

折り返し翻訳を利用した言い換え生成

松本武尊¹ 村上仁一²¹ 鳥取大学工学部電気情報系学科 ² 鳥取大学 工学部¹b19t2108k@edu.tottori-u.ac.jp²murakami@tottori-u.ac.jp

概要

言い換え技術は、機械翻訳や対話システムなど様々なことに応用されている [1]。本研究では、ニューラル機械翻訳 (NMT) を用いて、折り返し翻訳を行うことにより、自動で言い換え生成を行う。機械翻訳を用いた言い換え生成は、誤訳などにより精度が低下する。この問題に対して、本研究では入力文を折り返し翻訳した結果が一致した場合のみ言い換えとする。実験の結果、高い精度で大量の言い換え文を得ることに成功した。

1 はじめに

言い換え生成は、同一言語の翻訳と捉えることが可能である。そのため、機械翻訳を用いて言い換え生成を行うことができる。しかし、精度があまり良くない。その原因として、翻訳機の精度に大きく依存してしまうことがあげられる。そこで本研究では、折り返し翻訳を行い元の文と折り返し翻訳をされた文を比較する。そして、2つの文が一致したものを採用することによって誤訳が減ると考えられる。そこで、NMT を用いて折り返し翻訳を行い、言い換え生成を行い、精度の調査を行う。

2 従来手法

言い換え生成は、様々な研究がなされてきた。機械翻訳を用いた手法では、対訳文の一方を翻訳することによって、翻訳された結果ともう一方を言い換えとしている。[2] では、言い換え認識の評価用コーパスの構築のために、複数の翻訳機を用いて同じ英語文を翻訳し、複数の日本語訳を得ている。これらの日本語訳は、言い換えとなる。図 1 に例を示す。しかし、翻訳誤りなどによって言い換えとして不適切なものが含まれる。そこで、誤訳を回避するために、BLEU 値を用いて翻訳精度が高い日本語対訳を言い換え文とすることで、言い換え生成を行っ

た。しかし、BLEU 値を用いても、誤訳を十分に回避できない。

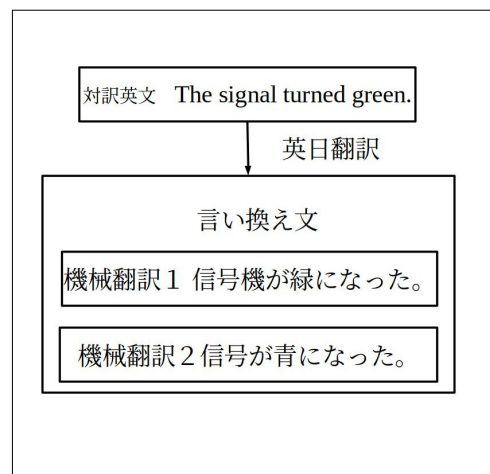


図 1 従来手法の例

3 提案手法

2つの文をそれぞれ翻訳し、翻訳結果が一致したとき、2つの文は言い換え文だと予想される。そこで、図 2 のように NMT を用いて折り返し翻訳を行い、翻訳結果が一致した場合、言い換え文とする。それにより、誤訳が減り言い換え精度が向上すると考えられる。以下に提案手法の例を示す。

1. 対訳英文 “The signal turned green.” を英日翻訳する
2. 1 の結果の “信号機が緑になった。” を日英翻訳する
3. 日英翻訳をした結果の “The signal turned green.” と対訳英文 “The signal turned green.” を比較し、一致しているかを調べる
4. 一致しているとき、英日翻訳した結果の “信号機が緑になった。” と対訳日本語文の “信号は青になった” を言い換え文とする

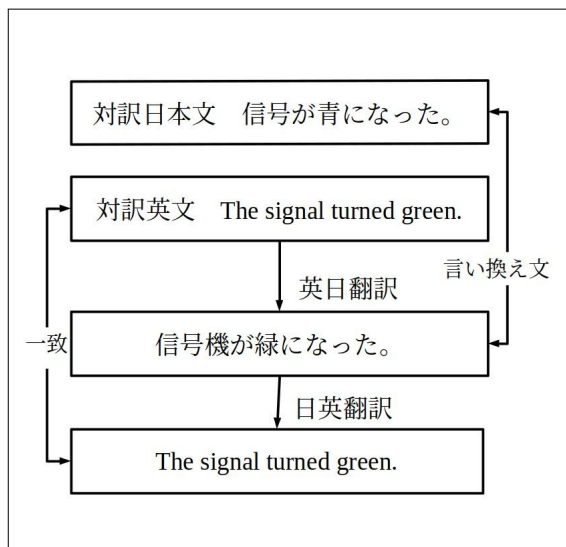


図2 提案手法の例

4 評価方法

本研究では、人手で評価を行う。言い換えは表現が多様であり言い換えの評価は曖昧である。そこで、2つの意味が完全に等しくなくとも、大きな差がなければ正解とする。

5 実験

本研究では、2種類の実験を行う。また、この2種類の実験で生成された言い換え文からランダムに20文抽出を行い、人手で評価を行う。

1. 折り返し翻訳を行う実験 (提案手法)
2. 折り返し翻訳を行わない実験 (ベースライン)

5.1 実験条件

本研究では、OpenNMT[3]を用いて機械翻訳を行う。また、実験データは単文の日英対訳文163,188文を学習及び入力として用いる。

5.2 実験手順

1. OpenNMTに日英対訳学習文約16万文を用いて学習する
2. 英日翻訳を行う
3. 2の結果を基に日英折り返し翻訳を行う
4. 対訳英文と3の結果が同一であるとき、対訳日本語と2を言い換え文とする
5. ベースラインと提案手法についてランダムに抽出した20文の人手評価を行う

5.3 実験結果

実験結果を表1に示す。本研究では、対訳日本語と一致したものは、同一文とみなし、言い換えに含まない。

	提案手法	ベースライン
言い換え文	25003	140957
同一文	17626	22231

表1より、折り返し翻訳を行うと、得られる言い換え文の数が大きく低下する。

5.4 出力例

表2に提案手法、表3にベースラインの出力結果を示す。

対訳英語文	対訳日本語	言い換え文	評価
Dreadful incidents happened one after another .	恐ろしい事件が次々に起きた。		
		すさまじい事件が次々に起きた。	○
He is good for nothing .	彼は何の役にも立たない。		
		それは何の役にも立たない男だ。	×

対訳英語文	対訳日本語	言い換え文	評価
She contemplated me for a long time .	長い間彼女はじっと私を見つめていた。		
		しばらく私を見つめていた。	○
He is working as a clerk in this office .	彼はこの会社の事務を扱っている。		
		それは事務の事務員で働いている。	×

表2, 3より、言い換えが不正解である文は、主語が正しく訳されていない場合がある。

5.5 人手評価

表4に入力文20文に対して得られた言い換え文の人手評価結果を示す。

	提案手法	ベースライン
言い換え文	20	20
正解	17	9
不正解	3	11
正解率	0.850	0.450

表 1 より、提案手法では得られた言い換え文は約 2 万 5000 文である。また、表 4 より、正解率は 0.850 である。ベースラインと比較すると、大幅な改善が見られる。

6 考察

6.1 4-Best

本研究では、大量の言い換え文を得るために、4-Best で調査をおこなう。以下に実験手順を示す。

1. OpenNMT を日英対訳学習文約 16 万文を用いて学習する
2. 英日翻訳を行い、1 文ずつ日本語文を 4 文出力する
3. 英日翻訳を行った結果の日本語文 1 文に対して、日英折り返し翻訳を行い、4 文出力する
4. 日英折り返し翻訳を行った結果、4 文の内 1 つでも対訳英文と同一であるとき、対訳日本語文と英日翻訳された結果を言い換えとする
5. ベースラインと提案手法についてランダムに抽出した 20 文の人手評価を行う

6.1.1 4-Best における実験結果

表 5 に 4-Best における実験結果を示す。ここで、学習文、入力文は提案手法と同じものを利用している。

表 5 4-Best における実験結果

	4-Best
言い換え文	134574
同一文	33442

表 1 と表 5 を比較すると、得られた言い換え文が増加したことが確認できる。

6.1.2 4-Best における出力例

表 6 に 4-Best の場合の出力例を示す。なお、言い換え文同士が同一になった場合や、折り返し翻訳が一致しない場合がある。そのため、入力文 1 文に対して 4 文出力されない場合がある。

表 6 4-Best における出力例

対訳英語文	The noise jangled on my nerves .	評価
対訳日本語文	その騒音で神経がひどくいら立った。	
言い換え文	ひどい騒音で神経がいら立った。	○
言い換え文	騒音で神経がひどくいら立った。	○
言い換え文	ひどい騒音で神経がひどくいら立った。	○
対訳英語文	The mother scolded her child harshly .	評価
対訳日本語文	母親は子供を厳しく叱った。	
言い換え文	その母親は子供を厳しく叱った。	○
言い換え文	それらの母親は子供を厳しく叱った。	×

表 6 より、言い換えが正解した文と不正解な文が混在している場合があることが確認できる。

6.1.3 4-Best における人手評価

表 7 に 4-Best における入力文 20 文に対して得られた言い換え文の人手評価結果を示す。

表 7 4-Best における入力文 20 文に対する出力の評価

	4-Best
言い換え文	49
正解	42
不正解	7
正解率	0.857

表 4 と表 7 を比較すると、得られた言い換え文が増加しており、正解率も増加している。

6.1.4 4-Best における 1 文に対する人手評価

複数の出力の内、正解した言い換えがある場合、入力文に対して言い換えを得られたとしたときの結果を表 8 に示す。

表 8 4-Best における正解した言い換えがある場合の評価

	4-Best
言い換え文	20
正解	20
失敗	0
正解率	1.000

表 8 より、4-Best では、入力文全てに対して言い換え文を得られた。また、表 5 より、4-Best では得られた言い換え文は約 13 万 5000 文である。表 7 より、正解率は 0.857 である。

6.2 4 システム

NMT は、学習を行う際に乱数が用いられる。そのため、同じ実験条件でも出力が異なる [4]。その

ため、1つのモデルで折り返し翻訳が一致しなかった文が複数のモデルで翻訳を行うことにより、折り返し翻訳が一致する可能性がある。これらから、言い換え文の増加を試みる。本研究では、4つのモデルを用いて実験を行い、N-Bestは4とする。これを4システムと呼ぶ。

6.2.1 4システムにおける実験結果

表9に4システムにおける実験結果を示す。ここで、学習文、入力文は提案手法と同様のデータを利用している。

表9 4システムにおける実験結果

4システム	
言い換え文	362917
同一文	115799

表5と表9を比較すると、得られた言い換え文がさらに増加していることが確認できる。

6.2.2 4システムにおける出力例

表10に4システムにおける出力例を示す。4-Bestと同様に、最大出力数は16文である。しかし、言い換え文同士が同一文の場合や、折り返し翻訳が一致しない場合がある。よって、入力文1文に対して16文出力されない場合がある。

表10 4システムにおける出力例

対訳英語文	The climate here doesn't agree with me .	評価
対訳日本語文	この気候は私には合わない。	
言い換え文	当地の気候は私の体に合わない。	○
言い換え文	この気候は私の体に合わない。	○
言い換え文	こんなこの気候は私には合わない。	×
対訳英語文	There was no real meat in his speech .	評価
対訳日本語文	彼の演説には意味のある内容がなかった。	
言い換え文	それは彼の演説には本当の肉類はいなかった。	×

表10より、大きく意味が異なっている文も見られる。

6.2.3 4システムにおける人手評価

表11に4システムにおける入力文20文に対して得られた言い換え文の人手評価結果を示す。

表11 4システムにおける入力文20文に対する出力の評価

4システム	
言い換え文	71
正解	58
失敗	16
正解率	0.817

表7と表11を比較すると、4システムでは、複数出力された言い換え文同士や、対訳日本語文と同一な文が増加している。そのため、正解率が減少したと考える。

6.2.4 4システムにおける1文に対する人手評価

4-Bestと同様に正解した言い換えがある場合、入力文に対して言い換え文を得られたとしたときの結果を表12に示す。

表12 4システムにおける正解した言い換えがある場合の評価

4システム	
言い換え文	20
正解	17
失敗	3
正解率	0.850

表8と表12を比較すると、精度が低下することが確認できる。また、表9より、4システムでは得られた言い換え文は約36万3000文である。表12より、正解率は、0.817である。

6.3 今後の課題

本研究では対訳コーパスが存在していることを前提にしている。そのため、今後の課題は、任意の入力に対しての言い換え生成である。

7 おわりに

本研究では、折り返し翻訳を行った結果が一致したことに着目して言い換え生成を行った。結果より、精度の向上が確認できた。また、考察より、4-Bestでは人手評価で正解率が0.857であり、約13万5000文の言い換え文を得た。今後として、任意の入力に対して言い換え生成を行い、生成された言い換え文を用いて、機械翻訳の精度向上を行いたい。

参考文献

- [1] Jianing Zhou and Suma Bhat. Paraphrase generation: A survey of the state of the art. In **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 5075–5086, 2021.
- [2] 鈴木由衣, 梶原智之, 小町守. 複数の機械翻訳を用いた言い換え認識の評価用コーパス構築に向けて. 言語処理学会第 23 回年次大会, pp. 609–612, 2017.
- [3] OpenNmt. <https://pypi.org/project/OpenNMT-py/2.0/>.
- [4] 矢野貴大, 村上仁一. ニューラル機械翻訳に乱数が与える影響.
- [5] 乾健太郎, 藤田篤. 言い換え技術に関する研究動向. 自然言語処理, Vol. 11, No. 5, pp. 151–198, 2004.
- [6] 小町守, 梶原智之. 言語学習支援のためのニューラル言い換え生成. 人工知能, Vol. 34, No. 4, pp. 451–459, 2019.
- [7] 村上仁一, 藤波進. 日本語と英語の対訳文対の収集と著作権の考察. 第一回コーパス日本語学ワークショップ, pp. 119–130, 2012.
- [8] 藤田篤ほか. 言い換え技術の研究動向: 分類体系, 知識獲得, 応用. 研究報告自然言語処理 (NL), Vol. 2013, No. 6, pp. 1–1, 2013.