

BERT の Masked Language Model を用いた 二文間の接続関係の推定

趙一 (茨城大学大学院理工学研究科情報工学専攻) *

曹銳 (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) ¶

Estimation of Connection between Two Sentences by Using the Masked Language Model of BERT

Yi Zhao (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 の Masked Language Model を利用すると、文中のマスクされた単語を推定することができる。また BERT は2つの文 s_1 と s_2 を入力できる。そこで s_2 の直前にマスク単語を置き、”[CLS] s_1 [SEP] [MASK] s_2 [SEP]” の形の文を BERT に入力し、BERT から推定される [MASK] の単語から文 s_1 と文 s_2 の接続関係を推定することを試みる。具体的にいくつかの接続詞を準備し、[MASK] がそれら接続詞となる確率を求めることで二文間の接続関係を推定する。実験では [MASK] の部分が実際に対象の接続詞であった文とその前文を利用して、マスクされた接続詞を推定できるかどうかの実験を行った。また教師あり学習で本タスクを解いた場合との比較も行った。

1. はじめに

二文間の論理的関係を推定する処理は対話、要約、文生成など様々なタスクで必要とされる処理であり、自然言語処理の重要な問題となっている。例えば対話ではユーザの発話間の関係を認識することで、論理的な繋がりのある文を返答することができる。

本論文では BERT (Devlin et al. (2019)) の Masked Language Model を利用して、与えられた二文間の関係を推定することを試みる。

* 20NM718L@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

BERT の Masked Language Model を利用すると、文中のマスクされた単語を推定することができる。また BERT は 2 つの文 s1 と s2 を入力することができる。そこで s2 の直前にマスク単語を置き、”[CLS] s1 [SEP] [MASK] s2 [SEP]” の形の文を BERT に入力し、BERT から推定される [MASK] の単語から文 s1 と文 s2 の接続関係を推定する。

具体的にはいくつかの接続詞を準備し、[MASK] がそれら接続詞となる確率を求めることで二文間の接続関係を推定する。実験では [MASK] の部分が実際に対象の接続詞であった文とその前文を利用して、マスクされた接続詞を推定できるかどうかの実験を行った。また教師あり学習で本タスクを解いた場合との比較も行った。

2. 関連研究

BERT の Masked Language Model を利用した研究としてはスペルチェックがある。BERT の Masked Language Model を利用すれば特定箇所に入る文字や単語を推定できるので、原文と推定した文字や単語とを比較することでスペルチェックを行うことができる。論文 (Zhang et al. (2020)) では誤り検出用ネットワークと誤り訂正用ネットワークからなるスペルチェックのシステムを構築した。その誤り検出用ネットワークに BERT の Masked Language Model を利用したソフトマスキング手法を用いている。

二文間の接続関係を推定する研究は従来より多くの研究 (山本和英・齋藤真実 (2008), Lin et al. (2009), Lan et al. (2013) など) があるが、ディープラーニングの手法を利用したものとしては論文 (大塚淳史ほか (2015)) がある。ここでは再帰的なニューラルネットワークを用いて文節や文の概念ベクトルを作成する手法である Recursive AutoEncoder (Socher et al. (2011)) を用いた概念ベクトルによる文間の接続関係推定手法を提案している。

3. BERT による接続関係の推定

2 つの文 s1 と s2 の接続関係を推定するには s1 と s2 に最も適切に当てはまる接続詞を推定すればよい。この接続詞の推定に BERT の Masked Language Model を利用する。具体的には s2 の直前にマスク単語を置き、以下の形の token 列を BERT に入力し、BERT から推定される [MASK] の単語から文 s1 と文 s2 の接続関係を推定する。

[CLS] (s1 の token 列) [SEP] [MASK] (s2 の token 列) [SEP]

BERT の Masked Language Model を利用すると、語彙リストにある全ての単語について [MASK] の位置にその単語が現れる確率が得られる。本論文では表 1 に示した 21 種類の接続詞を対象にして、[MASK] の位置にその接続詞が現れる確率を求め、そのなかから最も高い確率を持つ接続詞を二文間の接続関係と見なす。これら 21 種類の接続詞は毎日新聞 5 年分の記事を調べ、基本的に頻度の高いものから選択している。

しかし, また, だが, ただ, 一方, でも, そして, それでも, ところが, だから, 例えば, そこで, ただし, しかも, つまり, たとえば, まずは, が, それで, 実は, なお

表 1 対象とした接続詞

図 1 に本手法の動作例を示す. 図では以下の二つの文 s1 「私は犬が好き.」と s2 「猫は嫌い.」の接続関係を推定している. s1 と s2 は token id 列に変換され, [CLS] (token id は 2), [SEP](token id は 3) 及び [MASK] (token id は 4) の token id が付与されて BERT に入力される. BERT からは各 token の位置に BERT の語彙リストにある 32,000 単語の各単語が現れる確率が得られる. 対象単語とする 21 個の接続詞に対して, その接続詞が [MASK] 位置に入る確率を調べ, 最も高い確率の接続詞を推定結果とする.

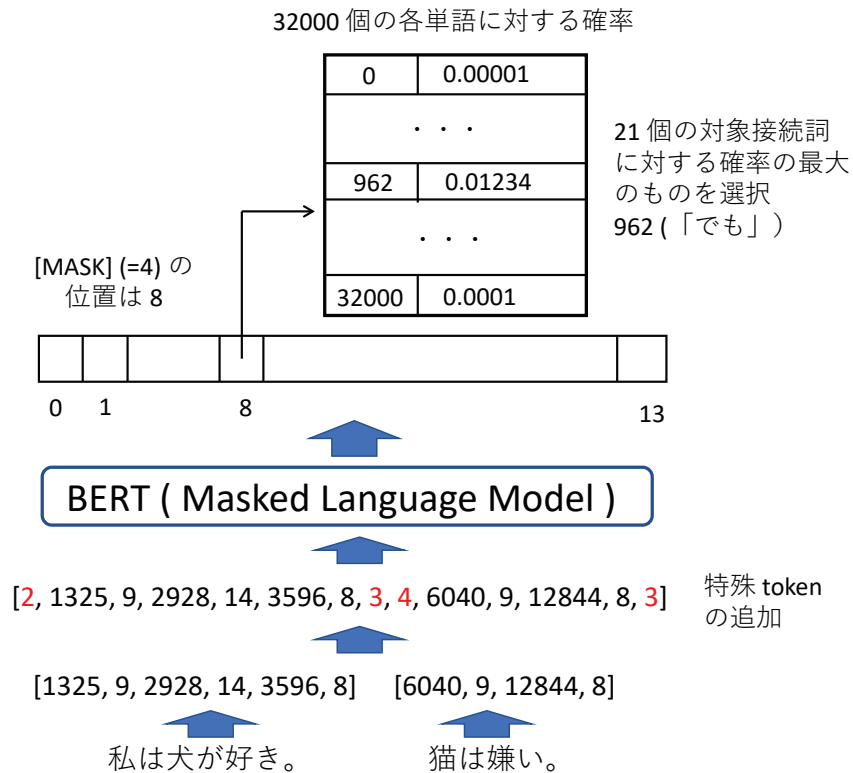


図 1 手法の実行例

4. 実験

表 1 の各接続詞が文頭に現れる文 (s2) をコーパスから検索し, その直前の文 (s1) と組にしてランダムに 100 組取り出した. ここでコーパスとしては 1993 年から 1998 年までの毎日新聞記事を用いた. つまり 21 種類の接続詞に関して各 100 個のテストデータが準備したことになる.

実験で用いた日本語 BERT モデルとしては、以下のサイトで公開されているものを利用した。

<https://github.com/cl-tohoku/bert-japanese>

上記のサイトには 4 つのモデルが登録されているが、その中の BERT-base_mecab-ipadic-bpe-32k を利用した。

実験の結果を表 2 に示す。1 列目が対象となる接続詞であり、2 列目が 100 個のテストデータ中正しくその接続詞を推定できた数である。3 列目は誤りのパターンの上位 3 個を示したものである。

対象単語	正解数	誤りのパターン
しかし	71	また (15), そして (4), つまり (2)
また	85	しかし (5), そして (4)
だが	0	しかし (65), また (20), そして (4)
ただ	4	しかし (50), また (27), ただし (10)
一方	28	また (49), しかし (16), なお (3)
でも	15	しかし (42), だから (16), そして (14)
そして	51	また (22), しかし (17), だから (5)
それでも	17	しかし (39), また (22), そして (11)
ところが	0	しかし (69), また (17), そこで (4)
だから	18	そして (28), しかし (17), また (13)
例えば	31	また (37), しかし (9), そして (6)
そこで	43	また (17), そして (17), しかし (14)
ただし	15	しかし (35), また (33), なお (6)
しかも	2	また (44), しかし (35), そして (7)
つまり	16	しかし (23), そして (20), また (17)
たとえば	9	また (38), 例えば (17), そして (13)
まずは	22	そして (27), また (16), そこで (10)
が	0	しかし (64), そして (10), また (9)
それで	10	そして (26), だから (14), しかし (13)
実は	5	また (32), しかし (24), だから (16)
なお	27	また (41), しかし (14), ただし (12)
合計	469	

表 2 実験結果

単純に総和を取って正解率を測ると $0.233(= 469/2100)$ という値であった。ランダムに推定すると正解率は $0.048(= 1/21)$ なので、BERT を利用した効果は確認できる。

5. 考察

5.1 接続詞のグループ化

本実験で対象とした接続詞の中には意味が同じであるものがいくつか含まれている。例えば、「例えば」と「たとえば」は表記が違っただけで同じ接続詞である。このように同じ意味の接続詞はグループ化して、そのグループの接続詞を推定できれば正解と見なす方が、応用上は有用だと思われる。

ここでは筆者らの主観に基づいて、表 3 のように接続詞をグループ化して、実験の正解数を計り直した。この結果を表 3 に示す。

グループ番号	対象単語グループ	正解数
1	例えば, たとえば	62
2	しかし, だが, ただ, でも, ところが, ただし, が, それでも	532
3	つまり, だから	53
4	また, 一方, なお	235
5	そして, それで, そこで	156
6	しかも	2
7	まずは	22
8	実は	5
合計		1,067

表 3 グループ化した接続詞に対する正解率

正解率を測ると $0.508(= 1067/2100)$ である。最大頻度による正解率は $0.381(= 800/2100)$ であり、教師なしによる識別の結果であることを考えれば、高い正解率であると考えられる。

5.2 対象外の接続詞

本稿では手法の可能性を試すために 21 種類の接続詞だけを対象としたが、対象としなかった接続詞も多くある。以下にその一覧を示す。

そのうえで、それだけに、だからこそ、そもそも、すると、ちなみに、いや、けれど、それから、それなのに、次いで、だけど、さて、それにしても、こうして、なのに、あ
るいは、それでは、なぜなら、けれども、それとも、従って、ましてや、それなら、だからといって、それに、それどころか、というのも、そのうえ、次に、だって、すな
わち、したがって、ところで、それでいて、そうして、じゃあ、そうしたら、よって、ゆ
えに、ともあれ、そしたら、そうすると、だとすれば、ともすれば、そうなると、または、かたや、ならば、では、でなければ、でないと、さもなければ、けど、かくして、それと、されど、かつ、惜しむらくは、だとすると、じゃ、しかしながら、否、もしくは、ですから、つまるところ、そのうえに、さりとて、尚、故に、なので、で、だけれども、そういや、お
よび、おなじく、おそれながら、いっぽう、いえ

下線の引かれている接続詞は利用した BERT の語彙集合では一単語と認識できないものである。BERT で一単語として認識できない単語については、マスク単語として推定することはできない。接続詞に対応するような接続表現（例えば「その結果として」など）についても同様である。

ただし本論文の目的は2文間の関係を認識することであり、接続詞あるいは接続表現を推定することではないので、本質的な欠点ではないと考える。

5.3 教師あり学習による推定

本論文では2文間の接続詞を推定するために BERT を利用している。BERT は既存のモデルを利用しているため、結果的に本手法では学習を必要としていない。ただし本タスクに対しては、ラベル付きデータを自動で構築できるために、教師あり学習の枠組みで解くことも可能である。ここでは BERT の feature based と fine tuning の方法それぞれにより本タスクを解いた場合の正解率を測る。

図2は利用したネットワーク図である。feature based の手法の学習では図2の W の層のパラメータだけを学習する形であり、fine tuning の手法の学習では図2の W の層のパラメータの他に、BERT 自体のパラメータも学習する形である。どちらの手法でも先頭にある特殊トークンの [CLS] のベクトルから1層（図の W）の Classification 層を経て表3に示された8種類の接続関係を識別する。

先の実験で利用した 2,100 個の文ペアの半分 1,050 個を訓練データとして利用し、残り半分をテストデータとして利用した。学習は 100 エポックまで行い、各エポックの学習終了時にその時点のモデルを保存し、そのモデルによりテストデータに対する正解率を測った。結果を図3に示す。

feature based の手法での正解率は約 0.335、fine tuning の手法での正解率は約 0.501 となった。どちらの手法でも本手法の正解率 0.508 を超えることはできなかった。

ただしこれは訓練データの量の問題であり、訓練データの量を増やすことで本手法以上の正解率は出せると考えている。この確認は今後の課題である。

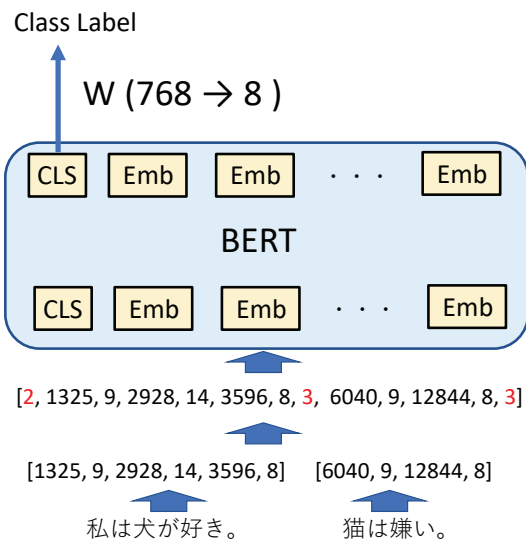


図2 BERT の識別による推定

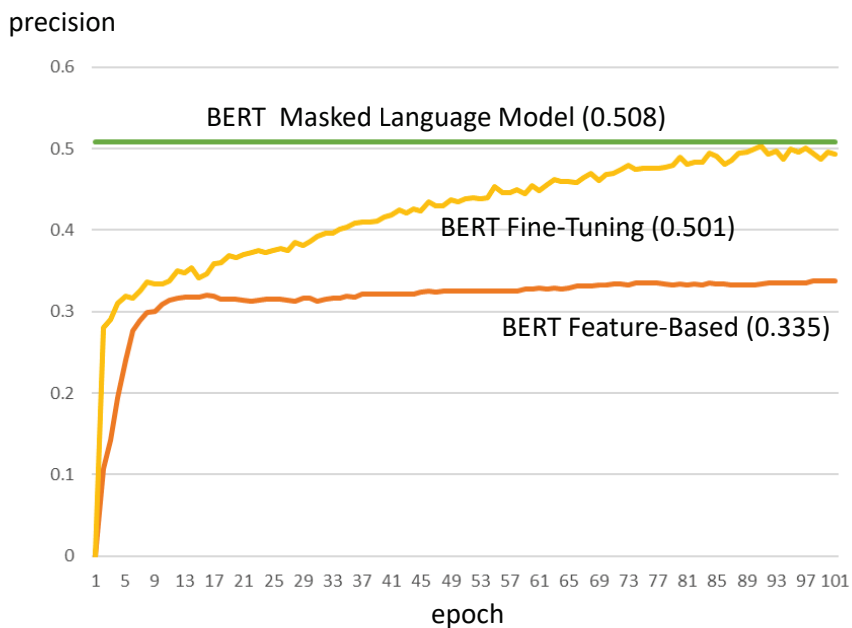


図3 教師あり学習による推定との比較

6. おわりに

本論文では BERT の Masked Language Model を利用して二文間の接続関係を推定することを試みた。1 文目と 2 文目の間に [MASK] トークンを置き、BERT の Masked Language Model を利用してマスクされた単語が対象とした接続詞になる確率を求めることでその推定を行う。実験では 21 種類の接続詞を対象にして、8 カテゴリの接続関係を推定した。接続詞

の推定の正解率は 0.223 であった。8 カテゴリの接続関係の推定の正解率は 0.508 であり、BERT による教師あり学習よりも良い結果であった。[MASK] トークンが各接続詞となる確率を求めることで、二文間の接続関係を推定するというアプローチを用いる場合、訓練データを増やすことは容易であるため、訓練データを増やすことで正解率を上げてゆくことは可能である。今後、本アプローチでどの程度まで推定を正しくできるのかを確かめたい。

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.
- Shaohua Zhang, Haoran Huang, Jicong Liu, and Hang Li (2020). “Spelling Error Correction with Soft-Masked BERT.” *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 882–890.
- 山本和英・齋藤真実 (2008). 「用例利用型による文間接続関係の同定」 自然言語処理, 15:3, pp. 21–51.
- Ziheng Lin, Min-Yen Kan, and Hwee Tou Ng (2009). “Recognizing implicit discourse relations in the Penn Discourse Treebank.” *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 343–351.
- Man Lan, Yu Xu, and Zheng-Yu Niu (2013). “Leveraging synthetic discourse data via multi-task learning for implicit discourse relation recognition.” *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 476–485.
- 大塚淳史・平野徹・宮崎千明・東中竜一郎・牧野俊朗・松尾義博 (2015). 「Recursive AutoEncoder を用いた文間の接続関係推定」 人工知能学会全国大会論文集, JSAI2015, pp. 4K15–4K15.
- Richard Socher, Eric H Huang, Jeffrey Pennin, Christopher D Manning, and Andrew Y Ng (2011). “Dynamic pooling and unfolding recursive autoencoders for paraphrase detection.” *Advances in neural information processing systems*, pp. 801–809.