Text Summarization using Abstractive Approach: A Study on Enhancing the Quality of Generated Summaries with Deep Learning Techniques

Group - 8

Bhavana Prasad Kote

Rashmi Shree Veeraiah

Sawan Shivanand Beli

Department of Applied Data Science, San Jose State University

DATA 270: Data Analytics Processes

Dr. Eduardo Chan

February 17, 2023

Project Proposal

Introduction

Background

Nowadays, with the abundance of data and information on the internet, sifting through entire articles or documents to identify the main points and crucial details can be a daunting and time-consuming task. Therefore, it is necessary to generate a summary of the text that helps when the document is too lengthy. This calls for a trained model that can automatically generate summarized text given a document as input. There are two categories of text summarization which are classified as extractive text summarization and abstractive text summarization. Extractive summarization extracts salient information from the text and generates a summary without replacing words and phrases present in the original text, whereas the abstractive method understands the semantics of text and generates text summary by creating new sentences using significant information from the source text. This method of summarization is similar to the style of representation by human beings.

Our research aims to utilize Deep Learning models to produce abstractive summaries of news articles, which will allow readers to save time from reading lengthy articles, or even skipping them altogether, by providing a condensed version without compromising the original meaning. While there are numerous models that use both extractive and abstractive methods for text summarization, we propose a model that builds on existing ideas and approaches by improving the abstractive method to generate higher quality summaries. Specifically, we aim to enhance the abstractive approach to attain higher proficiency in terms of the quality of the generated summaries, specifically in terms of their fluency, coherence, and accuracy, thus setting our model apart from the existing ones.

Problem Statement:

Despite significant advances in abstractive text summarization, there remain challenges related to coherence and repetition in generated summaries.

Proposed Models:

Historically, most summarization attempts have centered around using an extractive approach, which involves identifying essential sentences or sections from the source text and reproducing them to form a summary. However, by utilizing deep learning techniques, it is possible to create relatively smooth and coherent abstractive summaries. To achieve this, the proposed solution is to leverage a recurrent neural network for the task of abstractive summarization.

In addition to RNN, we aim to develop two novel models, outlined as follows. The first model uses a transformer-based architecture with multiple encoder inputs and decoder outputs. The encoder inputs would comprise document-level, sentence-level, and contextualized embeddings, while the decoder outputs would generate abstractive summaries via a beam search algorithm. To address issues such as coherence and repetition, reinforcement learning could be incorporated in the model. In summary, these proposed improvements would result in a more robust and accurate abstractive text summarization model. It is possible that the CNN/Daily Mail dataset and the DUC2004 dataset could be used for this purpose, or other similar datasets.

The second model uses Pegasus, which utilizes an encoder-decoder architecture based on transformers. The encoder can be used to mask tokens whereas, decoder to generate these sentences. It works based on self supervised learning.

Literature Review

While much work has been devoted to using various algorithms for extractive summarization, little attention has been paid to abstractive summarization. S. Chopra et al. (2016) proposed an attentive recurrent architecture for this purpose, in which each word in the input sequence is linked to an aggregate embedding vector. This vector is a combination of the word's position and its context within the sentence. Stochastic gradient descent was used with mini-batches of 32 to minimize loss, and all models were implemented using the torch library. The decoder utilized both Elman RNN and Long-Short Term Memory (LSTM) architecture to determine which would perform better on the Gigaword dataset. The method solely relied on the encoder-decoder model, originally proposed and used in context of machine translation by Bahdanau et al., 2014.

M. Qiu et al., (2022 [v3]) proposed SueNes, a weakly supervised strategy to evaluate the quality of single-document summarizing using negative sampling, is proposed in this study. The approach creates negative summaries by picking sentences at random from the original content and evaluating their coherence and significance. SueNes achieves good correlation with human judgements and beats alternative evaluation algorithms on numerous datasets, according to the authors.

Other related work by P. Kouris et al., (2021) presents a novel approach for abstractive text summarization that uses deep learning and semantic content generalization. The proposed technique employs a pre-trained neural network to generate summaries that capture the essential information of the source text. Additionally, the approach utilizes a content generalization module that enhances the coherence and consistency of the generated summaries. The effectiveness of the approach is demonstrated through experiments on various datasets, where it outperforms other state-of-the-art methods in terms of ROUGE scores and semantic similarity. The paper's contribution to the field of abstractive text summarization is significant, and the proposed approach shows promise for generating high-quality summaries of large textual content.

Esmaeilzadeh et al., (2019) used different models including LSTM- Encoder- Decoder architecture with attention mechanism, pointer-generator mechanism, coverage mechanism and transformers to generate abstractive summaries with LSTM encoder-decoder along with attention mechanism being the baseline model. They used CNN/ Daily Mail dataset to perform this study. The main motive behind this work was to address the issue of inability of baseline model to handle out-of-vocabulary words. Using adagrad optimizer, the models were trained upon hyperparameter tuning. The results of summarization from above four models were compared with the source text and found that LSTM encoder- decoder with attention, pointer-generator and coverage mechanism showed better accuracy rate solving the issue of incorrectly handled words and reiteration of previously generated summary.

Y. Chen and Q. Song (2021) came up with a splice strategy to achieve abstractive text summarization by utilizing the summary results from extractive method and abstractive method, which was further combined and used as an input for generating an abstractive summary by incorporating the same model utilized for the abstractive kind. They used the classic TextRank algorithm proposed by Mihalcea and T. P (2004). and BART algorithms and introduced the TextRank- BART model for the study. TextRank model was utilized to extract important sentences from the source dataset and BART model to generate summary. The fusion of both the resulting summaries were again put into the BART model to obtain summary high accuracy. It was found that the Rouge-1, Rouge-2, and Rouge-L average R score was improved by 1.5, 0.5 and 1.3 compared to the scores from BART alone.

J. Chen and F. You (2020) offered a summarization model based on semantic similarity. The model used LSTM encoder and decoder with attention mechanism proposed by Bahdanau et al., (2016) where encoder represents the source text and decoder represents the generated summary text. Semantic similarity is calculated between the source and summary through the function they built using Jaccard coefficient method to check the semantic relevance in order to achieve loss function minimization. Higher Jaccard coefficient value indicates higher similarity of semantics. For the Jaccard function, the final state of the encoder and decoder were used to represent source semantic vector and summary semantic vector. It was observed that the results from the proposed model were of higher quality.

Machine learning or deep learning models for extractive summarization select important sentences or phrases from a source document to create a condensed summary. Nevertheless, human summarization entails rewording the text in a way that preserves the original meaning while improving clarity. Although it is challenging for a trained model to match the fluency and clarity of a human summary, there are abstractive summarization techniques that are specifically developed to achieve this goal.

In the book Machine Learning for Text, Aggarwal C.C. (2022) discusses two distinct techniques for achieving abstractive summarization. The first approach involves generating an extractive summary of the target text and then refining it to enhance its presentation before producing the final output. However, this method has only achieved modest success and can even negatively impact the evaluation metrics of the summary. The second method entails applying deep learning techniques directly to the text to be summarized, resulting in an abstractive summary that has been more successful in achieving the desired outcome.

Several techniques are described, one of which involves the use of RNNs. Abstractive summarization benefits from the use of deep learning techniques, which generally produce more coherent summaries than other machine learning approaches. Abstractive summarization can be achieved through the use of a technique that employs sequence-to-sequence recurrent neural networks, where the input sequence corresponds to the original text and the output sequence corresponds to the summary. A bidirectional recurrent neural network with attention is used to carry out the transformation. To train the model, it is necessary to have both original text representations and abstractive summaries.

Another approach for abstractive summarization involves utilizing transformers. Transformers have an edge over recurrent neural networks because they can efficiently process longer sequences, which is especially beneficial for summarization tasks where a training pair often involves a complete article and its summary.

The paper 'Efficient Adaptation of Pretrained Transformers for Abstractive Summarization' utilizes a variant of the transformer language model called the GPT model for summarization. The work proposed two solutions called 'domain-adaptive' and 'source embeddings'. The pre-trained GPT model's knowledge wasn't utilized to start a summarization program directly because the language style of the GPT model's original training data was different from the language style used in the summarization data used in the research work. To solve the problem of the language difference between the GPT training data and the research work data, researchers introduced a new method called domain-adaptive training (DAT) that adjusts the transformer summarization model to the language used in the data. In end-task training, the model is trained to generate an accurate summary by only producing the right summary tokens, which increases the likelihood of creating the correct summary. To help the transformer distinguish between the text of the article and the text of the summary, a special kind of embedding is used, called "source-specific embedding". This embedding assigns a value to each token in the input that indicates whether it came from the article or the summary.
The study evaluated the outcomes using automatic metrics called ROUGE. The TransformerSM model outperformed all other models on all metrics, but it performed the best on the Rouge-L metric, which is usually the most significant metric for measuring the quality of summaries.

Fattah, M. A. (2014) proposes a method for improving the automatic text summarization of multiple documents. It uses a trainable summarizer that considers various features such as word similarity, text format, title, and cue phrases to select the most important sentences. The effectiveness of these features is studied, and they are combined to create three different summarizer models: a maximum entropy model, a naive Bayes classifier, and a support vector machine. These models are then merged into a hybrid model, which ranks the sentences in order of importance to produce the final summary. The method was evaluated using the DUC 2002 data corpus and found that the new approach had better results than other techniques when evaluated using the ROUGE score.

Methodology

Dataset Selection

Our research work will utilize the DUC dataset, which is readily accessible online.The Document Understanding Conferences (DUC) are a sequence of yearly events managed by NIST, with the aim of evaluating and advancing text summarization technology. Its dataset comprises 50 news articles from various fields like politics, sports, and business, along with a reference summary and numerous summaries generated by various systems. The primary objective of this dataset was to evaluate automatic text summarization systems, with an emphasis on extractive methods. The dataset's evaluation metrics included the commonly used measure ROUGE for assessing the quality of text summaries. The DUC 2004 dataset has had a significant impact on the development of text summarization research, having been widely employed in numerous studies.

Training dataset - The CNN/Daily Mail dataset is designed for text summarization and includes human-generated abstractive summary bullets. The dataset consists of 286,817 training pairs, 13,368 validation pairs, and 11,487 test pairs. The summaries were created as fill-in-the-blank questions, with corresponding news stories from CNN and Daily Mail websites as the passages that could provide the missing entity. The source documents in the training set have an average of 766 words and 29.74 sentences, while the summaries contain an average of 53 words and 3.72 sentences (Hermann et al., 2015).

id	article	highlights
92c514c913c	Ever noticed how plane seats appear to be getting smaller and smaller? With increasing numbers of people taking to the skies, some experts are questioning if having such packed out planes is putting passengers at risk. They say that the shrinking space on aeroplanes is not only uncomfortable - it's putting our health and safety in danger. More than squabbling over the arm rest, shrinking space on planes putting our health and safety in danger? This week, a U.S consumer advisory group set up by the Department of Transportation said at a public hearing that while the government is happy to set standards for animals flying on planes, it doesn't stipulate a minimum amount of space for humans. 'In a world where animals have more rights to space and food than humans,' said Charlie Leocha, consumer representative on the committee.¬†'It is time that the DOT and FAA take a stand for humane treatment of passengers.' But could crowding on planes lead to more serious issues than fighting for space in the overhead lockers, crashing elbows and seat back kicking? Tests conducted by the FAA use planes with a 31 inch pitch, a standard which on some airlines has decreased . Many economy seats on United Airlines have 30 inches of room, while some airlines offer as little as 28 inches . Cynthia Corbertt, a human factors researcher with the Federal Aviation Administration, that it conducts tests on how quickly passengers can leave a plane. But these tests are conducted using planes with 31 inches between each row of seats, a standard which on some airlines has decreased, reported the Detroit News. The distance between two seats from one point on a seat to the same point on the seat behind it is known as the pitch. While most airlines stick to a pitch of 31 inches or above, some fall below this. While United Airlines has 30 inches of space, Gulf Air economy seats have between 29 and 32 inches, Air Asia offers 29 inches and Spirit Airlines offers just 28 inches. British Airways has a seat pitch of 31 inches, while easyJet has 29 inches, Thomson's short haul seat pitch is 28 inches, and Virgin Atlantic's is 30-31.	Experts question if packed out planes are putting passengers at risk . U.S consumer advisory group says minimum space must be stipulated . Safety tests conducted on planes with more leg room than airlines offer .
2003841c7d	A drunk teenage boy had to be rescued by security after jumping into a lions' enclosure at a zoo in western India. Rahul Kumar, 17, clambered over the enclosure fence at the¬†Kamla Nehru Zoological Park in Ahmedabad, and began running towards the animals, shouting he would 'kill them'. Mr Kumar explained afterwards that he was drunk and 'thought I'd stand a good chance' against the predators. Next level drunk: Intoxicated Rahul Kumar, 17, climbed into the lions' enclosure at a zoo in Ahmedabad and began running towards the animals shouting 'Today I kill a lion!' Mr Kumar had been sitting near the enclosure when he suddenly made a dash for the lions, surprising zoo security. The intoxicated teenager ran towards the lions, shouting: 'Today I kill a lion or a lion kills me!' A zoo spokesman said: 'Guards had earlier spotted him close to the enclosure but had no idea he was planing to enter it. 'Fortunately, there are eight moats to cross before getting to where the lions usually are and he fell into the second one, allowing guards to catch up with him and take him out. 'We then handed him over to the police.' Brave fool: Fortunately, Mr Kumar fell into a moat as he ran towards the lions and could be rescued by zoo security staff before reaching the animals (stock image) Kumar later explained: 'I don't really know why I did it. 'I was drunk and thought I'd stand a good chance.' A police spokesman said: 'He has been cautioned and will be sent for psychiatric evaluation. 'Fortunately for him, the lions were asleep and the zoo guards acted quickly enough to prevent a tragedy similar to that in Delhi.' Last year a 20-year-old man was mauled to death by a tiger in the Indian capital after climbing into its enclosure at the city zoo.	Drunk teenage boy climbed into lion enclosure at zoo in west India . Rahul Kumar, 17, ran towards animals shouting 'Today I kill a lion!' Fortunately he fell into a moat before reaching lions and was rescued .

The DUC2004 dataset, which comprises 500 news articles and four summaries each, was created for testing purposes and designed for document summarization. The dataset is composed of 50 clusters of Text REtrieval Conference (TREC) documents, and each cluster contains an average of 10 documents. The collections included in the dataset are the AP newswire (1998-2000), New York Times newswire (1998-2000), and Xinhua News Agency (English version, 1996-2000) (DUC, 2004).

```

King Norodom Sihanouk on Tuesday praised agreements by Cambodia's
top two political parties _ previously bitter rivals _ to form a coalition
government led by strongman Hun Sen. In a short letter sent to news
agencies, the king said he had received copies of cooperation agreements
signed Monday that will place Hun Sen and his Cambodian People's Party
in firm control of fiscal and administrative functions in the government.
``The protocol on cooperation between the CPP and FUNCINPEC will certainly
bring peace and progress to our nation and people,'' Sihanouk wrote.
Uncompromising enemies just a few months ago, Hun Sen and FUNCINPEC
President Prince Norodom Ranariddh agreed Nov. 13 to form a government
at a summit convened by Sihanouk. The deal, which will make Hun Sen
prime minister and Ranariddh president of the National Assembly, ended
more than three months of political deadlock that followed a July
election narrowly won by Hun Sen. Key to the agreement was the formation
of a Senate as the upper house of Parliament, to be led by CPP President
Chea Sim, the outgoing head of the National Assembly. Sihanouk, recalling
procedures used in a past government, suggested Tuesday that he should
appoint the first two members of the upper house. The remaining senators,
he said, should be selected by a method agreed upon by the new government
and the National Assembly. Hun Sen said Monday that the CPP and FUNCINPEC
had agreed that the Senate would be half as large as the 122-seat
National Assembly. Other details of the Senate, including how much
power it will be given in the promulgation of legislation, have yet
to be ironed out by the two parties.
```

References

Bahdanau, D., Cho, K., & Bengio, Y. (2016). Neural Machine Translation by Jointly Learning to Align and Translate. arXiv: Computation & Language. https://doi.org/10.48550/arXiv.1409.0473

Chopra, S., Auli, M., & Rush, A. M. (2016). Abstractive sentence summarization with attentive recurrent neural networks. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 93-98).

Chen, J., & You, F. (2020). Text Summarization Generation Based on Semantic Similarity. 2020 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). https://doi.org/10.1109/ICITBS49701.2020.00210

Chen, Y., & Song, Q. (2021). News text summarization method based on Bart-TextRank model. 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). https://doi.org/10.1109/IAEAC50856.2021.9390683

DUC. (2004). DUC 2004 Evaluation. Retrieved from https://duc.nist.gov/duc2004/duc04-eval-summarization.htm

Esmaeilzadeh, S., Peh, G.X., & Xu, A. (2019). Neural Abstractive Text Summarization and Fake News Detection. arXiv.org. arXiv:1904.00788v2

Hermann, K. M., Kocisky, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M., & Blunsom, P. (2015). Teaching machines to read and comprehend. In Advances in neural information processing systems (pp. 1693-1701).

Kouris, P., Basios, M., Tzouvaras, V., & Malakasiotis, P. (2021). Deep Abstractive Text Summarization with Semantic Content Generalization. IEEE Access, 9, 171906-171923.

Kumar, R., Novak, J., & Tomkins, A. (2016). Structure and evolution of online social networks. In Foundations and Trends® in Web Science (Vol. 2, No. 3, pp. 127-256). Now Publishers, Inc.

Mihalcea, R., & P., T. (2004). Textrank, "Bringing order into text[C]". In Proceedings of the 2004 conference on empirical methods in natural language processing (pp. 404-411)

Qiu, M., Chen, D., Jin, P., & Ren, X. (2022). SueNes: A Weakly Supervised Strategy for Single-Document Summarization Evaluation. arXiv preprint arXiv:2204.03056v3.

Hoang, A., Bosselut, A., Celikyilmaz, A., & Choi, Y. (2019). Efficient Adaptation of Pretrained Transformers for Abstractive Summarization. *ArXiv: Computation and Language*. https://arxiv.org/pdf/1906.00138.pdf

Aggarwal, C.C. (2022). Text Summarization. In: Machine Learning for Text. Springer, Cham. https://doi.org/10.1007/978-3-030-96623-2_12

Fattah, M. A. (2014). A hybrid machine learning model for multi-document summarization. *Applied Intelligence*, *40*(4), 592–600. https://doi.org/10.1007/s10489-013-0490-0