

重文・複文翻訳における Transformer の性質

丸山京祐¹ 村上仁一²

¹ 鳥取大学大学院持続性社会創生科学研究科

² 鳥取大学工学部

¹m23j4056b@edu.tottori-u.ac.jp

²murakami@tottori-u.ac.jp

概要

本論文では、重文・複文の翻訳において対訳単文を学習することが精度に与える影響を Transformer と RNN を比較し調査した。その結果、自動評価指標では Transformer は RNN よりも対訳単文が翻訳精度向上に貢献することが分かった。しかし、人手評価で Transformer と RNN を比較した結果、自動評価指標の数値から期待されるほどの高い翻訳精度が得られなかったことを示す。

1 はじめに

近年、機械翻訳の分野において Transformer が高く評価されている [1]。Transformer はアテンション機構を持ち、単語同士の依存関係を学習できる。

そこで、重文・複文の翻訳について考えた。重文・複文の翻訳は複雑な文構造を翻訳することが重要である。文構造を翻訳するためには句や係り受けといった単語の依存関係を捉えることが必要になる。そのため、アテンション機構により単語の依存関係を適切に捉えることが翻訳精度向上に繋がる。また、単語の依存関係を学習するために対訳単文の学習を考えた。単文は単純な構造を持つため、単語の依存関係を学習しやすい。したがって、Transformer の学習コーパスに対訳単文が多く含まれることにより文構造を理解しやすくなる。そして、重文・複文の翻訳精度向上に繋がると推測できる。

本論文では、対訳単文を変化させながら学習することで重文・複文の翻訳精度に与える影響を Transformer と RNN を比較し調査した。

2 先行研究

本研究で考察する翻訳モデルについて説明する。

2.1 再帰的ニューラルネットワーク (RNN)

2014 年にニューラルネットワークを使った機械翻訳は提案された。特に再帰的ニューラルネットワーク (RNN) を用いて単語の語順を学習する seq2seq モデルが注目された [2]。

2.2 Transformer

Transformer は Vaswani らによって提案された [3]。従来の RNN の考え方に基づく機械翻訳モデルは長い文や複雑な文を処理することが難しい。その解決策として、アテンション機構を組み込んだ手法が導入されていた [4]。それに対して、Vaswani らは RNN に基づく手法では根本的な解決には限界があると考えた。そこで、アテンション機構のみに依存する手法としては提案された。

3 実験

本実験は、以下の手順で行う。

- 重文・複文のみの対訳コーパスに対訳単文を少しずつ加えていき、複数のコーパスを作成する
- 各々のコーパスを Transformer モデルと RNN モデルで学習し翻訳モデルを作成する
- 2 で作成した全ての翻訳モデルで重文・複文のテスト文を翻訳する
- それぞれの出力に対し自動評価と人手評価を行う

3.1 実験条件

本実験の実験条件を示す。日本語の分かち書きには mecab-python3 1.0.6 を使用する。辞書は unidic 1.1.0 を使用する。NMT の学習・翻訳には OpenNMT-py 2.3.0 を使用する。基本的に翻訳モデルのパラメータは Vaswani らの研究 [3]、RNN は Luong らの

研究 [4] に従う。ただし，Luong らの RNN はアテンションベース RNN モデルである。

表 1 実験条件

形態素解析ツール	mecab-python3 1.0.6
形態素解析辞書	unidic 1.1.0
NMT ツール	OpenNMT-py2.3.0

3.2 コーパス

使用するコーパスは電子辞書などから抽出した対訳文から作成された日英対訳単文コーパスと日英対訳重文・複文コーパスである [5]。表 2 に使用したコーパスの情報を示す。また，表 3, 4 にコーパスの一部を示す。

表 2 日英対訳データベース

対訳単文コーパス	学習文	163,188 文
	テスト文	16,328 文
対訳重文・複文コーパス	学習文	92,427 文
	テスト文	9,243 文

表 3 日英対訳単文コーパスの例文

1 ドル紙幣を崩した。	She broke a dollar bill into change.
あかりが一晩中ついている。	The lights have been burning all night.

表 4 日英対訳重文・複文コーパスの例文

毎日自転車に乗って通学しています。	I ride my bicycle to and from school every day.
毎日タイプライターで決まった分量だけの仕事をするということは文筆家には優れた修業法だ。	A daily stint at the typewriter is excellent discipline for a writer.

表 5 に実験で作成した学習用コーパスにおける，対訳単文と対訳重文・複文の数をそれぞれ示す。

表 5 学習用コーパス

対訳重文・複文	対訳単文	合計
92,427 文	0 文	92,427 文
92,427 文	20,399 文	112,826 文
92,427 文	40,797 文	133,224 文
92,427 文	81,594 文	174,021 文
92,427 文	163,188 文	255,615 文

4 実験結果

4.1 自動評価

本実験では，自動評価指標として BLEU[6]，METEOR[7]，TER[8]，RIBES[9] を使用した。Trans-

former で学習した結果を表 6 に示す。また，RNN モデルで学習した結果を表 7 に示す。テストデータは重文・複文コーパスのテスト文 9,243 文である。

表 6 Transformer モデルの自動評価

学習データ		自動評価指標			
重・複	単文	BLEU	METEOR	TER	RIBES
92,427 文	0 文	0.157	0.399	0.687	0.712
92,427 文	20,399 文	0.173	0.420	0.668	0.723
92,427 文	40,797 文	0.182	0.429	0.659	0.732
92,427 文	81,594 文	0.204	0.457	0.642	0.743
92,427 文	163,188 文	0.232	0.479	0.616	0.757

表 7 RNN モデルの自動評価

学習データ		自動評価指標			
重・複	単文	BLEU	METEOR	TER	RIBES
92,427 文	0 文	0.163	0.423	0.687	0.731
92,427 文	20,399 文	0.173	0.434	0.677	0.734
92,427 文	40,797 文	0.180	0.444	0.667	0.739
92,427 文	81,594 文	0.192	0.459	0.648	0.748
92,427 文	163,188 文	0.191	0.466	0.640	0.756

表 6, 7 より，RNN モデルの単文 163,188 文の BLEU 値を除き，対訳単文の増加に従い精度が向上している。Transformer と RNN の結果を比較すると，対訳単文を学習しないときは RNN がより高い精度を示す。しかし，対訳単文の量が増加するにつれて Transformer の精度の上昇がより大きくなっている。特に，重文・複文 92,427 文 + 単文 163,188 文の結果では Transformer は RNN より BLEU 値で 0.041 大きい値を示している。

4.2 人手評価

Transformer と RNN を比較した際，自動評価で特に大きい差があった重文・複文 92,427 文 + 単文 163,188 文で学習した結果の人手評価を行った。テスト文から 100 文を抽出し，著者を含む 5 人で対比較評価を実施した。kappa 値は 0.55 であった。評価基準を表 8 に示す。評価結果を表 9 に示す。

表 8 評価基準

Transformer	Transformer の方が良い
RNN	RNN の方が良い
○	どちらも翻訳が正しい
×	同程度の精度で翻訳が間違っている

表 9 より，人手評価においては transformer と RNN の精度は同等か RNN の方がわずかに精度が高い。

4.3 出力結果の例文

表 10, 11, 12, 13, 14 に出力結果の一部を示す。

表9 重文・複文 92,427 文 + 単文 163,188 文で学習した時の Transformer と RNN の対比較評価

評価者	Transformer	RNN	○	×
A	17	21	22	40
B	21	23	28	28
C	21	21	29	29
D	11	18	29	42
E	17	19	24	40
合計	87	102	132	179

表10 Transformer の出力が良い例

テスト文	そのような危機に直面しても彼女が冷静なのに驚いた。
参照文	He wondered at her composure in such a crisis .
Transformer	Even if we face such a crisis , she was surprised at her composure .
RNN	Her face in the face of such a crisis surprised her .

表10では、Transformer の出力は誤訳とは言えない。それに対して、RNN の出力は「直面する」を意味する動詞の「face」が名詞のような用法で2回出現し「冷静」は翻訳されていない。

表11 RNN の出力が良い例

テスト文	ご存じのように、自動車業界は法律の要求を満たすことができないと主張しています。
参照文	As you may know , the automotive industry claims that requirements of the Act can not be met .
Transformer	Will you know the automobile industry is a problem ?
RNN	As you know , we insist that the automobile industry will not meet the demands of law .

表11では、RNN の出力は参照文と比較すると主語が異なる。しかし、テスト文の情報では主語が「we(私達)」とも解釈できるため正しい翻訳であると判断した。それに対して、Transformer の出力は全く意味の異なる疑問文である。

表12 両方良い例

テスト文	合図が出るとすぐ彼らは席に着いた。
参照文	They took their seats at a given signal .
Transformer	As soon as the signal was given , they got to their seats .
RNN	They sat at their seats as soon as the signal was given .

表12では、どちらも正しく翻訳できている。「as soon as ~」の節の位置が異なるがどちらも文法は正しい。

表13 両方悪い例

テスト文	~でのデモは過去の抗議デモなどちっぽけに見えてしまうほどの規模になる見込みだ。
参照文	The scope of demonstrations in is expected to dwarf all past protests .
Transformer	In , its demonstrations are expected to be extremely small .
RNN	The number of demonstrators in is expected to grow even more than the last past protest protest .

表13では、Transformer の出力は「~でのデモ」がとて小さくなるという反対の意味になっている。それに対し、RNN の出力は前回のデモより大きくなるという意味になっている。また、RNN の出力は「protest」が連続して出現し文法が誤っている。これらの理由から同じ程度の翻訳間違いだと評価した。

表14 評価が分かれた例

テスト文	子供たちはみやげ物を買おうと観光客につきまとった。
参照文	The children tagged along at the tourists ' heels , trying to sell souvenirs .
Transformer	The children visited the tourists to do away with souvenirs .
RNN	The children followed the tourists to buy their gifts .

表14は評価者によって評価が分かれた例である。Transformer の出力は「つきまとう」を「visit (訪ねる)」、「売る」を「do away with (処理する、廃棄する)」と翻訳している。似た状況を表す言葉ではあるが、誤った翻訳になっている。それに対して、RNN の出力は「売る」を「buy (買う)」と翻訳しており、反対の意味になっている。判断が分かれた理由は、RNN の出力において反対の意味の部分がとて近いと判断した評価者と全く異なると判断した評価者がいた事である。

5 考察

5.1 実験結果の考察

Transformer はアテンション機構により単語の依存関係を学習できる。そこで、複雑な文の翻訳において単語同士の関係が理解しやすい単文を学習することが翻訳精度向上に重要であると仮定した。

表6,7より、自動評価では transformer は RNN に比べ対訳単文の量の増加による精度の向上がかなり大きいことが分かった。しかし、表9の人手評価では同程度か RNN の方が少し良い結果となった。

以上より、Transformer は対訳単文の学習により単語同士の依存関係を学習し句のレベルで正しく翻訳

しやすくなり自動評価の数値は向上する。しかし、表 11 からわかるように「you know」や「we insist that the automobile industry」といった句単位で翻訳ができて、文全体の文構造が間違っていれば正しい翻訳ができなくなる。つまり、単語同士の依存関係を学習し句単位の翻訳が強化されたことによって複雑な文構造の翻訳は精度が低下したと考えられる。

5.2 単語単位の翻訳による誤訳

本研究では、句単位の翻訳が強化されたことが複雑な文構造の翻訳に悪影響を及ぼしたと考察した。翻訳タスクにおいて、単語単位の翻訳により文全体では誤訳する問題が指摘されてきた。例えば、「彼は冗談が過ぎた。」の正しい翻訳は「His joke went too far.」である。それに対して、単語単位で翻訳すると「He passed joke.」となり、元の文とは異なる意味になってしまう。

本研究の結果より、Transformer は単語単位の翻訳による誤訳に近い問題が発生しやすくなると考えられる。

5.3 単文翻訳

本研究では、重文・複文の翻訳精度を調査する実験を行った。そこで、実験の結果が文構造の複雑さに影響されているのか調べるために、単文の翻訳実験も実施した。

この実験では表 2 の対訳単文コーパスのテスト文 9,243 文を 3 章と同じモデルで翻訳した。表 15, 16 に自動評価の結果を示す。また、重文・複文と同様に学習データが重文・複文 92,427 文 + 単文 163,188 文の RNN と Transformer の結果の人手評価を表 17 に示す。kappa 値は 0.50 であった。

表 15 Transformer モデルで単文翻訳の自動評価

学習データ		自動評価指標			
重・複	単文	BLEU	METEOR	TER	RIBES
92,427 文	0 文	0.157	0.378	0.684	0.707
92,427 文	20,399 文	0.178	0.414	0.647	0.729
92,427 文	40,797 文	0.194	0.437	0.628	0.747
92,427 文	81,594 文	0.215	0.466	0.608	0.763
92,427 文	163,188 文	0.240	0.496	0.579	0.783

表 16 RNN モデルで単文翻訳の自動評価

学習データ		自動評価指標			
重・複	単文	BLEU	METEOR	TER	RIBES
92,427 文	0 文	0.161	0.397	0.678	0.722
92,427 文	20,399 文	0.182	0.435	0.646	0.747
92,427 文	40,797 文	0.189	0.448	0.635	0.756
92,427 文	81,594 文	0.201	0.468	0.613	0.769
92,427 文	163,188 文	0.218	0.493	0.586	0.787

表 17 単文翻訳の人手評価結果

評価者	Transformer	RNN	○	×
A	12	23	39	26
B	17	16	44	23
C	15	22	39	23
D	10	14	59	17
E	14	25	38	23
合計	68	100	219	112

表 15, 16 より、自動評価指標では複文と同様に Transformer は RNN よりも単文を学習するほど精度の向上が大きい。そして、学習データが重文・複文 92,427 文 + 単文 163,188 文の数値は Transformer の方が高い。さらに、人手評価では重文・複文テスト文の結果と同様に Transformer が RNN と比べて精度が低いことがわかる。そのため、単純な文構造であっても一部が正しく全体が間違った翻訳が出力されてしまうことがわかる。

5.4 低資源の学習による影響

Google 翻訳や DeepL などの一般的に使用される機械翻訳ツールは大規模なコーパスで学習される。しかし、本実験は約 9 万文対、約 16 万文対と比較的小規模なコーパスを使用した。表 6, 7 からわかるように、RNN とは異なり Transformer は対訳単文を 8 万文対から 16 万文対に増加させたときでも自動評価指標の値に増加傾向が見られる。そのため、大規模なコーパスで学習していないことで Transformer のモデルの性質を検証しきれなかった可能性が考えられる。

6 おわりに

本論文では、Transformer に着目して実験を行った。複文の翻訳は複雑な文構造を翻訳することが重要である。そのため、アテンション機構により単語の依存関係を適切に捉えることが翻訳精度向上に繋がる。本研究では構造が単純な単文は単語の依存関係を理解しやすいため、Transformer の学習コーパスに含まれる対訳単文が複文の翻訳精度に大きく貢献すると仮説を立てた。

実験の結果、自動評価指標では RNN よりも精度が向上することが分かった。しかし、人手評価では期待する成果が得られなかった。理由としては、アテンション機構によって句レベルの翻訳精度が向上したことで句レベルの翻訳が優先され、文全体の構造が誤ることが原因であると推測される。

謝辞

人手評価に参加した以下の学生に感謝します。
名村 太一, 松本 武尊, 柳原 弘哉, 駿河 樹

参考文献

- [1] Tianyang Lin, Yuxin Wang, Xiangyang Liu, and Xipeng Qiu. A survey of transformers, 2021.
- [2] Nal Kalchbrenner and Phil Blunsom. Recurrent continuous translation models. In David Yarowsky, Timothy Baldwin, Anna Korhonen, Karen Livescu, and Steven Bethard, editors, **Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing**, pp. 1700–1709, Seattle, Washington, USA, October 2013. Association for Computational Linguistics.
- [3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. **Advances in neural information processing systems**, Vol. 30, , 2017.
- [4] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. Effective approaches to attention-based neural machine translation. **arXiv preprint arXiv:1508.04025**, 2015.
- [5] 村上仁一. 日英対訳データベースの作成のための1考察. 言語処理学会第17回年次大会発表論文集, D4-5, pp. 979–82, 2011.
- [6] 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, 2002.
- [7] Satyanjeev Banerjee and Alon Lavie. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In **Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization**, pp. 65–72, 2005.
- [8] Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. A study of translation edit rate with targeted human annotation. In **Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers**, pp. 223–231, 2006.
- [9] 平尾努, 磯崎秀樹, Kevin Duh, 須藤克仁, 塚田元, 永田昌明. Ribes: 順位相関に基づく翻訳の自動評価法. 言語処理学会第17年次大会発表論文集, pp. 1111–1114, 2011.