SSRG International Journal of Computer Science and Engineering
ISSN: 2348–8387 / https://doi.org/10.14445/23488387/IJCSE-V11I6P101

Volume 11 Issue 6, 1-5, June 2024

Original Article

Advancements in Deep Learning Architectures for Natural Language Processing Tasks

Ekambaram Kesavulu Reddy

Department of Computer Science, Sri Venkateswara University Tirupati, Andhra Pradesh, India.

Corresponding Author : ekreddysvu2008@gmail.com

Received: 12 April 2024 Revised: 18 May 2024 Accepted: 03 June 2024 Published: 19 June 2024

Abstract *- Recent years have been an active testing ground for artificial neural networks for language understanding, a very important aspect of NLP. In this respect, emerging NLP technologies are largely motivated by the rising requirements to cope with the issues raised by different NLP tasks, allowing the processing and analysis of large text data samples, uncovering complex language behaviors, as well as extracting valuable information from disorganized text. NLP (Natural Language Processing) has proven to be the most successful field of machine learning thanks to its capability to teach itself and detect all kinds of features on its own based on enormous amounts of data. In NLP tasks like language modelling, text classification, emotion analysis, and machine translation, RNNs, CNNs, and transformer-based models have been used in new ways. While NLP is generally agreed upon the difficulties it faces, the progress of technology also gives birth to unexpected challenges. Thus, two factors, namely the expanding collections of large text datasets and the pressing need for more accurate and time-saving NLP models that emerge as a consequence are giving rise to new kinds of deep learning models and techniques. Here, this paper analyzes as a whole the most recent achievement of neural architectures for natural language processing applications. From introducing current models and approaches in NLP, highlighting their strengths and weaknesses, and identifying the areas to be researched in the future, this paper will conduct this discussion.*

Then, this paper will go on and investigate the of one in NLP, together with the importance of constantly improving architectures which are responsible for tackling these hard tasks. Subsequently, it will talk about the recent breakthroughs in deep learning models namely RNNs, CNNs, transformer-based models and attention mechanisms will be discussed next. At last, this paper will cover the ever-evolving roofline in NLP research, including transfer learning, self-supervised learning, and multimodal learning. Moreover, this paper will also underline the current shortcomings of existing NLP models and locate the themes where research needs to be reevaluated. This article, through the deep learning architecture review for NLP, offered a full-range overview of the recent advancement in deep learning, and this article is developed as a valuable corpus for the researcher, practitioners, and students in the field of NLP.

Keywords *- Deep Learning, Natural Language Processing, CNNs, Recurrent Neural Networks, NLP Models, Transformed based models.*

1. Introduction

Natural Language Processing (NLP) is evolving along with computer and human language communication. As more industries seek to create smart systems that can comprehend and control human language, deep learning is the best technology for most NLP jobs. The depth of learning approach has solved NLP problems like text categorization, sentiment evaluation, machine translation, and speech recognition.

In contrast, NLP is dynamic. Continuous development presents new obstacles for more complex deep learning architecture. The rise of social media and exposure to personal content, whether written or spoken, has highlighted the need for models to analyses sarcasm, irony, and informal expressions. The growing demand for customized experiences drove conversation engagement and conversational AI systems development.

Thus, researchers are developing new neural networks using deep learning architecture for NLP tasks. Recent advances in deep learning architecture improve model capacities to show long-distance relations, process sequential data, and learn from large data sets. These new developments are suitable and effective for a wide range of NLP jobs; thus, they must be evaluated and analyzed.

Its goal is to summarise recent deep learning architectures used for NLP assignments. A deep learning paper will examine RNNs, CNNs, and Transformer-based models. A brief discussion of transfer learning and attention mechanisms in NLP will follow. Additionally, the study provides detailed information about these advancements and their potential uses in natural language processing.

2. Neural Networks

2.1. Recurrent Neural Networks

Due to its flexibility and efficiency in processing sequential data, Recurrent Neural Networks (RNNs) can readily process text, voice, and time series data. RNNs are widely used in NLP for language modelling, translation, and classification.

An RNN's hidden state memory lets it preserve states unchanged when processing fresh data, which may improve temporal dependencies. Thus, RNNs can handle sequence data with contextual connections between time steps from the input sequence representations.

As per Dixon, Halperin and Bilokon (2020), however, RNNs have disagreements that must be handled. The vanishing gradients issue stops RNN training after numerous time steps with poor outcomes. As gradients are rolled back into layers, they become smaller, making this method harder to solve. Another issue is long-term reliance, where the network can produce targets at later time steps based on current observations.

Scientists create complex RNN architecture models to overcome these issues. Long Short-Term Memory (LSTM), the most common architecture, uses a memory cell with numerous forget gates and input gates to transfer information between time steps selectively. LSTMs can learn long-term dependencies and address the disappearing gradients problem with an internal memory system (Franco et al., 2020).

3. Convolutional Neural Networks

3.1. Introduction to CNN's and Their Adaption for NLP Tasks

As per Media (2022), CNNs were originally developed for image recognition, but they have shown promise for NPL tasks in recent years. Convolutional Neural networks transform each word or character to a fixed-length vector for sequential data processing in NLP. This training helps the model identify important and unimportant words in a sentence or paragraph. Exploration of different CNN architectures, including Dilated CNNs and Temporal CNNs

Several types of CNN architectures have been used in NLP tasks, including.

3.1.1. Dilated CNNs

This type of specialist is also called recurrent neural networks (RNN) at the dilation stages. They are the dilative part of the model that maintains a broader time step; thus, the model is capable of capturing long-period dependences (Lee and Shin, 2020).

3.1.2. Temporal CNNs

As per Lee (2019), they do this with data that varies as it progresses, for example, speech, text, or time series data. They exploit the properties of convolutional layers to extract features from the input and then apply the recurrent layers to simulate the temporal dependencies.

3.2. Applications of CNNs in NLP

3.2.1. Text Classification

CNNs are designed for categorizing text into two or more groups, like spam vs. normal or hate vs. compassion. On-spam emails, positive vs. Negative reviews, etc.

3.2.2. Sentiment Analysis

Convolutional Neural Networks (CNNs) can decode text sentiment as positive, negative, or neutral.

The CNN model can be trained to categorize text into distinct categories functionally by continuously passing the text into the series of fixed-length vectors and feeding it into the CNN model. CNN model automatically acquires features from input data and makes judgements using them (Lan, 2020).

4. Transformed-Based Models

Over the past few years, the transformer model has changed Natural Language Processing (NLP). BERT, GPT-3, and Roberta lead many NLP tasks, outperforming RNNs; however, they make it difficult to compare. This section will highlight transformer-based models' benefits and weaknesses.

4.1. Overview of Transformer Architecture

As stated by Schmuck et al. (2020), a year later, in "Attention is All You Need" (1), Vaswani et al. talked about the transformer design. Transformer is an artificial neural network model that can do jobs like machine translation and text summarization that involve going from one sequence to another. The main thing that makes this transformer unique is its self-attention system, which lets it pay equal attention to many parts of the input sequence at once and judge their importance.

4.2. Self-Attention Mechanism

The self-attention mechanism powers transformer architecture. This makes this task easy and crucial for the network. Transformers process all inputs, unlike RNNs. Connects inputs to complicated context. The input sequence ("key," "value," and "query") is vectorized for self-attention. Keys and values are input data; queries are testing context.

The model soft-axes query-key dot product attention weights. Attention weights emphasize the main input sequence step because context matters. Dot creating attention weight particular weights and values gives a model output vector (Ali et al., 2020).

4.3. Pre-Traning and Fine-Tuning Strategies

As per Monarch (2021), transformer-based models are pre-trained on a large data set and then fine-tuned for NLP applications. Pretraining involves teaching the model to read a large text corpus or the full internet to understand context and meaning. In this situation, fine-tuning tailors the pre-trained model to NLP tasks like sentiment analysis and question answering. Transformer-based models require pre-training and fine-tuning to master. After pre-training with 30 million words and fine-tuning with NLP tasks, BERT has shown almost adequate performance (2).

4.4. Comparison of Different Transformer Variants

Many common transformer architectures have been introduced since its release more than a decade ago. These architectures all provide various levels of effectiveness. Some of the most popular transformer variants include: Some of the most popular transformer variants include: *BERT (Bidirectional Encoder Representations from Transformers)* As a pre-trained language model, BERT excels at several NLP tasks. Based on a multi-layer bidirectional transformer encoder. The model represents word meanings with contextualized expressions.

4.5.1.GPT (Generative Pretraining Transformer)

GPT generates text with a transformer-based pre-trained language model. It scored highest in text-generating tasks after targeting 45 million words.

4.5.2. RoBERTa (Robustly Optimized BERT Pretraining Approach)

RoBERTa is an altered variant of the BERT model, which is commonly used for online text classification. It was changed to use Bert architecture and classified texts well in benchmarks.

5. Transfer Learning In NLP

Transfer learning has revolutionized Natural Language Processing (NLP) by letting researchers and programmers employ pre-trained models later. Next, it will explain transfer learning's value and methodologies. Transfer learning applications are next.

5.1. Importance and Role of Transfer Learning in NLP

Transfer learning lets researchers utilize pre-trained model attributes for new tasks. Transfer learning is popular in NLP due to the increased data amount for training from scratch. Researchers spend more time on research since pre-trained models save time, resources, and computing power (Paleyes, Urma and Lawrence, 2022).

5.2.Transfer Learning Techniques

As per Qin and Chiang (2019), NLP uses transfer learning methods, including feature extraction and pre-trained model overwriting. Feature extraction creates a new model using appropriate pre-trained features. This strategy works when the task the model must master later differs from the pretraining model. Trained sentiment analysis models can extract features from new topic models. When model weights are modified to a new target, fine-tuning happens. The example works if the following job matches pre-training. Pre-trained text classification models can be supplemented with sentiment analysis.

5.3. ASE Studies

Several case studies demonstrate how transfer learning improves NLP across domains. BERT, a popular pre-trained language model, has achieved several NLP performance records. BET multi-layer bidirectional transformer encoders provide context-based sentence word representations. Sentiment analysis, question answering, and language translation can use dataset adjustment.

Discriminant BERT RoBERTa classifies text. In several text categorization benchmarks, RoBERTa surpassed trained models. Transfer learning improved low-resource language NLP. Researchers pretrained African language machine translation and text categorization models.

6. Attention Mechanisms

According to Taha, Cosgrave and Mckeever (2022), it appears that all NLP-specialized deep learning networks require attention. Models' ability to focus on language input that people cannot manually produce revolutionizes language processing. This paper will study NLP attention methods, focus on different sorts of attention, and show how they improve model performance and transparency.

NLP models focus on input words, phrases, and sentences using attention. Some features grow more important as the material lengthens, from paragraphs to articles. The model calculates each aspect's value and draws the necessary conclusion.

6.1. Types of Attention Mechanisms

There are two primary types of attention mechanisms used in NLP**.** These are the essential modules that the Transformer model comprises: 1) self-attention and 2) multi-head attention.

6.1.1.Self-Attention

The self-attention model finds hidden information in input data and highlights it in different areas. Imagine reading a phrase and focusing on certain words. Like self-attention, various data items are weighted based on their value to the current mode of action (Tekouabou et al., 2022).

6.1.2.Multi-Head Attention

Multi-head attention allows models to work in parallel with different attention heads focused on different subspaces at different points. A model with multi-head attention learns to focus on both the total input data and each portion.

6.1.3.Improving Model Interpretability And Performance

As stated by Ullal et al. (2021), attention mechanisms have given considerable steps in the domain of model transparency which also stands for greater performance in NLP work and discussions. Attention mechanisms, as one of the fundamental features of input data processing, allow for deep investigation of selected areas. Models use the approach in order to catch long-range dependencies and contextual linkages.

The precedent leads to a qualitative improvement in the model, which leads to an improved understanding of language and gives more accurate outcomes. Finally, attention processes play a significant role in limiting the effect of noise in the input data. Through selective emphasis on the input and paying attention to the most critical parts, models are able not only to perform more effectively but also to provide superior results.

In addition, the mechanism of attention supplies a guideline to understand and comprehend the model's decision-making process. Through attention weights analysis, it will be feasible to specify which sections from the input data, that the model is relying on and why. This system-transparency created is highly useful in the trustworthiness component of NLP systems.

In summary, attention mechanisms have now constituted a significant part of deep learning model recipes for natural language comprehension jobs. They help models learn topical topics without missing key sections of input data long-range dependencies and thus much better produce the proper outputs. By making attention processes more understandable and superb, they have changed the process of language processing into a revolution in the field of language understanding.

7. Self-Supervised Learning

Self-supervised learning predicts incoming data properties without labels or assistance. Natural language processing (NLP) has become a hot field because models can be trained on vast, cheap, and readily unlabeled datasets. Self-supervised learning has several benefits for NLP tasks, including.

7.1. Improved Performance

As stated by Vamathevan et al. (2019), self-supervised learning can pretrain machines on large levels of data to be able to extra-large self-learning high semantic and syntactic features and further fine-tuned the same for specialized NLP tasks, which ultimately paves the way for enhanced performance.

Self-supervision can take the matter of labelling data to a level of being time- and cost-consuming, without even mentioning that scarcity of labeled data is only the case for less-resourced languages.

7.2. Transfer Learning

Pre-trained models can be employed right away for a multitude of NLP tasks. Therefore, the transfer of learning knowledge gives an additional advantage in attacking the other tasks. The process of masking words and predicting them through a scaffold of the whole sentence, along with autoencoding, are the two major techniques of self-supervised learning in NLP.

A masked language model is trained so that it learns to predict the tokens in the subset (words or characters) that have been replaced by special tokens (mask). The model learns which token in context was replaced by the token that was given, so it is an encoder-only model. The approach makes it possible for the model to learn both the meaning and form relations between the tokens, leading to an increase in its ability to the language understanding (Zhang, 2021).

Even autoencoding trains a model to classify labels for the input data, taking into consideration the compressed indicators. In NLP, the process can be achieved by vectorizing a sentence or a paragraph into a lower-dimensional space and then decoding it back to the original phrase. This way, the model identifies the language phenomenon deep within, namely, syntax and semantics, and easy to generate text translation or summarization (Sarker et al., 2020).

8. Conclusion and Future Directions

This paper examines the latest work architecture deep learning for natural language processing advances. Deep learning's importance in NLP, sequential data problems, and evolutionary architectures' solutions were discussed. Then, it examines the development of RNNs, CNNs, and Transformer-constructed models. It is found that transfer learning, attention processes, and self-supervised learning improve NLP models. However, substantial computational resources and large volumes of labelled data are needed to explain why the model makes specific judgements in many circumstances. The next phase is to research new models with efficacy, interpretation, and adaptability for a variety of NLP tasks.

8.1. Natural Language Processing

As natural language processing advances, the next challenge is leveraging breakthroughs and creating the future. It presumably avoids NLP model robustness, accuracy, and interpretability. These advancements can be achieved by developing strategies to help models manage exceptional inputs without being over-computationally hard and by increasing transparency in model decision-making.

NLP models may integrate images, videos, and audio, among other developments. This will enable speech subject and subtlety consideration, improving NLP tasks. It can predict increased research on transfer learning and few-shot learning technologies that enable adaptation to new tasks and domains without training data (Uddin et al., 2019).

8.2. Recurrent Neural Networks

RNNs, which is the deep learning architecture type that exhibits a high performance when it comes to Natural Language Processing (NLP) activities, have the possibility of revolutionizing the way the research is done. RNNs are useful for language modelling, machine translation, and text classification since they handle sequential data. There are two basic RNN types: Simple RNNs and LSTM networks have stacking gradient-based faults. Simple RNNs update each time step using their own hidden state, but LSTM networks use a more complex design that can selectively forget or keep prior learning. Research shows that LSTM networks outperform Simple RNNs for several NLP applications. LSTM networks handle long-run factors well; hence, they perform better than Simple RNNs for language modelling. Machine translation used LSTM networks in word processing sequences and succeeded in the leading technology despite other attempts to catch it.

Acknowledgement

I am most thankful to my son Sri.E.Digvijaykumar Reddy, Walmart, Bangalore and my Daughter Smt.E.Spandhana, Oracle corporation of India, Bangalore, for supporting help, encouragement and cooperation at my home.

References

[1] Matthew F. Dixon, Igor Halperin, and Paul Bilokon, "Machine Learning in Finance," *Springer Cham*, vol. 1170, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] Muriel F. Franco et al., "SecRiskAI: A Machine Learning-Based Approach for Cybersecurity Risk Prediction in Businesses," *2022 IEEE 24th Conference on Business Informatics (CBI),* Amsterdam, Netherlands, pp. 1-10, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow*, 3rd Ed., O'Reilly Media, 2022. [Google Scholar] [Publisher Link]

[4] In Lee, and Yong Jae Shin, "Machine Learning for Enterprises: Applications, Algorithm Selection, And Challenges," *Business Horizons*, vol. 63, no. 2, pp. 157-170, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] Sebastian Raschka, *Python Machine Learning*, Packt Publishing, pp. 1-454, 2019. [Google Scholar] [Publisher Link]

[6] Guanghui Lan, "First-Order and Stochastic Optimization Methods for Machine Learning," *Springer Cham*, vol. 1, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Andrés Martínez et al., "A Machine Learning Framework for Customer Purchase Prediction in The Non-Contractual Setting," *European Journal of Operational Research*, vol. 281, no. 3, pp. 588-596, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Mahmood Asad Moh'd Ali et al., "Transforming Business Decision Making with Internet of Things (IoT) And Machine Learning (ML)," *2020 International Conference on Decision Aid Sciences and Application (DASA)*, pp. 674-679, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Robert (Munro) Monarch, "Human-in-the-Loop Machine Learning: Active Learning and Annotation for Human-Centered AI," *Manning*, pp. 1-325, 2021. [Google Scholar] [Publisher Link]

[10] Andrei Paleyes, Raoul Gabriel Urma, and Neil D. Lawrence, "Challenges in Deploying Machine Learning: A Survey of Case Studies," *ACM Computing Surveys*, vol. 55, no. 6, pp. 1-29, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] S. Joe Qin, and Leo H. Chiang, "Advances and Opportunities in Machine Learning for Process Data Analytics," *Computers and Chemical Engineering*, vol. 126, pp. 465-473, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[12] Iqbal H. Sarker et al., "Cybersecurity Data Science: An Overview from Machine Learning Perspective," *Journal of Big Data*, vol. 7, no. 41, pp. 1-29, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Ayman Taha, Bernard Cosgrave, and Susan Mckeever, "Using Feature Selection with Machine Learning for Generation of Insurance Insights," *Applied Sciences*, vol. 12, no. 6, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] Stephane Cedric Koumetio Tekouabou et al., "Reviewing the Application of Machine Learning Methods to Model Urban Form Indicators in Planning Decision Support Systems: Potential, Issues and Challenges," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 8, pp. 5943-5967, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Mithun S. Ullal et al., "The Role of Machine Learning in Digital Marketing," *Sage Open*, vol. 11, no. 4, pp. 1-12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[16] Shahadat Uddin et al., "Comparing Different Supervised Machine Learning Algorithms for Disease Prediction," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, pp. 1-16, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[17] Jessica Vamathevan et al., "Applications of Machine Learning in Drug Discovery and Development," *Nature Reviews Drug Discovery*, vol. 18, no. 6, pp. 463-477, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[18] Xiaofang Zhang, "Application of Data Mining and Machine Learning in Management Accounting Information System," *Journal of Applied Science and Engineering*, vol. 24, no. 5, pp. 813-820, 2021. [CrossRef] [Google Scholar] [Publisher Link]