Dan Waters · danwaters@my.unt.edu
University of North Texas
CSCE 5290 · Natural Language Processing
Fall 2021

Abstractive Text Summarization

CSCE 5290 Natural Language Processing | Project Increment 1

GitHub: https://github.com/danwaters/nlp-abstractive-text-summarization
Increment 1 video link: [link]

Introduction

Team

This is an individual project researched and implemented by Dan Waters (danwaters@my.unt.edu) and is discussed from my sole perspective.

Motivation and Significance

As the volume of data increases, humans have an increasing need to make decisions faster, with less time available to mentally absorb large swaths of text. Executives are notorious for preferring more concise information over large reports, simply due to time constraints and the need to make quick, well-informed decisions. Is today's technology capable of generating short summaries that an executive could base a decision on? That is the question I am to explore in this paper.

Significant problems still remain in summarization. Some approaches may produce factually incorrect information in sentences which are otherwise grammatically correct. Some of these difficult problems make summarization techniques, in the current state of the art, a "nice to have" and not yet something which can be taken as factual all the time. Thus, I will also provide a list of use cases for which abstractive summarization may fulfill a business need, while identifying gaps that prevent it from encompassing other business critical use cases.

Objective

This project's primary objective is to implement and empirically evaluate a few different approaches to abstractive text summarization, identify the strengths and weaknesses of each, and determine what can be improved.

Dan Waters · danwaters@my.unt.edu
University of North Texas
CSCE 5290 · Natural Language Processing
Fall 2021

Features

At a system level, the main feature can be simply described: the system generates a summary of some input text using abstractive summarization techniques. The system will be evaluated using ROUGE metrics[1] where possible.

Increment 1 Status

Overview

Since the original project proposal, I have learned a lot more about this task, and it became necessary to overhaul my research plans. We have also recently explored some key concepts in class that I did not previously know about, such as N-grams (just covered on October 27), which are critical to this project. In fact, understanding N-grams forms a basis for the ROUGE-1 and ROUGE-2 evaluation metrics, and I previously did not know how I would measure summarization results empirically. As a result of this pivot, I don't have any preliminary results of my own work yet, but will spend the next two increments doing so.

While I did experiment with my originally proposed GAN approach[2], it was yielding terrible results and is rather outdated, so I will not be discussing it here or pursuing GANs any further for this task. More modern approaches are discussed next.

Dataset

The dataset for this project is the DeepMind Q&A dataset[3] containing stories and questions for CNN and Daily Mail news corpora.

Related Work and Background

During this increment, I discovered a helpful meta-resource on Papers With Code[4] containing modern approaches and results specific to this task and dataset. This resource clearly documents the ROUGE performance of various models (Figure 1).

[1] "ROUGE (metric)." *Wikipedia,* Wikimedia Foundation, 17 September 2021, https://en.wikipedia.org/wiki/ROUGE_(metric)
[2] Liu, Linqing et al. "Generative Adversarial Network for Abstractive Text Summarization." Accessed 20 Sep 2021, https://arxiv.org/abs/1711.09357
[3] Cho, Kyunghun. "DeepMind Q&A Dataset." Accessed 20 Sep 2021, https://cs.nyu.edu/~kcho/DMQA/
[4] "Abstractive Text Summarization on CNN / Daily Mail." *Papers with Code,* Facebook AI. Accessed 25 October 2021, https://paperswithcode.com/sota/abstractive-text-summarization-on-cnn-daily

Dan Waters · danwaters@my.unt.edu
University of North Texas
CSCE 5290 · Natural Language Processing
Fall 2021

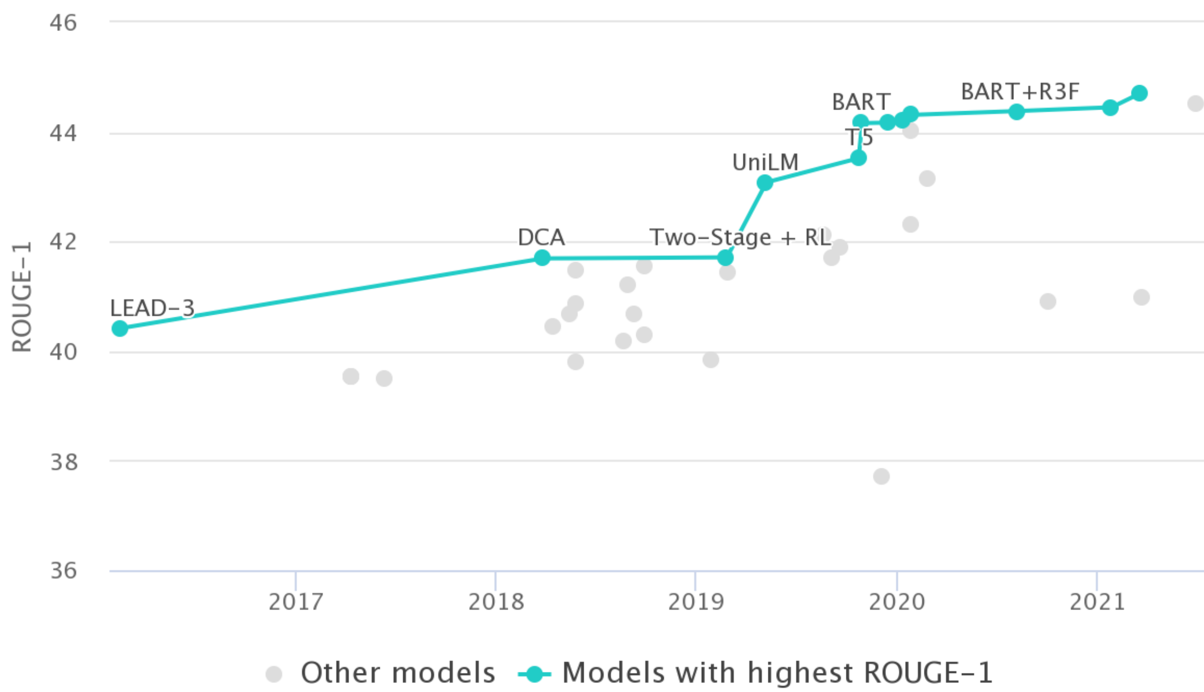

Figure 1: Abstractive Text Summarization on CNN / Daily Mail leaderboard. Source: Papers with Code

According to this leaderboard, the top models and their ROUGE-1 scores are:

Model	ROUGE-1 Score	Year
GMM-XXLarge	44.7	2021
BART + R-Drop	44.51	2021
MUPPET BART Large	44.45	2021

The benchmark score is LEAD-3 with a score of 40.42, based on sequence-to-sequence recurrent neural networks.

The papers supporting the top-performing models are far beyond my current understanding and involve some pre-trained models, so for the next increment I will be focusing on something less complex that I can implement myself in the time available.

Other approaches to this problem include sequence-to-sequence RNNs, pointer-generator networks, and transformers with attention. Transformers seem to be the favorite approach. Over the next two increments, I aim to explore all three approaches, starting with the techniques in the LEAD-3 model.

Limitations of Existing Solutions

Several limitations exist with the models available today. I have observed summarizations which:

- are grammatically correct, but factually incorrect,
- miss key context,
- are too long or too short,
- fall into repetitive sequences (particularly with the RNN approach).

A more detailed analysis of which models exhibit which behaviors will be provided in the next increments, as experimental results become available.

Detailed Design

To be completed in the next increment.

Analysis

To be performed in the next increment.

Implementation

To be completed in the next increment.

Preliminary Results

To be discussed in the next increment.

Project Management

High-Level Status Update

There is much more to do. As of October 31, I have only been able to research techniques, set up the data sets, and do some initial data cleaning, which I cover in my video. However, I believe I am set up to be successful in the next two increments of this project.

Work Completed

Dan Waters · danwaters@my.unt.edu
University of North Texas
CSCE 5290 · Natural Language Processing
Fall 2021

Task	**Description**
Dataset acquisition	Downloaded the four dataset components (CNN Stories, Daily Mail Stories, CNN Questions, Daily Mail Questions) and hosted them in my Google Drive where they can be downloaded into Google Colab
Data preprocessing	Using some ideas from this article[5], put together a notebook which preprocesses a given story. This is organized to be flexible so that other preprocessing techniques can easily be applied and modified.
Method investigation	Researched more modern techniques for abstractive text summarization using recent learnings about n-grams.

Work To Be Completed

Task	**Description**
Validate preprocessing	Make any iterative improvements to the preprocessing algorithm (e.g. embeddings, lemmatization or stemming if necessary, as required by the LEAP-3 approach)
Build sequence-to-sequence model	Build a LEAP-3 style Seq2Seq RNN
Evaluate sequence-to-sequence model	Evaluate LEAP-3 model against ROGUE-1 metric benchmark of 40.42
Build pointer-generator model	Use pointer-generator technique
Evaluate pointer-generator model	Evaluate pointer-generator technique against an acceptable benchmark

[5] Brownlee, Jason. "How to Prepare News Articles for Text Summarization." *Machine Learning Mastery*, 7 August 2017. Accessed 15 October 2021, https://machinelearningmastery.com/prepare-news-articles-text-summarization/

Dan Waters · danwaters@my.unt.edu
University of North Texas
CSCE 5290 · Natural Language Processing
Fall 2021

Build transformer model	Select a transformer / attention model to try
Evaluate transformer model	Evaluate transformer model against benchmark ROGUE-1 metric
Comparative analysis	Compare metrics and example summaries from each model to inform conclusions
Write conclusion	Use data from comparative analysis to make definitive statements about the methods explored during this project.

Dan Waters · danwaters@my.unt.edu
University of North Texas
CSCE 5290 · Natural Language Processing
Fall 2021

Resources

All work thus far has been committed to this GitHub repository:
https://github.com/danwaters/nlp-abstractive-text-summarization

Project proposal can be found here:
https://github.com/danwaters/nlp-abstractive-text-summarization/blob/main/CSCE%205290%20Project%20Proposal%20-%20Draft%20(Waters).pdf

This document (Increment 1):
https://github.com/danwaters/nlp-abstractive-text-summarization/blob/main/CSCE%205290%20Project%20Increment%201%20(Waters).pdf

References

1. "ROUGE (metric)." Wikipedia, Wikimedia Foundation, 17 September 2021, https://en.wikipedia.org/wiki/ROUGE_(metric)
2. Liu, Linqing et al. "Generative Adversarial Network for Abstractive Text Summarization." Accessed 20 Sep 2021, https://arxiv.org/abs/1711.09357
3. Cho, Kyunghun. "DeepMind Q&A Dataset." Accessed 20 Sep 2021, https://cs.nyu.edu/~kcho/DMQA/
4. "Abstractive Text Summarization on CNN / Daily Mail." Papers with Code, Facebook AI. Accessed 25 October 2021, https://paperswithcode.com/sota/abstractive-text-summarization-on-cnn-daily
5. Brownlee, Jason. "How to Prepare News Articles for Text Summarization." Machine Learning Mastery, 7 August 2017. Accessed 15 October 2021, https://machinelearningmastery.com/prepare-news-articles-text-summarization/