

どんな研究	大規模言語モデルを用いた機械翻訳では、原文の一部が翻訳されない「訳抜け」、原文にない内容が訳文に生成される「ハルシネーション」等の誤りが発生します。本研究では原文と訳文で同じ意味を持つ単語のペアを見つける単語対応技術を用いて誤りを軽減する手法を考案しました。
どこが凄い	Wikipediaから同じ対象に言及する異なる言語の文のペアを収集してモデルを訓練することで、 世界最高精度の単語対応技術 を実現しました。単語対応を用いて、訳文の単語が原文の単語をより多くカバーするように翻訳を制御することでカバー率が約5%向上し、訳抜け等が減少しました。
めざす未来	今後は、代名詞の指示対象、文体、語彙などの 一貫性を保ちながら 、流暢に長いテキストを生成できる大規模言語モデルの特長を活かしつつ、原文に対する訳文の 忠実度をさらに高めて 、特許・法律・医療など 厳密さを重視する分野 に適用可能な機械翻訳技術をめざします。

機械翻訳と単語対応

大規模言語モデルを用いた機械翻訳の課題: 訳抜けと湧き出し

入力 (原文) There was some damage to cars and buildings in Belgrade due to the snow

出力1 (訳文) セルビアの首都ベオグラードは雪で被害を受けた

訳抜け
ハルシネーション (湧き出し)

原文と訳文の単語対応が分かれば、訳抜けと湧き出しの検出が可能

スパン予測を用いた単語対応

1. データ作成: 同じ対象に言及する文のペアをWikipediaから収集

Wikipediaの言語間リンク (Wikiデータ) を利用

足利義満は、室町時代前期の …
Ashikaga Yoshimitsu was the third shogun …

足利義満在任期間、南北朝獲得統一…
Ashikaga Yoshimitsu was the third shogun …

Ashikaga Yoshimitsu war ein japanischer …
Ashikaga unit les cours du Nord et du …

2. 事前訓練: 同じ対象を言及する部分文字列 (スパン) を予測

Wikipediaデータに対するスパン予測の損失 (誤り) を最小化

Start End

訓練済み多言語モデル (Transformerエンコーダ, BERTなど)

Ashikaga Yoshimitsu was the … [SEP] 足利義満は室町前期の

3. 単語対応: 同じ意味を持つ部分文字列 (スパン) を予測

与えられた文のペアにおいて、すべての単語の対応先を一つずつ予測

Start/End

訓練済み多言語モデル (Transformerエンコーダ, BERTなど)

This is a pen [SEP] これはペンです

事前訓練の効果
(世界最高精度を達成)

日英単語対応精度: 77.6% → 83.7%
中英単語対応精度: 86.7% → 91.5%

大規模言語モデルを用いた機械翻訳

1. 対訳データを用いた継続事前訓練 (翻訳に必要な知識を付与)

原文と訳文を交互に並べたテキストの次単語予測の損失 (誤り) を最小化

is a pen これはペンです [EOS]

大規模言語モデル (Transformerデコーダ, Llamaなど)

This is a pen これはペンです

原文 訳文

2. 翻訳用の指示チューニング (ユーザの指示に従うように訓練)

プロンプトに対する応答の次単語予測の損失 (誤り) を最小化

プロンプトの次単語予測は無視

は京都が好き [EOS]

大規模言語モデル (Transformerデコーダ, Llamaなど)

英語から日本語への翻訳 英語: I love Kyoto 日本語: 私は京都が好き

プロンプト (指示+原文) 応答 (訳文)

3. 翻訳用の選好最適化 (ユーザの好みを反映するように訓練)

訳文の単語が原文の単語に対応する割合 (カバー率) を単語対応から計算

入力 (原文) There was some damage to cars and buildings in Belgrade due to the snow

出力2 (訳文) ベオグラードでは雪のせいで車や建物に被害が出た

出力2 (優先) ベオグラードでは雪のせいで車や建物に被害が出た

出力1 (非優先) セルビアの首都ベオグラードは雪で被害を受けた

単語カバー率 >

- 単語カバー率が高い訳文を低い訳文より出力として優先するように直接選好最適化(DPO)を用いて大規模言語モデルをファインチューン
- 単語カバー率が約5%向上し、訳抜け・ハルシネーションが減少

関連文献

- [1] Q. Wu, M. Nagata, Y. Tsuruoka, “WSPAlign: Word alignment pre-training via large-scale weakly supervised span prediction,” in *Proc. The 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023)*, 2023.
- [2] Q. Wu, M. Nagata, Z. Mao, Y. Tsuruoka, “Word alignment as preference for machine translation,” in *Proc. The 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024)*, 2024.

連絡先

永田 昌明 (Masaaki Nagata) 協創情報研究部 言語知能研究グループ