

Chloe Quinto

HW 3 - Adversarial Machine Translation Output

I pledge my honor that I have abided by the Stevens Honor System - Chloe Quinto

For this assignment, I translated from English to Filipino and vice versa.

	Input Sentence	Machine Translation	Correct Translation Output
1	Find your patience before I lose mine	Hanapin ang iyong pasensya bago mawala ang minahan ko	Hanapin mo ang pasensya mo bago mawala aking pasensya
2	I'm down in the dumps	Bumaba ako sa mga dump	Malungkot ako
3	I'm dying to see the new movie	nagmamtay ako upang makita ang bagong pelikula	Hindi ako makapaghintay na panoorin ang bagong pelikula
4	Para po	Para po	Stop the bus!
5	Ano'ng chika mo	What a chika you are	What's your gossip?
6	sayang	Honey	What a waste
7	Nanay tatay gusto kong tinapay Ate kuya gusto kong kape	Mom dad I want bread I really want coffee	Mom dad I want bread Sister brother I want coffee
8	Alab ng puso sa dibdib mo'y buhay	A heart attack on your chest is alive	With fervor burning Thee do our souls adore
9	hay naku	hay naku	Oh my!
10	bahala ka	it's up to you	Whatever

Discussion

Generally speaking, Google Translate fails to translate phrases/sentences that convey sarcasm, colloquialism, and idioms. As an example of sarcasm, “Find your patience before I lose mine” translates to “Find your patience before I lose mine”, however the mine used in the Filipino translation refers to the mining of coal and other minerals. As an example of idioms, the usage of the phrase “I’m down in the dumps” generally means “I’m depressed”. Google Translate takes this literally and basically says in Filipino “I went down into the dumps”. As an example of colloquialism, several phrases such as “hay naku” would generally mean “Oh my!”, yet Google Translate does not translate this and even suggests translating this Filipino to Japanese -> English.

An interesting find that I didn’t expect was the childhood rhyme in row 7. In the second line, “Ate kuya gusto kong kape” should translate to “sister, brother I want coffee”. However, it translates to “I really want coffee”. I think this had to do with the word “ate” could have been interpreted as “ate” as in “I just ate” in English. I guess since “ate” came first in the sentence, it also took precedence in “kuya” and

overshadowed the fact that the word “brother” was mentioned.

Ideas of Improvement

I have a few ideas of improving the neural network. One of them is modifying the dataset. Google Translate uses data from the literature that it has collected over the years. However, that does not necessarily cover nuanced areas of speech and text such as sarcasm, colloquialism, and idioms. The dataset could include speech taken from current news platforms and articles. Maybe more informal texts found in a variety of neutral (not specific to a topic) forums could possibly help. I would not suggest taking data from text messages as shorthand notation could interfere with the result.

Another improvement is to maybe add residual layers to deep LSTM RNN. There is a paper Residual LSTM: Design of Deep Recurrent Architecture for Distant Speech Recognition that found that residual LSTM provides spatial shortcut path from lower layers for efficient training of deep networks with multiple LSTM layers. It also provides a reduction of 10% of network parameters in comparison with highway LSTMs.