

分類器による英文前置詞誤り訂正の学習法

小田 幹雄

久留米工業高等専門学校 制御情報工学科

oda@kurume-nct.ac.jp

1 はじめに

第二言語習得の学習者 (L2 学習者) にとって、母語である第一言語と第二言語の語族が異なれば、文法、語彙、文化を背景とした語義が異なり、言語習得に苦勞と時間を要する [9]。日本人が習得する第二言語の割合は、その多くを英語が占めるが、日本語と英語は、文法、語義、語彙のすべての点で大きく異なり、日本人英語学習者の苦勞は、容易に推察できる。

第二言語の学習効果を向上させるためには、自然言語処理技術を応用し、教育的な観点から学習を支援することが重要である。自然言語処理技術として、コーパスを用いた深層学習を用いる手法が、近年盛んに研究されている [11]。L2 学習者支援のための自然言語処理の目的は、文章生成支援または作成した文章の誤り検出/訂正であり、文章誤り検出/訂正には、綴り、連語、文法及び語義の誤り検出/訂正が含まれる。この中で、連語誤りと文法誤りの支援が、日本人の英語学習者のために特に求められている。なお、これらの技術は、学習支援だけでなく、小論文の自動得点システムの開発にも利用されている [2]。

本稿では、L2 学習者支援システムの開発に向けて、日本人が文章を作成するとき、特に間違いが起りやすい前置詞の誤り訂正の精度向上を目的とする。Bidirectional Encoder Representations from Transformers (BERT) を前置詞の誤り訂正分類器として用いたとき、コーパス学習の改良法を提案し、その推定精度を考察する。

2 先行技術

文法誤り訂正の研究は、Shared Task と呼ばれる競技により、大きく発展してきた。主な Shared Task として、前置詞と冠詞を扱った HOO-2012 Shared Task、文法誤り全体を扱った CoNLL-2014 Shared Task や近年の BEA-2019 Shared Task がある [6][4]。文法誤り訂正手法には、規則に基づく手法、分類器を用いる手法または機械翻訳システム (NMT) を用いる手法があ

る。前置詞誤り訂正の従来研究には、例えば、計算言語学の手法の調査研究や最大エントロピー分類器を用いた研究がある [3][7]。近年は、NMT により文法誤りを統一的に扱う研究が盛んであり、BEA-2019 Shared Task では、transformer や fairseq 等によるエンコーダデコーダモデル (seq2seq) を用いる方法が主流である。また、複数の候補文を出力し、並べ替えをする手法を付加する方法も有効である [5][1]。

3 前置詞誤り訂正学習法の改良

英文法の誤り訂正に関して、単文中にある前置詞の誤り訂正を目的とする。文中にある前置詞の誤りを訂正する学習モデルとして BERT を用い、学習法の改良を提案する。

3.1 BERT

BERT は、Transformer を改良した双方向の自己注意機構をもつ学習モデルであり、まず、大規模なコーパスを用いて学習し、言語モデルを構築する (事前学習) [8]。さらに、分類タスクに応じて、後段のパラメータを学習する (転移学習)。自己注意機構は、トークンを検索クエリに変換し、キーとの類似度に応じてキーに対応する値を出力する。検索クエリ、キー及び値は、損失関数に基づき事前学習で求められる。事前学習では、 n 個のトークン系列 $\mathbf{x} = (x_1, \dots, x_n)$ を入力するとき、事前に設定された確率に基づき、 x_i をマスクトークンまたは他のトークンに置き換え、出力系列 $\mathbf{y} = (y_1, \dots, y_n)$ 中、 $y_i = x_i$ となるようパラメータが更新される。これにより、言語モデルが構築される。BERT は、RNN と異なり、トークンを一斉に入力するため、トークンの位置情報を対応トークンに接続する。また、複数種類のトークン系列を入力することも可能であり、各トークンにセグメント情報を接続する。

さらに、BERT は、転移学習により最終層を学習し、分類器を構成する。分類器の出力は、出力系列 $\mathbf{z} = (z_1, \dots, z_n)$ であり、入力系列の各トークンのクラス分類が可能である。本稿の学習法では、1 種類の

トークン系列 $\mathbf{x} = ([CLS], x_1, \dots, x_n)$ のみ入力し, z_0 の出力を学習する.

3.2 前置詞分類学習の予備実験

文法誤り訂正をもつ学習モデルを構成する場合, どの位置に誤りがあるか不明であり, また, 複数の誤りが含まれることもあるため, 一般に, seq2seq を用いることが多い. 本稿では, 前置詞の語彙が限られることを利用し, 前置詞の誤り訂正を分類問題とし, その学習法の改良を試みる.

まず, 事前学習モデルとして, BERT 開発者が Wikipedia を用いて学習した bert-base-uncased のパラメータを用い, British National Corpus (BNC) コーパスで転移学習した¹. BNC は, CLAWS part-of-speech tagger により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であり, CLAWS の主要カテゴリーに対する品詞精度は, 96–97% である. BNC コーパス中, 品詞タグが前置詞である文のみ訓練データに用い, 文中に複数の前置詞が含まれる場合は, 複数の教師学習データを生成した. また, 対象とする前置詞として前置詞句は除き, BNC コーパスに含まれる 110 種類の前置詞とし, 10M の文数のコーパスを作成した. さらに, 前置詞の推定に必要な文脈は, 主に前後の動詞または/および名詞であるため, 学習コストを低下するために, 前置詞の前後の文脈は, それぞれ 10 トークンのみ用いた [10].

前置詞の誤り訂正を分類問題として実行するため, 入力文中の対象となる前置詞を [MASK] トークンとし, [MASK] トークンを置き換えるトークン分類問題とした. このことから, 本学習モデルの訂正対象前置詞は, 異なる間違った前置詞が含まれる誤りおよび前置詞が不要な誤りである. しかしながら, 学習対象を単純化し, 提案手法の評価を明確にするために, 前置詞が不要な誤りは対象外とした. なお, 転移学習モデルは, Thomas 氏らのソースコードを用い, 表 1 のとおりの設定である².

ここで, L2 学習者コーパスである the Cambridge English Write & Improve (W&I, 3.7K の文数) のデータに基づき, L2 学習者が誤用する前置詞を図 1(a) に混同行列 (正解前置詞 – 誤り前置詞) で示す. 同図は, L2 学習者が 'in' および 'of' を誤用する確率が高いことを示している. なお, 同コーパスには, CEFR レベル A, B および C が含まれる. 同様に, BERT 学習モデ

表 1: 学習モデルのハイパーパラメータ

| Hyperparameter | value |
|-----------------------|--------|
| No. of parameters | 110M |
| Max sequence length | 64 |
| Vocabrary size | 30K |
| No. of layers | 12 |
| Hidden dimension size | 768 |
| No. of heads | 12 |
| No. of train epochs | 5 |
| Learning rate | 2.0e-5 |
| Batch size | 100 |

ルを W&I で評価したときの混同行列を同図 (b) に示す. 同図より, BERT 学習モデルも 'in' と 'of' を誤用する傾向が見られた. さらに, BNC コーパスは, 表 2 に示すとおり, 'in' および 'of' の出現頻度が, 他の前置詞の平均数よりかなり多く, 訓練データの統計的性質も影響していると考えられる. なお, 同図中, 前置詞の順番は, W&I コーパス中の評価対象前置詞の出現頻度に基づいている.

3.3 前置詞学習法の改良

前置詞 'in' および 'of' を他の前置詞と混同する確率を低下させるために, 入力文 \mathbf{x} に対する推定出力を

$$M_3(\mathbf{x}) \in \{ 'in', 'of', [OTHERS] \} \quad (1)$$

とした分類器 M_3 を用いる学習法 (3-class 学習法 $M_3 + M_m$) を提案する. BERT M_3 は, $\{ 'in', 'of', [OTHERS] \}$ の 3 分類問題を学習し, BERT M_m は, 'in' と 'of' を除く前置詞を学習する. M_3 の推定結果 $M_3(\mathbf{x}) \in \{ 'in', 'of' \}$ のとき, $M_3(\mathbf{x})$ を推定前置詞, そうでないとき, $M_m(\mathbf{x})$ を推定前置詞とする. 訓練から推定までをアルゴリズム 1 に示す.

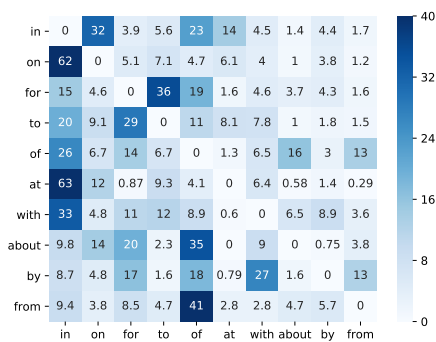
4 実験

4.1 実験方法と結果

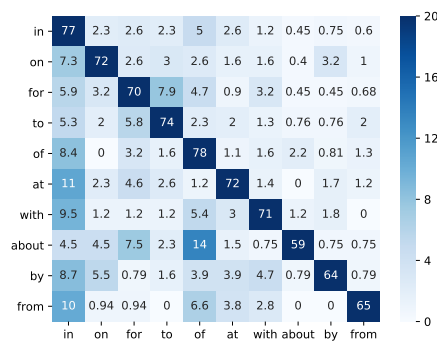
表 1 の設定により, 提案学習法 BERT $M_3 + M_m$ をそれぞれ転移学習した. なお, 学習モデル全体として, パラメータ数は分類層だけ 2 倍に増加することに注意する. BERT $M_3 + M_m$ の W&I コーパスによる評価結果を表 3 に示す. 同表は, 学習モデルの 10 種の前置詞に対する再現率 (R), 適合率 (P), $F_{0.5}$ スコアお

¹<https://github.com/google-research/bert/>

²<https://github.com/huggingface/transformers/>



(a) L2 学習者 (W&I 訓練データ)



(b) 予備実験の学習モデル

図 1: 混合行列 (行は正解前置詞, 列は L2 学習者が記述または予備実験の学習モデルが出力した誤り前置詞を示す。数値は, 正解前置詞ごとの誤り率 [%] である (対角成分は正解率 [%]).)

アルゴリズム 1: Training and prediction by a 3-class BERT Classifier

Input: BNC corpus for training, sentence \mathbf{x} for evaluating

Output: Predicted preposition

- 1 Train 3-class BERT M_3
- 2 Train multi-class BERT M_m
- 3 Check,
 - (I) If $M_3(\mathbf{x}) \in \{\text{'in'}, \text{'of'}\}$, then $M_3(\mathbf{x})$ is a predicted preposition,
 - (II) otherwise $M_m(\mathbf{x})$ is a predicted preposition.

および全体の精度 (AC) を示している。なお, 提案学習法の亜種として, $\{\text{'in'}, \text{'of'}, \text{'to'}, [\text{OTHERS}]\}$ を分類する 4-class 学習法 $M_4 + M_m$ の結果も示している。

4.2 考察

提案学習法では, とくに誤推定しやすい 'in' および 'of' とそれ以外の前置詞の学習頻度をできるだけ平滑化するために, 前置詞を $\{\text{'in'}, \text{'of'}, [\text{OTHERS}]\}$ にカテゴリー分けし, BERT M_3 により第 1 段の学習を行なった。その結果, 表 2 に示すとおり, 第 1 段階の BERT M_3 の学習では, $[\text{OTHERS}]$ が多数派となった。その後, 第 2 段階の BERT M_m で推定し, 最終的な推定前置詞を決定する。表 3 より, 提案学習モデルは, 全体の精度が若干低下したが, 前置詞 'in' および 'of' の適合率について, 若干の向上が見られた。適合率は, 文法誤り訂正のタスクにおいて, より重要な指数であ

表 2: 訓練データセット中の各前置詞の割合

| Data \ Prep. | 110-class | 3-class | 4-class | Reduct. |
|--------------|-----------|---------|---------|-------------|
| AVERAGE | 0.93 | 33.3 | 25.0 | 0.93 |
| in | 16.9 | 16.9 | 16.9 | 9.8 |
| of | 27.2 | 27.2 | 27.2 | 10.3 |
| to | 8.9 | - | 8.9 | 12.8 |
| [OTHERS] | - | 55.9 | 47.0 | - |

る。なお, 前置詞を $\{\text{'in'}, \text{'of'}, \text{'to'}, [\text{OTHERS}]\}$ に 4 分類した BERT M_4 を用いた学習モデルについて, 前置詞 'to' の適合率の向上は見られなかった。

さらに, 前置詞 'in' および 'of' の学習コーパス中の発生頻度が, 他の前置詞と混同する影響を調べるために, 表 2 中の Reduct のとおり, 'in' および 'of' を含む学習データを一部削除した。全体の精度は, 0.691 と予備実験と変わらなかった。訓練データセットで不足するデータを人工的に作成し追加することにより推定精度が向上することは知られているが, 訓練データを削減しても推定精度が変化しないことがわかった。

5 おわりに

本稿では, L2 学習者支援システムの開発に向けて, 前置詞の誤り訂正を BERT 分類器で学習するとき, L2 学習者および学習モデルの混同行列の性質を利用した学習法の改良を行なった。その結果, 全体の精度は若干低下したが, 注目した前置詞 'on' および 'of' につい

表 3: 提案学習モデルの前置詞誤り訂正の能力

| 前置詞 | 訓練データ 中の比率 [%] | 予備実験 | | | $M_3 + M_m(\text{in, of, [OTHERS]})$ | | | $M_4 + M_m(\text{in, of, to, [OTHERS]})$ | | |
|-------|-------------------|--------------|-------|-----------|--------------------------------------|--------------|-----------|--|--------------|-----------|
| | | R | P | $F_{0.5}$ | R | P | $F_{0.5}$ | R | P | $F_{0.5}$ |
| in | 16.9 | 0.771 | 0.679 | 0.695 | 0.719 | 0.700 | 0.703 | 0.731 | 0.696 | 0.703 |
| on | 5.6 | 0.715 | 0.810 | 0.789 | 0.725 | 0.774 | 0.764 | 0.727 | 0.789 | 0.776 |
| for | 7.6 | 0.703 | 0.726 | 0.721 | 0.703 | 0.703 | 0.703 | 0.703 | 0.703 | 0.703 |
| to | 8.9 | 0.738 | 0.738 | 0.738 | 0.743 | 0.728 | 0.731 | 0.720 | 0.735 | 0.732 |
| of | 27.2 | 0.776 | 0.667 | 0.686 | 0.774 | 0.687 | 0.702 | 0.757 | 0.690 | 0.703 |
| at | 4.6 | 0.722 | 0.798 | 0.782 | 0.722 | 0.793 | 0.778 | 0.722 | 0.801 | 0.784 |
| with | 5.8 | 0.714 | 0.600 | 0.620 | 0.726 | 0.589 | 0.612 | 0.726 | 0.589 | 0.612 |
| about | 1.2 | 0.586 | 0.743 | 0.705 | 0.602 | 0.690 | 0.670 | 0.617 | 0.633 | 0.669 |
| by | 4.6 | 0.638 | 0.618 | 0.622 | 0.591 | 0.615 | 0.610 | 0.575 | 0.635 | 0.622 |
| from | 3.8 | 0.651 | 0.616 | 0.623 | 0.660 | 0.560 | 0.578 | 0.642 | 0.553 | 0.569 |
| TOTAL | | $AC = 0.691$ | | | $AC = 0.680$ | | | $AC = 0.680$ | | |

て、適合率は若干改善した。今後の課題として、分類器の学習に関して、L2 学習者の混同行列を他の観点からさらに利用する検討が挙げられる。

謝辞

本研究は、JSPS 科研費 18K00904 の助成を受けたものである。

参考文献

- [1] Hiroki Asano, Masato Mita, Tomoya Mizumoto, and Jun Suzuki. The aip-tohoku system at the bea-2019 shared task. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pp. 176–182, 2019.
- [2] Yigal Attali and Jill Burstein. Automated essay scoring with e-rater® v. 2. *The Journal of Technology, Learning and Assessment*, Vol. 4, No. 3, 2006.
- [3] Timothy Baldwin, Valia Kordoni, and Aline Villavicencio. Prepositions in applications: A survey and introduction to the special issue. *Computational Linguistics*, Vol. 35, No. 2, pp. 119–149, 2009.
- [4] Christopher Bryant, Mariano Felice, Øistein E Andersen, and Ted Briscoe. The bea-2019 shared task on grammatical error correction. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pp. 52–75, 2019.
- [5] Yo Joong Choe, Jiyeon Ham, Kyubyong Park, and Yeol Yoon. A neural grammatical error correction system built on better pre-training and sequential transfer learning. *arXiv preprint arXiv:1907.01256*, 2019.
- [6] Robert Dale, Ilya Anisimoff, and George Narroway. Hoo 2012: A report on the preposition and determiner error correction shared task. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pp. 54–62. Association for Computational Linguistics, 2012.
- [7] Rachele De Felice and Stephen G Pulman. A classifier-based approach to preposition and determiner error correction in l2 english. In *Proceedings of the 22nd international conference on computational linguistics (Coling 2008)*, pp. 169–176, 2008.
- [8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [9] Merrill Swain and Sharon Lapkin. Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied linguistics*, Vol. 16, No. 3, pp. 371–391, 1995.
- [10] Alexander S Yeh and Marc B Vilain. Some properties of preposition and subordinate conjunction attachments. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 2*, pp. 1436–1442. Association for Computational Linguistics, 1998.
- [11] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. Recent trends in deep learning based natural language processing. *IEEE Computational intelligence magazine*, Vol. 13, No. 3, pp. 55–75, 2018.