An Adaptive Approach for Generalized Text Summarization using Optimized Transformers

Nazhim Kalam
Computer Science and Engineering
University of Westminster
London, UK
nazhimkalam@gmail.com

Torin Wirasingha
Department of Computing
Informatics Institute of Technology
Colombo, Sri Lanka
torin.w@iit.ac.lk

Abstract—This research explores the ways in which to get an optimal solution using Transformer for abstractive text summarization and yet making a generalized solution that can be adapted with respect to any domain (be it hotels, movies, restaurants) and increase its performance as the system gets used over time.

Index Terms—Natural Language Processing, Machine Learning, Deep Learning, Recall-Oriented Understudy for Gisting Evaluation, Inductive logic programming

I. INTRODUCTION

Today, there is a lot of textual material available, including news stories and reviews. Text summarizing helps us quickly find the key elements of the full piece by minimizing the quantity of text. [1]. Transformers in NLP is a novel architecture that aims to solve sequence-to-sequence tasks while handling long-range dependencies with ease. It has surpassed competing neural models like CNN (Convolutional Neural Nets) and RNN (Recurrent Neural Nets) in terms of performance to appear as the dominant architecture for natural language processing [2].

A. User Reviews

A user/customer review is typically referred to be written feedback from a customer who has used a product or service. Consumers frequently use user ratings and reviews to drive their purchasing decisions. Because the review data is unstructured, it becomes more challenging for consumers to compare and understand lengthier reviews [3]

User and customer reviews are extremely important to major corporations like tourism and hospitality, as they constitute the primary engine for the country's economic growth and development. where tourists from over the world may blog about their experiences and share their reviews online in numerous formats [4].

B. Corporate Advantage

It is also known that it costs at least five times as much time and money to acquire a new customer as it does to keep an existing one, so it is important to learn how to foster customer loyalty to the brand, business, or service that is being offered. Customer satisfaction is essential to the survival of corporate industries. Understanding client expectations through their feedback or reviews helps business industries grow and fix faults [5].

On the other hand, companies like Netflix or Amazon Prime can use movie summaries to help users understand their watching pattern or their interest. Likewise, movie-related industries need to allow customers to quickly scan the summary and quickly decide whether they should be watching it or not [6].

C. Text Summarization

With the massive accumulation of information/data on the internet nowadays, it is extremely difficult to extract relevant information from numerous textual documents. The goal of text summarizing is to provide a condensed yet meaningful version of lengthy textual content [7].

We all know that text summarization has several uses in a variety of internet-based fields, including search engines that are used for querying and e-commerce sites that utilize sentiment analysis to determine client satisfaction with items [1].

However, in the movie industry, consumers may utilize text summarization to simplify customer reviews of movies, which are often lengthy and time-consuming to read. This enables users to make better decisions when they decide whether to watch a certain movie [6].

D. Abstractive and Extractive Techniques

Generally, text summarization is classified into two which are; abstractive text summarization and extractive text summarization, however, the approach for creating a hybrid model for text summarization is possible [8]. The abstractive text summarization technique aims to produce sentences on its own and then uses them to provide a coherent summary. Therefore, the summary's content will vary from the original context yet still convey the same idea [1]. Additionally, it is well-recognized that a strong abstractive summary encompasses the input's key details and is linguistically fluent [9].

The extractive text summarizing method focuses on picking out key phrases or groups of phrases from the original input content and combining them to produce a concise yet insightful text summary. It is determined which sentences should be included as parts of the summary based on the statistical and linguistic characteristics of the sentences [10]. A hybrid system

is one that combines various strategies to produce a single system. However, hybrid text summarizing systems do exist, for instance, using a combination of extractive and abstractive summarization can be utilized to generate a hybrid system that uses encoder-decoders [11]

E. NLP with Deep Learning

NLP is a method for computers to intelligently and effectively analyze, comprehend, and derive meaning from human language, as opposed to other approaches that only focus on the interactions between human language and computers. Deep learning techniques are increasingly being used in the field of AI compared to traditional machine learning approaches due to their success rates in handling difficult high-computing learning tasks [12].

In today's NLP, machine learning is prominent, but for the most part, it only involves numerically optimizing the weights of characteristics and representations that have been created by humans. Deep learning aims to investigate how computers can utilize data to create features and representations suitable for challenging interpretation tasks [13].

F. Transformers

The open-source library Transformers contain modern transformer architectures that have been thoroughly developed and are integrated by a common API. Pretraining has enabled the efficient use of this capacity for a wide range of activities, and these designs have permitted the construction of higher-capacity models. Transformers are designed to be easy for practitioners, expandable for researchers, and quick and reliable in industrial deployments [2].

It has been demonstrated that the modern generation of pre-trained language models based on transformers is rather competent at identifying syntactic signals like noun modifiers, possessive pronouns, prepositions, or co-referents, as well as semantic cues like entities and relations [14]. Hugging Face Hub offers a variety of transformer designs, including BERT, GPT2, T5, PEGASUS, and many others. The figure below represents the daily average for unique downloads of the pre-trained transformer model architectures between Oct 2019 to May 2020 [2].

[1] the research compares various other researchers' approaches taken in order to perform abstractive text summarization, these techniques include the use of transformers and other neural networks approach such as CNN and LSTM RNN networks. The research comparison table below only includes the approaches of transformers used, taken in abstractive text summarization.

G. Hyperparameter Tuning

Finding the ideal collection of parameter values to train an algorithm using in order to build a model relevant to the dataset is known as hyperparameter tuning [15]. The calculation of the performance improvement that may be obtained by changing the value of each of the considered hyperparameters from the original value to the value indicated in the target configuration set by the tuning strategy is where hyperparameters make the biggest contribution to improving algorithm performance [16].

There are several hyperparameters that play a significant role in performance enhancement; however, not all the parameters do so; just a select handful do, for example, learning rate, weight decay, number of epochs, batch size, and warm-up ratio. As a result, giving critical hyperparameters a higher priority is crucial [17].

Automated framework tools, such as Optuna, an open-source framework for hyperparameter optimization built on the Python programming language, do hyperparameter tweaking. The application of numerous hyperparameter optimization techniques, including Grid Search, Random Search, TPE, and CMA-ES algorithms, was made easier by this framework [16].

H. Generalization

Generalization now plays a significant part in resolving issues in numerous fields that are linked to the same issue. The capacity of a model to generalize to new, previously unobserved data that comes from the same distribution as the model's original data is known as a generalization [1].

Generalization is a useful strategy for starting with the foundation and improving or specializing in one's field as more unseen domain data becomes available. Therefore, the generalized solution will be able to adapt to even unseen domain data, making this solution solve a common problem in multiple domains [18].

I. Data Expansion

The quality of a machine learning or deep learning model depends on a number of factors, one of which is the amount and quality of data fed during model training. There are several approaches to increase or expand your available data, one of which is data augmentation (making use of existing data points to create new data points). Making use of new data from the user end by saving as the model is used is another way of exposing new data for model retraining [19]. When generalized models are required to adapt to become domain-specific, model retraining will be considered with new data used by the specialized domain as the application is used.

II. MOTIVATION TO ENHANCE ABSTRACTIVE TEXT SUMMARIZATION PERFORMANCE

The identified problem can also be applied to several other domains which require improving the quality of abstractive text summarization using the advanced approaches of deep learning, not only specific movie reviews, this is why a generalized solution was thought of initially [20].

As mentioned in the work of [1], syntactic and semantic issues with text summarization were the main issues that researchers were concerned with solving. and with respect to their research by exploring multiple deep learning techniques, they concluded that Transformer based models (T5 model) outperformed in all NLP tasks, this encourages the author to go deeper into the field of transformers optimization in order

to enhance the quality of text summarization and address the constraints associated with the summarizing of movie reviews.

In the domain of movie review summarization, currently, there is no research done using the latest deep learning approaches (such as Transformers) to solve this problem, standard machine & deep learning algorithms such as Naïve Bayes, and RNN has been used, and the usage of advanced deep learning approaches can be utilized in order to enhance the quality/accuracy of the text summarization.

Deep learning models take longer to train, but they provide greater accuracy since they can simultaneously automate feature extraction and classification, whereas machine learning algorithms require feature selection at first. Therefore, applying deep learning techniques will help to improve the quality of text summarization and help the user in making better decisions [1].

III. Existing Work

A. Text Summarization Systems

There were multiple studies done previously in the area of text summarization, regarding both abstractive and extractive text summarization. [6] research is related to the domain of movie reviews summarization which is the same as this project domain, where the author has developed an automatic approach to summarize lengthy movie reviews along with a feature where the users are allowed to quickly recognize the positive and negative aspects of the movie with respect to the review process with. The text summarization approach taken by the author is an extractive approach, where sentence score ranking plays a major role in creating the summary.

The study of [21] is towards the domain of e-commerce but yet related to text summarization for customer reviews on the products they sell, so the purpose is that allow other customers to make better purchasing decisions on products, therefore the hassle of going through all the reviews to making a purchasing decision can be reduced to save time, the abstractive approach is considered to create the summary, which is a better choice of approach.

The research of [4] is another extractive approach for text summarization, where the author develops a solution for generating personalized aspect-based opinion summaries using a dataset that consists of a large collection of online tourist reviews. In addition, the author has gone a step further to personalize the summary's qualities by using the user's interest. However, using abstractive summarization would be a more effective strategy but also challenging when user interest customization is considered because the sentences have been created using their own words rather than with any sentence ranking technique.

[10] research is a comprehensive comparison study with benchmarking results of various pre-trained transformer architectures such as BART, BERT, T5, PEGASUS, etc... for abstractive text summarization which is an abstractive approach. This study includes the various types of datasets used to explore each model, with the evaluations as benchmarking results. The author has also concluded the best-performing

transformer architecture as T5 by comparing the evaluation results of the study.

The study conducted by [1] is also an abstractive approach to text summarization with the addition of proper grammar and no repeated words used a deep learning approach with RNN and likewise [1] research also relates to an experimenting study with various deep learning approaches for abstractive text summarization along with the evaluation benchmarking with a goal in search for the best deep learning approach for the problem.

B. Algorithmic approaches for Text Summarization

The study of [6] starts by first focusing on feature extraction, then transforming reviews into vector spaces, and applying the Naive Bayes machine learning method for review classification utilizing an undirected weighted graph-based ranking algorithm to rank score for each review phrase in graph and then, in order to construct the extractive summary, the highest scoring sentences are selected. However, the author has limited the use of sophisticated deep learning algorithms to improve performance by solely using standard machine learning approaches to tackle the problem.

[21] research made use of the seq2seq model for text summarization along with the attention mechanism for improved accuracy and the concept net number batch word embedding model, which is superior to the glove. Utilizing a 1D convolutional layer, a max pooling layer, an LSTM layer, and finally a fully connected layer at the very end. However, the author's use of generic deep learning algorithms to handle this problem introduces a new constraint that prevents performance from being improved using the most recent deep learning strategy for NLP-related problems, transformers.

The study of [1] focuses on the author's study utilizing the encoder-decoder model with the attention layer to produce text summaries with good syntax and no repeated words. The creation of an encoder-decoder model with gated recurrent units and training it to provide an abstract summary of a piece of writing. Although the author employed deep learning, its application in production required real-time training so that it could be updated with the most recent content over time.

C. Usage of Transformers

[10] The research employed pre-trained models such as Pipeline BART, BART modified, T5, and PEGASUS to deal with text summarization as a part of the comparison study done. The ROUGE Scores were used as the evaluation measures. During the experiments, the author employed transformer designs; however, the hyperparameters used were defaulted and might be tuned for better performance. The constraints consist of concentrating on developing more reliable models that can further expand the method to produce summaries of varying lengths and applicable for multi-document summarization.

[1] The author explores deep learning methods in the broad text summarization domain to determine which method—among a collection that includes RNN, CNN, and

Transformers—performs best. The author also considers metrics for model evaluations including BLEU and ROUGE, despite using sophisticated deep learning algorithms, the author was unable to undertake hyperparameter tuning to improve the method and obtain a better outcome.

IV. Proposed Solution Approach

The diagram below shows how the approach for creating a generalized model is considered and executed using the optimized transformer models.
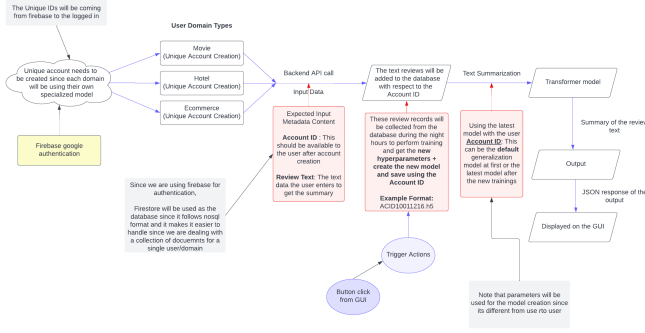

Fig. 1. Proposed Generalized Abstractive Summarization System Architecture

V. Evaluation

A machine learning model's performance, as well as its advantages and disadvantages, are understood through the process of model evaluation, which employs many evaluation measures. During the early stages of research, it's critical to evaluate models to determine their efficiency.

ROUGE also known as Recall-Oriented Understudy for Gisting Evaluation. Measures are made by comparison between an automatically generated summary/translation against a group of reference summaries (generally human-created summaries) [22]. ROUGE measures the recall, (according to how frequently the terms from the summaries created by humans appeared in those computers - generated.)

$$ROUGE - 1 = \frac{(\Sigma_{Reference\ Summary}\ \Sigma_{unigram}\ Count_{match}(unigram))}{(\Sigma_{Reference\ Summary}\ \Sigma_{unigram}\ Count(unigram))}$$

Fig. 2. Rouge-1 Score Equation

$$ROUGE - 2 = \frac{\Sigma_{s\in\{RefSummaries\}}\Sigma_{bigrams\ i\in S}\ \min\ (count(i,X),count(i,S))}{\Sigma_{s\in\{RefSummaries\}}\Sigma_{bigrams\ i\in S}\ count(i,S)}$$

Fig. 3. Rouge-2 Score Equation

$$ROUGE - L_{(candidate,references)} = max_k\{ROUGE - L_{single}(candidate,references_k)\}$$

Fig. 4. Rouge-L Score Equation

BLEU also known as Bilingual Evaluation Understudy is a metric used for the evaluation of the quality of machine-generated text by comparing it with a reference text that is supposed to be generated. [23]. BLEU measures the precision (as to how many words in the generated summaries appeared in the human-generated summaries)

$$BLEU = \frac{Number\ of\ words\ in\ the\ summary\ which\ are\ in\ gold\ standard}{Total\ number\ of\ words\ in\ the\ summary}$$

Fig. 5. BLEU Score Equation

The ROUGE score was used as the final evaluation metric for this research since the weightage of it is the best metric for this research as proven by previous work.

BART, T5 & Pegasus transformer architecture was experimented with for this research, out of which Bart gave the optimal results whereas T5 came close however Pegasus failed, there can be several reasons why Pegasus failed one of which can be due to the difference in the model architecture and the type of problem it can solve with respect to the dataset.
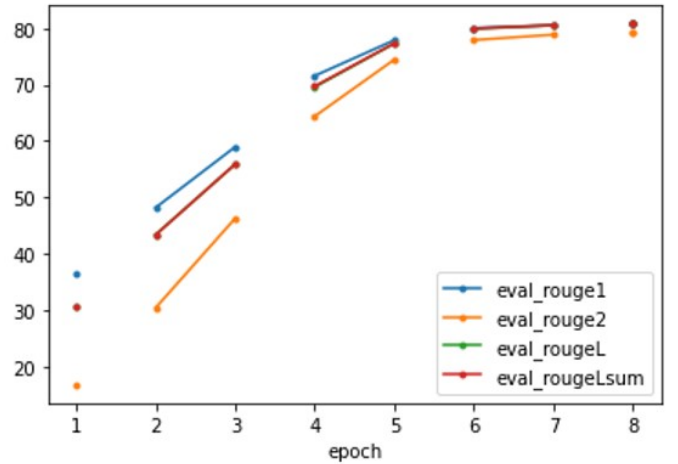

Fig. 6. Validation accuracy by the number of epochs - BART Model

ROUGE1 of 80.8, ROUGE2 of 79.42, ROUGE L of 80.8, and ROUGELSUM of 80.8 was the optimal evaluation metric result achieved from the BART model giving the best result.

VI. Future Enhancements

Future Enhancements for text summarization models can be achieved through the use of transformer hybridization. This approach could lead to improved performance, enabling the models to better handle complex natural language processing tasks. Another potential improvement is the inclusion of text paraphrase models for user reviews.

This would help to address the potential issue of inaccuracies in user-generated content. Additionally, applying keyword extraction for sentiment classification in review summaries could provide valuable insights for domain users to improve their services.
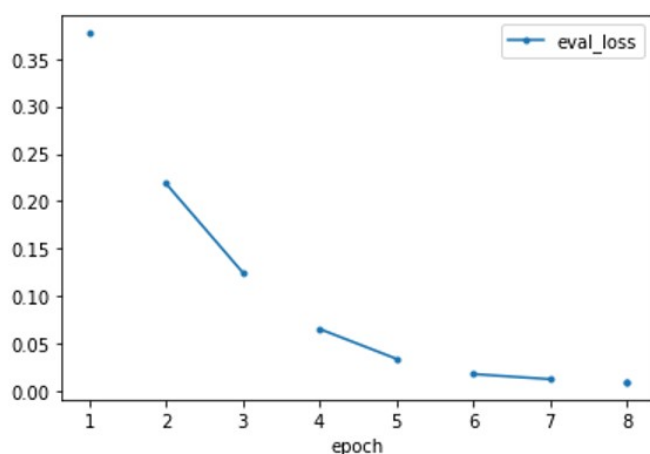
Fig. 7. Validation loss by the number of epochs - BART Model

By identifying which keywords contributed to the sentiment classification, businesses can understand their customers better and make informed decisions to enhance their services. These advancements can significantly benefit the field of natural language processing and improve the accuracy and efficiency of text summarization models.

VII. CONCLUSION

The conclusion of this study finds that the author was able to design, build, and evaluate an adaptive generalized abstractive text summarization system using optimized transformers and automated hyperparameter tuning and model retraining with respect to any domain.

REFERENCES

[1] A. G. Etemad, A. I. Abidi, and M. Chhabra, "A review on abstractive text summarization using deep learning," *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 2021. DOI: 10.1109/icrito51393.2021.9596500.

[2] T. Wolf, L. Debut, V. Sanh, *et al.*, "Transformers: State-of-the-art natural language processing," *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 2020. DOI: 10.18653/v1/2020.emnlp-demos.6.

[3] J. J. McAuley and J. Leskovec, "From amateurs to connoisseurs," *Proceedings of the 22nd international conference on World Wide Web*, 2013. DOI: 10.1145/2488388.2488466.

[4] R. Mukherjee, H. C. Peruri, U. Vishnu, P. Goyal, S. Bhattacharya, and N. Ganguly, "Read what you need: Controllable aspect-based opinion summarization of tourist reviews," *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020. DOI: 10.1145/3397271.3401269.

[5] A. Pizam and T. Ellis, "Customer satisfaction and its measurement in hospitality enterprises," *International Journal of Contemporary Hospitality Management*, vol. 11, no. 7, pp. 326–339, 1999. DOI: 10.1108/09596119910293231.

[6] A. Khan, M. A. Gul, M. Zareei, *et al.*, "Movie review summarization using supervised learning and graph-based ranking algorithm," *Computational Intelligence and Neuroscience*, vol. 2020, pp. 1–14, 2020. DOI: 10.1155/2020/7526580.

[7] T. Shi, Y. Keneshloo, N. Ramakrishnan, and C. K. Reddy, *Neural abstractive text summarization with sequence-to-sequence models*, Sep. 2020. [Online]. Available: https://arxiv.org/abs/1812.02303.

[8] A. F. Alsaqer and S. Sasi, "Movie review summarization and sentiment analysis using rapidminer," *2017 International Conference on Networks amp; Advances in Computational Technologies (NetACT)*, 2017. DOI: 10.1109/netact.2017.8076790.

[9] H. Zhang, J. Xu, and J. Wang, *Pretraining-based natural language generation for text summarization*, Apr. 2019. [Online]. Available: https://arxiv.org/abs/1902.09243.

[10] V. Gupta and G. S. Lehal, "A survey of text summarization extractive techniques," *Journal of Emerging Technologies in Web Intelligence*, vol. 2, no. 3, 2010. DOI: 10.4304/jetwi.2.3.258-268.

[11] M. Kirmani, N. Manzoor Hakak, M. Mohd, and M. Mohd, *Hybrid text summarization: A survey*, Jan. 1970. [Online]. Available: http://link.springer.com/10.1007/978-981-13-0589-4_7.

[12] M. M. Lopez and J. Kalita, *Deep learning applied to nlp*, Mar. 2017. [Online]. Available: http://arxiv.org/abs/1703.03091.

[13] R. Socher, Y. Bengio, and C. D. Manning, *Deep learning for nlp (without magic)*. [Online]. Available: https://aclanthology.org/P12-4005.

[14] A. M. Brasoveanu and R. Andonie, "Visualizing transformers for nlp: A brief survey," *2020 24th International Conference Information Visualisation (IV)*, 2020. DOI: 10.1109/iv51561.2020.00051.

[15] X. Liu and C. Wang, *An empirical study on hyperparameter optimization for fine-tuning pre-trained language models*, Jun. 2021. [Online]. Available: http://arxiv.org/abs/2106.09204.

[16] J. Joy and M. P. Selvan, "A comprehensive study on the performance of different multi-class classification algorithms and hyperparameter tuning techniques using optuna," *2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS)*, 2022. DOI: 10.1109/ic3sis54991.2022.9885695.

[17] S. Engdahl, *Blogs*, 2008. [Online]. Available: https://aws.amazon.com/blogs/machine-learning/bring-your-own-hyperparameter-optimization-algorithm-on-amazon-sagemaker/.

[18] *[pdf] domain generalization with mixstyle*. [Online]. Available: https://www.semanticscholar.org/reader/4f6eafafc9563a5b904535078df7e74afe39ef59.

[19] N. H. Barna and H. Heickal, "An automatic abstractive text summarization system," *Dhaka University Journal of Applied Science and Engineering*, vol. 6, no. 2, pp. 39–48, 2022. DOI: 10.3329/dujase.v6i2.59217.

[20] P. Kouris, G. Alexandridis, and A. Stafylopatis, "Abstractive text summarization based on deep learning and semantic content generalization," *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019. DOI: 10.18653/v1/p19-1501.

[21] R. Boorugu and G. Ramesh, "A survey on nlp based text summarization for summarizing product reviews," *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2020. DOI: 10.1109/icirca48905.2020.9183355.

[22] C.-Y. Lin, *Rouge: A package for automatic evaluation of summaries*. [Online]. Available: https://aclanthology.org/W04-1013.

[23] J. Steinberger and K. Ježek, *Evaluation measures for text summarization*. [Online]. Available: http://www.cai.sk/ojs/index.php/cai/article/view/37.