Translation of Textbooks

Because NTCIR QALab tasks target the world history subject of Japanese university entrance examinations, they provide Japanese world history textbooks for knowledge bases (KB). In English task, we can use the machine translated (MT) textbooks (statistical MT, in 2015).

We tried the latest neural MT to the Japanese textbooks, however, evaluation score got worse compared with statistical MT (from 0.063 to 0.056 in ROUGE-1, for NTCIR QALab-2 phase-1 dataset). The latest neural MT mistranslated some difficult terms.

We propose the following two-step approach to build knowledge bases by utilizing Wikidata.

1. Pre-translation
To translate nouns correctly, the nouns in the Japanese textbooks are translated into English by a bilingual world history term corpus which is developed by using Wikidata. As a result, Japanese-English mixed texts are generated.

2. Machine Translation
The pre-translated texts are translated by the latest neural MT (Bing Translator), and full English texts are obtained.

Evaluation and Discussion

Our system took the first place in ROUGE-1 and ROUGE-2 of the four end-to-end submissions. However, the number of grammatical and semantic errors were relatively greater than those of other submissions. We consider that Japanese-English mixed texts sometimes cause problems during MT process.

Baseline and Sentence Scoring

Our system is a fork of Sakamoto’s multilingual QA system (FeliscatusZero-multilingual [1]) for world history essays.

We added an English open textbook (Boundless.com) since ROUGE-1 improved.