

Fine-Tuning と混成的な逆翻訳サンプリングに基づく NMT の双方向反復的教師なし適応の改善

森田 知熙

秋葉 友良

塚田 元

豊橋技術科学大学

morita.tomohiro.al@tut.jp

akiba.tomoyoshi.tk@tut.jp

tsukada.hajime.hl@tut.jp

1 はじめに

機械翻訳においては、学習に使用する対訳コーパスの量と質が最終的な翻訳精度に直結する。しかしながら、対訳コーパスは単言語コーパスに比べて収集、構築が困難である。そのため、豊富な単言語コーパスを活用し、ドメイン内での翻訳精度を向上させるドメイン適応の研究が盛んに行われている。

中でも、Sennrich ら [1] は単言語コーパスを利用した半教師あり学習手法として単言語コーパスを逆翻訳して疑似対訳コーパスを生成し、翻訳モデルを学習することで翻訳精度の向上を達成した。

さらに Hoang ら [2], Zhang [3] らはこの手法を双方向に繰り返し行う、つまり疑似対訳コーパスの生成と翻訳モデルの学習を2つの言語間で交互に繰り返すことでさらなる翻訳精度の向上を達成している。この手法では、疑似対訳コーパスにより翻訳精度が改善された翻訳モデルにより更に質の高い疑似対訳コーパスを生成できるため、これを繰り返すことで翻訳精度を大きく改善することができる。また、森田ら [4] 藤澤ら [5], Jin ら [6] はこれらの手法のドメイン適応での有効性を評価している。

本論文では、単言語コーパスを利用した繰り返しの半教師あり学習によるドメイン適応手法において、さらに fine-tuning により、以前の翻訳モデルから再学習する方法を提案する。また、繰り返し学習の際、疑似対訳コーパスの単調化を防ぐため、beam search による逆翻訳文と random sampling による逆翻訳文を組み合わせることで疑似対訳コーパスの質を維持しつつ多様性を増加させる逆翻訳手法を提案する。

科学技術論文ドメインから特許ドメインへの適応について英日翻訳、日英翻訳の実験を行った結果、fine-tuning を行わない既存の繰り返し学習手法に対して、英日方向で+2.59 BLEU が向上した。さらに、

random sampling と beam search により生成した疑似対訳コーパスを併用することで、最終的な BLEU 値はそれぞれ英日方向で 40.58、日英方向で 38.22 となり、完全な対訳コーパスで教師あり学習を行った Oracle モデルの BLEU 値 (41.95, 40.04) に迫る翻訳精度を達成した。

2 関連研究

ニューラル機械翻訳におけるドメイン適応手法は、主にターゲットドメインに対訳コーパスがあることを前提とした教師あり学習手法と、単言語コーパスを利用する半教師あり学習手法を用いたものに大別され、本論文は後者に該当する。

対訳コーパスに比べ収集、構築が容易な半教師あり学習は様々な手法が提案されている。Gülçehre ら [7] はターゲット側の単言語コーパスから学習した言語モデルを利用する方法を提案している。Cehng ら [8] はオートエンコーダを用いて単言語コーパスと対訳コーパスを組み合わせる学習方法を提案している。Zhang ら [3] も同様にオートエンコーダを用いて翻訳と文の並べ替えを同時に学習するマルチタスク学習手法を提案している。これらの手法は単言語コーパスを用いるために翻訳システムの変更が必要となる。

一方で、翻訳モデルの大規模な変更を必要としない半教師あり学習手法も多く提案されている。Currey ら [9] はターゲット側の文をそのままソース側にコピーして入力する方法を提案している。Sennrich ら [1] は少量の対訳コーパスから学習したターゲット側の翻訳モデルで単言語コーパスを逆翻訳し、疑似対訳コーパスとして利用する方法を提案している。今村ら [10][11][12] は、Sennrich らの手法を拡張し、逆翻訳の際に random sampling を用いることで、疑似ソース文に多様性を持たせることで翻訳品質が向上することを報告している。Graça ら

[13] は、今村ら [10][11] の提案を踏まえ、N-ベストサンプリングにより探索空間を制限することで高品質な疑似対訳コーパスのみを生成する方法を提案している。

Zhang ら [3] や Hoang ら [2] は Sennrich らの手法を拡張し、ターゲット側だけでなくソース側の翻訳モデルも同様の手順で性能を改善することで、双方向の翻訳モデルの性能を改善する。さらに二つの性能が改善されたモデルで単言語コーパスをもう一度逆翻訳し、疑似対訳コーパスの質を向上させることができる。これを繰り返し、二つの言語間で双方向に疑似対訳コーパスの生成と翻訳モデルの更新を同時に繰り返し行うことにより翻訳精度のさらなる改善を達成している。また、Jin ら [6] や藤澤ら [5]、森田ら [4] はこれらの手法のドメイン適応での実験を行っており、その有効性を評価している。これらの手法では、各繰り返しで直前のモデルで生成した疑似対訳コーパスを用いて一から次の翻訳モデルを学習している。

3 提案法

本論文で提案する手法は、Zhang ら [3]、Hoang ら [2]、Jin ら [6]、森田ら [4] や藤澤ら [5] の双方向反復的教師なし適応手法をもとに、fine-tuning を取り入れたものである。

従来の双方向反復的教師なし適応手法 [4] では、同時に両方向の翻訳モデルの性能を改善するが、実際には両方向の翻訳モデルは交差する関係にあり、独立した二つの系列として交わることはない。

そのため、二つの系列どうして翻訳精度に差が生じ、また、優れた系列側の知識をもう一つの系列の性能改善に活かさないという問題があった。藤澤ら [5] は異なる系列の疑似対訳コーパスをもう一つの系列の疑似対訳コーパスと結合することで翻訳精度を改善できることを報告しているが、学習データが増大するために、繰り返し学習全体の時間の増大を避けることができない。

本論文で提案する手法は、翻訳モデルを学習する際、一からモデルを学習するのではなく一つ前のモデルから fine-tuning することで異なる二つの系列の疑似対訳コーパスの知識を学習することができる。また、学習データのサイズは従来手法 [4] と変わらないため、繰り返し学習全体の時間コストも増加しないという利点がある。

源言語を X、目的言語を Y とし、X から Y への翻

訳を X-Y、Y から X への翻訳を Y-X とする。提案法の手順は以下のとおりである。(図 1)

- 1 ドメイン外対訳コーパス (D_X^{out}, D_Y^{out}) から両方向の翻訳モデル ($Model_{X-Y}0, Model_{Y-X}0$) を学習する。
- 2 以下の手順で $Model_{X-Y}(i)$ を更新する。
 - 2.1 $Model_{Y-X}i$ で単言語コーパス D_Y^{in} を翻訳し、疑似対訳コーパス ($D_X^{in'}, D_Y^{in}$) を得る。
 - 2.2 疑似対訳コーパス ($D_X^{in'}, D_Y^{in}$) と対訳コーパス (D_X^{out}, D_Y^{out}) を混合し、 $Model_{X-Y}(i)$ から fine-tuning を行い $Model_{X-Y}(i+1)$ を得る。
- 3 以下の手順で $Model_{Y-X}i$ を更新する。
 - 3.1 $Model_{X-Y}i$ で単言語コーパス D_X^{in} を翻訳し、疑似対訳コーパス $D_Y^{in'}, D_X^{in}$ を得る。
 - 3.2 疑似対訳コーパス ($D_Y^{in'}, D_X^{in}$) と対訳コーパス (D_Y^{out}, D_X^{out}) を混合し、 $Model_{Y-X}i$ から fine-tuning を行い $Model_{Y-X}(i+1)$ を得る。
- 4 $i \leftarrow i+1$ としてステップ 2 に戻る。

また、今村ら [12] は、逆翻訳による半教師あり学習手法において、random sampling により疑似原文を複数生成することで翻訳精度が向上することを指摘しているが、学習データが増大するために、学習時間が増加するという欠点がある。さらに、Edunov ら [14] は、初期対訳コーパスサイズが小さい場合、逆翻訳器の翻訳精度が低いために、beam search 以外の逆翻訳手法は逆効果になることを報告している。

そこで、疑似対訳コーパスを生成する際に、半分を beam search、もう半分を random sampling により逆翻訳し、これらを混合する Combining Sampling を提案する。この手法は、逆翻訳器の性能が低い場合でも、ノイズが小さい beam search による翻訳文が混在するため疑似対訳コーパスの質の低下を防ぎつつ多様性の増加を期待できる。さらに、学習データサイズが増加しないため、繰り返し学習においてもデメリットなしで使用できる。

なお、学習データの質の偏りを防ぐため、逆翻訳の際に 1 行ごとに交互にサンプリング方法を切り替え、また上記の手順で翻訳モデルの更新を行うたびに単言語コーパスの偶数行と奇数行を入れ替える。

4 実験

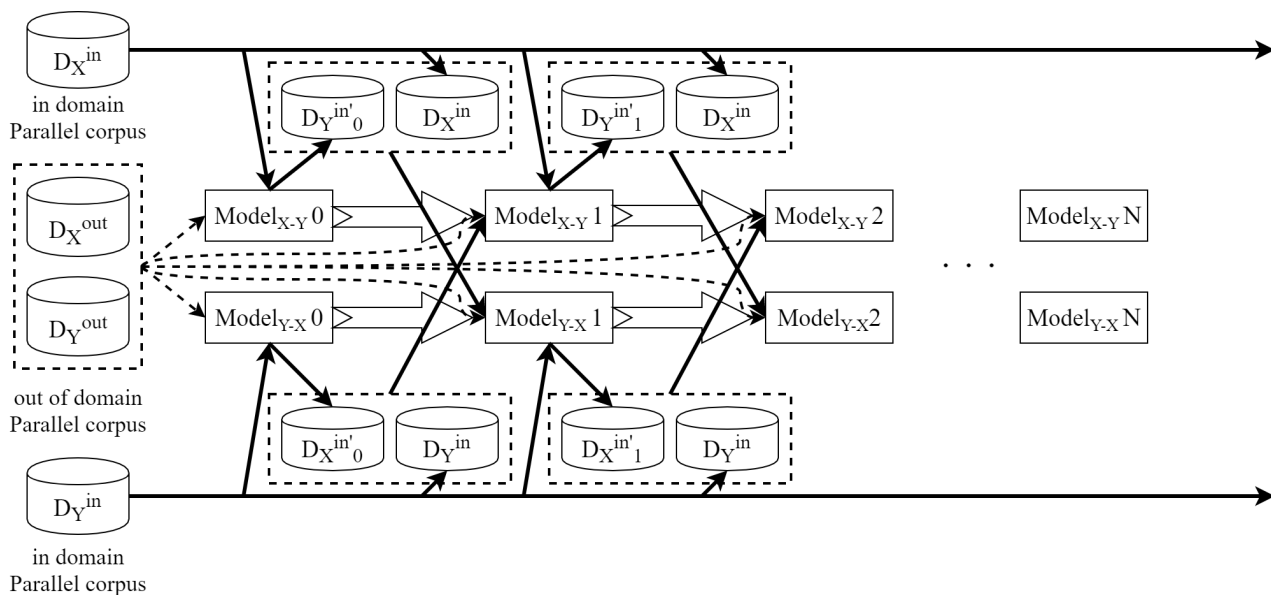


図1 fine-tuning を導入した双方向反復的教師なし適応手法の流れ

4.1 データセット

ドメイン外対訳コーパスとして Asian Scientific Paper Excerpt Corpus(ASPEC)[15] を、ドメイン内の単言語コーパスとして NTCIR8-PATENT を使用した。

ASPEC コーパスには訓練データが約 300 万文対するが、この内、先頭から 100 万文対をドメイン内対訳コーパスとして使用した。

PATENT コーパスは約 318 万文対の訓練データの内、英語と日本語の単言語コーパスが対訳にならないように、先頭の 10 万文対を除いたものを半分に分割し、先頭の 154 万文を英語の単言語コーパス、後ろの 154 万文を日本語の単言語コーパスとして使用した。事前の処理として、両コーパスのすべてのデータを NFKC 正規化し、英語側のコーパスは Moses[16] に付属するトークナイザーでトークナイズしたあと、Moses[16] の truecaser で大文字小文字の表記の統一を行った。また、学習前の事前処理として SentencePiece[17] により各データをサブワード化した。SentencePiece のモデルの学習には、ドメイン外対訳コーパスとドメイン内単言語コーパスを結合したものを用了。

4.2 実験方法

ドメイン外対訳コーパスである ASPEC コーパスを使用し、英日、日英方向の翻訳モデル (Model0) を学習する。次に、学習した日英翻訳モデルで、ドメイン内単言語コーパスである NTCIR8-PATENT コー

パスを逆翻訳し疑似対訳コーパスを作成する。この対訳コーパスと ASPEC コーパスを混合し、新たに英日方向の翻訳システムを学習する。英語の単言語コーパスに対しても同様の手順で疑似対訳コーパスを生成し、新たに日英翻訳モデルを学習する。これを Model1 とし、以降翻訳システムを学習するごとに Model2, Model3, ...ModelN とする。ただし、Model2 以降は、翻訳システムを学習する際に、一つ前の同方向のモデルから fine-tuning を行う。Model2 の英日翻訳システムは、Model1 の英日翻訳システムから fine-tuning し、日英方向の翻訳モデルは Model1 の日英翻訳モデルから fine-tuning する。Model3 の翻訳システムも同様に Model2 の翻訳システムから fine-tuning し、以降これを繰り返す。

ベースラインとして、fine-tuning を行わない通常の繰り返し学習を行ったモデルと性能を比較する。また、逆翻訳の方法による疑似対訳コーパスの質の違いも調べるため、以下の 5 つのモデルで性能の比較を行う。

Oracle NTCIR コーパス約 300 万文を完全な対訳コーパスとして ASPEC コーパスと結合し通常の教師あり学習を行う。

Baseline fine-tuning を行わず、通常の繰り返し学習を行う [4]。beam search により逆翻訳を行う。

Best Sampling 通常の beam search により逆翻訳を行う。

Random Sampling 尤度の高い文ではなく、モデルの事後確率による random sampling により文を

生成し逆翻訳を行う。

Combining Sampling 上記の二つの方法を組み合わせて疑似対訳コーパスを作成する。英日方向では、単言語コーパスの偶数行を random sampling, 奇数行を beam search により逆翻訳し、日英方向では奇数行を random sampling, 偶数行を beam search により逆翻訳する。ただし、翻訳システムの更新が Model1, Model2, ...ModelN と進むごとに偶数行と奇数行を入れ替える。

4.3 実験条件

翻訳システムとして、OpenNMT-py[18] の Transformer を使用した。エンコーダ、デコーダともに6層とし、隠れ層の次元を512とした。最適化にはAdamを使用した。サブワードの分割サイズは16Kとし、OpenNMT-py側の語彙サイズは30Kとした。学習ステップは、ASPECコーパスのみから学習を行う初期モデルは10000ステップ、以降のASPECコーパスとPATENTコーパスから学習を行うモデルは25000ステップとした。また、初期モデルでは1000ステップ、以降のモデルでは2500ステップごとに翻訳モデルを保存し、性能評価はBLEU[19]により行った。英日翻訳では、テストセットの翻訳結果をデトークナイズしたあとに、Meachにより分かち書きし、評価を行った。開発データでのBLEUが最も高いモデルをテストデータでの性能評価、及び単言語コーパスの逆翻訳に用いた。

4.4 実験結果

表1 各条件でのモデルのBLEUの最大値
NTCIR

model	En-Ja	Ja-En
ドメイン適応なし (Model 0)	17.45	18.01
Senrich[1](Model 1)	31.68	31.68
Baseline	35.91	35.44
Fine tuning + Best Sampling	38.50	33.15
Fine tuning + Random Sampling	34.13	30.91
Fine tuning + Combining Sampling	40.58	38.22
Oracle	41.95	40.04

各条件において Model9 まで学習したなかで、NTCIR コーパスの開発データに対する BLEU 値が最も高い Model のテストデータに対する BLEU を表1に示す。各条件でのすべてのモデルの BLEU 値は付録に示す。fine-tuning を行わないモデルに比べ、beam search による fine-tuning を行ったモデル (Fine tuning + Best Sampling) では英日方向では+2.59 BLEU

値が向上したが、日英方向では-2.29 BLEU 値が下がった。しかし、二つの逆翻訳手法を組み合わせたモデル (Combining Sampling) は最も大きく BLEU 値が向上し、fine-tuning を行わないモデルに比べ英日方向で+4.67, 日英方向で+2.78 BLEU 値が向上した。

random sampling で逆翻訳をしたモデルではベースラインよりも BLEU 値が低くなっている。これは、random sampling では逆翻訳文の質が低下し、ステップが進むごとに疑似対訳コーパスの質が低くなるためであると考えられる。同様の結果は、先行研究 [14] において、初期モデルの性能が低い場合は random Sampling が有効でないことが示されている。

一方で Combining Sampling で翻訳精度が大きく向上していることから、beam search による最もノイズの少ない翻訳文を混ぜることにより、初期モデルの精度が低い場合でも疑似対訳コーパスの質の低下を防ぐことができていると考えられる。

以上から、Combining Sampling は、従来の逆翻訳手法 [10][11][12][13] では翻訳精度の改善が困難であった初期モデルの精度が低い場合でも翻訳精度を改善することができると言える。

ただし、今回は beam search と Random sampling による逆翻訳文の割合を 1:1 として実験を行ったがより最適な混合割合がある可能性は残っている。

5 おわりに

本論文では、繰り返し学習によるドメイン適応手法で、fine-tuning を行うことでさらに翻訳精度が向上することを確認した。これは、異なる系列の翻訳モデルの知見を利用することができるためであると考えられ、手法は異なるが藤沢ら [5] の主張と一致する。また、疑似対訳コーパスを生成する際に beam search と random sampling による逆翻訳した文を混合することで翻訳精度がより大きく改善され、通常の教師あり学習を行った Oracle モデルに迫る翻訳精度を得た。

今後の研究としては beam search と random sampling による逆翻訳文の適切な混合割合を調査したい。また、ドメイン外対訳コーパスに対しても同様に逆翻訳文を混合することで翻訳精度が向上するかを実験したい。

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] Rico Sennrich, Barry Haddow, and Alexandra Birch. Improving neural machine translation models with monolingual data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 86–96, Berlin, Germany, August 2016. Association for Computational Linguistics. <https://www.aclweb.org/anthology/P16-1009>.
- [2] Vu Cong Duy Hoang, Philipp Koehn, Gholamreza Haffari, and Trevor Cohn. Iterative back-translation for neural machine translation. In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*, pp. 18–24, Melbourne, Australia, July 2018. Association for Computational Linguistics. <https://www.aclweb.org/anthology/W18-2703>.
- [3] Zhirui Zhang, Shujie Liu, Mu Li, Ming Zhou, and Enhong Chen. Joint training for neural machine translation models with monolingual data. *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32, No. 1, Apr. 2018. <https://ojs.aaai.org/index.php/AAAI/article/view/11248>.
- [4] 森田知熙, 秋葉友良, 塚田元. 双方向の逆翻訳を利用したニューラル機械翻訳の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. 1451–1454. Information Processing Society of Japan, March 2019. https://www.anlp.jp/proceedings/annual_meeting/2019/pdf_dir/A6-1.pdf.
- [5] 藤澤兼太, 秋葉友良, 塚田元. 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. 言語処理学会第26回年次大会 発表論文集, pp. 744–747. Association for Natural Language Processing, March 2020. https://www.anlp.jp/proceedings/annual_meeting/2020/pdf_dir/P3-26.pdf.
- [6] Di Jin, Zhijing Jin, Joey Tianyi Zhou, and Peter Szolovits. A simple baseline to semi-supervised domain adaptation for machine translation, 2020. <https://arxiv.org/abs/2001.08140>.
- [7] Çağlar Gülçehre, Orhan Firat, Kelvin Xu, Kyunghyun Cho, Łoic Barrault, Hwei-Chi Lin, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. On using monolingual corpora in neural machine translation. *CoRR*, Vol. abs/1503.03535, , 2015. <http://arxiv.org/abs/1503.03535>.
- [8] Yong Cheng, Wei Xu, Zhongjun He, Wei He, Hua Wu, Maosong Sun, and Yang Liu. Semi-supervised learning for neural machine translation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1965–1974, Berlin, Germany, August 2016. Association for Computational Linguistics. <https://www.aclweb.org/anthology/P16-1185>.
- [9] Anna Currey, Antonio Valerio Miceli Barone, and Kenneth Heafield. Copied monolingual data improves low-resource neural machine translation. In *Proceedings of the Second Conference on Machine Translation*, pp. 148–156, Copenhagen, Denmark, September 2017. Association for Computational Linguistics. <https://www.aclweb.org/anthology/W17-4715>.
- [10] 今村賢治, 藤田篤, 隅田英一郎. 単言語コーパスと逆翻訳を用いたエンコーダー・デコーダーの訓練法. 言語処理学会第24回年次大会 発表論文集, pp. 29–32. Information Processing Society of Japan, March 2018. https://www.anlp.jp/proceedings/annual_meeting/2018/pdf_dir/B1-3.pdf.
- [11] Kenji Imamura, Atsushi Fujita, and Eiichiro Sumita. Enhancement of encoder and attention using target monolingual corpora in neural machine translation. In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*, pp. 55–63, Melbourne, Australia, July 2018. Association for Computational Linguistics. <https://www.aclweb.org/anthology/W18-2707>.
- [12] 今村賢治, 藤田篤, 隅田英一郎. サンプリング生成に基づく複数逆翻訳を用いたニューラル機械翻訳. 人工知能学会論文誌, Vol. 35, No. 3, pp. A-JA9_1–9, 2020. https://www.jstage.jst.go.jp/article/tjsai/35/3/35_A-JA9/_article/-char/ja/.
- [13] Miguel Graça, Yunsu Kim, Julian Schamper, Shahram Khadivi, and Hermann Ney. Generalizing back-translation in neural machine translation. In *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)*, pp. 45–52, Florence, Italy, August 2019. Association for Computational Linguistics. <https://www.aclweb.org/anthology/W19-5205>.
- [14] Sergey Edunov, Myle Ott, Michael Auli, and David Grangier. Understanding back-translation at scale. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 489–500, Brussels, Belgium, October–November 2018. Association for Computational Linguistics. <https://www.aclweb.org/anthology/D18-1045>.
- [15] Toshiaki Nakazawa, Manabu Yaguchi, Kiyotaka Uchimoto, Masao Utiyama, Eiichiro Sumita, Sadao Kurohashi, and Hitoshi Isahara. ASPEC: Asian scientific paper excerpt corpus. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pp. 2204–2208, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA). <https://www.aclweb.org/anthology/L16-1350>.
- [16] Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pp. 177–180, Prague, Czech Republic, June 2007. Association for Computational Linguistics. <https://www.aclweb.org/anthology/P07-2045>.
- [17] Taku Kudo and John 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, Brussels, Belgium, November 2018. Association for Computational Linguistics. <https://www.aclweb.org/anthology/D18-2012>.
- [18] Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander M. Rush. Opennmt: Open-source toolkit for neural machine translation. *CoRR*, Vol. abs/1701.02810, , 2017. <http://arxiv.org/abs/1701.02810>.
- [19] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics. <https://www.aclweb.org/anthology/P02-1040>.

A 付録

表 2 Fine tuning + Best Sampling の各繰り返しごとの BLEU 値
ASPEC NTCIR

Model	ASPEC				NTCIR			
	dev		test		dev		test	
	En-Ja	Ja-En	En-Ja	Ja-En	En-Ja	Ja-En	En-Ja	Ja-En
Model 1	39.24	25.93	38.84	25.99	33.33	28.63	32.78	29.00
Model 2	40.25	26.44	39.98	26.63	37.33	32.39	36.84	32.35
Model 3	40.64	26.81	40.82	27.06	37.83	32.78	37.71	33.19
Model 4	40.68	27.20	41.11	27.05	38.57	32.90	38.21	33.22
Model 5	40.94	26.97	41.32	27.39	38.58	32.93	38.53	33.40
Model 6	41.19	27.02	41.12	27.55	38.36	33.20	38.64	33.63
Model 7	40.88	27.15	41.08	27.04	38.53	33.18	37.90	33.22
Model 8	41.36	26.95	41.45	27.19	38.59	33.21	38.50	33.15
Model 9	41.33	27.07	41.38	27.22	38.50	33.18	38.45	33.15

表 3 Fine tuning + Random Sampling の各繰り返しごとの BLEU 値
ASPEC NTCIR

Model	ASPEC				NTCIR			
	dev		test		dev		test	
	En-Ja	Ja-En	En-Ja	Ja-En	En-Ja	Ja-En	En-Ja	Ja-En
Model 1	38.71	27.67	38.90	27.83	32.77	28.90	32.57	30.72
Model 2	39.99	26.43	40.15	27.03	33.69	29.54	33.29	30.91
Model 3	40.22	26.88	40.83	27.65	34.58	28.78	34.13	29.69
Model 4	40.77	26.73	41.18	27.80	32.66	28.82	32.77	30.95
Model 5	40.84	26.78	41.05	27.55	33.63	27.91	33.90	30.19
Model 6	40.79	26.91	41.07	28.19	32.59	27.39	31.48	29.37
Model 7	41.14	27.21	41.35	28.42	32.28	27.97	32.06	29.53
Model 8	41.50	27.31	41.67	27.96	32.14	25.95	32.18	26.95
Model 9	41.26	27.23	41.53	28.17	30.38	27.30	30.26	29.03

表 4 Fine tuning + Combining Sampling の各繰り返しごとの BLEU 値
ASPEC NTCIR

Model	ASPEC				NTCIR			
	dev		test		dev		test	
	En-Ja	Ja-En	En-Ja	Ja-En	En-Ja	Ja-En	En-Ja	Ja-En
Model 1	37.70	27.07	37.21	27.61	34.56	30.57	33.60	31.64
Model 2	39.12	27.99	39.20	28.71	37.24	33.03	36.86	34.22
Model 3	39.52	28.38	39.13	29.18	38.66	35.20	38.30	37.03
Model 4	39.89	28.40	39.95	29.24	39.46	35.07	39.00	36.61
Model 5	40.51	28.50	40.58	29.42	39.70	36.51	39.69	38.06
Model 6	40.67	28.88	40.70	29.10	40.10	36.62	40.18	37.91
Model 7	40.78	29.05	40.63	28.89	40.18	36.72	40.03	38.34
Model 8	41.01	29.00	40.66	29.38	40.39	36.76	40.58	37.94
Model 9	41.15	28.94	40.88	29.30	40.19	36.88	40.21	38.22