

# アンケートデータに対するテンソルを用いた解析手法の性能評価

正井 宏明<sup>†</sup> 吉川 大弘<sup>†</sup> 古橋 武<sup>†</sup>

名古屋大学大学院 工学研究科<sup>†</sup>

## 1 はじめに

アンケート調査は、企業が市場の動向の把握や、消費者の潜在ニーズを発見するために行う有効な手段の1つである。このアンケート調査では、評価尺度法やSD法 [1] などが広く用いられているが、その課題として、トータル回答数（評価対象数 × 質問項目数）が多くなりやすく、回答数が多すぎると、回答者のやる気をそいでしまうことが挙げられる。そのため、意味の類似する質問をできるだけ除外する必要がある。

本稿では、回答者が評価対象に抱いた印象の類似性把握や、類似質問の集約を目的とした解析方法について検討する。評価尺度法やSD法で得られたアンケートデータは、図1のように質問 × 評価対象 × 回答者の3階のテンソル（3次元配列）データで表現することができる。従来の解析では、これらのデータに対して、表1のように2次元の行列に変換して多変量解析を行ってきた。本稿では、通常の行列の分解手法である特異値分解を3階のテンソルへ拡張した3-mode SVD[2]を用いて、アンケートデータを効果的に圧縮し、解析を行う手法について検討する。

## 2 テンソルを用いたアンケート解析法

### 2.1 3-mode SVD の概要

まず、テンソルを図1のようにモードごとに展開してそれぞれ行列の形にする。次に各モードごとの行列  $X_{mode}$  に対して、 $X_{mode} = U_{mode} \Sigma_{mode} V_{mode}^T$  のように特異値分解（SVD）を行い、基底行列  $U_{mode}$  を獲得する。さらに  $\Sigma_{mode}(\sigma_i)$  で求められる寄与率を考慮して、各モードの基底行列  $U_{mode}$  に対して1~数列を残し、それ以外を0とした縮小基底行列  $\hat{U}_{mode}$  を用いると、近似テンソル  $\hat{A}$  は、式(1)のように表現できる。

$$\hat{A} = S \times_1 \hat{U}_{質問} \times_2 \hat{U}_{対象} \times_3 \hat{U}_{回答者} \quad (1)$$

式(1)におけるコアテンソル  $S$  や演算などの詳細は、文献 [2] などを参照されたい。

### Performance Evaluation of Analysis using Tensor for Questionnaire Data

Hiroaki Masai<sup>†</sup>, Tomohiro Yoshikawa<sup>†</sup>, Takeshi Furuhashi<sup>†</sup> Nagoya University<sup>†</sup>

表 1: 従来手法の解析方法

	従来手法 1	従来手法 2
PCA の方法	回答者平均の行列に適用	図 1 の $X_{回答者}$ 行列に適用
問題点	個人間の回答差異が消失	質問と対象を区別せず解析
主成分得点	質問の類似度	回答者の類似度
主成分負荷量	対象の類似度	質問・対象の類似度

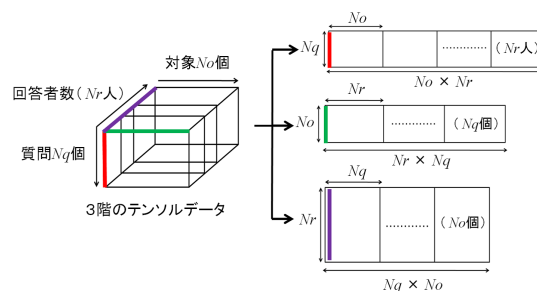


図 1: 各モードの展開行列

### 2.2 提案手法

本稿では、式(1)の近似テンソル  $\hat{A}$  を、図1の  $X_{回答者}$  の行列に展開した行列にPCAを適用する手法を提案する。提案手法により、対象間の類似性把握や質問の集約の検討を行う。

## 3 実験と考察

### 3.1 サンプルデータ

初めに、表2に示す質問5 × 評価対象3 × 回答者20（評点：1~5）のサンプルデータを用いて、提案手法に対する検討を行う。本サンプルデータの特徴を以下に示す。

- (1) 回答者 No.1 ~ 16 は、対象1と対象2の回答傾向が類似していて、対象3の回答傾向とほぼ逆である。
- (2) 回答者 No.17 ~ 20 は、対象2と対象3の回答傾向が類似していて、対象1の回答傾向とほぼ逆である。
- (3) Q1とQ2がすべて逆の評点である。
- (4) それぞれの対象で、Q1とQ5、Q2とQ3とQ4が類似している。
- (5) 回答者 No.1 ~ 8 と回答者 No.9 ~ 16 の回答傾向がほぼ逆である。

表 2: サンプルデータ

回答者	対象1					対象2					対象3				
	Q1	Q2	Q3	Q4	Q5	Q1	Q2	Q3	Q4	Q5	Q1	Q2	Q3	Q4	Q5
No.1	1	5	5	5	1	1	5	4	4	2	5	1	2	2	4
No.2	1	5	4	4	2	1	5	5	5	1	5	1	1	1	5
No.3	1	5	4	5	1	1	5	4	5	1	5	1	2	1	4
No.4	1	5	5	4	2	1	5	5	4	1	5	1	1	2	5
No.5	1	5	3	4	1	1	5	3	4	2	5	1	2	2	5
No.6	2	4	5	4	2	2	4	4	4	1	4	2	2	1	4
No.7	2	4	4	5	1	2	4	5	5	2	4	2	1	1	4
No.8	2	4	5	5	2	2	4	4	4	1	4	2	2	2	4
No.9	5	1	2	2	4	5	1	1	1	5	1	5	4	4	2
No.10	5	1	1	1	5	5	1	2	2	4	1	5	5	4	1
No.11	5	1	2	1	5	5	1	1	2	4	1	5	4	5	2
No.12	5	1	2	2	4	5	1	2	1	5	1	5	5	4	2
No.13	5	1	3	2	4	5	1	3	2	4	1	5	5	5	1
No.14	4	2	1	2	5	4	2	1	1	4	2	4	4	4	2
No.15	4	2	2	1	4	4	2	2	1	4	2	4	5	4	2
No.16	4	2	1	2	5	4	2	1	2	5	2	4	5	5	2
No.17	1	5	4	5	1	5	1	2	2	5	5	1	2	2	5
No.18	1	5	5	4	2	5	1	2	1	4	5	1	2	1	4
No.19	1	5	4	4	2	5	1	1	1	5	5	1	1	2	5
No.20	2	4	4	4	1	4	2	1	2	4	4	2	2	2	4

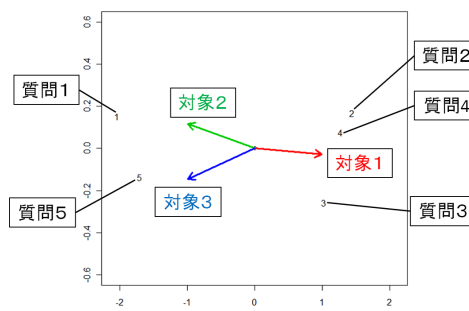
### 3.2 サンプルデータの解析

図 2(a) より, 従来手法 1 では, 3.1 で示した (1) の特徴が (5) の特徴によって相殺されてしまい, (2) の特徴を強調した形が表れていることがわかる. またこの結果のみからでは, 質問の集約を行うことが困難であることがわかる. また図 2(b) より, 従来手法 2 の主成分負荷量では, 3.1 で示した (3) と (4) の特徴がよく表れていることがわかり, 質問の集約を行うことは可能であると考えられる. しかし, 対象と質問の関係が消失してしまう解析法であることで, (1) の特徴である対象 1 と対象 2 の相関があまり見えず, 対象間の類似性を把握することが困難であることがわかる.

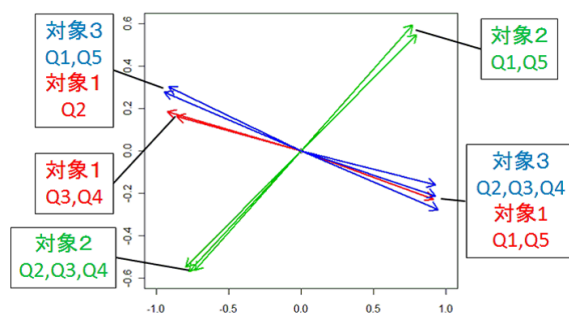
一方で, 図 2(c) の提案手法の主成分負荷量においては, 3.1 で示した (1)(3)(4) の特徴が, 従来手法 1,2 と比較して明確に表れていることがわかり, 質問の集約を行うことも容易であると考えられる. なお対象 1 と対象 2 の主成分負荷量がややずれているのは, (2) の特徴によるものだと考えられる.

### 4 おわりに

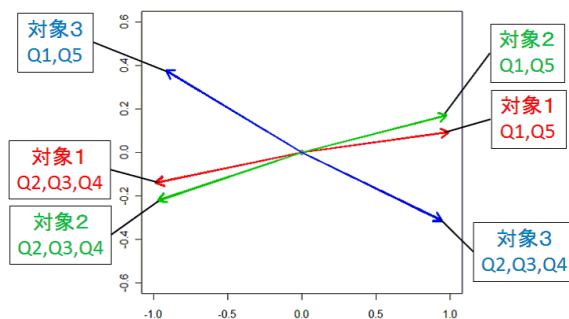
本稿では, テンソル (3-mode SVD) を用いてアンケートデータを効果的に圧縮し, 解析を行う手法について, サンプルデータに対する性能評価を行った. 従来の解析手法と比較し, 提案手法に基づく解析が, 質問の集約や対象間における類似性の把握が行いやすくなることを示した.



(a) 従来手法 1 の biplot



(b) 従来手法 2 の主成分負荷量散布図



(c) 提案手法 (rank = (2, 2, 20) 近似) の主成分負荷量

図 2: 従来手法と提案手法の比較

### 参考文献

- [1] Osgood, C.E., Suck, G.J., and Tannenbaum P.H.: The Measurement of Meaning, University of Illinois Press, 1957
- [2] M. A. O. Vasilescu, and D. Terzopoulos: "Multilinear Analysis of Image Ensembles: Tensorfaces", in the Proc. ECCV-02, Copenhagen, Denmark, pp.447-460, 2002