

研究終了報告書

「自然言語処理の真価を引き出す言い換え生成」

研究期間：2019年10月～2022年3月

研究者：梶原 智之

1. 研究のねらい

東京五輪や大阪万博に向けて、日常会話の機械翻訳の精度向上が急務である。しかし、翻訳器の訓練に必要な100万文対を超える日常会話の対訳データをあらゆる言語対で用意することは困難である。ドメインを限定しなければ大規模な対訳データを利用できるが、訓練データと異なるドメインの翻訳は大きく精度を損なう。本研究では、ドメインに依存せず翻訳器の性能を十分に引き出すために、入力文を翻訳しやすい表現へ言い換える。

2. 研究成果

(1) 概要

安定して高品質な自然言語処理を実現するために、入力文の前編集(言い換え)に取り組んだ。代表的な応用である機械翻訳に関する実験を主に実施したが、タスクによらず自然言語処理の技術を広く改善するための前編集の技術を確立することが目的である。ACT-Xの研究期間には、言い換えの有効性を確認するための予備実験(研究テーマA)、言い換えによって機械翻訳の性能を改善する研究(研究テーマB)、言い換えが必要な入力文を検出する研究(研究テーマC)、言い換え技術そのものの改良に関する研究(研究テーマD)の大きく4つの研究を実施した。

(2) 詳細

研究テーマA: 多様な言い換えの生成と翻訳品質の推定

まず、理想的な言い換えによって自然言語処理モデルの性能をどの程度改善できるのかを調査した。入力文Sに対して言い換え辞書を用いて語句を言い換えた100種類の言い換え文Pを生成し、SおよびPをそれぞれ自然言語処理モデルに入力し、出力Tを得た。T₀からT₁₀₀までの出力の中から、正解に最も近い出力Tを選んだところ、テキスト分類とテキスト生成の両方のタスクにおいて、大幅な性能の改善を確認できた。具体的には、表1に示すように、英語からドイツ語への機械翻訳と英語文の感情極性を推定する5クラス分類において、語句の言い換えに基づく前編集が有望であるとわかった。

表1: 研究テーマAの実験結果

	機械翻訳(英語→ドイツ語)	感情極性分類(英語)
ベースライン(Transformer / BERT)	BLEU = 18.6	Accuracy = 54.8
ベースライン + 前編集	BLEU = 28.7	Accuracy = 84.9

研究テーマB: 言い換えラティスに基づくテキスト生成

本テーマでは、機械翻訳やスタイル変換などのテキスト生成タスクにおいて、入力文の言い換えによる性能改善に取り組んだ。研究テーマAと同様に、言い換え辞書を用いた語句の言い換



え生成を行い、入力文に対して複数の言い換えを得る。これらの言い換えをラティスと呼ばれるグラフで表現し、このグラフを用いたテキスト生成を行う。人間が書いたテキストの中の個々の表現は、それだけでなくはならない表現とは限らないが、提案手法ではこのような表現の揺れ（語彙選択の曖昧性）を吸収できる。例えば、提案手法では入力文中の表現が「愛大」であっても「愛媛大学」であっても同じラティスが生成されるため、翻訳器はどちらの表現を含む文が入力されても同じ翻訳文を生成できる。また、翻訳器は複数の表現の中から好きな表現に注目しつつ翻訳を行えるため、翻訳品質の改善も期待できる。

英語から日本語への機械翻訳およびカジュアルな英語からフォーマルな英語へのスタイル変換の2つのタスクにおいて、提案手法の有効性を確認するための実験を行った。ベースラインモデル(seq2seq)は、テキスト生成のために標準的に用いられる Transformer である。提案手法(lattice2seq)は、ベースラインモデルのエンコーダ部分をラティス用に拡張したモデルである。表2の実験結果から、言い換えラティスの利用によって機械翻訳とスタイル変換の両方のタスクにおいて、テキスト生成の品質改善を確認できる。本テーマは情報通信機構の藤田篤氏との共同研究として実施し、言語処理学会第27回年次大会において富士通株式会社よりスポンサー賞(研究成果3-3)を受賞した。

表 2：研究テーマ B の実験結果

	機械翻訳(英語→日本語)	スタイル変換(カジュアル英語→フォーマル英語)
seq2seq	BLEU = 37.9	BLEU = 64.3
lattice2seq	BLEU = 40.0	BLEU = 67.3

研究テーマ C: 言い換えるべき入力文の検出

本テーマでは、テーマ A やテーマ B の言い換えによって翻訳品質の改善が見込める入力文の検出に取り組んだ。2つの対訳データ(AとB)を用意し、対訳データ A で翻訳器を訓練する。この翻訳器で対訳データ B の入力文を翻訳し、出力文と正解文を比較して各出力文の翻訳品質を評価する。そして、対訳データ B の入力文の中から、高品質な翻訳ができた入力文と低品質な翻訳になってしまった入力文をそれぞれ抽出し、言い換えが必要な入力文(そのままでは低品質な翻訳になる)か不要な入力文(そのままでも高品質な翻訳ができる)かを分類するための訓練データとする。

入力文に対して言い換えが必要か不要かの2値分類器を訓練し、言い換えが必要と判定された入力文と言い換えが不要と判定された入力文をそれぞれ翻訳したところ、表3のような実験結果が得られた。自作の翻訳器(Transformer)と公開されている翻訳器(Google 翻訳)のそれぞれについて、言い換えが必要と判定した入力文に対する翻訳品質は言い換えが不要と判定した入力文に対する翻訳品質よりも著しく低いため、言い換えるべき入力文の検出に成功していることがわかる。このようにして言い換えるべきと判定された入力文を、テーマ A やテーマ B の方法で言い換えることによって、機械翻訳の性能を効果的に改善できる。この成果は、情報処理学会第84回全国大会で発表(研究成果3-5)した。

表 3 : 研究テーマ C の実験結果

	Transformer	Google 翻訳
言い換え必要と判定した入力文の翻訳(日本語→英語)	BLEU = 9.4	BLEU = 5.0
言い換え不要と判定した入力文の翻訳(日本語→英語)	BLEU = 20.6	BLEU = 15.6

研究テーマ D: 言い換え生成技術の高度化

言い換え生成の技術を改良するための研究にも取り組んだ。単語単位の言い換えに関する研究(研究成果 1-1, 3-2)と、文単位の言い換えに関する研究(研究成果 3-1, 3-4)の 2 つに取り組んだ。

単語単位の言い換えについては、日本語の単語を難解な表現から平易な表現に言い換えることを考え、言い換え辞書や評価用データセットを整備した。言い換え辞書は、日英の対訳文上で自動的に単語アライメントを求め、同じ英単語に対応する日本語の単語同士を言い換えとして収集し、各単語の難易度を推定して難解な表現と平易な表現の対を得た。評価用データセットは、クラウドソーシングを用いて文脈中での単語の言い換えを収集した。この評価用データセットを用いて日本語の言い換え技術の評価した結果、深層学習(word2vec や BERT)に基づく言い換え手法よりも、我々の言い換え辞書に基づく手法の方が高い Precision を達成できることがわかった。

表 4 : 研究テーマ D の実験結果 (単語の言い換え)

	Precision	Accuracy
word2vec	0.251	0.130
BERT	0.261	0.261
言い換え辞書	0.273	0.227

文単位の言い換えについては、カリキュラム学習の技術を用いて高品質な言い換え生成を実現した。言い換え生成のための訓練用データは、最も豊富な英語でも数十万文対という規模でしか存在せず、数百万から数千万文対の訓練用データにアクセスできる機械翻訳などのタスクとは状況が異なる。小規模な訓練用データを有効活用するために、我々はカリキュラム学習の技術の活用を考えた。自然言語処理におけるカリキュラム学習では、単語頻度や文長に応じて、解きやすい事例(高頻度や短文)から解きにくい事例(低頻度や長文)へと学習を進める方法が採用されてきた。本研究では、言い換えタスクに特化した解きやすさの指標として、言い換え前後での編集距離の大きさを検討した。つまり、多くの単語を編集するような言い換えは解きにくく、多くの単語をコピーするような言い換えは解きやすいと考える。実験の結果、編集距離に基づくカリキュラム学習が、カリキュラム学習を用いないベースラインや単語頻度や文長などに基づく既存のカリキュラム学習よりも高い言い換え性能を達成できることを確認した。

表 5 : 研究テーマ D の実験結果 (文の言い換え)

	ドメイン A	ドメイン B
カリキュラム学習なし	BLEU = 69.8	BLEU = 75.0



頻度によるカリキュラム学習	BLEU = 70.1	BLEU = 74.6
文長によるカリキュラム学習	BLEU = 69.8	BLEU = 74.9
編集距離によるカリキュラム学習	BLEU = 70.3	BLEU = 75.4

3. 今後の展開

ACT-X の研究期間には、入力文に対して多様な言い換えを生成することで機械翻訳をはじめとする自然言語処理の性能を改善する手法を提案した。今後の展開としては、大きく以下の2つがある。

- (1) 性能改善のために有効な最良の言い換え表現の推定
- (2) ブラックボックスモデルに適用可能な前編集方法の確立

前者については、現状の多様な言い換えの生成には多くの計算コストが必要になるため、有効な前編集を施した1文を生成できることが望ましい。後者については、現状ではラティスに基づく機械翻訳モデルを自前で訓練することを想定しているが、Google 翻訳や DeepL 翻訳などのブラックボックス型の自然言語処理モデルに対しても適用可能な前編集へ拡張することで、社会実装につなげていきたい。

4. 自己評価

語句の言い換えに基づく前編集の技術および前編集が有効な文を検出する技術を提案し、機械翻訳をはじめとする自然言語処理の性能改善を確認できた。これらの技術は、機械翻訳のようなテキスト生成と感情極性分類のようなテキスト分類に広く適用可能な汎用的な技術であり、自然言語処理の多くの応用タスクの性能改善のために今後適用されることが期待できる。研究目的は概ね達成できたが、主要な成果をまとめた論文が論文誌や国際会議に未採択であるため、公開を急ぎたい。

5. 主な研究成果リスト

- (1) 代表的な論文(原著論文)発表

研究期間累積件数: 1件

1. 梶原智之, 西原大貴, 小平知範, 小町守. **日本語の語彙平易化のための言語資源の整備**. 自然言語処理, 27 巻, 4 号, pp.801-824, 2020.

日本語の語句の言い換えの研究基盤となる言語資源を構築し、公開した。特に、難解な表現から平易な表現への言い換えに焦点を当て、大規模な言い換え辞書を自動構築した。また、クラウドソーシングを用いて文脈に適した語句の言い換えを収集した評価用データセットも構築し、日本語の語句の言い換え手法を評価した。

- (2) 特許出願

なし

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

1. Sora Kadotani, Tomoyuki Kajiwara, Yuki Arase, Makoto Onizuka. **Edit Distance Based Curriculum Learning for Paraphrase Generation**. In Proceedings of the ACL-IJCNLP 2021 Student Research Workshop (ACL-IJCNLP 2021 SRW), pp.229-234, 2021.
2. Daiki Nishihara, Tomoyuki Kajiwara. **Word Complexity Estimation for Japanese Lexical Simplification**. In Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020), pp.3107-3113, 2020.
3. 西原大貴, 梶原智之, 荒瀬由紀, 藤田篤. **言い換えラティスを用いたテキスト生成の性能改善**. 言語処理学会第 27 回年次大会, pp.1729-1733, 2021. 【スポンサー賞受賞】
4. 門谷宙, 梶原智之, 荒瀬由紀, 鬼塚真. **編集距離に基づくカリキュラム学習を用いたスタイル変換**. 人工知能学会第 35 回全国大会, 2021.
5. 野口夏希, 梶原智之. **日英機械翻訳におけるモデルと入力文の相性判定**. 情報処理学会第 84 回全国大会, 2022.