

どんな研究

近年、機械翻訳技術が発展し高精度で流暢な翻訳結果が得られるようになりました。しかし従来の方法ではどの訳語を用いるかを指定しながら高い精度を実現することができないため、翻訳文が意図しない表現となることがありました。本展示では指定した表現を用いた翻訳を行うために**機械翻訳の操作性を向上させる技術**を紹介します。

どこが凄い

私たちは**翻訳文に指定した語句を必ず含む機械翻訳手法**を新たに開発し、従来の方法に比べてより高速に高い精度で翻訳できるようになりました。また、この技術を用いて国際ワークショップWAT2021で開催された日英翻訳のコンペティションに参加し、**世界一位の成績を収めました**。

めざす未来

法務や特許、科学技術などの分野の文書を翻訳する際には、専門用語や固有名詞などの語句が文書内で一貫して同じ表現で翻訳されることが強く求められます。本技術を用いることで、それらの語句に対して訳語を指定できるようになり、**表現の一貫性を保つことができるようになります**。将来的にはユーザとの対話的な操作による翻訳を行うことで、**ユーザにとって望ましい表現を用いた翻訳の実現をめざします**。

指定した語句を含む機械翻訳



"Red cow is very strong."

「赤い牛は非常に強い」
を英語に翻訳すると？



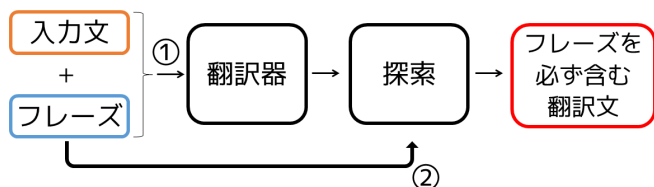
"Red **bull** is very strong."

"bull"という単語を使って



- ユーザの好きな表現を翻訳で用いたい
- 特許や法務の翻訳では専門用語などの訳語が文書内で一貫して同じであることが求められる
- 一般的な機械翻訳では出力される翻訳文に対して**表現などの指示を与えるのは難しい**
→ **機械翻訳の操作性を向上させる必要がある**
- フレーズを指定**するとそのフレーズを**必ず含む**翻訳ができる機械翻訳手法を提案
 - 高速に動作して**高い翻訳精度**を達成
 - WAT2021 語彙制約付き翻訳コンペで**世界1位**

提案手法のポイント



- ① 翻訳器に 翻訳したい文と 指定したフレーズを一緒に入力して 翻訳器がフレーズを**できるだけ含む**翻訳をするように学習する
- ② 翻訳器の出力をもとにフレーズを**必ず含む**翻訳文を探索する

ポイント:

翻訳器がフレーズを出力するように学習する ① ことにより後段の翻訳文の探索 ② が**効率化**

- ①だけではフレーズが**必ず含まれる保証がない**
- ②だけでは**翻訳精度が低く計算時間もかかる**

提案手法の効果

科学技術論文データによる実験

- 専門用語に対して**人間が正解の表現を指定した時**の英日/日英翻訳の性能評価を実施

手法	翻訳精度 BLEU (高い方が良い)	
	英語→日本語	日本語→英語
一般的なNMT [†]	44.64*	29.30*
①のみ	53.79*	41.88*
②のみ	45.38	23.22
提案手法	55.49	43.33

[†]フレーズの情報を利用していない

*翻訳結果が指定したすべてのフレーズを含んでいない

- ①と②を組み合わせることで **フレーズを指定できる**ことに加えて **翻訳精度を大きく改善**
→ 適切に正しいフレーズを指定できれば **人間と同等の精度での翻訳が実現可能**
- 指定したフレーズを必ず含む翻訳では 提案手法が ②のみよりも**3倍以上高速**に動作することを確認

関連文献

[1] K. Chousa, M. Morishita, "Input augmentation improves constrained beam search for neural machine translation: NTT at WAT 2021," in *Proc. of the 8th Workshop on Asian Translation (WAT2021)*, pp. 53–61, 2021.

連絡先

帖佐 克己 (Katsuki Chousa) 協創情報研究部 言語知能研究グループ
Email: cs-openhouse-ml@hco.ntt.co.jp