**CS 224N Final Project Proposal**

1. Team

|  |  |
| --- | --- |
| **Name** | **SUID** |
| Stylianos Serghiou | sstelios |
| Peter Li | peter888 |
| Apurva Pancholi | apurva03 |

1. Mentor

We discussed our project with Richard Rocher, but we did not formally arrange for him or any other TAs to be our mentors.

1. Problem description

Creating summaries of long texts, whether that pertains to medical research articles or patient records, is an essential, yet arduous and time-consuming aspect of medicine. We are interested in exploring state-of-the-art algorithms capable of addressing such complicated summarization tasks. To do so, we have agreed with Richard Rocher to work on replicating his recent paper on arXiv: “A Deep Reinforced Model for Abstractive Summarization.”

Though RNN-based encoder-decoder models have achieved good performance on short input and output sequence for abstractive summarization, these models often include repetitive and incoherent phrases. Furthermore, models trained only with supervised learning often assume ground truth is provided at every step during training, exhibiting exposure bias. The approach detailed in the paper we aim to replicate attempts to address those problems.

1. Data

Modiﬁed version of the CNN and Daily Mail dataset [1]

Collection of New York Times articles published between 1996 and 2007 [2]

1. methodology/algorithm

Our aim is to use the exact algorithm and methodology used in the paper we are aiming to replicate Briefly, as per the paper, this would be a neural network model with intra-attention that attends over the input and continuously generated output separately. In addition, this would also include a new training method that combines standard supervised word prediction and reinforcement learning.

1. related work

[1] Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. Teaching machines to read and comprehend. In Advances in Neural Information Processing Systems, pp. 1693–1701, 2015.

[2] Evan Sandhaus. The new york times annotated corpus. Linguistic Data Consortium, Philadelphia, 6(12):e26752, 2008.

[3] Romain Paulus, Caiming Xiong & Richard Socher - A Deep Reinforced Model For Abstractive Summariztaion *arXiv:1705.04304v3 [cs.CL] 13 Nov 2017*

1. Evaluation plan

We plan to use the evaluation mechanism described in the paper we aim to replicate. In particular, we aim to use ROUGE (Recall-Oriented Understudy for Gisting Evaluation)-1, ROUGE-2 and ROUGE-L. If possible, we will also try to have 100 summaries rated by human evaluators on Amazon Mechanical Turk, as was done in the original paper.