

例題2 意味空間 (Semantic Space) を体験

以下、フィクションです

- 時は1942年、あなたは英国海軍情報部の言語心理分析官 (コード名: くまのPoohさん)
- 分析対象はドイツ軍潜水艦U-Boatが発信する大量の暗号文
- 沈没したU-Boatから暗号解読のコードブックを入手することに成功したが、大部分の文字が消えていた
- 現時点で判明しているのは、(1)「V1, V2, …」は他動詞、(2)「N1, N2, …」は目的語、(3)「N1 = Katze = Cat」

Q) 今回の任務は「N19」が何を意味する名詞か推理すること

1. 他動詞と目的語の共起頻度をカウントし、その共起頻度表を「例題2」と命名
2. 「例題2」を主成分分析にかける => 意味空間を描く
3. 意味空間における単語の布置を注意深く観察する

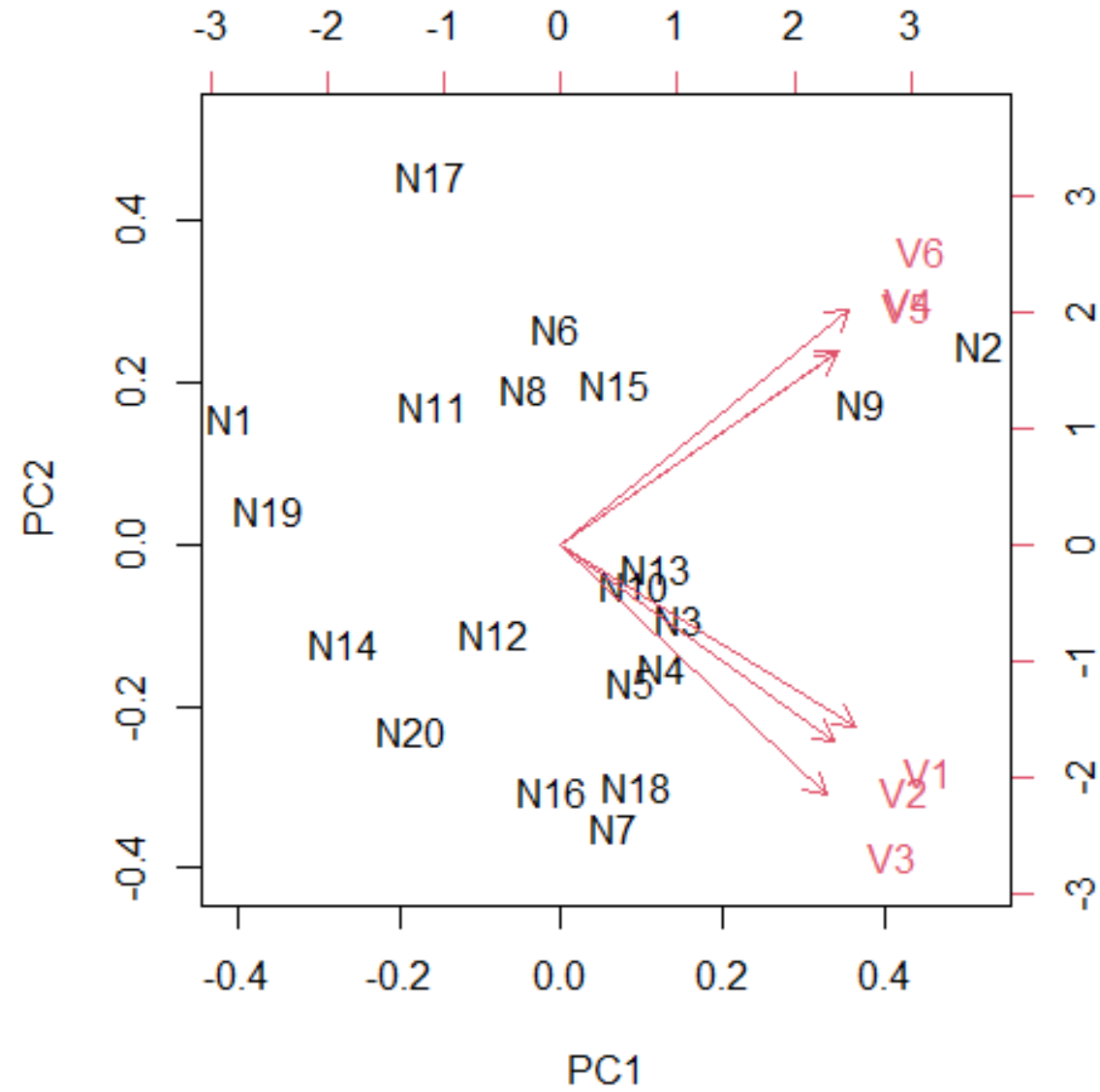
例題2

横山詔一2021作成

	V1	V2	V3	V4	V5	V6
N1	0	10	10	20	0	10
N2	30	30	40	50	30	40
N3	30	20	40	20	10	30
N4	30	40	20	10	20	20
N5	20	40	30	20	10	20
N6	10	20	20	20	20	30
N7	30	30	40	30	0	10
N8	20	20	10	20	10	30
N9	30	30	30	40	30	30
N10	30	20	30	30	10	20
N11	20	10	10	20	10	20
N12	20	20	30	20	0	20
N13	30	30	20	20	20	20
N14	20	20	10	10	10	0
N15	20	30	10	20	20	30
N16	30	30	30	10	0	20
N17	10	10	0	30	10	30
N18	30	30	40	10	10	20
N19	20	0	10	0	10	10
N20	10	30	30	0	10	10

N19とN1は似ていないように見える

「例題2」を主成分分析し、結果を2次元空間に布置



- 人工知能による単語の意味処理は Distributional Hypothesis (分布仮説) に立脚している。これは「**同じ文脈に出現する単語は、似たような意味を持つ傾向がある**」という仮説である (Harris, 1954)
- 上記の仮説に基づく意味論を「**分布意味論 (Distributional Semantics)**」という
- 今回は他動詞と目的語 (名詞) の共起頻度をカウントしたデータを主成分分析 (Principal Component Analysis: **PCA**) で解析

- 多くの単語をなるべく少ない次元の空間に布置 => これを「分布意味論に基づく意味空間」という
 - ✓ PCAにより空間の次元数を低減でき, データ操作が楽になる
 - ✓ PCAにより単語の潜在的意味(人間には理解不能!)を数値で表現できるようになる

Q) 今回の任務は「N19」が何を意味する名詞か推理すること

1. 「N19」に意味が近いのは「N1 = Katze = Cat」
2. よって「N19」は「Hund = Dog」かもネ

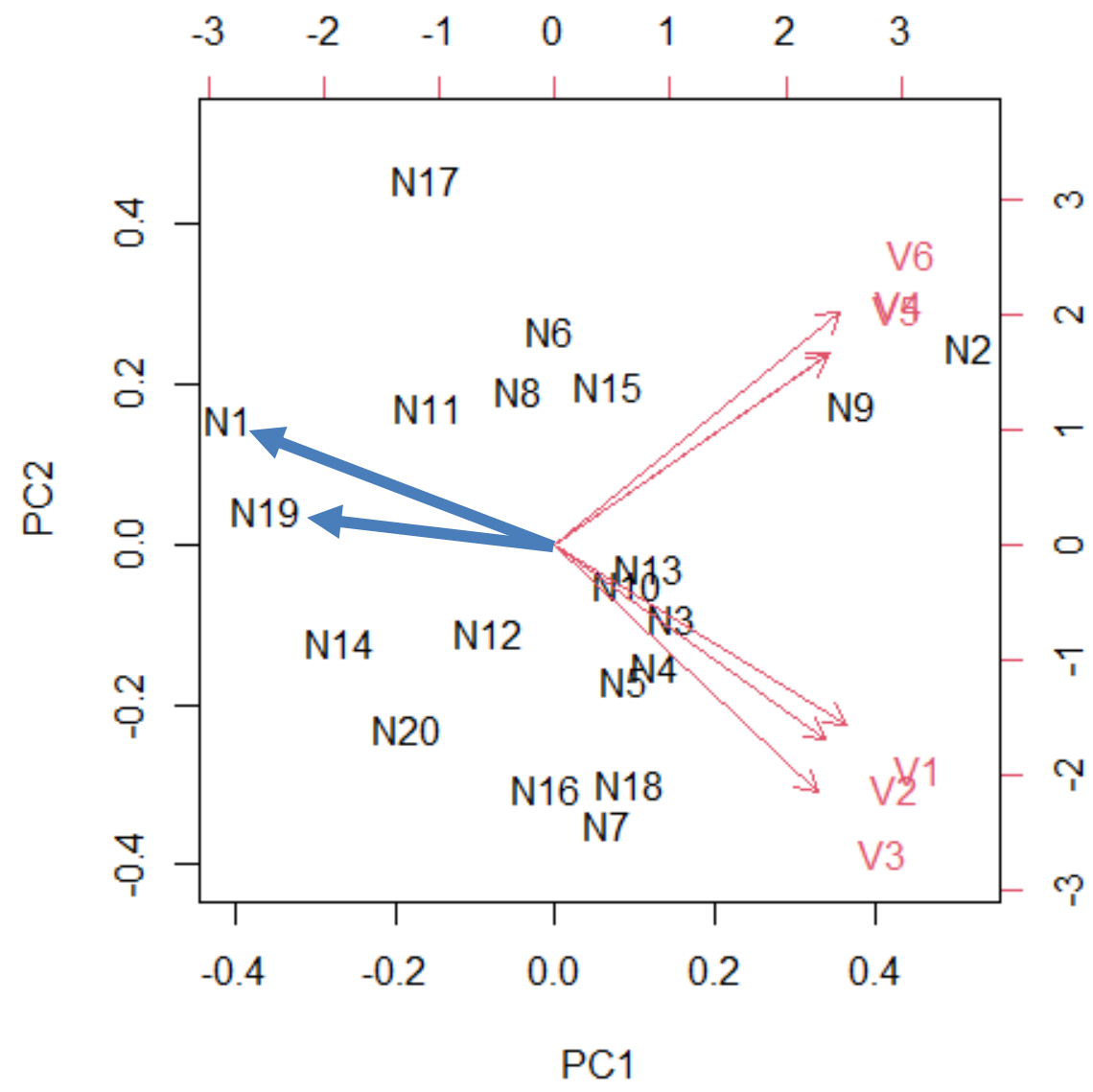
- 人工知能が語の意味を学習し, 理解するときの基礎にPCAがある

例題2

分布意味論に基づく意味空間と 単語のベクトル表現

- 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
- 単語の意味をどう表現・記述するか
 1. 国語辞典には語釈, 用例, 品詞などが記述されている
 2. 人工知能は意味空間の中心(原点)から単語の座標に矢印を引く \Rightarrow 単語をベクトルで表現・記述する方法(数値で表現できる), 次のスライド参照

「例題2」を主成分分析し、結果を2次元空間に布置→単語のベクトル表現の例



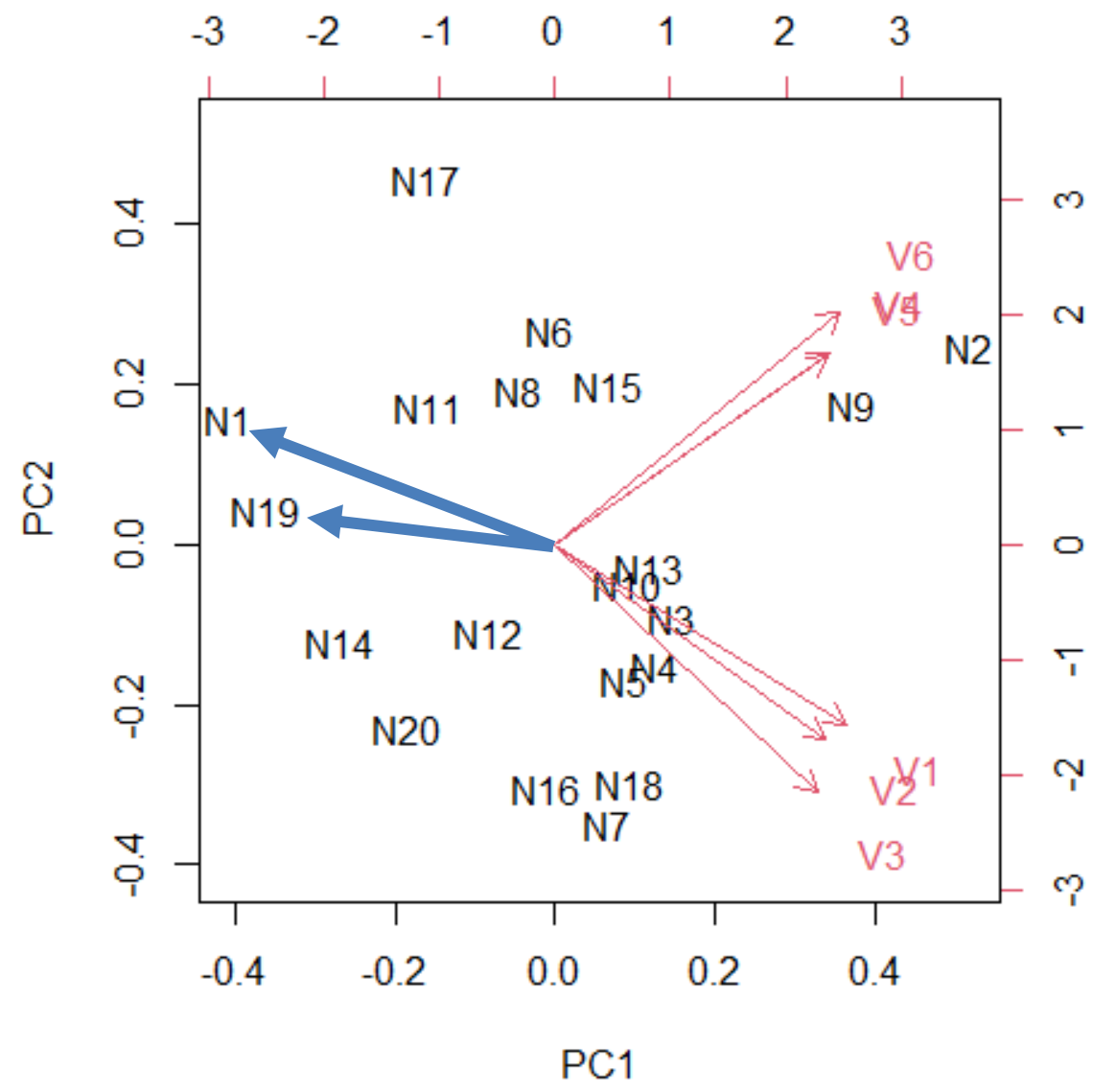
N1の意味はN1ベクトルで、
N19の意味はN19ベクトル
で表現・記述されている

例題2

分布意味論に基づく意味空間と 単語埋め込み (Word Embedding)

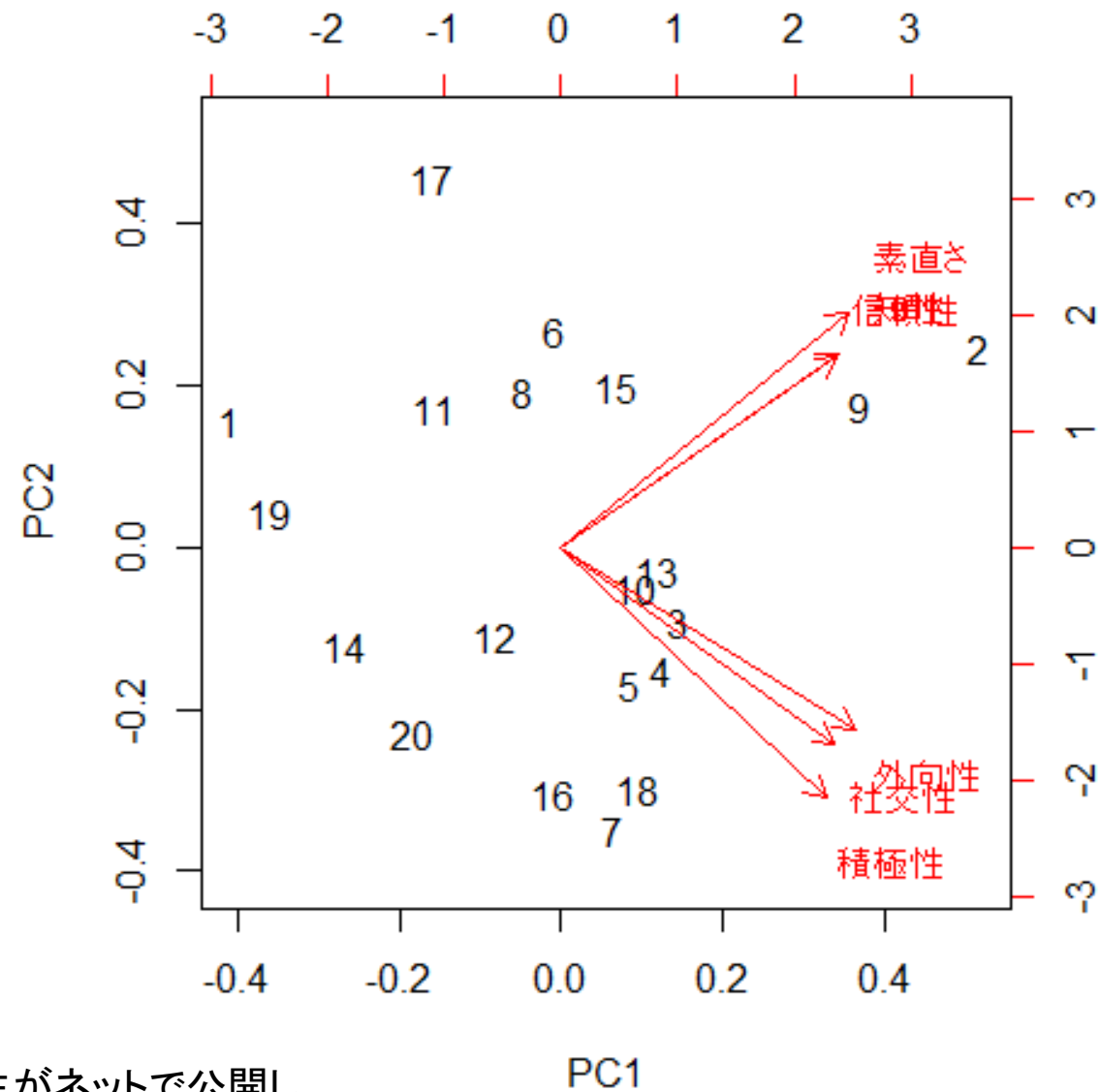
1. 単語のベクトル表現を「分散表現」とか「**単語埋め込み**」ということがある
 - 単語分散表現のイメージ => 単語の意味はその単語自体にあるのではなく、周囲の単語との共時性, 布置, 文脈のなかに散らばって存在する
 - **単語埋め込みのイメージ** => **単語のベクトル表現には周囲の単語との共時性, 布置, 文脈の情報のエッセンスが圧縮されて埋め込まれている**
2. ベクトル表現だとベクトルの足し算や引き算が簡単にできる
3. その結果, ご利益が! 驚くような性能を発揮
 - ✓ たとえば「王様 - 男性 + 女性 = 女王」などが瞬時にできる

「例題2」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



N1の意味はN1ベクトルで、
N19の意味はN19ベクトル
で表現・記述されている

「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



「性格検査」データは小塩真司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

「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)

