

# BERT による単語埋め込み表現の分散値を用いた 語義の広がり分析

欧陽恵子 (茨城大学大学院理工学研究科情報工学専攻) \*  
曹鋭 (茨城大学大学院理工学研究科情報工学専攻) †  
白静 (茨城大学大学院理工学研究科情報工学専攻) ‡  
馬ブン (茨城大学大学院理工学研究科情報工学専攻) §  
新納浩幸 (茨城大学大学院理工学研究科情報工学専攻) ¶

## Analysis of Polysemy using Variance Values for Word Embeddings by BERT

Yanghuizi Ou (Graduate School of Science and Engineering, Ibaraki University)

Rui Cao (Graduate School of Science and Engineering, Ibaraki University)

Jing Bai (Graduate School of Science and Engineering, Ibaraki University)

Wen Ma (Graduate School of Science and Engineering, Ibaraki University)

Hiroyuki Shinnou (Graduate School of Science and Engineering, Ibaraki University)

### 要旨

BERT が出力する単語の埋め込み表現は、その単語が現れた文の文脈に依存している。このためある単語  $w$  の用例を収集し、BERT により得られる単語  $w$  に対する埋め込み表現から、それらの分散値を計算すると、その値は単語  $w$  の語義の広がりに対応すると考えられる。そこで多義語「頭」「意味」「核」「記録」「言葉」「胸」と単義語「生産」「政治」「意識」「抗議」「成績」を対象にそれら単語の用例を収集し、分散値の比較を行った。多義語に対しては大きな分散値、単義に対しては小さな分散値が出るのが予想される。また BERT のどの位置の階層が最も語義の広がりにも影響しているかも調査した。しかしこれら実験から、上記の分散値では語義の広がりを測定するのは困難であることがわかった。その原因について考察する。

### 1. はじめに

本論文では BERT (Devlin et al. (2019)) が出力する単語の埋め込み表現と語義との関係を調べるために、多義語と単義語に分けて、それら埋め込み表現の分散値、つまり語義の広がり进行比较する。BERT が出力する単語の埋め込み表現が語義を表現しているなら、多義語に対する分散値は大きく、単義語に対する分散値は小さくなるのが予想できる。

BERT は言語の事前学習モデルであり、基本的には入力された単語列を対応する単語の埋

---

\* 19NM705X@vc.ibaraki.ac.jp

† 18ND305G@vc.ibaraki.ac.jp

‡ 19ND301R@vc.ibaraki.ac.jp

§ 19ND302H@vc.ibaraki.ac.jp

¶ hiroyuki.shinnou.0828@vc.ibaraki.ac.jp

め込み表現列に変換する。このとき得られる単語の埋め込み表現は word2vec (Mikolov et al. (2013)) などから得られる分散表現のように固定したベクトルではなく、その単語が現れた文脈、つまり入力された単語列に依存している。この点からある単語  $w$  を含む文  $s$  を BERT に入力し、 $w$  に対応する埋め込み表現  $e_w$  を得たとき、 $e_w$  は  $s$  内における  $w$  の語義を表していると考えられる。本論文では  $e_w$  を収集し、その分散値  $V_w$  を得ることで  $w$  の語義の広がり（多様性）を調べることで、BERT が出力する単語の埋め込み表現と語義との関係を考察する。特に注目するのは単義語  $w$  に対する  $e_w$  の分散値  $V_w$  である。 $e_w$  が語義を表しているのであれば、 $e_w$  の分散値  $V_w$  は非常に小さいはずである。これを多義語  $w$  に対する  $e_w$  の分散値  $V_w$  との比較から確認する。

また BERT は概略 Multi-head Attention の層を 12 層重ねたモデルであり、各層毎に単語  $w$  に対する埋め込み表現  $e_w$  が得られる。今第  $i$  層の単語  $w$  に対する埋め込み表現を  $e_w^{(i)}$  とおく。つまり  $e_w = e_w^{(12)}$  である。 $e_w$  がある程度の広がり（分散値）を持っていたとしても、 $e_w^{(1)}$  は分散表現に近い形なので、その広がり（分散値）は小さく、層を経るに従い徐々に広がりが増していくと考えられる。本研究では各階層ごとに分散値  $V_w^{(i)}$  を求め、どの階層で分散値が最も大きく増加するかを確認する。これによってどの階層が語義を特定するのに寄与しているかが考察できる。

実験では多義語として「頭」「意味」「核」「記録」「言葉」「胸」の 6 単語、単義語「生産」「政治」「意識」「抗議」「成績」の 5 単語を対象にして分散値  $V_w^{(i)}$  を求めることで、BERT の出力する単語埋め込み表現の語義の広がり（分散値）の分析した。

しかし実験から、上記した分散値では語義の広がり（分散値）を測定するのは困難であることがわかった。その原因についても考察する。

## 2. 関連研究

BERT の出力する埋め込み表現と語義との関連性に関する研究としては、BERT を利用した語義曖昧性解消 (Word Sense Disambiguation; WSD) の研究がある。論文 (曹鋭ほか (2019)) では BERT の出力する埋め込み表現を特徴ベクトルとして利用して、教師あり学習により WSD を行っている。また all-words WSD を行った論文 (Du et al. (2019)) や 3 つの BERT ベースの WSD モデルを提案した (Huang et al. (2019)) などの研究もある。更に論文 (谷田部梨恵・佐々木稔 (2020b)) では WSD の解決のために、用例文内の単語が同じ語義で使われているかどうかを判定するために、BERT の出力する埋め込み表現を利用している。

BERT の出力する埋め込み表現をクラスタリングする研究は、語義の広がりに関する研究と関連している。論文 (山内崇史ほか (2020)) では、BERT の出力する埋め込み表現をクラスタリングすることにより領域表現を獲得している。この領域表現が本研究における語義のクラスタに対応している。この領域表現を単語間の意味的類似度推定タスクおよび単語間の関係推定タスクに利用して、それらの精度向上を果たしている。

もっと一般的に BERT の出力する埋め込み表現と言語現象との関連付けを研究するために、論文 (浅原正幸・加藤祥 (2020a)) では、『現代日本語書き言葉均衡コーパス』(BCCWJ) に対して BERT の出力する埋め込み表現を付与したコーパス BERTed-BCCWJ を構築している。

### 3. 単語埋め込み表現の分散値と語義の広がり

単語  $w$  を含む文  $s$  を  $n$  個集め、それらを  $s_1, s_2, \dots, s_n$  とする。これらの文を BERT に入力する。  $s_i$  中の  $w$  に対応する BERT から得られる埋め込み表現を  $e_{w_i}$  とする。  $e_{w_i}$  の平均ベクトルを  $\bar{e}_w$  とし、  $e_w$  の分散値  $V_w$  を以下で定義する。

$$V_w = \frac{1}{n} \sum_{i=1}^n \|\bar{e}_w - e_{w_i}\|^2$$

単語  $w$  の語義の広がりを  $V_w$  によって測ることにする。

また BERT の第  $i$  層目の出力内の単語  $w$  に対する埋め込み表現を  $e_w^{(i)}$  とおき、その分散値を  $V_w^{(i)}$  とおく。 BERT は 12 層からなるので、  $V_w^{(1)}$  から  $V_w^{(12)} = V_w$  が得られる。これらの値を確認することで、語義が明確になる階層位置を考察する。

### 4. 実験

#### 4.1 単義語と多義語との語義の広がりの差

対象とした単語は、多義語として「頭」「意味」「核」「記録」「言葉」「胸」の 6 単語、単義語として「生産」「政治」「意識」「抗議」「成績」の 5 単語である。各単語の用例は BCCWJ 及び毎日新聞の '93 から '98 年の記事からランダムに取り出した。取り出した用例の数と得られた分散値について多義語は表 1、単義語は表 2 に示す。

多義語 6 単語に対する分散値の平均は 186.14、単義語 5 単語に対する分散値の平均は 176.43 であり、単義語の分散値の方が多義語の分散値よりも小さくなっている。ただし統計的な有意差はなく、この実験結果からは予想していた結果は得られなかった。

表 1 多義語の用例数と実験結果

単語	頭	意味	核	記録	言葉	胸
用例数	82	91	191	151	184	74
分散値	197.07	241.28	112.96	213.29	179.72	172.50

表 2 単義語の用例数と実験結果

単語	生産	政治	意識	抗議	成績
用例数	110	434	82	42	33
分散値	163.81	162.73	204.19	162.00	189.38

#### 4.2 BERT の各階層における語義の広がりの変化

BERT の各階層における語義の広がりの変化を調べた。多義語の結果を図 1 に、単義語の結果を図 2 に示す。

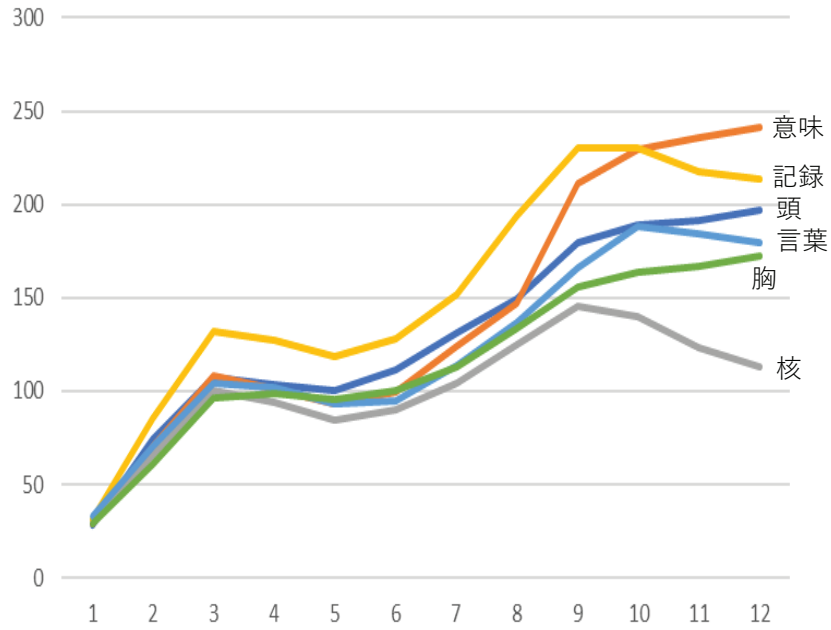


図1 BERTの各階層における分散値（多義語）

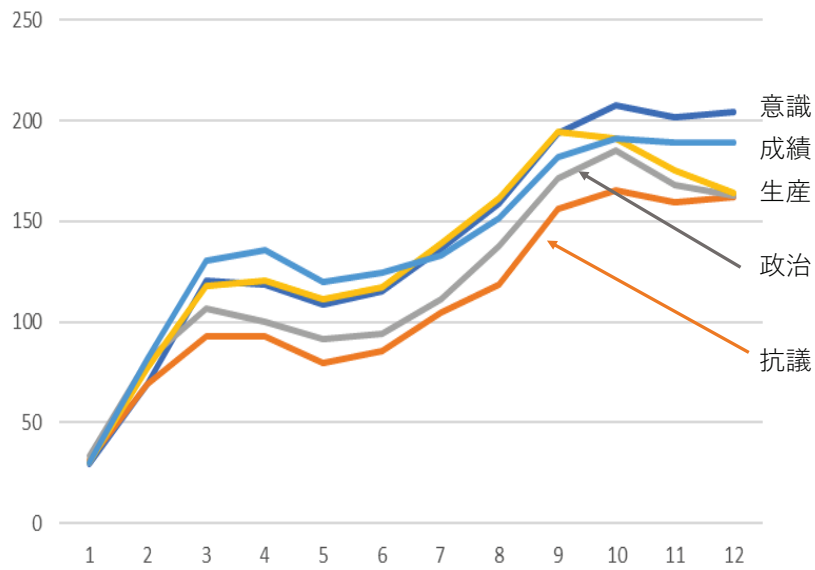


図2 BERTの各階層における分散値（単義語）

こちらの場合も多義語と単義語に大きな差は見る事ができなかった。また語義の広がりを示す分散値は階層が上がるに従って徐々に大きくなってゆくと考えられる。しかしどちらの場合も単調に分散値が上昇するという訳ではなく、第4層目と第9層目辺りに勾配が下降したり平坦になるような様子が見られる。

## 5. 考察

実験では多義語に対する分散値と単義語に対する分散値に大きな違いはなかった。これは語義の広がりをも単に分散値から計算したことによるものだと考えられる。当初、語義に対する埋め込み表現は図3のような位置関係になると予想していた。図4の場合は、確かに多義語に対する分散値は大きく、単義語に対する分散値は小さくなる。

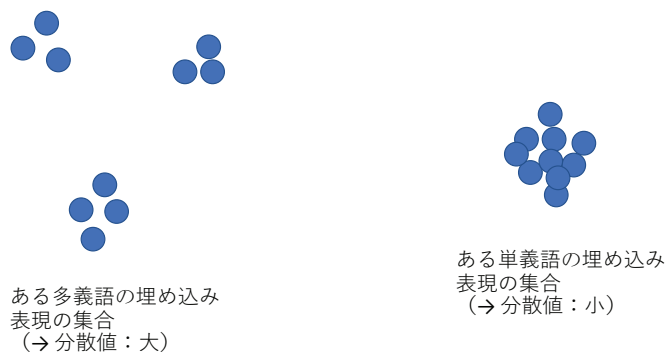
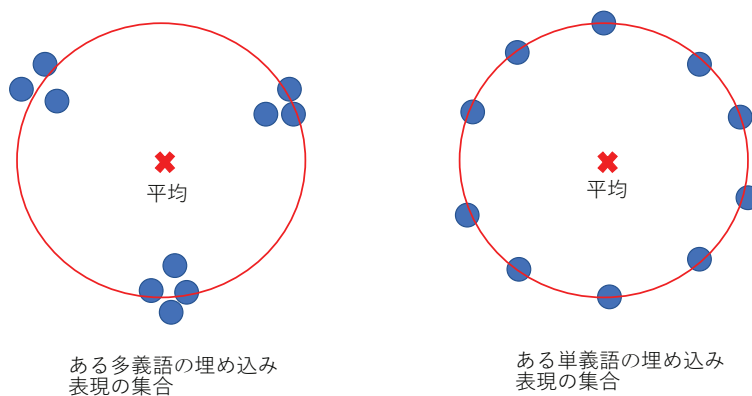


図3 予想していた多義語と単義語の埋め込み表現の位置関係

しかし実際は例えば図4のような位置関係になっていたと考えられる。図のような場合は、多義語に対する分散値と単義語に対する分散値に大きな違いは生じない。



分散値は同等

図4 現実の多義語と単義語の埋め込み表現の位置関係例

語義の広がりを確認するために、BERT から出力される対象単語  $w$  に対する埋め込み表現  $e_w$  の集合内の全ペアに対してその距離を求め、その平均の距離を測定した。結果を表3と表4に示すが、埋め込み表現間の平均距離に関しても多義語と単義語間では大きな差は生じていない。

本論文では語義の広がりをも埋め込み表現の位置関係で測ることができると考えていたが、結

表 3 多義語の埋め込み表現間の平均距離

単語	頭	意味	核	記録	言葉	胸
分散値	197.07	241.28	112.96	213.29	179.72	172.50
平均距離	19.97	22.09	15.08	20.72	19.01	18.70

表 4 単義語の埋め込み表現間の平均距離

単語	生産	政治	意識	抗議	成績
分散値	163.81	162.73	204.19	162.00	189.38
平均距離	18.21	18.06	20.33	18.21	19.75

果的にはそれができていない。原因として以下の3つが考えられる。

#### (1) 語義のクラスタが小さいとは限らない

単語  $w$  の用例を集めて、BERT から出力される  $w$  の埋め込み表現  $e_w$  の集合を作り、そこからクラスタリングすれば語義のクラスタが作成される。当初、この語義のクラスタが小さいことを想定していた。単義語  $w$  に対しては  $e_w$  の集合が語義のクラスタ自体を表すことになるが、前章の実験はその大きさが特に小さくはないことを示している。

#### (2) 語義のクラスタ間距離が大きいとは限らない

語義のクラスタがある程度の大きさを持っていたとしても、語義のクラスタ間距離が大きければ、多義語  $w$  に対する  $e_w$  の集合の分散値は大きくなるはずである。当初、この語義のクラスタ間距離が大きいと予想していた。しかし前章の実験ではそのような結果は示されなかった。

本来、語義のクラスタ間距離は語義の距離に対応しているので、それらの値も多様性があり、一様に大きいとは限らない。

#### (3) コーパスから得た用例が多義語になっていない可能性もある

本実験で用いた多義語に対して実際に語義が異なる用例を収集できているかを確認できていない。

上記の3点の他に、そもそもBERTが出力する埋め込み表現自体が語義を表現しているのかどうかも確認の必要がある。BERTは、本来、設定したタスクを利用してFine-Tuningを行って利用するものであり、タスクに応じてBERTの重みが調整され、その出力がタスクに適したものになる。このため素のBERTが出力した埋め込み表現が直接語義を表現できているのかどうかは不明である。

今後は上記の点に注意し、BERTの出力する埋め込み表現と語義との関係を調べていきたい。

## 6. おわりに

本論文ではBERTが出力する単語の埋め込み表現と語義との関係を調べるために、単義語と多義語に分けて、それら埋め込み表現のなす広がり測ることを行った。

具体的には、単語  $w$  の用例を収集し、BERT により得られる単語  $w$  に対する埋め込み表現から、それらの分散値を単語  $w$  の語義の広がりとして定義した。多義語「頭」「意味」「核」「記録」「言葉」「胸」と単義語「生産」「政治」「意識」「抗議」「成績」を対象にそれら単語の用例を収集し、分散値を測った。BERT が出力する単語の埋め込み表現が語義を表現しているのなら、多義語に対する分散値は大きく、単義語に対する分散値は小さくなることが期待されたが、両者の分散値に大きな差は生じなかった。予想に反した結果に対する原因を考察した。今後はその考察をもとに、BERT の出力する埋め込み表現と語義との関係を調べていきたい。

## 文 献

- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” *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.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean (2013). “Distributed Representations of Words and Phrases and their Compositionality.” C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 26*. pp. 3111–3119.
- 曹鋭・田中裕隆・白静・馬ブン・新納浩幸 (2019). 「BERT を利用した教師あり学習による語義曖昧性解消」 言語資源活用ワークショップ 2019 発表論文集.
- Jiaju Du, Fanchao Qi, and Maosong Sun (2019). “Using bert for word sense disambiguation.” *arXiv preprint arXiv:1909.08358*.
- Luyao Huang, Chi Sun, Xipeng Qiu, and Xuanjing Huang (2019). “GlossBERT: BERT for Word Sense Disambiguation with Gloss Knowledge.” *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. 3509–3514.
- 谷田部梨恵・佐々木稔 (2020b). 「BERT の学習済みモデルを用いた用例文ペアの同義判定」 言語処理学会第 26 回年次大会, P4-10.
- 山内崇史・梶原智之・荒瀬由紀 (2020). 「文脈を考慮した単語ベクトル集合からの単語領域表現」 言語処理学会第 26 回年次大会, B4-2.
- 浅原正幸・加藤祥 (2020a). 「BERTed-BCCWJ: 多層文脈化単語埋め込み情報を付与した『現代日本語書き言葉均衡コーパス』データ」 言語処理学会第 26 回年次大会, P2-5.