

通時的な類似度行列に基づく単語の意味変化の分析

木山朔¹ 相田太一² 小町守¹ 小木曾智信³ 高村大也⁴ 持橋大地^{3,5}
¹一橋大学 ²東京都立大学 ³国立国語研究所 ⁴産業技術総合研究所 ⁵統計数理研究所
hajime@scl.sds.hit-u.ac.jp

掲載号の情報

32 巻 4 号 p. 1189-1240.

doi: <https://doi.org/10.5715/jnlp.32.1189>

概要

単語の意味や関係性は時間とともに変化する。この現象を単語の通時的な意味変化という。複数時期にわたってどのように意味が変化するかを理解を目的とした研究は、意味変化の詳細な分析にとって重要である。しかし、隣接する時点間の変化点のみを検出する方法では、詳細な意味変化の分析には不十分である。また、BERT に基づく手法を用いて語義の割合を解析する方法には、計算コストが高くなるという問題が生じる。これらの課題に対処するために、本研究では軽量かつ高速な単語埋め込みを使用し、単語埋め込みに基づく単語の通時的な類似度行列を利用することで、複数の時系列にわたって意味変化のパターンを捉えるフレームワークを提案する。任意の単語に対して、時系列の各時点間で埋め込み間の類似度を算出することで、その類似度の変化パターンを意味変化のパターンとみなし、分析を行うことが可能となる。また、得られた複数の単語の類似度行列をクラスタリングすることにより、教師なしで類似度行列のパターンが似ている単語を算出できる。分析結果より、意味変化している単語と同じパターンである単語や、社会的な要因によって変化したと考えられる単語と同じパターンである単語をグループ化できることを示した。また、英語と日本語のデータセットに対して、複数の時期間隔において実験を行い、フレームワークの有効性とその限界について議論した。¹⁾

参考文献

- [1] 木山朔, 相田太一, 小町守, 小木曾智信, 高村大也, 持橋大地. 単語の通時的な類似度行列による意味変化パターンの分析. 研究報告自然言語処理 (NL-262), 第 34 巻, pp. 1-16, Dec 2024.
- [2] Hajime Kiyama, Taichi Aida, Mamoru Komachi, Toshinobu Ogiso, Hiroya Takamura, and Daichi Mochihashi. Analyzing continuous semantic shifts with diachronic word similarity matrices. In **Proceedings of the 31st International Conference on Computational Linguistics**, pp. 1613-1631, Abu Dhabi, UAE, January 2025. Association for Computational Linguistics.

1) 本論文は、情報処理学会第 262 回自然言語処理研究発表会で発表した論文 [1] および The 31th International Conference on Computational Linguistics (COLING 2025) で発表した論文 [2] をもとに和訳と加筆、修正を行ったものである。