

NMT での学習データの単語数制約による翻訳精度の向上

細川楓¹ 村上仁一²

¹ 鳥取大学工学部電気情報系学科 ² 鳥取大学工学部

¹b20t2096x@edu.tottori-u.ac.jp ²murakami@tottori-u.ac.jp

1 概要

ニューラルネットワーク翻訳 (Neural Machine Translation: 以下, NMT) に関する研究はこれまでに多数行われている [1]. しかし, 極端に単語数の多いテスト文の翻訳は困難である. この原因として, 学習文とテスト文の単語数の差に着目した. 機械翻訳を回帰分析と考えると, 入力に対して学習文を基に出力の予測を行っている. そのため, 学習データ中に入力と同じ単語数の学習文が多いほど, より正確な出力の予測ができると考えた.

本論文では, テスト文の単語数と学習文の単語数を揃えて機械翻訳を行う. そして, テスト文に対する学習文の単語数による影響を調査する. 実験の結果, 提案手法による翻訳精度の向上が確認できた.

2 encoder-decoder モデル

NMT は基本的に encoder-decoder モデル [2] を利用している. encoder は入力文から単語を抽出し, decoder はそれを入力として翻訳を出力する. また, 先頭から単語が出力されていくが, 2 単語目以降, decoder は過去の出力も入力として受け取る.

3 回帰分析

回帰分析 [3] は出力と入力の関係を説明するモデルを導くことを目的としている. 回帰分析モデルは入力を独立変数 x , 出力を従属変数 y として, 以下の独立変数 x の n 次多項式

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots + a_nx^n + \epsilon \quad (1)$$

として表される. ここで a , b は回帰係数, ϵ は正規分布に従う誤差である.

4 回帰次数

回帰分析では回帰次数を決定する方法が問題となる. 機械翻訳においては入力文の単語数を回帰次数だと考えることができる [4]. encoder によって抽出

された入力文中の単語から, decoder によって翻訳が生成されるという encoder-decoder モデルの考え方より, 入力文中の単語が独立変数に対応するとみなせる. このことから, 入力文と同一の単語数の学習文で学習を行うことが妥当である.

5 先行研究

Kenton Murray と David Chiang[5] は, 翻訳モデルに文レベルの修正を導入することで翻訳精度の向上を試みた. NMT は短すぎる翻訳を生成する傾向がある. この問題に対し, 長さ補正として標準のモデルスコアに出力シーケンスの長さで調整された単語報酬と呼ばれる値から計算される定数単語報酬の追加を行うことで対処している.

出力文の単語数に着目したこの研究では, 長さ補正を行った場合, 補正を行わない場合と比べて翻訳精度が向上したことが示されている. このことから, 単語数という要素が翻訳精度に影響を与えているのではないかと考えられる.

6 提案手法

テスト文の単語数に学習文の単語数を揃えて機械翻訳を行う.

本研究では, 以下の手法を提案する.

1. 単語数を揃えたテスト文のコーパスを作成
2. テスト文と同じ単語数の学習文のコーパスを作成
3. 作成した学習文のコーパスを NMT で学習
4. 学習した NMT でテスト文を翻訳

7 実験

本実験では, OpenNMT-py[6] を用いて翻訳実験を行う. 手法ごとに異なる条件の学習文を学習させる. そして, 翻訳結果を比較する.

7.1 実験データ

実験には、電子辞書などの例文より抽出したコーパス [7] から単語数ごとに文を抽出して用いる。コーパス内のデータは単文となっている。学習データとして、コーパス内から手法ごとに条件に合致した対訳学習文を 2 万 4269 文選んで用いる。また、テストデータとして日本語のテスト文を 2302 文選んで用いる。コーパス中で単語数 9 の文が最も多かったため、実験では単語数 9 の文をテストデータとする。

提案手法では、単語数 9 の文だけで学習を行う。しかし、現実的にはテスト文と同一の単語数の学習文だけを集めるのは難しい。そこで、テスト文の単語数に近い範囲の単語数の学習文だけで学習を行う。範囲はテスト文の単語数の $2/3$ から $3/2$ とする。テスト文の単語数が 9 単語なので、学習文の範囲は 6 単語から 13 単語にする。

表 1 に本実験で行う 5 種類の手法と用いる学習データを示す。

表 1 本実験で行う 5 種類の手法と用いる学習データ

手法	学習データ
ベースライン	バラバラな単語数の学習データ
提案手法	単語数 9 の学習データ
幅制限	単語数 6~13 の学習データ
上カット	単語数 13 以下の学習データ
下カット	単語数 6 以上の学習データ

表 2, 3 に学習文 (単語数 9)、テスト文の例文を載せる。

表 2 単文対訳学習文の例文

1 か月で仕事がずいぶん溜まった。
In just one month a lot of work has piled up .

表 3 単文テスト文 (単語数 9) の例文

このビルは通りに面している。
彼は大学に在学している。

8 実験結果

8.1 自動評価結果

単語数 9 の単文テスト文 2302 文の翻訳結果の出力文に対して自動評価指標 BLEU[8], METEOR[9],

TER[10], RIBES[11] による自動評価を行った。表 4 に自動評価の結果を示す。

表 4 学習文 24,269 文でのテスト文 2302 文の自動評価の結果

	BLUE	METEOR	TER	RIBES
提案手法	0.1380	0.3745	0.6501	0.7315
幅制限	0.1271	0.3591	0.6600	0.7217
上カット	0.1138	0.3469	0.6711	0.7109
下カット	0.1153	0.3512	0.6707	0.7125
ベースライン	0.1116	0.3444	0.6685	0.7139

表 4 より、学習文の単語数がテスト文の単語数に近づくほど精度は向上していくことがわかる。

8.2 人手評価結果

単語数 9 の単文テスト文 2302 文を翻訳した結果から 100 文を抽出し、人手による対比較評価を行った。評価者は 1 名である。表 5 に対比較評価の結果を示す。

表 5 対比較評価の結果

評価者	A
提案手法 > ベースライン	20
提案手法 < ベースライン	15
どちらも正しい	7
どちらも間違い	58

表 5 より、提案手法の方がベースラインより良い結果となっている。

8.3 出力結果の例文

表 6 に翻訳例を示す。

表6 翻訳例

テスト文	その光景に私は感動した。
参照文	The scene touched my heart .
提案手法	I was impressed by the scene .
幅制限	The sight knocked me up .
上カット	I was impressed by that scene .
下カット	I was a thrill on the sight .
ベースライン	The scene looked me strange .
テスト文	本を机の上に置きなさい。
参照文	Place the book on the desk .
提案手法	Stick the book on the desk .
幅制限	Put the book on the desk .
上カット	Put the books on the desk .
下カット	Wrap the book on the desk .
ベースライン	Put this book on the desk .
テスト文	このピアノは調子が合っている。
参照文	This piano is in tune .
提案手法	This piano is in tune .
幅制限	This piano is in tune .
上カット	This piano is correct .
下カット	This piano is well-organized .
ベースライン	This piano is in tune .

8.4 実験のまとめ

実験結果より、学習文の単語数がテスト文の単語数に近いほど翻訳の精度が向上することがわかる。

9 考察

9.1 ベースラインと比較した提案手法の翻訳精度

表4, 5より、提案手法は自動評価と対比較評価においてベースラインよりも良い結果を出している。以下の表7に提案手法の方が良かった翻訳例を示す。

表7 提案手法の方が良かった翻訳例

テスト文	シャツの襟が黒ずんできた。
参照文	My shirt collar developed a grimy ring .
提案手法	My shirt collar developed a grimy ring .
ベースライン	The jacket has gotten on her shirt .
テスト文	大統領はその条約に調印した。
参照文	The President signed the treaty .
提案手法	The president signed the treaty .
ベースライン	The president was divided on the treaty .
テスト文	彼は講演の要点を書き留めた。
参照文	He noted down the main points of the lecture .
提案手法	He wrote down the gist of his lecture .
ベースライン	He asserted the roll of his speech .

しかし、対比較評価では提案手法のほうが良い結果が出ているものの、得られた正しい翻訳の数は少なかった。翻訳例100例のうち58例が提案手法とベースラインの両方で間違っていると評価された。この原因として、学習文の数が少なかったことを考えている。今回の実験では、提案手法のためにテスト文と同じ単語数の学習文を集めた。しかし、特定の単語数の学習文を大量に集めることは難しかった。そのため、学習文の不足により提案手法とベースラインの翻訳精度が十分でなかったために翻訳をうまく行えなかった文が多くなったとみている。

9.2 幅制限、ベースラインの学習データ中の単語数分布

幅制限とベースラインはどちらも単語数がバラバラな文を学習している。しかし、表4より、幅制限はベースラインよりも良い結果を出している。幅制限はベースラインと比較して、学習文をテスト文の単語数に近い単語数に限定して学習を行っている。そのため、原因として幅制限、ベースラインの学習データ中の単語数分布に違いがあることが考えられる。

図1に幅制限、図2にベースラインの学習データ中の単語数分布を示す。

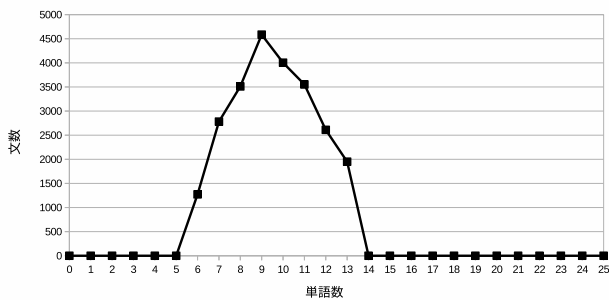


図1 幅制限の学習データ中の単語数分布

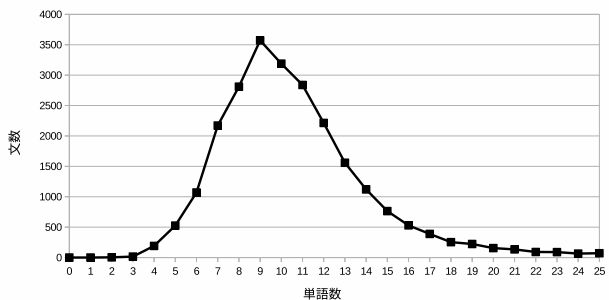


図2 ベースラインの学習データ中の単語数分布

上図より、幅制限、ベースラインの学習データ中の単語数分布は似た形をとっていることがわかる。しかし、ベースラインでは幅制限よりも含まれるデータの範囲が広い。このことから、ベースラインでは幅制限よりも大きい、または小さい単語数の学習文がノイズとなったために、翻訳精度が低下したとみている。

9.3 上カットと下カットの評価結果

表4より、上カットと下カットの自動評価結果はベースラインとあまり変わらない結果となっている。幅制限ではベースラインよりも良い結果が出ていることから、上カット、下カットのどちらかだけでは翻訳精度は向上しないことがわかる。テスト文の単語数と比較して、上カットには単語数の小さすぎる学習文、下カットには単語数の大きすぎる学習文が含まれている。表8に上カットでの単語数の小さすぎる単文対訳学習文の例文、表9に下カットでの単語数の大きすぎる単文対訳学習文の例文を載せる。このことから単語数の大小にかかわらず、テスト文の単語数に対して極端に単語数の差がある学習文が翻訳の際にノイズになっていると考えている。

表8 上カットでの単語数の小さすぎる単文対訳学習文の例文

日本語	英文
急げ。	Shake the lead out .
たまに来る。	He comes now and then .
まかせなさい。	Leave it to me .

表9 下カットでの単語数の大きすぎる単文対訳学習文の例文

日本語	英文
その派遣会社は臨時の秘書、アッセンブリー・ラインのアシスタントやガードマンだけでなく、医師、弁護士、臨床検査技師から、はては企業のトップまで派遣してくれる。	The temp agency offers not only temporary secretaries , assembly-line assistants and security guards but also doctors , lawyers and medical technicians even chief executives .
NATO軍は、中国大使館以前にも、コソボやそれ以外のユーゴ各地で、民間の列車や難民の車列、病院などへの誤爆を引き起こしている。	Indeed , even before the bombing of the Chinese Embassy , NATO planes had mistakenly hit vehicles and trains carrying civilians and refugees , and hospitals in Kosovo and other parts of Yugoslavia .
プラスチック容器も、千三百四十八市町村が四百九十三市町村に減り、収集量は二十三万九千トンが十万六千四百八トンになった。	The total weight of recycled plastic containers reportedly will drop from the planned 239,000 tons per year to 106,418 tons per year .

10 おわりに

本研究では、NMTによる機械翻訳においてテスト文と同じ単語数の学習文を学習させて翻訳を行った。その結果、以下のことがわかった。

- テスト文と学習文の単語数を揃えることで翻訳精度が向上する
- 学習文の単語数をテスト文の単語数に近い単語数に限定した場合でも翻訳精度は向上する
- テスト文の単語数と極端に差がある学習文は、単語数の大小にかかわらず翻訳精度が低下する原因となる

参考文献

- [1] Philipp Koehn and Rebecca Knowles. Six challenges for neural machine translation. **arXiv preprint arXiv:1706.03872**, 2017.
- [2] 中澤敏明. 機械翻訳の新しいパラダイム：ニューラル機械翻訳の原理. 情報管理, Vol. 60, No. 5, pp. 299–306, 2017.
- [3] 岩崎学. 統計的因果推論の視点による重回帰分析. 日本統計学会誌, Vol. 50, No. 2, pp. 363–379, 2021.
- [4] 村上仁一. 多数決による自己回帰モデルに基づく機械翻訳. 2024. 2024年度言語処理学会発表予定.
- [5] Kenton Murray and David Chiang. Correcting length bias in neural machine translation. **arXiv preprint arXiv:1808.10006**, 2018.
- [6] Opennmt - open-source neural machine translation, 2017. <https://opennmt.net/>.
- [7] 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.
- [8] 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.
- [9] Satyanjee 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.
- [10] Gregor Leusch, Nicola Ueffing, and Hermann Ney. A novel string-to-string distance measure with applications to machine translation evaluation. In **Proceedings of Machine Translation Summit IX: Papers**, 2003.
- [11] 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.