

Project Topic: Machine Translation using Encoder Decoder Model using Recurrent Neural Network (RNN)

1. What is the problem you are trying to solve? Write 1-2 paragraphs describing the task and what NLP areas it relates to.

Language translation has become very common with Google translation pioneering in this field. This project aims to use machine translation model to translate German language sentences to English. The idea is to use Recurrent Neural Network over the encoder-decoder NLP model to attain the end result of translating a sentence from German to English.

The application will read sentences that will have <EOS> tagging and will be encoded to an internal representation of lengths that are bounded. The decoder model will then be used to collect the words from the encoded input. The decoder model stops whenever the <EOS> tagging is reached. The encoder-decoder model uses the Long short term memory (LSTM) for first encoding and then decoding.

2. Example Application – 3 Points

Why are you trying to solve this problem? Why should people care about this problem?

Write 1 paragraph giving an example application of the task. For example.

The fundamental part of being human is the ability to communicate with one another. There are roughly 6,600 different languages around the world today. Machine translation using encoder decoder gives people from different countries and ethnic groups to convey the original tone or intent of message, taking into account cultural and regional differences between source and target languages. The major categories for machine translation applications are for industries in business use (like Government, Software and technology, Military and defense, Healthcare, Finance, E-commerce, Education, Media), online/ app for consumers use (Text-to-text, Text-to-speech, Speech-to-text, Speech-to-speech, Image-to-text). To meet these demands, many technological companies are investing on this. The recent advances in deep learning and neural network technologies led to increased quality of translations. The shift from phrase based approach (2006) used in google translator to the deep learning technologies (2016) resulted in 60% increased accuracy. Google now supports over 100 languages for translation.

3. Relevant Papers – 3 Points

1. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation- Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio
(Submitted on 3 Jun 2014 (v1), last

revised 3 Sep 2014 (this version, v3))

2. [Dahl et al.2012] George E. Dahl, Dong Yu, Li Deng, and Alex Acero. 2012. Context-dependent pretrained deep neural networks for large vocabulary speech recognition. IEEE Transactions on Audio, Speech, and Language Processing, 20(1):33–42.
3. [Socher et al.2011] Richard Socher, Eric H. Huang, Jeffrey Pennington, Andrew Y. Ng, and Christopher D. Manning. 2011. Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. In Advances in Neural Information Processing Systems 24.

Team Members:

Vinit Kumar - vxk170008
Aniket Ashok - axa170068
Suraj Raghavendra Vadvadgi - srv180000