

# 3相データにおける分析モデルの性能比較

2007MI038 服部基統

指導教員 松田眞一

## 1 はじめに

現実世界でのデータを対象に因子分析を行う場合、3相データを使用するほうが適切であることも多いように感じる。そこで、複数ある3相データの分析モデルについて性能がどのように違っているのかを比較し、3相因子分析を有効に行えるようになることを本研究の目的とする。

## 2 3相因子分析

この章は柳井ら [2] と奥村ら [1]、および村上 [3] を参考にした。

### 2.1 PARAFAC

$m$  個の因子が存在するとき、 $F$  を  $n \times m$  の共通因子負荷量行列、 $W_k$  を  $m \times m$  の対角行列、 $\Lambda$  を  $p \times m$  の共通因子負荷量行列、 $\mathcal{E}_k$  を誤差行列とすると

$$X_k = FW_k\Lambda' + \mathcal{E}_k \quad (k = 1, 2, \dots, q) \quad (1)$$

と表現されるモデルを PARAFAC モデルという。

なお、本研究では自作のプログラムにより解析を行った。

### 2.2 Tucker

個体因子数を  $a(a = 1, 2, \dots, s)$ 、変数因子数を  $b(b = 1, 2, \dots, t)$ 、条件因子数を  $c(c = 1, 2, \dots, u)$  とし、 $A$  を  $q \times u$  の条件  $\times$  条件因子の因子負荷量行列、 $B$  を  $p \times t$  の変数  $\times$  変数因子の因子負荷量行列、 $F$  を  $n \times s$  の個体  $\times$  個体因子の因子負荷量行列、 $G'$  を  $s \times ut$  の超行列とすると

$$X = FC(A \otimes B)' \quad (2)$$

ただし、

$$C = G'K_{tu} \quad (3)$$

を Tucker の 3 相因子分析モデル (Tucker 3 モデル) という。

### 2.3 Tucker の近似解法

村上 [3] では条件因子を用いない Tucker 2 モデルを主成分分析を反復させる「段階的主成分分析」というアルゴリズムで求めて因子の意味付けを行っている。

これに対して、奥村・大澤 [1] は主成分分析ではなく因子分析を反復させる「段階的因子分析」とも呼ぶべきアルゴリズムによって因子の意味付けを行っている。

## 3 性能比較

### 3.1 データ

本研究において分析モデルの性能比較をする際、3相データを得るために職業イメージに関するアンケートを作成した。ここでは、職種数が個体、イメージの形容詞数が変数、そして回答者数が条件と、3相データの各パラメータが対応するようにしてある。

### 3.2 比較方法

PARAFAC モデルを用いた因子分析、奥村・大澤 [1] の Tucker による「段階的因子分析」の 2 通りでアンケートデータを分析する。それぞれの結果に対して因子の意味付けを行っていき、これら 2 つのモデルの共通点や相違点などを考察していく。

### 3.3 PARAFAC モデルを用いた分析結果

因子数を 3 として分析をした。実行結果は以下の通り。まず、変数に関する共通因子負荷量行列は、

	Factor1	Factor2	Factor3
派手だ	-0.454	1.149	-1.150
頼もしい	-0.337	-0.516	-1.119
気楽だ	1.237	1.180	-0.452
感覚的だ	1.419	1.616	-0.626
安定している	-0.331	-1.179	-1.080
現代的だ	0.135	0.032	-1.100
知的だ	-0.818	-0.300	-1.427
優雅だ	-0.216	0.873	-1.022
庶民的だ	2.133	-1.135	-0.125
面白い	0.919	1.262	-0.715
忙しい	-0.057	-0.718	-1.233
収入が良い	-1.181	-0.366	-1.471
男女平等だ	0.815	0.386	-0.900
自由だ	1.351	1.572	-0.581

次に、個体に関する共通因子得点行列は、

	Factor1	Factor2	Factor3
政治家	0.342	-0.256	-1.001
画家	1.243	1.694	-0.869
医者	0.373	-1.063	-1.190
学者	0.823	-0.359	-1.064
弁護士	0.412	-0.646	-1.153
プログラマー	1.040	-0.808	-1.008
商社マン	0.965	-0.920	-0.988
証券マン	0.904	-0.918	-0.998
銀行マン	0.941	-1.211	-0.992
アナウンサー	0.588	0.614	-1.131
デパート店員	1.382	-0.587	-0.845
工場労働者	1.489	-1.338	-0.724
雑誌編集者	1.040	0.532	-0.998
小・中・高教員	1.279	-1.271	-1.048
カメラマン	1.196	1.468	-0.875

よって、因子の意味付けは次のように考えられる。変数の因子は、変数の因子は、

- 第 1 因子：「庶民的なイメージ」
- 第 2 因子：「自由をとるか」  $\leftrightarrow$  「安定をとるか」
- 第 3 因子：「社会的地位が高いイメージ」(反転)

また、個体についての因子は、

- 第1因子：「仕事内容がイメージしやすい職業」
- 第2因子：「自由なイメージの職業」↔「堅いイメージの職業」
- 第3因子：「エリート職」

となる。さらに、条件ごとの各因子の効き具合を表す感性行列をクラスター分析にかけると、2群に分けることができ、

- 第1群：「平等にこだわりたい人」
- 第2群：「職種も給料もこだわりたい人」

となった。

### 3.4 Tucker モデルを用いた分析結果

変量についての因子数3、条件についての因子数3として分析をした。実行結果は以下の通り。

変量についての共通因子負荷行列は、

	Factor1	Factor2	Factor3
派手だ	-0.017	0.482	0.164
頼もしい	0.633	0.189	0.092
気楽だ	-0.672	0.165	0.255
感覚的だ	-0.394	0.092	0.488
安定している	0.448	0.283	-0.325
現代的だ	0.361	0.225	0.187
知的だ	0.621	0.435	0.112
優雅だ	0.015	0.576	0.199
庶民的だ	-0.088	-0.568	-0.107
面白い	-0.156	0.259	0.568
忙しい	0.648	0.049	-0.115
収入が良い	0.489	0.698	-0.329
男女平等だ	0.043	0.037	0.194
自由だ	-0.583	0.218	0.437

続いて、条件（職業）についての共通因子負荷量行列は、

	Factor1	Factor2	Factor3
政治家	0.607	0.103	-0.530
画家	-0.873	0.276	-0.159
医者	0.856	-0.090	-0.201
学者	0.009	0.262	-0.038
弁護士	0.616	-0.075	-0.178
プログラマー	0.053	-0.024	0.282
商社マン	0.530	0.273	0.305
証券マン	0.658	0.408	0.373
銀行マン	0.680	0.411	0.249
アナウンサー	0.033	0.691	-0.330
デパート店員	-0.142	0.496	0.387
工場労働者	-0.126	-0.092	0.542
雑誌編集者	-0.408	0.452	-0.273
小・中・高教員	0.061	-0.027	0.154
カメラマン	-0.787	0.454	-0.225

よって、因子の意味付けは次のように考えられる。変量の因子について、

- 第1因子：「インテリなイメージ」↔「自由なイメージ」
- 第2因子：「裕福なイメージ」↔「庶民的なイメージ」

- 第3因子：「直観的なイメージ」

また、条件についての因子は、

- 第1因子：「インテリ職」↔「芸術職」
- 第2因子：「感性を生かす職業」
- 第3因子：「仕事が大変な職業」

となった。さらに、個体の感性行列をクラスター分析にかけると2群に分かれ、

- 第1群：「比較的ラクな方法でお金を稼ぎたい人」
- 第2群：「仕事にはやりがいを求める人」

となった。なお、各群の因子の対応は、第1群では

- 変量第1因子 — 条件第1因子、条件第3因子
- 変量第2因子 — 条件第1因子
- 変量第3因子 — 条件第3因子

と対応し、第2群では、

- 変量第1因子 — 条件第1因子
- 変量第2因子 — 条件第3因子
- 変量第3因子 — 条件第2因子

が対応した。

## 4 まとめ

2つのモデルの相違点として因子数の決定が挙げられる。Tucker モデルが変量と条件についての2つ因子数が決定できることに対して、PARAFAC モデルは共通で1つの因子数しか決定できない。しかし、どちらも個体、変量、条件それぞれの相互作用をみることができ3相データの利点を生かしていると言える。ただ、3相データにおける変量や条件が多い場合にはTucker モデルを用いた方が細かい操作ができるため便利であると考えられる。また、各モデルで同じような解釈ができる因子もあれば、異なった解釈ができる因子もあるため、理想としては両モデルを用いて分析を行うことがデータに隠れている潜在因子を見つけ出すためには必要だと考えられる。

## 5 おわりに

PARAFAC モデルを用いた因子分析のプログラム作成には終始悩まされた。そのため、本研究は3相因子分析理論の理解が中心となってしまった。しかし、3相データを用いた分析は現実世界での多くの事象に対して効果を発揮することがわかった。よって今後も引き続きプログラムの改良を続け、3相因子分析をはじめ、他の3相データ分析も有効に行えるようにしていきたい。

## 参考文献

- [1] 奥村高志, 大澤健一 『交通標識の認識に対する統計的研究』, 2001 .
- [2] 柳井晴夫, 繁樹算男, 前川真一, 市川雅教 『因子分析-その理論と方法-』, 1990 .
- [3] 村上隆: 3相データの階層的な主成分分析 『人間行動の計量分析-多変量データ解析の理論と応用』, pp.71-94 , 1990 .