

# ニューラル機械翻訳における双方向反復的教師なし適応の改善

藤澤 兼太      秋葉 友良      塚田 元

豊橋技術科学大学

fujisawa@nlp.cs.tut.ac.jp, akiba@cs.tut.ac.jp, tsukada@brain.tut.ac.jp

## 1 はじめに

近年、機械翻訳分野において、ニューラルネットワークを用いたニューラル機械翻訳 (Neural Machine Translation: NMT) が高い性能を示しており、機械翻訳における代表的な手法になっている。

NMT の翻訳性能は、対訳コーパスのサイズや質に大きく影響を受ける。そのため、対訳コーパスの構築が難しい分野 (ドメイン) においては、良質な翻訳モデルの学習が困難という問題点がある。

低リソースなドメインにおいて対訳コーパスの構築は困難であるが、対訳でない単言語コーパスを構築することは比較的容易である。そのため、翻訳モデルの学習に単言語コーパスを活用する様々な手法が提案されている。Lample ら [1] は、対訳コーパスを用いずに、両言語の単言語コーパスのみを用いて NMT モデルの教師なし学習を行う手法を提案した。Gülçehre ら [2] は、目的言語側の単言語コーパスのみを用いて学習した言語モデルを、NMT モデルに効果的に統合する手法を提案した。Sennrich ら [3] は、ドメイン内の単言語コーパスを逆翻訳した後、元の単言語コーパスと合わせることで疑似的な対訳コーパスを構築し、学習に用いる手法を提案した。森田ら [4]、Hoang ら [5] および Zhang ら [6] は Sennrich らの手法を拡張し、単言語コーパスの逆翻訳及び疑似対訳コーパスの構築を双方向かつ反復的に行うことで、翻訳性能のさらなる向上を達成した。

本研究では、森田らが提案したドメイン適応手法の性質を調査するとともに、手法の改善を図ることを目的とする。具体的にはまず、(1) 文の分割単位により効果は変わるか; (2) ドメイン外対訳コーパスのサイズにより性能はどう変わるかの2点について実験を行い、考察する。次に、同一の単言語コーパスから構築した疑似対訳コーパスを複数結合して学習に用いる手法を提案し、翻訳性能の向上を図る。

実験の結果、文の分割単位は単語単位よりもサブワード単位の方が性能が高くなることがわかった。ド

メイン外対訳コーパスのサイズによる性能比較では、対訳コーパスのサイズが小さいほど手法による効果が高くなるという結果が得られた。疑似対訳コーパスの拡張による翻訳性能の改善実験では、異なる系列 (2.2 節で解説) のモデルと混ぜることにより翻訳性能が向上するということがわかった。

## 2 双方向反復的教師なし適応

### 2.1 手法の流れ

2つの言語をそれぞれ  $X$ ,  $Y$  と表し、 $X$  からの  $Y$  への翻訳を  $X \rightarrow Y$ 、 $Y$  から  $X$  への翻訳を  $Y \rightarrow X$  と表す。森田らの手法 [4] では以下の手順によりモデルを学習していく (図 1)。

1 ドメイン外の対訳コーパス ( $C_X^{out}$ ,  $C_Y^{out}$ ) を用いて、 $X \rightarrow Y$  方向の翻訳モデル  $Model_{X \rightarrow Y}^0$  と  $Y \rightarrow X$  方向の翻訳モデル  $Model_{Y \rightarrow X}^0$  を学習する。

2 以下の手順により、モデルを再学習する ( $i$  の初期値は 0)。

2.1  $Model_{Y \rightarrow X}^i$  を用いて単言語コーパス  $C_Y^{in}$  を翻訳し、疑似対訳コーパス ( $C_X^{in'}$ ,  $C_Y^{in}$ ) を得る。疑似対訳コーパスを対訳コーパス ( $C_X^{out}$ ,  $C_Y^{out}$ ) と混合し、 $Model_{X \rightarrow Y}^{i+1}$  を学習する。

2.2  $Model_{X \rightarrow Y}^i$  を用いて単言語コーパス  $C_X^{in}$  を翻訳し、疑似対訳コーパス ( $C_Y^{in'}$ ,  $C_X^{in}$ ) を得る。疑似対訳コーパスを対訳コーパス ( $C_Y^{out}$ ,  $C_X^{out}$ ) と混合し、 $Model_{Y \rightarrow X}^{i+1}$  を学習する。

3  $i \leftarrow i + 1$  とし、ステップ 2 に戻る。

### 2.2 学習の系列

双方向反復的教師なし適応では、英日翻訳モデルを学習した後はそれを用いて日英翻訳モデルの学習用

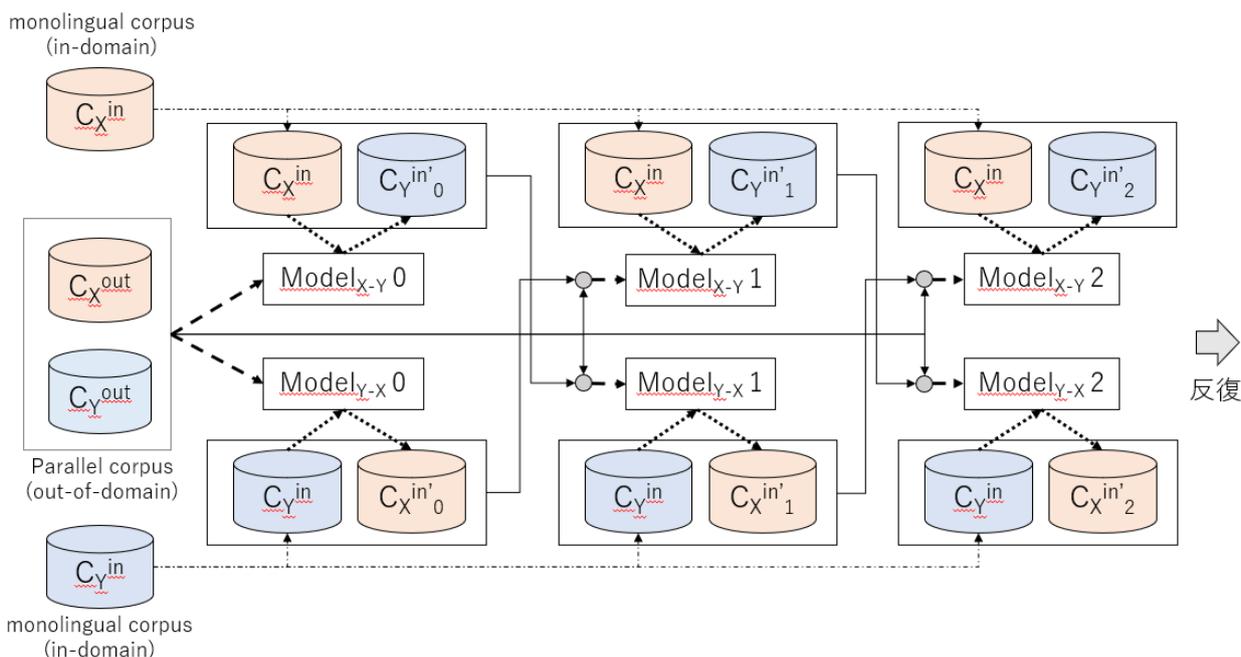


図 1: 双方向反復的教師なし適応の流れ

の疑似対訳コーパスを作成するというように交差して学習が進んでいく。そのため、片側の翻訳方向には  $Model 0, Model 2, Model 4 \dots$  と学習が進んでいく系列と、 $Model 1, Model 3, Model 5 \dots$  と学習が進んでいく系列の 2 つの系列が現れる。この 2 つの学習系列はそれぞれ独立して学習が進んでいくということが、この手法において重要な性質の一つになる。

### 3 実験

#### 3.1 実験設定

##### 3.1.1 データセット

ドメイン外の対訳コーパスには、Asian Scientific Paper Excerpt Corpus (ASPEC)[7] の英日対訳コーパスを用いた。ドメイン内の単言語コーパスには、NTCIR8 PATMT の英日対訳コーパスから、互いに対訳にならないように英語 150 万文、日本語 150 万文を抽出して使用した。テストデータには NTCIR8 PATMT より 899 対、開発データには NTCIR8 PATMT より 2000 対を用いた。全てのテキストデータに対して NFKC 変換と小文字化を行い、トークナイザによる文の分割を行った。単語単位の分割には、日本語に対しては MeCab、英語に対しては SMT システム Moses のスク

リプトを使用した。サブワード単位の分割には、Sentence Piece[8] の BPE ベースのものを使用し、分割する語彙の上限は 8000 とした。Sentence Piece のモデルの学習には、ドメイン外の対訳コーパスとドメイン内の単言語コーパスを結合したものをを用いた。Sentence Piece のモデルを学習する際には、日本語は MeCab による事前分割をしていない原文を使用し、英語は Moses により事前分割したものをを使用した。

##### 3.1.2 実験方法

2 章に記述した手法に従い、まずドメイン外 (ASPEC) の英日対訳コーパスを使用して双方向の  $Model 0$  を学習する。次に  $Model 0$  によりドメイン内 (NTCIR8 PATMT) の単言語コーパスを翻訳し、ドメイン外対訳コーパスと合わせて疑似対訳コーパスを構築する。この疑似対訳コーパスを用いて両方向の翻訳モデル  $Model 1$  を学習する。同様の動作を反復して  $Model 10$  まで学習し、各モデルの性能を BLEU スコア [9] により評価する。10 エポック学習を行い、エポックごとに出力されたモデルの中から最も開発データに対する accuracy が高いモデルを選択した。

ニューラル機械翻訳システムの基盤には、OpenNMT[10] を使用した。エンコーダを 1 層

表 1: 条件ごとのモデルの BLEU 値

	tokenizer	En-Ja	Ja-En
1M	MeCab/Moses	34.74	33.23
	Sentence Piece	34.60	<b>34.71</b>
500K	MeCab/Moses	32.87	32.60
	Sentence Piece	<b>34.52</b>	<b>34.41</b>
100K	MeCab/Moses	29.84	30.12
	Sentence Piece	<b>30.91</b>	29.88

の双方向 500 次元 LSTM, デコーダを 1 層の単方向 500 次元 LSTM にし, 学習アルゴリズムには学習率を 0.001 とした Adam を用いた.

## 3.2 双方向反復的教師なし適応手法の性質

### 3.2.1 実験条件

対訳コーパスのサイズを 100 万文 (1M), 50 万文 (500K), 10 万文 (100K) と変え, 実験を行った. また, 文の分割単位を MeCab/Moses または Sentence Piece と変えて実験を行った.

### 3.2.2 実験結果

表 1 に, 各条件において *Model 10* まで学習した中で, BLEU スコアの最高値を示す. 1pt 以上の差に注目すると, コーパスサイズが 1M の場合, 日英では Sentence Piece が 1.48pt 高い. コーパスサイズが 500K の場合, Sentence Piece が英日, 日英それぞれ 1.65pt, 1.81pt 高い. 100K の場合では, 英日では Sentence Piece が 1.07pt 高い. 以上の結果から Sentence Piece の方が平均的に性能が高いことがわかる. また, コーパスサイズが 1M の場合と 500K の場合の BLEU スコアの差が英日, 日英両方において小さくなっている. この結果から, 初期の対訳コーパスのサイズを 1M より更に大きくした場合でも, 改善幅は徐々に小さくなっていくことが予想される.

次に, 対訳コーパスのサイズによって, 手法の効果の大きさに影響が出るのかについて検証する. トークナイザに Sentence Piece を用いた場合における, コーパスサイズごとの BLEU スコアの上昇の様子を表 2 に示す. BLEU スコアを示す列の括弧内の値は *Model 0* からの上昇値である. また, *Model n* の  $n$  の値は, 10 回の反復学習を行った中で最も性能の高かったモデル

表 2: 対訳コーパスサイズごとの手法の効果

	Model	En-Ja	Ja-En
1M	<i>Model 0</i>	13.52	18.26
	<i>Model 1</i>	29.14(+15.62)	29.90(+11.64)
	<i>Model n</i> ( $n = 9$ )	34.60(+21.08)	34.71(+16.45)
500K	<i>Model 0</i>	11.64	14.77
	<i>Model 1</i>	27.82(+16.18)	27.52(+12.75)
	<i>Model n</i> ( $n = 7$ )	34.52(+22.88)	33.52(+18.75)
100K	<i>Model 0</i>	7.44	9.28
	<i>Model 1</i>	21.00(+13.56)	19.30(+10.02)
	<i>Model n</i> ( $n = 10$ )	30.91(+23.47)	29.88(+20.60)

のものになるため, コーパスサイズごとに違うことに注意されたい. 英日, 日英どちらにおいても対訳コーパスのサイズが小さくなるほど上昇の幅が大きくなっていることがわかる. よって, 対訳コーパスのサイズが小さい場合に特に手法の効果が現れると考えられる.

## 3.3 疑似対訳コーパスの拡張による改善

### 3.3.1 実験条件

前節での実験では, 単言語コーパスを逆翻訳して構築した疑似対訳コーパスは一つしか学習に用いていなかった. この節では, 同一の単言語コーパスから逆翻訳により構築した疑似対訳コーパスを複数結合し, 学習に用いる手法を提案する. 具体的には, あるモデルにより構築された疑似対訳コーパスに対して, 異なる系列のモデルにより構築された疑似対訳コーパスを結合する場合と, 同じ系列のモデルにより構築された疑似対訳コーパスを結合する場合を実験し, 比較した.

今回の実験では, ドメイン外対訳コーパスサイズを 100K とした場合の, 英日翻訳モデル *Model 10* を学習する際に拡張を行うことを考えた. つまり, 日英翻訳モデル *Model 9* で構築した疑似対訳コーパスに, 別の日英翻訳モデル *Model n* ( $n < 9$ ) により構築した疑似対訳コーパスを追加し, 学習を行った. 実験結果を表 3 に示す. 表内の *Model 9'* と *Model 9''* は, *Model 9* を学習する際にエポックごとに出力されたモデルのうち, *Model 9* を除く上位 2 つのモデルである. 表 3 の () 内に示した日英 BLEU スコアに注目しながら結果を比較すると, 日英 BLEU スコアが比較的高い同じ系列 (2.2 節) のモデル (*Model 7*, *Model 5* などの奇

表 3: 疑似対訳コーパスの拡張実験 (ドメイン外対訳コーパスサイズ 100K の場合)

Je-En Model for Back-Translation(BLEU)	En-Ja
<i>Model 9</i> (29.88)	30.91
+ <i>Model 9'</i> (30.63)	31.1
+ <i>Model 9''</i> (29.74)	31.24
+ <i>Model 8</i> (29.43)	<b>32.23</b>
+ <i>Model 7</i> (29.51)	30.99
+ <i>Model 6</i> (27.98)	<b>31.74</b>
+ <i>Model 5</i> (28.14)	31.04
+ <i>Model 4</i> (26.11)	<b>31.47</b>
+Model 8(29.43) + Model 7(29.43)	32.3
+Model 8(29.43) + Model 6(27.98)	32.67

数モデル)による疑似対訳コーパスを追加した場合よりも、異なる系列のモデル (*Model 8*, *Model 6* などの偶数モデル)による疑似対訳コーパスを追加した場合の方が、英日 BLEU スコアが高くなっている。以上の結果から、異なる系列による疑似対訳コーパスと混合することにより、効果的に性能を向上させることができると考えられる。

表 3 の下部に、疑似対訳コーパスを 2 つ追加した場合の結果を示している。学習データがかなり増えるため計算コストは高くなるものの、同様に BLEU スコアが向上することを確認した。

## 4 結論

本研究では、森田ら [4] の手法に対して、手法の性質の調査と、改善するための手法の提案を行った。

文の分割単位による効果の調査実験から、文の分割単位はサブワード単位 (Sentence Piece) の方が性能が高くなるという結果が得られた。ドメイン外対訳コーパスサイズによる手法の効果に関する実験からは、サイズが小さいほど手法の効果 (性能の向上) が表れやすく、サイズがある程度大きくなると最終的な翻訳性能の差は小さくなることがわかった。

手法を改善するために、同一の単言語コーパスから作成した疑似対訳コーパスを複数用いる手法を提案した。実験の結果、同じ系列のモデルによる疑似対訳コーパスと混ぜるよりも、異なる系列による疑似対訳

コーパスと混ぜることにより、性能が効果的に改善されるということがわかった。

今後の研究としては、単言語コーパスのサイズが増加または減少した場合の手法の効果について調査したい。また、複数の疑似対訳コーパスの混合を反復時に毎回行う手法について実験を行い、更なる改善を図りたい。

謝辞 本研究は JSPS 科研費 19K11980 および 18H01062 の助成を受けた。

## 参考文献

- [1] G. Lample, L. Denoyer, and M. Ranzato. Unsupervised machine translation using monolingual corpora only. CoRR, Vol. abs/1711.00043. 2017.
- [2] Ç. Gülçehre, O. Firat, K. Xu, K. Cho, L. Barrault, H. Lin, F. Bougares, H. Schwenk, and Y. Bengio. On using monolingual corpora in neural machine translation. CoRR, Vol. abs/1503.03535, 2015.
- [3] R. Sennrich, B. Haddow, and A. Birch. Improving neural machine translation models with monolingual data. CoRR, Vol. abs/1511.06709, 2015.
- [4] 森田知熙, 秋葉友良, 塚田元. 双方向ニューラル機械翻訳の反復的な教師なし適応の検討. 言語処理学会第 25 回年次大会 発表論文集, pp. 14511454, 2015.
- [5] V.C.D. Hoang, P. Koehn, G. Haffari, T. Cohn, “Iterative back-translation for neural machine translation”, In Proceedings of the 2nd Workshop on Neural Machine Translation and Generation, pp.18–24, 2018.
- [6] Z. Zhang, S. Liu, M. Li, M Zhou, and E Chen, “Joint training for neural machine translation models with monolingual data”, In Proceedings of AAAI, 2018.
- [7] T. Nakazawa, M. Yaguchi, K. Uchimoto, M. Utiyama, E. Sumita, S. Kurohashi, and H. Isahara. ASPEC: Asian scientific paper excerpt corpus. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2016), pp. 2204–2208, 2016.
- [8] T. Kudo and J. Richardson. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pp. 66-71, 2018.
- [9] K. Papineni. BLEU : a method for automatic evaluation of machine translation. 40th Annual meeting of the Association for Computational Linguistics, pp. 311-318, 2002.
- [10] G. Klein, Y. Kim, Y. Deng, J. Senellart, and A.M.Rush. OpenNMT: Open-source toolkit for neural machine translation. CoRR abs/1701.02810, 2017.