

文脈を反映した単語分散表現による順序効果のモデル化 Modeling Order Effects with Context-Reflecting Word Embeddings

前田 晃弘^{†*}, 鳥居 拓馬[‡], 日高 昇平[†]
Akihiro Maeda, Takuma Torii, Shohei Hidaka

[†] 北陸先端科学技術大学院大学, [‡] 東京電機大学, * 日本学術振興会特別研究員 DC
Japan Advanced Institute of Science and Technology, Tokyo Denki University, JSPS
akihiro.maeda@jaist.ac.jp

概要

単語類似度の知覚が単語を提示する順序に影響される現象（順序効果）を説明するため、文脈に対応する部分空間を反映した新たな単語分散表現のモデルを提案する。実験データを用いて部分空間の違いが非対称な類似度評価をもたらすか否か検証する。

キーワード：非対称な類似性、文脈、順序効果、ベクトル空間モデル、形式概念分析、量子認知、射影測定

1. はじめに

1.1 順序効果を示す現象

順序効果とは、人間の評価や判断が刺激の提示される順序に影響されることを指す。順序効果が発現する現象は知覚や意思決定の場で観察されることが知られている。例えば、Tversky (1977) は「中国は北朝鮮に似ている」と「北朝鮮は中国に似ている」の二文を比較するとき、実験被験者の大半が後者が際立つ (prominent) と回答したことを明らかにした。判断が対象の提示順序に左右されることは、類似性の知覚に方向性があり非対称である可能性を示唆する。

非対称な類似性を説明する理論は、心理学や認知科学の立場からこれまで提案されてきた。Tversky (1977) は、概念を特徴の集合と捉え、比較対象である二つの概念の集合の差から非対称な類似性が知覚されると主張した。一方、量子認知の立場に立つ Bussemeyer & Bruza (2012) は、非可換性を備えた量子確率理論を用いて、異なる部分空間（特徴に対応すると考えられる）の次元数から順序効果が現れる数理モデルを提案している。

1.2 単語分散表現とベクトル空間モデル

概念や単語の関係性に関する研究手法の一つに、個々の単語を一つのベクトル（分散表現）として表現するベクトル空間モデルがある。ベクトル空間モデルでは、単語や概念間の関係性をベクトル間の幾何的關係として捉えて、類似度をベクトル間の（ユークリッド）距離や角度（コサイン類似度）により定量化する。

対称性は距離が満たす公理の一つである。そのた

め、ユークリッド距離を用いた類似度評価は必然的に対称性を示す。また、コサイン類似度も二つのベクトルの角度の関数であるので対称性を示す。従って、2点の間の距離や角度を類似度とするベクトル空間モデルでは、順序効果を示す現象をモデル化することは不可能である。

1.3 先行研究の課題と本研究の提案

Martínez-Mingo et al. (2024) は量子確率モデルにおける類似度評価をベクトル空間モデルに適用した上で、コーパスから得られたデータを用いた順序効果の実証研究を試みている。単語ベクトルを部分空間へ分解するアルゴリズムによりベクトル空間モデルから量子確率モデルへの橋渡しをした上で、量子確率において射影測定と呼ばれる測定プロセスを用いて類似度を評価している。同提案には、1) 部分空間を抽出する手法の理論的根拠が乏しい、2) 類似度の評価を射影測定として捉える妥当性が検証されていない、3) 結果的に、提案モデルが順序効果の再現に失敗している、との課題を残している。

これまで著者らは、単語共起分布の数理的構造を解明するために、単語共起行列から数学的な内部構造（形式概念と呼ばれる）を抽出するアルゴリズムを提案し、それらが解釈可能な範疇あるいは文脈に対応することを実証的に明らかにした (Maeda et al., 2024)。その成果を踏まえ、本研究では、文脈に対応する部分空間を用いた新たな単語の分散表現を提案した上で、非対称な類似度を説明するモデルの構築を試みる。

2. 提案モデル

2.1 単語ベクトルと共起行列の内部構造

単語ベクトルは、コーパス中の単語共起頻度から学習され、文脈における単語分布の構造を反映していると考えられている (Lenci & Sahlgren, 2023)。Maeda et al. (2024) は、束論の一つである形式概念分析を適用して、形式概念の定義を満たす共起行列の部分行列を抽出するアルゴリズムを提案した。その上で、形式概念である部分行列に含まれる単語群が、人間に解釈

可能な範疇を構成することを実証的に明らかにするとともに、形式概念が文脈を反映することを指摘した。その提案手法は多義語の曖昧性解消にも有効であり、例えば単語 *tie* が持つ3つの異なる意味 (*necktie*, *fasten*, *draw*) とその文脈 (表1) に対応して、共起行列の中に形式概念の定義を満たす3つの部分行列が存在しており、単語 *tie* に対応する行ベクトルはそのいずれにも含まれることを明らかにした。

意味	文脈	形式概念に含まれる単語例
necktie	clothing	tie, jacket, pants
fasten	fixing act	tie, cable, loose
draw	competition	tie, teams, playoffs

表1: 多義語 *tie* を構成する形式概念

これは、多義語の単語ベクトルが異なる意味・文脈に対応する部分ベクトルを内包することを示唆する。

2.2 文脈を反映した単語分散表現のモデル化

本研究では、単語ベクトルが内包する文脈に対応した部分ベクトルを基底ベクトルとして、それらの張る部分空間により単語を表現することを提案する。

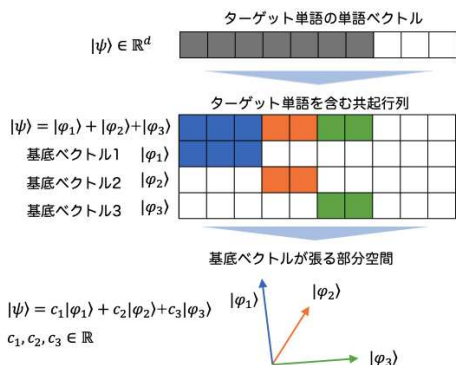


図1: 単語ベクトルから部分空間表現への変換

単語ベクトルから部分空間表現へ変換する具体的な手順を図1に示す。図の上段では、ターゲット単語が d 次元の行ベクトルとして、量子確率のケット表現を用いて $|\psi\rangle$ により表現されている。色はベクトルの成分が閾値以上であることを示す。中段は、4つの単語 (行) ベクトルを垂直に並べた行列を示し、一行目はターゲット単語である。色は閾値以上であることを示し、同色の行は同じ成分が揃って閾値以上であることを示す。行を適宜入れ替えた時に現れる同色の部分行列は形式概念に対応し、対応する単語が文脈を共有していることを示す (Maeda et al., 2024)。式 $|\psi\rangle = |\phi_1\rangle + |\phi_2\rangle + |\phi_3\rangle$ は、単語ベクトルが3つの部分行列 (形式概念) をそれぞれ構成する部分ベクトルに分解できることを示す。これらの部分ベクトルを基底として張る部分空間を

$|\psi\rangle = c_1|\phi_1\rangle + c_2|\phi_2\rangle + c_3|\phi_3\rangle$ ($c_1, c_2, c_3 \in \mathbb{R}$) として構成する。図1の下段は、ターゲット単語の部分空間表現である。

2.3 提案モデルの妥当性

部分空間表現はその背景にある文脈の構造を反映した表現を与えるという利点がある。特に、多義語における曖昧性解消や、意味合成の際の文脈に沿った意味抽出 (Selectional preferences; Lenci & Sahlgren, 2023) をモデル化でき、表現力を拡張すると考えられる。

また、通常の単語分散表現がコーパス中の異なる文脈の比率を反映したものである一方、部分空間として捉える表現は文脈を反映する基底ベクトルの任意の線形結合を同一視するので、コーパス中の文書構成に依存しないことが期待される。

3. 類似度評価方法の検討

次に、二つの単語の部分空間表現に対する類似度評価の方法を検討する。Martínez-Mingo et al. (2024) は、Pothos & Busemeyer (2011) に従い、類似度評価として量子確率論の射影測定を適用している。

3.1 量子確率における射影測定

量子確率は、量子系の状態や測定を記述する量子力学のための確率論である。単位ベクトルで表現される状態を正規直交基底へ射影することを測定として、特定の測定値を得る確率を記述することができる (石坂ら, 1999)。量子現象において同じ測定を連続して行うと同じ測定値が得られるという反復仮説をモデル化するため冪等性を持つ直交射影が導入されている。

Pothos & Busemeyer (2011) は、単語間の類似性評価を、量子確率における2回の連続した測定として定式化する。測定前の観測者の心的状態を初期状態 $|\psi\rangle$ として、各単語に対応する測定基底へ順々に射影を行い、量子確率に基づいて計算された確率を、単語間の順序づけられた類似性評価とみなす (図2)。具体的に

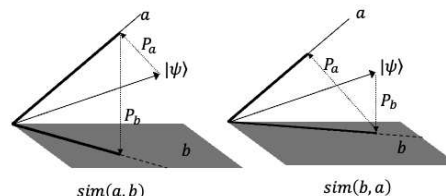


図2: 量子確率を用いた類似性評価の非対称性: 二つの部分空間への射影順序が異なる。

は、文 “Word a is similar to word b” に対する観測者の同意を、 $a \rightarrow b$ の順の類似性評価として、式 (1) により定義する。

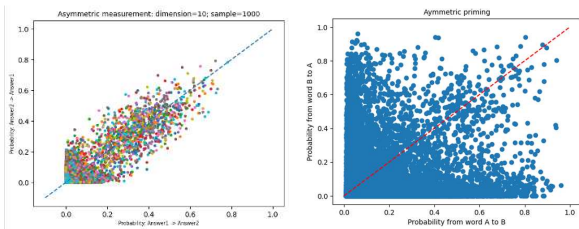
$$\text{sim}(a, b) := \|\mathbf{P}_b|\psi_a\rangle\|^2 \|\mathbf{P}_a|\psi\rangle\|^2 \quad (1)$$

なお、 $\mathbf{P}_a, \mathbf{P}_b$ は、それぞれ単語 a, b を構成する基底への射影行列であり、 $|\psi_a\rangle$ は、初期状態 $|\psi\rangle$ を \mathbf{P}_a で射影したのち、正規化して得られる新たな状態である。

3.2 射影測定による類似度評価の問題点

射影測定による類似度評価の妥当性は、先行研究において十分に検討されていない。順序効果を説明するベクトル空間モデルを構築すると本研究の立場から、その妥当性を検証するため、射影測定の数学的な挙動を調べる二つのシミュレーションを行った。

■シミュレーション 1 射影測定自体の挙動を調べるため、まず同じ個数の基底からなる 2 つの部分空間の間で計算される類似度の非対称性を調べた (図 3)。左図 3a は、 $d = 10$ 次元の正規直交基底を 2 組準備し、それぞれから単語 a, b の基底を一つずつランダムに抽出、また状態ベクトルもランダムに抽出した上で、式 (1) に従って計算された二つの値 $\text{sim}(a, b), \text{sim}(b, a)$ を (x, y) 座標上にプロットした。



(a) 射影測定による数値結果 (b) 人間の単語連想データ

図 3: 非対称な類似度の数値範囲

図 3b は、単語連想における非対称性を示した評価実験データ (USF データ; Nelson et al., 2004) を図示したものである。単語 a を刺激として単語 b を連想する確率を x 軸、反対に単語 b から単語 a を連想する確率を y 軸にプロットしている。順序効果は単語を順番に想起して意味認知を行う内的プロセスを反映していると仮定すれば、単語連想データを類似度評価の代替指標とすることは妥当と考えられる。USF データの具体例を表 2 に示す。 *japan* を見て *china* を連想する確率が、反対方向の確率の 5 倍近くと非対称的であり順序効果が観察される一方、男性・女性を表す単語対は高い対称性を示している。

類似度	単語 A	単語 B	$Pr(A \rightarrow B)$	$Pr(B \rightarrow A)$
対称的	girls	boys	0.50	0.50
	king	queen	0.77	0.73
非対称	assist	help	0.84	0.02
	japan	china	0.24	0.05

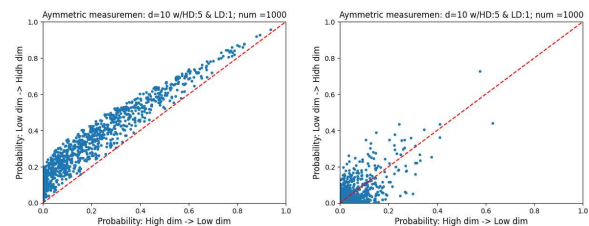
表 2: 単語連想の対称性を示す USF データ例

図 3 の結果は、数値ペアの取りうる範囲が、射影測

定 (図 3a) では、実験データ (図 3b) と比べて限定されることが明らかである。数学的にも、ベクトルの幾何学的配置より $\text{sim}(a, b)$ と $\text{sim}(b, a)$ の間に一定の制約があることを示すことができる。従って、この数値実験の条件下での射影測定では、観測される範囲の非対称性を説明できないと考えられる。

■シミュレーション 2 次に、部分空間をなす基底数の違いから非対称な類似性を説明できるかという観点より射影測定の妥当性を検討する。

シミュレーションの手順を以下に示す。二つの単語 a, b が、それぞれ m, n ($m > n$) 個の d 次元直交基底 $\{|u_i\rangle\}_{i=1}^m$ と $\{|v_i\rangle\}_{i=1}^n$ の張る部分空間により表現されるとして、ランダムに選択する。射影測定のための射影行列は、 $\mathbf{P}_a = \sum_{i=1}^m |u_i\rangle\langle u_i|$, $\mathbf{P}_b = \sum_{i=1}^n |v_i\rangle\langle v_i|$ である。次に初期状態 $|\psi\rangle$ を、シナリオ 1) 比較される二単語に対して中立的な位置にある初期状態を選択するよう、 $\|\mathbf{P}_a|\psi\rangle\|^2 = \|\mathbf{P}_b|\psi\rangle\|^2$ を満たす $|\psi\rangle$ を設定するケース (Pothos & Busemeyer, 2011) と、シナリオ 2) ランダムに抽出したベクトルを $|\psi\rangle$ とするケースにより設定する。最後に式 (1) により二単語間の類似性を評価する。



(a) 中立的な位置に置かれた初期状態 (b) ランダムに抽出された初期状態

図 4: 部分空間の違いに起因する類似度非対称性

左図 4a はすべての点が 45 度線より上に位置しており、すべてのサンプリングで $\text{sim}(a, b) < \text{sim}(b, a)$ であることを示す。これは基底数の少ない部分空間から多い部分空間への類似性評価が、顕著性の低い単語から高い単語への類似度知覚 (「北朝鮮から中国への類似性」) に対応して、反対方向の類似度よりも高いという Tversky (1977) の報告と整合する。一方、右図 4b は、初期状態をランダムに抽出した場合、類似度の非対称性の方向が一意に定まらないことを示唆する。また、人間の単語連想データの分布構造 (図 3b) と異なる。

3.3 部分空間の多寡による順序効果の説明可能性

前 2 節までの結果は、量子確率の射影測定を用いた類似性評価では、実際に観察される順序効果を十分に説明できない可能性を示唆する。また、評価前の初期

状態は評価対象である2つの単語に対して中立であるとの条件(シナリオ1のケース)は,その初期状態を特定するために,評価対象を事前に認知していることを暗黙に前提している。こうした暗黙の前提に基づく論法は,情報の提示順に対する認知過程の違いを説明する目的に照らせば,不整合であると考えられる。

一方,節1.で述べたように,先行研究は非対称な類似性を単語に含まれる特徴(文脈)の数の違いに求める。特に,非対称性の方向に関して,顕著性の高い概念は多くの特徴を有するので,特徴の少ない単語から多い単語への類似度が反対方向よりも大きいと考える。実際,表2のデータ例のうち,非対称な類似度を示す単語対では単語Aよりも単語Bの方が顕著性が高く,より多くの特徴(部分空間)を持つと予想される。実データの分布(図3b)の示す構造的特徴を再現する評価方法が求められる。

■シミュレーション3 これらに代わる類似度評価の方法として,初期状態を一方の単語の部分空間を構成する基底のセントロイドとして,他方の部分空間への射影測定を適用するプロセスを考える。図5に示すように,これにより初期状態の中立条件に依存せずに,低次の部分空間から高次の部分空間への類似度が高くなる傾向を示す結果を得ることができる。また,非対称性を表す分布の形状も図3bにより類似した傾向を示す。

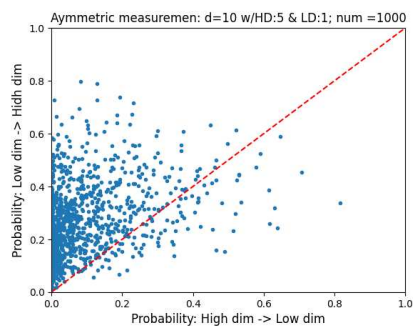


図5: セントロイドを状態とした時の非対称性

4. 考察と結論

本研究は,単語間の類似度評価が非対称性を示す順序効果を説明するために,先行研究が暗黙的に前提としてきた単語表現と類似度評価方法の仮定について,前者については単語共起行列中の数学的構造(形式概念)を用いた部分空間表現を提案するとともに,後者については先行研究で用いられた射影測定の連続適用に代わる新たな方法が必要であることを示した。

これらの検討を踏まえた今後の研究として,提案モデルに基づき実データから単語の部分空間表現を構築

して,USFデータに見られた非対称性の分布構造の再現を目指す。部分空間による表現は単語分散表現に内在する文脈構造を明示的に扱うことを可能にする。これは,量子確率が扱う「重ね合わせ」を想起させるが,射影測定に代わる測定が必要であった。類似度評価の知覚プロセスを測定としてモデル化することの認知的妥当性を検証するとともに,射影測定に限らない一般測定(石坂ら,1999)の設計可能性を検討する。

同様に,部分空間を扱える可能性のある数学として,幾何代数の活用が考えられる。幾何代数では多重ベクトルを扱うことができるので,例えば, m 個の基底からなる部分空間を m -vector として表現することが可能である。幾何代数は,幾何積と縮約(内積を一般化したもの)という演算を備えており,単語を表す二つの多重ベクトル間の演算を用いてそれらの関係の特徴づけることができる。特に,一重ベクトル同士の縮約は内積計算であり,単語の類似度評価にコサイン類似度が有効であることから,その一般化を考える。

謝辞

本研究は科研費基盤研究B(一般)JP23H0369,挑戦的研究(萌芽)JP22K19790,JST さきがけJP-MJPR20C9,JST CREST JPMJCR23P4,JSPS KAKENHI 24KJ1202の助成を受けて行われた。

文献

- [1] Tversky, A. (1977). Features of similarity. *Psychological Review* 84(4), 327–352.
- [2] Busemeyer, J., & Bruza, P. (2012). Quantum models of cognition and decision. *Cambridge University Press*
- [3] Martínez-Mingo, A., & Martínez-Huertas, J., & Olmos, R., & Jorge-Botana, G. (2024). Quantum projections on conceptual subspaces: A deeper dive into methodological challenges and opportunities. *Science Progress* 107(2),.
- [4] Maeda, A., & Torii, T., & Hidaka, S. (2024). Decomposing Co-occurrence Matrices into Interpretable Components as Formal Concepts. *The 62nd Annual Meeting of the Association for Computational Linguistics*, .
- [5] Lenci, A., & Sahlgren, M. (2023). Distributional semantics. *Cambridge University Press*.
- [6] Pothos, E., & Busemeyer, J. (2011). A quantum probability explanation for violations of symmetry in similarity judgments. *Proceedings of the annual meeting of the cognitive science society*, 33(33).
- [7] 石坂 智・小川 朋宏・河内 亮周・木村 元・林 正人(1999). 量子情報科学入門 共立出版
- [8] Nelson, D., & McEvoy, C., & Schreiber, T. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36(3) 402–407.