

CSCI-GA.2565-001 Machine Learning: Project Proposal

Team: Aman Chopra, Niharika Sinha, Siddhartha Singh

Project: Abstractive Text Summarization using Multilayer Bi-directional LSTM and Attention (using Beam Search to increase efficiency)

Which problem are you solving:

The amount of textual information is constantly increasing and it becomes increasingly important to summarise the data to get the gist of the data. The two famous techniques for the same are extractive and abstractive summarization. While extractive summarisation focuses on just extracting the important sentences and concatenating them, abstractive summarization is a better approach as it involves paraphrasing using novel sentences. We are planning to implement a abstractive model using Bi-directional LSTM with an attention layer. Since the problem is computationally expensive we are exploring concepts such as Beam Search to increase the efficiency of the model. Increasing the efficiency will definitely be traded off with a decrease in accuracy (or maybe not?). We will also try to understand these effects in our experiments.

What have people tried before?

There are a lot of papers starting from unidirectional LSTM to bidirectional LSTM without attention layer to with different attention mechanisms. For example, [1] uses different models with unidirectional and bi-directional LSTM for abstractive text summarization. [2] explored stacked bi-directional LSTM (with encoder-decoder architecture) with an attention layer on Amazon's database of food and product reviews.

[1]

<https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2749095.pdf>

[2] Kovačević A., Kečo D. (2022) Bidirectional LSTM Networks for Abstractive Text Summarization. In: Ademović N., Mujčić E., Akšamija Z., Kevrić J., Avdaković S., Volić I. (eds) Advanced Technologies, Systems, and Applications VI. IAT 2021. Lecture Notes in Networks and Systems, vol 316. Springer, Cham. https://doi.org/10.1007/978-3-030-90055-7_21

If applicable, which data will you use? Do you have access to this data?

We are planning to use the freely available CNN DailyMail English-language dataset containing just over 300k unique news articles as written by journalists at CNN and the Daily Mail. We have also sent the access request for the Gigaword dataset which is a Headline-generation on a corpus of article pairs from Gigaword consisting of around 4 million articles.

What do you expect to find?

We expect to explore Word2Vec algorithms like skipgram and find the impact of different architectures and hyperparameters like state size, dropout, and activation functions of the LSTM model on the accuracy of the dataset. We also expect to land into issues of Out of Vocabulary words. We will try to find the effect of Beam search on the efficiency and accuracy of the model.

What do you expect to be challenging?

The implementation of the existing architecture of the given problem statement in itself is challenging. On top of that, we are using different mechanisms like adding an attention layer and using concepts such as Beam Search. A challenging part would be to take care of the Out of Vocabulary words; taking care of which in itself is a good project topic. If

time permits we might atleast provide some solutions to the problem. We also expect the computation time to be huge for the the size of the dataset.

What's an example of an experiment that you will try to run? What are the possible results of such an experiment?

We are planning to do many experiments one of which can be experimenting with the number of layers in the multilayer stacked LSTM. One possible result might be as we increase the number of layers, the accuracy might increase up to a certain layer but the task will be computationally expensive. Since the dataset is huge and the computational resource required for training will be a lot. Still we have planned to use ROGUE to evaluate score even if it is for a smaller dataset. Another experiment will be to find the effect of using Beam Search on the efficiency and accuracy of the model.

If applicable, how can you tell whether your idea works or not?

We are planning to use the two popular evaluation metrics: ROUGE((Recall-Oriented Understudy Gisting Evaluation) & BLEU((Bilingual Evaluation Understudy). Both of these metrics are designed to efficiently evaluate computer-generated texts and compare the similarity of the generated summary to the reference summary, each method scoring similarity between 0 and 1. This will tell us how good our model is and whether the changes in the model work or not.

What would someone else in the class learn from the project?

We think that abstractive text summarization is an important topic. Not only does it cover important aspects of Natural Language Processing, but it also explores encoder-decoder architecture and LSTM networks. The

architecture and advancements on these topics will surely be a good learning experience.

References:

<https://github.com/harvardnlp/sent-summary>

https://huggingface.co/datasets/cnn_dailymail

<https://huggingface.co/datasets/gigaword>

https://humboldt-wi.github.io/blog/research/information_systems_1920/nlp_text_summarization_techniques/

<https://machinelearningmastery.com/attention-long-short-term-memory-recurrent-neural-networks/>

<https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2749095.pdf>

https://link.springer.com/chapter/10.1007/978-3-030-90055-7_21#citeas

<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9528391>

<https://github.com/DeepsMoseli/Bidirectional-LSTM-for-text-summarization->

<https://medium.com/hackernoon/multilayer-bidirectional-lstm-gru-for-text-summarization-made-easy-tutorial-4-a63db108b44f>