

TED 講演の英日翻訳と日英翻訳の検討

足立十一郎 山本一公 中川聖一

中部大学 工学部 情報工学科

{ep19002-1545@sti, kazumasayamamoto@isc, nakagawa@isc}.chubu.ac.jp

概要

我々は、MIT の講義や TED 講演を対象に、テキスト入力や音声入力の英日機械翻訳の研究を行ってきた。本稿では、Transformer の翻訳モデルを用いて、英日・日英共通の双方向モデル、TED の英-西の英語側コーパス、および日本語話し言葉コーパス CSJ の模擬講演を使用したデータ拡張、大量の書き言葉対訳コーパスである ASPEC を用いた転移学習などによる改善について報告する。また、TED 講演の BLEU 値が如何に低いかを複数人による翻訳結果の比較と人間による評価により検討した結果について報告する。

1 はじめに

我々は、MIT の講義や TED 講演を対象に、テキスト入力や音声入力の英日機械翻訳の研究を行ってきた[1,2,3]。機械翻訳技術の進展とともに、統計的機械翻訳手法、系列変換によるニューラル機械翻訳手法を検討してきたが、今回、Transformer による翻訳手法を導入し、改善を図ったので報告する。

TED 講演の英日機械翻訳は、講演が話し言葉であるため文単位が曖昧であり、また学習に使用するパラレルデータが少ないという問題があり、100 万ペアある ASPEC のような書き言葉の翻訳と比べて格段に難しい[2,3]。TED 講演の英日翻訳に関しては、IWSLT のワークショップで、2021 年には英日の瞬時テキスト翻訳が、2022 年には英日の瞬時音声翻訳がタスクに加えられている[4,5]。Fukuda らは、Transformer の学習に TED の英日 22 万文のパラレルコーパス以外に、WMT20 news の 1800 万文ペアを用いて、IWSLT2021 Dev セットに対してテキスト入力のオフライン翻訳で 16.8、瞬時翻訳で 15.6 の BLEU 値を得ている[6]。瞬時翻訳は、オフライン翻

訳に比べて BLEU 値で 10%~20%程度低下する。また、音声入力の音声翻訳は、音声認識性能に大きく依存するが、テキスト翻訳に比べて、BLEU 値で 10%~20%低下する[4,5]。

パラレルデータ不足を補うために単言語コーパスによるデータ拡張手法がある[7]。これは、単言語コーパスを翻訳することで疑似的なパラレルコーパスを作成し、ベースとなるパラレルコーパスと混合して学習に用いる手法である。山岸らは、本手法の単言語コーパスとして、日本語話し言葉コーパス CSJ の学会講演の書き起こしを使用した[8]。本研究では、TED の英語-スペイン語のパラレルコーパスの英語側の単言語と CSJ の日本語単言語の模擬講演コーパスを使用する。

同じく山岸らは、Fukuda らと同様に、ドメインの異なる大規模パラレルコーパスを用いた転移学習を TED 翻訳に適用した[8]。具体的には、IWSLT2010 の TED のテストデータに対して、書き言葉の 100 万文英日ペアの ASPEC や字幕データの 280 万文ペアの JESC からの転移学習で、それぞれ 14.19, 15.47 の BLEU 値を得ている[7,9]。本研究でもこの手法を利用する。

低リソース言語のモデルの学習手法として、マルチリンガル学習がある[10]。この手法は、多言語のパラレルデータで多言語翻訳モデルを学習し、エンコーダの出力である表現を汎用的な意味空間に写像するものと考えられ、小規模のパラレルデータでこのモデルを適応学習する方法である。この類似な方法として、双方向翻訳モデルが提案されている[11]。これは、対訳コーパスからソースとターゲットを順逆両方向の2つのパラレルデータを構築して、順逆両方向の翻訳モデルを共通なモデルとして学習する。Thanh らは、TED 講演の英語-ベトナム語の双方向翻訳に適用し、有効性を示している[12]。本研究ではパ

ラレルコーパスの規模とこの手法の有効性との関係を明らかにする。最後に TED 講演の翻訳の困難さを BLEU と人手による評価で検討する。

2 翻訳モデル

2.1 LSTM モデル [13]

LSTM(Long-short Term Memory)はある入力単語シーケンス $\{x_1, x_2, \dots, x_i\}$ から別の出力単語シーケンス $\{y_1, y_2, \dots, y_j\}$ に変換するリカレントネットワークによる sequence-to-sequence モデルである。エンコーダとデコーダは LSTM によって構成される。LSTM とは再帰型ニューラルネットワーク(RNN)の一種で時系列データも考慮したモデルである。LSTM は忘却ゲート、入力ゲート、出力ゲートによって RNN で発生する勾配消失問題の改善を行っている。

本研究では LSTM モデルの翻訳モデルとして 1 層の双方向の LSTM をエンコーダ、1 層の単方向の LSTM をデコーダに利用する。またエンコーダの出力の全体ではなく一部に注目する注意機構を用いた。

2.2 Transformer モデル [14]

Transformer モデルは再帰型ニューラルネットワークの様に時系列データを用いて学習を行う。出力を求める際は RNN ではなくセルフアテンション機構を用いる。エンコーダは同じ構成のエンコーダの積み重ねによって構成されており、それらのエンコーダはセルフアテンションとフィードフォワードネットワーク(FFNN)により構成される。デコーダも同じ構成のデコーダが積み重ねられている。一つのデコーダはセルフアテンションと FFNN に加えてその間に注意機構が入っている。標準モデルはエンコーダ 6 層、デコーダ 6 層であるが、最適な数は学習データ量に依存する。IWSLT の 22 万文ペアの英日パラレルコーパスで 20 epoch 学習した場合は、6 層-6 層で 0.80, 5 層-5 層で 1.94, 4 層-4 層で 10.36, 3 層-4 層で 13.66, 3 層-3 層で 13.22 の BLEU 値であった。本研究ではエンコーダとデコーダは 3 層とした。

2.3 単方向翻訳モデルと双方向翻訳モデル

単方向翻訳モデルの場合は、原言語と目的言語のペアで学習することで翻訳モデルを作成する。双方向翻訳モデルの場合は原言語と目的言語のペアとそれを入れ替えた目的言語と原言語のペアから成るデータセットによって学習する。一つの共通モデルで

英日、日英翻訳を実現する。

3 翻訳モデルの改善

3.1 データ拡張

本実験で使用する TED 講演の IWSLT (International Workshop on Spoken Language Translation) のコーパスには英語と日本語の対訳コーパスが少ないため、英語または日本語の単言語コーパスをベースモデル g で翻訳することで英語と日本語の疑似対訳コーパスを作成し、翻訳モデル学習の追加の学習データとすることでデータ拡張を行う。データ拡張には IWSLT2018 英語-スペイン語ペアの英語側コーパスと CSJ(Corpus of Spontaneous Japanese)日本語コーパスの模擬講演を用いた。

3.2 ASPEC(書き言葉)からの転移学習

100 万文ペアからなる ASPEC (Asian Scientific Paper Excerpt Corpus)コーパスにより単方向翻訳モデル、双方向翻訳モデルを学習し、そのパラメータを初期値として IWSLT 及びデータ拡張したデータセットで翻訳モデルを学習する。なお、ASPEC コーパスによる翻訳モデル学習の際には、TED 語彙ファイルに合わせて学習を行った。

4 翻訳実験

4.1 使用するデータ

本研究では IWSLT2016 英日対訳コーパスを使用した。単言語コーパスとして IWSLT2018 英語-スペイン語ペアの英語側と日本語の単言語コーパスである CSJ をデータ拡張に用いた。表 1 に学習文数、開発文数、テスト文数を示す。

4.2 各翻訳モデルの仕様

LSTM モデルの学習には OpenNMT[15]Ver. 0.7.0 を使用した。エンコーダは 1 層の双方向 LSTM、デコーダは 1 層の単方向 LSTM を使用した。LSTM の隠れ層の次元数と単語分散表現の次元数は 1024 とした。学習の最適化は SGD、学習率はファインチューニング時は 0.01562、それ以外は 1.0、学習 step 数はファインチューニング時は開発データで一番性能が良かった step 数を選び、それ以外は単方向翻訳モデルでは 80000、双方向翻訳モデルでは 100000 とした。

表 2: TED 講演の翻訳実験結果(BLEU) (無印は 5 つのモデルの平均値、*印は単一の固定 seed モデル)

| 翻訳モデル | 使用データセット | LSTM | | Transformer | |
|----------------------|----------------|--------|--------|-------------|--------|
| | | 英日翻訳 | 日英翻訳 | 英日翻訳 | 日英翻訳 |
| 単方向 | ベースライン | 9.26 | 9.21 | 13.68* | 12.15* |
| 双方向 | ベースライン | 10.43 | 10.37 | 13.43* | 12.12* |
| ASPEC 転移学習 単方向 | ベースライン | 9.52* | 9.94* | 14.42* | 14.10* |
| | +IWSLT2018+CSJ | 10.43* | 11.24* | 14.66* | 13.58* |
| ASPEC 転移学習 双方向 | ベースライン | 11.19* | 11.14* | 13.65* | 12.37* |
| | +IWSLT2018+CSJ | 12.36* | 11.76* | 14.39* | 12.65* |
| | +fine tuning | 11.95* | 11.67* | 15.02* | 14.18* |

表 1: 使用データセット

| データセット | 学習文数 | 開発文数 | テスト文数 |
|-----------------------|----------------------|-----------------|------------------|
| IWSLT2016 (英語-日本語) | 223,108 (1863 講演) | 871 (8 講演) | 1,194 (12 講演) |
| IWSLT2018 (英語) | 80,222 (862 講演) | | |
| CSJ (日本語) | 21,9229 (1565 講演) | | |
| ASPEC (英語-日本語) | 1,000,000 (24 分野) | 1,790 (2 分野) | 1,812 (24 分野) |

バッチサイズは 64 で epoch 数に換算すると単方向のベースラインは約 23epoch である。語彙サイズはソース・ターゲットそれぞれ 50000 語とした。

Transformer モデルの学習には fairseq[16]Ver.0.12.2 を使用した。エンコーダとデコーダの層数はそれぞれ 3 層、単語分散表現の次元数は 1024 とした。学習の最適化は Adam、学習率は 0.0005、学習 epoch 数は 20 まで学習した。

4.3 評価方法

翻訳モデルの性能評価には BLEU を用いる。本研究では 4-gram までの値を計算し、翻訳モデルの評価に使用した。また、LSTM モデルについては一部を除いてランダム初期値のため、5 つのモデルを学習して BLEU 値の平均を求めた。

TED の日本語訳を用いた参照訳は直訳ではなく意識が多く、機械による訳と相当異なる結果になる。そこで、翻訳業者 A と B に依頼して、新たな参照文として使用し、比較する。

5 翻訳実験結果

5.1 翻訳モデルの比較

LSTM と Transformer による各翻訳モデルの主な評価結果を表 2 に示す(詳細な結果は付録に示す)。LSTM と Transformer を比べると全体で Transformer の方が BLEU 値で 3 程度性能が良い結果となった。

5.2 単方向翻訳モデルと双方向翻訳モデル

LSTM モデルではベースラインから双方向翻訳モデルにすることによって BLEU 値が+1.17 改善された。Transformer モデルでは単方向翻訳モデルと双方向翻訳モデルを比較すると BLEU 値の差はあまりなかった。単方向翻訳モデルと双方向翻訳モデルの優劣は学習データ量に依存するかを調べるために ASPEC で評価した。表 3 に結果を示す。Transformer ではすべての場合で単方向翻訳モデルが良かったが、LSTM モデルにおいて 100 万文では単方向翻訳モデルが良く、50 万文では同等、25 万文の場合は双方向翻訳モデルが性能を上回った。

表 3: ASPEC による単方向、双方向翻訳モデルの翻訳結果

| 学習データ | LSTM | transformer |
|------------|-------------|-------------|
| | 英日 / 日英翻訳 | 英日 / 日英翻訳 |
| 100 万文、単方向 | 36.0 / 26.6 | 39.4 / 27.5 |
| 100 万文、双方向 | 34.3 / 25.1 | 38.0 / 26.4 |
| 50 万文、単方向 | 33.1 / 24.1 | 38.0 / 26.6 |
| 50 万文、双方向 | 33.6 / 23.9 | 36.1 / 25.2 |
| 25 万文、単方向 | 28.9 / 19.9 | 35.1 / 24.6 |
| 25 万文、双方向 | 30.2 / 22.0 | 34.2 / 24.4 |

5.3 単言語データからのデータ拡張

— 英語コーパス 対 日本語コーパス —

(a)英日翻訳

表 2 の「英日翻訳」欄に各モデルの英日翻訳結果

を示す。LSTM モデルの単方向翻訳モデルではデータ拡張を行うことでベースラインから BLEU 値が改善されることが分かる。IWSLT2018 と CSJ によるデータ拡張を比較すると、CSJ によるデータ拡張の方が良くなった。双方向翻訳モデルの場合は、CSJ よりも IWSLT2018 によるデータ拡張の方が良い結果となった。Transformer モデルでは双方向翻訳モデル+IWSLT2018 によるデータ拡張で BLEU 値が向上した。

データ拡張による学習後、元の英日コーパス（英日翻訳用）もしくは日英コーパス（日英翻訳用）でファインチューニングをすると、LSTM モデルでも Transformer モデルでも IWSLT2018 によるデータ拡張をしたモデルで改善が見られた。

(b) 日英翻訳

表 2 の「日英翻訳」欄に各モデルの日英翻訳結果を示す。LSTM の双方向翻訳モデルでは IWSLT2018 のデータ拡張により多くの場合で BLEU 値が改善された。一方、CSJ によるデータ拡張ではあまり改善が見られなかった。Transformer モデルの双方向翻訳モデルでは IWSLT2018 によるデータ拡張で BLEU 値が改善された。

ファインチューニングをすると、英日翻訳の結果と同じ傾向で LSTM モデルでも Transformer モデルでも IWSLT2018 によるデータ拡張をしたモデルで改善が見られた。

5.4 ASPEC モデルからの転移学習

表 2 に示すように ASPEC モデルからの転移学習で、LSTM でも Transformer モデルでも多くの場合で BLEU 値が改善された。データ拡張と fine tuning により日英翻訳で 15.02 の BLEU 値が得られた。

5.5 人手による評価

(a) 二つの翻訳業者 A, B によって翻訳された IWSLT の 12 講演のうち 3 講演のテスト文 254 文を参照文とした場合と IWSLT2016 (TED) 参照文と比較した。また、TED 参照文に翻訳業者 A, B の翻訳文を加えたマルチリファレンスとした結果を表 5 示す。翻訳業者同士の比較でも BLEU 値は低い。特に翻訳業者 A の BLEU 値が低い。マルチリファレンスでの BLEU 値は 24.03 と大きく改善した。TED の翻訳は人間でも揺れが大きい。TED 参照文は意識であり、機械翻訳に対するターゲットとしては必ずしも適切とは言えない。IWSLT2021 では、直訳に近い参照文でも評

価され、BLEU スコアが 10%~20%向上、2つのマルチリファレンスだと 50%以上向上している[4]。我々の場合は、3つのマルチリファレンスで 100%以上向上した。

(b) 二つの翻訳業者による翻訳文と Transformer モデル (CSJ によるデータ拡張+ASPEC による転移学習をした単方向翻訳モデル) による機械翻訳結果を、TED 参照文と比較して男女 2 名 (C, D) の人手によって 1 文ずつ 5 点満点で評価した結果を表 6 に示す。評価基準を以下に示す。

- 5: TED 参照文の意味と同じ意味である。
- 4: TED 参照文の意味と概ね同じ意味である。
- 3: TED:参照文の意味と大まかには似ており、意図はほぼ通じる。
- 2: TED 参照文の意味と一部やや外れている。
- 1 TED:参照文の意味とかなり離れていて、意図は正しく伝わらない。

結果は、翻訳業者 B > 翻訳業者 A > Transformer と、表 4 の BLEU 値と同じ傾向であり、BLEU 値と人手による評価は相関があると思われる。

表 5: 人手による翻訳との比較 (TED 3 講演)

| 翻訳文 | 参照文 | BLEU |
|-------------|-----------|-------|
| 翻訳業者 A | TED 参照文 | 13.75 |
| 翻訳業者 B | TED 参照文 | 27.85 |
| Transformer | TED 参照文 | 11.23 |
| Transformer | 翻訳業者 A | 13.05 |
| Transformer | 翻訳業者 B | 11.87 |
| Transformer | マルチリファレンス | 24.03 |

表 6: 人手による TED 参照文と翻訳結果の比較

| 翻訳文 \ 評価者 | 男 C | 女 D | 平均 |
|-------------|------|------|------|
| 翻訳業者 A | 3.97 | 3.86 | 3.92 |
| 翻訳業者 B | 4.30 | 4.26 | 4.28 |
| Transformer | 2.87 | 3.01 | 2.94 |

6 まとめ

本研究では、22 万文ペアと小規模データでは双方向翻訳モデルが単方向翻訳モデルよりも翻訳性能が向上した。データ拡張やファインチューニング、転移学習を組み合わせるとさらに翻訳性能が向上した。ただし、Transformer のベースラインの性能が高く、Transformer に対する改善効果は比較的小さい。最良モデルの英日で 15.02、日英で 15.49 の BLEU 値が得られた。

謝辞

本研究は、豊橋技術科学大学の秋葉友良准教授および秋葉研究室の学生諸君にご協力をいただいた。また、JSPS 科研費 18H01062 の研究助成を受けた。

参考文献

- [1] 後藤 統興, 山本 一公, 中川 聖一: 音声認識誤りを考慮した英語講義音声の日本語への音声翻訳システムの検討, 言語処理学会第 23 回年次大会, pp.1054-1057, 2017.
- [2] 佐橋 広也, 秋葉 友良, 中川 聖一: 科学技術論文抄録と講義の英日機械翻訳のリスクアラインの検討, 言語処理学会第 25 回年次大会, pp. 1165-1168, 2019.
- [3] 松永 凌: トピック分散表現と逆翻訳によるデータ拡張を用いた英日機械翻訳, 中部大学卒業論文, 2022.
- [4] Antonios Anastasopoulos, Ondřej Bojar, et.al: Findings of the IWSLT 2021 evaluation campaign, Proc. IWSLT-2021, pp.1-29, 2021.
- [5] Antonios Anastasopoulos, Loïc Barrault, et.al: Findings of the IWSLT 2022 evaluation campaign, Proc. IWSLT-2022, pp. 98-158, 2022.
- [6] Ryo Fukuda, et al.: NAIST English-to-Japanese simultaneous translation system for IWSLT 2021 simultaneous text-to-text task, Proc. IWSLT-2021, pp. 39-45, 2021.
- [7] Rico Sennrich, Barry Haddow, Alexandra Birch: Improving neural machine translation models with monolingual data, Proc. ACL, pp. 86-96, 2016.
- [8] 山岸 勇輝, 秋葉 友良, 塚田 元: 逆翻訳を用いたデータ拡張と転移学習を利用した英日講演字幕翻訳の改善, 言語処理学会第 26 回全国大会, pp.1277-1280, 2020.
- [9] 山岸 勇輝, 秋葉 勇輝, 塚田 元: 入力側単言語資源と転移学習を利用による講演字幕を対象とした英日ニューラル機械翻訳の改善, 言語処理学会第 28 回全国大会, pp.1317-1321, 2022.
- [10] Raj Dabre, Chenhui Chu, Anoop Kunchukuttan: A survey of multilingual neural machine translation, ACM Computer Surveys, vol.53, no.3, 2020.
- [11] Liang Ding, Di Wu, Dachen Tao: Improving neural machine translation by bidirectional training, arXiv:2109.07780, 2021.
- [12] Bui Tuan Thanh, 秋葉 友良, 塚田 元: 双方向翻訳モデルと反復逆翻訳を用いた低資源言語に対するニューラル機械翻訳の性能向上, 言語処理学会第 28 回全国大会, pp. 1756-1760, 2022.
- [13] Ilya Sutskever, Oriol Vinyals, Quoc V. Le: Sequence to sequence learning with neural networks, Advances in Neural Information Processing Systems (NIPS 2014), 27, 2014.
- [14] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin: Attention is all you need. Conference on Neural Information Processing Systems (NIPS2017), 30, 2017.
- [15] Open Source Neural Machine Translation in PyTorch, 2023 年 1 月 12 日, <https://github.com/OpenNMT/OpenNMT-py>.
- [16] Facebook AI Research Sequence-to-Sequence Toolkit written in Python, 2023 年 1 月 12 日, <https://github.com/facebookresearch/fairseq>.

付録 A: TED 講演の翻訳実験結果 (BLEU) (無印は 5 つのモデルの平均値、*印は単一の固定 seed モデル)

| 翻訳モデル | 使用データセット | LSTM | | Transformer | |
|---------------------|-------------------------|--------|--------|-------------|--------|
| | | 英日翻訳 | 日英翻訳 | 英日翻訳 | 日英翻訳 |
| 単方向翻訳モデル | ベースライン | 9.26 | 9.21 | 13.68* | 12.15* |
| | +IWSLT2018(疑似日本語-英語) | 9.64 | 10.71 | 13.50* | 13.69* |
| | +CSJ(疑似英語-日本語) | 10.12 | 9.51 | 13.39* | 12.58* |
| | +IWSLT2018+CSJ | 10.25 | 10.58 | 13.83* | 14.29* |
| 双方向翻訳モデル | ベースライン | 10.43 | 10.37 | 13.43* | 12.12* |
| | +IWSLT2018(疑似日本語-英語) | 10.85 | 11.76 | 13.52* | 13.74* |
| | +IWSLT2018(英語-疑似日本語) | 10.93 | 10.37 | 14.40* | 12.61* |
| | +IWSLT2018(疑似日-英,英-疑似日) | 11.14 | 11.83 | 14.40* | 14.50* |
| | +fine tuning | 11.23* | 12.04* | 14.50* | 15.00* |
| | +CSJ(疑似英語-日本語) | 10.89 | 10.27 | 13.75* | 11.61* |
| | +CSJ(日本語-疑似英語) | 10.76 | 10.43 | 13.67* | 10.78* |
| | +CSJ(疑似英-日、日-疑似英) | 10.82 | 10.35 | 13.47* | 11.39* |
| | +fine tuning | 10.64* | 10.64* | 14.48* | 12.25* |
| + IWSLT2018+CSJ | 11.20 | 11.23 | 14.09* | 12.20* | |
| ASPEC 転移学習 単方向翻訳モデル | ベースライン | 9.52* | 9.94* | 14.42* | 14.10* |
| | +IWSLT2018(疑似日本語-英語) | 9.38* | 12.02* | 14.08* | 15.49* |
| | +CSJ(日本語-疑似英語) | 10.78* | 9.96* | 14.52* | 13.01* |
| | +IWSLT2018+CSJ | 10.43* | 11.24* | 14.66* | 13.58* |
| ASPEC 転移学習 双方向翻訳モデル | ベースライン | 11.19* | 11.14* | 13.65* | 12.37* |
| | +IWSLT2018(疑似日-英,英-疑似日) | 11.65* | 12.57* | 14.39* | 14.61* |
| | +CSJ(疑似英-日、日-疑似英) | 11.94* | 11.21* | 14.43* | 11.73* |
| | +IWSLT2018+CSJ | 12.36* | 11.76* | 14.39* | 12.65* |
| | +fine tuning | 11.95* | 11.67* | 15.02* | 14.18* |

付録 B: 翻訳例

| | |
|--|--|
| 原文 | and so now I' m working on , how can I get these mental images in my mind out to my computer screen faster ? |
| IWSLT2016 参照文 | 今 取り組んでいるのは頭の中にあるイメージを即コンピュータ画面に表示する方法です |
| 業者 A の翻訳文 | 今は、自分の頭の中にあるこれらのメンタルイメージをどうやってより速くコンピュータの画面へと移すことができるか、について取り組んでいます。 |
| 業者 B の翻訳文 | そして今、私はこれらの心の中のイメージをどうすればより速く、コンピュータのスクリーンに映し出せるか？という課題に取り組んでいます。 |
| Transformer 単方向翻訳モデル ベースライン | 今私はコンピュータ画面に対して精神的なイメージを得るにはどうすればいいのでしょうか？ |
| Transformer 双方向翻訳モデル ASPEC 転移学習 +データ拡張+fine tuning | どうやってコンピュータの画面に頭の中でこうした画像を届けることができるのか？ |