

BERTによる英文前置詞誤り修正支援の一手法

中谷 聡[†] 上野 史^{††} 太田 学^{††}

[†] 岡山大学工学部情報系学科 〒700-8530 岡山県岡山市北区津島中 3-1-1

^{††} 岡山大学学術研究院自然科学学域 〒700-8530 岡山県岡山市北区津島中 3-1-1

E-mail: [†]p2xo3pxw@s.okayama-u.ac.jp, ^{††}{uwano, ohta}@okayama-u.ac.jp

あらまし 英語の学習者が適当な英文前置詞を選択するのは難しい。本研究では、複数の自然言語処理タスクで成果を上げている自然言語処理モデルである Bidirectional Encoder Representations from Transformers (BERT) を用いて、入力された英文に含まれる英文前置詞誤りの修正候補をユーザに提示する手法を提案する。具体的には、英文中の前置詞を検出し、BERT の Masked Language Model (MLM) を利用し単語の出現確率を求め、それを利用して誤った前置詞を検出し、適切な前置詞を予測する。実験では、提案手法により前置詞誤りを自動で検出、修正し、それらの精度を評価した。誤り検出の F 値は、正解と異なる前置詞を使用する置換誤りの検出で 0.933、不要な前置詞を使用する挿入誤りの検出で 0.870、必要な前置詞が抜け落ちる欠落誤りで 0.844 だった。修正精度は、置換誤りの場合 0.929、欠落誤りの場合 0.926 だった。またオンラインの英文修正サービスである Grammarly とこれらの結果を比較した結果、提案手法は、F 値が置換誤りの検出で 32.8 ポイント、挿入誤りの検出で 13.5 ポイント Grammarly を上回った。修正精度は置換誤りの修正で 23.7 ポイント、欠落誤りの修正で 12.6 ポイント Grammarly を上回った。

キーワード BERT, 前置詞誤り, 誤り検出, 誤り修正

1 はじめに

英語を母語としない人にとって前置詞の選択は難しい。前置詞の選択に迷うような場合、通常、辞書で調べるが、期待するような例文が存在しない場合もあれば、期待するものと似た例文があってもその例文の前置詞が適切ではない場合もある。また期待する意味とは異なる前置詞も、文法的には間違っていない。そのため前置詞の意味を取り違えているような場合、書き手は修正が必要なことに気づきにくく、英文修正サービスでも修正対象にならない場合がある。そこで本稿では、文脈を考慮することが可能であり、複数の自然言語処理タスクで成果を上げている自然言語処理モデルの Bidirectional Encoder Representations from Transformers (BERT) [1] を利用した英文前置詞誤りの修正支援手法を提案する。

本研究では前置詞誤りの種類を三つ定義する。置換誤りは異なる前置詞を使用する誤り、挿入誤りは不要な箇所に余分な前置詞を使用する誤り、欠落誤りは必要な前置詞が抜け落ちる誤りである。本稿では、それぞれの前置詞誤りを検出し、修正候補を提示する手法を提案する。置換誤りでは、複数の修正候補の前置詞をユーザに提示する。挿入誤りでは、不要な前置詞を示す。欠落誤りでは、必要な箇所に必要な前置詞の候補をユーザに複数提示する。

本稿の構成は以下のとおりである。2 節では、関連研究と既存の英文修正サービスについて述べる。3 節では、提案する BERT を利用した前置詞誤りの検出と修正について述べる。4 節では、修正支援のためのインタフェースについて述べる。5 節では、提案手法の評価実験とその結果について述べる。6 節では、本稿のまとめと今後の課題について述べる。

2 関連研究

2.1 BERT

BERT [1] は自然言語処理における事前学習モデルの一つで、複数の自然言語処理タスクで成果を上げ注目された。図 1 に BERT の概略図を示す。BERT に英文を入力すると単語に分割され、単語がトークンとして入力される。例えば “I live in Okayama.” では “I”, “live”, “in”, “Okayama”, “.” に分割され、五つのトークンが BERT の入力になる。図 1 では E_i が入力を表し、 T_i が出力を表している。Trm は、attention を用いた機械学習モデルの transformer [2] である。従来の言語モデルの一つである ELMo [3] では left-to-right モデルと right-to-left モデルをそれぞれ独立に動かし、それぞれの出力を足し合わせることで双方向の読み込みを実現していた。一方 BERT は同一モデル内で注目する単語の周囲の文脈を双方向から学習し、前後の関係性を考慮することができる。BERT では事前学習の際に、一部の単語をマスクし、その単語を予測する Masked Language Model (MLM) と、2 文を入力として受け取り、後の文が前の文に続く文か否かを予測する Next Sentence Prediction (NSP) を用いて学習する。事前学習をすることで、入力シーケンスをそのまま双方向から読み込み学習の際に予測すべき単語が先読みされることを防いでいる。

MLM では入力シーケンスの単語の 15% を [MASK] トークンに置き換える。[MASK] トークンに置き換えられた単語を、マスクされなかった単語によって与えられる文脈に基づいて予測する。例えば以下の文 (1) の場合、文中から 15% の確率で単語を選ぶ。選ばれた単語が “in” と “brother” の場合、それらをマスクした文 (2) を作成する。文 (2) の [MASK] トー

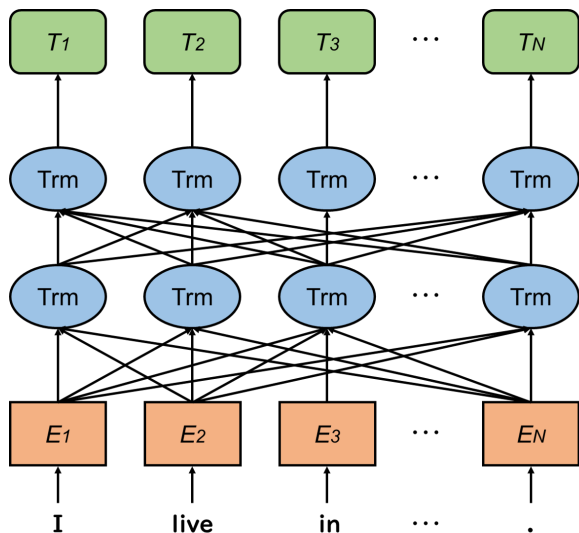


図1 BERTの概略図[1]

クンを推測する。

(1) I live in Okayama with my brother since 2022.

(2) I live [MASK] Okayama with my [MASK] since 2022.

2.2 英文前置詞誤りの修正

有富ら[4]は、検索エンジンにより英文前置詞誤りを自動検出して修正するシステムを提案した。前置詞を含む英文を入力として与えると、システムが前置詞の前後の句からなる検索フレーズを自動生成して検索する。そして検索結果から前置詞の出現確率を計算し、結果をユーザーに提示する。

三田[5]は、同じ前置詞でも“be on the table”では「場所」，“meet on Monday”では「時」のように、前置詞前後の単語による潜在的な意味関係があることに注目し、前置詞の潜在的意味関係を表現する Semantic Relation を推定した。そして、その Semantic Relation を前置詞誤り訂正のための新たな素性として利用する前置詞意味解析モデルを提案し、CoNLL2014 Shared Taskの評価セットで10分割交差検定による評価実験を行った。その結果、潜在的意味関係を利用するモデルの方が利用しないモデルと比べて前置詞の誤り訂正の適合率が0.906から0.1019に、再現率が0.1386から0.1610に、F値が0.1095から0.1248に、高くなることを確認した。

2.3 BERTを用いた英文誤りの修正

小田[6]は、BERTを用いた英文前置詞の置換誤りの修正法を提案し、さらにそれを改良した。まず、トークンの品詞が付加されたアノテーション付きコーパスのBritish National Corpusで転移学習し、前置詞を[MASK]トークンとし、[MASK]トークンに入る前置詞を予測した。次に第二言語学習者の誤りに多い前置詞“in”、“of”の誤り修正精度の向上のため、“in”、“of”、その他の前置詞の3分類問題を学習した。その結果、転移学習をした方法と比べ、3分類問題を学習した方法は、第二言語学習者コーパスW&Iで10種類の前置詞に対する誤り訂正の

適合率が“to”では0.679から0.700に、“of”では0.667から0.687に高くなることを確認した。

また誤りの修正でないが、文章の空所に当てはまる単語や文を予測するタスクの一つである空所補充問題は、機械学習モデルの評価に用いられるタスクである。三木ら[7]は、BERTのMLMを利用し、英文空所補充問題の解答を四つの選択肢の中から予測する問題の解法を提案した。三木らの提案手法は、2000年度から2016年度の大学入試センター試験の問題のうち、空所が1か所のみの問題を使った評価実験において、92.3%の正答率を示した。

2.4 Grammarly

Grammarly[8]は、入力した英文の様々な項目の修正が行えるオンラインサービスである。英文を入力すると、修正が必要と予想される箇所が修正の種類によって色分けされて表示される。色分けされた箇所をクリックすると修正候補が表示され、修正候補をクリックすることで文章の修正が行える。修正できる内容は、Grammarlyの無料版で文法、スペルミス、句読点の位置であり、プレミアム版ではそれらに加えて可読性、より適切な単語の選択、盗用チェックなどがある。他のアプリとの連携も可能であり、ユーザはメールを書いたりSNSでメッセージを打ったりする際に修正の提案を受けることができる。

3 BERTを利用した前置詞誤りの検出と修正

ここではまず3.1節で置換、挿入、欠落の三種類の前置詞誤りについて説明する。つづく3.2節は提案する前置詞誤りの検出と修正の概要について述べる。3.3節ではBERTによる置換誤りと挿入誤りの検出のために、前置詞をマスクした文の作成方法について述べる。3.4節ではBERTによる欠落誤りの検出のために、[MASK]トークンを挿入した文の作成方法について述べる。3.5節ではBERTによるMLMを利用した[MASK]トークンに入る単語の出現確率の予測について述べる。3.6節ではBERTの出力による前置詞誤りの自動検出について述べる。3.7節ではBERTの出力を利用した前置詞誤りの修正候補の提示について述べる。

3.1 前置詞誤りの種類

本研究では、以下の三種類の前置詞誤りの検出と修正を行う。

・置換誤り

異なる前置詞を使用する誤り

誤り： I play tennis **in** Sunday.

正解： I play tennis **on** Sunday.

・挿入誤り

不要な前置詞を使用する誤り

誤り： They discussed **about** the new coronavirus.

正解： They discussed the new coronavirus.

・欠落誤り

必要な前置詞が抜け落ちる誤り

誤り： I often take care my brother.

正解： I often take care **of** my brother.

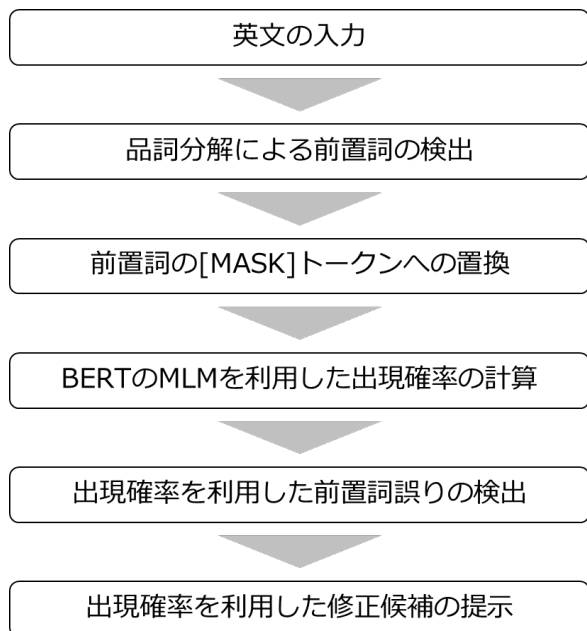


図 2 置換誤りと挿入誤りの修正支援の概要

3.2 概 要

BERT による前置詞誤りの検出と修正の流れを図 2 と図 3 に示す。図 2 が置換誤りと挿入誤り、図 3 が欠落誤りの修正支援の概要である。図 2 の置換誤りと挿入誤りの検出と修正では、まず入力された英文を品詞分解する。次に前置詞を一つずつ [MASK] トークンに置き換えた英文を作成する。次に BERT の MLM を利用し、[MASK] トークンに入る単語の出現確率を求める。次に単語の出現確率を基に、前置詞が誤りかどうかを検出する。最後に [MASK] トークンに入る単語の出現確率を基に、修正候補を提示する。

一方、図 3 の欠落誤りの検出と修正では、まず入力された英文を品詞分解する。そしてその英文の一部の単語を除くすべての単語の直前に [MASK] トークンを挿入した英文を作成する。この挿入場所については 3.4 節で詳しく述べる。次に MLM を利用し、[MASK] トークンに入る単語の出現確率を求める。次に単語の出現確率を基に、欠落誤りを検出する。また単語の出現確率を基に、挿入すべき前置詞を提示する。

3.3 置換誤りと挿入誤りの検出のためのマスクした文の生成

BERT の MLM を利用するために、入力文の前置詞を [MASK] トークンに置き換える。まず、入力された英文を、Python の自然言語処理のライブラリである NLTK [9] を用いて品詞分解する。NLTK で用いられる品詞タグセット Penn Treebank P.O.S Tags [10] では、前置詞を含む品詞タグは IN と TO の 2 種類がある。NLTK で品詞分解すると、入力された英文の単語とその単語に対応する品詞タグが出力として得られ、IN には前置詞と従属接続詞、TO には単語 “to” が含まれる。この単語 “to” は前置詞としてだけでなく不定詞としても用いられる。本研究では、前置詞のみを検出するため、品詞タグ IN については、前置詞の用法ではない従属接続詞を検出対象から除外し、品詞タグ TO については、to の次に動詞の原形

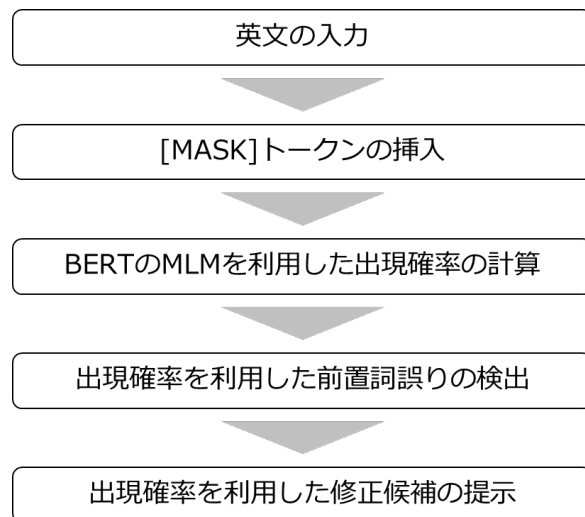


図 3 欠落誤りの修正支援の概要

が来るとき、検出対象から除外する。

次に、検出した前置詞を元に、入力英文の一部をマスクした文を生成する。これは、各前置詞を [MASK] トークンに置き換えて生成するため、英文が前置詞の数だけ生成される。例えば “I visit to Okayama on 2021.” という文からは、“to” と “on” をそれぞれマスクした以下の 2 文が作成される。

- (1) I visit [MASK] Okayama on 2021.
- (2) I visit to Okayama [MASK] 2021.

3.4 欠落誤り検出のためのマスクした文の生成

BERT の MLM を利用するために、入力文に [MASK] トークンを挿入する。ここでは [MASK] トークンが挿入された英文が、挿入する [MASK] トークンの数だけ生成される。この [MASK] トークンは不定詞 to、動詞の原形、記号、助動詞、等位接続詞、間投詞の直前と限定詞、助動詞の直後には挿入しないが、その他の語の直前には挿入する。例えば “I should study English.” という文からは以下の 2 文が生成される。

- (1) [MASK] I should study English.
- (2) I should study [MASK] English.

3.5 BERT による [MASK] トークンに入る単語の出現確率の予測

BERT の出力として、[MASK] トークンに入る単語とその出現確率を得る。本研究では、前置詞以外の単語も含めた出現確率の大きい上位 5 単語とその出現確率を求める。この上位 5 単語に前置詞がないときは Cambridge Dictionary [11] に主要な前置詞として記載されている 54 の前置詞の出現確率も求める。

3.6 前置詞誤りの自動検出

3.5 節で得られた単語とその出現確率を基に、前置詞誤りを検出する。三種類の誤りについての判定基準は以下の通りである。

● 置換誤りの検出

元の前置詞の出現確率が 0.1 未満かつ元の前置詞とは異なる前置詞の出現確率が 0.1 以上の場合、元の前置詞を置換誤りとし

て検出する。

- 挿入誤りの検出

出現確率上位 5 単語に含まれる前置詞の出現確率の合計が 0.2 以下かつ最も出現確率の高い前置詞の出現確率が 0.1 未満の場合、元の前置詞を挿入誤りとして検出する。

- 欠落誤りの検出

出現確率が 0.3 以上の前置詞があり、かつ出現確率上位 5 単語に含まれる前置詞の出現確率の合計が 0.9 以上の場合、その [MASK] トークンの箇所に欠落誤りがあるとする。

置換誤りと挿入誤りの検出では、判定結果は置換誤り、挿入誤り、誤り無しのいずれかになる。なお置換誤りの前置詞が挿入誤りと判定されることや、挿入誤りの前置詞が置換誤りと判定されることがある。誤り無しは非検出となる。欠落誤りの検出の場合、判定結果は欠落誤りか誤り無しどちらかになる。誤り無しは非検出となる。

3.7 前置詞誤りの修正候補の提示

3.5 節で得られた単語の出現確率を基に、修正候補を提示する。具体的には以下のようにする。

- 置換誤りの修正候補の提示

出現確率が 0.1 以上の前置詞を、大きいものから最大三つ提示する

- 欠落誤りの修正候補の提示

出現確率が 0.1 以上の前置詞を、大きいものから最大三つ提示する

なお挿入誤りは検出すれば、その前置詞の削除を促すため、修正候補は提示しない。

4 インタフェース

図 4 に実装した英文前置詞誤り修正支援プログラムのインタフェースの画面を示す。なお図 4 は置換誤りの修正の様子を示している。この図の最上部のボックスが英文を入力する部分である。その下のボックスには、元の英文で前置詞の修正が必要と判定された箇所を、大括弧に置き換えた英文が表示される。修正が不要と判定されたときは、「入力された英文に前置詞誤りはありません」と表示される。その下には、修正候補の前置詞が出現確率の高い順に評価とともに表示される。この評価の基準は表 1 に示す。その横のボックスには、置換誤りまたは欠落誤りと判定された場合は、修正候補の前置詞で修正した英文が表示される。挿入誤りの場合は、その前置詞を削除した英文が表示される。

左下の「実行」は前置詞誤りの検出を開始するボタンである。右下の「次の修正箇所を表示」は前置詞の修正箇所が複数あるとき、すべての前置詞誤りについて次の修正箇所と修正候補を表示するボタンである。右下の「修正箇所を最初から表示」を押すと、「次の修正候補を表示」ボタンで次の修正箇所を表示した後でも、最初の修正箇所を表示する。

図 4 は、「I study the history at Okayama.」という英文を入力したときの実行結果の画面である。この文では前置詞 “at”

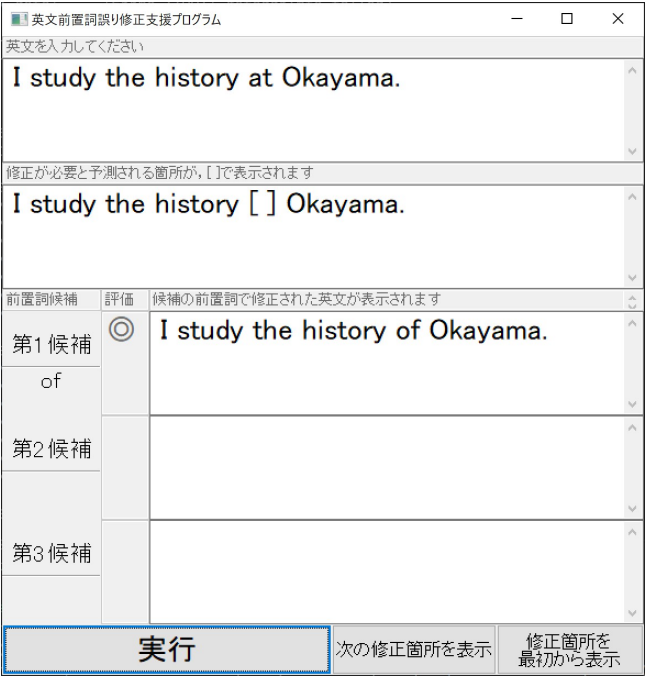


図 4 置換誤り修正の実行結果の画面

表 1 表示する修正候補の評価とその基準

評価	基準
◎	置換誤りまたは欠落誤りと判定された場合で、修正候補の内の出現確率が 0.5 以上の前置詞
○	置換誤りまたは欠落誤りと判定された場合で、修正候補の内の出現確率が 0.2 以上 0.5 未満の前置詞
△	置換誤りまたは欠落誤りと判定された場合で、修正候補の内の出現確率が 0.1 以上 0.2 未満の前置詞
-	挿入誤りと判定された前置詞

が誤りとして検出され、前置詞 “of” が修正候補として提示されている。

5 評価実験

ここでは提案手法に対する評価実験について述べる。5.1 節では実験結果で用いる評価指標について説明する。5.2 節で提案手法の誤り検出の適切性を評価する。5.3 節で誤りの検出と修正の精度を評価する。5.4 節と 5.5 節では、検出と修正の精度について提案手法と Grammarly を比較をする。5.6 節では、5.2 節から 5.5 節で述べた実験の結果に基づいて、提案手法について考察する。

5.1 評価指標

前置詞誤り検出の評価指標として、再現率 (Recall)、適合率 (Precision)、再現率と適合率の調和平均の F 値を用いる。これらの指標は以下の式で表される。

$$\text{再現率 (Recall)} = \frac{TP}{TP + FN} \quad (1)$$

$$\text{適合率 (Precision)} = \frac{TP}{TP + FP} \quad (2)$$

$$F \text{ 値} = \frac{2 \cdot \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} \quad (3)$$

なお TP , FN , FP は以下の通りである。

- TP : 正しく前置詞誤りと判定した前置詞誤り
- FN : 誤って前置詞誤りと判定しなかった前置詞誤り
- FP : 誤って前置詞誤りと判定したが誤りでない前置詞

また、置換誤りと欠落誤りの修正の評価指標として、以下の修正精度を用いる。

$$\text{修正精度} = \frac{\text{正しく誤りを修正できた前置詞の数}}{\text{誤りとして正しく検出された前置詞の数}} \quad (4)$$

なお挿入誤りは、正しく検出すれば修正できるため、検出で評価し修正では評価しない。

5.2 前置詞誤り検出の適切性の評価

実装したインタフェースに前置詞を含む誤りのない英文を入力として与え、前置詞誤りを何も検出しないか実験により確かめる。テストデータに三つの書籍 [13] [14] [15] から集めた計 100 文を使用し、この中には 232 個の前置詞が含まれる。置換誤りと挿入誤りの検出では、これら 232 個の前置詞を [MASK] トークンに置換した 232 文を評価に用いる。欠落誤りの検出では、3.4 節で説明した方法で、元の 100 文に対し、欠落誤り検出のための [MASK] トークンを挿入した英文を生成し、得られた 1263 文を評価に用いる。評価は、置換誤りと挿入誤りの検出では、それぞれ前置詞の置換や削除が必要ないと判定されれば正解である。欠落誤りの検出では、前置詞の挿入が不要と判定されれば正解である。

表 2 に、誤りのない前置詞を [MASK] トークンに置換した 232 文に対する置換誤りと挿入誤りの検出結果を、表 3 に、誤りのない箇所に [MASK] トークンを挿入した 1263 文に対する欠落誤りの検出結果を示す。表 2 から、置換誤りで 0.914 の正解率、挿入誤りで 0.970 の正解率が得られた。表 3 から、欠落誤りで 0.988 の正解率が得られた。

5.3 前置詞誤りの検出と修正の精度評価

一定割合の前置詞誤りを含む英文を提案手法に与え、その前置詞誤りを検出して、修正できるかを実験により評価する。テストデータは 5.2 節と同じ 100 文を用いる。ここでは、この 100 文へそれぞれの種類の前置詞誤りを加えた以下の各 100 文をインタフェースへの入力とする。

- 置換誤りの評価のための 100 文

元の 100 文に含まれる前置詞 232 の内、無作為に選んだ 116 を他の前置詞に変更したもの。

- 挿入誤りの評価のための 100 文

元の 100 文それぞれで、無作為に選んだ前置詞を無作為に選んだ位置に一つ挿入したもの。

- 欠落誤りの評価のための 100 文

元の 100 文それぞれで、無作為に選んだ前置詞を一つ削除したもの。

なお、欠落誤りの評価では、前置詞を一つ削除して生成した 100 文で評価するが、3.4 節で説明したように、提案手法は

表 2 誤りのない英文に対する置換誤りと挿入誤りの検出結果

	誤りのない前置詞 (232)		
	非検出	誤検出	正解率
置換誤りの検出	212	20	0.914
挿入誤りの検出	225	7	0.970

表 3 誤りのない英文に対する欠落誤りの検出結果

	前置詞が不要な箇所 (1263)		
	非検出	誤検出	正解率
欠落誤りの検出	1248	15	0.988

[MASK] トークンを挿入した文を複数生成して BERT へ入力する。これは、欠落誤り検出のために [MASK] トークンを挿入する箇所が一つの入力文に対して複数あるためである。本実験では、合計 1163 文が生成された。

提案手法を利用した英文前置詞誤りの検出と修正の正誤判定は以下を行う。

- 検出の正誤判定

入力文の誤り箇所では修正が提示されたとき、前置詞誤りを検出したとする。置換誤りの箇所を挿入誤りと検出したり、挿入誤りの箇所を置換誤りと検出したりしたときも、検出したと判定する。欠落誤りの検出では、欠落誤り以外の箇所では欠落誤りが提示されたときは誤検出となる。なお、挿入誤りの評価のための 100 文で、挿入誤りを検出すれば、修正もできることになる。

- 修正の正誤判定

置換誤りと欠落誤りの修正の場合は、提案手法が提示する修正候補の中に正解前置詞が含まれていた場合に修正できたと判定する。

表 4 に、置換誤り評価用 100 文を使った実験における、置換誤りの 116 の前置詞と誤りのない 116 の前置詞に対する置換および挿入誤りの自動検出結果を示す。

表 5 に、挿入誤り評価用 100 文を使った実験における、挿入誤りの 100 の前置詞と誤りのない 232 の前置詞に対する挿入および置換誤りの自動検出結果を示す。

表 6 に、欠落誤り評価用 100 文に欠落誤り検出のための [MASK] トークンを挿入することで生成された 1163 文に対する、欠落誤り検出結果を示す。元の英文から欠落誤りとして削除した 100 の前置詞の箇所では、欠落誤りを検出するとそれは正しい検出で、誤りのない 1063 の箇所では欠落誤りを検出するとそれは誤検出となる。

表 7 に、それぞれの前置詞誤りの検出性能をまとめる。表 7 に示すように、すべての種類の前置詞誤り検出の結果で、再現率は 0.8 以上だった。一方、適合率は 0.9 より低く、特に挿入誤りでは 0.7 より低かった。

表 8 に、欠落誤り評価用 100 文に対する欠落誤り検出結果を示す。表 8 の数値は文の数であり、合計すると 100 になる。表 8 より、誤りのない箇所では欠落誤りを検出した英文は、誤検出を合計した値になるため、100 文のうち 15 文である。

表 9 に、前置詞誤りとして検出できた前置詞に対して、修正を試みた結果を示す。置換誤りと欠落誤りで 0.8 以上の修正精

表 4 置換誤り評価用 100 文に対する置換および挿入誤りの検出結果

誤っている前置詞 (116)		誤りのない前置詞 (116)			
検出		非検出	非検出	誤検出	
置換誤り	挿入誤り			置換誤り	挿入誤り
110	4	2	95	15	6

表 5 挿入誤り評価用 100 文に対する挿入および置換誤りの検出結果

挿入した前置詞 (100)		元の前置詞 (232)			
検出		非検出	非検出	誤検出	
挿入誤り	置換誤り			置換誤り	挿入誤り
94	6	0	181	18	33

表 6 欠落誤り評価用 100 文に [MASK] トークンを挿入し生成された 1163 文に対する欠落誤りの検出結果

削除した前置詞 (100)		前置詞が不要な箇所 (1063)	
検出	非検出	非検出	誤検出
80	20	1048	15

表 7 前置詞誤りの検出性能

	再現率	適合率	F 値
置換誤り評価用データ	0.983	0.844	0.908
挿入誤り評価用データ	1.000	0.662	0.797
欠落誤り評価用データ	0.800	0.842	0.821

表 8 欠落誤り評価用 100 文に対する欠落誤りの検出結果 (数値は文の数)

		前置詞が不要な箇所	
		非検出	誤検出
前置詞が必要な箇所 (1 文につき 1 箇所)	検出	69	11
	非検出	16	4

表 9 検出した置換誤りと欠落誤りに対する修正性能

	修正できた	修正できなかった	修正精度
置換誤り	100	14	0.877
欠落誤り	74	6	0.925

度を示した。

5.4 Grammarly との検出と修正の性能比較

Grammarly は、英文を入力すると自動で英文の間違いを検出し、修正候補を提示する。そこで、Grammarly と提案手法で前置詞誤りの検出と修正についてその精度を比較する。Grammarly は、より多くの修正候補を提示するプレミアム版を使用する。テストデータは 5.3 節でを使用したものと同じ、三種類の前置詞誤り評価用の各 100 文である。Grammarly の自動検出、自動修正の正誤判定は以下に行う。

- 自動検出の正誤判定

入力文の前置詞の誤り箇所で見つけた修正候補が提示されたとき、前置詞誤りを検出したとする。それ以外の箇所で見つけた前置詞の修正が提示されたときは誤検出となる。挿入誤りの場合は、挿入した前置詞の削除を提示されたとき、検出できたとする。

- 自動修正の正誤判定

Grammarly が提示する誤り箇所の修正候補中に元の英文の前置詞（正解）がある場合、修正できたとする。

表 10 置換誤り評価用と挿入誤り評価用の各 100 文に対する置換誤りと挿入誤りの検出性能の比較

		再現率	適合率	F 値
置換誤り 評価用データ	Grammarly	0.526	0.984	0.685
	提案手法	0.983	0.844	0.908
挿入誤り 評価用データ	Grammarly	0.630	0.940	0.754
	提案手法	1.000	0.662	0.797

表 11 欠落誤り評価用 100 文に対する欠落誤りの検出結果 (Grammarly | 提案手法)

		前置詞が不要な箇所	
		非検出	誤検出
前置詞が必要な箇所 (1 文につき 1 箇所)	検出	72 69	0 11
	非検出	26 16	2 4

表 12 検出した置換誤りと欠落誤りに対する修正性能の比較

		修正できた	修正できない	修正精度
置換誤り	Grammarly	52	9	0.852
	提案手法	100	14	0.877
欠落誤り	Grammarly	59	13	0.819
	提案手法	74	6	0.925

表 10 に、置換誤り評価用と挿入誤り評価用の各 100 文に対する、提案手法と Grammarly の置換誤りと挿入誤りの自動検出性能を示す。表 10 に示すように、提案手法と Grammarly を比較すると、置換誤り評価用 100 文では、提案手法が再現率は 45.7 ポイント、F 値は 22.3 ポイント上回っている。一方、適合率は Grammarly が 14.0 ポイント上回っている。挿入誤り評価用 100 文では、再現率は提案手法が 37.0 ポイント上回っている。一方、適合率は Grammarly が 27.8 ポイント上回っている。

表 11 に、欠落誤り評価用 100 文に対する提案手法と Grammarly の欠落誤りの自動検出の結果を示す。表 11 の数値は表 8 と同じ文の数である。表 11 から、提案手法と Grammarly を比較すると、欠落誤りがある箇所で見つけた欠落誤りを検出した文は、Grammarly が 72 文に対して提案手法は 80 文であり、提案手法の方が欠落誤りを多く検出した。一方、誤りのない箇所で見つけた欠落誤りを誤検出した文は、Grammarly の 2 文に対して提案手法は 15 文であるため、Grammarly の方が誤りのない箇所で見つけた欠落誤りを検出することは少ない。

表 12 に、提案手法と Grammarly の置換誤りと欠落誤りの修正性能の比較を示す。表 12 に示すように、提案手法が、置換誤りの修正精度で 2.5 ポイント、欠落誤りの修正精度で 10.6 ポイント Grammarly を上回っている。

5.5 TOEIC®参考書の問題による Grammarly との検出と修正の精度比較

ここでは TOEIC®の文法参考書 [12] の問題を用いて 5.3 節と 5.4 節と同様の実験を行い、Grammarly と提案手法で前置詞誤りの検出と修正の性能を比較する。

実験に使用するデータは、[12] の書籍の空所補充問題の空所を補充した文の内、元の空所または空所以外の箇所に見つけた前置詞を含む無作為に選んだ 30 文である。この 30 文には 58 の前置詞

が含まれている。実験では、三種類の誤りの対する自動検出と修正の性能を評価するため、以下のような各種類の誤りを含む英文を生成して入力とした。

- 置換誤りの評価のための 30 文

30 文に含まれる前置詞 58 の内、無作為に選んだ 29 を他の前置詞に変更したもの。

- 挿入誤りの評価のための 30 文

30 文それぞれで、無作為に選んだ前置詞を無作為に選んだ位置に一つ挿入したもの。

- 欠落誤りの評価のための 30 文

30 文それぞれで、無作為に選んだ前置詞を一つ削除したもの。

なお、提案手法による欠落誤り検出では、前置詞を一つ削除して生成した 30 文に対し、3.4 節で示した箇所に [MASK] トークンを挿入することで 314 文が生成された。

表 13 に、TOEIC®参考書の問題文の置換誤り評価用 30 文を使った実験における、置換誤りの 29 の前置詞と誤りのない 29 の前置詞に対する置換および挿入誤りの自動検出結果を示す。

表 14 に、TOEIC®参考書の問題文の挿入誤り評価用 30 文を使った実験における、挿入誤りの 30 の前置詞と誤りのない 58 の前置詞に対する挿入および置換誤りの自動検出結果を示す。

表 15 に、TOEIC®参考書の問題文の欠落誤り評価用 30 文に、欠落誤り検出のための [MASK] トークンを挿入することで生成された 314 文に対する欠落誤りの検出結果を示す。欠落誤りとして削除した 30 の前置詞の箇所で、欠落誤りを検出するとそれは正しい検出で、誤りのない 284 の箇所で欠落誤りを検出するとそれは誤検出となる。

表 16 に、TOEIC®参考書の問題文から作成した置換誤り評価用と挿入誤り評価用の各 30 文に対する、提案手法と Grammarly の置換誤りと挿入誤りの自動検出性能を示す。表 16 に示すように、提案手法と Grammarly を比較すると、置換誤りでは、提案手法が再現率は 53.3 ポイント、F 値は 32.8 ポイント上回っている。一方、適合率は Grammarly が 9.7 ポイント上回っている。挿入誤りでは、提案手法が再現率は 40.0 ポイント、F 値は 13.5 ポイント上回っている。一方、適合率は Grammarly が 17.8 ポイント上回っている。

表 17 に、TOEIC®参考書の問題文の、欠落誤り評価用 30 文に対する提案手法と Grammarly の欠落誤りの自動検出結果を示す。欠落誤りがある箇所で欠落誤りを検出した文は、Grammarly が 25 文に対して提案手法は 27 文であり、提案手法の方が欠落誤りを多く検出した。一方、誤りのない箇所で欠落誤りを誤検出した文は、Grammarly の 2 文に対して提案手法は 6 文であり、Grammarly の方が欠落誤りを誤検出することは少ない。

表 18 に、TOEIC®参考書の問題文の、置換誤り評価用と欠落誤り評価用の各 30 文に対する、提案手法と Grammarly の置換誤りと欠落誤りの修正性能の比較を示す。表 18 に示すように、提案手法が、置換誤りの修正精度で 23.7 ポイント、欠落誤りの修正精度で 12.6 ポイント上回っている。

表 13 TOEIC®参考書の問題文の置換誤り評価用 30 文に対する置換および挿入誤りの検出結果

誤っている前置詞 (29)		誤りのない前置詞 (29)			
検出		非検出	非検出	誤検出	
置換誤り	挿入誤り			置換誤り	挿入誤り
28	0	1	26	3	0

表 14 TOEIC®参考書の問題文の挿入誤り評価用 30 文に対する挿入および置換誤りの検出結果

挿入した前置詞 (30)		元の前置詞 (58)			
検出		非検出	非検出	誤検出	
挿入誤り	置換誤り			置換誤り	挿入誤り
27	3	0	49	6	3

表 15 TOEIC®参考書の問題文の欠落誤り評価用 30 文に対する欠落誤りの検出結果

削除した前置詞 (30)		前置詞が必要ない箇所 (284)	
検出	非検出	非検出	誤検出
27	3	277	7

表 16 TOEIC®参考書の問題文の置換誤りと挿入誤り評価用の各 30 文に対する置換誤りと挿入誤りの検出性能の比較

		再現率	適合率	F 値
置換誤り 評価用データ	Grammarly	0.433	1.000	0.605
	提案手法	0.966	0.903	0.933
挿入誤り 評価用データ	Grammarly	0.600	0.947	0.735
	提案手法	1.000	0.769	0.870

表 17 TOEIC®参考書の問題文の欠落誤り評価用 30 文に対する欠落誤りの検出性能 (Grammarly | 提案手法)

		前置詞が不要な箇所	
		非検出	誤検出
前置詞が必要な箇所 (1 文につき 1 箇所)	検出	24 22	1 5
	非検出	4 2	1 1

表 18 TOEIC®参考書の問題文の評価用 30 文において検出した置換誤りと欠落誤りに対する修正性能の比較

		修正できた	修正できない	修正精度
置換誤り	Grammarly	9	4	0.692
	提案手法	26	2	0.929
欠落誤り	Grammarly	20	5	0.800
	提案手法	25	2	0.926

5.6 考 察

5.6.1 前置詞誤り検出の適切性

5.2 節の実験で、誤りのない箇所が前置詞誤りと判定された理由として以下の原因が挙げられる。誤りのない文に対する置換誤りの検出では、元の前置詞以外で出現確率の高い別の前置詞があることが挙げられる。例えば、“The costs of special programs designed to bring out the humanity in prisoners are too high for implementation throughout the nation.” という誤りのない文の “in” を [MASK] トークンに置換して、出現確率を求めると、“in” の出現確率は 0.004 で二番目に高いが “of” の出現確率は 0.992 であるため、これを置換誤りと誤判

定する。誤りのない文に対する挿入誤りの検出では、前置詞の箇所における前置詞以外の品詞の単語の出現確率が高く、前置詞の出現確率が低いことが挙げられる。例えば、“Most smart phones have tools for basic tasks as well as for multimedia functions.”という誤りのない文の二つ目の“for”を[MASK]トークンに置換して、出現確率を求めると、“basic”が0.158, “some”が0.118, “other”が0.100, になる。“for”の出現確率は0.083になり四番目に大きい、挿入誤りの検出基準の値0.1より小さい。そのため提案手法はこれを挿入誤りと誤判定する。誤りのない文に対する欠落誤りの検出では、前置詞の意味も持つ前置詞とは異なる品詞の単語が欠落していると判定し、挿入を指示されることがある。例えば、“The man who shaped some of the greatest technological innovations of our time never graduated from college.”という誤りのない文の“shaped”と“some”の間に[MASK]トークンを挿入した文に対し、前置詞の意味も持つ副詞の“up”が欠落していると提案手法が誤検出する場合がある。このような間違いは、修正後の文を形態素解析し、修正した単語が前置詞か調べることで解決できる可能性がある。

5.6.2 検出と修正の精度

前置詞誤りの種類とは関係なく提案手法が誤検出する例がある。それは前置詞誤りの箇所以外の前置詞誤りを検出しようとするとき、前置詞誤りの箇所の影響を受ける場合である。例えば、“Air pollution creates holes in without the protective ozone layer of the stratosphere, thereby allowing harmful ultraviolet radiation to reach Earth’s surface.”という文では“without”が挿入誤りの前置詞である。しかし、この文を入力すると、“in”が挿入誤り、“without”が置換誤りの修正候補として検出された。

提案手法による置換誤りと欠落誤りの修正で、誤りを適切に修正できなかった原因として、自動修正の判定基準の妥当性の問題がある。修正候補は出現確率の高い候補を最大三つ提示するが、四番目に出現確率が高い前置詞が正解である例があった。

6 ま と め

本稿では、BERTにより英文前置詞誤りの修正を支援する手法を提案した。提案手法では、前置詞の置換誤りと挿入誤りの検出のために、まず入力文中の前置詞を[MASK]トークンに置換する。また欠落誤りの検出のために一部の単語を除くすべての単語の直前に[MASK]トークンを挿入する。次にそれら[MASK]トークンを含む文をBERTへの入力とし、BERTのMLMを利用して[MASK]トークンに対応する単語の出現確率を求める。そしてその出現確率を利用して前置詞誤りを検出し、修正候補を提示する。本研究ではまた、英文を入力すると前置詞誤りを自動で検出し、前置詞の修正候補を提示するインタフェースを実装した。

実験では、提案手法の性能を評価するため、提案手法の前置詞誤りの検出の適切性と、前置詞誤りの検出と修正の性能を確かめた。さらに、提案手法と既存の英文修正サービスである

Grammarlyと検出と修正の性能を比較した。その結果、提案手法は、F値が置換誤りの検出で0.908, 挿入誤りの検出で0.797となり、それぞれGrammarlyを22.3ポイント, 4.3ポイント上回った。修正精度は置換誤りで0.877, 欠落誤りで0.925となり、それぞれGrammarlyを2.5ポイント, 10.6ポイント上回った。一方、誤り検出の適合率は置換誤りで0.844, 挿入誤りで0.662となり、それぞれGrammarlyを14.0ポイント, 27.8ポイント下回った。評価実験に用いる英文を変えて同様の実験を行い、提案手法とGrammarlyを比較しても同様の結果が得られた。前置詞誤りの検出と修正に関して、提案手法は全般的にGrammarlyを上回ったが、検出の適合率のみGrammarlyを下回った。そのため、誤検出の削減が今後の課題と考えている。

文 献

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” arXiv:1810.04805, 2018.
- [2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser and Illia Polosukhin, “Attention Is All You Need,” In Advances in Neural Information Processing Systems, Vol. 30, pp. 5998–6008, 2017.
- [3] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. arXiv:1810.04805, 2018.
- [4] 有富隼, 太田学: “検索エンジンによる英文前置詞誤り修正支援,” DBSJ Journal Vol. 9, No. 1, pp. 70–75, 2010.
- [5] 三田雅人, “文脈を考慮した前置詞誤り訂正のための潜在意味関係の利用,” 修士論文 NAIST-IS-MT1451103, 2016.
- [6] 小田幹雄, 分類器による英文前置詞誤り訂正の学習法. 言語処理学会第26回年次大会, pp. 97–100, 2020.
- [7] 三木一弘, 太田学, “BERTを用いた英文空所補充問題の一解法,” DEIM2020, G2-4, 2020.
- [8] Grammarly, <https://app.grammarly.com/>
- [9] Bird, Steven, Edward Loper and Ewan Klein (2009), *Natural Language Processing with Python*. O’Reilly Media Inc.
- [10] Penn Treebank P.O.S Tags, https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn-treebank-pos.html
- [11] Cambridge Dictionary, <https://dictionary.cambridge.org/ja/grammar/british-grammar/prepositions>
- [12] 花田徹也, “1 駅 1 題 TOEIC®L&R TEST 文法特急,” 朝日新聞出版, 2021.
- [13] 金谷健一, “理数系のための技術英語練習帳,” 共立出版, 2018.
- [14] Judy L. Miller, Robert F. Cohen, “Longman Academic Reading Series 3: Reading Skills for College,” Pearson Education Inc., 2017.
- [15] Alice Oshima, Ann Hogue, “Longman Academic Writing Series 4: Paragraphs to Essays, Fifth Edition,” Pearson Education Inc., 2017.