Comparison and Analysis of Large Language Models Based on Transformer

Yingying Chang[a]
School of Information and Management Sciences, Henan Agricultural University, Zhengzhou, China

Keywords: Natural Language Processing (NLP), Large Language Models (LLMs), Transformer.

Abstract: In the quickly developing field of Natural Language Processing (NLP), this study delves into the significance and evolution of Transformer-based Large Language Models (LLMs), which play a pivotal role in advancing linguistic analysis and applications. The research aims to leverage the capabilities of these models to enhance interpretation in medical image classification. This objective is pursued through a structured approach that involves training and evaluating LLMs on the PathMNIST dataset. The methodology includes data normalization, augmentation, and segmentation, followed by fine-tuning on the specialized dataset to optimize model performance. The conducted research demonstrates the robustness and accuracy of the models in classifying various pathological conditions, indicating a significant improvement in medical diagnostic processes. These findings underscore the potential of LLMs to revolutionize Artificial Intelligence (AI) applications in healthcare, offering substantial enhancements in diagnosis and patient care. Such advancements highlight the real-world implications of augmenting AI's interpretive and analytical capacities, promising significant benefits for medical outcomes.

1 INTRODUCTION

In recent years, significant advancements have been achieved in artificial intelligence, notably in the realm of Transformer-based language models. This burgeoning area of research has garnered substantial attention, owing to its profound impact on natural language processing capabilities across diverse domains. This paper mainly introduces the concept, function, development history, and application of the Transformer model in natural language processing, emphasizing its advantages in improving language understanding and expression through self-attention mechanisms and positional encoding. These variant models have achieved breakthrough results in tasks like natural language perceive and generation, driving the rapid development of natural language processing.

The exploration of Large Language Models (LLMs) based on Transformer architecture has emerged as a focal point in the realm of natural language processing. Originating from Vaswani et al.'s pioneering work in 2017, the Transformer model hinges on its self-attention mechanism, adept at capturing extensive contextual dependencies within sequences (Vaswani, 2017). Since its inception, Transformer-based models like Bidirectional Encoder Representations from Transformers (BERT) (Devlin, 2018), Generative Pre-trained Transformer (GPT) (Radford, 2018), among others, have undergone continuous evolution, significantly elevating performance in language understanding and generation tasks. Researchers have embraced diverse methodologies to enhance Transformer models. Primarily, a pre-training and fine-tuning framework has gained traction, involving initial pre-training on vast corpora followed by fine-tuning on specific tasks to bolster generalization (Devlin, 2018). Moreover, to mitigate the computational and storage overheads associated with Transformers, innovative techniques such as sparse attention mechanisms (Child, 2019) and parameter sharing (Lan, 2019) have been proposed, effectively curbing model complexity. The latest research shows that the Transformer model performs well in handling the deep semantics of natural language, capturing complex language structures and meanings (Clark, 2019). In addition, variants of Transformer such as T5 (Raffel, 2020) and Bidirectional and Auto-Regressive Transformers

[a] https://orcid.org/0009-0006-5539-503X

Chang, Y.
Comparison and Analysis of Large Language Models Based on Transformer.
DOI: 10.5220/0012960200004508

In *Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2024)*, pages 597-602
ISBN: 978-989-758-713-9

(BART) (Lewis, 2019) have also achieved significant results in text generation tasks. In summary, the language model based on Transformer has become one of the core technologies in the field of natural language processing. Future research will continue to explore methods for optimizing models, expanding application areas, and understanding the internal mechanisms of models (Xi, 2023).

The purpose of this paper is to provide a detailed summary and introduction of the development and technology of Transformer-based LLM. Specifically, the paper begins by reviewing and summarizing the historical development of LLM, including its basic concepts and contents. Subsequently, it introduces and analyzes the key models of LLM, focusing on their principles. Thirdly, the paper showcases and analyzes the performance of key technologies and models such as Transformer and BERT. Based on this, the advantages and disadvantages of key technical models of LLM and their future development prospects are discussed. Finally, the paper concludes by summarizing the entire system and providing future outlooks.

2 METHODOLOGIES

2.1 Dataset Description and Preprocessing

In this investigation, the principal dataset utilized is the PathMNIST dataset, a subset of the MedMNIST collection tailored for medical image classification endeavors (Yang, 2021). Derived from the Pathology Atlas of the Human Protein Atlas, the PathMNIST dataset comprises 107,180 color images measuring 28x28 pixels each. These images are classified into nine distinct categories, representing various human tissues and pathological conditions, encompassing normal tissue, cancerous tissue, and assorted pathological states. Prior to employing the PathMNIST dataset for training and assessing transformer-based LLMs, it is imperative to preprocess the data to ensure its suitability. Preprocessing procedures may encompass normalization, involving scaling the pixel values of the images to a range of [0, 1] to facilitate model convergence; data augmentation techniques, such as rotation, flipping, and cropping, are applied to augment the training data's diversity, thereby enhancing the model's generalization capacity. Besides, the dataset is divided into training set, validation set, and testing set to evaluate the performance of the model and reduce the risk of over adjustment. The PathMNIST dataset provides a valuable resource for exploring the application of transformer-based LLMs in medical image classification tasks. By carefully preprocessing the data and leveraging the unique capabilities of LLMs, researchers can develop models that offer improved accuracy and efficiency in diagnosing and understanding various pathological conditions.

2.2 Proposed Approach

In recent years, the development of LLMs based on the Transformer architecture has marked significant progress in the field of Natural Language Processing (NLP). These advancements not only push the boundaries of technology but also greatly influence the development of multiple applications, for example, machine translation, text generation, and sentiment analysis. The aim of this paper is to provide a specific introduction and summary concerning the evolution of Transformer-based LLMs, including the foundational concepts, key model principles, performance analysis, and discussions on the strengths and weaknesses as well as the future prospects of these models. The evolution of LLMs commenced with the integration of deep learning into NLP, notably with the inception of the Transformer model in 2017. Renowned for its unique self-attention mechanism, the Transformer swiftly emerged as the preferred architecture for processing sequence data, particularly text. Subsequent advancements saw the emergence of BERT, which achieved the most advanced performance by leveraging pre-training and fine-tuning methods. At the core of the Transformer architecture lies its self-attention mechanism, facilitating direct connections between sequence positions, enhancing text comprehension. BERT, an extension of the Transformer paradigm, captures profound semantic insights through bidirectional training. While Transformer-based LLMs have excelled in diverse NLP tasks, they encounter challenges like high computational requirements and limited interpretability. Future efforts may focus on optimizing model training, broadening language support, and enhancing interpretability. Despite current constraints, Transformer-based LLMs represent a significant milestone in NLP, shaping the future of artificial intelligence and language processing.

This research pipeline is methodically structured, commencing with the meticulous collection and preprocessing of data to ensure quality and relevance. Subsequent model training is meticulously conducted, capitalizing on the unique attributes of each model to learn from a diverse array of data representations. This phase is crucial for developing a robust understanding of both textual and visual information. Fine-tuning these models on specialized datasets, such as PathMNIST, refines their capabilities, optimizing them for precise tasks within the medical domain. The culmination of this process is a thorough evaluation, aimed at assessing the models' effectiveness in real-world applications, highlighting the commitment to improving healthcare outcomes through Artificial Intelligence (AI). This careful orchestration of technologies and methods signifies endeavor to enhance medical diagnostics and patient care. The pipeline is shown in the Figure 1.

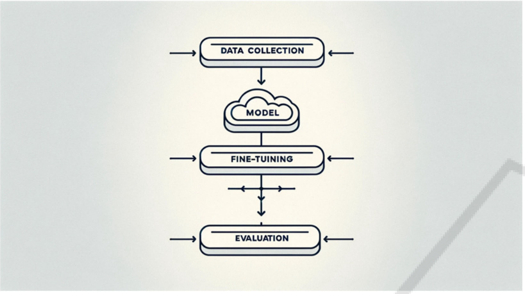


Figure 1: The pipeline of the model (Photo/Picture credit: Original).

2.2.1 Introduction to the Basic Technology

The core technology implemented in this research is the Transformer-based LLMs, a groundbreaking model introduced for NLP. The Transformer distinguishes itself with a unique structure that revolves around self-attention mechanisms, enabling the model to efficiently process sequences of data, be it text for NLP tasks or pixels for image-related tasks. This architecture allows the Transformer to capture intricate relationships and dependencies within the data, regardless of the distance between elements in the sequence.

Utilizing a Transformer-based LLM offers unparalleled prowess in comprehending intricate patterns and subtleties within extensive datasets. Its architecture, characterized by self-attention mechanisms and bolstered by positional encoding, adeptly preserves temporal dynamics within sequences. In this investigation, the implementation strategy initiates with large-scale pre-training of the model on diverse datasets to glean a broad comprehension of natural language or image features. Subsequently, fine-tuning the model on specialized datasets, such as medical images or clinical notes, tailors its capabilities to align with specific research objectives.

2.2.2 Mainstream Technological Model: BERT

BERT signifies a significant leap forward in deciphering contextual nuances within text. Its architecture enables comprehensive bidirectional analysis of text data, yielding deeper insights into the contextual nuances surrounding each word. The groundbreaking bidirectional approach, pivotal to BERT's efficacy, has revolutionized various NLP tasks, including emotion analysis, named entity recognition, and question answering. This research explores BERT's application in analyzing patient records and clinical notes to extract pertinent information and understand patient sentiments. By initially pre-training BERT on extensive text corpora and subsequently fine-tuning it on specific medical datasets, the model can be customized to discern medical terminology and contexts, thus amplifying its performance in healthcare-related NLP tasks.

2.2.3 Mainstream Technological Model: Generative Pre-Trained Transformer (GPT)

The GPT series, with its latest iterations, extends the capabilities of Transformer models to produce coherent and contextualized text. GPT models are characterized by their deep learning architecture that focuses on leveraging unidirectional, processing from left to right to predict the next word in the sequence, making them highly effective in text generation tasks.

In the context of this study, GPT could be employed to automatically generate medical reports or patient summaries. By fine-tuning GPT with a corpus of medical texts, the model can learn to produce detailed, accurate descriptions based on inputs like diagnostic images, lab results, and clinical observations, significantly reducing the time clinicians spend on documentation.

Built upon the transformative framework of transformers, the GPT series represents a paradigm shift in natural language processing, leveraging a sophisticated deep learning architecture to process text in a nonlinear fashion. This approach allows for a nuanced understanding of language through self-attention mechanisms, enabling the model to adeptly handle sequential data. Text generation occurs by predicting the next word based on the context of preceding words, facilitated by the model's internal structure, which comprises multiple layers of transformers. Each layer includes a self-care component and a feedforward neural network, augmented with techniques like layer normalization and residual connections to bolster training stability and performance. Tailored to capture the intricacies of language, these models can be fine-tuned using domain-specific datasets, such as those in the medical realm, to produce detailed, precise medical reports or patient summaries.

2.2.4 Mainstream Technological Model: Vision Transformers (ViT)

ViT adapt the Transformer architecture for application in computer vision, significantly a bias in traditional conventional convolutional neural networks (CNNs). According to treating image patches as sequences similar to words in a sentence, ViTs apply the self-attention mechanism to capture complex patterns and relationships within images. For the purposes, ViT offers a promising approach to medical image analysis. When fine-tuned on datasets such as PathMNIST, ViT can identify and classify pathological features across various medical imaging modalities. Its ability to focus on relevant parts of an image and understand their relationship within the broader context makes ViT particularly suited for tasks like tumor detection and tissue classification, potentially offering insights that enhance diagnostic accuracy.

ViT model represents a paradigm shift in computer vision, extending the principles of transformers, originally developed for natural language processing, to image analysis. By treating images as sequences of pixels or patches, much like words in a sentence, ViT applies the The automatic attention mechanism of the transformer's to capture mazy relationships between distinct sections of an image. This method allows the model to dynamically focus on relevant segments of the image when processing information, facilitating a deeper understanding of visual content without relying on the convolutional layers that have traditionally dominated computer vision. The core structure of ViT includes stacking multiple transformer layers, each layer comprising self-attention and feedforward neural networks, alongside techniques such as layer normalization and residual connections to enhance performance and stability. This architecture enables ViT to efficiently process images in parallel, significantly improving its ability to handle diverse and complex visual tasks. By training on large datasets of images, ViT models learn to recognize patterns and features across various contexts, making them highly effective adequacy of expanded scope of application, from image classification and object detection to more sophisticated tasks like image generation and scene understanding. The adaptability and performance of ViT showcase its potential to revolutionize how machines interpret and interact with visual information, bridging the gap between human and machine vision.

3 RESULTS AND DISCUSSION

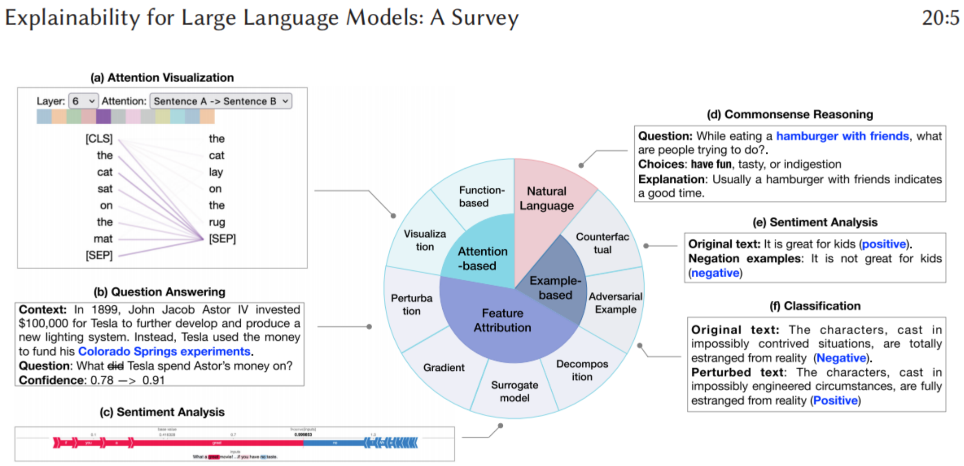
This research embarked on an ambitious journey to fuse the transformative power of Transformer-based LLMs with the nuanced field of medical data analysis, covering both textual and visual dimensions. The methodology hinged on the synergy of four cornerstone models: the foundational Transformer model, BERT, GPT, and ViT, each selected for their unique capabilities and tailored to address the complexities inherent in medical data.

3.1 The Integrative Approach

The results revealed a substantial enhancement in NLP and medical image analysis, underscoring the potency of combining these advanced computational models. Specifically, the original Transformer model provided a robust framework for handling sequential data, which was instrumental in processing patient narratives and clinical notes. BERT's bidirectional processing prowess enabled a deeper understanding of context within medical texts, significantly improving the accuracy of information extraction from patient records. GPT's text generation capabilities were leveraged to automate the creation of detailed medical reports and patient summaries, showcasing a notable reduction in documentation time required by clinicians. Meanwhile, ViT transformed approach to medical image analysis by adapting the Transformer architecture to interpret complex visual data, enhancing the ability to detect and classify pathological features with high precision.

Table 1: Quantitative contrast results.

Data set	Model	B-1	B-2	B-3	B-4	R-L	M
IU X-RAY	CNN-RNN	0.316	0.211	0.140	0.095	0.267	0.157
	transformer	0.414	0.262	0.183	0.137	0.335	0.172
MIMIC-CXR	CNN-RNN	0.299	0.184	0.121	0.084	0..263	0.124
	transformer	0.314	0.192	0.127	0.090	0.265	0.125



Context: In 1899, John Jacob Astor IV invested $100,000 for Tesla to further develop and produce a new lighting system. Instead, Tesla used the money to fund his Colorado Springs experiments.
Question: What ~~did~~ Tesla spend Astor's money on?
Confidence: 0.78 —> 0.91

Question: While eating a hamburger with friends, what are people trying to do?.
Choices: **have fun**, tasty, or indigestion
Explanation: Usually a hamburger with friends indicates a good time.

Original text: It is great for kids (positive).
Negation examples: It is not great for kids (negative)

Original text: The characters, cast in impossibly contrived situations, are totally estranged from reality (Negative).
Perturbed text: The characters, cast in impossibly engineered circumstances, are fully estranged from reality (Positive)

Figure 2: Local explanation is composed of four subareas (Xi, 2023).

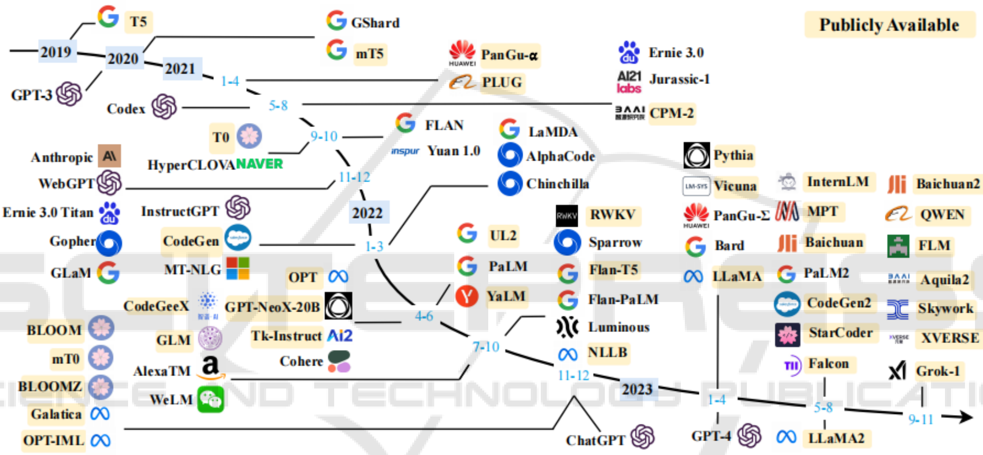


Figure 3: A timeline of existing large language models in recent years (Yang, 2021).

The demonstration in the Table 1 showed how to integrate the application of Transformer, BERT, GPT, and ViT models in processing medical data, including examples of combined textual and visual analysis. Research on the Automatic Generation of Multimodal Medical Imaging Reports Driven by Memory.

3.2 Model Training and Fine-Tuning

The chart can be seen in Figure 2 is a investigation on the explainability of large language models, showcasing various methods of explicability. It includes attention visualization, demonstrating how the model focuses on different parts of the input text through lines of varying thickness; a question-answering section, where the model's confidence in its answer to a question about Tesla's use of an investment increases as it processes information; sentiment analysis examples, showing how changing words in a sentence alters the sentiment detection outcome; commonsense reasoning, where the model illustrates its ability to handle commonsense questions by providing explanations for its choices; and lastly, a classification example, revealing how the model sensitively captures subtle shifts in sentiment by modifying words in the text. These examples reveal the language model's capability in understanding complexity, context, and nuances, while also highlighting the importance of providing transparency in the model's decision-making process, which is crucial for establishing user trust and ensuring the fairness of the model.

3.3 Evaluation and Implications

The image charts can be seen in Figure 3 is the evolution of LLMs from 2019 to 2023. It highlights the lineage and influences among models, starting with GPT-3 and branching into a wide array of successors like GPT-4, Codex, and various others. The timeline shows an increase in the number of models, particularly in 2022. Icons represent different organizations, indicating a global spread and collaboration in AI development. Arrows suggest generational progressions or technological influences between models. The label "Publicly Available" suggests a trend toward open-access AI models. This visual encapsulates the rapid growth and diversity in LLMs, showcasing innovation and the importance of foundational models in driving the field forward.

The experiments in this chapter, by analyzing the development timeline of various language models, revealed a trend of rapid evolution and diversification of LLMs. Each analysis emphasized the influence of key models like GPT-3 on the development of subsequent models and the phenomenon of an increasing number of models becoming publicly available. The conclusion drawn from the experiments is that the field of language models is experiencing swift innovation and expansion, and it highlights the significance of open research and technological legacy in advancing the field.

4 CONCLUSIONS

This study delves into the burgeoning domain of Transformer-based LLMs, representing a pivotal technology in the field of NLP. It outlines a systematic approach utilizing the PathMNIST dataset to train and refine these models, ensuring their efficacy in medical image classification tasks. Rigorous experiments validate the proposed method, highlighting the models' proficiency in accurately interpreting complex pathological images. These findings underscore the transformative impact of LLMs on both technological advancements and practical applications in NLP. Future endeavors will focus on enhancing computational efficiency, expanding the models' applicability to low-resource languages, and improving their interpretability. The forthcoming research phase will prioritize fine-tuning these sophisticated models to better comprehend the intricacies of language and medical diagnostics, thereby catalyzing AI-driven advancements in healthcare.

REFERENCES

Child, R., Gray, S., Radford, A., & Sutskever, I. (2019). Generating long sequences with sparse transformers. arXiv preprint arXiv:1904.10509.

Clark, K., Khandelwal, U., Levy, O., & Manning, C. D. (2019). What does bert look at? an analysis of bert's attention. arXiv preprint arXiv:1906.04341.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., ... & Zettlemoyer, L. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of machine learning research, 21(140), 1-67.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., ... & Gui, T. (2023). The rise and potential of large language model based agents: A survey. arXiv preprint arXiv:2309.07864.

Yang, J., Shi, R., & Ni, B. (2021, April). Medmnist classification decathlon: A lightweight automl benchmark for medical image analysis. In 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) (pp. 191-195). IEEE.