科学技術論文抄録と講義音声の英日機械翻訳のリスコアリングの検討

佐橋 広也 † 秋葉 友良 † 中川聖一 ††
† 豊橋技術科学大学 †† 中部大学
{sahashi, akiba}@nlp.cs.tut.ac.jp, nakagawa@tut.jp

1 はじめに

近年, ウェブ上において利用可能な講義映像が増加している (例 MITOpenCourseWare (MITOCW)). しかし, これらの講義映像は一般的に、その言語が母国語でない学生にとって学習意欲や効率を減少させる外国語である。この問題に対して、字幕付きの講義映像は有効である [3]。翻訳タスクにおいては、ニューラル機械翻訳 (NMT) が目覚ましい発展を遂げており、従来の統計的翻訳機械翻訳 (SMT) の性能を上回っている。しかし NMT は SMT に比べ、学習に必要なパラレルコーパスの量が十分でなければ、翻訳性能を向上させることが難しく、翻訳の語彙サイズについても制限を持つ。そのため機械翻訳の候補をリスコアリングし、性能を向上させる研究が複数報告されている。SMT ではラティスデコーディングを使用した翻訳候補のリスコアリングが行われている [4]。NMT では翻訳候補を別の NMT で再度スコアを評価しリスコアリング方法が報告されている [7]。

本稿では先ず大規模な科学技術論文抄録 (ASPEC) パラレルコーパスで学習した NMT と SMT の翻訳結果を比較し、逆翻訳と文ベクトル類似度に基づく 2 つのリスコアリング手法による結果を報告する。次に、パラレルコーパスがない MIT の講義音声翻訳に対して、音声認識誤り対処法及び中規模な TED パラレルコーパスと大規模な ASPEC パラレルコーパスを用いた結果について報告する。

2 翻訳システム

2.1 SMT

SMT の翻訳モデルは原言語の単語列から目的言語の単語列へ翻訳される確率を計算するモデルである (図 1)。翻訳確率は 2 つの言語間の単語またはフレーズ単位で計算される。フレーズ単位の翻訳確率を言語間の翻訳確率とするために、学習コーパスからフレーズテーブルを学習する。原言語文 $\boldsymbol{F}$ に対応する目的言語文 $\boldsymbol{E}$ の単語アライメントを $\boldsymbol{a}$ としたとき、計算式は以下のように表すことができる。

$$\hat{\boldsymbol{E}} = \arg\max_{\boldsymbol{E}} P(\boldsymbol{E}|\boldsymbol{F}) = \arg\max_{\boldsymbol{E}} \frac{P(\boldsymbol{F}|\boldsymbol{E})P(\boldsymbol{E})}{P(\boldsymbol{F})} \quad (1)$$

$$P(\boldsymbol{F}|\boldsymbol{E}) = \sum_{\mathbf{a}} P(\boldsymbol{F}, \boldsymbol{a}|\boldsymbol{E}) \quad (2)$$

ここで、a は単語のアライメントを示し、$P(\boldsymbol{E})$ は目的言語の言語モデル、$P(\boldsymbol{F}|\boldsymbol{E})$ は翻訳モデルと呼ばれる。

以前、我々は、目的言語の言語モデルを種々用意して複数の翻訳文候補をリスコアリングする手法を試みた [5]。

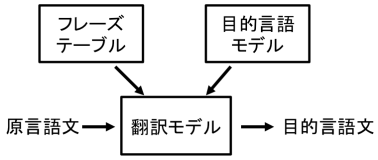

図 1: SMT のブロック図

2.2 NMT

NMT の主流であるエンコーダ-デコーダモデルについて説明する。原言語 $\boldsymbol{F}$ の入力文を単語レベルの埋め込みベクトルに変換してエンコーダへ入力する。エンコーダから出力される分散表現は入力文の意味や構造を捉えた文ベクトルとなる。文ベクトルをデコーダに入力した場合、最初の目的言語の単語 $\boldsymbol{e}_1$ を出力確率によって予測する。次の単語を予測するために、出力された単語を入力として与え、終端記号が予測されるまで単語の予測を繰り返し、最終的に目的言語文 $\boldsymbol{E}$ を出力する (図 2)[1]。単語の予測の際にそれぞれどの原単語に対して注目するかを与えるために、エンコーダから出力される単語ベクトルに重みをかけるアテンション機構によって制御する [2]。θ をモデルのパラメータとしたとき、デコーダの計算式は、以下のように表すことができる。

$$P(\boldsymbol{E}|\boldsymbol{F};\theta) = \prod_{j=1}^{J} P(e_j|\boldsymbol{F}, \boldsymbol{E} < j; \theta) \quad (3)$$

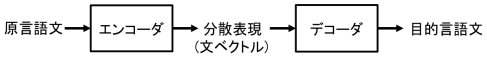

図 2: NMT のブロック図

3 リスコアリング

3.1 ベクトル空間によるリスコアリング

リスコアリングの評価指標として、目的言語文の翻訳候補文と、原言語文の入力文の２つの文ベクトルの類似度は有用であると考えられる。しかし、入力文のベクトルと翻訳文のベクトルの類似度によって翻訳候補をリスコアリングする際、両言語のベクトル空間が同一の意味空間になっている必要がある。そこで原言語のベクトル表現と目的言語のベクトル表現を同一の意味空間に写像する二つの手法を提案する。最終的には、これらの二つの手法を組み合わせて使用する。

(a) 英日同一分散表現の学習

一つ目の手法は同一の意味空間へ写像するために、複数の種類の言語を入力とした NMT を作成する。一つは通常の翻訳システムである原言語文から目的言語文を予測するモデル (および目的言語文から原言語文を予測するモデル)、もう一つはオートエンコーダのシステムで、目的言語文から目的言語文を予測するモデルである。この２つのモデルは同じ目的言語文を予測するため、エンコーダの最終状態の隠れベクトル (文ベクトル) 表現は同じである (図 3)。そのため２種類のモデルを同一のエンコーダデコーダモデルで学習する。これによって原言語のベクトル表現と目的言語のベクトル表現を同一に写像することが可能となる。但し、この同一空間ベクトル間での日英の文の類似度は実験の結果、不十分であるので次の日英間の文ベクトルのマッピングで高精度化を図る。

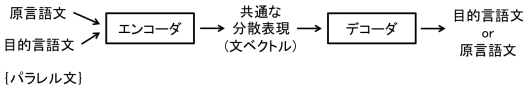

図 3: 同一の分散表現を持つエンコーダ

(b) 英日分散表現のマッピング [8]

二つ目の手法として原言語ベクトルを同一意味空間の目的言語ベクトルに変換するために線形変換と非線形変換 (ニューラルネットワーク) を使用する (図 4)。原言語の文ベクトルを入力とし、目的言語の文ベクトルを出力とするために文ベクトルの両言語ペアを学習に使用する。文献 [1] では単語レベルの線形マッピングを試みている。本研究では非線形変換を 3 層の隠れ層を持つニューラルネットワークで実現した。本マッピング手法だけでも日英の文ベクトルの比較は可能であるが、実験の結果、マッピングは意外に難しく、リスコアリングは不十分であった。そこで前節で述べた同一文ベクトル空間表現された日英文ベクトルに対してマッピングを行う。

3.2 逆翻訳によるリスコアリング [9]

原言語文を使用したリスコアリングの方法として逆翻訳によるリスコアリングを行う。NMT と SMT の翻訳候補文を日英方向の翻訳モデルを使用し、原言語方向へ翻訳する。その後、原言語文と逆翻訳文の BLEU 又は BLEU+1 を文ごとに導出し、NMT と SMT の翻訳文をリスコアリングする。文類似度の評価尺度として文単位の BLEU では 4gram 一致が文中にない場合、計算することができない。その為少ない ngram 一致等も考慮した指標方法である BLEU+1 を使用する。

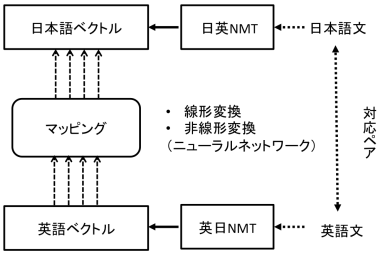

図 4: 文ベクトル間のマッピングの構成

4 翻訳システムの改善

4.1 バイトペアエンコーディング (BPE)[6]

バイトペアエンコーディング (BPE) は、NMT の問題の 1 つである未知語の入力を改善する手法である。高頻度に出現する単語は 1 語として扱い、低頻度に出現する単語は分割して、短い単位の連結で表現する。これにより、未知語が短い単位の連結で表現できる様になり、未知語の数が減少する。

音声認識器によって誤認識される単語は、意味表現としては大きく違うことがあるが、文字単位ではよく類似している場合がある。単語が短い単位に分割されることによって、単語としては誤認識された場合でも分割部分の単位の文字が正解である場合、翻訳文を予測する際の入力への意味的情報が増加すると考え、誤認識の影響が軽減できると期待できる。

4.2 コーパスの追加

(a) 大規模コーパスの追加

NMT では、経験上、少なくとも数十万ペアのパラレルコーパスが必要である。講義音声のような翻訳タスクでは、十分なパラレルコーパスが利用できない場合が多い。そこで、ドメインが多少異なっても、大量のパラレルコーパスを追加利用することは有用と考えられる。本研究が対象とする MIT 講義音声翻訳タスクではパラレルコーパスはほとんどなく (開発データとテストデータのみ有り)、代理用に TED コーパスを利用する。しかし、TED コーパスのパラレル文は 14 万文と少なく十分に NMT を学習することができないため、ASPEC コーパスを追加コーパスとして利用することも試みる。

(b) 音声の誤認識コーパスの追加 [10]

音声翻訳の大きな問題は音声認識誤りである。そこで、誤認識が含まれた入力文に対しての翻訳性能を

改善するために、クリーンなコーパスに誤認識されたコーパスを追加する。誤認識されたコーパスは、実際の認識器による誤りを用いる手法と、擬似的に誤りを作成する 2 つの手法を使用する。実際の誤りコーパスは発話音声を自動音声認識システム (ASR) によって認識する。誤りのバリエーションを増やすために学習コーパスの異なる複数の音響モデル・言語モデルを使用する。擬似的な誤りコーパスは正解書き起こしから作成される。音素変換表を使用し、正解確率と誤認識のパラメータを用いて、入力文の音素列に対して各音素を挿入・削除・置換する。

5 評価実験

5.1 実験条件

(a) 科学技術論文抄録 (ASPEC) コーパス

SMT の翻訳モデルの作成ツールには Moses を用い、パラメータの調整には Moses に搭載されている MERT ツールを用いる。学習の語彙サイズは日本語が 20404 語、英語は 32240 語である。SMT は最良の翻訳結果と予測される 1 ベスト翻訳結果以外に、複数の翻訳候補を 1000 ベストまで出力する。NMT の語彙サイズは両言語 10,000 語に制限し、エンコーダは双方向 LSTM(500+500 次元)、デコーダは翻訳性能の向上のため、アテンション機構を持った LSTM(1000 次元) で構成される。学習のエポック数は 10 で、実験に使うモデルは評価データによって決める。

リスコアリングの文ベクトルの比較にはコサイン類似度を採用し、入力文 (英語) とコサイン類似度が高い翻訳文 (日本語) を最良の翻訳結果として選択する。リスコアリングは上限値として SMT と NMT の 2 つの翻訳候補で、1 文ごとに BLEU の最も高い方の文を選択する場合 (オラクル) も行った。バイトペアエンコーディング (BPE) は文献 [6] を使用する。BPE の分割サイズを 30000 とする。

(b) TED コーパスと MIT 講義音声コーパス

講義音声の音声翻訳のタスクとして MIT の講義音声コーパス [10] を用いる。テストデータは MITOCW から 2 話者の 159 発話 (話者 1: 65 発話, 話者 2: 94 発話) を使用する。しかし、日本語正解文の付随する MIT 講義音声のパラレルコーパスが無いため (開発セット文とテスト文にはパラレコーパスを作成) 学習用には話し言葉のドメインとなる講演 TED の書き起こし 145032 文をコーパスとして使用する。

実際の音声認識誤りを作成するための ASR は、DNN-HMM を用いた [10]。また疑似音声認識誤りを生成するための音素の正解確率は 85、90、95、100% で実行し、挿入・削除誤りはそれぞれ 0 または 5% で実行している [10]。音声の置換誤りは、正解 音素に近い 3 つの音素に順に (100 − 正解音素確率) の 1/2、1/3、 1/6 の確率で起こるものとする。

表 1: ASPEC コーパスによる翻訳実験結果 (BLEU)

		BPE ベース (30000)				単語ベース	
NMT 候補数	0	1	5	1	5	1	5
SMT 候補数	1	0	0	1	30	0	0
オラクル	25.83	38.14	43.36	39.41	44.56	35.44	40.16
逆翻訳	-	-	**38.51**	36.73	36.78	-	35.61
ベクトル空間によるリスコアリング							
エンコーダ入力：日英、デコーダ出力：日本語、アテンションあり							
オラクル	-	-	40.39	36.92	38.66	-	38.43
マッピング	-	-	38.02	33.72	33.17	-	35.23
エンコーダ入力：日英、デコーダ出力：日本語、アテンションなし							
オラクル	-	-	39.70	36.42	37.47	-	38.21
マッピング	-	-	**38.51**	34.24	33.50	-	35.48
エンコーダ入力：日英、デコーダ出力：英語、アテンションなし							
オラクル	-	-	40.81	37.23	39.83	-	38.44
マッピング	-	-	**38.40**	34.35	33.17	-	35.51

5.2 翻訳実験結果

(a) ASPEC コーパス

SMT と NMT による英日翻訳と翻訳候補のリスコアリングの評価実験を行った。学習コーパスとテスト文には学術論文抄録のコーパスである ASPEC コーパスの英語-日本語ペアを使用する。翻訳品質でソートされた上位 100 万文を使用し、テスト文は用意された 1812 文を使用する。

翻訳結果の第一候補の BLEU は SMT で 25.8、単語単位の NMT で 35.4、BPE 単位で 38.1 であった。これは、以前報告したロイター記事の翻訳結果 [8](SMT:20.09、NMT:21.97) と傾向が異なっている。ロイター記事ではパラレルコーパスが 5 万文と非常に少なかったため、SMT と NMT の性能差はほとんどなかった。このため SMT と NMT の翻訳結果のオラクルなリスコアリングで大きく性能が向上した (BLEU:27.05)[8]。それに対して、ASPEC ではパラレルコーパスが 100 万文と多く、SMT よりも NMT の性能が圧倒的に良くなっている。

隠れベクトルを共通化した文ベクトルを出力する NMT はのアテンションの有無、入力は英語と日本語、出力は英語または日本語と複数のモデルを用意し、マッピングネットワークを作成するための文ペアのサイズは 100 万文とした。リスコアリングによる ASPEC コーパスで学習した SMT または NMT の翻訳結果の BLEU を表 1 に示す。

BPE 単位の NMT は単語単位よりも翻訳性能が良い。オラクルの結果から、SMT の翻訳文と NMT の翻訳文は質が異なるものの、相補的に利用すれば BLEU が 1.3 上昇することを示しているが、NMT の翻訳候補の 5 ベストのオラクルでは 7.2 上昇している。BPE 単位による NMT の翻訳候補 5 ベストを逆翻訳によりにリスコアリングした場合、BPE ベースラインよりも 0.37 改善した。つまり、翻訳の質が高い翻訳結果を絞り込んでリスコアリングすることが翻訳結果の改善につながると考えられる。

ベクトル空間のマッピングに使用する単語単位の NMT は、アテンション機構を使用する場合としない場合で比較を行った。また、BPE 単位の BLEU では

表 2: TED コーパスによる音声翻訳実験結果 (BLEU)

コーパス	TED	+ASR	+擬似的	+ASR +擬似的	+ASR (複数)	+擬似的 (複数)	+ASPEC (500000)
総文数	145032	212402	287021	354391	1422329	574540	645032
入力	単語ベース						
書き起こし	7.33	6.59	5.78	7.33	6.61	7.30	7.84
音声認識	5.30	5.72	5.15	6.46	5.33	6.26	6.83
入力	BPE ベース (30000)						
書き起こし	6.71	7.63	6.88	7.97	6.29	7.51	8.72
音声認識	5.13	6.08	5.96	6.77	5.31	6.36	7.34

ベースラインよりも逆翻訳によるリスコアリングと同程度に上昇した。

(b) MIT 講義の音声翻訳実験結果

BPE を適用した場合と誤りコーパスをパラレルコーパスに追加した場合の NMT による翻訳結果の BLEU を表 2 に示す。実際の音声認識誤りコーパスは 1 つの ASR を使用した場合と、複数の ASR を使用した場合で行い、擬似的な誤りコーパスにおいても 1 つの疑似 ASR、複数の疑似 ASR で実験を行い、2 つのコーパスを組み合わせた実験も行った。

SMT による BLEU は書き起こし入力では 10.3、音声入力では、7.7 となった。講義音声翻訳では NMT よりも SMT のほうが性能が良かった。この理由は、NMT の学習に用いたパラレルコーパスが TED の 14 万文であったため、言語モデルが十分に学習できなかったためである。(SMT では目的言語モデルを利用)。

BPE を適用した場合、書き起こし入力も音声認識入力においても単語単位のベースラインを上回るには至らなかった。これは、元々の BLEU が低いためであると考えられる。実際の音声認識誤りコーパスを加えることにより、音声認識入力の場合、ベースラインから 0.42、BPE では 0.95 の上昇が見られた。擬似的な誤りコーパスのみを追加した場合では、精度が上昇しなかったが、実際の音声認識誤りコーパスと組み合わせることにより、単語単位の NMT で 1.16 の BLEU の上昇、BPE では 1.64 の BLEU の上昇が見られた。そのため、音声認識の誤りコーパスの追加は実際の誤り、擬似的な誤りどちらも音声翻訳において有用であると考えられる。書き起こし入力文では BPE において 1.26 の上昇が見られた。複数のバリエーションを持つコーパス増大により、NMT のモデルの精度自体が上昇したと考えられる。

ASPEC コーパスの 50 万文ペアをパラレルコーパスに追加することで、コーパスの増大により BPE において、書き起こし入力で 2.01、音声入力では 2.21 の BLEU の上昇が見られた。これはドメインが異なっても大量のパラレルコーパスの利用は効果があることを示している。

6 おわりに

本研究では英日翻訳において NMT と SMT、複数の翻訳結果のリスコアリングの検討、音声認識誤りに考慮したコーパスの追加手法を検討した。逆翻訳による自動リスコアリングにおいて、ベースラインから最大の 0.39 の BLEU の上昇が得られることを示した。さらに音声認識の誤認識を持った誤りコーパスを追加することにより音声翻訳実験においてベースラインから、1.47 の上昇が見られた。

謝辞

本研究は JSPS 科研費 25280062 及び,18H01062 の助成を受けた。

参考文献

[1] M. Artetxe, G. Labaka, and E. Agirre. Learning bilingual word embeddings with (almost) no bilingual data. In *Proc. ACL2017*, pp. 451–462, 2017.

[2] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[3] Norioki Goto, Kazumasa Yamamoto, and Seiichi Nakagawa. English to Japanese spoken lecture translation system by using DNN-HMM and phrase-based SMT. In *2015 2nd International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, pp. 1–6, Aug 2015.

[4] E. Matusov, S. Kanthak, and H. Ney. On the integration of speech recognition and statistical machine translation. In *Proc. INTERSPEECH2005*, 2005.

[5] K. Sahashi, N. Goto, H. Seki, K. Yamamoto, T. Akiba, and S. Nakagawa. Robust lecture speech translation for speech misrecognition and its rescoring effect from multiple candidates. In *Proc. ICAICTA2017*, pp. 1–6, 2017.

[6] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1715–1725. Association for Computational Linguistics, 2016.

[7] 今村賢治, 隅田英一郎ほか. 双方向リランキングとアンサンブルを併用したニューラル機械翻訳における複数モデルの利用法. 情報処理学会研究報告自然言語処理 (NL), Vol. 2017, No. 9, pp. 1–8, 2017.

[8] 佐橋広也, 西村友樹, 秋葉友良, 中川聖一. 統計的翻訳とニューラル翻訳による翻訳候補の文の分散表現に基づくリスコアリングの検討. 言語処理学会第 24 回年次大会発表論文集 (NLP), Vol. 2018, pp. 260–263, 2018.

[9] 佐橋広也, 西村友樹, 秋葉友良, 中川聖一. 統計的翻訳とニューラル翻訳に基づく翻訳候補文の分散表現と逆翻訳によるリスコアリングの検討. 情報処理学会研究報告自然言語処理 (NL), Vol. 2018, pp. 1–5, 2018.

[10] 後藤統興, 山本一公, 中川聖一. 英日講義音声翻訳に対する音声認識誤りを考慮したパラレルコーパスの利用. 情報処理学会研究報告音声言語情報処理 (SLP), Vol. 2016, pp. 1–7, 2016.