Natural Language Processing with deep learning

Muhammad Umer Bin Yaqoob0
dept. Electronic Engineering

***Abstract*—This study presents an in-depth exploration of NLP through the evolution from rule-based systems to advanced deep learning methodologies. It emphasizes the transition from statistical methods to the deployment of neural networks, such as RNNs and LSTM networks, which are proficient in handling sequential data and recognizing temporal and contextual nuances in text. With the advent of attention mechanisms and the Transformer architecture, NLP models have significantly increased in efficiency and efficacy. These days companies receive hundreds of applications against each job opening which adds to the work of sorting through all these resumes to find out candidates that are required for the job. In this research we have proposed a DistilBERT based deep learning classification technique to predict job titles based on the resume data so that companies that need to speed up the recruitment process they don't have to go through a lengthy process of sorting through resume documents. Results suggest that if we have enough dataset to train the model on each job title it can significantly reduce the time taken by this process.**

I. Introduction

Natural Language Processing (NLP) has progressively grappled with the complexities of human language through the lens of computational intelligence. Initially, the field of NLP depended on a foundation of rules crafted by linguists to decode and interpret text, which proved to be fragile and constrained due to the dynamic and intricate nature of language. A pivotal transition occurred with the introduction of statistical methods, where machine learning algorithms began to discern linguistic patterns within extensive text datasets for language prediction. The true metamorphosis, however, was catalyzed by the emergence of deep learning approaches within the field of NLP. The deep learning revolution in NLP was characterized by the deployment of intricate neural networks with multiple layers. This revolution was further propelled by the deployment of architectures such as Recurrent Neural Networks [] (RNNs) and Long Short-Term Memory [2] (LSTM) networks, which were adept at handling data sequences and recognizing temporal relationships and contextual nuances within text. Innovations continued with the advent of attention mechanisms and the transformative Transformer architecture, which facilitated the simultaneous processing of text sequences, thereby markedly enhancing the efficiency and efficacy of NLP models. This gave rise to models like Bidirectional Encoder Representations from Transformers [3] (BERT), Generative Pretrained Transformer [4] (GPT), and its variants [5], which have set new standards in NLP tasks through their ability to understand context and generate human-like text. Despite these advancements, challenges in NLP persist. Ambiguity, context-sensitivity, and the subtleties of human language make NLP a complex field. Deep learning addresses these challenges by leveraging large datasets and computational power to learn nuanced patterns in language. Transfer learning [6] and unsupervised pre-training allow models to develop a broad understanding of language before being fine-tuned on specific tasks, enabling them to handle a wide range of NLP challenges. However, issues such as the need for vast amounts of annotated data, the interpretability of models, and the computational resources required remain areas of active research and development.

II. Background and Related Work

The trajectory of NLP methodologies has been marked by significant milestones, evolving from rule-based systems [7] to advanced deep learning techniques. Initially, NLP systems were constructed around linguistic rules developed by experts, which allowed for the parsing and interpretation of text but lacked scalability and struggled with the nuances of language. The emergence of machine learning in NLP brought statistical models to the forefront, utilizing algorithms like decision trees [8] , Naïve Bayes [9], and Support Vector Machines [10] to learn from data. However, these models were limited in their ability to capture context and sequence information inherent in language. The integration of deep learning into NLP has been transformative, introducing models with the capacity to identify complex patterns within voluminous datasets. RNNs and LSTMs have been pivotal in addressing the sequential nature of text, facilitating the identification of inter-sentence and inter-paragraph dependencies. Convolutional Neural Networks [11] (CNNs), traditionally associated with image processing, have been repurposed for NLP tasks, enabling the detection of textual patterns and the execution of classification tasks. The unveiling of the Transformer architecture, as expounded in the seminal work "Attention is All You Need [12]," constituted a paradigm shift within NLP. This architecture eschewed the sequential processing of RNNs and LSTMs in lieu of attention mechanisms, granting models the ability to appraise the significance of words across a sentence or document, independent of sequential position. This advancement in parallel processing substantially enhanced the models' operational efficiency and effectiveness. BERT further extended this framework by integrating bidirectional context, allowing for the comprehension of contextual relationships surrounding each token from both directions, thereby achieving significant gains in language processing tasks. The development trajectory of NLP continued with the GPT series, which augmented the Transformer's architecture to emphasize unsupervised learning and the generation of contextually coherent text. Successive iterations of GPT have seen an increase in model size and complexity, resulting in architectures with extensive parameter sets capable of performing a broad spectrum of language-related tasks without the necessity for task-specific pre-training. These developments have been instrumental in shaping the current landscape of NLP research, which is persistently pushing the boundaries of language comprehension and generative capabilities, with the aim of developing more sophisticated AI systems. The field of NLP is characterized by continuous innovation, with research efforts dedicated to addressing the challenges related to model efficiency, interpretability, and the ethical deployment of NLP technologies.

III. Foundations of Deep Learning

Inspired by the neural networks of the human brain, artificial neural networks [13] (ANNs) are structured as a series of nodes, termed "neurons," interconnected by "synapses" that carry signals. Inputs are received by these neurons, processed via an activation function, and the resulting output is relayed forward. The perceptron, a basic ANN form, comprises a singular neuron with tunable weights and bias, and is adept at performing linear classifications. Neurons within more complex, multi-layered networks are organized into distinct layers: the input layer accepts the initial data, hidden layers perform computations, and the output layer delivers the final verdict. The connection strengths, or weights, modulate the impact of these inputs on the outputs, with network training entailing the adjustment of these weights to minimize output errors against expected results. Deep learning, a specialized branch of machine learning, is characterized by its use of multi-layered neural networks, facilitating a tiered approach to feature learning. This stratified model architecture permits the extraction of increasingly complex patterns from basic elements, a concept known as feature hierarchy. Early network layers might identify rudimentary features such as edges in visual data or phonemes in audio processing, while deeper layers synthesize these to discern more intricate forms like shapes or specific lexicon. Unlike traditional machine learning techniques that necessitate handcrafted feature extraction, deep learning architectures inherently and dynamically learn these feature hierarchies from the data. The training of deep neural networks is underpinned by back propagation [14], an algorithm that calculates the gradient of the loss function concerning the network's weights. The network's forward pass circulates the input to produce an output, which is then evaluated against the target output by the loss function. Subsequently, back propagation undertakes a reverse pass to determine the loss function's gradient, informing weight adjustments to reduce the loss, thereby optimizing the network's performance.

References

[1] L. R. Medsker and L. Jain, "Recurrent neural networks," Des. Appl., vol. 5, no. 64–67, p. 2, 2001. This reference discusses Recurrent Neural Networks (RNNs) which are fundamental to deep learning in NLP, particularly for their ability to handle sequential data such as text.

[2] J. Cheng, L. Dong, and M. Lapata, "Long short-term memory-networks for machine reading," ArXiv Prepr. ArXiv160106733, 2016.

[3] S. Alaparthi and M. Mishra, "Bidirectional Encoder Representations from Transformers (BERT): A sentiment analysis odyssey," ArXiv Prepr. ArXiv200701127, 2020.

[4] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, and others, "Improving language understanding by generative pre-training," 2018.

[5] K. Roumeliotis and N. Tselikas, "ChatGPT and Open-AI Models: A Preliminary Review," Future Internet, vol. 15, no. 6, p. 192, 2023, doi: 10.3390/fi15060192.

[6] L. Torrey and J. Shavlik, "Transfer learning," in Handbook of research on machine learning applications and trends: algorithms, methods, and techniques, IGI global, 2010, pp. 242–264.

[7] M. Santaholma, "Grammar sharing techniques for rule-based multilingual NLP systems," in Proceedings of the 16th Nordic Conference of Computational Linguistics (NODALIDA 2007), 2007, pp. 253–260.

[8] L. Rokach and O. Maimon, "Decision trees," Data Min. Knowl. Discov. Handb., pp. 165–192, 2005.

[9] K. P. Murphy and others, "Naive bayes classifiers," Univ. Br. Columbia, vol. 18, no. 60, pp. 1–8, 2006.

[10] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," IEEE Intell. Syst. Their Appl., vol. 13, no. 4, pp. 18–28, 1998.

[11] P. Kim and P. Kim, "Convolutional neural network," MATLAB Deep Learn. Mach. Learn. Neural Netw. Artif. Intell., pp. 121–147, 2017.

[12] A. Vaswani et al., "Attention is all you need," Adv. Neural Inf. Process. Syst., vol. 30, 2017.

[13] A. Abraham, "Artificial neural networks," Handb. Meas. Syst. Des., 2005.

[14] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," nature, vol. 323, no. 6088, pp. 533–536, 1986.

[15] J. Ramos and others, "Using tf-idf to determine word relevance in document queries," in Proceedings of the first instructional conference on machine learning, Citeseer, 2003, pp. 29–48.

[16] S. Anbhawal, "Resume Dataset." 2023. [Online]. Available: https://www.kaggle.com/datasets/snehaanbhawal/resume-dataset

[17] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," ArXiv Prepr. ArXiv191001108, 2019.