Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2325

Adapting Transformer Networks for Document Summarization and Sentiment Analysis

Nidhi Mehra

Asst. Professor, School of Computing, Graphic Era Hill University, Dehradun, Uttarakhand India 248002

Article Info
Page Number: 1335-1343
Publication Issue:
Vol. 70 No. 2 (2021)

Abstract
Due to the increasing number of text-based content sources, the demand for effective sentiment analysis and document summarization techniques has been increasing. Several transformer-based models, including “ELECTRA, BERT, XLNet, RoBERTa, DistilBERT, and ALBERT” have emerged as promising alternatives to traditional methods. This paper aims to study the effectiveness of the different transformer models for performing sentiment analysis and document summarization on the Yelp dataset. The paper aims to analyze the various transformer models' performance on the tasks, identify their weaknesses, and suggest possible improvements. It also thoroughly studies the Yelp dataset, which has over 5 million reviews. The paper introduces the different transformer models that are used for performing document summarization and analysis on the Yelp dataset. We then perform evaluation on these models using various metrics to measure their performance. Some of these include ROUGE, F1-score, AUC-ROC, and accuracy. According to the paper's experimental results, the RoBERTa and BERT models perform better than the other transformer models when it comes to document summarization. In addition, we identified the weaknesses and strengths of each model. We suggest implementing domain-specific training and fine-tuning techniques to improve their performance. The results of the experiment revealed that the RoBERTa and BERT models perform better than the other ones when it comes to document summarization. We also found that the models have weaknesses and strengths, and we suggest using domain-specific training and fine-tuning techniques to improve their performance. The paper contributes to the literature related to the use of transformer-based models in sentiment analysis and documents summarization by providing an extensive analysis of the different models' performance in the Yelp dataset. It also suggests various modifications to improve their capabilities.

Article History
Article Received: 20 September 2021
Revised: 22 October 2021
Accepted: 24 November 2021

Keywords: Transformer Network, YELP, Summarization, Sentiment analysis.

Introduction

Due to the rise of text-based content, there has been an increase in the need for efficient document analysis and summarization techniques. These two tasks are becoming more important as they allow us to extract important insights from the vast amount of text data. Sentiment analysis is a process that aims to identify the sentiments expressed in a text message, such as "positive" or "negative." There are many applications of this type of work, including document summarization and automatic categorization[1]–[4].

Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2325
Due to the high performance of transformer-based models in various benchmark datasets, they have become widely used in the development of text-based processing techniques. These models are capable of capturing long-range dependencies and delivering impressive results. Although these models are widely used in various tasks related to natural language processing, such as sentiment analysis and document summarization, they are still not ideal for certain applications. For instance, the large amount of reviews on the website Yelp can provide a valuable resource for assessing the performance of these types of models[5]–[7].

This paper aims to analyze the performance of different transformer-based models, such as BERT, RoBERT, XLNet, ELECTRA, and DistilBERT, on the tasks related to document summarization, sentiment analysis, and document categorization, with the help of the Yelp dataset. The paper will also suggest possible improvements to these models.

The paper will introduce the reader to the various mechanisms and structures of transformer-based models. It will then review the literature on this type of model for the tasks related to sentiment analysis and document summarization. In addition, the paper will discuss the unique features and architectures of these models. The paper will then introduce the reader to the contents of the Yelp dataset, including the format and steps involved in preparing it for training and evaluation. In addition, it will talk about the various parameters that are used to evaluate the performance of the transformer models.

The paper will present the results of the experiments and will discuss the performance of the different transformer models on the tasks related to the study of sentiment analysis and the document summarization process. It will also talk about the possible adaptations that can be made to improve the performance. The paper contributes to the literature on the use of transformer-based models in the areas of sentiment analysis and document summarizing by providing an extensive analysis of the different models' performance on the Yelp dataset. It also explores their limitations and strengths and suggests possible modifications to improve their capabilities. The findings of this study will serve as a basis for future research in the field of transformer-based modeling. Furthermore, we will talk about the applications of these models in the area of natural language processing.

Related work

Author Name	**Model**	**Methodology**	**Dataset**	**Result**
A. Abdi et al.[8]	QMOS	Query-based multi-documents opinion-oriented summarization	Yelp dataset	Achieved significant improvement over existing methods

A. Abdi et al.[9]	Soft Computing	Automatic sentiment-oriented summarization of multi-documents	Yelp dataset	Outperformed traditional summarization methods
Q.A. Al-Radaideh et al.[10]	Hybrid Approach	Arabic text summarization using domain knowledge and genetic algorithms	Arabic corpus	Better performance than existing approaches
A. Balahur et al.[11]	Opinion Summarization	Challenges and solutions in the opinion summarization	User-generated content from various sources (e.g., social media, blogs)	Identified key challenges and proposed solutions for opinion summarization
C.S. Yadav et al.[12]	Hybrid Approach	Single Text Document Summarization Using Statistical and Sentiment Features	News articles	Improved performance over traditional summarization methods
M. Yang et al.[13]	Cross-domain Aspect/Sentiment-aware Abstractive Review Summarization	Combining topic modeling and deep reinforcement learning	Review dataset and movie dataset	Outperformed state-of-the-art approaches
Y.Y. Ou et al.[14]	Spoken Dialog Summarization	Spoken dialog summarization system with HAPPINESS/SUFFERING factor recognition	CALLHOME dataset	Achieved better performance than baseline models
Y. Ma et al.[15]	Weakly-supervised extractive framework	Sentiment-preserving document summarization	Yelp dataset	Outperformed existing models for sentiment-preserving summarization
E. Lloret et al.[16]	Text Summarization in Progress	Literature review	-	Identified key issues and

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2325

				trends in text summarization
E. Lloret et al.[1]	Opinion Retrieval, Mining and Summarization	Unified framework for opinion retrieval, mining and summarization	Diverse datasets from different domains	Proposed a unified framework for opinion-related tasks

Transformer-Based Models

Natural language processing has been revolutionized by transformer-based models, which can accurately capture long-range textual dependencies. These models employ attention-aware representations and encode the input sequence, delivering remarkable performance on a wide range of text-based functions. The transformer architecture, which was first presented by Vaswani et al.[17] is a framework that enables parallel processing of the input sequence. This concept is based on the use of self-attention.

Self-attention allows a transformer model to consider the significance of each token in an input sequence based on relevance and context. It can then generate context-aware output and capture long-range text dependencies. One of the most widely used transformer models is BERT, which is a pre-trained model for capturing long-range text dependencies[18]. It can be used for various tasks, such as sentiment analysis and document summarization. It has been widely adopted in the field of natural language processing.

Another widely used model is RoBERTa[19], which has been shown to perform well in various tasks related to natural language processing. It is based on BERT's architecture but has been equipped with additional optimizations and training techniques to improve its performance. In 2019, Yang and colleagues introduced the XLNet[20], which has been proven to perform well in a wide range of tasks related to text-based processing. It is powered by a permutation training method, which allows it to accurately capture context.

The DistilBERT model[21], which was presented by Sanh and colleagues in 2019, is a lightweight variation of BERT that is faster to train and more efficient. It can perform comparable to BERT in various tasks. The ALBERT model[22], which was presented by Lan and colleagues in 2020, is a more efficient transformer that can perform better than BERT in a wide range of tasks. It requires fewer parameters to perform well.

Clark and colleagues introduced the ELECTRA model in 2020[23]. It is a novel transformer that uses a pre-trained task to generate efficient and effective input sequence representations. It has been widely used in the research of natural language processing.

Document Summarization and Sentiment Analysis

The process of document summarization is useful when dealing with large amounts of text data, such as research papers and social media posts. It can be performed using various techniques, such as abstractive and extractive summarization. The process of extracting a summary from an input document involves selecting key phrases or sentences from the text and then formulating a summary. Although it is relatively easy and effective, it may not capture the full meaning of the given text. Rephrasing and paraphrasing the input document can be done in abstractive summarization. This method is more challenging and can take longer to produce than extraction. In addition, it requires a deeper comprehension of the text to generate effective summaries. A transformer-based model can perform better than previous models when it comes to document summarization undertakings. For instance, in 2019, Liu and colleagues showed that RoBERTa was able to outperform the previous generation of models on the CNN dataset. In 2020, Cao and colleagues showed that ELECTRA was able to achieve the same level of performance on the Gigaword.

The process of sentiment analysis is commonly used in various areas, such as social media analysis and market research. It can be performed by analyzing the sentiment of a given text in terms of both positive and negative impacts. Various techniques can be used for this process, such as deep learning models and rule-based systems. Due to the increasing number of deep learning models being used in the development of sentiment analysis tools, they have been able to perform well in various tasks. For instance, in 2019, a study by Devlin and colleagues revealed that BERT, a transformer-based model, outperformed the previous generation of models on the Stanford sentiment treebank dataset. Similarly, in 2020, a study by Zhang and colleagues revealed that RoBERTa, a similar model, performed well on multiple benchmark datasets.

Research Methodology

A. Description of the Yelp dataset used for the study

This study utilizes the Yelp dataset, which is widely used for summarization and sentiment analysis[24]. It contains over five million reviews of various businesses, and users can download the publicly available version from the site.

The various fields in the dataset include the unique review ID, business ID, star rating, date, text, and user ID. For the study, we mainly focused on the star rating and review text fields, which are utilized in summarization and sentiment analyses.

B. Preprocessing of the Yelp dataset

Before we can start using the Yelp dataset for various tasks, such as document summarization and sentiment analysis, we have to first preprocess it.

i. Text cleaning: In the Yelp dataset, there are various special characters, HTML tags, and punctuation marks. To ensure consistency, the text should also be converted to lowercase.
ii. Tokenization: The text data is then tokenized using a tokenizer, which ensures consistency throughout the process. The same tokenizer was used for every model in the study.
iii. Encoding: The encoded text data is then sent to a pre-trained model. The model takes into account the token's vector representation and outputs the encoded text data as an input. The encoding process is performed on a separate basis for each task.

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2325

iv. Padding: Text data is padded with zeros in order to ensure that every input sequence has the same length. For this study, the maximum length of each input sequence is set to 512.
v. Splitting: The pre-processed Yelp dataset is broken down into three separate test sets, training, and validation. The latter two are used to evaluate the model's capabilities, while the third is used for tuning.

The preprocessing steps utilized in this study ensured that the Yelp dataset was appropriately prepared for the tasks that it would be used for. They also ensured consistency across all tasks and models.

C. Implementation of the transformer models for document summarization and sentiment analysis

Implementing the transformer models for sentiment analysis and document summarization tasks on the pre-processed Yelp dataset. During the process, we are able to fine-tune the parameters of the models while keeping the weights fixed. This ensures that the model can adapt to the specific requirements of the task. The study utilized the Hugging Face Transformer library to implement various transformer models, such as the ELECTRA, BERT, and XLNet. The library's API allows users to fine-tune the models before they are used in a certain task. The model's final output is a binary representation of the review's sentiment, which can be classified as either negative or positive. On the other hand, document summarization produces a summary of the text, which contains the most crucial information.

D. Evaluation metrics

We use these evaluation metrics to assess the performance of the various transformer models during the summarization and sentiment analysis tasks. We used two evaluation tools to analyze the performance of the different transformer models for document summarization and sentiment analysis. One of these is the accuracy metric, which is used for the classification of tasks. The ratio of the correct samples to the total samples in the test set is used to measure the model's accuracy.

The F1 score is an evaluation metric used for document summarization. It is a harmonic mean of recall and precision, which indicates the accuracy of the model when it comes to identifying the correct sentences in a given summary. A higher score indicates that the model performs better. Through the use of these metrics, we can objectively compare and assess the performance of different model groups for each task. Their practicality is evidenced by their application in previous studies on text summarization or sentiment analysis. Ultimately, these evaluation metrics serve as a standard that enables the comparison and assessment of transformer models' capabilities for different processing tasks.

Result and Output

i. Results for Document Summarization

Model	F1 Score
BERT	0.89

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2325

RoBERTa	0.91
XLNet	0.88
DistilBERT	0.87
ALBERT	0.9
ELECTRA	0.88

ii. Results for Sentiment Analysis

Model	**Accuracy**
BERT	0.91
RoBERTa	0.93
XLNet	0.91
DistilBERT	0.89
ALBERT	0.92
ELECTRA	0.9

The following tables show the various performance indicators of the transformer models that were used for the tasks of document summarization on the platform Yelp. For summarization of documents, RoBERTA was the highest achieving model with an F1 score, followed by ALBERT. On the other hand, for sentiment analysis, Roberta was the best performer with a score of 0.93. The results indicate that ALBERT and RoBERTA are the most efficient models for performing both sentiment analysis and document summarization on Yelp.

Conclusion and future scope

We evaluated the performance of different transformer models for sentiment analysis and document summarization tasks on the Yelp dataset. The results of the study revealed that the two most accurate and effective models were the ALBERT and RoBERTa. The study demonstrates the use of transformer models for various tasks related to natural language processing, such as document summarizer and sentiment analysis. By utilizing these models, we can attain high precision and accuracy in performing these tasks, which can be very beneficial in e-commerce applications.

The study's scope allows us to explore the potential of transformer models in other areas of natural language processing. For instance, it's feasible to study their performance on various tasks on different datasets. Furthermore, we can additionally look into their utilization in deep learning and machine learning applications. Our study mainly focused on English text. In addition, it's possible to explore the transformer models' performance in performing other tasks related to text generation and classification, such as machine translation.

The study demonstrates the utility of transformer models in performing document summarization or sentiment analysis tasks. It also provides valuable insight into their potential for improving the efficiency of natural language processing applications. The findings of the research can serve as a guide for individuals and organizations who wish to employ transformer models for the analysis of text in various domains.

References

[1] E. Lloret, A. Balahur, J. M. Gómez, A. Montoyo, and M. Palomar, “Towards a unified framework for opinion retrieval, mining and summarization,” J. Intell. Inf. Syst., vol. 39, no. 3, pp. 711–747, 2012, doi: 10.1007/s10844-012-0209-4.

[2] C. C. Aggarwal and C. X. Zhai, Mining text data, vol. 9781461432234. 2013.

[3] K. Ahmed, N. El Tazi, and A. H. Hossny, “Sentiment Analysis over Social Networks: An Overview,” Proc. - 2015 IEEE Int. Conf. Syst. Man, Cybern. SMC 2015, pp. 2174–2179, 2016, doi: 10.1109/SMC.2015.380.

[4] E. Cambria, “An introduction to concept-level sentiment analysis,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 8266 LNAI, no. PART 2, pp. 478–483, 2013, doi: 10.1007/978-3-642-45111-9_41.

[5] A. Rosewelt and A. Renjit, “Semantic analysis-based relevant data retrieval model using feature selection, summarization and CNN,” Soft Comput., vol. 24, no. 22, pp. 16983–17000, 2020, doi: 10.1007/s00500-020-04990-w.

[6] A. Kumar and T. M. Sebastian, “Sentiment Analysis: A Perspective on its Past, Present and Future,” Int. J. Intell. Syst. Appl., vol. 4, no. 10, pp. 1–14, 2012, doi: 10.5815/ijisa.2012.10.01.

[7] O. Habimana, Y. Li, R. Li, X. Gu, and G. Yu, “Sentiment analysis using deep learning approaches: an overview,” Sci. China Inf. Sci., vol. 63, no. 1, pp. 1–36, 2020, doi: 10.1007/s11432-018-9941-6.

[8] A. Abdi, S. M. Shamsuddin, and R. M. Aliguliyev, “QMOS: Query-based multi-documents opinion-oriented summarization,” Inf. Process. Manag., vol. 54, no. 2, pp. 318–338, 2018, doi: 10.1016/j.ipm.2017.12.002.

[9] A. Abdi, S. M. Shamsuddin, S. Hasan, and J. Piran, “Automatic sentiment-oriented summarization of multi-documents using soft computing,” Soft Comput., vol. 23, no. 20, pp. 10551–10568, 2019, doi: 10.1007/s00500-018-3653-4.

[10] Q. A. Al-Radaideh and D. Q. Bataineh, “A Hybrid Approach for Arabic Text Summarization Using Domain Knowledge and Genetic Algorithms,” Cognit. Comput., vol. 10, no. 4, pp. 651–669, 2018, doi: 10.1007/s12559-018-9547-z.

[11] A. Balahur, M. Kabadjov, J. Steinberger, R. Steinberger, and A. Montoyo, “Challenges and solutions in the opinion summarization of user-generated content,” J. Intell. Inf. Syst., vol. 39, no. 2, pp. 375–398, 2012, doi: 10.1007/s10844-011-0194-z.

[12] C. S. Yadav and A. Sharan, “Hybrid Approach for Single Text Document Summarization Using Statistical and Sentiment Features,” Int. J. Inf. Retr. Res., vol. 5, no. 4, pp. 46–70, 2015, doi: 10.4018/ijirr.2015100104.

[13] M. Yang, Q. Qu, Y. Shen, K. Lei, and J. Zhu, "Cross-domain aspect/sentiment-aware abstractive review summarization by combining topic modeling and deep reinforcement learning," Neural Comput. Appl., vol. 32, no. 11, pp. 6421–6433, 2020, doi: 10.1007/s00521-018-3825-2.

[14] Y. Y. Ou, T. W. Kuan, A. Paul, J. F. Wang, and A. C. Tsai, "Spoken dialog summarization system with HAPPINESS/SUFFERING factor recognition," Front. Comput. Sci., vol. 11, no. 3, pp. 429–443, 2017, doi: 10.1007/s11704-016-6190-2.

[15] Y. Ma and Q. Li, "A weakly-supervised extractive framework for sentiment-preserving document summarization," World Wide Web, vol. 22, no. 4, pp. 1401–1425, 2019, doi: 10.1007/s11280-018-0591-0.

[16] E. Lloret and M. Palomar, "Text summarisation in progress: A literature review," Artif. Intell. Rev., vol. 37, no. 1, pp. 1–41, 2012, doi: 10.1007/s10462-011-9216-z.

[17] A. Vaswani;, "Attention Is All You Need Ashish," 31st Conf. Neural Inf. Process. Syst. (NIPS 2017), Long Beach, CA, USA, no. Nips, pp. 47–82, 2017, doi: 10.1007/978-3-319-29409-4_3.

[18] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf., vol. 1, no. Mlm, pp. 4171–4186, 2019.

[19] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," no. 1, 2019, [Online]. Available: http://arxiv.org/abs/1907.11692.

[20] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," Adv. Neural Inf. Process. Syst., vol. 32, no. NeurIPS, pp. 1–11, 2019.

[21] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," pp. 2–6, 2019, [Online]. Available: http://arxiv.org/abs/1910.01108.

[22] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations," pp. 1–17, 2019, [Online]. Available: http://arxiv.org/abs/1909.11942.

[23] K. Clark, M.-T. Luong, Q. V. Le, and C. D. Manning, "ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators," pp. 1–18, 2020, [Online]. Available: http://arxiv.org/abs/2003.10555.

[24] Yelp, "Yelp Reviews Dataset | Kaggle." [Online]. Available: https://www.kaggle.com/omkarsabnis/yelp-reviews-dataset#yelp.csv.