ構造（後に嗜好競合構造データと呼ぶ）も反映できない。

本研究ではアンケートデータの構造を考慮して行列データとして定式化する。次に、得られたデータを解析するための新しい解析手法を提案し、その有効性を明らかにする。

## 2. 従来手法

階層クラスタリング手法では被験者から得た点数を質問項目ごとに平均し、そのベクトルをもとにクラスタリングを行う。

【例1】 イメージアンケート

アンケートにより表1の結果が得られたとする。このデータから「ア、イ、ウ、エ」の4人をクラスタリングする。

個体 4人 「ア、イ、ウ、エ」  
 質問項目 2つ 「①、②」  
 被験者 3人 「a,b,c」

表1 得られたアンケートデータ

	ア		イ		ウ		エ	
	①	②	①	②	①	②	①	②
a	-3	2	0	1	-2	4	0	2
b	-3	2	0	1	-2	4	0	2
c	3	0	0	1	2	0	0	2
d	3	0	0	1	2	0	0	2
合計	0	4	0	4	0	8	0	8
平均	0	1	0	1	0	2	0	2

この手法では“合計”または“平均”を用いて解析をしている。ところが、これでは被験者における項目の値のばらつき、すなわち行方向や列方向のベクトルが考慮されておらず、得られたデータが十分に活用されていない。

そこで次章では被験者ごとの評価値のばらつきを考慮した方法で解析を行う。

## 3. 提案手法

### 3.1 アンケートデータの定式化

質問項目をベクトルとして考えるのではなく、行列の形を残して距離構造を取ることを考える。問題を以下のように定式化する。クラスタリングの対象となる個体を  $x_1, x_2, \dots, x_n$  とする。

個体の集合  $X = \{x_1, x_2, \dots, x_n\}$

ここで

被験者の集合  $P = \{P_1, P_2, \dots, P_p\}$

質問項目の集合  $Q = \{Q_1, Q_2, \dots, Q_q\}$

個体  $x_k$  ( $k=1, 2, \dots, n$ ) は単なる記号もしくは個体に關するデータと同一視できる。一般に  $x_k$  に関して、 $p \times q$  個の実変数が存在しているとすると

$$x_k = \begin{pmatrix} x_k^{11} & x_k^{12} & \cdots & x_k^{1q} \\ x_k^{21} & & & \\ \vdots & & & \\ x_k^{p1} & & & x_k^{pq} \end{pmatrix} \quad (3.1)$$

とおける。 $x_k$  は  $p \times q$  次元実行列である。

課題は、このようにして行列の形で与えられるデータを、目的に合わせて適切にクラスタリングする手法を提案することである。

【定義】 嗜好競合構造データ

被験者  $i$  の個体  $x_k, x_l$  に対する評価の差  $\sum |x_k^{ij} - x_l^{ij}|^2$  が大きいアンケートデータを嗜好競合構造を持つデータと定義する。

□

### 3.2 提案手法

ここでは、「被験者」と「質問項目」の両方のデータ、すなわち行列の各要素全てのデータを考慮し、この構造を踏まえた上でクラスタリングを行う方法を提案する。

3.1で述べた嗜好競合構造データを被験者ごとのアンケート調査結果を平均化せず、そのままのデータでクラスタリング手法に用いる。

手順1

個体のうち2つの  $x_k, x_l$  を取り出し、それらの距離  $d(k, l)$  を測る。距離が最小の行列同士を同じクラスタ内に入れる。

距離  $d(k, l)$  の測り方

$$d(k, l) = \sum_{i=1}^m d_i(k, l) (\bar{x}_k^i - \bar{x}_l^i)^2 + \sum_{j=1}^m d_i(k, l) \quad (3.2)$$

ここで

$$d_i(k, l) = \sqrt{\sum_{j=1}^n (x_k^{ij} - x_l^{ij})^2} \quad (3.3)$$

手順2

手順1でできたクラスタも含め、新たにクラスタリングをおこないクラスタを生成する。

## 4. 予備実験

【例2】 例1の個体ア、イ、ウ、エの4人をクラスタリングする。

### 4.1 従来手法

クラスタリングには統計ソフトを使用し、階層クラスタリング手法を用いた。個体の平均評価値のユークリッド距離を求めている。

<結果>

ユークリッド距離

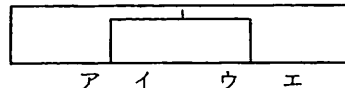


図1 従来手法でのクラスタリング

クラスタリングは  $\{\text{ア}, \text{イ}\}$   $\{\text{ウ}, \text{エ}\}$  となった。

### 4.2 提案手法

3.2の解析手法を用いてクラスタリングを行う。

具体的には、表1を用いて被験者の個体に対する点数を行列として表したまま、その行列間の距離を式(3.2)、(3.3)を用いて計算し、この距離構造に基づいてクラスタリングを行う。

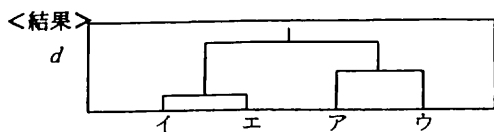


図2 提案手法でのクラスタリング

#### 4.3 考察

従来手法でのクラスタリング→ {ア, イ} {ウ, エ}

提案手法でのクラスタリング→ {イ, エ} {ア, ウ}

従来手法では一方向のデータしか扱うことができないために2次元である行列データの解析においては、やむを得ず被験者の平均値、すなわち各列ベクトルの平均値を算出するなどして1次元に変換して解析を行っている。よって平均値での解析はできるが平均する前のデータの数値やそのばらつき方について生データをそのまま生かすことができない。

これに比べ、提案手法では各列ベクトルの平均値に加え、列ごとのユークリッド距離も考慮している。そのため、列ベクトルの大きさと列中の数値のばらつきの2つが演算要素となっており、これらの要素に見合った解析がなされている。

提案手法のクラスタリング {イ, エ} は列ごとの平均値に違いがあるものの、同一被験者の項目ごとの評価のばらつきが小さいため同一クラスに分類されたものと考えられる。{ア, ウ} に関しても列ごとの平均値に違いがあるのであるが、同一被験者における項目ごとの評価のばらつきが大きいこと同一クラスに分類されたものだと考えられる。

よって嗜好競合構造データに対し、有効な手法であるといえる。

#### 5. 実験

「有名人のイメージ」について実験を行った。これはマスメディアで活躍するタレント・有名人(個体)の知名度・イメージ評価(質問項目)データを体系的に収集することによって、テレビ番組やCMのキャスティングをより効果的なものとするための基礎データになる[1]。

##### アンケート方法

有名人男女10人について15項目のイメージを7段階(-3, -2, -1, 0, 1, 2, 3)で評価してもらった。有名人の選び方としては誰もが知っている、個性豊かである、幅広い年齢層、最近話題を呼んだ、などを念頭においた。

##### ①有名人男女(10人)

安室奈美恵、遠藤久美子、叶姉妹、椎名林檎、広末涼子、稲川淳二、梅宮辰夫、及川光博、木村拓哉、諸星和己

##### ②イメージ項目(15項目)

センスのよさ、新しさ、安心感、経済的な感じ、飽きがこない、心地よい、家庭的だ、親しみがもてる、権威・格が感じられる、明るい、生き生きした、高級感、さわやか、信頼感、夏らしい

##### ③被験者 21人

大学生、大学院生、助手

#### 5-1 従来手法の結果

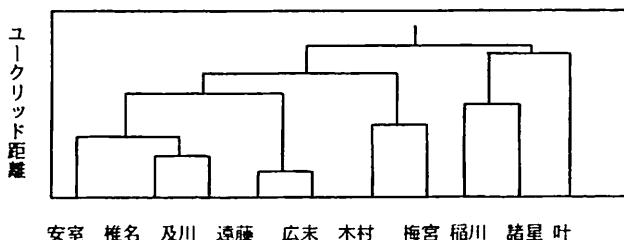


図3 従来手法でのクラスタリング

#### 5-2 提案手法の結果

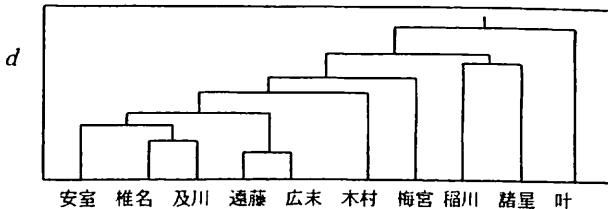


図4 提案手法でのクラスタリング

#### 6. 考察

従来手法と提案手法の結果の違いについてクラスタリングに顕著な差を示した、{梅宮, 木村} について述べる。

すなわち、従来手法においては{梅宮, 木村}という独立したクラスが存在しているのに対し、提案手法では存在していない。原因としては表2に示した通り、特に“センスの良さ”、“さわやか”の項目においてdが大きいことが挙げられる。

表2 梅宮と諸星のデータ

梅宮 X-木村 Y	センスのよさ	さわやか	家庭的だ	安心感
平均 X	0	-0.4762	1.28571	0.2857
平均 Y	2	1.2381	0.33333	0.5238
平均 X-平均 Y	2	1.71429	0.95238	0.2381
d	62.44998	38.592	16.4049	10.19
平均とdの差	-60.44998	-36.878	-15.453	-9.952
相関係数	0.05383819	0.30283	0.47037	0.0965

従来手法では列の平均値、すなわちイメージ項目ごとに被験者の評価点数を平均化した値で解析したために被験者ごとの評価のばらつきが考慮されていなかった。しかし、本研究では列の平均値に加えて、列ごとのユークリッド距離 $d_i(k, l)$ を導入することにより、列ベクトルの方向、すなわちイメージ項目における被験者ごとの評価点数についても解析の対象とした。その結果イメージ項目での被験者ごとの評価の違いも考慮できるようになった。

本研究で定義した「嗜好競合構造データ」とはいわゆる消費者によって好みが極端に分かれるような構造をもつデータである。アーティストやスポーツのプロチーム、ビールなどの嗜好品は、好みは人によって二極化したり、一方を好む人間は他方に良いイメージを持たないなどの傾向がある。本研究ではそのようなデータに対する解析法を提案し、実験的にその有効性を示した。

今回は従来手法に列ごとの式(3.3)で定義したユークリッド距離 $d_i(k, l)$ を加えたが、今後は行ごとのユークリッド距離も取り入れることにより、本研究におけるアンケートデータでいうところの同じ被験者でのイメージ項目ごとの評価点数の違いについても考慮したい。

#### 7. むすび

個体に対し、嗜好に基づく被験者の評価値が大幅に変動するような嗜好競合構造データに対し、新しく個体間距離 $d(k, l)$ を導入することによりクラスタリングを行う手法を提案した。個体間の平均評価値に基づくクラスタリングに比べデータの構造を加味しており興味ある結果を得ることが可能となった。得られる結果を適用分野の知見に基づいて解釈・考察することにより、有効な性質を引き出すことができる。

#### 参考文献

- [1] 小泉秀昭 “ブランド構築における有名人広告の戦略的考察” 日経広告研究所187号
- [2] 宮本定明(1999): “クラスタ分析入門” 森北出版