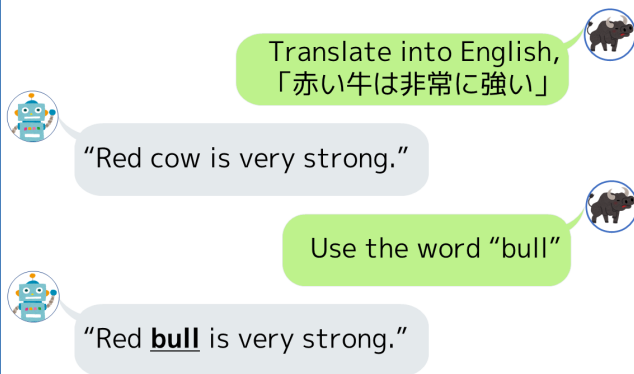


Abstract

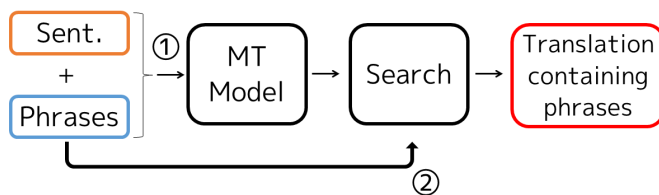
Although the recent neural machine translation achieves excellent performance, controlling its output expressions is still challenging. We propose a lexically constrained neural machine translation, **a method whose translations contain user-specified phrases**. Our method **improves translation performance** while **saving inference time** and was ranked first in the international competition at WAT 2021. When translating documents in domains such as legal, patent, and scientific, the translation of proper nouns and technical terms is strongly required to be the same expressions throughout the document. Our method will contribute to **ensuring consistency in translation** by user-specifying expressions.

Lexically Constrained Machine Translation



- Translating with favorite expressions is essential.
- For translation of patent and technical papers, the translation of proper nouns is required to be the same expression throughout the document.
- **Controlling outputs** of machine translation (MT) is still challenging.
 - The **controllability** of MT needs to be improved.
- We propose an MT method whose **translation contains given specified phrases**.
 - It achieves **high translation accuracy** and **works fast**.
 - It won **1st place** in the competition at WAT 2021.

Proposed Method



- ① Input a sentence and **specified phrases** into the model, and train the model **to output given phrases**.
- ② Search a translation **containing given all phrases** based on model outputs.

Point:

Learning the model to output phrases (①) makes the latter search step (②) **more efficient**.

- Only ① **cannot guarantee that the translation contains all given terms**.
- Only ② is **less accurate and works slower**.

Experiment

w/ scientific paper dataset

- Evaluate the performance when **a human gives the appropriate expression for technical terms as specified phrases**.

Method	Translation Accuracy BLEU (higher is better)	
	En→Ja	Ja→En
General MT [†]	44.64*	29.30*
Only ①	53.79*	41.88*
Only ②	45.38	23.22
Proposed	55.49	43.33

[†] w/o phrases information

* translations do not contain all given phrases

- **By combining ① and ②, our method can specify phrases and improve translation accuracy.**
 - Our method can **yield a human-parity score** when we specify the appropriate phrases.
- Comparing two methods whose translations contain all given terms, we confirmed our method works **more than three times faster** than only ②.

References

[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.

Contact

Katsuki Chousa / Linguistic Intelligence Research Group, Innovative Communication Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp