

単語分散表現を用いた新型コロナウイルスによる意味変化検出

高橋寛武 石原祥太郎¹ 白井穂乃¹¹ 株式会社日本経済新聞社

hiromu.takahashi56@gmail.com {shotaro.ishihara,hono.shirai}@nex.nikkei.com

概要

本稿では日本語の新聞記事データを題材に、新型コロナウイルス感染症で意味が変化した単語を検出する手法を提案する。最初に、実データを用いて時系列の推移で単語分散表現の性能が悪化する可能性を検証し、単語の意味変化を検出する意義を議論した。次に先行研究に独自の工夫を施した手法を用いて、意味が変化した「コロナ」「まん延」といった単語を導出した。最後に単語分散表現全体の変化度合いを計算する方法を提案し、新型コロナウイルス感染症による単語の意味変化の大きさを確認した。

1 はじめに

新型コロナウイルス感染症の世界的な流行を背景に、日常生活にもさまざまな変化が生じている。変化したものの一つに言葉の使い方があり、新しい生活様式に応じた新語や流行語が登場した。2020年度の「ユーキャン新語・流行語大賞」では、新型コロナウイルス感染症の集団感染を避けるための行動を表す「3密（密閉、密集、密接）」が年間大賞に選ばれた¹⁾。「密」という語自体は従来から存在していたが、新しい意味を伴い日常で使用されている。

このような単語の意味変化は **semantic shifts** や **semantic change** と呼ばれ、言語学・社会学的な意義だけでなく、自然言語処理・検索などの情報科学の観点でも関心が寄せられている [1]。たとえば近年は大規模事前学習言語モデルが飛躍的な発展を遂げているが、このような言語モデルの性能は単語の意味変化に伴い悪化することが示されている [2][3]。

人間が扱う言語の変化については古くからさまざまな観点で研究されてきた [4]。近年は大規模なコーパスが利用可能になると共に、単語分散表現 [5] を用いて単語の意味変化を捉えようとする取り組みが加速した [6]。たとえば Gulordava ら [7] は、1960年代と 1990年代のコーパスにおける分布を比較し

1) <https://www.jiyu.co.jp/singo/index.php?eid=00037>

「sleep」という単語が睡眠障害に関してより否定的な意味合いになったという文化的な変化を確認した。単語分散表現の設計についても、コーパスを時間軸で分割するのではなく通時的に捉えたり [8]、時系列性に加えて文脈も考慮したり [9] と、意味変化の影響を軽減する取り組みも存在する。新型コロナウイルス感染症関連では Guo ら [10] が、Twitter²⁾ から抽出したコーパスを時系列で分析し、単語の意味の変化を観察した。日本語での研究事例は少なく、単語の意味変化の検出手法は確立されていない。同様に、時系列に沿った意味変化に伴う単語分散表現の性能悪化の可能性についても議論が少ない。

本研究では日本語の新聞記事データを題材に、新型コロナウイルス感染症で意味が変化した単語を検出する手法を構築する。第2章では初めに、時系列に沿った意味変化に伴う単語分散表現の性能悪化について検証し、近年の自然言語処理領域の文脈で単語の意味変化を検知する意義を議論する。次に第3章では提案手法を説明し、第4章では実験結果について述べる。最後に第5章では本稿を総括する。

2 単語の意味変化に伴う性能劣化

本章では、時系列に沿った意味変化に伴う単語分散表現の性能悪化を検証する。具体的には新型コロナウイルス感染症の流行前後における性能の違いと時系列ごとの性能の違いという2つの設定で予備実験を行う。

2.1 word2vec モデルの構築

単語分散表現の獲得には、先行研究でも数多く利用されている word2vec [5] を利用する。コーパスには、株式会社日本経済新聞社が提供するニュース配信サービス「日経電子版」³⁾の新聞記事データを用いる。取得期間はサービスが開始された2010年3月23日～2021年11月30日である。予備実験では、

2) <https://twitter.com>3) <https://www.nikkei.com/>

特定の期間を切り出して学習用に用いることで、単語分散表現の違いを検証する。

word2vec モデルの構築には、gensim[11] を利用する。テキストの前処理として HTML タグや URL を削除し、分かち書きには MeCab[12]、辞書には mecab-ipadic-NEologd[13] を用いる。

2.2 word2vec モデルの評価

日経電子版の新聞記事データを用いて、日経電子版の新聞記事データ（2010年3月23日～2019年10月31日）で学習した word2vec モデルの性能を検証する。比較対象には、WikiEntVec[14]、白ヤギ⁴⁾、chiVe[15] の3つの word2vec モデルを用いる。

モデルの比較には、記事に含まれる単語から記事のジャンルを予測する多クラス分類問題を設定する。キーワードそれぞれの単語分散表現の平均 [16] を入力とし、分類器には勾配ブースティング決定木の LightGBM[17] を用いる。キーワードは、記事本文から固有表現抽出した名詞を中心に、記者や編集者が人手で付与している。

ジャンルは8種類で、評価指標は正答率とした。検証用データセットとしては2020年1月1日～2021年11月30日の新聞記事データを使った。日経電子版の新聞記事データで構築した word2vec モデルは最も高い正答率を示し、今回構築した word2vec モデルは日経電子版の新聞記事データを用いた分析で使うモデルとして有用だと示唆された。評価実験の詳細は、Appendix A に示す。

2.3 予備実験 1: 流行前後の性能の変化

1つ目の予備実験では、新型コロナウイルス感染症の流行前後の word2vec モデルの性能の違いを比較する。具体的には、記事に付与されたキーワードを入力として、コロナ後の記事データを対象に記事のジャンルを予測するタスクに取り組む。これにより「新型コロナウイルス感染症の流行前後では単語の意味変化が多く、コロナ後の記事データを分析する場合、word2vec モデルとしてコロナ前モデルよりもコロナ後モデルを利用する方が予測性能が高い」という仮説を検証する。入力は2.2節の設定と同様、キーワードそれぞれの単語分散表現の平均を用い、分類器には LightGBM を用いる。

コロナ前モデルは2018年1月1日～2019年8月

表 1 新型コロナウイルス感染症の流行前後の性能の変化

	共通語以外含む	共通語のみ	すべて
前	0.7927	0.8291	0.8193
後	0.8019	0.8336	0.8258

表 2 2021年の記事データに対する性能の変化

年	word2vec のみ	word2vec と LightGBM を過去のデータで学習
	過去のデータで学習	
2011	0.8036	0.1886
2012	0.8060	0.1102
2013	0.8090	0.3768
2014	0.8087	0.3989
2015	0.8113	0.2234
2016	0.8157	0.4092
2017	0.8180	0.2610
2018	0.8193	0.3946
2019	0.8233	0.4684
2020	0.8284	0.5412

31日、コロナ後モデルは2020年1月1日～2021年8月31日の日経電子版の新聞記事データで学習する。検証用データセットとしては2021年9月1日～2021年11月30日の新聞記事データを用いる。ここでコロナ前後の変化を評価するため、共通語を定義する。共通語は2つの word2vec モデルで語彙が共通している語である。キーワードが全て共通語の19356記事（共通語のみ）、一つでも共通語ではない語を含む4258記事（共通語以外含む）、両者を合わせた23614記事（すべて）の3観点で性能を評価する。また、評価指標は正答率とする。乱数の seed を変えて10回実験した平均の結果を表1に示す。いずれの観点でも、コロナ後モデルの方が性能が高いという結果が得られた。特に共通語のみをキーワードに含む場合、つまり単語の意味変化による影響が最も強く出る条件下で性能差が確認された。

2.4 予備実験 2: 時系列の性能劣化の推移

2つ目の予備実験では Kutuzov ら [1] と同様に、各年1年分のデータで学習させたモデルを使って時系列の性能の変化を検証する。タスクは2.3節の予備実験と同じく、キーワードを入力として、LightGBM を用いて記事のジャンルを予測する。

検証では LightGBM の学習用にも過去のデータを使うか否かで2つの実験を設定する。過去のデータを使う設定（word2vec と LightGBM を過去のデータ

4) <https://github.com/shiroyagicorp/japanese-word2vec-model-builder>

で学習)では, word2vec モデルを各年のデータで学習し, LightGBM の学習・検証に 2021 年 1 月 1 日～11 月 30 日のデータを用いる. 過去のデータを使わない設定 (word2vec のみ過去のデータで学習) では, word2vec モデルと LightGBM を各年のデータで学習し, 検証には 2021 年 1 月 1 日～11 月 30 日のデータを利用する. 結果は表 2 の通りである. どちらの設定でも概ね学習時期が過去になるほど性能が劣化することが確認できた.

2.5 単語の意味変化を検知する意義

2 つの予備実験を通じ, word2vec モデルの性能が時間の経過に伴い劣化し, その一つの要因は単語の意味変化にあると示唆された. 近年の自然言語処理の領域では, 大規模なコーパスで事前学習した単語分散表現を活用する場面が多い. 単語の意味変化の検知は, 単語分散表現を活用するモデルの性能劣化を議論できる点で意義のある取り組みと言える.

3 提案手法

本章では, 提案手法である, 意味が変化した単語の検出手法について説明する. 新型コロナウイルス感染症における単語の意味変化を Twitter コーパスで分析した Guo ら [10] の先行研究を基に単語の意味変化の度合いを計算し, 本研究の目的に応じて単語の出現頻度による絞り込みを実施する.

3.1 ベクトル空間のマッピング

最初に, 単語の意味変化を検出するため, 学習済みの 2 つの word2vec モデルのベクトル空間に回転を施し, 同一の空間に落とし込む手法について説明する. まず, 意味変化の割合は単語頻度の負の累乗に従うという考え方 [18] に基づき, 登場頻度の高い単語の意味は時間の経過とともに変化せず局所的な構造は保存されていると仮定を置く. この仮定の下, Guo ら [10] が提案する手法を利用する. Guo ら [10] の手法は, 学習済みの 2 つの word2vec モデルを入力とし, 回転行列を導出して座標軸を揃える手法である. 具体的には, word2vec モデル i で出現頻度が上位 1000 件の単語行列 W_i と, word2vec モデル j で同じ 1000 個の単語から作った行列 W_j から, 特異値分解の応用で最適化問題を解く.

座標軸を揃えるマッピング処理を実施した後, 各モデルにおける単語分散表現のコサイン類似度を計算することで, 単語の意味変化の度合いを観測でき

る. しかし片方向のみの変換では類似度は平均して低くなるため [19], Guo ら [10] は逆方向に同様の処理を適用し, 単語の安定性 (*stab*) という観点で単語の意味変化を捉えている. モデル i とモデル j を比較した単語 w の *stab* を以下のように定義する. *stab* が小さいほど 2 つの word2vec モデルのベクトル空間での違いが大きく, 単語の意味変化しているとみなす.

$$stab(w) = \frac{sim_{ij}(w) + sim_{ji}(w)}{2}$$

$$sim_{ij}(w) = \cos - sim(R^{ji}R^{ij}V_w^i, V_w^j)$$

ただし, $\cos - sim$ はコサイン類似度, R^{ji} はモデル j から i のマッピングに用いた回転行列, V_w^i は単語 w のモデル i での単語分散表現を意味する.

3.2 単語の出現頻度による絞り込み

提案手法では, *stab* を利用して意味変化した単語を列挙する際, 出現頻度の情報を用いて, 過去のモデルでのみ数多く登場している一部の用語を除外する. 具体的には, コロナ後 (2020～2021 年) の word2vec モデルでの出現頻度 (100 万語当たりの調整頻度) の年間平均, およびコロナ前 (2010～2019 年) の年間平均と比較した増加率を各単語について計算し, それぞれの閾値で絞り込みを実施する. 単語の出現頻度による補正は, 自然言語処理分野では TF-IDF [20] などで伝統的に利用されている.

3.3 モデル全体の変化度合いの計算

単語の *stab* を利用したモデル全体の変化度合いの計算方法を提案する. モデル全体の変化度合いはモデル内の単語の変化度合いと相関しているとみなし, 全単語の *stab* の平均値を採用する. この値が小さいほど, モデル全体の変化度合いが大きい.

4 実験

提案手法の有効性を検証するため, 実際に新型コロナウイルス感染症の流行前後の word2vec モデルをマッピングする.

4.1 ベクトル空間のマッピング

実験ではコロナ前とコロナ後の 2 つのモデルを用いた. コロナ前モデルは 2010 年 3 月 23 日～2019 年 10 月 31 日, コロナ後モデルは 2020 年 1 月 1 日～2021 年 8 月 31 日の日経電子版の新聞記事データで



図1 マッピング後のベクトル空間の例

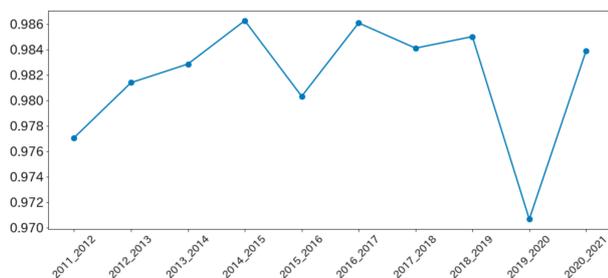


図2 モデル全体の変化度合いの年次推移

学習した word2vec モデルである。マッピングには、Hamilton ら [18] が公開している実装⁵⁾を用いた。

マッピング後のベクトル空間について、「コロナ」という単語に焦点を当てて可視化した結果を図 1 に示す。可視化のために、t-SNE[21]を用いて 300 次元のベクトル空間を 2 次元に圧縮した。左下の青色の四角で示した「コロナ」周辺の単語はコロナ前、右上の赤色の丸で示した「コロナ」周辺の単語はコロナ後の単語を表している。コロナ前は工業や飲食物の文脈で利用されていた単語が、コロナ後に感染症や災害の文脈に変わっていったことが読み取れる。

4.2 単語の出現頻度による絞り込み

stab が小さかった単語の一覧を確認すると、新型コロナウイルス感染症をはじめエボラ出血熱やデング熱など、感染症関連の単語ばかりが登場している (Appendix B)。提案手法である、単語の出現頻度の情報を用いて絞り込みを適用した結果を以下に示す。具体的には「コロナ後の word2vec モデルでの

5) <https://github.com/williamleif/histwords>

出現頻度の年間平均が 15 回より多い」かつ「増加率が 500 %より大きい」という条件を適用した。

単語の出現頻度による絞り込み

感染拡大, 新型コロナウイルス, 感染, 感染症, 感染者, ウイルス, ウイルス感染, ワクチン, 感染力, 変異, 新型肺炎, まん延, パンデミック, 感染予防, WHO, 接種, 世界保健機関, 感染リスク, 重症, 感染者数, 陽性, 新規感染者, 禍, 抗体, 感染対策, 緊急事態宣言, PCR 検査, ウイルス対策, コロナ, 陰性, 消毒, 隔離, 軽症, 外出自粛, 濃厚接触者, 検体, 病床, 医療従事者, 終息, 保健所, デルタ, アストラゼネカ, オンライン診療, クラスター, 武漢, マスク, ファイザー, 療養, 死者数, To, 2020, 回分, テレワーク, 経済活動, 自粛, 無観客, 人出, 往来, 収束, 休校, 巣ごもり, 都府県, リモートワーク, 発令

4.3 モデル全体の変化度合いの計算

最後に、モデル全体の変化度合いについて、1 年ごとのデータで学習した word2vec モデルを用いて検証する。2010~2021 年の各年のモデル全体の変化度合いの年次推移を図 2 に示す。たとえば 2010 年と 2011 年の変化度合いは、学習用に 2010 年と 2011 年のデータを用いた word2vec を比較することで算出した。ただし 2021 年については 11 月 30 日までのデータを利用している。2019 年と 2020 年の変化度合いが最も大きく、新型コロナウイルス感染症による単語の意味変化の大きさが確認できた。

5 おわりに

本稿では、日本語の新聞記事データを題材に、新型コロナウイルス感染症で意味が変化した単語を検出する手法を提案した。最初に、時系列に沿った単語の意味変化による分散表現の性能悪化を定量的に確認した。次に Guo ら [10] の研究を追試し、単語の出現頻度による独自の工夫を施した。最後に、単語分散表現全体の変化度合いを計算する方法を提案し、実験を通じて新型コロナウイルス感染症による単語の意味変化の大きさを確認した。

意味変化した単語を検出する手法は、新型コロナウイルス感染症以外の文脈でも適用できる可能性がある。ただし単語の出現頻度による絞り込みの際には閾値を適切に設定する必要がある、閾値の設定方法については今後の課題である。

参考文献

- [1] Andrey Kutuzov, Lilja Øvrelid, Terrence Szymanski, and Erik Velldal. Diachronic word embeddings and semantic shifts: a survey. In **Proceedings of the 27th International Conference on Computational Linguistics**, pp. 1384–1397, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics.
- [2] Shoetsu Sato, Jin Sakuma, Naoki Yoshinaga, Masashi Toyoda, and Masaru Kitsuregawa. Vocabulary adaptation for domain adaptation in neural machine translation. In **Findings of the Association for Computational Linguistics: EMNLP 2020**, pp. 4269–4279, Online, November 2020. Association for Computational Linguistics.
- [3] Kokil Jaidka, Niyati Chhaya, and Lyle Ungar. Diachronic degradation of language models: Insights from social media. In **Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics**, pp. 195–200, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [4] L. Bloomfield. **Language**. Holt, Rinehart and Winston, New York, 1933.
- [5] Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In **Workshop Track Proceedings of 1st International Conference on Learning Representations**, Scottsdale, Arizona, USA, May 2013.
- [6] Elizabeth Closs Traugott. Semantic change. In **Oxford Research Encyclopedia of Linguistics**. March 2017.
- [7] Kristina Gulordava and Marco Baroni. A distributional similarity approach to the detection of semantic change in the Google Books ngram corpus. In **Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics**, pp. 67–71, Edinburgh, UK, July 2011. Association for Computational Linguistics.
- [8] Alex Rosenfeld and Katrin Erk. Deep neural models of semantic shift. In **Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 474–484, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.
- [9] Valentin Hofmann, Janet Pierrehumbert, and Hinrich Schütze. Dynamic contextualized word embeddings. In **Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics**, 2021.
- [10] Yanzhu Guo, Christos Xypolopoulos, and Michalis Vazirgiannis. How COVID-19 is changing our language : Detecting semantic shift in twitter word embeddings. February 2021.
- [11] Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In **Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks**, pp. 45–50, Valletta, Malta, May 2010. ELRA.
- [12] Takumitsu Kudo. Mecab : Yet another part-of-speech and morphological analyzer. 2005.
- [13] Taiichi Hashimoto Toshinori Sato and Manabu Okumura. Implementation of a word segmentation dictionary called mecab-ipadic-neologd and study on how to use it effectively for information retrieval (in japanese). In **Proceedings of the Twenty-three Annual Meeting of the Association for Natural Language Processing**, pp. NLP2017–B6–1. The Association for Natural Language Processing, 2017.
- [14] Masatoshi Suzuki, Koji Matsuda, Satoshi Sekine, Naoaki Okazaki, and Kentaro Inui. A joint neural model for Fine-Grained named entity classification of wikipedia articles. **IEICE Transactions on Information and Systems**, Vol. E101.D, No. 1, pp. 73–81, 2018.
- [15] 真鍋陽俊, 岡照晃, 海川祥毅, 高岡一馬, 内田佳孝, 浅原正幸. 複数粒度の分割結果に基づく日本語単語分散表現. 言語処理学会第 25 回年次大会, pp. NLP2019–P8–5. 言語処理学会, 2019.
- [16] Dinghan Shen, Guoyin Wang, Wenlin Wang, Martin Renqiang Min, Qinliang Su, Yizhe Zhang, Chunyuan Li, Ricardo Henao, and Lawrence Carin. Baseline needs more love: On simple Word-Embedding-Based models and associated pooling mechanisms. In **Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics**, pp. 440–450, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [17] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A highly efficient gradient boosting decision tree. In **Advances in Neural Information Processing Systems**, Vol. 30. Curran Associates, Inc., 2017.
- [18] William L Hamilton, Jure Leskovec, and Dan Jurafsky. Diachronic word embeddings reveal statistical laws of semantic change. In **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics**, pp. 1489–1501, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [19] Hosein Azaronyad, Mostafa Dehghani, Kaspar Beelen, Alexandra Arkut, Maarten Marx, and Jaap Kamps. Words are malleable: Computing semantic shifts in political and media discourse. In **Proceedings of the 2017 ACM on Conference on Information and Knowledge Management**, pp. 1509–1518, Singapore, Singapore, November 2017. Association for Computing Machinery.
- [20] Juan Ramos. Using tf-idf to determine word relevance in document queries. In **Proceedings of the first instructional conference on machine learning**, Vol. 242, pp. 29–48. Citeseer, 2003.
- [21] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. **Journal of machine learning research**, Vol. 9, No. 11, 2008.
- [22] Yuya Sakaizawa and Mamoru Komachi. Construction of a Japanese word similarity dataset. In **Proceedings of the Eleventh International Conference on Language Resources and Evaluation**, Miyazaki, Japan, May 2018. European Language Resources Association.
- [23] Keisuke Inohara and Akira Utsumi. JWSAN: Japanese word similarity and association norm. **Language Resources and Evaluation**, pp. 1–29, 2021.

表3 word2vec モデルの比較

	モデル			
	日経	WikiEntVec	白ヤギ	chiVe
次元数	300	200	50	300
語彙数	493,531	1,015,474	335,476	3,644,628
JWSD-adv	0.281	0.182	0.155	0.255
JWSD-verb	0.251	0.149	0.223	0.260
JWSD-noun	0.274	0.250	0.203	0.310
JWSD-adj	0.287	0.158	0.257	0.404
JWSAN-2145	0.627	0.642	0.580	0.701
JWSAN-1400	0.499	0.499	0.416	0.541
NIKKEI	0.934	0.896	0.896	0.925

A word2vec モデルの評価実験

日経電子版の新聞記事データで構築した word2vec モデルの性能を他のモデルと比較した結果を表3に示す。モデルは以下の4つを用いた。

- 日経: 日経電子版の新聞記事データ (2010年3月23日~2019年10月31日) で学習
- WikiEntVec: 日本語 Wikipedia で学習
- 白ヤギ: 日本語 Wikipedia で学習
- chiVe: 国立国語研究所の日本語ウェブコーパス (NWJC) で学習

各モデルの評価には日本語単語類似度データセット「Japanese Word Similarity Dataset (JWSD)」[22]と、日本語類似度・関連度データセット「JWSAN」[23]、および日経電子版の新聞記事データを使った。JWSD は単語の類似度を0~10の値で付与したデータセットで、品詞は形容詞 (JWSD-adv)・動詞 (JWSD-verb)・名詞 (JWSD-noun)・副詞 (JWSD-adj) の4種類ある。JWSAN は名詞・動詞・形容詞の類似度・関連度のデータセットで、類似度と関連度がそれぞれ1~7の値で付与されている。すべての単語ペア2145組のデータセット (JWSAN-2145) と、分散表現に適したデータに厳選した1400組のデータセット (JWSAN-1400) が存在する。評価指標にはスピアマンの順位相関係数⁶⁾を用いた。

日経電子版の新聞記事データを用いたタスク (NIKKEI) では、記事に含まれる単語からジャンルを予測した。ジャンルは、企業・暮らし・国際・スポーツ・マーケット・経済・社会・政治の8種類で、評価指標は正答率を用いた。検証用データセットとしては2020年1月1日~2021年11月30日の新聞記事データを使った。

6) <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.spearmanr.html>

日経電子版の新聞記事データで構築した word2vec モデルは、その他のモデルと比較し遜色ない性能を示していると分かる。特に NIKKEI タスクでは4つのモデルの中で最も高い正答率で、今回構築した word2vec モデルは日経電子版の新聞記事データを用いた分析で使うモデルとして有用だと示唆された。

B stab が小さい単語上位100件

stab が小さい単語上位100件には、新型コロナウイルス感染症をはじめエボラ出血熱やデング熱など、感染症関連の単語ばかりが登場した。これらの単語のうち一部はコロナ後モデルの語彙としての登場頻度が多くなかったため、本研究の目的を考慮し、登場頻度に基づく絞り込みを実施した。

stab が小さい単語上位100件

感染拡大, 新型コロナウイルス, 感染, コロナウイルス, 感染症, 感染者, ウイルス, インフルエンザウイルス, ニパウイルス感染症, エボラ出血熱, ウイルス感染, MERS, デング熱, リフトバレー熱, ジカ熱, 豚コレラ, エボラ熱, 重症熱性血小板減少症候群, エボラウイルス, クリミア・コンゴ出血熱, ワクチン, アフリカ豚コレラ, 出血熱, 鳥インフル, 耐性菌, 感染力, ツツガムシ病, 鳥インフルエンザ, 変異, インフルエンザ, HBV, 中東呼吸器症候群, ラッサ熱, ポリオウイルス, ジカウイルス感染症, 国立感染症研究所, マラリア, 感染ルート, 新型肺炎, エボラウイルス病, 結核, 院内感染, アデノウイルス, 病原体, 原虫, 麻疹, 熱帯病, 感染経路, エボラ, マイコプラズマ, 蚊, チクングニア熱, インフル, はしか, 病原性, 髄膜炎菌, フィラリア症, 新型インフルエンザ, HIV, ハマダラカ, 感染源, 麻疹ウイルス, 黄熱, ヒト免疫不全ウイルス, 新興感染症, エイズ, 多剤耐性菌, まん延, 空気感染, H7N9, ムンプス, パンデミック, ノロウイルス, 日本脳炎, マールブルグ, プリオン病, 感染地域, 狂犬病, 水疱瘡, 感染予防, マーズ, 薬剤耐性菌, C型肝炎, 呼吸器感染症, 豚流行性下痢, ロタウイルス, 成人T細胞白血病, デングウイルス, E型肝炎, 寄生虫, 日本紅斑熱, フィラリア, 世界的大流行, WHO, 抗生物質, A型肝炎, 伝染病, 急性感染症, SFTS, 流行