Multi Politeness-Domain Neural Machine Translation for Japanese and Korean

Henry Li Xinyuan, Jerry Chen, Ray Lee

Autumn 2021

1 Training Data

1.1 Choice of Corpus

Countless corpuses of Japanese exist on the Internet, yet the ones that would be suitable for our needs are far and few between. There is typically a strong correlation between formality and context, which is not bad news for us since relying purely on morphology to label formality would have problems of its own. However, we want to avoid introducing into our corpus large chunks of sentences with the same context in the same formality domain, lest any of our models learns to classify contexts rather than formality. Many such examples of bad corpuses exist, such as the corpus of Japanese legal documents: in Japanese, all legal documents are written in informal form (contrary to what one might assume); we must be very careful when using such corpuses by balancing and mixing them with corpuses from other sources and with different formality domains. Examples of good corpuses include the subtitle corpus, although the translation quality of some of the sentences in that corpus has been questioned.

1.2 Politeness Labels

We designed our model to be able to handle both translation and formality classification. While extracting a representation for politeness from the automatically extracted features in a neural network pipeline isn't impossible, that is not what we are trying to achieve. Rather, we would train our model under a supervised learning scheme, where each sentence has a corresponding ground truth translation and formality label attached.

One of the earliest roadblocks we faced is the scarcity of such sentence-formality pairs. Such corpuses are extremely difficult to find in sufficient quantities that would allow for adequate training of a neural classification model. Human annotation is unfortunately not so accessible for Japanese (in terms of pricing) as some other languages. As such, we devised a number of ways to generated such sentence-formality pairs.

1.2.1 Procedural Generation of Politeness Labels

Fortunately for us, this is a topic that had been studied previously by Feely et al. [FHG19], which in turn was based on the Kyoto Text Analysis Toolkit (KyTea) [NNM11]. In their case, Japanese was the destination language; the task was to generate Japanese sentences that would match the formality levels of the input. The authors identified verb suffixes and copulas as the keys for indentifying sentence formality, with short-form corresponding to informal sentences and long-form corresponding to formal ones. They published a conversion script which would identify and convert any informal verb suffixes into formal verb suffixes, and vice versa.

We start from their script and make two important modifications. First we adapt their conversion script into a classificaiton script. Next, we observe that while long-short verb form is generally a good identifier for sentence formality, there are also situations where a verb can be grammatically constrained to being short-form despite being used in a formal context. Analysing the grammar of a Japanese sentence is rather difficult and is something we would avoid to do; as such, we infer that a sentence that has any long-form verb is always formal whereas a sentence with short-form verbs may still be either formal or informal, with the probability of the sentence being informal increasing with the number of short-form verbs in the sentence (that relation need not be linear).

We also take advantage of the fact that the formality level in a single document should remain the same. This observation can serve multiple purposes:

1 A sanity check for the outputs of our script;

2 A tool to tune our priors for formality classes: specifically, the formality label that we assign to sentences with no long-form words;

3 Simplify our calculation for classifying sentences that had been segmented into documents.

For 2, formally, we are interested in minimising the cross-entropy loss between true (document-level) formality labels and inferred formality labels from the number of long and short term verbs in the sentences:
$min(-\sum_s^S lab(s) \cdot \log(f(s)))$
Where s denote some sentence in the set of all pssible sentences S (or a sample of it, as represented by our corpus); $lab(s)$ denotes the true (document-level) formality label for sentence s; and $f(s)$ being the formality label generation function we come up with, which can be further be broken down into some function $g(n_s, n_l, l)$ which takes the count of short and long form verbs in the sentence n_s and n_l, as well as the sentence length l, as input.

1.2.2 Using Pre-trained Japanese Language Model for Politeness Labelling

While Japanese is not nearly as high-resource as English (GPT-3 was never specifically trained on Japanese, for example), there are nevertheless some available pre-trained Japanese language models that are available. One of the best recent models that were developed is the Japanese BERT trained at Tohoku University. Similar to the original BERT, this model performs a mask-filling task on Japanese sentences.

The obvious way of making use of a pre-trained language model for politeness labelling would be to fine-tuned it on the new task. However, that would require correctly labelled data, leading us to a chicken-and-egg problem. Alternatively we could use the labels that we generated in section 1.2.1, although then we would be constrained by the quality of our previous scoring function.

Thus we propose a slightly modified approach: for each sentence, we identify and mask each verb (along with its suffix conjugation) and each copula one at a time. We feed the masked sentences to the language model. We then score each sentence based on whether the language model filled the masks with long forms or short forms.

1.2.3 Other Techniques for Politeness Labelling

Some other techniques for politeness labelling of Japanese and other languages had been proposed, and we will discuss them briefly here. Dugan [Dug20] proposed generating politeness labels from the corresponding English translation, an idea we didn't find convincing due to the inherent problems associated with inferring formality from English which has relatively few clear markers for formality, not the mention the problem with noise introduced by otherwise perfect translations with incorrect formality.

2 Training Task

2.1 Politeness-domain Identification

References

[Dug20] Liam Dugan. *Learning Formality from Japanese-English Parallel Corpora*. 2020. URL: http://liamdugan.com/static/thesis-bb3055b500391857a7b237a22a6f18e5.pdf.

[FHG19] Weston Feely, Eva Hasler, and Adrià de Gispert. "Controlling Japanese Honorifics in English-to-Japanese Neural Machine Translation". In: *Proceedings of the 6th Workshop on Asian Translation*. Hong Kong, China: Association for Computational Linguistics, 2019, pp. 45–53. URL: https://www.aclweb.org/anthology/D19-5203.

[NNM11] Graham Neubig, Yosuke Nakata, and Shinsuke Mori. "Pointwise Prediction for Robust, Adaptable Japanese Morphological Analysis". In: *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT)*. Portland, Oregon, USA., 2011. URL: http://www.phontron.com/paper/neubig11aclshort.pdf.