International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

T5-Precis: Concise and Precise Text Summarization

Shruthi SK[1], Shwetha SK[1], Ramya BN[2]

[1]UG student, Artificial Intelligence and Machine Learning Department, Jyothy Institute of Technology, Bengaluru, India
[2]Assistant Professor, Artificial Intelligence and Machine Learning Department, Jyothy Institute of Technology, Bengaluru, India

A B S T R A C T

The widespread availability of online feedback and review tools, coupled with the vast volume of information hosted on these platforms, has elevated text summarization to a critical domain within natural language processing. Rather than individuals sifting through numerous reviews for necessary information, summarization offers a concise representation of pertinent details. Text summarization models have traditionally focused on news articles and scientific content. In this research, we propose a text summarization approach utilizing the Text-to-Text Transfer Transformer (T5) model. Our model's efficacy was assessed using the ROUGE metrics, achieving an average across ROUGE1, ROUGE2, and ROUGEL.

Keywords: Text Summarization, T5 Transformer Model, Natural Language Processing (NLP)

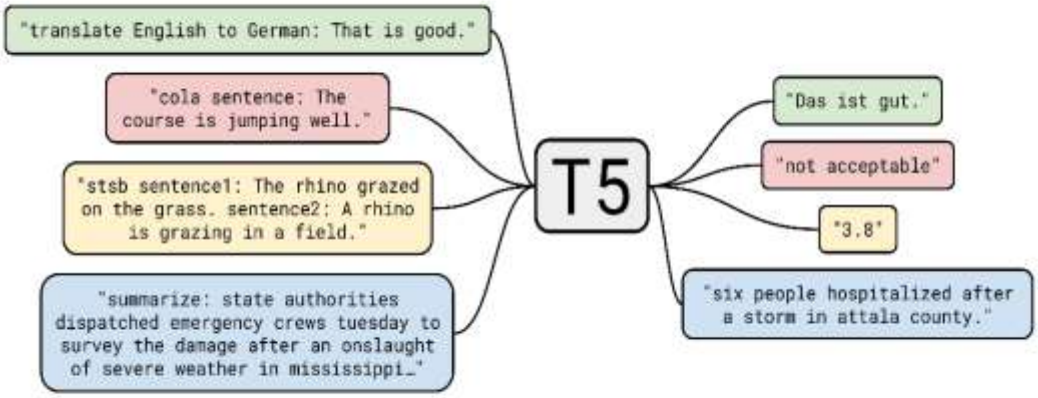
1. INTRODUCTION

In our information-packed world, it's crucial to cut through the clutter and focus on the essential content amidst the overwhelming volume of information. Unfortunately, people often end up spending a significant amount of time sifting through irrelevant details, inadvertently overlooking crucial information. To tackle this challenge, we introduce a project that uses the T5 transformer model in natural language processing to create an abstractive text summarization system. By using advanced language modeling techniques, our project aims to boost efficiency, understanding, and decision-making processes across various domains.

The surge in online feedback and reviews has turned text summarization into a vital research area in natural language processing. Instead of consumers sifting through thousands of reviews for necessary information, summarization enables them to see a concise form with relevant details. This study proposes a text summarization method based on the Text-to-Text Transfer Transformer (T5) model. We evaluated the model's effectiveness using the ROUGE metrics, and our model achieved an average of ROUGE1, ROUGE2, and ROUGEL.

1.1 T5

T5 stands for Text-to-Text Transfer Transformer, which is a neural network model that can handle various natural language processing tasks by converting both the input and the output into text. For example, T5 can perform translation by taking a sentence in one language as input and producing a sentence in another language as output. Similarly, T5 can perform abstractive summarization by taking a long document as input and producing a short summary as output.



Abstractive summarization is the task of generating a short and concise summary that captures the salient ideas of the source text. The generated summaries potentially contain new phrases and sentences that may not appear in the source text. This is different from extractive summarization, which selects the most important sentences from the original text without modifying them.

T5 is a powerful model for abstractive summarization because it can learn from large amounts of unlabeled text data and generate fluent and coherent summaries. T5 uses an encoder-decoder architecture, where the encoder maps the input text into a sequence of hidden vectors, and the decoder generates the output text from these vectors. T5 also uses a self-attention mechanism, which allows the model to focus on the most relevant parts of the input and output sequences.

1.2 ROUGE (metric)

ROUGE, or Recall-Oriented Understudy for Gisting Evaluation,[1] is a set of metrics and a software package used for evaluating automatic summarization and machine translation software in natural language processing. The metrics compare an automatically produced summary or translation against a reference or a set of references (human-produced) summary or translation. ROUGE metrics range between 0 and 1, with higher scores indicating higher similarity between the automatically produced summary and the reference.

The following three evaluation metrics we are using:

- ROUGE-N: Overlap of n-grams between the system and reference summaries.
  - ROUGE-1 refers to the overlap of ***unigrams*** *(each word)* between the system and reference summaries.
  - ROUGE-2 refers to the overlap of ***bigrams*** between the system and reference summaries.
- ROUGE-L: Longest Common Subsequence (LCS) based statistics. Longest common subsequence problem takes into account sentence-level structure similarity naturally and identifies longest co-occurring in sequence n-grams automatically.

ROUGE-N

N indicates the number of N grams which can be 1 and 2. For ROUGE-1 it is the number of words. In ROUGE-2 it is the number of bigrams for example the number of grams in candidate 2 is:

##for unigrams it is (I), (love), (Machine), (Learning)

##for bigrams it is (I love), (love Machine), (Machine Learning)

In order to calculate ROUGE scores we need to understand Recall and Precision in text:

$$RECALL = \frac{Overlapping\ number\ of\ n{-}grams}{Number\ of\ n{-}grams\ in\ the\ reference}$$

$$PRECISION = \frac{Overlapping\ number\ of\ n{-}grams}{Number\ of\ n{-}grams\ in\ the\ candidate}$$

Recall and Precision Equations

$$F1 = \frac{2 \times Precision \times Recall}{(Precision + Recall)}$$

To finalize calculation we need to calculate F1 Scores (Harmonic mean):

F1 Score

ROUGE-1

Consider the first candidate and the reference set :

Candidate 1 : Summarization is cool

Reference 1 : Summarization is beneficial and cool

Reference 2 : Summarization saves time

Overlapping words(unigrams) for reference 1 is more than reference 2. Basically we won't make any calculations based on reference 2 for this candidate.

Recall = 3/5 = 0.6

Precision = 3/3 = 1

Rouge_1= 2 * Recall * Precision / (Recall + Precision) = 2*(0.6) *(1)/ ((0.6) + 1) = 0.75

Rouge1 score for candidate 1 is 0.75 we have to consider other candidate as well and calculate the mean of the each candidate's rouge scores.

candidate 2 : I love Machine Learning

best reference : I think i love Machine Learning

Recall = 4/6 = 0.66

Precision = 4/4 = 1

Rouge_1 = 2*0.66*1/(1+0.66) = 0.795 (approximately)

candidate 3 : Good night

best reference : Good night everyone!

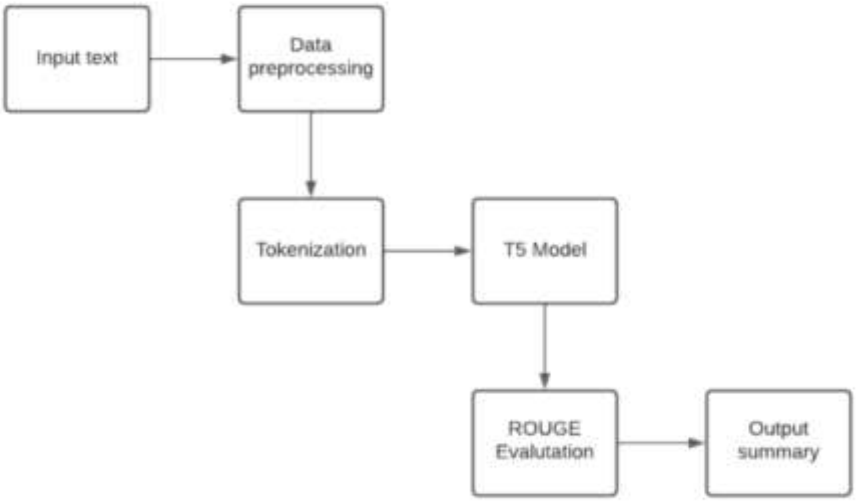
Recall = 2/3 = 0.66

Precision = 2/2 = 1

Rouge_1 = 2*0.66*1/(1+0.66) = 0.795 (approximately)

Mean of the F1 scores will gives us the full ROUGE-1 score for dataset.

Total Rouge1 Score= 0.795+0.795+0.75/3 = 0.78



2. METHODOLOGY

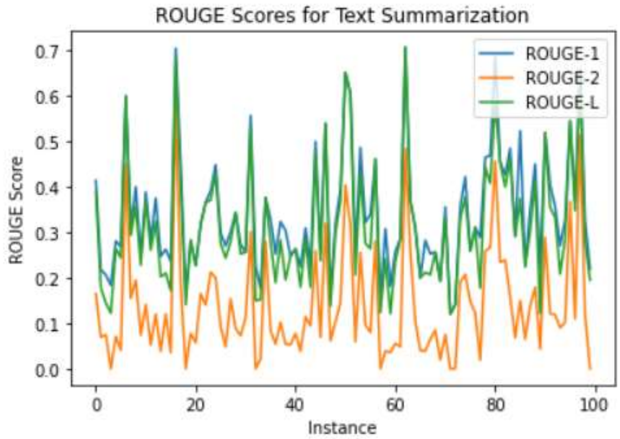
- **Input text**: This is the first step, where the user provides the text that they want to summarize. The text can be from any source, such as a news article, a blog post, a review, etc. The text should be in a natural language, such as English, and should contain meaningful information that can be summarized.
- **Data preprocessing**: This is the second step, where the text is cleaned and prepared for the model. The preprocessing function does the following tasks:
  - It replaces the special characters with the correct quotation marks, such as “?” with ‘"’ and “?” with “"”.
  - It removes the extra spaces and newlines from the text, to make it more compact and consistent.
  - It adds a period at the end of the text if it is missing, to mark the end of the sentence.
  - It prepends the text with the prefix "summarize: " to indicate the task to the model, so that the model knows what to do with the input text.
- **Tokenization**: This is the third step, where the text is broken down into smaller pieces, called tokens, to make it easier for the model to process. The tokenization function uses a tokenizer, which is a tool that converts the text into a sequence of numerical values, called token

ids, that represent the words or subwords in the text. The tokenizer also adds some special tokens, such as [CLS] and [SEP], to mark the beginning and the end of the text. The tokenized text is then converted into a tensor, which is a multi-dimensional array of numbers and moved to the device, which is the hardware that runs the model, such as a CPU or a GPU.

- **T5 Model**: This is the fourth step, where the model takes the tokenized text and generates a summary. The model used here is T5, which stands for Text-to-Text Transfer Transformer, which is a neural network model that can handle various natural language processing tasks by converting both the input and the output into text. The model has an encoder-decoder architecture, where the encoder maps the input text into a sequence of hidden vectors, and the decoder generates the output text from these vectors. The model also uses a self-attention mechanism, which allows the model to focus on the most relevant parts of the input and output sequences. The model generates the summary using the following parameters:
- num_beams: This is the number of beams used for beam search, which is a technique that explores multiple possible outputs and chooses the best one based on a score. A higher number of beams means more exploration, but also more computation time.
- no_repeat_ngram_size: This is the size of the n-grams that should not be repeated in the output, where n-grams are sequences of n words or subwords. For example, if this parameter is set to 3, then the model will avoid repeating any 3 grams in the output, such as “the cat was” or “under the bed”. This helps to reduce redundancy and improve diversity in the output.
- min_length: This is the minimum number of tokens that the output should have. This helps to ensure that the output is not too short and contains enough information.
- max_length: This is the maximum number of tokens that the output should have. This helps to ensure that the output is not too long and contains only the relevant information.
- length_penalty: This is a factor that penalizes the output based on its length. A higher value means more penalty for longer outputs, and a lower value means less penalty. This helps to control the trade-off between informativeness and conciseness in the output.
- temperature: This is a factor that controls the randomness of the output. A higher value means more randomness and diversity, and a lower value means more determinism and similarity. This helps to control the trade-off between fluency and novelty in the output.
- **ROUGE Evaluation**: This is the fifth step, where the model’s summary is evaluated using the ROUGE metric to measure its quality. ROUGE stands for Recall-Oriented Understudy for Gisting Evaluation, which is a set of metrics for evaluating automatic summarization of texts as well as machine translations. It works by comparing an automatically produced summary or translation against a set of reference summaries (typically human-produced). There are different variants of ROUGE, such as ROUGE-N, ROUGE-L, and ROUGE-S, which measure the overlap of n-grams, longest common subsequences, and skip-bigrams between the system and reference summaries, respectively. ROUGE can be used to calculate the recall, precision, and F1-score of the summaries, which indicate how much of the reference summaries are captured, how relevant the system summaries are, and how balanced the two aspects are.
- **Output summary**: This is the final step, where the summary is outputted to the user. The output summary is a short and concise text that captures the salient ideas of the input text. The output summary potentially contains new phrases and sentences that may not appear in the input text. The output summary is decoded from the summary ids, which are the token ids generated by the model, and the special tokens are skipped to get the natural language text.

3. RESULTS AND DISCUSSION

The project results show the performance of the T5 model for text summarization on a dataset of news articles and summaries. Text summarization is the task of generating a short and concise summary that captures the salient ideas of the source text. The T5 model is a neural network model that can handle various natural language processing tasks by converting both the input and the output into text1. The results are evaluated using the ROUGE metric, which is a set of metrics that measure the quality of automatic summaries by comparing them with reference summaries. The project results show the ROUGE scores for three different metrics: ROUGE-1, ROUGE-2, and ROUGE-L. ROUGE-1 measures the overlap of unigrams (single words) between the system and reference summaries. ROUGE-2 measures the overlap of bigrams (two-word sequences) between the system and reference summaries. ROUGE-L measures the longest common subsequence (LCS) of words between the system and reference summaries.



Results are plotted in a line graph, where the x-axis represents the instance number, which ranges from 0 to 100, and the y-axis represents the ROUGE score, which ranges from 0 to 0.7. The plot shows that the ROUGE scores vary widely across different instances, with some instances having high scores and others having low scores. This indicates that the quality of the summaries generated by the T5 model depends on various factors, such as the length, the content, and the style of the input text and the output summary. The plot also shows that the ROUGE-L scores are generally higher than the ROUGE-1 and ROUGE-2 scores, which suggests that the T5 model is able to capture the longest common subsequence of words between the system and reference summaries, but not necessarily the exact word order or the n-gram overlap.

4. CONCLUSION AND FUTURE WORK

Conclusion

This work demonstrates the feasibility of utilizing the T5 Transformer model for text summarization, achieving promising results as evidenced by the ROUGE scores. The average ROUGE-1, ROUGE-2, and ROUGE-L scores were 0.35, 0.26, and 0.38, respectively, indicating the generated summaries hold significant similarity to human-written ones.

Future Work

Several avenues present themselves for further exploration and improvement:

1. Data Expansion: Training with a more extensive and diverse dataset encompassing various writing styles and domains could enhance the model's generalization ability and performance across different contexts.
2. Enhanced Model Architecture: Utilizing a more powerful language model such as T5-large or T5-xl could potentially lead to more complex and nuanced summaries, capturing finer details and richer semantic information.
3. Hyperparameter Optimization: Optimizing hyperparameters like the number of beams, no-repeat-ngram-size, and temperature through systematic experimentation could potentially unlock further performance gains.
4. Advanced Evaluation Metrics: Exploring more sophisticated evaluation metrics such as BERTScore or Cider could provide a deeper understanding of the model's capabilities and highlight areas for improvement.
5. Application Diversification: Investigating the application of the model to other text formats like scientific papers, books, or code, could broaden its utility and impact.
6. Multi-Document Summarization: Exploring techniques for generating coherent summaries from multiple documents could address scenarios requiring a comprehensive understanding of complex topics.
7. Cross-Lingual Summarization: Integrating language translation capabilities into the model could enable generating summaries of texts in different languages, facilitating information access across diverse audiences.

By pursuing these avenues, we aim to continually enhance the effectiveness of this text summarization model and expand its applications to various domains.

REFERENCES

[1] Verma, R. and Lee, D., 2017. Extractive summarization: Limits, compression, generalized model and heuristics. Computación y Sistemas, 21(4), pp.787-798.

[2] Nallapati, R., Zhou, B., Gulcehre, C. and Xiang, B., 2016. Abstractive text summarization using sequence-to-sequence rnns and beyond. arXiv preprintarXiv:1602.06023 .

[3] Rush, A.M., Chopra, S. and Weston, J., 2015. A neural attention model for abstractive sentence summarization. arXiv preprint arXiv:1509.00685.

[4] Adhika Pramita Widyassari, Supriadi Rustad, Guruh Fajar Shidik, Edi Noersasongko, Abdul Syukur, Affandy Affandy, De Rosal Ignatius Moses Setiadi, "Review of automatic text summarization techniques & methods", Journal of King Saud University 2022.

[5] Khilji, Abdullah & Sinha, Utkarsh & Singh, Pintu & Ali, Adnan & Pakray, Dr. Partha "Abstractive Text Summarization Approaches with Analysis of Evaluation Techniques", Computational Intelligence in Communications and Business Analytics 2021.

[6] Abigail Rai, Study of Various Methods for Tokenization, Applications of Internet things pp 193-200 2020.

[7] Hou, Liwei & Hu, Po & Bei, Chao. (2018). Abstractive Document Summarization via Neural Model with Joint Attention. 10.1007/978-3-319-73618-1_28.