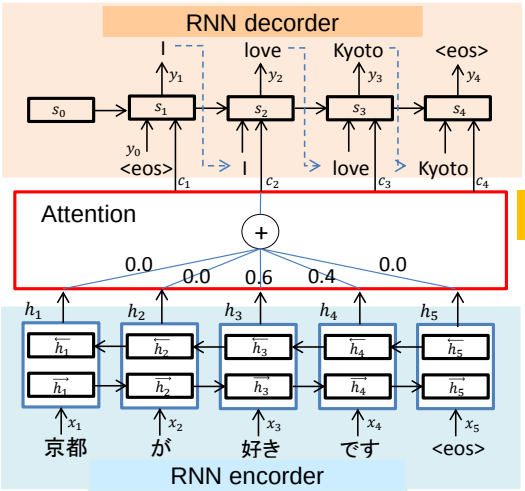
10 Can computer translate considering context?

- Testing neural machine translation's understanding of context -

Abstract

This project researches methods for evaluating a machine translation system, specifically whether the system could correctly understand the context and produce appropriate translation when multiple sentences were input. We collected examples of correct translations in which their translations changed when preceding sentences were different. They were then classified according to their linguistic phenomenon, which is the key to understanding different meanings. Tests of contextual understanding were then performed.
We have created the first database that systematically collects examples that are necessary for understanding context when translating from Japanese to English. In a Japanese sentence, the subject or object is often omitted. Different meanings of the same word are represented by different kanji characters. We therefore need a novel approach to Japanese-English translation that is different from previous research on English to French translation.

System of Neural Machine Translation



Preliminary experiment of neural machine translation using context (preceding sentence)

Translation accuracy increased but we could not understand the machine translation process.

Translation unit	Editorial	Speech	Subtitles
One sentence at a time (max 50 words)	15.17	10.04	11.00
Two sentences at a time (max 100 words)	16.36	10.30	11.37

Test of contextual understanding by neural machine translation

- ✓ Four sentences consisting of the preceding sentence and its translation and the current sentence and its translation were prepared.
- ✓ The performance of the model was evaluated by calculating the rate of correct translations against the rate of mistranslations (we were able to analyze errors by each language phenomenon).

	Anaphor / Ellipsis		Cohesion / Consistency		
	Pronoun	Article	Correspondence	Repetition	Disambiguation
Method 1	*** %	### %	…	…	…
Method 2	+++ %	--- %	…	…	…

Disambiguation test: Search for words that have multiple meanings from Japanese-English dictionary

	Source language	Target language
Preceding sentence	昨日、渋谷へ行った。	I went to Shibuya yesterday.
Input sentence	すごい人だった	Correct: There are a lot of people. Incorrect: He was a great man.

Correspondence/Repetition test: Search for structural constraints from author's introspection.

	Source language	Target language
Preceding sentence	いい時計ですね。	It's a nice clock.
Input sentence	この時計は父の形見なんです。	Correct: This clock is a memento of my father. Incorrect: This watch is a memento of my father.

Pronoun test: Create pronoun in English translation using data of analysis of omitted subject or object in a Japanese sentence.

	Source language	Target language
Preceding sentence	申し訳ありませんが、先生は午後少し遅れているんです。	I'm afraid that the doctor is running a bit late this afternoon.
Input sentence	診察するまでに20分ほどかかると思います。	Correct: It might be about 20 minutes before he can see you. Incorrect: It might be about 20 minutes before we can see you.

References

[1] M. Morishita, J. Suzuki, M. Nagata, "NTT Neural Machine Translation Systems at WAT 2017," in Proc. *The 4th Workshop on Asian Translation* (WAT-2017), 2017.
[2] R, Bawden, R. Wennrich, A. Birch, B. Haddow, "Evaluating Discourse Phenomena in Neural Machine Translation," in Proc. *The 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics* (NAACL-2018), 2018.

Contact

Masaaki Nagata Linguistic Intelligence Research Group, Innovative Communication Laboratory