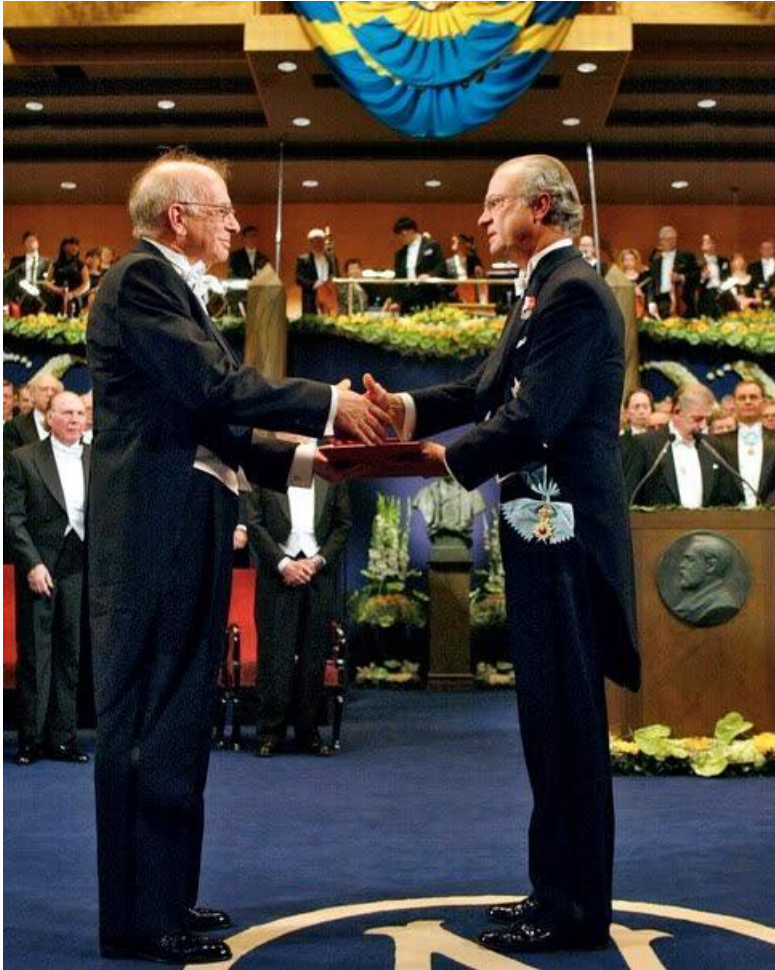


類似性判断 (similarity judgement) に関するモデルについて



- **Kahneman** & Tversky (1979) による計量経済学の研究が2002年にノーベル経済学賞を受賞
- Kahneman & Tverskyは両者とも**認知心理学者**
- 惜しいことにTverskyは受賞前に死去
- Tverskyは1992年に日本心理学会第56回大会で講演, 通訳は繁榘算男先生(青学 繁榘江里先生のお父上)
- Tverskyは類似性判断に関するモデル(feature contrast model)でも有名 → [資料教材pdf](#)

例題1

主成分分析(因子分析)のどこが好きなん？

以下、フィクションです

- あなたは米国FBI心理分析官
- きょうの分析対象は国家機密データ, コード名は「性格検査結果」
- 「性格検査結果」をスパイが国外に持ち出そうとしているとの通報あり
- そこで, スパイ容疑者宅に踏み込み「例題1」のデータを押収
- 「例題1」は「性格検査結果」とまったく似ていない(暗号化されているのでは?)
- そのため, スパイ容疑者は国家機密なぞ知らぬ存ぜぬ, わたしは善良な米国市民だと言い張っている

Q1) 「例題1」と「性格検査結果」の類似度(どのくらい似ているか)を示してください

コード名「性格検査結果」（実は公開データ, 出典は後で明示します）

番号	外向性	社交性	積極性	知性	信頼性	素直さ
1	3	4	4	5	4	4
2	6	6	7	8	7	7
3	6	5	7	5	5	6
4	6	7	5	4	6	5
5	5	7	6	5	5	5
6	4	5	5	5	6	6
7	6	6	7	6	4	4
8	5	5	4	5	5	6
9	6	6	6	7	7	6
10	6	5	6	6	5	5
11	5	4	4	5	5	5
12	5	5	6	5	4	5
13	6	6	5	5	6	5
14	5	5	4	4	5	3
15	5	6	4	5	6	6
16	6	6	6	4	4	5
17	4	4	3	6	5	6
18	6	6	7	4	5	5
19	5	3	4	3	5	4
20	4	6	6	3	5	4

スパイ容疑者が持っていた「例題1」

作家20人の作品コーパスから得た使用率の6変数データだと主張

1	著者ID	格助詞は[.]	格助詞に	動詞思う	助動詞だ	ら抜き	名詞止め
2	1	0.150	0.250	0.100	0.100	0.040	0.040
3	2	0.300	0.400	0.175	0.150	0.070	0.070
4	3	0.300	0.250	0.125	0.125	0.070	0.060
5	4	0.300	0.200	0.150	0.175	0.050	0.050
6	5	0.250	0.250	0.125	0.175	0.060	0.050
7	6	0.200	0.250	0.150	0.125	0.050	0.060
8	7	0.300	0.300	0.100	0.150	0.070	0.040
9	8	0.250	0.250	0.125	0.125	0.040	0.060
10	9	0.300	0.350	0.175	0.150	0.060	0.060
11	10	0.300	0.300	0.125	0.125	0.060	0.050
12	11	0.250	0.250	0.125	0.100	0.040	0.050
13	12	0.250	0.250	0.100	0.125	0.060	0.050
14	13	0.300	0.250	0.150	0.150	0.050	0.050
15	14	0.250	0.200	0.125	0.125	0.040	0.030
16	15	0.250	0.250	0.150	0.150	0.040	0.060
17	16	0.300	0.200	0.100	0.150	0.060	0.050
18	17	0.200	0.300	0.125	0.100	0.030	0.060
19	18	0.250	0.150	0.125	0.075	0.040	0.040
20	19	0.200	0.150	0.125	0.150	0.060	0.040
21	Who?	0.300	0.200	0.125	0.150	0.070	0.050

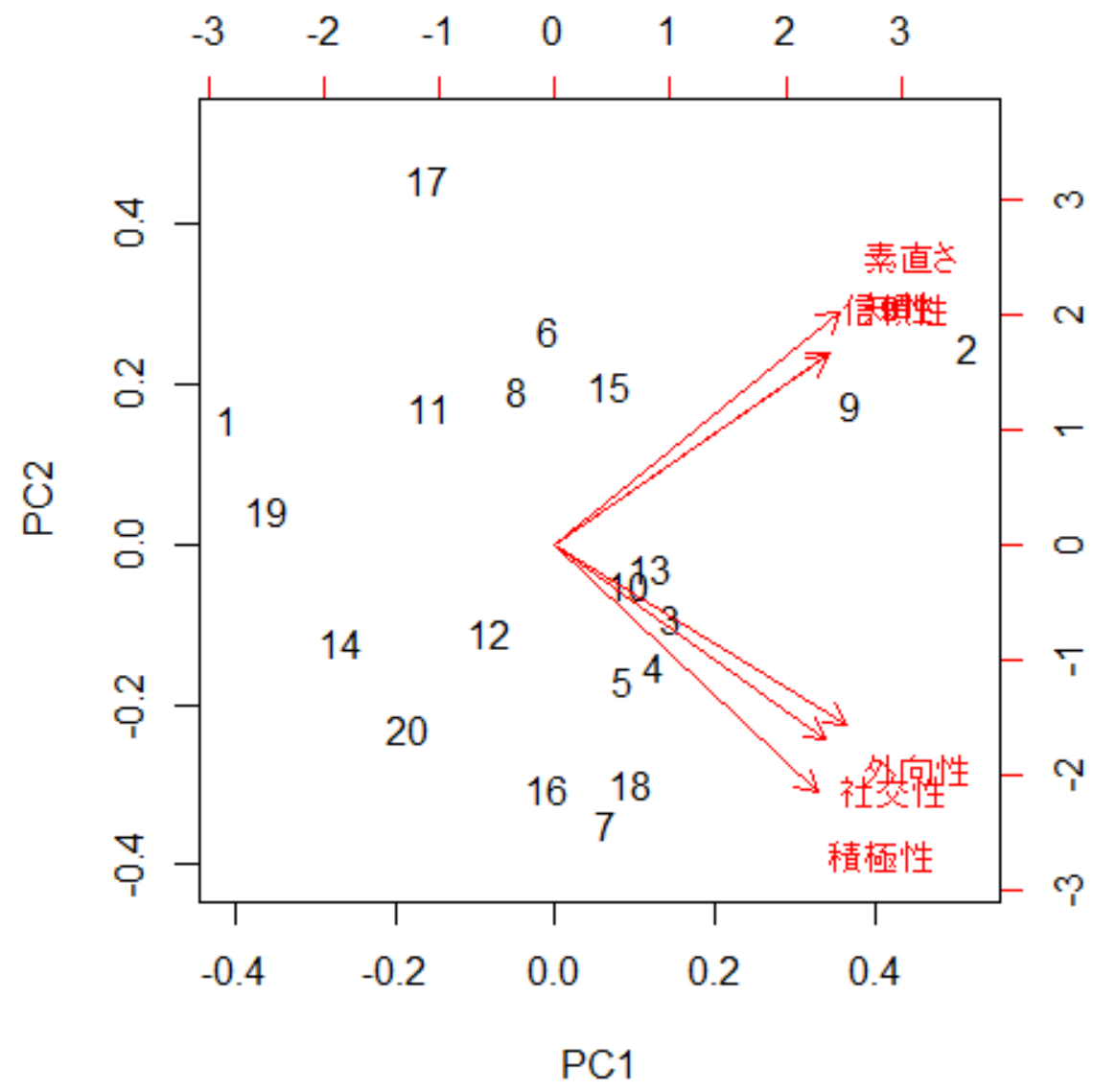
「性格検査結果」の一部と「例題1」の一部を並べてみると

1. まったく似ていない感じもするが、行や列を入れ替えているのかも
2. 「性格検査結果」と「例題1」の眼に見えない構造を視覚化したい

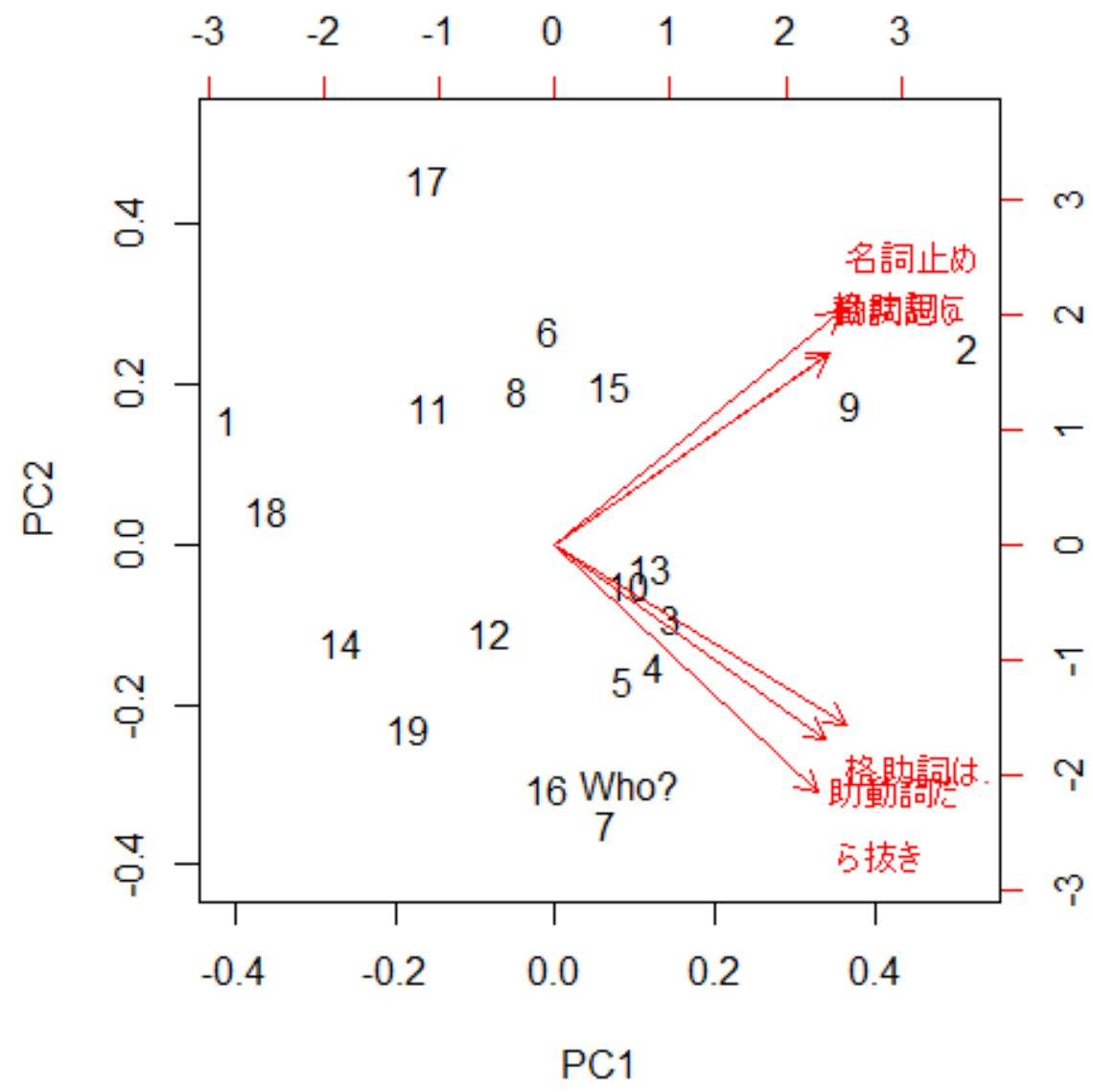
番号	外向性	社交性	積極性	知性	信頼性	素直さ
1	3	4	4	5	4	4
2	6	6	7	8	7	7
3	6	5	7	5	5	6
4	6	7	5	4	6	5
5	5	7	6	5	5	5
6	4	5	5	5	6	6
7	6	6	7	6	4	4

1	著者ID	格助詞は[.]	格助詞に	動詞思う	助動詞だ	ら抜き	名詞止め
2	1	0.150	0.250	0.100	0.100	0.040	0.040
3	2	0.300	0.400	0.175	0.150	0.070	0.070
4	3	0.300	0.250	0.125	0.125	0.070	0.060
5	4	0.300	0.200	0.150	0.175	0.050	0.050
6	5	0.250	0.250	0.125	0.175	0.060	0.050
7	6	0.200	0.250	0.150	0.125	0.050	0.060

「性格検査結果」を主成分分析し、結果を2次元空間に布置→次のスライドと比較



「例題1」を主成分分析→前のスライドと比較(18, 19, 20, Who?, ら抜き等に注目)



例題1

主成分分析(因子分析)のどこが好きなん？

- 「例題1」は「性格検査結果」とまったく似ていない
- 主成分分析(Principal Component Analysis: **PCA**)で解析し、結果を2次元空間に布置すると、両者はきれいに一致することが判明

Q1) 「例題1」と「性格検査結果」の類似度(どのくらい似ているか)を示してください

A1) 両者の潜在構造(眼には見えない構造)は同じ、**類似度は100%**

- ✓ 人工知能が語の意味を学習し、理解するときの基礎に**PCA**がある
- ✓ 「性格検査」データは小塩真司(おしお・あつし)先生がネットで公開しているものです
http://www.f.waseda.jp/oshio.at/edu/data_b/top.html 心理データ解析 第8回(2)分析例1(主因子解・バリマックス回転)