

ニューラル機械翻訳モデルを用いたマルチソース文法誤り訂正

Multi-source neural grammatical error correction

曹 国林 *¹ 高村 大也 *^{2*3} 奥村 学 *^{2*4}
Cao Guolin Hiroya Takamura Manabu Okumura

*¹東京工業大学工学院 *²東京工業大学科学技術創成研究院
School of Engineering, Tokyo Institute of Technology Institute of Innovative Research, Tokyo Institute of Technology

*³産業技術総合研究所 *⁴理化学研究所 AIP
AIST RIKEN AIP

This is the first attempt to use a multi-source encoder-decoder model for the grammatical error correction task (GEC). In addition to the possibly erroneous sentence written in a second language, our model uses the sentence written in the mother tongue of the learner. With our model, we achieved up to 1.13 GLEU score increases than the single source baseline model.

1. はじめに

文法誤り訂正は、特に第二言語学習者が書いた文における誤りを検出し正しい文に訂正する課題である [9]。高性能な文法誤り訂正システムは、語学学習者に正しい指導を与えることができ、語学学習の効率が上がることが期待される。

既存の文法誤り訂正課題では、規則による手法、統計的翻訳モデル (SMT) を使った手法と、ニューラル機械翻訳モデル (NMT) を使った手法が提案されている。提案手法に関わらず、従来の文法誤り訂正システムは、語学学習者の書いた文だけを入力とし、正しい文に訂正することを目指している。しかし、誤り文だけでは、正しい訂正ができない場合もある。例えば、表 1 の入力に示したように、英語を学習している日本人が “I am **leaving** in Tokyo.” という文法的誤りを含む文を書いたとする。このような文に対し、“I am **living** in Tokyo.” と “I am leaving Tokyo.” という少なくとも二つの可能な正解文が考えられる (表 1 の可能な出力)。どちらが正しいかは、学習者が伝えたい内容に依る。例えば、学習者が伝えようとする内容が母国語で “私は東京に住んでいる” (表 2 の入力) だと分かると、誤り文を正しく “I am living in Tokyo.” に訂正する有用な手がかりとなる。ここで、母国語文は正しく書けるとする仮定は妥当だと言えよう。

そこで本研究では、誤り文と共に同じ意味の母国語文を入力とし、文法誤り訂正を行うマルチソースの文法誤り訂正モデルを提案する。また、Allen の研究で [4] 同じデータセットに対し、NMT による手法が SMT による手法より高い GLEU スコアに達成した為、本研究では NMT による手法を選択している。その結果、提案モデルはベースラインとなる誤り文のみを入力とするニューラル文法誤り訂正モデルより、高い GLEU スコアを達成している。

2. 関連研究

2.1 ニューラル文法誤り訂正

ニューラル機械翻訳モデルによる文法誤り訂正手法は、文法誤り訂正を、誤りのある文から正解文に「翻訳」する課題
連絡先: 曹国林, 東京工業大学工学院情報通信系 奥村・高村研究室, 横浜市緑区長津田町 4259 R2-728, 045-924-5295, kokurin@lr.pi.titech.ac.jp

表 1: 曖昧な誤り訂正の例

入力	可能な出力
I am leaving in Tokyo.	I am living in Tokyo.
	I am leaving Tokyo.

表 2: 母国語文の情報を用いた訂正の例

入力	可能な出力
I am leaving in Tokyo.	I am living in Tokyo.
私は東京に住んでいる.	

として捉えている。即ち、翻訳モデルを使い、入力した誤り文 $X = [x_1, \dots, x_n]$ に対し、正解文 $Y = [y_1, \dots, y_m]$ を直接生成することである。ここで x_n と y_m は、それぞれ誤り文と正解文の各単語に対応する。ニューラル文法誤り訂正モデルは Yuan [7] により初めて提案されている。

2.2 マルチソースニューラル翻訳モデル

Zoph はニューラル機械翻訳モデルをもとに、同じ意味を持つフランス語文とドイツ語文を当時に入力し、英語へ翻訳するマルチソースニューラル翻訳モデルを提案している [8]。マルチソースニューラル翻訳モデルは、encoder-decoder の encoder の部分を、二つに拡張した翻訳モデルである。

本研究では、Zoph の提案したモデルをもとに、母国語の情報を活用するマルチソース文法誤り訂正モデルを提案する。

2.3 和文英訳における文法誤り訂正

和文英訳における文法誤り訂正は、学習者が訳した学習文から、誤りを検出し正しい文に訂正する課題である [9]。その応用例として Duolingo*¹ などのサービスがある。語学学習者が決められた文に対し翻訳を行い、システムは翻訳文の間違いを訂正し、語学のライティング支援を提供している。

このような研究は、母国語文からの翻訳文に対する訂正であり、母国語文の意味保存が訂正の軸である。本研究では言語学習者が書いた誤り文に対し、母国語を補助情報として訂正を行う。

*¹ <https://www.duolingo.com/>

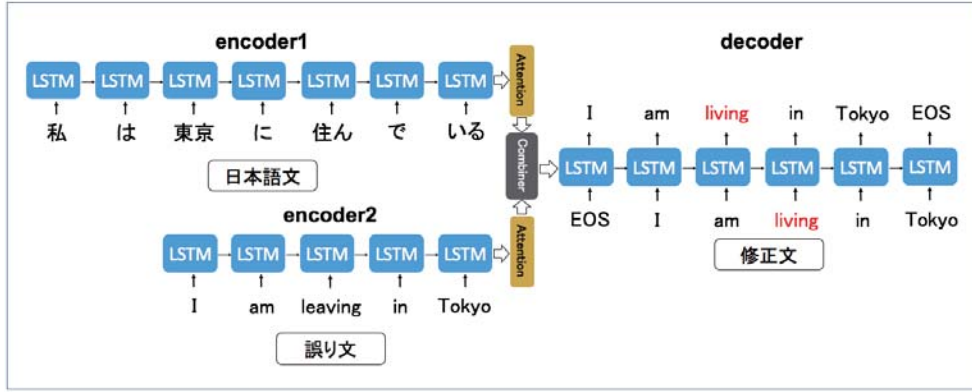


図 1: マルチソース文法誤り訂正モデル

3. ニューラル文法誤り訂正モデル

3.1 ベースラインモデル

ベースラインモデルである、注意機構付き両方向 LSTM 文法誤り訂正モデルについて説明する。

まず、encoder は英語文 $X = [x_1, \dots, x_n]$ の各要素 (単語) を単語ベクトルで表現した $[e_1^x, \dots, e_n^x]$ に対し、式 (1) と式 (2) により前向き隠れ状態と逆向き隠れ状態を計算し、式 (3) により連結させ入力単語 x_i の隠れ状態 h_i^x を計算する:

$$\vec{h}_i^x = f(\vec{h}_{i-1}^x, e_i^x), \quad (1)$$

$$\overleftarrow{h}_i^x = f(\overleftarrow{h}_{i+1}^x, e_i^x), \quad (2)$$

$$h_i^x = [h_i^x; \overleftarrow{h}_i^x]. \quad (3)$$

次に、decoder では式 (4) により decoder の入力単語 y_j に対する隠れ状態ベクトルを計算する:

$$h_j^y = f(h_{j-1}^y, e_j^y). \quad (4)$$

ここで e_j^y は入力単語 y_j の単語ベクトルである。

本研究では Luong ら [1] が提案したモデルを適用し、decoder における注意機構の計算に用いる隠れ状態ベクトル \tilde{h}_j を計算する。

注意機構とは、decoder の注意力の隠れ状態 \tilde{h}_j を計算する際、encoder の隠れ状態を考慮する機構である。encoder の各隠れ状態に対し、式 (5) によって計算されるスコアの加重平均を計算し、式 (6) に示す文脈ベクトルを計算する:

$$a_i^j = \frac{\exp(h_j^{yT} h_i^x)}{\sum_{i=1}^n \exp(h_j^{yT} h_i^x)}, \quad (5)$$

$$c_j = [h_1^x \dots h_n^x]^T [a_1^j \dots a_n^j]. \quad (6)$$

注意機構によって計算された文脈ベクトルと、decoder の現在の隠れ状態 h_j^y を連結させたものに、重み W_a をかけ、式 (7) により decoder の最終隠れ状態 \tilde{h}_j を計算する:

$$\tilde{h}_j = W_a [h_j^y; c_j]. \quad (7)$$

\tilde{h}_j を利用し、式 (8) によって正解文中の各単語 y_j を予測させる:

$$p(y_j | y_{<j}, X) = \text{softmax}(\tilde{h}_j). \quad (8)$$

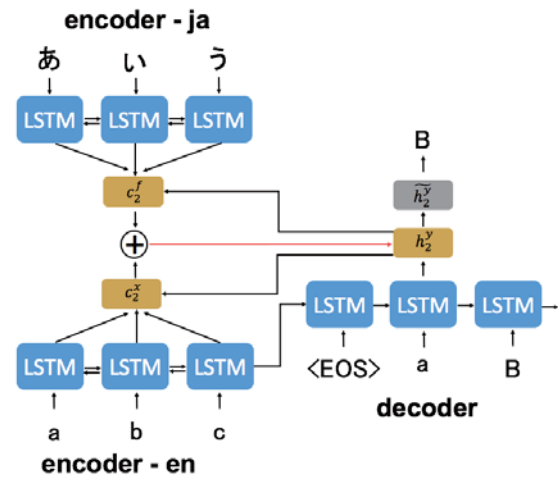


図 2: マルチ加算注意機構モデル。英語文「abc」と日本語文「あいう」を同時に入力した場合、「aBc」に訂正する decoder の出力一部である。ここで ja は日本語で en は英語を表す。

3.2 提案モデル

本研究では、Zoph のマルチソースニューラル翻訳モデルをもとに、二つの異なる注意機構をもったマルチソースニューラル文法誤り訂正モデルを提案する。2つの提案モデルの概要は図 1 に示す。

3.2.1 マルチ加算注意機構モデル

図 2 にマルチ加算注意機構モデルの詳細を示している。マルチ加算注意機構モデルは、英語文 $X = [x_1, \dots, x_n]$ と、母国語文 $F = [f_1, \dots, f_l]$ を入力とする 2つの encoder を持っている。それぞれの encoder は、入力文を隠れ状態 $h^x = [h_1^x, \dots, h_n^x]$ と $h^f = [h_1^f, \dots, h_l^f]$ に変換する。次に、英語文の最終の隠れ状態である h_n^x を decoder に渡し、decoder の隠れ状態の初期値とする。また、二つの encoder の隠れ状態に、式 (5)、式 (6) に示した注意機構を適用し、二つの文脈ベクトル c_j^x と c_j^f を計算する。

最後に式 (6) によって計算された英語文と母国語文の文脈ベクトル c_j^x と c_j^f に対し、式 (9) により decoder の最終隠れ状態の計算手法を提案する:

$$\tilde{h}_j = W_{sum} [(h_j^y); (c_j^x + c_j^f)]. \quad (9)$$

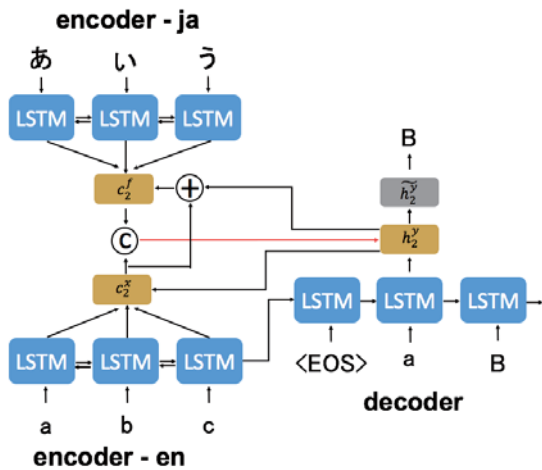


図 3: マルチ階層注意機構モデル. 英語文「abc」と日本語文「あいう」を同時に入力した場合、「aBc」に訂正する decoder の出力一部である. ここで ja は日本語で en は英語を表す. ここで⊕は二つのベクトルを連結することを意味する.

式 (9) によって計算された \tilde{h}_j を利用し, 式 (8) によって正解文中の各単語ベクトル y_j を予測させる.

3.2.2 マルチ階層注意機構モデル

Yang ら [6] の研究で, 階層注意機構モデルは文書分類を行う際, 入力文からより分類結果と関連する単語情報を選択できることを示している. 図 3 にマルチ階層注意機構モデルの詳細を示している. 本研究では階層注意機構モデルのこのような特徴を利用し, decoder が各単語を予測する際, 日本語入力文から, 関連する単語情報をよりよく選択するように学習させるため, マルチ階層注意機構モデルを提案している. マルチ階層注意機構モデルは, decoder の現在の隠れ状態 h_j^y と英語文の文脈ベクトル c_j^x を足した結果を利用し (式 (10)), 母国語文の文脈ベクトルを計算する仕組みである (式 (11)):

$$a_k^{f,j} = \frac{\exp((h_j^y + c_j^x)^T h_k^f)}{\sum_{k=1}^n \exp((h_j^y + c_j^x)^T h_k^f)}, \quad (10)$$

$$c_j^f = [h_1^f \dots h_l^f]^T [a_1^{f,j} \dots a_l^{f,j}]. \quad (11)$$

最後に英語文と母国語文の文脈ベクトル c_j^x と c_j^f に対し, 式 (12) にもとづき \tilde{h}_j を計算し, 式 (8) によって正解文中の各単語ベクトル y_j を予測させる:

$$\tilde{h}_j = W_{hier} [h_j^y; c_j^x; c_j^f]. \quad (12)$$

4. 実験

本研究の全てのモデルは Github にある, PyTorch の seq2seq のチュートリアルコード*2 をもとに実装を行っている.

*2 <https://github.com/spro/practical-pytorch/blob/master/seq2seq-translation/seq2seq-translation.ipynb>

表 3: ベースラインと二つのマルチソースニューラル文法誤り訂正モデルの GLEU スコア (%)

モデル	開発データ	評価データ
入力文のまま	59.65	58.21
baseline	60.59	58.76
Multi-source-sum	61.41	59.54
Multi-source-hier	61.92	59.89
日本語文	34.33	33.42

表 4: ベースラインのモデルと提案モデルの訂正例
例 1

誤り文	Please correct in writing .
日本語入力文	私の文章を訂正してください.
正解文	Please correct my writing .
ベースライン	Please correct in writing .
提案モデル:sum	Please correct my writing .
提案モデル:hier	Please correct my writing .

例 2

誤り文	I am leaving in Tokyo .
日本語入力文	私は東京に住んでいる.
正解文	I am living in Tokyo .
ベースライン	I am leaving in Tokyo .
提案モデル:sum	I am living in Tokyo .
提案モデル:hier	I am living in Tokyo .

4.1 実験データ

まず, 本研究では誤り文, 正解文と母国語文の三つの文が揃ったデータが必要である. しかし, 既存のデータは誤り文と正解文のペアになっているデータしか存在しない. そこで, 本研究では既存の Lang-8 Learner Corpora[5] から, 正解文を日本語に翻訳し, 対応する日本語母国語文のデータを作成した. 正解文の翻訳には Google Translation*3 を使用し, 翻訳した日本語文は MeCab*4 で形態素解析をしている. 実験では, 学習用データ 1,070,000 文, 開発データ 18,452 文と評価用データ文 10,227 文を使用した. また encoder と decoder は一層のニューラルネットワーク構造を使い, 隠れ状態は 300 次元に設定している.

4.2 実験結果

今回の実験では, GLEU[2] を用いて各モデルの評価を行う. GLEU は機械翻訳課題での評価手法である BLEU[3] を文法誤り訂正課題向けに拡張したものであり, その計算方法は式 (13) に示している. ここで p^* は計算された GLEU スコアを表し, A と B は文を表す. また $count_{A,B}(ngram)$ は, A における $ngram$ の出現頻度と, B における $ngram$ の出現頻度のうち小さい方を返す. また $count_C(ngram)$ は, C における $ngram$ の出現頻度を返す. 即ち GLEU では, 入力文 S において訂正されるべき ($ngram$) が, システムの出力文 C で訂

*3 <https://cloud.google.com/translate/?hl=ja>

*4 <http://taku910.github.io/mecab/>

$$p^* = \frac{\sum_{ngram \in C \cap R} count_{C, R}(ngram) - \sum_{ngram \in C \cap S} \max[0, count_{C, S}(ngram) - count_{C, R}(ngram)]}{\sum_{ngram \in C} count_C(ngram)} \quad (13)$$

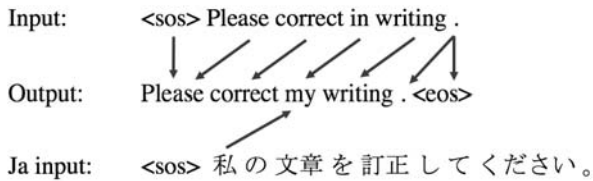


図 4: 各単語を出力する際の注意機構の働き。Ja は日本語を表す。<eos> と <eos> はそれぞれ文頭と文尾を意味する。矢印は、各単語を出力する際に、注意機構の重みが最大となる入力単語を表している。ただし、日本語文については、“my” 以外はすべて <eos> が最大であったため、それらは図では省略している。

正されていない場合、減点することでモデルを評価する。ここで R は正解文を意味する。

実験結果は表 3 に表す。同じ実験設定で実験を行った結果、二つの提案モデルともベースラインのモデルを上回る精度を得ている。また、マルチ階層注意機構モデルはベースラインのモデルと比較して、GLEU スコアにおいて 1.13 ポイントの上昇を達成している。

また、今回の実験用データを使い、日本語文から正解文への翻訳を行なっている。その結果、開発データと評価データに対しそれぞれ 34.33 と 33.42 の GLEU スコアを達成している。

4.3 出力に対する考察

ここで、ベースラインモデルと提案モデルの出力結果に対して分析を行う。表 4 に二つの誤り文 “Please correct in writing” と “I am leaving in Tokyo .” の入力に対し、各モデルの出力結果を示す。ベースラインモデルは二つの入力文を正しく訂正出来ていないが、両提案モデルは正しく訂正を行っている。これは提案モデルが日本語入力の情報を利用し、訂正を行っているからだと考えられる。

図 4 にマルチ階層注意機構モデルが表 4 の例 1 の入力に対し、各単語を予測する際の注意機構の行動を示している。システムの各出力単語の母国語入力文における $a_k^{f,j}$ の値を確認した結果、単語 “my” を予測する時、日本語入力文の単語 “私” の情報がもっとも重要視されている。

5. おわりに

本研究は、文法誤り訂正課題において、初めて言語学習者の母国語の情報を利用した研究である。また、実験において、マルチソースニューラル文法誤り訂正のモデルの有効性を検証した。提案モデルは、GLEU スコアにおいてベースラインモデルを上回っていることを示した。

今後の課題として、水本 [10] により提案されたコーパスの構築手法を使い、機械翻訳によって生成された母国語ではなく、語学学習者が書いた母国語文の含むコーパスに対しモデルを適用することが挙げられる。

参考文献

- [1] Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention-based neural machine translation. *CoRR*, 2015.
- [2] Courtney Napoles, Keisuke Sakaguchi, Matt Post, and Joel R. Tetreault. GLEU without tuning. *CoRR*, 2016.
- [3] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, pp. 311–318, Philadelphia, Pennsylvania, USA, July 2002.
- [4] Allen Schmalz, Yoon Kim, Alexander M. Rush, and Stuart M. Shieber. Adapting sequence models for sentence correction. *CoRR*, 2017.
- [5] Toshikazu Tajiri, Mamoru Komachi, and Yuji Matsumoto. Tense and aspect error correction for esl learners using global context. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 198–202, Jeju Island, Korea, July 2012.
- [6] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1480–1489, San Diego, California, June 2016.
- [7] Zheng Yuan and Ted Briscoe. Grammatical error correction using neural machine translation. In *HLT-NAACL*, 2016.
- [8] Barret Zoph and Kevin Knight. Multi-source neural translation. *CoRR*, 2016.
- [9] 永田亮. 語学学習支援のための言語処理. 自然言語処理シリーズ 11. コロナ社, 2017.
- [10] 水本智也. 語学学習 SNS の添削ログからの母語訳付き学習者コーパスの構築に向けて. 第 6 回コーパス日本語学ワークショップ予稿集, 2014.