

日本語 PromptBERT におけるプロンプトベース対照学習の有効性とプロンプトの性質の調査

芝山 直希^{1,a)} 古宮 嘉那子^{2,b)} 新納 浩幸^{1,c)}

概要: PromptBERT は文の埋め込み表現を構築する手法である。たとえば、「[X] means [MASK]」のような入力文 [X] と [MASK] トークンを組み合わせたテンプレートを用意しておき、[MASK] トークンに対する埋め込み表現を BERT から推定し、それを入力文の埋め込み表現とみなすという手法である。このように簡易な手法でありながら高い性能を示すことから近年注目されている。また PromptBERT はテンプレートを組み合わせて対照学習を行うことで、モデルの性能を高められることも大きな特徴である。ただし、単純にテンプレートから作られる埋め込み表現を対比させるだけでも対照学習は行えるはずだが、PromptBERT のオリジナルの論文では、推定に利用したテンプレート自身を BERT に入力した場合の [MASK] トークンの埋め込み表現を、テンプレート自身が保有する情報によるノイズとみなし、推定した入力文の埋め込み表現からその値を減じて対照学習を行っている（オリジナルの手法）。本論文では我々が考案した日本語テンプレートによる日本語 PromptBERT に対して、通常の対照学習とオリジナルの手法を比較した実験を行い、テンプレートのノイズを除去する効果を確認する。また PromptBERT で利用するテンプレートの [MASK] トークンから推定される単語を調査することで、プロンプトの性質を考察する。

1. はじめに

文の埋め込み表現は、自然言語処理の多くのタスクで必要になるため、従来より様々な構築手法が提案されてきた。近年では BERT (Bidirectional Encoder Representations from Transformers)[1] やその基盤となる Transformer[6] を利用した構築手法の研究が活発である。文の埋め込み表現を構築する標準的手法は SentenceBERT [5] であるが、2022 年に提案された PromptBERT[4] は、英語の STS タスクでの評価において SentenceBERT などの近年提案された構築手法を大きく上回る性能を示した。PromptBERT は事前定義された「[X] means [MASK]」のような「[MASK] トークンと挿入箇所 [X] を含むテンプレート」の [X] に入力文 s を挿入して文を作成し、作成した文を BERT に入力した時に出力される [MASK] トークンの埋め込み表現を、文 s の埋め込み表現とする簡易な手法である。

日本語で PromptBERT の運用を試みる場合、課題と確認すべき点が存在する。まず、PromptBERT の性能が利

用するテンプレートに依存することは明らかであり、性能の高い日本語 PromptBERT を構築するには適切なテンプレートを考案する必要がある。また PromptBERT を用いれば、2つのテンプレートから構築される文の埋め込み表現を直接対比させることで通常の対照学習が行えるが、PromptBERT を提案した論文 [4] ではテンプレートの文脈・情報量によるノイズを除く形の工夫を行った対照学習を行っている（以降、これをオリジナルの手法と呼ぶ）。英語においては、オリジナルの手法と、2つの埋め込み表現を獲得し対比させる「通常の対照学習」を採用した手法である SimCSE[2] や ConSERT[8] との比較があるが、日本語においては通常の対照学習を採用したモデルとの比較は行われていない。日本語においても通常の対照学習との差異を確認すべきである。

本論文では、前述した点のうち、特に後者の確認すべき点に着目し、日本語 PromptBERT におけるプロンプトベースの対照学習の有効性を調査した。具体的には、論文 [4] で提案された手法と、それを改変し単純化した通常の対照学習手法を比較した。日本語のテンプレートとしては、既存研究 [10] のテンプレートを利用した。また、PromptBERT は [MASK] トークンの埋め込み表現を文の埋め込み表現として出力する。この性質を利用し、テンプレート内の [MASK] トークンから推定される単語を調査することで、

¹ 茨城大学
Ibaraki University

² 東京農工大学
Tokyo University of Agriculture and Technology

a) 21nd303a@vc.ibaraki.ac.jp

b) kkomiya@go.tuat.ac.jp

c) hiroyuki.shinnou.0828@vc.ibaraki.ac.jp

プロンプトの性質を調査した。

2. 関連研究

本論文は PromptBERT の提案論文 [4] の教師なし学習手法とオリジナルの手法を一部改変し単純化した手法の 2 つを比較し、考察する。

本論文に先立ち、日本語 PromptBERT を扱った論文に、論文 [10] がある。論文 [10] では、日本語 PromptBERT 用に「[X] means [MASK]」の代わりとなりうるような日本語のテンプレートを複数提案し、構築した日本語 PromptBERT を既存手法と比較している。その結果、「[MASK] は [X] である」という構文のテンプレートが日本語 PromptBERT において最も有効であったことを示した。また、BERT の訓練時に利用されるタスクの性質自身を考慮してテンプレートを作成・使用することで、もともとの BERT モデルの性能を評価できる可能性と、テンプレートの文法的正しさが文の埋め込み表現の安定性に影響する可能性について考察している。

また、オリジナルの手法は対照学習を利用した手法である。対照学習とは、類似したデータを近く、異なるデータを遠くするように特徴量を学習する自己教示型の機械学習である。PromptBERT 以外に文の埋め込み表現の構築に対照学習を利用した研究としては、SimCSE[2] と ConSERT[8] がある。SimCSE の教師なし学習は、入力文の埋め込み表現を 2 回、異なる Dropout マスクを適用して出力することで、文の埋め込み表現を 2 つ獲得し、対照学習を行う手法である。また、ConSERT は、入力文の埋め込み表現群から別の埋め込み表現群を作成し、元の埋め込み表現群と作成された埋め込み表現群のそれぞれから文の埋め込み表現を作成し対比させることで、対照学習を行う。さらに、教師あり学習と組み合わせた学習手法の提案を行っている。

3. PromptBERT

PromptBERT は以下の 3 工程により文の埋め込み表現を作成する。この手法を適用し、追加学習を行わないモデルを Prompt based BERT と呼ぶことにする。

- (1) 入力文挿入箇所 [X] と [MASK] トークンを含むテンプレート (プロンプト) を作成・準備する。
- (2) テンプレート内の [X] を入力文に置換し、それを BERT に入力する。
- (3) BERT から推定される [MASK] トークンの埋め込み表現を文の埋め込み表現として抽出する。

更に PromptBERT では「Prompt based contrastive learning with template denoising」[4] と名付けられた対照学習を用いて Prompt based BERT の性能を高めている。その手順を以下に示す。

- (1) テンプレートを 2 つ (t_1 と t_2) 準備する。
- (2) t_1 と t_2 と BERT を用いて入力文 s に対する文の埋め

込み表現 h_1, h_2 を得る。

- (3) t_1 及び t_2 内の [X] を入力文に置換せずに t_1 及び t_2 自身を BERT に直接入力し、[MASK] トークンの埋め込み表現 h'_1, h'_2 を得る。この時、[X] より後ろのトークンの position ids を入力文の該当箇所の position ids と同じ値にする。この処理により、[X] を入力文に置換して得た埋め込み表現 h_1 及び h_2 と置換せずに得た埋め込み表現 h'_1 及び h'_2 において、同じ表層のトークンの position ids がそれぞれ一致するようになる。
- (4) $h_1 - h'_1$ と $h_2 - h'_2$ を用いて対照損失を計算し、この 2 つを近づけるように学習する。

オリジナルの手法の具体的な挙動の例を示す。例えば入力文 s を「彼はいい奴だ。」とする。更に、入力文の文頭に [CLS] トークンが、文末に [SEP] トークンが自動的に挿入されると仮定する。

- (1) t_1 に「[X][MASK]」を、 t_2 に「[X] は [MASK] だ。」を使うと仮定する。
- (2) t_1, t_2 に s を挿入すると、「彼はいい奴だ。[MASK]」と「彼はいい奴だ。は [MASK] だ。」という表現を得る。これらを入力として埋め込み表現を計算し、[MASK] トークンの埋め込み表現 h_1, h_2 を得る。
- (3) 内蔵する BERT に「[X][MASK]」($=t_1$)、「[X] は [MASK] だ。」($=t_2$) を入力し、[MASK] トークンの埋め込み表現 h'_1, h'_2 を得る。この時、 s が「彼、は、いい、奴、だ。」というトークン列に分割されたと仮定すると、入力時の t_1 の position ids は [0, 1, 6, 7] に、 t_2 の position ids は [0, 1, 6, 7, 8, 9] に調整される。
- (4) $h_1 - h'_1$ と $h_2 - h'_2$ を用いて対照損失を計算し、この 2 つを近づけるように学習する。

本手法において最も特徴的な工夫が「テンプレート自身の埋め込み表現を入力文の埋め込み表現から減算する」という点である。テンプレート t 内の [X] を入力文 s に置換するという操作は、 s に t が持つ情報を追加し、文脈を変化させる性質を持つ。 t は s と同じ自然言語で記述されるため、 s と t に s を挿入した文 $t(s)$ の間に、文脈及び情報量の差異が生じるのは回避出来ない。この差異が各単語の埋め込み表現に影響を与え、 s の埋め込み表現として使用される [MASK] トークンの埋め込み表現に対するノイズになる可能性がある。これを低減し、[MASK] トークンに不適切な意味が付与される事を回避する試みが [X] を置換しない時の t の埋め込み表現、すなわち「テンプレート自身の埋め込み表現」 $t([X])$ で減算する行為である。本手法では [X] 以降のテンプレート部分の position id を $t(s)$ のそれと一致させることで、 $t(s)$ と同じトークン長の、一切の情報・文脈を持たない文の埋め込み表現を獲得し、「Template denoising」を行っていると考えられる。このノイズ除去によって、異なるテンプレートの元で作成された 2 つの埋め込み表現から、ノイズになり得る「それぞれのテンプレート固有の情報・文脈」

の影響が低減され、効果的な対照学習が期待できる。

4. プロンプトベース対照学習の有効性調査

オリジナルの手法では、テンプレートの [X] を入力文に置換した際の埋め込み表現 h から、置換を行わなかった際の埋め込み表現 h' を減算して対照学習を行っているが、理論上はこの減算処理を行わなくても対照学習が可能である。本節では減算処理を省略するように改変した手法を通常の対照学習手法とし、この手法で作成した日本語 PromptBERT とオリジナルの手法で作成した日本語 PromptBERT を比較する事で、両手法の有効性を調査した。

本論文における通常の対照学習として、以下に示す手順の学習手法を定義する。

- (1) テンプレートと学習対象となる日本語 PromptBERT をそれぞれ 2 つ準備する。
- (2) テンプレート 1 を採用した学習対象 1 で、文の埋め込み表現 h_1 を得る。
- (3) テンプレート 2 を採用した学習対象 2 で、同様に文の埋め込み表現 h_2 を得る。
- (4) 以下の式に基づいて損失を計算し、学習対象 1、2 を最適化する。 N は文の埋め込み表現の次元数、 h_{ni} は文の埋め込み表現 h_n の i 番目の次元の値である。

$$loss = \sum_{i=1}^N \| h_{1i} - h_{2i} \|^2$$

4.1 実験設定と使用データ・テンプレート

本調査において、日本語 PromptBERT の構築及び評価による比較の際に使用するデータセットとして、”Japanese Realistic Textual Entailment Corpus” (JRTE)[11][3]^{*1*}2 を採用した。当該データセットは株式会社リクルートが公開・提供している、「じゃらんクチコミデータ」を元に作成された二値分類型含意関係推論データセットである。本論文において使用したテンプレート、本論文内での略称、及び対照学習時に対とするテンプレートの一覧を表 1 に示す。^{*3}t1, t2 と t3, t4 はそれぞれ [MASK] と [X] の位置が異なる「似た構文」となっている。

上記 4 つのテンプレートを用いて通常とオリジナルの 2 つの手法で対照学習を行い、構築した日本語 PromptBERT、及び対照学習を行っていない Prompt based BERT について、JRTE のテストデータに対する正解率を評価した。なお正解率は、5 つの異なるシード値を用いてモデルを評価した際の最大値と最小値の平均である。日本語 PromptBERT 及び日本語 Prompt based BERT の元

^{*1} <https://github.com/megagonlabs/jrte-corpus>

^{*2} 本論文では言語処理学会第 26 回年次大会にて構築されたもの [11] を使用した。

^{*3} 例：出力に使用するテンプレートに t2 を採用する場合、t1 と組み合わせて対照学習を行った。

表 1 使用したテンプレート及び対照学習ペア一覧と略称

テンプレート	略称	対となるペア	引用元等
[MASK] とは、[X] である。	t1	t2	先行研究 [10]
[X] とは、[MASK] である。	t2	t1	先行研究 [10]
[X] は [MASK] を意味する。	t3	t4	元論文の直訳
[MASK] は [X] を意味する。	t4	t3	元論文の直訳を元に考案

とするモデルには、東北大学が公開している日本語 BERT v1^{*4}を使用した。

4.2 調査結果・考察

前小節で述べた評価実験の結果を表 2 及び図 1 に示す。

表 2 JRTE による各モデルの正解率 (5 シードの最大最小平均)

モデル	[MASK] とは、[X] である。	[X] とは、[MASK] である。	[X] は [MASK] を意味する。	[MASK] は [X] を意味する。
	t1	t2	t3	t4
Prompt based BERT	0.6575	0.4473	0.4349	0.5950
PromptBERT (通常の対照学習)	0.6550	0.5310	0.4928	0.6706
PromptBERT (オリジナルの手法)	0.7062	0.6778	0.6971	0.7146

表 2 の全てのテンプレートにおいて、オリジナルの手法が、通常の対照学習以上のモデル改善能力を示した。PromptBERT では、テンプレートの [X] を入力文に置換することから、テンプレートを構成する [X] より後ろのトークンの position id が入力文によって異なる。オリジナルの手法はこの悪影響を抑える工夫が施されているが、通常の対照学習手法では行っていない。よって、通常の対照学習手法は、テンプレートによる入力文の変形に伴う文脈変化の影響を受けた可能性があると考えられる。また、Prompt based BERT を含む全てのモデルで t1, t4 を採用したモデルが、他のテンプレートを使用したモデルと比べて高い性能を示した。t1 は、日本語 BERT v1 が事前学習に使用したコーパスである日本語 Wikipedia において、頻出する構文である「(見出し語) は、(説明文)」と類似している [10]。よって、t1 の構文の知識は事前学習で獲得している為、t1 は既知のテンプレートであると言える。t4 は t1 と類似しているため、事前学習で得た知識を使用・応用出来たテンプレートが高い性能を示したと考えられる。し

^{*4} <https://huggingface.co/cl-tohoku/bert-base-japanese>

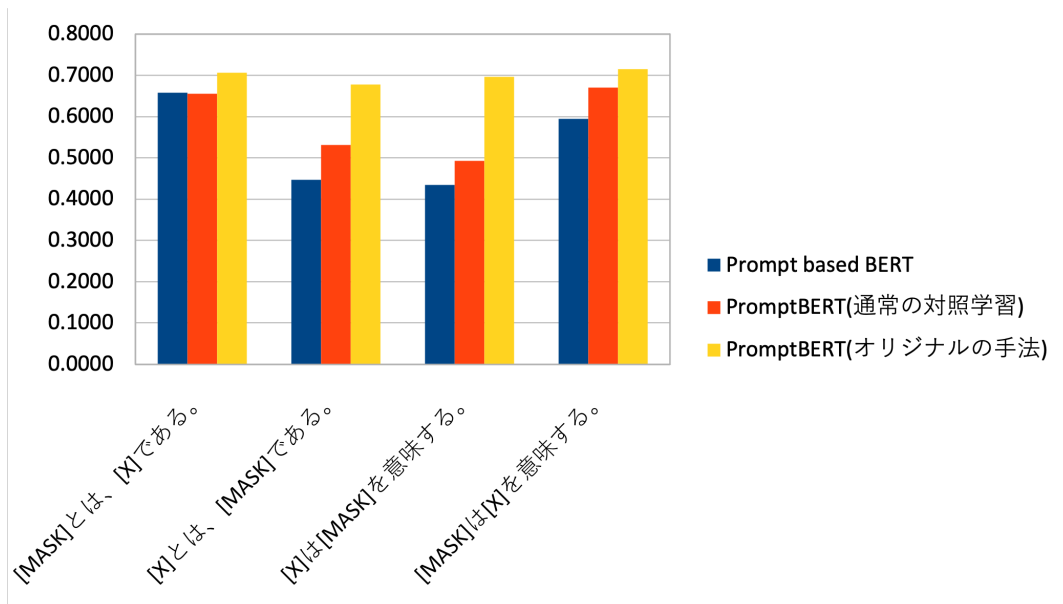


図 1 JRTE による各モデルの正解率 (5 シードの最大最小平均, グラフ)

しかし通常的手法においては、t1 を使用した場合のみ性能を改善出来ていない。通常的手法を用いても、t4 を使用した場合において性能を改善出来ている点を考慮すると、通常の対照学習手法は「モデルが事前学習で得た知識を応用できる未知のテンプレート」を用いた場合は性能を改善出来るが、既知のテンプレートを使用した際の性能改善を苦手とする可能性がある。

5. プロンプトの性質の調査

PromptBERT において、文の埋め込み表現は [MASK] トークンの埋め込み表現である。これは、採用するテンプレートの [MASK] トークンを調査することで、テンプレートの性質を調査出来るという事を意味する。特に空所推定タスクに適応したモデルを用いることで、入力文を挿入した時テンプレートがどのような意味を [MASK] トークンに付与しているか可視化することが出来る。本論文で採用した全てのテンプレートは [X] と [MASK] が意味的に等しくなる構文であり、文章を一語要約するタスクまたは意味から単語を検索するタスクのいずれかで見做すことも出来る (論文 [10])。よって、空所推定モデルによって可視化した「入力文挿入時のテンプレートの [MASK] トークン」と入力文との関連性を考察することで、テンプレートの基本性能や性質への理解を深めることが出来ると考えられる。本論文では、前節で使用したテンプレートに対し、空所推定タスクで fine-tune した日本語 BERT を用いた空所推定を行い、主観評価によるプロンプトの性質調査を試みた。

5.1 使用データセット及び調査手順

本節において、空所推定タスクの学習元データとして JSTS の訓練データを使用した。JSTS は近年作成・公開

された汎用的日本語ベンチマーク JGLUE[9] に収録されている、2 文間の類似度を求めるタスクのデータセットである。当該ベンチマークは文書分類、含意関係推論、質問応答などの複数種類のデータセットで構成されており、本論文執筆時点においては訓練データ、及び開発データのみが公開されている*5。JGLUE にはドメインに対する依存性が存在しないため、JSTS に含まれる文を使用することで一定以上の汎用性を有する空所推定モデルの作成が可能だと考えられる。当該データセットを用いて、東北 BERT v1 に対する Masked Language Model による追加学習を 3 epoch 分実行した。その後、学習後のモデルと JRTE テストデータに出現する文を用いて、入力文を挿入したテンプレートに対する空所推定を行った。

5.2 調査結果に対する主観評価

前小節で述べた調査による空所推定の結果の中で、興味深いと感じたものを表 3 に抜粋する。

これらの推論結果は主に「特定の表現に反応していると思われるもの」(キッチン、数字など)「意味的圧縮がされていると考えられるもの」に分けられ、前者は特に無関係である場合 (英語、キッチンなど) とそうでない場合に分けられる。使用したテンプレート毎に頻出する推論語は異なる場合があるため、各テンプレートが何らかの性質または傾向を有している可能性がある。

より詳細な調査を行うためには客観的な指標が必要となるが、テンプレートであるが故の制約が存在する。[MASK] トークンに対する正解が入力文によって異なり、正解とな

*5 GitHub リポジトリ (<https://github.com/yahoojapan/JGLUE>) には、テストデータはリーダーボードと同時に公開されるとある。

表 3 入力文挿入後のテンプレートに対する興味深い空所予測実例

入力文	テンプレート	推論結果
接客はとても丁寧でした。	t1	キッチン
岩盤浴は、良かったです。	t1	浴
お掃除もきちんとされていまし た。	t1	キッチン
清掃がしっかりされています。	t1	トイレ
御祝いのケーキありがとうございました!	t2	ケーキ
岩盤浴は、良かったです。	t2	雪
また、温泉も最高です。	t2	英語
大浴場でリラックスできました。	t2	英語
御祝いのケーキありがとうございました!	t3	クリスマス
肌がスベスベになりました。	t3	赤ちゃん
清潔感ばっちり!	t3	浄化
枕元にコンセントがあれば良かっ た。	t3	赤ちゃん
岩盤浴は、良かったです。	t4	キッチン
コストパフォーマンスは最高です!	t4	数字
大浴場でリラックスできました。	t4	男性
とにかく、清潔でした。	t4	トイレ

る語も定義されていない点である。この点を解消できなければ、プロンプトの性質の調査に採用できる評価指標が限られてくるだろう。一方で、各推論候補の確信度はプロンプトに対する調査でも問題なく取得できる。また、ラベルなし空所推定データセットに適用できる手法・指標であれば問題なく適用できると考えられる。

5.3 追加調査：日本語 PromptBERT による空所推定

作成した日本語 PromptBERT に内蔵されている BERT を用いた、追加学習を伴わない空所推定による調査も試験的にを行った。この調査には t1 を出力用プロンプトとして採用し、シード値を 503 として 4 節の評価実験で作成した日本語 PromptBERT を用いた。5.1 節と同様に、JRTE テストデータに出現する文を空所推定時の入力文として使用した。

追加学習を行っていないのも相まって、関連の薄い語やサブワードが推論結果として挙がったが、推論結果として採用された語には「入力文に依存していると思われる偏り」が存在した。この点から、日本語 PromptBERT も各テンプレートに依存する「何らかの埋め込み表現の圧縮手法」を会得している可能性が考えられる。

6. おわりに

本論文では、JRTE データセットを用いて構築した日本語 PromptBERT を評価・比較する事で、PromptBERT の提案論文 [4] の手法と異なる 2 種のテンプレートを使用して行う単純な対照学習手法の有効性を調査した。その

結果として、提案論文の手法が、2 種のテンプレートで行う単純な対照学習手法よりも高い改善能力を示した。また、PromptBERT がテンプレート中に含まれる [MASK] トークンの埋め込みを文の埋め込み表現とみなしているという性質を利用して、主観評価によるプロンプトの性質の調査を行った。結果として各テンプレート及び日本語 PromptBERT は、テンプレート依存の何らかの性質・埋め込み表現圧縮手法を持つ可能性がある事を発見した。

今後は確信度を用いたプロンプトの性質調査を行いつつ、客観的評価の方法・指標を検討していきたい。

謝辞

本研究は 2022 年度国立情報学研究所公募型共同研究 (22FC04)、JSPS 科研費 22K12145 の助成を受けています。

参考文献

- [1] Devlin, J., Chang, M., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (2018).
- [2] Gao, T., Yao, X. and Chen, D.: SimCSE: Simple Contrastive Learning of Sentence Embeddings, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Online and Punta Cana, Dominican Republic, Association for Computational Linguistics, pp. 6894–6910 (online), DOI: 10.18653/v1/2021.emnlp-main.552 (2021).
- [3] Hayashibe, Y.: Japanese Realistic Textual Entailment Corpus, *Proceedings of the 12th Language Resources and Evaluation Conference*, Marseille, France, European Language Resources Association, pp. 6827–6834 (online), available from <https://aclanthology.org/2020.lrec-1.843> (2020).
- [4] Jiang, T., Huang, S., Zhang, Z., Wang, D., Zhuang, F., Wei, F., Huang, H., Zhang, L. and Zhang, Q.: PromptBERT: Improving BERT Sentence Embeddings with Prompts (2022).
- [5] Reimers, N. and Gurevych, I.: Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, (online), available from <https://arxiv.org/abs/1908.10084> (2019).
- [6] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I.: Attention is all you need, *Advances in neural information processing systems*, Vol. 30 (2017).
- [7] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q. and Rush, A. M.: Transformers: State-of-the-Art Natural Language Processing, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Online, Association for Computational Linguistics, pp. 38–45 (online), available from <https://www.aclweb.org/anthology/2020.emnlp-demos.6> (2020).
- [8] Yan, Y., Li, R., Wang, S., Zhang, F., Wu, W. and Xu, W.: ConSERT: A Contrastive Framework for Self-

- Supervised Sentence Representation Transfer (2021).
- [9] 栗原健太郎, 河原大輔, 柴田知秀: JGLUE: 日本語言語理解ベンチマーク, 言語処理学会 第 28 回年次大会, pp. 2023-2028 (2022).
 - [10] 芝山直希, 古宮嘉那子, 新納浩幸ほか: 対照学習を利用した日本語 PromptBERT の構築, 研究報告自然言語処理 (NL), Vol. 2022, No. 3, pp. 1-6 (2022).
 - [11] 林部祐太: 知識の整理のための根拠付き自然文間含意関係コーパスの構築, 言語処理学会第 26 回年次大会論文集, pp. 820-823 (2020).

付 録

A.1 PromptBERT の実装について

本論文において、PromptBERT 及び Prompt based BERT は既存研究 [10] にて実装したプログラムを使用した。このプログラムでは、PromptBERT (Prompt based BERT) をライブラリ「SentenceTransformers[5]」の Transformer クラスを親クラスとする独自クラス Prompted-Transformer と、[MASK] トークンを抽出する独自クラス PromptPooling の 2 つを組み合わせる作成する SentenceTransformer クラスとして実装している。

A.2 実験設定の詳細

本論文内の 4 節で行った実験におけるハイパーパラメータを表 A.1 に、各シード値での評価結果を表 A.3 に示す。

表 A.1 対照学習手法の評価実験におけるハイパーパラメータ

パラメータ	値
Seed 値	503, 823, 2022, 8291240, 8301417
Epoch 数	25
バッチサイズ	32
学習率	5e-05
max_seq_length	128
評価間隔	40 steps
評価タイミング	評価間隔毎、epoch 終了時
評価用クラス	BinaryClassificationEvaluator (SentenceTransformers[5])
モデル保存	JRTE dev データで評価し、最も優秀な時点のモデルを保存
Temperature (論文 [4] の手法のみ)	0.1

5 節において、ライブラリ Huggingface Transformers[7] の BertForMaskedLM, DataCollatorForLanguageModeling, TrainingArguments, Trainer クラスを使用して空所推定モデルを構築・訓練した。モデル訓練時の各クラスの引数のうち、訓練に関わるハイパーパラメータでありデフォルトでない値を設定したものを表 A.2 に抜粋する。ここに示されていないハイパーパラメータはデフォルト値を使用した。

表 A.2 空所推論モデル構築時のハイパーパラメータの値 (変更したもののみ抜粋)

クラス	引数	値
TrainingArguments	per_device_train_batch_size	16
TrainingArguments	seed	503
TrainingArguments	data_seed	503

表 A.3 各シード値でのモデルの評価結果

モデル	Prompt	503	823	2022	8291240	8301413
Prompt based BERT	[MASK] とは、[X] である。	0.6575	0.6575	0.6575	0.6575	0.6575
	[X] とは、[MASK] である。	0.4346	0.4327	0.4321	0.4626	0.4470
	[X] は [MASK] を意味する。	0.4219	0.4419	0.4105	0.4592	0.4426
	[MASK] は [X] を意味する。	0.5950	0.5950	0.5950	0.5950	0.5949
PromptBERT (通常の対照学習)	[MASK] とは、[X] である。	0.6302	0.6601	0.6633	0.6798	0.6592
	[X] とは、[MASK] である。	0.5227	0.5350	0.5110	0.5510	0.5407
	[X] は [MASK] を意味する。	0.4629	0.4926	0.4979	0.4926	0.5227
	[MASK] は [X] を意味する。	0.6564	0.6551	0.6608	0.6528	0.6884
PromptBERT(オリジナルの手法)	[MASK] とは、[X] である。	0.6976	0.7180	0.7095	0.6944	0.7007
	[X] とは、[MASK] である。	0.6952	0.6604	0.6885	0.6677	0.6853
	[X] は [MASK] を意味する。	0.6799	0.6803	0.7012	0.6791	0.7150
	[MASK] は [X] を意味する。	0.7129	0.7137	0.7302	0.6991	0.7266