

主成分距離行列シルエットクラスタリングによる潜在因子ラベル付けモデル

Model to Name for Latent Variable by Silhouette Clustering Applied to Principal Component Distance Matrix

大槻 明¹

Akira Otsuki

1. はじめに

従来の因子分析モデルでは、抽出因子数の正当性を確認するところまでは統計的に行われているが、抽出された潜在因子のラベル付け（潜在因子の解釈や意味付け）は、分析者の主観で行われているケースが多い。しかし、分析者主観が入る時点で潜在因子の解釈に対する科学的根拠が損なわれてしまう可能性がある。

そこで本研究では、潜在因子のラベル付けに科学的根拠を持たせるための補助的なアプローチとして、主成分得点からユークリッド距離を求めて距離行列を作成し、この距離行列を対象に、シルエット分析を用いたクラスタリングを行うことで、潜在因子にラベル付けを行うモデルについて考案した。なお、この提案モデルは因子分析に限らず多くの距離行列データに適用することができると考えられる。

第 4 章、第 5 章で既存手法と比較分析した結果、従来の手法では、潜在因子数は推定できても、特に、高負荷量の因子を対象とした潜在因子のラベル付けを行うことが難しかった。対して本提案モデルでは、シルエット分析を用いて最適なクラスタ数 K に分割したうえで潜在因子数を抽出するモデルであるため、従来手法よりも科学的根拠を持たせる形でラベル付けを行うことが可能であることを示した。

2. 従来の因子分析モデルにおける潜在因子の解釈

因子分析の結果解釈に関する論述は、古くは 1968 年に Armstrong ら[1]が論文「On the interpretation of factor analysis」の中で行っており、この論文では、研究者は因子分析の解釈に例を用いて説明するのみであり、この例には因子分析の解釈に係る信頼性や妥当性が省略されていると主張されている。最近でも、John T.[2]が 1992 年～2002 年 The Journal of Educational Research で公開された論文を調査し、教育研究学分野における因子分析の解釈に係る問題について論じている。具体的には、分析者は因子分析の結果を評価するために十分な情報を提供していないことが問題であると指摘している。つまり、分析者の恣意的な解釈により、因子分析の結果が解釈されていることを示唆している。

日本に目を向けると、柳井[3]が、1998 年から 1999 年までに教育心理学研究・心理学研究で発表された論文を調査した結果について論じているが、主に因子分析法の利用をめぐる諸問題（因子回転の問題）についてしか論じられていない。つまり、日本においては潜在因子の解釈に関する議論はこれまで盛んには行われてこなかったと考えられる。

潜在因子数の推定に関する手法としては、最尤法を使用する場合には、多変量正規分布から乖離している変数を削除する、といった多変量正規分布を条件とすることがよく知られている[4]。また、堀[5]のまとめによると、スクリーテストや平行分析なども良く用いられている。さらに、Lee ら[6]は歪度と尖度から分析対象の変数を選択している。

以上をまとめると、本論文で取り上げた先行研究では、潜在因子を抽出する手法については数多く研究されてきたが、抽出された潜在因子に対するラベル付け（解釈や意味付け）に深く言及している研究は少なかった。つまり、抽出された潜在因子を元に、分析者の主観でラベル付けが行われるケースが多いと考えられる。具体例を示すと、中川ら[7]の論文では、因子負荷量 0.3 以上の因子を潜在因子として抽出しており、松岡ら[8]の論文では、因子負荷量 500 以上の因子を潜在因子として抽出している。三保ら[9]の論文では、「因子パターンの値が各因子で ± 0.40 以上であった 39 項目を 4 因子の解釈の対象項目とした」とされている。このように、分析者の主観が入る時点で潜在因子の解釈に対する科学的根拠が損なわれてしまう可能性があると考えられる。

ゆえに本研究では、潜在因子のラベル付けに科学的根拠を持たせるための補助的なアプローチとして、主成分得点からユークリッド距離を求めて距離行列を作成し、この距離行列を対象に、シルエット分析を用いたクラスタリングを行うことで、潜在因子にラベル付けを行うモデルについて提案する。

3. 分析対象データ

厚生労働省の平成 25 年若年者雇用実態調査の概況[10]で公開されている表 22 のデータを用いることとした。この理由は、同表は、最終学校卒業後初めて勤務した会社をやめた主な理由について、男女、年齢及び最終学歴などの別ごとにまとめられたものであり、ここから、従来手法と本モデルで、初めて勤務した会社をやめた主な理由の潜在因子を推定できるかについて比較検証するためである。実際に分析に用いたデータを付録 A に示す。

4. 従来の因子分析モデルによる潜在因子数の推定及びラベル付け

4.1 潜在因子数の推定

第 2 章で述べた先行研究の手法[6-9]を参考に因子分析を行う。まず、潜在因子数の決定であるが、中川ら[7]の論文では、「5)抽出因子数決定のための基準」の中で述べているように、「固有値が 1 よりも大きいもので、因子負荷量 0.3 以上の潜在因子」としている。松岡ら[8]の論文では、因子負荷量 500 以上の高い負荷量を解釈基準とした、とされている。三保ら[9]の論文では、「第 1 因子から第 7 因子にかけて 10.21, 6.33, 3.78, 2.33, 2.16, 1.67, 1.44 となった。3 因子から 5 因子の結果を比較して、解釈可能性から、因子数を 4 とし、主因子法で共通性を推定した。」としか書かれておらず、結局はこの論文の図 1 で示されている 4 分野を設定しているだけであった。つまり、先行研究では、分析者の主観により潜在因子の抽出方法（閾値）がバラバラである。そこで、堀[6]のまとめを参考に平行分析

¹ 日本大学経済学部 Nihon University College of Economics

及びスクリーテストを用いることとした。平行分析の結果を図 1 に、スクリーテストの結果を図 2 にそれぞれ示す。

図 1 の x が主成分分解、つまり共通性の推定をしないときのスクリープロットを表している。そして、△ が因子分析のスクリープロットであり、ここから因子数を決めることができる。具体的には、点線(二種)が平行分析を表しており、これは同じデータセットの乱数表から作られる相関行列や、データセットの数値をリサンプリングして作られる相関行列のスクリープロットである。このラインより上に来る因子が有意と判断されるため、平行分析からは 3 因子と判断される。

また、スクリーテスト (図 2) は、相関行列の固有値を固有値順位に対してプロット (スクリープロット) し、これに最下位固有値から傾向線を引き、その傾向線から離れる固有値の順位が因子数となる。そして、三保らの手法を参考にすると、3 因子から 5 因子の結果を比較した場合、解釈可能性から因子数は 3 か 4 と仮定される。また、中川らの「固有値が 1 以上」の基準を参考にしても、因子数は 3 か 4 と仮定される。

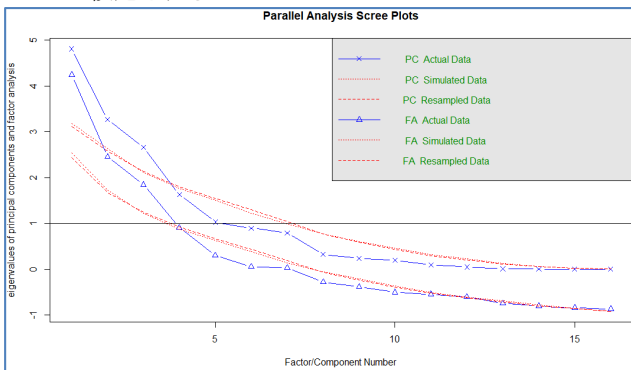


図 1. 平行分析による因子数の決定

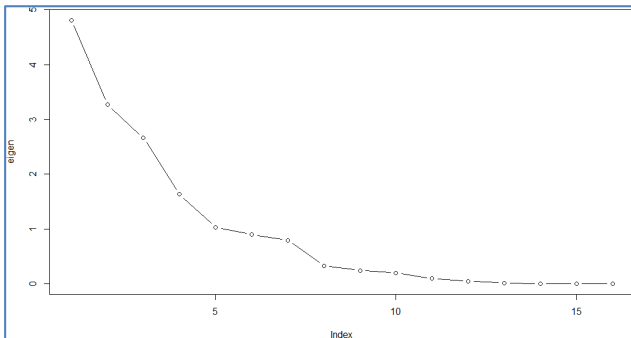


図 2. スクリーテストによる因子数の決定

4.2 潜在因子のラベル付け

前節から、因子数を 3 及び 4 として因子分析を行った。なお、因子得点を求める方法には回帰法 (トンプソン法) を、また回転法には promax を用いて因子分析を行った。この結果を表 1~表 5 に示す。

前述のように、中川ら[7]の論文では、因子負荷量 0.3 以上の因子を潜在因子として抽出しており、松岡ら[8]の論文では、因子負荷量 .500 以上の高い負荷量の因子を潜在因子として抽出している。三保ら[9]の論文では、「因子パターンの値が各因子で ± 0.40 以上であった 39 項目を 4 因子の解釈の対象項目とした」とされている。つまり、先行研究では分析者の主観によって潜在因子抽出の基準はバ

ラバラである。そこで、中川ら[7]の論文では、多くの論文をサーベイした結果、因子負荷量 ± 0.30 以上が妥当として潜在因子の解釈を試みていたため、本研究でもこの閾値を採用することとした。なお、例えば表 1 の x7 のように、複数の因子が因子負荷量 ± 0.30 となる場合は高い方の因子を採用した。さらに、相関分析では、相関係数が .70 以上で「強い相関がある」と定義されている[11]。このように高い負荷量で見ると科学的根拠が増すと考えられるため、負荷量 ≥ 0.7 を閾値とした場合についても検証した。

まず、因子数 3 のケースで因子負荷量 ± 0.30 の基準を用いた場合の潜在因子のラベル付けであるが (表 1~2)、Factor1 (潜在因子) は「労働環境と自己都合の混在」というラベルを、Factor2 は「労働環境」のラベルを付けることがそれぞれ可能であると考えられる。しかし、Factor3 は潜在因子が「労働環境 (x1)」、「自己都合 (x11)」の 1 つずつしかなく、2 つ以上の変数が揃わないとラベル付けをする根拠に乏しいと考えられるため、Factor3 はラベル付けを行うことが困難である。なお、表 2, 3, 5, 8 の 2 列目に x1~x16 の上位カテゴリを便宜的に作成しているが、これらの上位カテゴリが無くては上述及び下記に述べる解釈 (ラベル付け) と同様の解釈ができると考えられる。

次に、因子数 3 のケースで因子負荷量 ± 0.70 の基準を用いた場合の潜在因子のラベル付けであるが (表 1, 3)、Factor1 は、因子負荷量 ± 0.30 以上の時と同様に「労働環境と自己都合の混在」というラベルを付けることが可能であると考えられる。しかし、Factor2 を同様の閾値でみた場合は、閾値を超える因子数は、「労働環境 (x5)」と「自己都合 (x16)」の 1 つずつしかない。同様に、Factor3 も閾値を超える因子数も、「労働環境 (x1)」、「自己都合 (x11)」の 1 つずつしかないため、これらも前述と同様にラベル付けを行う根拠に乏しい結果となった。

表 1. 初めて勤務した会社を辞めた理由尺度の因子分析と尺度構成結果 (因子数 3)

	Factor1	Factor2	Factor3	Ave	StDev
x1	-0.41	0.15	<u>0.93</u>	0.22	0.55
x2	<u>0.45</u>	0.12		0.29	0.17
x3	<u>0.75</u>	-0.11	0.14	0.26	0.36
x4	<u>-0.58</u>	0.25	0.21	-0.04	0.38
x5	0.50	<u>1.08</u>	0.24	0.61	0.35
x6	<u>0.33</u>	0.30		0.32	0.02
x7	-0.34	<u>0.60</u>	0.16	0.14	0.39
x8	<u>-0.85</u>	-0.28		-0.56	0.28
x9	0.42	<u>-0.56</u>	0.29	0.05	0.43
x10	<u>-0.84</u>	0.15		-0.34	0.49
x11	-0.21	-0.10	<u>-0.90</u>	-0.40	0.35
x12	0.14	-0.15	0.20	0.06	0.15
x13	<u>-0.79</u>			-0.79	0.00
x14	<u>1.01</u>	0.61	-0.11	0.50	0.46
x15	<u>0.62</u>			0.62	0.00
x16	0.14	<u>0.73</u>	-0.26	0.20	0.41

表 2. 潜在因子の解釈 (因子数 3, 因子負荷量 ≥ ± 0.30)

Factor1	労働環境	x3 責任のある仕事を任されたかった
	労働環境	x4 ノルマや責任が重すぎた
	労働環境	x6 賃金の条件がよくなかった
	労働環境	x8 人間関係がよくなかった
	自己都合	x2 自分の技能・能力が活かされなかった

	自己都合	x10 健康上の理由
	自己都合	x13 独立して事業を始めるため
	自己都合	x14 家業をつぐ又は手伝えるため
	自己都合	x15 1つの会社に長く勤務する気がなかったため
Factor2	労働環境	x5 会社に将来性がない
	労働環境	x7 労働時間・休日・休暇の条件がよくなかった
	労働環境	x9 不安定な雇用状態が嫌だった
	自己都合	x16 倒産、整理解雇又は希望退職に応じたため
Factor3	労働環境	x1 仕事が自分に合わない
	自己都合	x11 結婚、子育てのため

表 3. 潜在因子の解釈 (因子数 3, 因子負荷量 $\geq \pm 0.70$)

Factor1	労働環境	x3 責任のある仕事を任されたかった
	労働環境	x8 人間関係がよくなかった
	自己都合	x10 健康上の理由
	自己都合	x13 独立して事業を始めるため
	自己都合	x14 家業をつぐ又は手伝えるため
Factor2	労働環境	x5 会社に将来性がない
	自己都合	x16 倒産、整理解雇又は希望退職に応じたため
Factor3	労働環境	x1 仕事が自分に合わない
	自己都合	x11 結婚、子育てのため

次に、因子数 4 のケースで因子負荷量 ± 0.30 の基準を用いた場合の潜在因子のラベル付けであるが (表 4~5), Factor4 に該当する因子が存在していない時点で、潜在因子数 4 は妥当ではないと考えられる。具体的に、Factor4 で因子負荷量 $\geq \pm 0.30$ の変数は x7 と x14 しかないが、x7 は Factor2 の因子負荷量が最も高いため Factor2 に所属することになり、x14 は Factor1 の因子負荷量が最も高いため Factor1 に所属することになる。結果 Factor4 に所属する因子は存在しないことになる。ゆえに、因子数 4 のケースで因子負荷量 $\geq \pm 0.70$ の基準を用いた場合の検証は行わないこととした。

表 4. 初めて勤務した会社を辞めた理由尺度の因子分析と尺度構成結果 (因子数 4, 因子負荷量 $\geq \pm 0.30$)

	Factor1	Factor2	Factor3	Factor4	Ave	StDev
x1	-0.25	0.12	<u>0.93</u>	0.21	0.25	0.43
x2	<u>0.89</u>	-0.17	0.57		0.43	0.45
x3	<u>0.96</u>	-0.18	0.13		0.30	0.48
x4	0.10	<u>0.97</u>			0.54	0.43
x5	0.28	<u>1.00</u>	0.27		0.51	0.34
x6	0.15	0.26	0.29	0.23	0.23	0.05
x7	-0.24	<u>0.53</u>	0.14	0.38	0.21	0.29
x8	<u>-0.69</u>	-0.25	0.11		-0.28	0.33
x9	0.50	<u>-0.54</u>	0.25	-0.16	0.01	0.40
x10	-0.50	0.11	<u>0.55</u>		0.05	0.43
x11	-0.13	-0.12	<u>-0.93</u>	0.21	-0.24	0.42
x12	-0.19	0.29	<u>-0.51</u>		-0.14	0.33
x13	<u>-0.77</u>				-0.77	0.00
x14	<u>0.69</u>	0.57	-0.10	-0.31	0.21	0.42
x15	<u>0.55</u>	0.21			0.38	0.17
x16	<u>0.69</u>	-0.22			0.23	0.46

表 5. 潜在因子の解釈 (因子数 4, $\geq \pm 0.30$)

Factor1	労働環境	x3 責任のある仕事を任された
---------	------	-----------------

		かった
	労働環境	x8 人間関係がよくなかった
	自己都合	x2 自分の技能・能力が活かされなかった
	自己都合	x10 健康上の理由
	自己都合	x13 独立して事業を始めるため
	自己都合	x14 家業をつぐ又は手伝えるため
	自己都合	x15 1つの会社に長く勤務する気がなかったため
	自己都合	x16 倒産、整理解雇又は希望退職に応じたため
Factor2	労働環境	x4 ノルマや責任が重すぎた
	労働環境	x5 会社に将来性がない
	労働環境	x7 労働時間・休日・休暇の条件がよくなかった
	労働環境	x9 不安定な雇用状態が嫌だった
Factor3	労働環境	x1 仕事が自分に合わない
	自己都合	x11 結婚、子育てのため
	自己都合	x12 介護、看護のため
Factor4		NA (該当する因子が無い)

5. 提案モデル

5.1 主成分分析及び累積寄与率

本モデルで主成分分析を用いる理由は、n 個の観測データを (合成変数) に要約し、クラスタリングをするための距離行列を作成するためである。データを要約する手法には、主成分分析以外にも、対応分析[12]や MDS (多次元尺度構成法) [13]などが存在するが、まず、対応分析はカテゴリカル (質的) データを対象とした手法であり、本研究が対象とするデータは量的データである。また、主成分分析はデータの散らばり具合の情報量として保持して合成変数を作成し、それを軸としてデータを要約するのに対し、MDSは対象間の距離を維持したまま小さな次元に要約する。つまり、主成分分析では合成変数が何かしらの意味を持つのに対し MDS での軸は特に意味はもたない。本研究では、潜在因子を抽出することを目的とするため、合成変数に何かしらの意味を持たせることが可能である主成分分析を採用することとした。

まず、第 3 章の対象データ (付録 A) を元に主成分分析を行い、累積寄与率を調べて第 n 主成分得点までを使うかを調査する。表 6 に累積寄与率を求めた結果を示しているが、今回は累積寄与率が 90% を超える第 5 主成分得点までを使用することとした。

表 6. Cumulative Proportion

PC1	PC2	PC3	PC4	PC5	PC6	PC7	...	PC16
0.40	0.67	0.82	0.87	0.92	0.95	0.97	...	1.00

5.2 主成分得点からユークリッド距離を計算

潜在因子は、付録 A であれば説明変数 (x1~x16) から算出するため、付録 A の行列を転置したうえで、第 5 主成分までの得点を用いて各説明変数間のユークリッド距離 (Euclidean Distance: ED) を求める。例として説明変数 x1, x2 間のユークリッド距離 ED(x1,x2)は次式により求められる。なお、pc1~5は第 1~第 5 主成分得点を表す。

$$ED(x1, x2) = \sqrt{(x1_{pc1} - x2_{pc1})^2 + (x1_{pc2} - x2_{pc2})^2 + \dots + (x1_{pc5} - x2_{pc5})^2} \quad (1)$$

そして、式 1 によって算出された x1~16 間のユークリッド距離を元に、表 7 に示すような主成分得点距離行列を作成する。

表 7. 主成分得点距離行列のイメージ

	x1	x2	...	x16
x1	ED(x1,x1)	ED(x1,x2)	...	ED(x1,x16)
x2	ED(x2,x1)	ED(x2,x2)	...	ED(x2,x16)
...
x16	ED(x16,x1)	ED(x16,x2)	...	ED(x16,x16)

5.3 主成分距離行列を対象としたクラスタリング

前節の主成分得点距離行列を対象に、シルエット分析 [14] を用いてクラスタ数 K を調べる。シルエット分析は、クラスタ内の面が適切に分割されているかどうかを検証するための手法であり、シルエット (silhouette) 値を用いて分析する。シルエット値は、他のクラスタの点と比べてその点が自身のクラスタ内の他の点にどれくらい相似しているかを示す尺度であり、i 番目の点のシルエット値 S_i は、次式のように求められる。

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (2)$$

つまり、平均シルエット値が最も大きい時にクラスタ分割を行うと、クラスタ内凝集度が最も高く、また他のクラスタとの分散度が最も高い状態でクラスタリングを行うことができる。今回は、図 3 に示す通り、最も平均シルエット幅が大きい数値は 3 であったため、K=3 としてクラスタリングを行った。この結果を図 4 に示す。

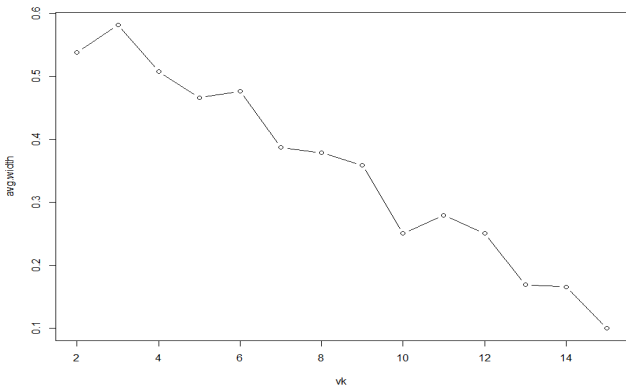


図 3. シルエット値の推移

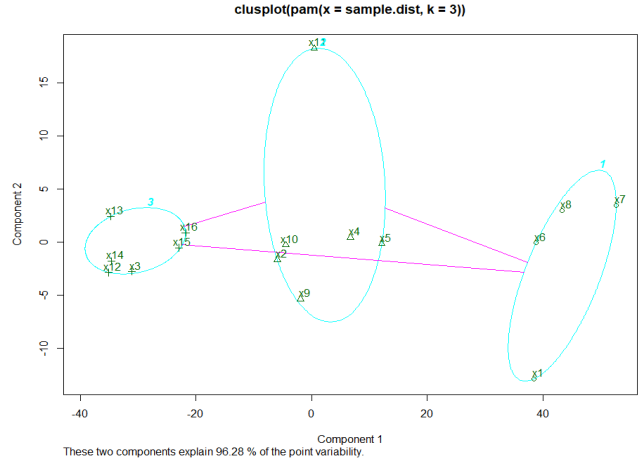


図 4. クラスタリングの結果 (K=3)

図 4 の各クラスタのメンバをリスト化したものが表 7 である。表 8 から Cluster1 は労働環境に問題を感じて退職したメンバが多いクラスタであることが明らかになり、Cluster3 は自己都合による退職したメンバが多いクラスタであることが明らかになった。さらに、Cluster2 は労働環境に加え自己都合による退職したメンバのクラスタであることが明らかとなった。以上から、「労働環境(Cluster1)」と「労働環境及び自己都合の混在(Cluster2)」及び「自己都合(Cluster3)」という 3 つの潜在因子が抽出できたと考えられる。今後の課題としては、Cluster2 のように複数の意味が混合する潜在因子が特定できた場合に、Cluster2 内のスコアの分散を見るなど、複数の意味が混合する潜在因子の詳細な解釈が可能となるアプローチについて研究していきたい。

表 8. クラスタ内メンバリスト (K=3)

Cluster1	労働環境	x1 仕事が自分に合わない
	労働環境	x6 賃金の条件がよくなかった
	労働環境	x7 労働時間・休日・休暇の条件がよくなかった
	労働環境	x8 人間関係がよくなかった
Cluster2	労働環境	x4 ノルマや責任が重すぎた
	労働環境	x5 会社に将来性がない
	労働環境	x9 不安定な雇用状態が嫌だった
	自己都合	x2 自分の技能・能力が活かされなかった
	自己都合	x10 健康上の理由
	自己都合	x11 結婚、子育てのため
Cluster3	労働環境	x3 責任のある仕事を任されたかった
	自己都合	x12 介護、看護のため
	自己都合	x13 独立して事業を始めるため
	自己都合	x14 家業をつぐ又は手伝うため
	自己都合	x15 1つの会社に長く勤務する気がなかったため
	自己都合	x16 倒産、整理解雇又は希望退職に応じたため

5.4 クラスタリングの結果 (潜在因子の) 解釈

表9~11は、付録Aを前節のクラスタ1~3ごとに分割し、「①Cluster1~3の平均」列、「①/①~③の合計」列を追加したものである。

そして、次式の IoC_{ij} (Interpretation of Cluster)により、各クラスタの評価を行った。

$$IoC_{ij} = \frac{ave_{ij}}{\sum(ave_{i=1\sim3j})} \quad (3)$$

i はクラスタ番号1~3を表し、 j は1, 2, ..., 17の値(番号)を表す。つまり j は付録Aの行サンプルa~qに対応している。例えば、 $j=1$ はaを、 $j=17$ はqをそれぞれ表す。なので、 ave_{ij} は i 番目のクラスタの j 行目のサンプルの全変数(x1~x16)の平均スコアを表す。また、 $\sum(ave_{i=1\sim3j})$ は、 j 行目のサンプルのクラスタ1~3の平均スコアの合計を表す。つまり、 IoC_{ij} はクラスタ1~3の合計に占める各クラスタの割合を、行サンプルの j 行の単位で算出する式である。なお、表9~11では、 ave_{ij} は「①Cluster1~3の平均」の列に、 IoC_{ij} は「①/①~③の合計」の列にそれぞれ表示されている。また、出典元[10]によると、付録Aの値は選択割合であると述べられているため、 IoC_{ij} の値が高いほど、a~qの全被験者属性の選択率が高い変数の集まりであると考えられる。ゆえに、クラスタ1~3の重要度は、「クラスタ1>クラスタ2>クラスタ3」という解釈ができ、さらに、全クラスタ(1~3)の中で、クラスタ1の割合が過半数以上を占めていたため、a~qの全被験者属性が退職した理由に最も起因している潜在要因は「労働環境」であることが明らかとなった。

表9. IoC_{ij} を用いたクラスタ1の分析結果

	x1	x6	x7	x8	①Cluster1 の平均	①/①~③ の合計
a	22.0	22.1	22.7	15.7	20.63	0.65
b	16.1	14.7	21.8	22.8	18.85	0.61
c	23.3	20.1	23.8	22.7	22.48	0.70
d	18.7	18.3	22.6	22.1	20.43	0.63
e	17.4	17.3	21.8	16.7	18.30	0.60
f	19.0	17.9	20.3	19.6	19.20	0.63
g	23.4	19.6	22.3	26.7	23.00	0.66
h	14.3	21.5	20.8	24.0	20.15	0.63
i	19.7	16.4	25.3	15.5	19.23	0.60
j	14.5	16.8	11.6	14.1	14.25	0.57
k	18.9	20.1	24.7	18.2	20.48	0.64
l	18.6	16.0	19.8	21.0	18.85	0.61
m	27.3	20.0	21.8	24.2	23.33	0.66
n	21.1	20.5	27.3	19.8	22.18	0.65
o	19.9	23.3	25.5	19.2	21.98	0.61
p	12.3	18.2	24.7	15.6	17.70	0.57
q	7.7	18.8	19.8	20.0	16.58	0.54

表10. IoC_{ij} を用いたクラスタ2の分析結果

	x2	x4	x5	x9	x10	x11	②Cluster2 の平均	①/①~③ の合計
a	8.9	10	15.2	8.7	7.3	3.0	8.85	0.28
b	7.1	12	9.9	8.7	9.0	15.0	10.28	0.33
c	5.6	9.4	12.4	8.0	9.0	5.0	8.23	0.26
d	7.2	10.3	13.2	9.8	8.6	8.1	9.53	0.30
e	8.5	12.1	11.9	8.1	7.9	11.9	10.07	0.33
f	6.6	9.3	11.6	9.3	7.0	10.3	9.02	0.30
g	6.7	11.5	13.8	7.4	9.7	8.3	9.57	0.27
h	8.2	11.7	8.3	8.6	9.5	15.6	10.32	0.32
i	10.1	14.1	13.9	6.9	9.5	8.7	10.53	0.33
j	10.6	5.1	9.6	18.7	4.9	4.1	8.83	0.35
k	8.7	11.0	14.5	8.2	7.6	5.4	9.23	0.29
l	7.2	11.2	10.2	9.1	8.9	13.6	10.03	0.33
m	8.8	14.2	13.1	11.3	10.9	4.0	10.38	0.29

n	9.6	11.6	10.1	10.5	8.5	8.2	9.75	0.29
o	9.7	15.6	14.9	9.1	8.4	13.1	11.80	0.33
p	6.6	8.7	16.2	8.1	6.9	16.3	10.47	0.34
q	8.8	9.2	13.6	6.6	8.2	17.5	10.65	0.35

表11. IoC_{ij} を用いたクラスタ3の分析結果

	x3	x12	x13	x14	x15	x16	③Cluster 3の平均	①/①~③ の合計
a	2.1	1.3	0.9	1.9	4.1	4.1	2.40	0.08
b	1.4	0.5	1.0	0.3	3.8	4.7	1.95	0.06
c	0.8	2.6	0.1	0.2	3.2	2.2	1.52	0.05
d	1.9	0.7	0.4	0.9	5.0	4.7	2.27	0.07
e	1.9	0.6	1.6	1.4	3.4	4.8	2.28	0.07
f	1.4	0.5	0.4	0.8	4.6	5.0	2.12	0.07
g	1.8	0.6	1.3	1.2	4.1	4.7	2.28	0.07
h	1.9	0.5	2.7	0.2	1.9	2.5	1.62	0.05
i	2.0	1.6	1.0	1.6	2.9	4.0	2.18	0.07
j	5.0	1.7	0.2	2.2	1.5	1.3	1.98	0.08
k	2.5	0.6	0.7	1.6	4.3	5.5	2.53	0.08
l	1.0	1.2	1.2	0.5	3.6	3.4	1.82	0.06
m	1.9	0.2	0.1	0.7	3.4	4.7	1.83	0.05
n	1.8	2.2	0.0	0.6	4.9	4.2	2.28	0.07
o	3.1	0.1	0.4	0.9	6.4	3.7	2.43	0.07
p	1.9	1.3	1.3	2.6	4.5	6.5	3.02	0.10
q	1.8	0.6	6.4	2.0	4.5	6.0	3.55	0.12

6. おわりに

潜在因子の解釈に関する論述は、古くは1968年にJ. Scottらの論文から行われているが、従来の因子分析では、因子回転や因子数の選定に関する議論は数多く行われてきたが、抽出された潜在因子の解釈は分析者の主観で行われてきたケースが多い。しかし、主観が入ってしまう時点で科学的根拠が大きく損なわれてしまう可能性がある。

ゆえに、本研究では、主成分得点得点からユークリッド距離を求めて主成分得点距離行列を作成し、シルエット(silhouette)分析を用いてこの距離行列から最適なクラスタ数Kを抽出することで、最適な潜在因子の集合知を得るアプローチについて提案した。

今後の課題としては、複数の意味が混合する潜在因子が特定できたときに、その解釈が可能となるアプローチについて研究していきたい。また、本研究で実証したデータは個票データではなく要約されたデータであったため、今後は個票データでも本モデルの有用性を確認していきたい。

参考文献

- [1] Armstrong, J. S., & Soelberg, P.: On the interpretation of factor analysis. Retrieved from https://repository.upenn.edu/marketing_papers/14 (1968).
- [2] John T. Pohlmann: Use and Interpretation of Factor Analysis in The Journal of Educational Research: 1992-2002, The Journal of Educational Research, Volume 98, Issue 1, pp.14-23, 2004.
- [3] 柳井晴夫: “因子分析法の利用をめぐる問題点を中心にして”, The Annual Report of Educational Psychology in Japan, Vol. 39, pp.96-108 (2000).
- [4] Boomsma, A: The robustness of LISREL against small sample sizes in factor analysis models. In K. G. Jöreskog & H. Wold (Eds.), Systems under indirect observation: Causality, structure, prediction (part 1) (pp. 149-173). Amsterdam: North-Holland, 1982.
- [5] 堀啓造: 因子分析における因子数決定法, 香川大学経済論叢 第77巻 第4号, pp.35-37, 2005.
- [6] Lee, S-M., Terada, M., Shimizu, K., & Lee, M-H.: Comparative Analysis of Work Values Across Four Nations. Journal of Employment Counseling, 54, 132-144, 2017.
- [7] 中川有加, 西田みゆき, 柳井晴夫: “日本の看護学研究における因子分析法の利用”, 聖路加看護大学紀要 (31), pp.8-16 (2005).
- [8] 松岡緑, 西田真寿美, 関文恭: 因子分析による看護学生の老人像に関する研究, 九州大学医療技術短期大学部紀要 8, pp.35-43, 1981.
- [9] 三保紀裕・清水和秋: 大学進学理由と大学での学習観の測定 — 尺度の構成を中心として — キャリア教育研究, 29, pp.43-55, 2011.
- [10] 厚生労働省: 平成25年若年者雇用実態調査の概況の表 22, https://www.mhlw.go.jp/toukei/list/dl/4-21c-jyakunenkyouou-h25_gaikyou.pdf.
- [11] 首都大学東京授業用コンテンツ: 「相関分析」, <https://infolit.uec.tmu.ac.jp/lit/contents/office2010/statistics/04/>, 2019.5.3閲覧.
- [12] Michael Greenacre: Correspondence Analysis in Practice, Chapman and Hall, 2016.
- [13] 高根 芳雄: 多次元尺度法, 東京大学出版会, 1980.
- [14] Rousseeuw, P.J., ” Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”. J. Comput. Appl. Math., 20, pp.53-65, 1987.

付録 A. 性・年齢階級・最終学歴・雇用形態・初めて勤務した会社での勤続期間階級、最終学校卒業後初めて勤務した会社をやめた主な理由別在学していない若年労働者割合[10]

(「その他」や「不明」は潜在因子の意味解釈が困難なため除外した。またデータに-が含まれている項目や本分析に関係が無いと考えられる項目も除外した。)

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	
	仕事 が自 分に 合わ ない	自 分 の 技 能 ・ 能 力 が 活 か せ ら れ な か っ た	責 任 の あ る 仕 事 を 任 さ れ た か っ た	ノ ル マ や 責 任 が 重 い	会 社 に 将 来 性 が な い	賃 金 の 条 件 が よ く な か っ た	労 働 時 間 ・ 休 日 ・ 休 暇 の 条 件 が よ く な か っ た	人 間 関 係 が よ く な か っ た	不 安 定 な 雇 用 状 態 が 嫌 だ っ た	健 康 上 の 理 由	結 婚 、 子 育 て の た め	介 護 、 看 護 の た め	独 立 し て 事 業 を 始 め る た め	家 業 を つ ぐ 又 は 手 伝 う た め	社 に 長 く 勤 務 す る 気 が な か っ た た め	1 つ の 会 社 に 長 く 勤 務 す る 気 が な か っ た た め	倒 産 、 整 理 解 雇 又 は 希 望 退 職 に 応 じ た た め
a 男	22.0	8.9	2.1	10.0	15.2	22.1	22.7	15.7	8.7	7.3	3.0	1.3	0.9	1.9	4.1	4.1	
b 女	16.1	7.1	1.4	12.0	9.9	14.7	21.8	22.8	8.7	9.0	15	0.5	1.0	0.3	3.8	4.7	
c 20～24歳	23.3	5.6	0.8	9.4	12.4	20.1	23.8	22.7	8.0	9.0	5.0	2.6	0.1	0.2	3.2	2.2	
d 25～29歳	18.7	7.2	1.9	10.3	13.2	18.3	22.6	22.1	9.8	8.6	8.1	0.7	0.4	0.9	5.0	4.7	
e 30～34歳	17.4	8.5	1.9	12.1	11.9	17.3	21.8	16.7	8.1	7.9	11.9	0.6	1.6	1.4	3.4	4.8	
f 高校卒	19.0	6.6	1.4	9.3	11.6	17.9	20.3	19.6	9.3	7.0	10.3	0.5	0.4	0.8	4.6	5.0	
g 専修学校 (専門課程)修了	23.4	6.7	1.8	11.5	13.8	19.6	22.3	26.7	7.4	9.7	8.3	0.6	1.3	1.2	4.1	4.7	
h 高専・短大卒	14.3	8.2	1.9	11.7	8.3	21.5	20.8	24.0	8.6	9.5	15.6	0.5	2.7	0.2	1.9	2.5	
i 大学卒	19.7	10.1	2.0	14.1	13.9	16.4	25.3	15.5	6.9	9.5	8.7	1.6	1.0	1.6	2.9	4.0	
j 大学院修了	14.5	10.6	5.0	5.1 0	9.6	16.8	11.6	14.1	18.7	4.9	4.1	1.7	0.2	2.2	1.5	1.3	
k 正社員	18.9	8.7	2.5	11.0	14.5	20.1	24.7	18.2	8.2	7.6	5.4	0.6	0.7	1.6	4.3	5.5	
l 正社員以外	18.6	7.2	1.0	11.2	10.2	16.0	19.8	21.0	9.1	8.9	13.6	1.2	1.2	0.5	3.6	3.4	
m 6か月～ 1年未満	27.3	8.8	1.9	14.2	13.1	20.0	21.8	24.2	11.3	10.9	4.0	0.2	0.1	0.7	3.4	4.7	
n 1年～2 年未満	21.1	9.6	1.8	11.6	10.1	20.5	27.3	19.8	10.5	8.5	8.2	2.2	0.0	0.6	4.9	4.2	
o 2年～3 年未満	19.9	9.7	3.1	15.6	14.9	23.3	25.5	19.2	9.1	8.4	13.1	0.1	0.4	0.9	6.4	3.7	
p 3年～5 年未満	12.3	6.6	1.9	8.7	16.2	18.2	24.7	15.6	8.1	6.9	16.3	1.3	1.3	2.6	4.5	6.5	
q 5年～10 年未満	7.7	8.8	1.8	9.2	13.6	18.8	19.8	20.0	6.6	8.2	17.5	0.6	6.4	2.0	4.5	6.0	