

令和 3 年度茨城大学大学院理工学研究科情報工学専攻 修士

学位論文

BERT の Masked Language Model を用いた 2 文間の接  
続関係の推定

所属 情報工学専攻

著者 趙一 (20NM718L)

指導教員 新納浩幸教授

令和 4 年 2 月 3 日 (水)

## BERT の Masked Language Model を用いた 2 文間の接続関係の推定

### 著者

趙一 (20NM718L)

### 指導教員

新納浩幸教授

### 論文要旨

BERT (bidirectional encoder representations from transformers) [1] は 2018 年に GOOGLE によって開発された, 自然言語処理(NLP)のための Transformer に基づく事前学習モデルである. 自然言語処理タスクは, QA (Question Answering) や NLI (Natural Language Inference) などのタスクが含まれている.

BERT は教師なし事前学習によって NLP タスクを高い精度で解くことができる. 事前学習タスクは, NSP (Next Sentence Prediction) と MLM (Masked Language Model) である. Next Sentence Prediction は, 文の連続性を学習する. Masked Language Model は, 単語トークンの埋め込み表現を学習する. 処理したいテキストをトークンに変換し, 適切なトークン表現を出力するための学習手法である. BERT の MLM では, 入力トークンの 15% をランダムにマスクし, マスクしたところに正確なトークンを予測する.

例えば, MLM を利用した研究としてスペルチェックがある. また, 2 文間の接続関係の推定には MLM の手法が利用されている他, 深層学習モデルの RNN (Recurrent neural network) と CNN (Convolutional Neural Network) が広く使われている.

本研究では, 2 文間の接続関係を推定するタスクを対象に, BERT と Masked Language Model による教師なし学習によって文間の接続関係を推定する手法を提案する. 具体的には, まず文 2 (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 に入力する. 次に, 接続関係を表す接続詞を準備し, BERT が推定した [MASK] に対応する単語がそれらの接続詞となる確率を求めることで 2 文間の関係を推定する. 実験では [MASK] の部分が実際に接続詞であったものを対象にして正しく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. 実験結果としては, BERT による教師あり学習よりも良い結果であった.

Master's Thesis in Scholastic 2021, Major in Computer and  
Information Sciences,  
Graduate School of Science and Engineering, Ibaraki University

## **Determining the Logical Relation between Two Sentences by Using the Masked Language Model of BERT**

### **Author**

ZHAO YI (20NM718L)

### **Adviser**

Prof. Hiroyuki Shinnou

### **Abstract**

Bidirectional encoder representation from transformer (BERT)[1] was proposed by GOOGLE in 2018. BERT is a Transformer based pre-training model of Natural Language Processing (NLP). NLP tasks include QA (Question Answering), NLI (Natural Language Inference) and so on.

Unsupervised pre-training BERT performs NLP tasks with high accuracy. The pre-training tasks are NSP (Next Sentence Prediction) and MLM (Masked Language Model). In NSP, the BERT trains the representation of a sentence. In MLM, the BERT trains the representation of words. 15% of input tokens are masked, and the model predicts the correct token for mask tokens.

MLM is applied not only to pre-training task, but also to learning of downstream tasks. For example, studies using MLM includes spell checking and inferring conjunctive relationships between two sentences.

In this study, we propose a method for estimating the connection relationship between two sentences by unsupervised learning with MLM. we first place a [mask] just before sentence2 (s2) and input the form "[CLS] s1 [SEP] [MASK] s2 [SEP]" into BERT. Next, the relationship between the two sentences is inferred the conjunction relationship of the conjunctions that BERT predicted as corresponding to mask token. we evaluated whether the proposed method can correctly estimate the connection relations from the conjunctions predicted for mask tokens. Then, we compared the results with those obtained by solving this task with supervised learning. The experimental results show that the proposed method performs better than BERT supervised learning.

# 目次

第 1 章	序論	5
第 2 章	関連研究	7
2.1	BERT	7
2.2	MLM を利用したスペルチェック	9
2.3	マルチタスク学習アプローチによる 2 文間関係の推定	10
2.4	2 文間の接続関係の推定	11
第 3 章	提案手法	12
第 4 章	実験	14
4.1	実験データ	14
4.2	日本語 BERT 事前学習モデル	14
4.3	実験結果	15
第 5 章	考察	17
5.1	接続詞のグループ化	17
5.2	対象外の接続詞	17
5.3	ニューラルネットワークに基づく言語モデル	19
5.4	教師あり学習による推定	19
第 6 章	結論	22
	参考文献	24

# 第 1 章

## 序論

2018 年末に Google が発表した事前学習済みモデル BERT は自然言語処理システムの性能を大きく向上させた。BERT の利用によって自然言語処理システムの性能が大幅に向上し、タスクによっては人間の能力を超えている。従来、自然言語処理は書き言葉を対象とした教師あり学習による研究が主であった。また、教師あり学習ではタスク処理に大量のデータが必要である。現在、自然言語処理のさまざまなタスクで BERT を組み込んだシステムが利用されている。

自然言語理解は、つまり、意味的なテキストの類似度、意味的な関連性など、2つの文の関係を理解するためのタスクであり、これは自然言語理解の基本である。NLI (Natural Language Inference) は、2つの文の関係が含意関係、矛盾、または中立であるかどうかを判断するタスクである。意味的な関連性は、2つの文の関連性を推定する。この関連性では、似た文と矛盾する文のペアの両方が関連していると見なす。2つの文の関係を理解する必要のあるタスクは、自然言語処理 (NLP) タスクのさまざまなアプリケーションに、利用される。例えば、ドキュメントの要約、テキスト生成、セマンティック検索、ダイアログシステム、質問応答システム、テキスト表現の評価、機械翻訳などのアプリケーションに活用されている。

本論文では教師あり学習ではなく、教師なし学習による、BERT を利用した 2 文間の接続関係の推定を行っている。例えば (文 1) 「私は犬が好き。」 (文 2) 「猫は嫌い。」であるとき、文 1 と文 2 の接続関係が「逆接」であることを推定する。このようなタスクは自然言語処理において重要である。例えば、対話ではユーザの発話間の関係性を認識することで、論理的な繋がりのある文を返答することができる。

提案手法の具体的な方法は、BERT に内在する Mask Language Model (MLM) を利

用する. 文 2 の先頭に MASK 単語を追加し, 「(文 1) + MASK + (文 2)」を BERT に入力し, MASK 単語に入る最も適切な接続詞を MLM によって推定する. そして, MLM が推定した接続詞の種類によって接続関係を推定する.

考察では教師なし MLM による実験に加えて, 教師あり学習手法で本タスクを処理する場合との比較も行った.

## 第 2 章

# 関連研究

### 2.1 BERT

BERT (Bidirectional Encoder Representations from Transformers) は、自然言語処理の事前学習モデルである。Transformer の Multi-head Attention を採用したことで、自然言語の文脈を双方向での学習する。大規模データセットを用いて事前学習した BERT モデルに対し、自然言語処理の応用タスクに対する fine-tuning を行うことで様々な応用タスクを高精度で解くことができる。事前学習する時に用いられるタスクは複数の単語が [MASK] トークンに置き換えられた文から元の単語を予測する Masked Language Model と、入力された 2 文が連続しているかどうかを判定する Next Sentence Prediction の 2 つである。

BERT のモデルの概要を図 2.1 に示す。Masked Language Model では入力文の一部を [MASK] トークンや別の単語のトークンに置換し、置換前の入力文のトークンをラベルとして使用する。確率で [MASK] トークンへの置換を行わず、別の単語のトークンに置換する処理が行われる。このようなタスクによって、単語トークンに対する BERT の出力を最適化する。また、置換の単位をトークンから単語へと変更し、対象となった単語のサブワードを含めた「単語全体」を [MASK] トークンに置換する Whole Word Masking という手法が採用される場合がある。

Next Sentence Prediction では [SEP] トークンで区切られた 2 文を入力し、それらが連続した文であるかどうかを識別する。識別は、[CLS] トークンに対する出力によって行う。各タスクの概要を図 2.2 及び図 2.3 にそれぞれ示す。

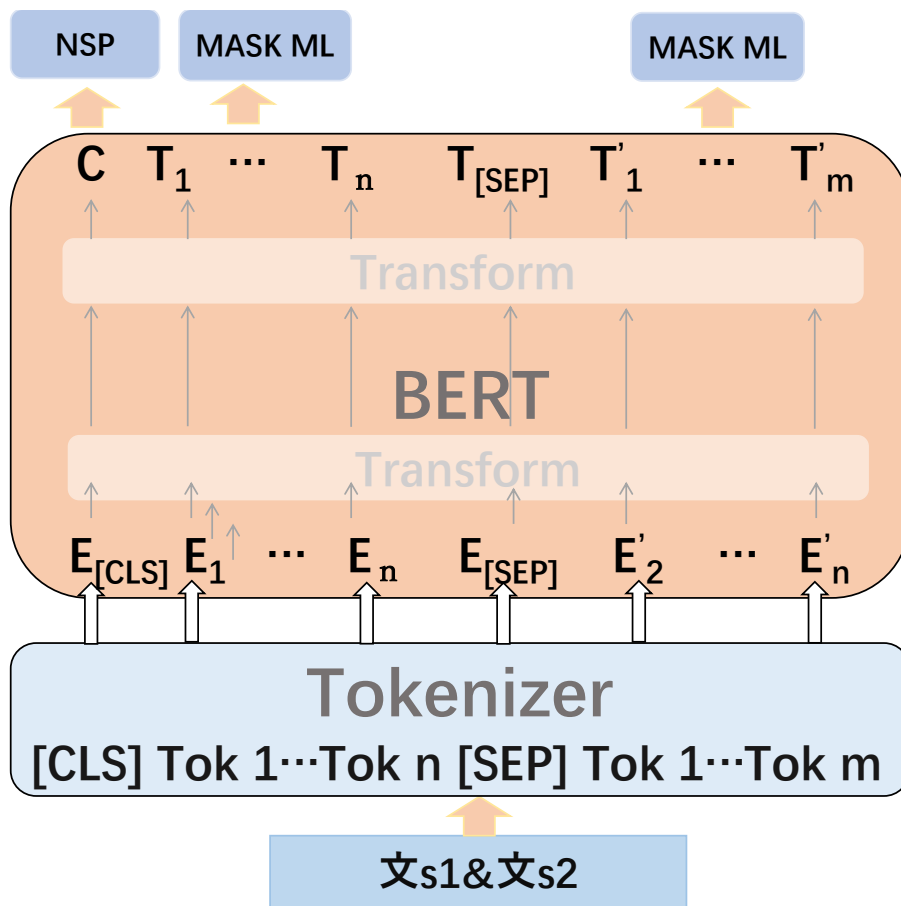


図 2.1: BERT モデルの概要)

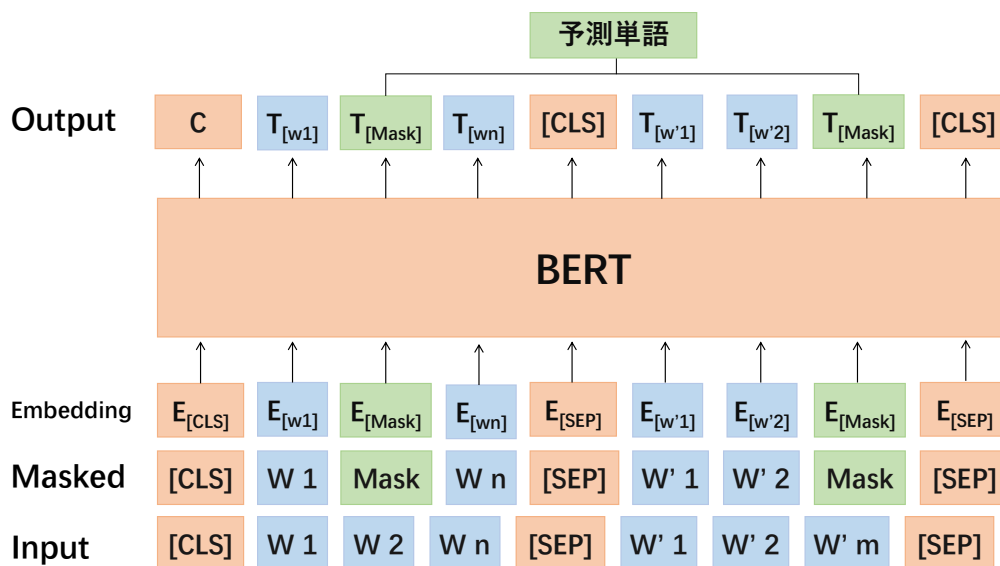


図 2.2: Masked Language Model タスク (概要)

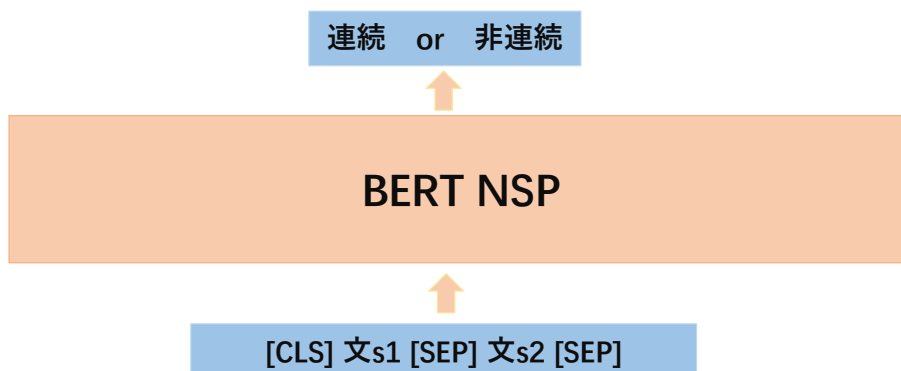


図 2.3: Next Sentence Prediction タスク (概要)

## 2.2 MLM を利用したスペルチェック

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 には言語理解のための知識を習得する一定の能力があるため、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。これを利用して、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。Zhang らの研究 [2] 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 (図 2.4)[2] 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。Zhang は Chinese Spelling Check Task (CSC) [3] コーパスの中で SIGHAN 部分と News Title コーパスと 5 million news titles コーパスを利用し、それぞれの実験を行っていた。

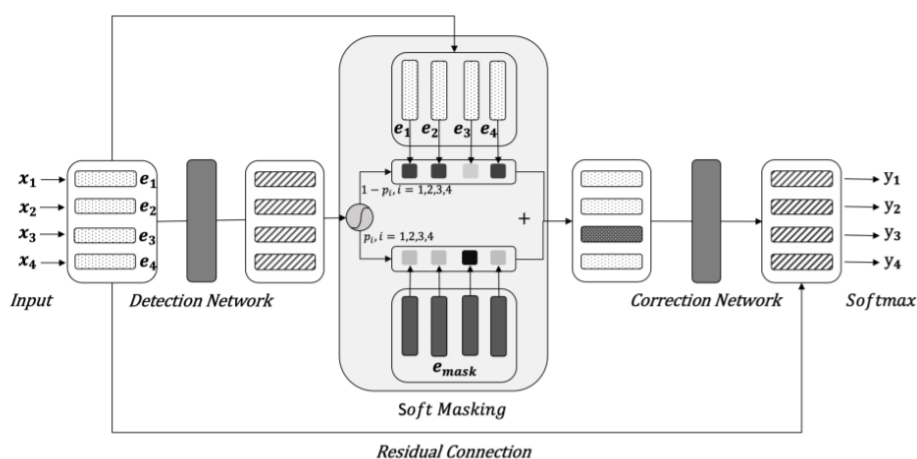


図 2.4: soft-marked BERT のアーキテクチ

Soft-marked BERT の中には Bi-GRU 検出ネットワークと BERT に基づく補正ネットワークという二つのネットワークが含まれている。検出ネットワークは間違っただけの確率を予測する。補正ネットワークは間違っただけの修正確率を予測し、前者はソフトマスキングを使用し、その予測を後者に渡す。BERT の事前トレーニングでは、予測のために文字の 15% のみがマスクされているため、学習済みモデルには十分なエラー検出機能を備えていない。検出ネットワークの出力を BERT の入力トークンがマスクされる確率とし、エラーの可能性が高い場合、マスクされる可能性が高くなる。

Zhang らは、Soft-Masked BERT を、BERT のみを使用する方法を含むいくつかのベースラインモデル方法と比較した。データセットとして、SIGHAN のベンチマークデータセットを使用した。また、ニュースタイトルと呼ばれる大規模で高品質のデータセットを作成した。このデータセットにはニュース記事のヘッドラインが含まれており、以前のデータセットの 10 倍の大きさである。実験結果は、Soft-Masked BERT が、両方のデータセットの精度でベースラインモデルを大幅に上回っていることを示している。この研究は、主に 2 つの貢献がある。

1. CSC タスクを解決するための新しいニューラルネットワーク構造で Soft-Masked BERT を提案したこと。
2. Soft-Masked BERT の有効性を実験で検証すること。

## 2.3 マルチタスク学習アプローチによる 2 文間関係の推定

2 文間の接続関係を推定するタスクには、2 つの課題がある。1 つは言語をベクトル表現で表す方法であり、もう 1 つは文間関係の特徴量を表すことが困難である点である。bag-of-words モデルのようなテキスト表現方法の分布はスパースになってしまう。[4]

単語埋め込み技術は、one-hot encoding のような高次元でスパースなベクトルを低次元で密に分散されたベクトルに変換することから、単語表現のスパース性の問題を解決するために使用されている。ここで、単語の埋め込み表現は文の意味的な情報と構文情報を埋め込み、言語の規則を表す。これらの特性のため、単語の埋め込みは、再帰型ニューラルネットワーク (RNN) [5,6]、畳み込みニューラルネットワーク (CNN) などのディープニューラルネットワークの優れた初期シードとして広く使用されている。

ディープニューラルネットワークは、さまざまな分野で優れた表現力を示している。広く研究されているのは二つのアプローチがある。エンコーディングベースとソフトア

ライメントモデルである。エンコーディングベースのモデルは、単語ベクトルを文書ベクトルに結ぶときに、各文を個別的にエンコードする。ソフトアラインメントモデルは、主に文のペアを共同でエンコードするメソッドで使用される。この手法は、Parikh [7] によって最初に提案され、2つの文を互いに比較することによって同時にエンコードされる方法を利用している。

深層学習で二文間の接続関係の推定に HongSeok Choi がマルチタスクフレームワークという手法を提案した。エンコーディングラベルでより多くの特徴があるため、エンコーディングベースモデルより良いパフォーマンスできる。ただし、深層学習には大量のデータが必要になるため、データ量が少ない場合は過学習する可能性がある。いくつかのマルチタスク学習 (MTL) アプローチは、過学習に対処する手法として提案された [8-10]。MTL は、複数の関連するタスクを同時に学習し、各タスクに積極的な効果を生み出すアプローチである。MTL は、複数のタスクでトレーニングすることにより、より一般的な表現を学習する [11]。複数のタスクによるデータ拡張の効果により、モデルの過学習が防止される。

HongSeok Choi の実験では、他のアプローチと比較してメリッドがあることが示されている。MTL アプローチは外部ツールやタスク固有のパラメーターを必要としない。プレーンテキストを読み取り、単一のユニバーサル表現を生成し、複数のタスクを予測するだけである。実験結果は提案されたモデルが、文のペアを共同でエンコードすることにより、他のエンコードベースのモデルよりも優れていることを示している。

## 2.4 2文間の接続関係の推定

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 [12] 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大塚らの研究 [13] がある。そ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 Auto Encoder [14] 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。

## 第3章

# 提案手法

2つの文を  $s_1$  と  $s_2$  と表す。文  $s_1$  と  $s_2$  の接続関係を推定したい場合には  $s_1$  と  $s_2$  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。具体的には利用したコーパスを文型から整理し、一行づつ” で終わらせる。  $s_1$  の直前に [CLS] を置き、  $s_1$  の終端に [SEP] を置き、  $s_2$  の直前にマスク単語を置き、また  $s_2$  の終端に [SEP] を置く。この処理によって得られるトークン列を図 3.1 に示す。



図 3.1: BERT に入力 token 列

整理した token 列を BERT に入力し、BERT が推定した [MASK] に対応する単語から文  $s_1$  と文  $s_2$  の接続関係を推定する。

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 21 種類の接続詞を対象とした。表 3.1 にその 21 種類の接続詞を示す、[MASK] の位置にその接続詞が現れる確率を求め、その中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。

図 3.2 に本手法の動作例を示す。図では以下の二つの文  $s_1$  「私は犬が好き。」と  $s_2$  「猫は嫌い。」の接続関係を推定している。まず、文  $s_1$  と文  $s_2$  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



## 第 4 章

# 実験

### 4.1 実験データ

表 3.1 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.

### 4.2 日本語 BERT 事前学習モデル

本研究では日本語に対応した事前学習モデルとして, 東北大学乾研究室が以下のサイトで公開している日本語事前学習モデルを使用する.

<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. このモデルは, 訓練コーパスを日本語 Wikipedia(約 3000 万文) で構成されている. モデルの構造は BERT BASE と同じ 12 層, 語彙数は 32000 としている. モデルは, BASE と同じ構成を使用し, 訓練される.

この事前学習モデルの入力となるテキストは, 同じく東北大学乾研究室が公開している MeCab 形態素解析器と mecab-ipadic-NEologd 辞書を使用し, テキストを文に分割する. 例として「私は犬が好き。」の文では, 「私」「は」「犬」「が」「好き」「。」のように

扱われる。

### 4.3 実験結果

実験の結果を表 4.1 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。

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  $0.048(= 1/21)$  という値で，BERT を利用した効果は確認できる。

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 4.1: 実験結果

## 第 5 章

# 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。ここで、接続詞をグループ化する。同じ意味の接続詞はグループ化でき、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。

接続関係の種類に関しては従来多くの研究者がいくつか提案している [15]。日本語では、市川が文の接続関係を、「順接」、「逆接」、「添加」、「対比」、「転換」、「同列」、「補足」、「連鎖」の 8 つの類型に分類しているが [12]、本研究は筆者の主観に基づいて、表 5.1 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 5.1 に示す。1 列目は 8 個グループの番号であり、2 列目はグループした単語であり、3 列目は正解数である。

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

言語の構成は非常に複雑である。同じ意味の文は多くの言い方がある。つまり、接続詞又は文間で接続を表す言い方も多くある。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

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		1067

表 5.1: グループ化した接続詞に対する正解率

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, さもなければ, かくして, けど, 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

上記の接続詞は大体分けると2つの種類がある。下線で引かれている接続詞と引かれていない接続詞である。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）についても同様に利用したBERTの語彙集合では一単語と認識

できないである。

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 ニューラルネットワークに基づく言語モデル

ニューラルネットワークに基づく言語モデルは、異なる学習方法に応じて大きく分けると2つの種類がある。Feature-based と Fine-tune である。

Feature-based はニューラルネットワークを訓練することによって、単語の埋め込みを取得する。つまり、語彙の各単語を表すために、ニューラルネットワーク経てより適切なベクトルが取得される。モデル自体が使用されないが、モデル訓練から取得したパラメータを単語の埋め込みとして使用する。例としては、ELMO と word2vec である。

fine-tune は微調整を通じ、高いパフォーマンスが出るというモデルである。特定の下流タスクに応じて、元モデルにいくつか変更させ、現在のタスクに必要なものを最終に出力する。下流タスクと一致するために、これらの変更は通常モデルの最後のレイヤーにあるか、既存のネットワークに新しくネットワークを追加する。

### 5.4 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5.1と図5.2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5.1のWの層のパラメータだけを学習する形である。fine tuning の手法の学習では図5.2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5.1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

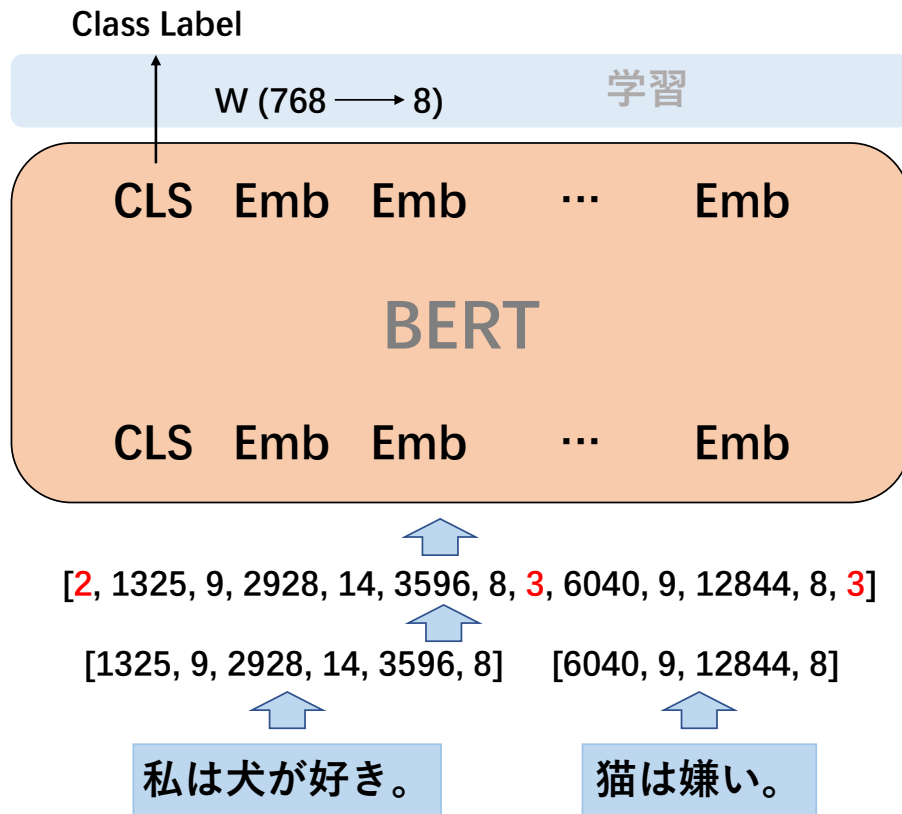


図 5.1: Feature-based による推定

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 5.3 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。

ただし今回の訓練データは 1050 個の文だけとしたので、非常に少ないと思われる。正解率が低いと訓練データの量の問題であると考え、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。

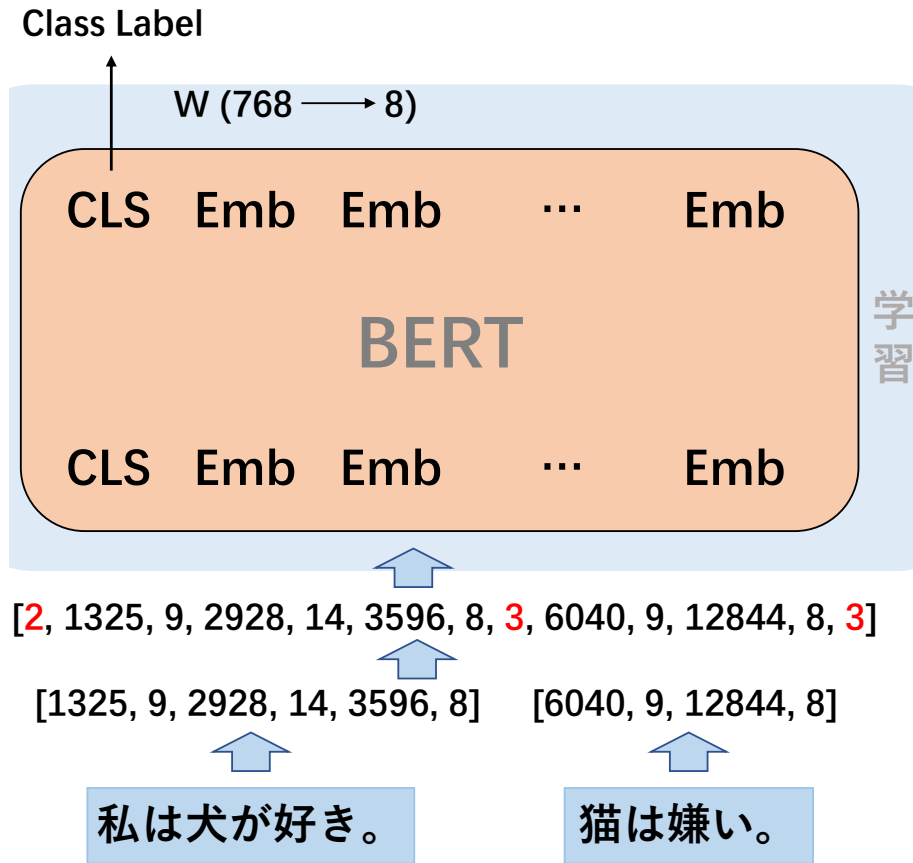


図 5.2: Fine-tuning による推定

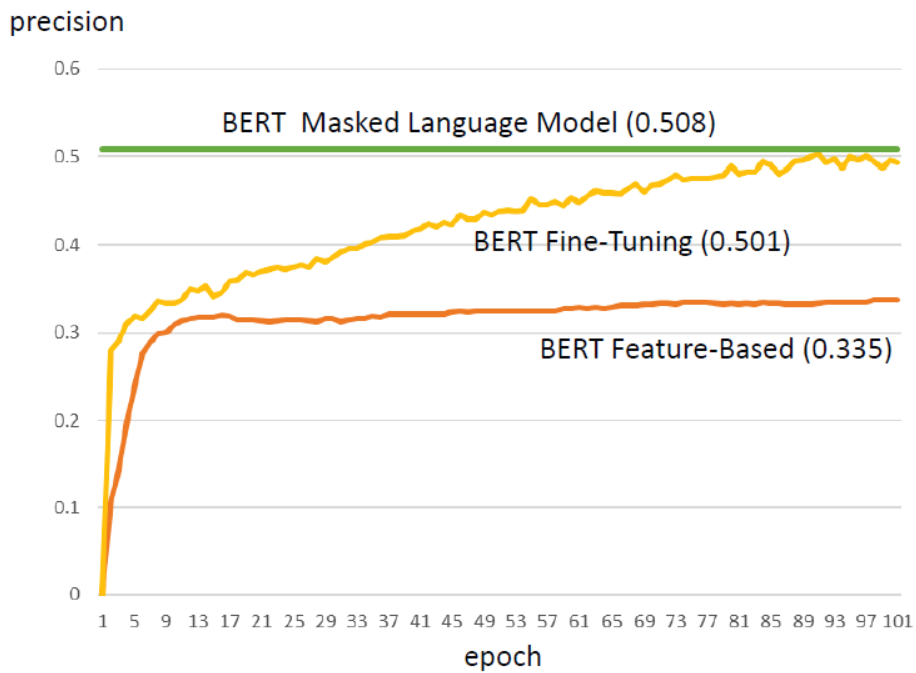


図 5.3: 教師あり学習による推定との比較

## 第 6 章

# 結論

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 カテゴリー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 カテゴリー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。

# 謝辞

本研究を進めるにあたって、多くのご指導を頂いた指導教員の新納浩幸教授に感謝致します。また、日常の議論を通して多くの知識、示唆を頂いた田中裕隆さん、新納研究室の皆様感謝します。

## 参考文献

- [1] 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.
- [2] 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.
- [3] Yuzhong Hong, Xianguo Yu, Neng He, Nan Liu, and Junhui Liu. (2019). “A fast, adaptable, simple, powerful chinese spell checker based on daedecoder paradigm.” In Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019),pages 160–169.
- [4] HongSeok Choi, Hyunju Lee (2019). “Multitask learning approach for understanding the relationship between two sentences” Information Sciences 485 (2019) 413–426
- [5] T. Mikolov, M. Karafiát, L. Burget, J. Cernocký, S. Khudanpur. “Recurrent neural network based language model” 11th Annual Conference of the International Speech Communication Association (ISCA), 2010, pp. 1045–1048 .
- [6] S. Hochreiter, J. Schmidhuber. “ Long short-term memory” Neural Comput, 9 (8) (1997) 1735–1780 .
- [7] A. Parikh, O. Täckström, D. Das, J. Uszkoreit. “ A decomposable attention model for natural language inference” Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for

- Computational Linguistics, 2016, pp. 2249–2255 .
- [8] D. Dong, H. Wu, W. He, D. Yu, H. Wang. “ Multi-task learning for multiple language translation ” Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL and IJCNL), Association for Computational Linguistics, 2015, pp. 1723–1732 .
- [9] R. Collobert, J. Weston. “ A unified architecture for natural language processing: deep neural networks with multitask learning ” Proceedings of the 25th International Conference on Machine Learning (ICML), ACM, 2008, pp. 160–167, doi: 10.1145/1390156.1390177 .
- [10] M.-T. Luong, Q.V. Le, I. Sutskever, O. Vinyals, L. Kaiser. “ Multi-Task Sequence to Sequence Learning ” arXiv preprint arXiv:1511.06114 (2015).
- [11] S. Ruder. “ An Overview of Multi-Task Learning in Deep Neural Networks. ” arXiv:1706.05098 (2017).
- [12] 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.
- [13] 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 Au-to Encoder 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.
- [14] 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.
- [15] 市川孝 (1978). “ 国語教育のための文章論概説. ” 教育出版,pp.65-67.