

Research-Paper-Summarization

Goda Nagakalyani

Introduction

We aim to create a system to summarize and simplify academic research papers which we hope can be extended to academic writing of any kind. Reading research papers is a skill that has to be learnt. The task is daunting and time consuming for new academics. Academics often spend hours pouring over research papers during the course of their research. We aim to optimize this by providing academics and researchers concise and informative summaries of papers. This can be extended to to simplify all manners of technical documentation, such as medical reports, audit reports, industrial product manuals, etc., based on the information used to train our system. We forecast that a successful summarization algorithm can reduce countless human work hours in large volumes thus opening avenue to new research and developments. Laypeople, hobbyists, people who do not have years of education in a particular field would be able to understand concepts and implement them if they choose to, promoting interdisciplinary progress.

Background

The main problem relating to text summarization arises from the severe lack of labeled training data. This hurdle has been overcome through this academic research paper: “A Supervised Approach to Extractive Summarisation of Scientific Papers” [1]. They give us a way to mine scientific papers and also create preliminary summaries that can be used as training data. Furthermore, they proposed multiple methods, both analytical and empirical to create text summaries using everything from scoring sentences based on word overlap to advanced methods such as Recurrent Neural Networks and LSTMs. This paper was the basis for the first half of our project, where we explored extractive methods to generate summaries.

Recurrent Neural Networks (RNNs), while not cutting edge, form the backbone of text generation today. RNNs help deal with the sequential nature of text while also accounting for the dependencies words have on each other [5]. However, they do not account for long term dependencies between words in text. Long Short Term Memory(LSTM) units overcome this by providing an attention mechanism to remember previous words in the passage. This allows for text to be generated from longer, more natural sentences and account for the contextual nature of words. This is the concept used in the second half of our project where we worked on generating summaries using abstractive methods.

The state of the art in today’s Natural Language Processing landscape is the use of Transformers namely BERT(Google) [7], GPT-2(OpenAI), et cetera. These improve upon the by LSTM concept of attention by using words that occur before and after a certain word. While our system does not compare with the state of the art today, they give a good idea of the differences between extractive and abstractive summarization methods. Implementing transformers is considered to be future work.

Project

Extractive Approach

While our main goal was to make summaries of academic papers, we experimented with different methods to achieve our results. We used a dataset created by Ed Collins, et al [1]. which was mined from ScienceDirect and contained summaries, called “Highlights”, penned by the authors themselves. Our initial approach was to score each sentence based on the words that overlapped between the main body of the paper and the author generated summaries. The top “n” sentences with the highest scores were picked as summary sentences. To do this, we used a Support Vector Machine Regressor to predict sentence scores based on document vectors generated by Gensim’s doc2vec function. An additional experiment was carried out using pre-trained Glove embeddings but the results were unsatisfactory. Our results were adequate, but did not produce very good summaries.

Generated Extractive Summary

...

[' Sustained low-intensity muscle activity has been associated with adducted and extended wrist postures during keyboard intensive tasks (Dennerlein and Johnson, 2006)',
'5, FDP=1, FDS=0N/cm2) (Brook etal',
' MCP adducted 3',
' The highest total muscle forces of three intrinsic muscles occurred in the adducted wrist posture']

...

Author Generated Highlights

...

['We quantified the effect of four wrist postures during tapping on resulting finger and wrist muscle stress (including both active and passive component)',
' Neutral wrist posture was the optimal option among four tested wrist postures when all muscles were considered',
' Extensor muscles exhibited higher muscle stresses than flexors',
' Wrist extensors stress remained higher than 4',
'5N/cm and wrist flexor stress remained below 0'

]

...

Rogue Scores for Extractive Model

We used ROUGE scores to quantify our results.

...

{'rouge-1': {'f': 0.2093023206246621,
'p': 0.23076923076923078,
'r': 0.19148936170212766},
'rouge-2': {'f': 0.05882352453479472,

```
'p': 0.06976744186046512,  
'r': 0.05084745762711865},  
'rouge-l': {'f': 0.1600446237017552,  
'p': 0.1794871794871795,  
'r': 0.14893617021276595}}  
...
```

Abstractive Approach

Next, we decided to move away from extractive summarization and experiment with abstractive summarization. Since our previous analytical method was ineffective, we decided to take an empirical approach by using a TensorFlow and Keras based Neural Network to generate our abstractive summaries. Our earlier approach did not account for word placement and inter-word dependencies. To account for this, we use LSTMs to encode context information. Keras also does not possess an inbuilt implementation of an Attention Layer, so we sourced a third-party implementation to create our model[8]. Lastly, we increased the number of layers in our Neural Network to three, to capture more latent information about words.

Highlights

```
...  
the process after bone loss group of 25 patients 17 females and 8 males were evaluated we used  
five methods for calculating fractal dimension in the region of interest fractal dimension  
decreases during the process after bone loss the methods of calculating fractal dimension give  
different results but consistent  
...
```

Predicted summary

```
...  
we propose a new model for the model of the model of the model we propose a new model for  
the model of the model we propose a new model for the model of the model we propose a new  
model for the model of the system topological pwr  
...
```

Rogue Scores for Abstractive Model

```
...  
{'rouge-1': {'f': 0.17532959684816446,  
'p': 0.3350513594592542,  
'r': 0.12188175690842054},  
'rouge-2': {'f': 0.030755802549186344,  
'p': 0.064583333333333334,  
'r': 0.02072526507309116},  
'rouge-l': {'f': 0.11545382682223433,  
'p': 0.29515996430470115,  
'r': 0.10673200005607746}}
```

...

Abstractive Model Loss Function

Summary

There are state of the art methods, such as transformers, which do a good job at summarization but none of which have been applied to the domain of research paper summarization. Our task was to summarize research papers using analytical and empirical methods, both extractive and abstractive respectively and we have tried to measure performances of both the tasks. We mined research papers under the topic of Computer Science from the ScienceDirect website using the method provided in the research paper “A Supervised Approach to Extractive Summarisation of Scientific Papers”[1]. We spent quite some time processing the data as the output from the mining algorithm was messy. We created two models for both extractive and abstractive methods, and we trained and tested them. The faults in both of our models were substantially different. While our extractive model produced grammatically coherent but irrelevant summaries, our abstractive model fell into an infinite loop, repeating the same words over and over again. Although our models did not perform well, more training time and computational power would definitely improve it. Also, we weren’t able to implement transformers, as that required more computation and time. The future scope of our project would be to implement BERT.

Conclusion

Through our experiments we have gained valuable insights into creating machine learning models for text summarization task. We created two different models for the different interpretations of the same task. We used heuristics for one model and deep recurrent neural network for the other and received vastly different results. More training data, compute power, and time for training would have certainly made our results better. We have compared both of our approach in terms of their performance and the generated summaries. The future scope of our project would be to train the models we have already created for more iterations and with all of our data. Additionally, we would like to implement more cutting-edge deep learning technologies like transforms especially BERT.

References

- [1] [A Supervised Approach to Extractive Summarisation of Scientific Papers, Ed Collins, Isabelle Augenstein, Sebastian Riedel, 2017 2](<https://arxiv.org/pdf/1706.03946.pdf>)

- [2] [Data-driven Summarization of Scientific Articles - Nikola I. Nikolov, Michael Pfeiffer, Richard H.R. Hahnloser, 2018](<https://arxiv.org/pdf/1804.08875.pdf>)

[3] [Text Summarization with Pretrained Encoders - Yang Liu and Mirella Lapata, 2019](<https://arxiv.org/pdf/1908.08345v2.pdf>)

[4] A Discourse-Aware Attention Model for Abstractive Summarization of Long Documents - Arman Cohan, Franck Dernoncourt, Doo Soon Kim, Trung BuiSeokhwan Kim, Walter Chang, Nazli Goharian, 2018

[5] Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) - Network Alex Sherstinsky, 2018

[6] Attention Is All You Need - Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, 2017

[7] Leveraging BERT for Extractive Text Summarization on Lectures - Derek Miller, 2018

[8][AttentionLayerCode](https://github.com/thushv89/attention_keras/blob/master/layers/attention.py)