

多様な翻訳候補から適切な翻訳文を選べます

<p>どんな研究</p>	<p>機械翻訳が幅広い翻訳候補を出力できれば、その中から文脈やTPOに合わせた適切な文を人が選ぶことでより品質の高い翻訳を容易に実現できます。しかし、従来の技術では似通った翻訳候補しか出力されないという問題がありました。この展示では、より多様な翻訳候補を生成する技術を紹介します。</p>
<p>どこが凄い</p>	<p>大量の対訳データから入力に近い様々な文を自動で検索し、それらを参考に多様な翻訳候補を生成する手法を考案しました。従来手法は翻訳精度が低下する問題がありましたが、本手法では検索で得られた文を活用することで翻訳精度を保ったまま、大幅に多様性を向上することが可能になります。</p>
<p>めざす未来</p>	<p>本技術が進展すれば、TPOに合わせた翻訳や単語の誤りなどを人が簡単に修正できるようになり、編集の負担が大幅に軽減できます。また検索に使うデータを変えるだけで多くの状況に合わせた翻訳候補を出力できるようになり、様々なニーズに適した高精度な機械翻訳の実現が期待できます。</p>

翻訳候補の多様性

このカメは主に何を食べますか？

- What does this **turtle** mainly eat? (ウミガメ)
- What does this **tortoise** mainly eat? (リクガメ)



文脈や状況によって適切な翻訳文は異なる
→ 機械翻訳で複数の翻訳候補が提示できると、より正確な翻訳を人間が選択でき、翻訳精度が高められる！

翻訳候補を出力する難しさ

従来手法では

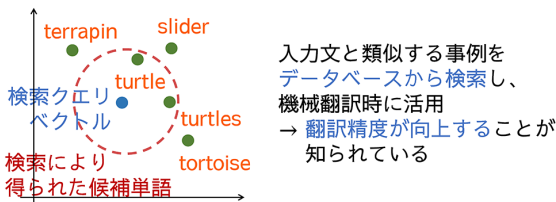
- 1 似たような翻訳候補が多くなってしまう
- 2 各候補の翻訳精度が下がってしまう

本研究のポイント

検索に基づく機械翻訳を応用することで、多様かつ高精度な翻訳候補を生成可能に！

検索に基づく多様な翻訳候補の生成

従来の検索に基づく翻訳手法

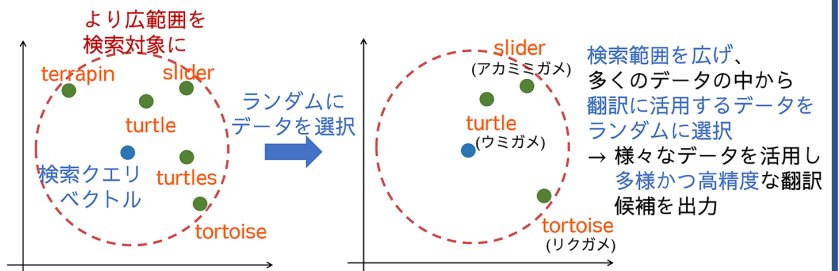


What does this **turtle** mainly eat?
What do these **turtles** mainly eat?

従来手法では検索で得られる候補単語が類似してしまうため似通った翻訳候補文ばかりが生成されてしまう

U. Khandelwa et al., "Nearest Neighbor Machine Translation," in Proc. International Conference on Learning Representations (ICLR), 2021.

多様な翻訳候補生成のための提案手法 [1]



What does this **turtle** mainly eat?
What does this **tortoise** mainly eat?
What does this **slider** mainly eat?

幅広い検索結果を使用することで多様かつ高精度な翻訳候補を生成可能に

提案手法の効果

多様性 (DP) ↑ 翻訳候補の精度 (BLEU) ↑

① 通常の機械翻訳器	31.4	42.6
② 従来の多様性を考慮する手法	35.9	40.0
③ 従来の検索に基づく手法	32.3	51.8
提案法A. ②と③の組み合わせ	42.0	48.6
提案法B. ② + ③ + ランダム選択	54.4	48.4

ランダム選択時には、検索範囲を通常の1.1~4倍に広げ適切な検索範囲を決定
DP: 出力された候補文にどれくらい異なった単語/フレーズが含まれているかを示す指標

実験結果

- ・③の検索に基づく従来手法は、翻訳時にデータベースを活用することで高い翻訳精度を達成
- ・しかし、これらの従来法①、②、③では、翻訳候補が似通ってしまい多様性の評価値が低い
- ・提案法AとBでは、翻訳精度を一定程度維持しつつ、高い多様性評価値を達成
→ 特に、検索時にランダムにデータを選択する提案法Bでは、検索に基づく手法による高い翻訳精度とランダム選択に基づく多様性向上の両方のメリットが得られる

関連文献

[1] 西田悠人, 森下睦, 上垣外英剛, 渡辺太郎, "摂動を加えたkNN機械翻訳による多様な翻訳候補の生成," 言語処理学会第29回年次大会, 2023.

連絡先

森下 睦 (Makoto Morishita)
協創情報研究部 言語知能研究グループ