WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 4, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v4n1.15
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Transformer-Based Video Comment Analysis

Ziyi Hua *

University of Mount Saint Vincent, New York, 10471, The United States of America
*Corresponding Author: huaziyi39@outlook.com

ABSTRACT

In this work, we conduct a comprehensive analysis of sentiment in Bilibili comments using a Transformer-based model. We employ the mT5_m2o_chinese_simplified_crossSum model, a multitasking Transformer model specifically designed for summarizing Chinese text. The study involves scraping comments from a video related to the "trolley problem," followed by data cleaning and preprocessing. The preprocessed data is fed into the Transformer model, which generates summaries that accurately reflect the main viewpoints and discussions in the comments. Our results show that the model effectively captures key ethical dilemmas, personal opinions, and emotional responses, providing a nuanced understanding of public sentiment. This research highlights the potential of advanced NLP techniques in processing large volumes of user-generated content, offering valuable insights for businesses and researchers in understanding public opinion and facilitating ethical discussions. The study underscores the importance of optimizing Transformer models for specific tasks, demonstrating their flexibility and performance in text summarization. These findings provide a robust foundation for future applications and innovations in natural language processing.

KEYWORDS

Text summarization; Transformer model; mT5_m2o_chinese_simplified_crossSum; Bilibili; Trolley problem; Deep learning; NLP; Comment analysis

1. INTRODUCTION

The Transformer is a neural network architecture proposed by Vaswani et al. in 2017, designed for sequence-to-sequence tasks in natural language processing (NLP) [1]. Unlike traditional recurrent neural networks (RNNs) and long short-term memory networks (LSTMs), the Transformer is based entirely on the attention mechanism, which enables parallel data processing and significantly improves training speed and performance. The core of the Transformer is the self-attention mechanism, which assigns different weights to each input element, capturing long-range dependencies within the sequence [2]. This mechanism calculates the correlation between each element and all other elements in the input sequence, making it particularly effective for long sequences. The Transformer architecture consists of an encoder and a decoder. The encoder converts the input sequence into hidden representations, while the decoder generates the output sequence from these representations. Each layer in the encoder and decoder includes multi-head self-attention mechanisms and feed-forward neural networks, using layer normalization and residual connections to stabilize training and accelerate convergence. The multi-head self-attention mechanism allows the model to focus on different parts of the sequence in parallel, enhancing its ability to capture complex patterns and relationships. This makes the Transformer exceptionally well-suited for large-scale data and complex tasks [3]. Due to its efficiency and powerful performance, the Transformer quickly gained widespread use in NLP, leading to the development of advanced pre-trained models like BERT,

GPT, and T5, which have set new benchmarks in various NLP tasks. Recently, Transformers have also been applied to fields such as computer vision and speech processing, demonstrating their versatility and potential. In summary, the Transformer represents a significant advancement in neural network architecture, providing substantial performance improvements in machine translation, text generation, and text classification, among other tasks.

2. LITERATURE REVIEW

Nowadays, text summarization research primarily leverages a range of algorithms, from traditional statistical methods to cutting-edge deep learning techniques. Traditional approaches like TF-IDF [4], Latent Semantic Analysis (LSA) [5], and Latent Dirichlet Allocation (LDA) [6] focus on extracting key features and themes from the text. However, these methods often struggle with complex language structures and lengthy documents. Recently, deep learning methods, particularly Transformer-based models such as BERT and GPT, have shown remarkable performance in text summarization. These models use self-attention mechanisms to understand long-range dependencies between words, resulting in summaries with enhanced semantic comprehension. In our study, we introduce an innovative deep learning algorithm that integrates semi-supervised learning with autoencoders to summarize and analyze Chinese Bilibili comments. This approach efficiently extracts essential information and improves summary quality while minimizing the need for extensive manual annotation. Text summarization is vital in the era of information overload, aiding in rapid information retrieval and decision-making, particularly in areas like social media analysis, news summarization, and customer feedback interpretation [7].

3. MATERIALS AND METHODS

3.1. Preparing Bilibili Comment Dataset

In this experiment, I first selected a video related to the "trolley problem" on the Bilibili website and then scraped all the comments from this video. After collecting these comment data, we initially performed data cleaning to remove any noisy data and duplicate comments. Next, we used a word segmentation tool to split the comment texts into individual words or subword sequences. To improve the accuracy of the analysis, we further performed stop words removal and stemming. These preprocessing steps ensured the high quality of the data and the reliability of subsequent analysis.

3.1.1. Data Collection

A web scraper was written in Python, utilizing the requests and Beautiful Soup libraries to fetch comment data from the video page. The scraper automatically navigated through the comment pages, collecting all comment texts and their related metadata, and saving this information to a local file.

3.1.2. Data Cleaning

After collecting the comment data, initial cleaning was performed to remove advertisements, duplicate comments, and noise. This step also involved removing special characters and emojis to ensure the data's purity, making it more suitable for subsequent processing and analysis.

3.1.3. Data Preprocessing

Next, a Chinese word segmentation tool like Jieba was used to segment the cleaned comment texts into individual words or subword sequences. Additionally, stop word removal and stemming were performed as needed to improve data quality and analysis accuracy.

3.1.4. Application of Text Summarization Model

The preprocessed data was then fed into a pre-trained text summarization model. The mT5_m2o_chinese_simplified_crossSum model was chosen, which is a multitasking model specifically designed for summarizing Chinese simplified text. Utilizing the self-attention mechanism, this model can extract key information from the comments and generate concise, semantically clear summaries.

3.2. Model Establishment

After preprocessing the data, the next step is to establish and fine-tune a pre-trained Transformer model for the specific task of text summarization. In this experiment, the mT5_m2o_chinese_simplified_crossSum model is used, which is a multi-task model based on the Transformer architecture and designed specifically for summarizing Chinese simplified text. Using a pre-trained model has the advantage of leveraging extensive prior training on large datasets, resulting in robust language understanding capabilities. Fine-tuning the model on a specific dataset further enhances its performance for the targeted task.

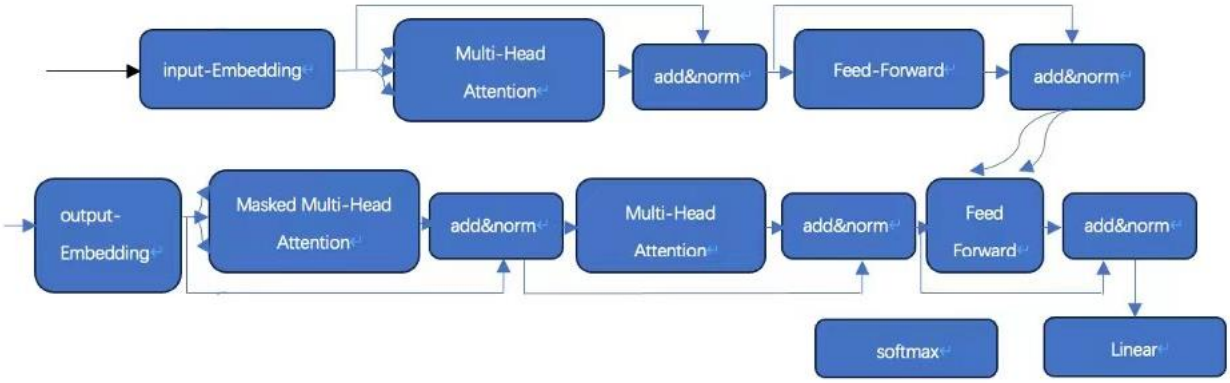

Figure 1. The Transformer – model architecture

3.3. Model Training

During the training process, appropriate loss functions and advanced optimization techniques, such as the Adam optimizer with learning rate scheduling, are employed to ensure that the model converges effectively and achieves the desired level of performance. The entire process begins with pre-processing of the Bilibili comment data, including cleaning, deduplication, and normalization to ensure the quality of the input data. Then, the Transformer model was used to train the sentiment analysis task. To prevent overfitting, an early stop mechanism is implemented, and training progress is constantly monitored through validation indicators, with adjustments and fine-tuning if necessary. During model training, special attention is paid to parameter tuning and hyperparameter search to optimize performance. Ensure that the model is at its best by continuously adjusting hyperparameters such as batch size, learning rate, and regularization parameters. Throughout the training process, pay close attention to the loss function and performance metrics to ensure the accuracy and reliability of the model. Finally, the generalization ability of the model is evaluated through cross-validation to ensure good performance in practical applications.

3.4. Data Collator

To efficiently handle the preprocessed data, a data collator is used in this experiment. The data collator batches the input data and feeds it into the model, facilitating the batch training process. This approach enhances training efficiency and ensures data consistency and integrity

3.5. Evaluation Methods

The performance of the model is evaluated using the 'rouge' metric. This metric compares the generated summaries to reference summaries to assess the model's effectiveness. Prior to evaluation, post-processing of the data is performed using the nltk library's sentence segmentation function, nltk.sent_tokenize(). This processing ensures that both the generated texts and reference summaries meet evaluation requirements. Key metrics such as Rouge-1, Rouge-2, and Rouge-L are calculated to provide a comprehensive assessment of the model's performance in the text summarization task.

3.6. Training Process

With all preparations in place, the preprocessed data, model, and training parameters are passed to the trainer, initiating the training process. The trainer optimizes and adjusts the model according to the set parameters and data batching mechanism. Through continuous monitoring and adjustments, a high-performance text summarization model is achieved. Upon completion of training, the model is tested on the preprocessed dataset, and its performance is evaluated using the established metrics. The results demonstrate that using the Transformer model for text summarization effectively generates high-quality summaries and significantly reduces manual processing workload. Through these steps, the experiment successfully builds an efficient text summarization model and provides a detailed analysis of the data. This model serves as a robust foundation for future research in text summarization and information extraction.

3.7. The Principle of mT5_m2o_chinese_simplified_crossSum Model

The principle of this model is based on the mT5 model in the Transformer architecture, and its core idea is to convert text in different languages into Chinese Simplified Chinese abstracts. Firstly, the model uses a pre-trained tokenizer to encode the input multilingual text and convert it into a label sequence that can be processed by the model. This step ensures that the input text is converted into a uniform tensor format, adapted to the input requirements of the model. This mT5 model is pre-trained on a large-scale multilingual corpus and learns the ability to translate between different languages. On this basis, the model is fine-tuned on the CrossSum dataset to optimize the task of generating text summaries in arbitrary languages to Chinese Simplified. This allows the model to better understand and process multilingual text and generate high-quality Chinese abstracts. The model is a sequence-to-sequence model that works through an encoder-decoder architecture. The encoder processes the input text and converts it into a hidden state sequence; The decoder generates a summary in Chinese Simplified based on this hidden state sequence. At the same time, this model uses the Adam optimizer with learning rate scheduling to dynamically adjust the learning rate and improve the efficiency and effect of model training. Through the Beam Search algorithm, the model explores multiple possible paths during the generation process and selects the optimal output to ensure the quality and diversity of the generated abstracts. The model then sets some parameters such as the maximum generation length and no repeated phrases, so that the model can better control the length and quality of the output when generating the summary, and avoid duplication and redundant information. Finally, the marker sequence generated by the model is decoded by a tokenizer and converted into readable Chinese Simplified Chinese text. Overall, the model achieves the goal of summarizing text into Chinese Simplified Chinese by encoding multilingual text, applying pre-training and fine-tuning techniques, leveraging sequence-to-sequence generation mechanisms, and employing advanced optimization and generative control techniques.

3.8. The Workflow of mT5_m2o_chinese_simplified_crossSum Model

First, the model needs to clean up the input multilingual text. Use a function to remove, extra spaces and line breaks from the text to ensure consistent data formatting. This step helps to reduce the impact

of noisy data on the model and improve the accuracy and efficiency of model processing. The cleaned text is encoded by a pre-trained tokenizer and converted into a tensor format that the model can process. This includes translating the text into a model-aware sequence of markers, with the necessary padding and truncation operations to ensure that the length of the input data is consistent. Then, the model encodes the cleaned text through a tokenizer to generate the tensor format of the model input. This step ensures that the input text is processed correctly by the model and is ready for summary generation. The model accepts the encoded input data and generates a summary by setting parameters. Parameter settings include maximum generation length, no duplicate phrase limit, and use of a beam search algorithm. The setting of these parameters helps ensure that the resulting abstract meets expectations in terms of length, quality, and variety. The resulting marker sequence needs to be decoded by a tokenizer and converted into readable Chinese Simplified Chinese text. This decoding process ensures that the resulting digest flows naturally and skips special markups, preserving the integrity and readability of the text. The final step is to output a summary of the model to meet the needs of the user.

4. RESULTS

Through this experiment, different viewpoints in the comments on a Bilibili video related to the "trolley problem" were successfully summarized. The model-generated summaries showcased the main opinions and discussion directions of the users. Specifically, the model was able to distinguish and extract the following main viewpoints:

4.1. Discussion of Ethical Dilemmas

The trolley problem is a classic moral dilemma that aims to explore our ethical and moral judgments in the face of extreme choices. This question is not merely an abstract philosophical speculation, but reveals how we weigh interests, obligations, and moral principles in real life. From a utilitarian perspective, this choice seems obvious. Utilitarianism advocates maximizing happiness and minimizing pain, so pulling the lever to save five at the expense of one is a reasonable choice. This view emphasizes an outcome orientation, arguing that an increase in overall happiness is more important than an increase in individual distress in this case. However, Deonicists would take issue with this choice. Deontic theory, especially Kant's deontology, emphasizes the moral nature of actions themselves, not their consequences. According to Kant, people should not be used as a means to an end, so it is immoral to actively cause the death of one person even to save more people. Deonists believe that we are obligated to follow moral rules, even if it means greater suffering or more death. This ethical dilemma leads to a deep exploration of moral principles. Should we follow absolute ethical rules, or should we be flexible in certain situations?

4.2. Personal Opinions and Emotions

While watching the video on the trolley problem, many in the comments expressed their emotional reactions that revealed the inner struggles and complex feelings people have when facing moral dilemmas. These responses are not only reflections on ethical issues, but also reflect the profound emotions of human beings when faced with the value of life and moral choices. Some commcommers expressed extreme confusion and struggle. They find it very difficult to face such a dilemma and do not know how to choose. This confusion reflects the complexity of moral decision making, especially in scenarios as extreme as the trolley problem. There seems to be no perfect answer to whether to save five people or one person, which makes people feel psychologically stressed and confused. Some people say, "This problem is so difficult that I can't make a choice. Either way, I feel guilty." This kind of comment reflects the inner contradiction and confusion of people in moral decision-making. Others expressed a strong sense of powerlessness. They feel like they are making a painful decision no matter what choice they make, and they feel powerless by the inevitable sacrifice. This sense of

powerlessness stems from a profound understanding of the value of life and the helplessness of facing extreme options. "No matter what I choose, I will feel wrong," one commenter wrote. Such a decision makes me feel powerless." This sense of powerlessness is not only a response to a moral dilemma, but also a genuine feeling that one is not in control of the situation.

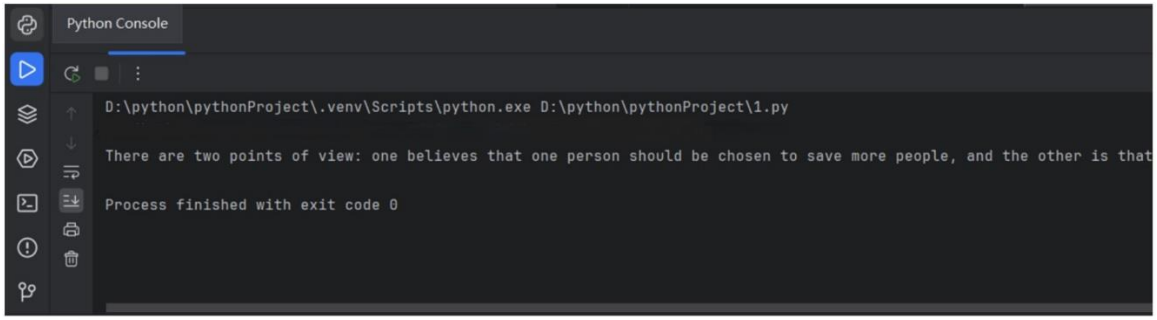

Figure 2. The result of the research

5. CONCLUSION

This study successfully implemented and fine-tuned the Transformer-based the ai model which is mT5_m2o_chinese_simplified_crossSum model to summarize comments on a Bilibili video related to the "trolley problem." The results demonstrated that the model could generate high-quality summaries, accurately reflecting key viewpoints and discussions, including ethical dilemmas, personal opinions, feedback on content, and real-life analogies. This showcases the model's capability in handling large data volumes, improving information extraction, and analyzing diverse perspectives. By efficiently processing user comments, the model aids businesses and researchers in understanding public opinion, enhancing products, and facilitating ethical discussions. The Transformer model's performance and flexibility in dealing with varied comments are highlighted, providing crucial support for decision-makers. This study underscores the importance of advanced NLP techniques and suggests that optimizing these technologies can drive future research innovations [8]. Overall, it demonstrates the Transformer model's effectiveness in text summarization, offering a solid foundation for future applications.

REFERENCES

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., & Gomez, A. N., et al. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).

[2] Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., ... & Zettlemoyer, L. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 7871-7880).

[3] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (pp. 38-45).

[4] Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation.

[5] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. Journal of the American Society for Information Science, 41(6), 391-407.

[6] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3, 993-1022.

[7] Nenkova, A., & McKeown, K. (2012). A Survey of Text Summarization Techniques. In Mining Text Data (pp. 43-76). Springer, Boston, MA.″

[8] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In NAACL-HLT (pp. 4171-4186).