

NLP Final Project Proposal

Ange Olson, Pavani Samala, Yaxin(Janet) Zhuang

What problem did you select and why did you select it?

The problem we are looking to solve is explanation and interpretability in generating news headings. Papers such as [Liu and Lapata, 2019](#) employ pre-trained models (e.g. BERT) to generate these headings, but fail to explain why, particularly for shortened summaries, specific words might be chosen as representative of the article over others. To solve this problem, we aim to use [LIME](#) to explain their presented model.

Developing computer-based summaries can help to eliminate human bias in determining the headline of an article, i.e. click-bait designed to catch readers' attention but not properly describe the article itself. Furthermore, explanation of the headline chosen allows readers to maintain a sense of trust that what they are going to read is true to the summary they are presented with.

What database/dataset will you use? Is it large enough to train a deep network?

We aim to use the [XSum dataset](#) provided through the [2019 paper](#) mentioned above, as it contains one-sentence labels for each news article, corresponding better to the quick summaries/titles we are looking to interpret. The dataset contains 226,711 articles—large enough for a deep network.

What deep network will you use?

We will be using the pre-trained BERT model (an encoder) with the additional transformer layers added by the paper's authors, accessed via [GitHub](#).

Will it be a standard form of the network, or will you have to customize it?

We will use a standard version of the network provided in the paper.

What framework will you use to implement the network? Why?

The pretrained models have been implemented by using Pytorch, so we will also be utilizing the same framework.

What reference materials will you use to obtain sufficient background on applying the chosen network to the specific problem that you selected?

We will be referring to the following papers and website for ideas and background information:

<https://arxiv.org/pdf/1908.08345v2.pdf>

<https://github.com/marcotcr/lime>

<https://paperswithcode.com/dataset/xsum>

<https://github.com/nlpyang/PreSumm>

How will you judge the performance of the network? What metrics will you use?

Similarly to the 2019 paper, we aim to use the ROUGE F1 scores (ROUGE-1, ROUGE-2, ROUGE-L).

Provide a rough schedule for completing the project.

11/8/22: Finish proposal, create GitHub

11/9/22-11/16/22: Download the code, re-create the results of the paper

11/17/22-11/23/22: Familiarize with LIME, work on interpretability of paper results

11/24/22-11/26/22: Break

11/27/22-11/30/22: Time permitting, test on another dataset—perhaps on more sensationalist news, like BuzzFeed

12/01/22: Start report

12/12/2022: Submit project