

# 通時的な単語の意味変化を捉える単語分散表現の同時学習

相田太一<sup>1</sup> 小町守<sup>1</sup> 小木曾智信<sup>2</sup> 高村大也<sup>3</sup> 持橋大地<sup>4</sup>  
東京都立大学<sup>1</sup> 国立国語研究所<sup>2</sup> 産総研/東京工業大学<sup>3</sup> 統計数理研究所<sup>4</sup>  
aida-taichi@ed.tmu.ac.jp komachi@tmu.ac.jp togiso@ninja1.ac.jp  
takamura.hiroya@aist.go.jp daichi@ism.ac.jp

## 1 はじめに

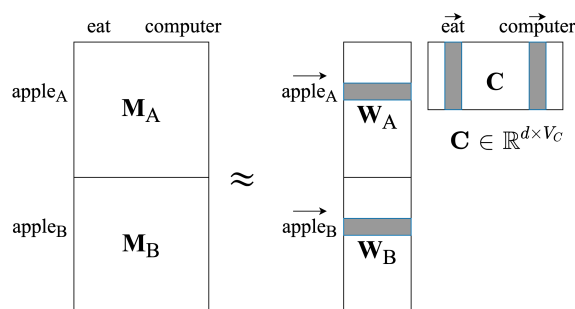
言語は時代とともに変化するものであり、現代社会においても、日々新しい単語が生まれている。既存の単語についても、ある単語が時間の経過とともに、全く異なる意味で使われる場合も少なくない。例えば、“gay”という単語は元々「陽気な」という意味で用いられていたが、近年では「同性愛」という意味が主に使われるようになった。このような通時的な単語の意味の変化を捉えることができれば、昔の時代の文書への分野適応や、辞書学での単語の意味変化に関する記述への利用などが期待できる。

近年では、通時的に学習した単語分散表現を用いて単語の意味変化を検出する手法が数多く提案されている。Kulkarni ら [1] や Hamilton ら [2] は任意の時期毎に学習した単語分散表現に対して、線形変換によって対応付けを行う手法を提案した。また、Yao ら [3] によって各時期の単語分散表現を同時に獲得する手法が提案された。これまでの手法は文脈を考慮していない単語ベクトルを用いているため、単語の用法毎の変化を調査できなかったが、BERT [4] などの事前学習済み言語モデルによって文脈を考慮した単語ベクトルを生成できるようになった。最近では英語や日本語などの代表的な言語で事前訓練済みのモデルが公開されたこともあり、BERT を用いた研究も行われている [5, 6, 7]。

しかし、こうした単語の意味変化を通時的に分析するための手法には、以下のような問題がある。

- 単語ベクトル間の関係を線形モデルで表せるという強い仮定をおいている [1, 2]
- ハイパーパラメータの設定に敏感である [3]
- 公開されているモデルの言語に依存する [5, 6, 7]
- 手法間での定量的な比較が行われている研究が少ない

こうした問題に対し、線形モデルで表現可能とい



$$\mathbf{M}_A, \mathbf{M}_B \in \mathbb{R}^{V_w \times V_c} \quad \mathbf{W}_A, \mathbf{W}_B \in \mathbb{R}^{V_w \times d}$$

図 1: Word2vec と SVD の等価性を利用した時期ごとの単語ベクトルの獲得 PMI-SVD<sub>joint</sub> の様子。

う仮定やハイパーパラメータに敏感であるという問題を解決するため、我々は Levy ら [8] により示された Word2vec と特異値分解の等価性を用い、図 1 のように通時的な単語分散表現を同時に獲得する手法を提案し、日本語の文書データに対して戦前と戦後における単語の意味変化を網羅的に分析した [9]。

本研究ではこの手法を拡張し、また実際に意味が変化した単語のリストを用いることで定量的な評価を行い、提案手法と先行研究の手法を比較する。(1) 定量的な評価を行った結果、提案手法は既存の手法と同等以上の性能を獲得した。また、(2) 実際に意味の変化した単語に対して定性的な評価を行った結果、事前訓練済みの BERT よりも効果的に意味変化を捉えていることを示した。

## 2 関連研究

Hamilton ら [2] は、学習した各時期の単語分散表現  $\mathbf{W}_t, \mathbf{W}_{t+1}$  を回転行列  $\mathbf{R}$  で対応付ける手法を提案した。

$$\mathbf{R}(t) = \operatorname{argmin}_{\mathbf{R}: \mathbf{R}\mathbf{R}^T = \mathbf{I}} \|\mathbf{W}_t \mathbf{R} - \mathbf{W}_{t+1}\|_F^2 \quad (1)$$

ここで、 $\|\cdot\|_F$  はフロベニウスノルムである。

また、Yao ら [3] は各時期で獲得した Positive PMI (PPMI) 行列に対して、以下の目的関数を最小化することで、各時期の単語分散表現  $\mathbf{W}_t$  および文脈語の

分散表現  $\mathbf{C}_t$  を同時に獲得する手法を提案した。

$$\begin{aligned} & \frac{1}{2} \sum_{t=1}^T \|\mathbf{M}_t - \mathbf{W}_t \mathbf{C}_t\|_F^2 + \frac{\gamma}{2} \sum_{t=1}^T \|\mathbf{W}_t - \mathbf{C}_t^T\|_F^2 \\ & + \frac{\lambda}{2} \sum_{t=1}^T \|\mathbf{W}_t\|_F^2 + \frac{\tau}{2} \sum_{t=1}^{T-1} \|\mathbf{W}_{t+1} - \mathbf{W}_t\|_F^2 \\ & + \frac{\lambda}{2} \sum_{t=1}^T \|\mathbf{C}_t\|_F^2 + \frac{\tau}{2} \sum_{t=1}^{T-1} \|\mathbf{C}_{t+1} - \mathbf{C}_t\|_F^2 \quad (2) \end{aligned}$$

ここで、 $\gamma, \lambda, \tau$  はハイパーパラメータである。

近年では、文脈を考慮した単語ベクトルを生成できる事前学習済み言語モデルである BERT を用いた研究が行われている。単語の語義単位の意味変化を扱う研究として、単語の語義ごとに辞書の例文からベクトルを計算するもの [5] や、BERT によって得られる単語ベクトルを語義単位のクラスタリングするもの [7] がある。また、単語の代表的な意味の変化を調べる研究として、BERT によって得られるベクトルを平均することで単語の代表的なベクトルを獲得するもの [6] がある。本研究では、単語の代表的な意味の変化に着目した比較および分析を行う。

### 3 提案手法

#### 3.1 準備：PMI-SVD [8]

まず、基盤となる手法である Levy ら [8] の単語分散表現学習手法について説明する。コーパス全体において、単語  $w$  とその周辺に現れる文脈語  $c$  との共起確率を  $p(w, c)$ 、 $w$  と  $c$  それぞれの出現確率を  $p(w), p(c)$  としたとき、単語ベクトルの学習は、Shifted Positive PMI (SPPMI)

$$M[w, c] = \max \left( \log \frac{p(w, c)}{p(w)p(c)} - \log k, 0 \right) \quad (3)$$

を要素とする  $V_W \times V_C$  の行列  $\mathbf{M}$  ( $V_W$  は対象語の語彙サイズ、 $V_C$  は文脈語の語彙サイズを示す) を、 $\mathbf{M} \approx \mathbf{W}\mathbf{C}$  と  $d$  次元に特異値分解したときの  $V_W \times d$  の行列  $\mathbf{W}$  の各行に等しい [8]。同様に、文脈語ベクトルは  $d \times V_C$  の行列  $\mathbf{C}$  の各列として獲得できる。 $\mathbf{W}$  の列数 (および  $\mathbf{C}$  の行数)  $d$  は単語ベクトルの次元数を示しており、以下本研究では  $d=100$  とした。式 (3) の定数  $k$  は Word2vec の負例サンプリングにおける負例数に相当し、以下  $k=1$  とした。

#### 3.2 PMI-SVD<sub>joint</sub> [9]

この方法を拡張すると、時期の違う単語ベクトルを同時に計算することが可能になる。文脈語の分散

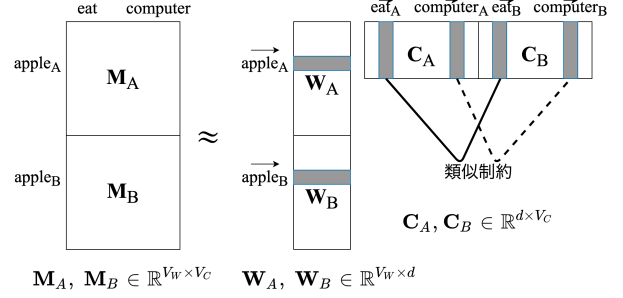


図2: 文脈語の意味変化も考慮する提案手法 PMI-SVD<sub>c</sub> の行列分解の模式図。

表現  $\mathbf{C}$  の各列である文脈ベクトルが変化しないと仮定すると、時期 A (たとえば明治時代) における PMI 行列を  $\mathbf{M}_A$ 、時期 B (たとえば平成時代) における PMI 行列を  $\mathbf{M}_B$  とすれば、 $\mathbf{M}_A$  と  $\mathbf{M}_B$  を縦に結合した  $\mathbf{M} = [\mathbf{M}_A; \mathbf{M}_B]$  も同様に

$$\begin{bmatrix} \mathbf{M}_A \\ \mathbf{M}_B \end{bmatrix} \approx \begin{bmatrix} \mathbf{W}_A \\ \mathbf{W}_B \end{bmatrix} \begin{bmatrix} \mathbf{C} \end{bmatrix} \quad (4)$$

と行列分解することができる (図 1)。このとき  $\mathbf{W}_A$  および  $\mathbf{W}_B$  の対応する行が、時期 A と時期 B の同じ単語の単語ベクトルとなり、後処理による近似的な対応づけは必要としない [9]。式 (4) の計算は、 $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$  と特異値分解を行った後で、 $\mathbf{W} = \mathbf{U}\mathbf{\Sigma}^{1/2}$ 、 $\mathbf{C} = \mathbf{\Sigma}^{1/2}\mathbf{V}^T$  とすることで行える。

#### 3.3 PMI-SVD<sub>c</sub>

上の手法では、時期が経過しても文脈語は意味が変化しないという仮定を置いていた。この仮定を避けるために、次に文脈語の意味変化も考慮するモデルを提案する。文脈語の意味変化を単純に考慮するのであれば、単語ベクトル行列と同じ数の文脈語ベクトル行列を作成すれば良い。しかし、各時期の PMI 行列  $\mathbf{M}_t$  を個別に行列分解するだけでは、時期間での対応が取れない。そこで、通常 of 行列分解 (式 (4)) の目的関数に、隣接する時期間で文脈語ベクトルが類似しているという制約項を次式で追加する。

$$\sum_{t=1}^T \|\mathbf{M}_t - \mathbf{W}_t \mathbf{C}_t\|_F + \tau \sum_{t=1}^{T-1} \|\mathbf{C}_{t+1} - \mathbf{C}_t\|_F \quad (5)$$

ここで  $\tau$  は、制約の強さを決めるハイパーパラメータである。これは Yao ら [3] のモデル (式 (2)) と比べて簡略化されており、ハイパーパラメータの探索に必要な試行数を大きく減らしつつ、以下のように実験的にも同等以上の性能を示した。

## 4 実験

擬似的に意味変化を生成したデータおよび、実際に意味が変化した単語のリストを用いた定量的な評価 [1] により提案手法と既存手法の比較を行った。

### 4.1 データ

日本語と英語のデータを用いて、2つの時期における単語の意味変化について実験を行った。日本語では、『日本語歴史コーパス』<sup>1)</sup>の一部として公開されている近代雑誌コーパス<sup>2)</sup>に、「昭和・平成書き言葉コーパス」として構築中の雑誌(『中央公論』『文藝春秋』)データを追加したものを、戦前(1895–1944年)と戦後(1945–1997年)に分けて用いた。英語では、Corpus of Historical American English<sup>3)</sup>(COHA)の1900年代と1990年代を用いた。

各時期の文書で100回以上出現する名詞・動詞・形容詞・副詞を分析対象の単語とした。また、文脈語には分析対象の単語と同じものを用いた。

### 4.2 比較手法

提案手法である PMI-SVD<sub>joint</sub>、PMI-SVD<sub>c</sub> と以下の既存手法を比較した。

- **PMI-SVD<sub>align</sub>** [2]: PMI-SVD を各時期で訓練し、回転行列で対応付けを行う (式 (1))。
- **Word2Vec<sub>align</sub>** [2]: PMI-SVD<sub>align</sub> と同様だが、PMI-SVD の代わりに Word2Vec skip-gram negative-sampling を訓練する。
- **Dynamic Word Embeddings** [3]: 式 (2) を最小化することで各時期の単語分散表現を獲得する。
- **BERT** [4]: 各時期の各単語を代表するベクトルは平均によって獲得した [6]。本実験では、huggingface<sup>4)</sup> で公開されている事前訓練された BERT を使用した。

### 4.3 評価

最初に、各モデルで語彙中の全ての単語について2つの時期間の余弦類似度が低い順にランキングを行い、リストを作成した。次に、このリストの上位  $k$  単語と、実際に意味が変化した単語のリストとの一致率を計算して評価した (Recall@ $k$ )。

1) [https://pj.ninjal.ac.jp/corpus\\_center/chj/](https://pj.ninjal.ac.jp/corpus_center/chj/)  
2) [https://pj.ninjal.ac.jp/corpus\\_center/cmj/woman-mag/](https://pj.ninjal.ac.jp/corpus_center/cmj/woman-mag/)  
3) <https://www.english-corpora.org/coha/>  
4) <https://github.com/huggingface/transformers>

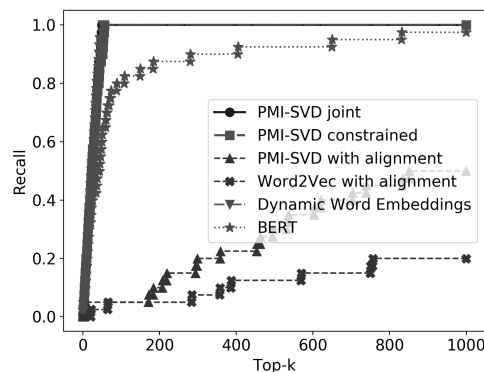


図 3: 日本語のデータで擬似的に意味の変化する単語を生成し、Recall@ $k$  で評価した結果。

表 1: 日本語のデータにおける平均再現率。

モデル	擬似 (4.4)	実際 (4.5)
PMI-SVD <sub>joint</sub>	<b>0.995</b>	0.621
PMI-SVD <sub>c</sub>	<b>0.995</b>	0.579
PMI-SVD <sub>align</sub> [2]	0.752	0.601
Word2Vec <sub>align</sub> [2]	0.574	0.525
Dynamic Word Embeddings [3]	<b>0.995</b>	0.509
BERT [6]	0.973	<b>0.660</b>

### 4.4 擬似的に生成した意味の変化する単語を用いた比較

まず、簡単な問題として、2つの時期間で意味が完全に变化する単語を擬似的に生成し [10]、その単語を用いて定量的な比較を行った。擬似的に意味が変化する単語は、各時期における余弦類似度の絶対値がいずれも 0.01 よりも小さい、なるべく無関係な単語ペアの集合の中から無作為に 50 ペアを抽出し、後の時期のコーパスの単語 (たとえば「虫」) を全て前の時期の単語 (たとえば「中隊」) に置き換えることで、2つの時期間で完全に意味が変化する単語を設定した。

日本語のデータにおいて、各手法について上位 1,000 単語までの範囲で Recall@ $k$  を評価した結果を図 3 に示す (英語でも同等の結果を確認した)。図 3 及び表 1 の平均再現率より、提案手法 PMI-SVD<sub>joint</sub>、PMI-SVD<sub>c</sub> は既存手法を上回る性能を示した。

### 4.5 実際に意味が変化した単語を用いた比較

次に、実際に意味が変化した単語を用いて定量的な比較を行った。日本語では間淵らが作成した単語リスト [11] を、英語では Kulkarni らが作成した単語リスト [1] を用いて Recall@ $k$  で評価した。日本語のデータにおける結果を図 4 に示す。付録 A に示したように、英語でも同様の結果となった。図 4 及び表 1 の平均再現率より、提案手法である PMI-SVD<sub>joint</sub> は簡単な手法でありながら既存の手法を上回り、事前学習済みの BERT に迫る性能を示した。

表 2: 日本語において意味変化した可能性が高いと予測された上位 10 単語 (1 文字の単語は除く).

順位	BERT		PMI-SVD <sub>joint</sub>	
	単語	説明	単語	説明
1	若く	匹敵する, 年齢が若い → 年齢が若い	公明	公明正大, 公正 → [組織名], 公正
2	ふれ	降る, 言及する, 抵触する → 言及する, 触る	行い	ふるまい → ふるまい, 実行
3	行い	ふるまい → ふるまい, 実行	欠け	物理的欠損 → 概念の欠損
4	公明	公明正大, 公正 → [組織名], 公正	キー	音楽, [人名] → 音楽, 物理・概念的な鍵
5	思い	思考, 動作でもたらされる感情 → 思考	覚え	記憶 → 記憶, 感じる
6	削除	文字や発言を消す → 文字や発言を消す	飛び	一足飛び, 移動 → 移動
7	在り	物理的に存在する → 概念として存在する	突如	副詞的用法 → 副詞的用法
8	参議	官職, 議事に参与する → 議員	不能	不可能 → ○○不能
9	欠け	物理的欠損 → 概念の欠損	構想	骨組みとなる考え → 骨組みとなる考え, ○○構想
10	幼稚	幼い → 幼稚園, 幼い	思惑	意図, 相場の変動を予測 → 意図

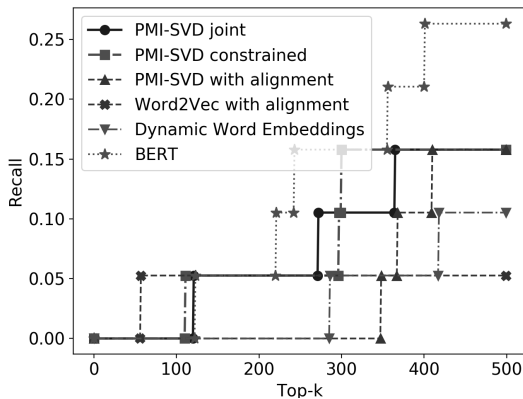


図 4: 日本語のデータで実際に意味の変化した単語を用いて Recall@k を評価した結果.

## 5 議論

ここでは、4 節で優れた結果を獲得していた BERT と提案手法である PMI-SVD<sub>joint</sub> について、2つの定性的な評価 [2] を行った。

まず、日本語について、それぞれのモデルにおいて意味変化した可能性が高いと予測された上位 10 単語を比較した (表 2)。この時、1 文字の単語は除外した。表 2 より、BERT は意味的な変化を捉え、PMI-SVD<sub>joint</sub> では固有名詞に関する変化を敏感に捉えていることがわかる。これは、BERT は与えられた文全体を考慮して単語ベクトルを計算しているのに対し、PMI-SVD<sub>joint</sub> は直近の単語の情報から単語

表 3: 「了解」、「要領」という単語と意味の近い各時期の周辺 5 単語 (1 文字の単語は除く)。

了解 (理解 → 承知)				要領 (要点 → 処理手段)			
BERT		PMI-SVD <sub>joint</sub>		BERT		PMI-SVD <sub>joint</sub>	
戦前	戦後	戦前	戦後	戦前	戦後	戦前	戦後
承諾	承諾	理解	納得	順序	順序	項目	記述
承知	承知	判断	承諾	格好	格好	一説	内容
納得	承認	推測	理解	便宜	教訓	詳細	趣旨
理解	同意	納得	同意	教訓	了解	大体	現状
断定	納得	判定	確認	消息	取扱い	引用	答弁

ベクトルを獲得しているためだと考える。

次に、実際に意味が変化した単語について、それぞれの手法で学習した単語分散表現において対象単語のベクトルに近い 5 単語 (1 文字の単語は除く) を比較した。表 3 は「了解」及び「要領」という単語についての結果である。PMI-SVD<sub>joint</sub> は意味変化に伴い周辺に「承知」に関する単語が出現し、変化前の「理解」に関する単語と共存する結果になっているが、BERT は意味が変化する前の戦前に「承知」に関する周辺単語が出現してしまっている。また、「要領」に関しても同様に、BERT は変化前の戦前から「うまく処理する手段」に関する周辺単語が出現していることがわかる。これは、BERT が時期を意識せずに訓練されており、変化後の単語の語義に強く影響されてしまったためだと考える。

## 6 おわりに

本研究では、以前の研究で提案した手法と先行研究について定量的・定性的な比較を行った。意味の変化した単語を用いて定量的な評価を行った結果、提案手法が従来の手法と同等またはそれ以上の結果を獲得することを示した。また、実際に意味の変化した単語に対して提案手法と BERT で定性的な比較を行った結果、提案手法がより効果的に単語の意味変化を捉えていることを示した。

今後は、BERT を分析対象のデータのみで訓練させ、対等な条件で性能の比較を行う予定である。また、比較する時期を 3 つ以上に増やし、単語の意味変化についてさらに詳細な比較・分析を行いたい。

**謝辞** 本研究は国立国語研究所の共同研究プロジェクト「現代語の意味の変化に対する計量的・統計力学的アプローチ」、同「通時コーパスの設計と日本語史研究の新展開」および JSPS 科研費 19H00531, 18K11456 の研究成果の一部を報告したものである。

## 参考文献

- [1] Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. Statistically significant detection of linguistic change. In *Proceedings of the 24th International Conference on World Wide Web, WWW'15*, p. 625–635, Republic and Canton of Geneva, CHE, 2015. International World Wide Web Conferences Steering Committee.
- [2] 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 (Volume 1: Long Papers)*, pp. 1489–1501, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [3] Zijun Yao, Yifan Sun, Weicong Ding, Nikhil Rao, and Hui Xiong. Dynamic word embeddings for evolving semantic discovery. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM'18*, p. 673–681, New York, NY, USA, 2018. Association for Computing Machinery.
- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [5] Renfen Hu, Shen Li, and Shichen Liang. Diachronic sense modeling with deep contextualized word embeddings: An ecological view. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 3899–3908, Florence, Italy, July 2019. Association for Computational Linguistics.
- [6] Matej Martinc, Petra Kralj Novak, and Senja Pollak. Leveraging contextual embeddings for detecting diachronic semantic shift. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 4811–4819, Marseille, France, May 2020. European Language Resources Association.
- [7] Mario Giulianelli, Marco Del Tredici, and Raquel Fernández. Analysing lexical semantic change with contextualised word representations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 3960–3973, Online, July 2020. Association for Computational Linguistics.
- [8] Omer Levy and Yoav Goldberg. Neural word embedding as implicit matrix factorization. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pp. 2177–2185. Curran Associates, Inc., 2014.
- [9] 相田太一, 小町守, 小木曾智信, 高村大也, 坂田綾香, 小山慎介, 持橋大地. 単語分散表現の結合学習による単語の意味の通時的変化の分析. 言語処理学会第 26 回年次大会 発表論文集, 2020.
- [10] Philippa Shoemark, Farhana Ferdousi Liza, Dong Nguyen, Scott Hale, and Barbara McGillivray. Room to Glo: A systematic comparison of semantic change detection approaches with word embeddings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 66–76, Hong Kong, China, November 2019. Association for Computational Linguistics.
- [11] 間淵洋子, 小木曾智信. 近現代日本語の意味変化分析のための単語データセット構築の試み. 言語処理学会第 27 回年次大会 発表論文集, 2021.

## A 英語の実験結果

まず、英語データにおける結果を以下に示す。

表 4: 英語において意味変化した可能性が高いと予測された上位 10 単語。

順位	BERT		PMI-SVD <sub>joint</sub>	
	単語	説明	単語	説明
1	leagues	長さの単位 → メジャーリーグ	tile	覆う, 素材 → 正方形の板
2	coach	馬車, 指導者 → 指導者	major	主要な → メジャーリーグ
3	chapter	本や時間の分割単位 → 本の分割単位	facility	能力 → 場所や装置
4	affirmative	肯定的 → affirmative action	moreover	副詞的用法 → 副詞的用法
5	web	クモの糸 → インターネット	frank	正直, [人名] → [人名]
6	kaiser	肯定 → パン, [会社名]	ad	[前置詞的用法] (ラテン語) → 広告
7	bush	茂み → [人名]	peruvian	ペルー人 → ペルー国の
8	bangs	[人名] → 前髪	shrinking	恐怖で萎縮する → 経済縮小
9	n	i dunno → i dunno, rock n roll	pro	[前置詞的用法] (ラテン語) → 専門家
10	facility	能力 → 場所や装置	coach	馬車, 指導者 → 指導者

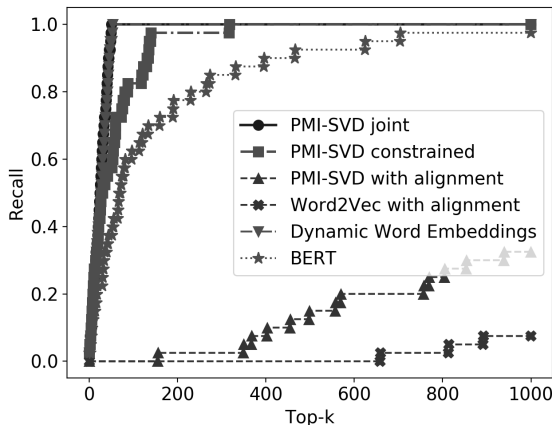


図 5: 英語のデータで擬似的に意味の変化する単語を生成し、Recall@k で評価した結果。

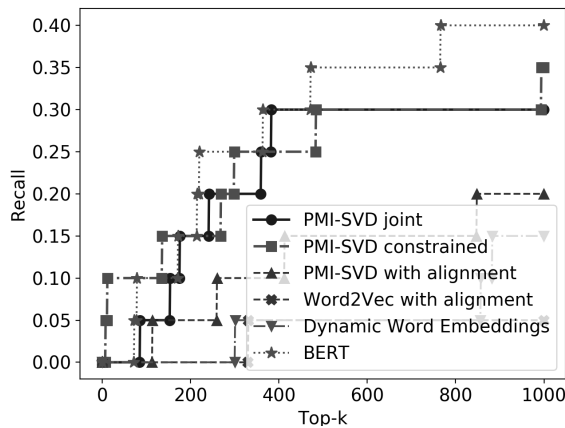


図 6: 英語のデータで実際に意味の変化した単語を用いて各手法に対して Recall@k で評価した結果。

表 5: 英語のデータにおける平均再現率。

モデル	擬似	実際
PMI-SVD <sub>joint</sub>	<b>0.997</b>	0.654
PMI-SVD <sub>c</sub>	0.993	0.652
PMI-SVD <sub>align</sub> [2]	0.714	0.514
Word2Vec <sub>align</sub> [2]	0.482	0.424
Dynamic Word Embeddings [3]	<b>0.997</b>	<b>0.509</b>
BERT [6]	0.981	<b>0.711</b>

表 6: gay という単語について、各時期のベクトルの周辺 10 単語。

gay (陽気な → 同性愛の)			
BERT		PMI-SVD <sub>joint</sub>	
1900s	1990s	1900s	1990s
<i>queer</i>	<i>queer</i>	<i>merry</i>	<i>sex</i>
<i>masculine</i>	<i>sexual</i>	<i>bright</i>	<i>mothers</i>
<i>sex</i>	<i>sex</i>	<i>happy</i>	<i>gang</i>
<i>male</i>	<i>male</i>	<i>sang</i>	<i>women</i>
<i>youthful</i>	<i>masculine</i>	<i>singing</i>	<i>young</i>
<i>handsome</i>	<i>unmarried</i>	<i>beautiful</i>	<i>blacks</i>
<i>romantic</i>	<i>discrimination</i>	<i>cheerful</i>	<i>adults</i>
<i>intimate</i>	<i>wed</i>	<i>lovely</i>	<i>group</i>
<i>naked</i>	<i>males</i>	<i>jolly</i>	<i>unhappy</i>
<i>wed</i>	<i>men</i>	<i>dance</i>	<i>disabled</i>

表 7: rock という単語について、各時期のベクトルの周辺 10 単語。

rock (鉱物 → 音楽)			
BERT		PMI-SVD <sub>joint</sub>	
1900s	1990s	1900s	1990s
<i>rocks</i>	<i>rocks</i>	<i>rocks</i>	<i>rocks</i>
<i>stone</i>	<i>stone</i>	<i>limestone</i>	<i>concerts</i>
<i>granite</i>	<i>rocky</i>	<i>crust</i>	<i>concert</i>
<i>cliff</i>	<i>cliff</i>	<i>layers</i>	<i>sand</i>
<i>rocky</i>	<i>granite</i>	<i>volcanic</i>	<i>singer</i>
<i>stones</i>	<i>stones</i>	<i>granite</i>	<i>symphony</i>
<i>limestone</i>	<i>limestone</i>	<i>cliff</i>	<i>stones</i>
<i>sand</i>	<i>sand</i>	<i>cliffs</i>	<i>volcanic</i>
<i>mountain</i>	<i>concrete</i>	<i>lava</i>	<i>comedy</i>
<i>cave</i>	<i>slide</i>	<i>sand</i>	<i>lava</i>

## B 日本語の評価に用いた単語リスト

次に、今回 4.5 節で用いた日本語で実際に意味の変化した単語リストを以下に示す。

適当, 故障, 優勝, 非常, 心持ち, 広告, 了解, 結構, 住居, 主婦, 要領, 全然, 風俗, 障害, 婦人, 貴族, 普通, 設備, 教授, 女性, 情報, 普段, 自然, モデル, とても, 衣