DEEP LEARNING APPLICATIONS IN TEXT USING TRANSFORMER ARCHITECTURES, LLM, AND NEURAL NETWORKS

S. Nandhini
Assistant Professor, Department of Computer Science
Stella Maris College (Autonomous), Chennai, India
nandu4v11@gmail.com
https://doi.org/10.34293/9789361639715.shanlax.ch.029

Abstract

Recent advancement in the domain of deep learning have changed the perception of pattern recognition and automation which has paved way for machines to process and understand the complex data with unprecedented performance in image, text and speech processing tasks. Methodologies in Neural networks like Convolutional Neural networks (CNN), Recurrent Neural Networks(RNN) and Transformer based models have been proven to perform extremely well in extracting hierarchical features, modelling long term dependencies and enabling the understanding of context across modalities. CNN architecture is very much aligned with image-based feature extraction and prediction while RNN architecture is employed in speech and time-series data for capturing temporal sequence modelling. Transformer based architecture is used for parallel processing of text with context aware text processing. The emergence of Large Language Models(LLM) diagnosis has been further extended in medical domain and real time surveillance for image recognition with text processing tasks such as summarization, sentiment analysis, semantic plausibility and machine translation. Speech recognition leverages voice assistants, transcription services and language learning platform. The contribution of these architecture redefines the intelligent system by bridging the gap between the human understanding and machine-driven automation. The contribution of these intelligent systems in medical diagnosis, natural language understanding and human computer interaction in achieving generalizable, adaptive and context aware solutions is remarkable and highly reliable for complex data processing.

Keywords: *Artificial Intelligence(AI), Neural Networks, Deep Learning, Transformer architecture, Natural Language Processing(NLP), Large Language Models(LLM)*

Introduction

Deep learning, a subset of machine learning has given a new dimension for pattern recognition, which allowed the machine to recognise pattern and structures in complex data without the need of explicit programming. Deep learning models learn hierarchical representation of data, in contract with standard algorithm which fail to do. Deep learning is involved in variety of domain including speech recognition, image classification, natural language processing and biometric identification because of its ability to recognise complicated pattern of data[1][11].

The detection of complex patterns is enhanced by deep architectures with numerous hidden layers which extracts the level of abstraction. The performance of neural network in solving real world issues with at most accuracy is been further expanded by artificial neural network. Artificial Neural Network, made up of several layers with interconnected neurons are the foundation of deep learning. Artificial neural network is inspired by the structure and function of the human neural system present in the human brain. ANN consist of input layer, one or more hidden layer and an output layer, which are made up of interconnected neurons arranged in layers. Input data is processed with the help of weights and activation function and the result if passed on to the next hidden layer. ANN is trained to learn the pattern from the raw data and can modify the connection wrights through predefined training enabling them to represent patters and relationship with the given data[11].

Deep learning models is trained to detect and recognise patters in text, audio, images etc. without manual engineering[2][11]. Early AI models relied on rule based methods customised for statistical models. With the computational power and ability to handle large datasets, deep learning have become the most sought after technology in modern artificial intelligence. Pattern recognition in deep learning has been instrumental in tasks such as biometric identification, sentiment analysis and voice assistants. Neural architectures with the clan of CNN for image recognition, RNNs and LSTM for sequence modelling and Transformers for parallel processing have promoted its capacity in all major data modalities[4].

Evolution of Deep Learning

The evolution of artificial neural network started from early 1940's where the researchers tried to model the basic functioning of the human brain in computational terms. The first model was known as McCulloch-Pitts neuron is a simple binary threshold unit that mimicked the firing behaviour of biological neurons. In the 1950s Frank Rosenblatt introduced the perceptron, the first learning algorithm for single layer neural network with binary classifications. In the late 1980's back propagation algorithm was introduced by Rumelhart, Hinton and Williams, enabled multi-layer networks to be trained efficiently, reigniting research in neural networks and pattern recognition[17].

After the evolution of deep learning, many notable designs have been developed for different workloads and data formats. The basic foundation of many application is Feed-Forward Neural Network(FNN) where passes in single direction from input layer to output layer traversing through hidden layers. Convolutional Neural Networks(CNN) was built in particular for the image processing. The convolutional layers are the basic building layers of the neural network which extracts the features from the images. Effective training on large image dataset is acquired with the help of CNN. Recurrent Neural Networks(RNN) uses feedback loops to store the user's input in the memory for sequential and time-dependant data.

ransformer Model have gained more popularity in recent times with pattern recognition involving language and beyond. Transformer architecture utilizes self-attention mechanism used for parallel processing of data with key capabilities of text generation, translation, summarization, answering questions and analysis. Transformer architecture was introduced in the year 2017 with the paper "Attention is all you need". Transformers enable the models to understand context over long sequences of data. This framework has led to state-of-the-art in natural language understanding and machine translation. Transformers has laid the foundation for LLMs like Generative Pre-trained Transformer (GPT), and Bidirectional Encoder Representation from Transformers(BERT) which has significantly improved accuracy, scalability and multimodal integration capabilities.

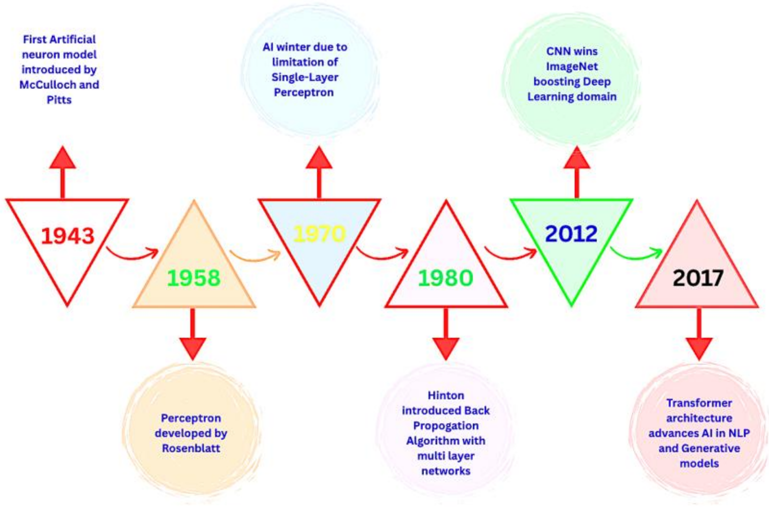

Figure 1: Evolution of Deep Learning

Fundamentals of Neural Network

The biological neural structures present in human brain serve as a motivation for neural network with computational models. The fundamentals of neural network are made up of interconnected layers of artificial neurones or nodes that carry basic calculations. The neurons receive input signals, applies weights and biases and then gives the output to the next layers available. Input layer, hidden layer and output layer are stacked in the modern neural networks to propose the deep neural network which are capable of learning the patterns from the hierarchical data representation [10][17].

Training. Neural network model involves teaching the model with large dataset with training and testing of the model to predict the accurate results by adjusting the internal parameters. The network takes the input data to the input layer, passes it through its layers to produce the output[17]. The output is compared to the true label or the target output with loss function. The loss function quantifies the difference between the predicted output and the actual target output providing a measure of error that enhances the learning process.

Another method to minimize the error is gradient descent where the core idea is to compute the gradient of the loss function with each weight and bias in the network, indicating the direction and magnitude which is adjusted to reduce the error. This gradient descent is done by backpropagation algorithm, which propagates error in backward to the network from output layer to the input layer [8]. By repeatedly adjusting the parameters with small proportionality, the network gradually learns to map the input to the targeted output.

LLM in Text Processing

Deep Learning has transformed the text processing by enabling the model to learn rich, contextual representation of language with large raw data. In the early deep learning approaches, Recurrent Neural Network (RNN) and their variant Long Short-Term Memory(LSTM) networks are used to capture sequential dependencies in text[1]. These models faced the limitation in parallelization and performed poorly with long range dependencies due to their sequential processing nature[4]. The development of transformer architecture revolutionized text processing by introducing self-attention mechanism that allowed models to weight each word with other word in sequence, regardless of distance. This architecture laid the foundation of large Language Model (LLM)

Large Language Modelling is a notable achievement in the field of artificial intelligence (AI), advancing the capabilities of natural language processing (NLP) systems. LLM is built to absorb, understand and write as like a human being[1]. LLM's are trained on large dataset with transformer based architecture and was proven to perform extremely well in a variety of application from making decision to conversational bots etc. LLM is built on Semantic understanding nature of LLM plays an important role in the field of education with wide range of task of summarizing, translating answering the question and creating the content based on the user's input[4]. LLM's are widely used to judge grammaticality, semantic plausibility. LLM enforces consistency and acts a repository of knowledge which is hard to guarantee. LLM works by predicting the next sequence of words based on the user's given context involving billions of parameters with fluency and adaptability.

Effectiveness of LLM in text Processing

Large Language Models(LLM), known as the best variety of text processing model so far with capabilities of capturing complex languages pattern and sematic correlations by using the transformer-based architecture with huge amount of training data[1]. This Transformer-based architecture aims to assess their efficacy in applications like sentiment analysis which detects emotional cues, semantic plausibility ensuring contextual correctness, machine translation, resulting in context-aware translation in various languages. LLM reduces vast amount of information into a short but consistent output. These assessments will prove the ability of LLM in providing significant human-like comprehension that goes beyond basic text analysis.

Transformer-Based Architecture Over Traditional Models

Conventional Neural Networks such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) have been employed for a long time in natural language processing but they have difficulties in dealing the data with parallelisation and dependency over time. The difficulties of these neural network architectures are redefined by the transformer-based architecture which serves as a foundation for LLM. Transformer-Based architecture use attention process which enables models to assess the relative value of words in text, independent of their distance from one another.

Transformer-based architecture uses self-attention mechanism to effectively capture dependencies that prevails while permitting parallelised training, optimising the computational time in contrast with RNNs which suffers from diminished gradients over time and sequential bottlenecks. These capabilities are the foundation of contemporary

Large Language Models as they scale well with huge parameters and large datasets. Transformers represent both local and global interaction via multi-head attention, providing greater understanding of context than CNN which are restricted to local feature extractions. Transformer based architecture goes beyond text enabling multimodal, voice or vision application where the traditional systems are still in process of learning to integrate. Transformer-based architecture redefined the intelligent system that can. Be efficient, scalable and context aware while positioning them as the dominant architecture in AI and ML fraternity[7].

Comparison between CNN, RNN and Transformer-based Architecture

Feature	Convolutional Neural Network	Recurrent Neural Network	Transformer-based Architecture
Domain in use	Image recognition, spatial feature extraction	Sequential data(Speech, text, time-series)	Text, vision, speech, multi-modal tasks
Processing of input	Local receptive fields(spatial features)	Sequential, step-by-step processing	Parallel processing of entire input data
Pros	Good at predicting local patterns(edges, textures)	Captures data with temporal dependencies suitable for sequence modelling	Capable of handling long term dependency data with context-awareness and scalability
Cons	Performs fairly with sequential or temporal data	Struggles with long sequence of data, vanishing gradient problems	Computationally expensive, requires large data sets and hardware configurations.
Processing time	Fast, highly parallelizable	Slow due to sequential nature of input	Fast processing (parallelizable with GPU)
Context Handling	Limited to local spatial context	Limited context but can be improved with LSTM/GRU)	Captures both local and global context
Scalability	Moderate, not ideal for large models	Difficult to scale up with large datasets	Scales efficiently with large parameters and billions of datasets
Application	Object detection, medical imaging and facial recognition	Speech recognition, language modelling, time-series prediction	Machine translation, LLM, multi-modal AI and summarization

Transformer Architecture in LLM

LLMs like GPT, PaLM and LLaMA have been proven to be having exceptional fluency, understanding of the context and versatility by utilizing the deep learning architecture with transformer framework. Transformers are used for parallel processing of data with key capabilities of text generation, translating the text into different languages, code generation and analysis[2].

A neural network design called Transformer is used in machine learning applications specifically in computer vision and natural language processing (NLP). The transformers architecture was presented in the 2017 publication "Attention is All You Need" by Vaswani et al. Transformers enables the models to understand the context of the long sequence word given as input. Transformer architecture uses tokens, the fundamental unit of text the models takes for processing. When working with natural language processing, the text is broken into smaller units called tokens[2]. The tokens can be a word, sub word or a character. Tokens are mapped to vectors which is known as embedding. The models uses these embeddings as inputs to the attention mechanism to learn about the relationship and context between the tokens across the input sequences. Token embedding and positional embedding are the types of embedding that is mapped to the model for processing.

The architecture of transformer was first proposed in a paper " Attention is all you need" which has become a breakthrough model and became the back bone to most of the natural language processing systems[9]. The transformer is made up of encoder-decoder architecture where both stacks are built from multi-head self-attention mechanism, a feed-forward neural network and residual connections with layer normalization. Multi-head attention is used to learn different kinds of relationship between the tokens while positional encoding inserts information about the sequence order

The transformer consist of two main parts which are encoder and decoder. Each part of encoder and decoder are made up of multiple layers including self-attention feed forward network with normalization mechanism. The job of encoder is to read the input and the decoder is used to produce the output[2][9].

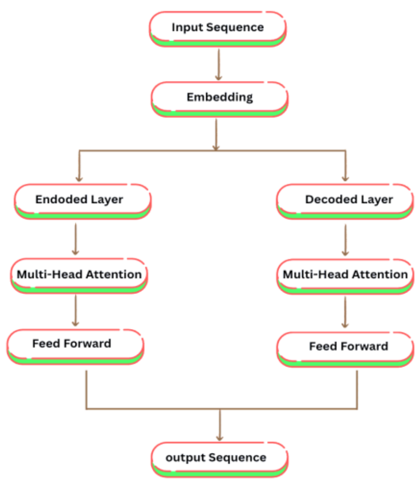

Figure 2 Transformer architecture

Encoder Architecture

The encoder architecture contains three main parts with positional embeddings. The foremost is multi-head attention which is used in encoder that allows the model to focus on different parts of the input simultaneously[11][10]. Multi head attention runs multiple attention mechanism at the same time and concatenate their output enabling the model to focus on different types of relationships[10]. The Attention score can be computed using the formula

$$\text{Attention Score}(Q, K, V) = \left[\frac{\text{softmax}(Q \times k)}{\sqrt{d^k}} \right] V$$

Where Q – query

K – Key,

V - Value

Decoder Architecture

The decoder architecture in transformer is the part of the transformer model which is responsible for generating the output sequences in tasks like text generation, translation, summarization[9]. This architecture is built on the encoder-decoder framework but can be used as a standalone for autoregressive generation.

The Decoder starts by taking the input token converted into embedding(mapped into vector). Positional coding has been added as the model knows the order of the token to be processed. The decoder architecture contains three main parts with masked self-attention, encoder-decoder attention, feedforward and normalization. The masked self-attention si used to prevent the model from peeking into the next token to be processed. The encoder-decoder attention focusses only on encoder's input. Feed forward network helps in generating the output token by token based on the past token encoder's input.

Multi-head (Q, K, V) = concat (head1,head2, …….,head h)W^0

Where head n = Attention(Qw^q_n, Kw^k_n, V w^v_n)

The working phase of LLM takes place in two phases which are training phase and inference phase. The training phase consist of collection of data, Pre-processing the raw data and making them fit for analysis, the model architecture and training of the model. Pretraining is the phase where the model learns general language patterns from the large pool of text data[9]. The objective of the pre-training phase is to predict the next token in a sequence with the previous input given. The data that is used for training is available in abundant from books, articles, internet sources, code repositories etc.

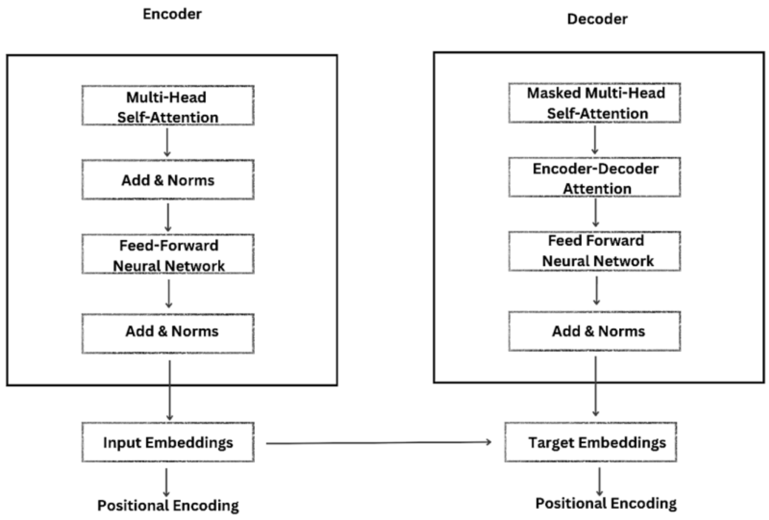

Figure 3 Encoder-Decoder Architecture

Once the model is trained, the model learns grammar, facts about the real world and statistical association between words. The model develops a basic understanding of the language structure. The second phase is inference phase where this phase is responsible for processing the input, generating the output with sampling of data for reference and post processing etc.

Application of LLM

Application	Description	Example Models/ tools
Text Summarization	LLMs can process the large document and extract the important points. This kind of model is very much helpful in reviewing the document, legal case briefs etc.	GPT-4, BART
Machine Translation	Translation of the content takes place without affecting the context and tone	PaLM, mBART
Content Generation	This tool is used to write a creative content custom made based on the user's input and	GPT-4, Jasper AI
Code Generation	Creates a code snippet or identifies the underlying error in the program along with explaining the programming logic.	Codex, GitHub, Copilot
Information Retrieval & Q/A System	Providing accurate answers to the question given by the user and often incorporates RAG(retrieval Augmented Generation) for accuracy	GPT-4 + RAG, Bing Copilot

Educational Tutoring & Adaptive Learning	LLM models are efficient enough in explaining the concept to the user based on the learner's level. Can be ablet o generate personalized exercises, e-learning platforms making learning more easy and interactive	GPT-4 , Khanmigo
Sentiment Analysis & Opinion Mining	LLM models can understand the human emotions based on the behaviour of the interaction with the model. It can detect the user's emotion, opinion and attitude in text and helps to give customized solution to the user in social sentiment trends	RoBERTa, GPT-4

Implications for Education, content generation and multimodal Applications

The rapid development of Transformer-based architecture and Large Language Models(LLMs) has significant implications in the field of education, context generation and different multi modal applications[3]. LLM can change the perspective of the learner by assisting with personalized tutoring systems, adaptive exercises based on the knowledge gained with personalized learning materials and interactive explanations tailor made for the learner's level of understanding. LLM based solutions can dynamically adapt with the learners 's speed and queries however oversight is necessary to address the accuracy, bias ensuring meaning learning outcomes.

LLMs have evolved in the process of generating professional, technical and creative materials in the field of content generation. These systems can provide consistency, context-aware outputs at scale in creative writing, journalism, drafting of legal papers and reports etc. This increases new possibilities for innovation, efficiency and accessibility but also opens up for questions about ownership, novelty and veracity of the content produced. Mechanism for critical evaluation are required to avoid excessive reliance and ensuring that individual creativity and judgement continues to be crucial[3].

LLMs are efficient in handling challenging tasks like medical diagnostics , monitoring, assistive technology and immersive AR/VR experiences that utilises text, language, sight and other modalities. These features shows how AI can be used to produce flexible, context-aware solutions. The important factors that need to be considered are computational expenses, privacy and fair deployment which are still a prevailing issue to be addressed. Multimodal applications represents major advancement in the direction of organic human-computer interaction by combining text, audio, vision and other modalities.

Conclusion

Large Language Model are an incredible development in the field of artificial intelligence which allows the machine to think as like as human in comprehending, producing and communicating with the language that is used by humans with the at most level of accuracy and scalability like never before. The ability of LLM to understand, produce and communicate with human vocabulary at large volumes and high precision is the major advancement in natural language processing[1]. With extensive training and transformer design, the model is enabled to perform well on variety of task including reasoning and creating a creative content, code generation and natural language processing and translation that is tailor made for the user's request[2]. However these models comes up with significant fall backs in spite of having capabilities which can be un done. The drawbacks cab be of restricted interoperability, prejudice, hallucinations, cost involved in developing and training a model and privacy issues.

Future Enhancement

Deep Learning aims to improve the model's accuracy, efficiency, adaptability while expanding the technology across industries. The on-going research in the field of deep learning tries to make more light weight, energy efficient models like model compression, quantization, reducing computational costs with accuracy. Development in LLM is making major advancements that can go beyond little tweaks to provide more human-friendly, reliable and adaptable systems[1]. Advancement in multimodal learning will enable seamless integration of text, speech, images etc. Continual and short learning to the model will allow the model to adapt quickly to the new environment with minimal re-training while preserving the privacy. These kinds of development will open a new path in the field of autonomous systems, personalized healthcare, scientific discovery, and immersive AR/VR environments. In order to provide secure intelligence, research may gear up concentrating on developing systems which are not just portable, cost-effective but also adjustable in real time through federated and lifetime learning.

References

1. Wu, J., Yang, S., Zhan, R., Yuan, Y., Chao, L. S., & Wong, D. F. (2025). A survey on llm-generated text detection: Necessity, methods, and future directions. Computational Linguistics, 51(1), 275-338.
2. Sharrab, Y. O., Attar, H., Eljinini, M. A. H., Al-Omary, Y., & Al-Momani, W. A. (2025). Advancements in Speech Recognition: A Systematic Review of Deep Learning Transformer Models, Trends, Innovations, and Future Directions. IEEE Access.
3. Shahzad, T., Mazhar, T., Tariq, M. U., Ahmad, W., Ouahada, K., & Hamam, H. (2025). A comprehensive review of large language models: issues and solutions in learning environments. Discover Sustainability, 6(1), 27.
4. Li, J., Li, J., Yang, G., Yang, L., Chi, H., & Yang, L. (2025). Applications of large language models and multimodal large models in autonomous driving: A comprehensive review.
5. Hassan, M., Kabir, M. E., Jusoh, M., An, H. K., Negnevitsky, M., & Li, C. (2025). Large Language Models in Transportation: A Comprehensive Bibliometric Analysis of Emerging Trends, Challenges and Future Research. IEEE Access.
6. Nafi'ah, J., & Faruq, D. J. (2025). Conceptualizing deep learning approach in primary education: Integrating mindful, meaningful, and joyful. Journal of Educational Research and Practice, 3(2), 225-237.

7. Viel, F., Maciel, R. C., Seman, L. O., Zeferino, C. A., Bezerra, E. A., & Leithardt, V. R. Q. (2023). Hyperspectral image classification: An analysis employing CNN, LSTM, transformer, and attention mechanism. IEEE access, 11, 24835-24850.
8. Fantin Irudaya Raj, E., & Balaji, M. (2022). Application of deep learning and machine learning in pattern recognition. In Advance Concepts of Image Processing and Pattern Recognition: Effective Solution for Global Challenges (pp. 63-89). Singapore: Springer Singapore.
9. Chitty-Venkata, K. T., Emani, M., Vishwanath, V., & Somani, A. K. (2022). Neural architecture search for transformers: A survey. IEEE Access, 10, 108374-108412.
10. Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2022). Transformers in vision: A survey. ACM computing surveys (CSUR), 54(10s), 1-41.
11. Han, K., Xiao, A., Wu, E., Guo, J., Xu, C., & Wang, Y. (2021). Transformer in transformer. Advances in neural information processing systems, 34, 15908-15919.
12. Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep learning--based text classification: a comprehensive review. ACM computing surveys (CSUR), 54(3), 1-40.
13. Xiong, R., Yang, Y., He, D., Zheng, K., Zheng, S., Xing, C., ... & Liu, T. (2020, November). On layer normalization in the transformer architecture. In International conference on machine learning (pp. 10524-10533). PMLR.
14. Kelleher, J. D. (2019). Deep learning. MIT press.
15. Karita, S., Chen, N., Hayashi, T., Hori, T., Inaguma, H., Jiang, Z., ... & Zhang, W. (2019, December). A comparative study on transformer vs rnn in speech applications. In 2019 IEEE automatic speech recognition and understanding workshop (ASRU) (pp. 449-456). IEEE.
16. Liang, H., Sun, X., Sun, Y., & Gao, Y. (2017). Text feature extraction based on deep learning: a review. EURASIP journal on wireless communications and networking, 2017(1), 211.
17. Wang, H., & Raj, B. (2017). On the origin of deep learning. arXiv preprint arXiv:1702.07800.
18. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.