*Research Article*

# The Survey: Advances in Natural LanguageProcessing using Deep Learning*

**Vamsi Krishna Vedantam,**
Advanced Analytics, Tech Mahindra, Copenhagen, Denmark

**Article History**: Received:11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

**Abstract—**Natural Language Processing using Deep Learning is one of the critical areas of Artificial Intelligence to focus in the next decades. Over the last few years, Artificial intelligence had evolved by maturing critical areas in research and development. The latest developments in Natural Language Processing con- tributed to the successful implementation of machine translations, linguistic models, Speech recognitions, automatic text generations applications. This paper covers the recent advancements in Natural Language Processing using Deep Learning and some of the much-waited areas in NLP to look for in the next few years. The first section explains Deep Learning architecture, Natural Language Processing techniques followed by the second section that highlights the developments in NLP using Deep learningand the last part by concluding the critical takeaways from my article.

***Index Terms*—**Natural Language Processing, Artificial Intel- ligence, CNN, RNN, LSTM, Attention, Transformer, Transfer Learning in NLP, Natural Language Understanding, Deep Gen- erative models.

## I. INTRODUCTION

Natural language processing(NLP) is a field of Artificial Intelligence where computer performs human-like activities such as understanding the meaning of the text, translate one language to other, recognize speech and convert to meaningfulactions, generate and summarize text, the sentiment of a topic,web search, segmentation of documents, radiology reports etc.Developing NLP applications is never an easy task as ma- chines expect humans to code them via programming language [1]. Traditional approaches in understanding human speech using programming languages often underperform because there are many complex areas to handle, such as dialects, the context of the topic and jargons used by speaker [1].Many machine learning based NLP algorithms evolved over the past decades that handles most of the limitations mentioned above to an extent. The advances in Deep Learning based NLP has been gaining much traction in recent years, which solves complex tasks that legacy models fail to achieve. Thanks to massive computational power and research community.

***Deep Learning*** is a statistical technique that determines patterns from the data, using neural networks [2]. Deep Learn- ing has multiple architectures such as Deep Neural Network, Convolutional Neural Network, Deep Belief Networks, Recur- rent Neural Network that are being used by computer science engineers across the world to develop right from simple image recognition to complex driverless car applications. State of the art systems developed using Deep Learning framework are capable of outperforming humans, and in most of the cases, these systems can handle where a human takes years to solve. Some people especially rookies into the field assume that morenumber of hidden layers in the neural network makes Deep Learning Networks but rather its ability to solve extremely challenging problems and its ability to retain properties from layers makes Deep Learning much reliable and powerful tool compared to shallow networks [3].

The Recent advancements in computing, deep learning approaches have obtained very high performance across many different NLP tasks. One of the recent developments in NLP space is Knowledge Graphs(KG). Knowledge Graph some- times also referred to Knowledge Base, is a vital source for understanding human knowledge represented as structured entities and relational concepts about them [4]. One way to construct knowledge graphs is by developing machine learningor deep learning models that extract the relationship between entities from large amounts of text from the internet [4]. Among all the NLP applications, deep learning based models such as sequence-to-sequence algorithms comprising RNNs have made significant advancements in machine translation in recent times and outperformed traditional approaches [5].

## II. DEEP LEARNING APPLIED TO NATURAL LANGUAGE PROCESSING

Deep in the context of Neural Networks is when lay- ers(N) are greater than two in a network. Deep learning architectures broadly categorised into Deep Neural Networks, Convolutional neural network(CNN), Recurrent Neural Net- works(RNN) and Long short-term memory(LSTM)/Gated re- current units (GRUs). This section gives a brief introductionto CNN, GRU and LSTM architectures.

### *A. Convolutional neural network(CNN):*

CNN is a supervised deep learning architecture used in several fields, including NLP. Convolutions in CNN preserves information between pixels by learning patterns of image data in the form of tiny and meaningful features such as edges. CNNs are used extensively in image recognition/classification,video processing and video captioning, as well as speech recognition and other NLP applications. Most of the times, it is essential to focus on whether or not information appears in particular areas rather than concentrating on precisely where it occurred [6]. With the growing challenges in the fields like

computer vision, deep learning models becomes very hard to handle due to the increase in complexity. The super-powerful models require vast amounts of data for training to get the best outcomes. As the size of the data increases in these complex models, the time it takes to train the model also increases tremendously, that led to the introduction of techniques for fast processing of CNNs [7]. The primary advantage of fast processing of CNNs is the Fourier transformations of filters where filters gets reused and process with different images in parallel in a mini-batch. The other benefit is that weights can be reutilized while weights gets updated into both input im- ages and filters simultaneously, and more importantly, inverse transformations are required only once per channel per image [8].

### *B. Recurrent Neural Networks(RNN):*

Recurrent Neural Networks are the most used Deep Learn- ing variant in NLP tasks. In traditional NN, all inputs and outputs are interdependent. While performing NLP tasks, it expects words in sequence, and the prediction of the next word has a strong influence from the previous set of words in a given sentence. Sometimes, words that occur in the latter part of the text also important in predicting the current word. Thus, both backward and forward directions of looking at the sentences are very critical. This form of handling words in both directions is called bidirectional RNN [6]. RNN makes use of its ability to remember all related words in the sentence and learns weights from backpropagation while training the model. However, RNNs has a major disadvantage when it comes to memorizing previous words in a sentence that are very far from the current word, which means, while training the network, weights that are supposed to update through backpropagation gets diminished as it passes to each previous step in the network. The sequence of words can have a different combination between input and output relationship such as one-to-one, one-to-many, many-to-one and many-to-many.

### *C. Long Short-Term Memory(LSTM):*

The most powerful architecture under deep learning for NLP is LSTM. Among different types of neural networks for embedding the sentences, Bi-directional LSTM is prominent and gives the best outcomes in language modelling, machine translation, question answering applications. Though it hassome of the limitations such as high computing power dueto non-parallel process within given sentence [10] and weaker capability in capturing words in long-range in a sentence that leads to weak performance in encoding longer sentences [11], LSTM still the best choice for most of the Natural Language Processing models. One way to overcome these limitations is to process all the words in sentence simultaneously at each recurrent step instead of treating each word individually in a sentence in hidden states [12]. This method of updating states by exchanging information recurrently termed as sentence state LSTM. Sentence encoding with this method can be achieved effectively after initial recurrent steps, while the number ofsteps that required for bi-directional LSTM scales with thelength of the sentence [12].

## III. LATEST ADVANCEMENTS IN NLP USING DEEP LEARNING:

Natural Language Processing has come a very long wayin solving some of the fundamental tasks using computers that most of the humans performs or sometimes struggles unless an expert in language to achieve language translation, summarizing the text, sentiment from the text etc. Next few years would be critical for further advancements for NLP, especially using deep learning. Rest of the sections in this paper covers some of the critical areas of focus for next-generation Natural Language Processing.

### *A. Attention:*

Attention mechanism is a method where the importance score of each element is calculated and based on this score encoding sequence is performed. This method applied in various Natural Language Processing applications such as text summarization, sentiment classification, question and an- swering tasks, etc.. and achieved significant improvement in the results [12]. Introduced into neural machine translationin 2014(by Bahdanau, Cho, and Bengio), this method hasgained prominent place in NLP tasks as this address some of the challenges from traditional RNNs like vanishing meaning of the text after multiple steps and explicit word alignment during encoding process. Attention models still based on RNNs while encoding the text, but during the decoding, it computes and assigns attention score to each of the hidden representation in the network [12]. By calculating attention score on each element, and ranking of high encoding value, both the challenges from RNN are addressed. There are many variations in attention mechanism-based models like memory-based attention, multi-dimensional attention, self attention, Hierarchical attention etc.. Achieving great results on manyof the NLP tasks, attention based models widely accepted in various sectors. Another advantage with attention models is determining feature importance from neural network models [13]. Traditionally, understanding neural network models are very hard and some times impossible hence called as black- box models, Attention mechanism solves this mystery problemas well to an extent.

### *B. Transformer:*

First proposed by Ashish Vaswani in ”Attention All You Need”, this method is gaining popularity over the last few years. Though this is based on Attention mechanism, Trans- formers works by taking very less time to train through par- allelizing the process and manage the dependencies between input and output and continuous recurrence [14]. In traditionalmodels, several operations that are required to link signals between input and output increases the distance and makes themodel more difficult to learn the dependencies. This approach shortens the distance to a fixed number of operations by averaging weighted positions and multi-head based attention approach [14]. Transformer compares each word with the rest of the words in the sentence while computing the next possiblerepresentation of a word and generates attention value for all possible words in sentence [15].

### *C. Neural Machine Translation:*

Neural Machine Translations primarily works on the princi- ples of encoding and decoding. In Machine Translations using Neural Networks, each sentence in the source language is encoded into a fixed vector. The encoded vector then gets decoded to language to which sentence has to be converted. Though this architecture works best in most of the scenarios, there are some limitations such as performance of encoder- decoder worsens while Neural network tries to compress required information in long sentences of a source language into a fixed vector. A solution to overcome this challenge is proposed in the paper [16]. In this proposed approach, rather than encoding an entire input sentence to a fixed vector, it creates vectors in sequence while encoding the input sentence and selects a subset of the vectors adaptively decoding the sentence. This approach also works best while translating longsentences by freeing neural network model from the need of compressing entire sentence to one single vector [16].

### *D. Machine reading comprehension:*

One of the key areas of advancements in NLP is machine’s ability to read the text and extract meaning and context out of it. This gave way forward for machine reading compre- hension where machine reads the text(passage) and answers the questions from the text. Researchers and developers were struggling to analyse clean and reliable text data that canbe used for Machine Reading Compression.Stanford(SQuAD)had released open dataset that address need for large andhigh quality dataset [17]]. Researchers are working towards taking Machine Reading Comprehension models to next level, especially on questions and answers models. Like generating answers from the passage for the respective questions, Han Xiao and team had proposed ”Dual Ask-Answer Network for Machine Reading Comprehension” [18] which generates ques- tions from the text. Neural network models that are used for answering the questions using machine reading comprehension models, same neural network can be leveraged for question generation [18].

### *E. Natural language understanding(NLU):*

Natural Language Understanding is a subset in NLP that interprets the meaning of the text and communicates in sim- ilar to human-like interaction. The perfect example of NLU advances in recent years are Alexa, Siri, Google’s Assistant etc.. Recent advances in NLU, especially in NLP by using pre-trained models such as BERT in many applications that made these applications much successful. However, performance deteriorates when pre-trained models directly applied to fine- tune the conversational corpora. As proposed in the paper [19] ToD-BERT handles this by considering goal-oriented dialogues that work best for many conversational interactions.

Research communities, especially big players like Microsoft, Google and Amazon helping the developers around the world by releasing the pre-trained models. For example, MT-DNN is an open-source NLU toolkit that makes developers andresearches to train the NLP or NLU models using deep learning models [20].

*F. Deep generative models:*

Natural Language Processing using deep generative modelsare critical to incorporate vast linguistic features by leveragingunlabelled data and to learn patterns from the data. VariationalAuto-Encoders (VAEs) and Generative Adversarial Networks (GANs) are the two key deep generative models for Natural Language processing. Availability of text data in differentforms makes the NLP applications need of the hour, and at thesame, it also highlights limitation solve to complex problems. Latest deep learning techniques that are used in most of the complex NLP tasks are VAEs and GANs. Traditional deep learning algorithms are best suited on labelled text datasets. However, most of the text data for analysis are without labels. VAEs, which is unsupervised model, addresses this issue by encoding and decoding data the latent variables to reconstruct the text data [21]. Similar to VAEs, GANs are also prominent in solving NLP problems. GANs has a generator that generatesthe output using the semantic vector while false discriminator examples in the text [22].

*G. Transfer Learning:*

Transfer learning described as a set of methods that are generalized to reuse without training the model from scratch. Transfer Learning approach traditionally is the key area in Image recognition, Image classification etc.. Over the last few years, transfer learning in NLP is gaining popularity as these methods produce state-of-the-art results in NLP tasks. However, a lot more research is in progress to decide how well transferable neural networks can be leveraged [23]. Se- mantically similar items, word embeddings, MULT, and INIT, are the best scenarios where Transfer Learning can be applied [23]. Existing transfer learning approaches in NLP possess some limitations such as fine-tuning all the weights of the pre- trained model that could produce weak results when dealing with smaller datasets. Similarly, another approach under con- sideration is multi-task learning, but due to retraining of the entire pre-trained model is a big drawback in this approach. Recently, researchers have come up with 'adapters' which addresses catastrophic problems and fine-tuning of entire pre- trained weights [24].

*H. Knowledge and common sense:*

The amount of textual data in the form of web content and written in multiple languages is increasing exponentially everyday from all parts of the world. Due to this increase in text data, the advancements in NLP is heading towards processing semantic information from massive corpora to understand it in a more meaningful way. Such applications can be built using common sense semantic resources [25]. To achieve common sense reasoning tasks, it requires a model to go beyond traditional approaches like pattern recognition. Majority of thetimes, NLP applications requires an understanding of narra- tives that can be achieved only by commonsense reasoning with regards to the people [26]. The pragmatic inference has high potential to be used in many NLP applications that require exact involvement of humans intents and emotional reactions [26]. Natural Language systems still not immune to challengeslike understanding the text in a similar way how human fol- lows it along with common sense knowledge and interpreting the data [27]. One way to overcome these challenges is by introducing external knowledge with leveraging pre-trained models that produce an improved understanding of language and coherence of generated texts [28].

*I. Low-resource NLP tasks:*

There are several thousand languages that are in use in the world, out of these only few languages are being supported by the majority of Natural Language processing. The main challenge in leaving out most of these languages are due to the availability of resources, and these are termed as low resources languages. The next big step that NLP researchers are focussing is improving low resource NLP models. One of the well-noted approaches that many developers follow is applying text extraction simultaneously on multiple large set of languages [29]. Some of the recent advances include morphological analysis based linguistic models that perform identification, analysis and description of words that builts up from smaller pieces of words [30]. The domain adversarial learning helps in reducing overfitting in low resource models [31]. When new language has to be considered for NLP tasks, monolingual word vectors can be directly used rather than using multi-lingual pre-trained vectors [31].

## IV. CONCLUSION:

In this article, I presented a detailed survey of the re- cent advances in Natural Language Processing using Deep Learning. The success of Artificial Intelligence in next few years depends on how well text data is utilized by research community and developers to build NLP applications using Deep Learning and solve some of the complex problems that traditional approaches struggle to provide the right solution. These advancements in NLP is very crucial for various target sectors such as Healthcare, Banking, Financial Services and Insurance (BFSI), IT Telecommunication, Automotive, Media Entertainment, Retail, Transportation, Aerospace and many more. This research should help researchers, developers and student community to focus on some of the critical areas under Natural Language Processing in the coming years. This article also represents some of the prominent NLP scientists studies that could potentially change the way text data is handled and emphasizes the ways to overcome challenges with existing approaches in Natural Language Processing.

## REFERENCES

[1] Marc Moreno Lopez and Jugal Kalita “Deep Learning applied to NLP,” in Computation and Language (cs.CL) in arXiv:1703.03091 URL https://arxiv.org/abs/1703.03091

[2] Gary Marcus ”Deep Learning:A Criti-cal Appraisal” in cs.AI inarXiv:1801.00631URL:https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf

[3] Book: Neural Networks and Deep Learning The original online book :http://neuralnetworksanddeeplearning.com

[4] Book: Deep Learning in Natural Language Processing by Deng, Li, Liu, Yang (Eds.)

[5] Hang Li ”Deep learning for natural language processing: advantages and challenges” //academic.oup.com/nsr/article-abstract/5/1/24/4107792

[6] Daniel W. Otter, Julian R. Medina, and Jugal K. Kalita in ”A Survey of the Usages of Deep Learning for Natural Language Processing”, in arXiv:1807.10854v3 . URL: https://arxiv.org/pdf/1807.10854.pdf

[7] M. Mathieu, M. Henaff, Y. LeCun, Fast training of convolutional networks through ffts, in: Proceedings of the International Conferenceon Learning Representations (ICLR), 2014.

[8] Jiuxiang Gua, , Zhenhua Wangb, , Jason Kuenb , Lianyang Mab , Amir Shahroudyb , Bing Shuaib , Ting Liub , Xingxing Wangb , Li Wangb
, Gang Wangb , Jianfei Caic , Tsuhan Chenc in ”Recent Advances in Convolutional Neural Networks ”

[9] Andrej Karpathy blog in blog: http://karpathy.github.io/2015/05/21/rnn- effectiveness/

[10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In NIPS. pages 6000–6010.

[11] Philipp Koehn and Rebecca Knowles. 2017. Six challenges for neural machine translation. In Proceedings of the First Workshop on Neural Machine Translation. Vancouver, pages 28–39.

[12] Dichao Hu in ”An Introductory Survey on Attention Mechanisms in NLP Problems” in arXiv:1811.05544v1 [cs.CL] 12 Nov 2018

[13] Blaz Skrlj 1 and Saso D zeroski 1 and Nada Lavrac 1 and Matej Petkovic´ 1 in Feature Importance Estimation with Self-Attention Net- works. arXiv:2002.04464v1 [cs.LG] 11 Feb 2020

[14] Ashish Vaswani Google Brain avaswani@google.com Noam Shazeer Google Brain noam@google.com Niki Parmar Google Research nikip@google.com Jakob Uszkoreit Google Research usz@google.com Llion Jones Google Research llion@google.com Aidan N. Gomez
† University of Toronto aidan@cs.toronto.edu Łukasz Kaiser Google Brain lukaszkaiser@google.com in ”Attention is all you need” arXiv:1706.03762v5 [cs.CL] 6 Dec 2017

[15] Blog by Google AI: https://ai.googleblog.com/2017/08/transformer- novel-neural-network.html

[16] Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio in ”Neural Machine Translation by Jointly Learning to Align and Translate”. arXiv:1409.0473

[17] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, Percy Liang in ”SQuAD: 100,000+ Questions for Machine Comprehension of Text”. arXiv:1606.05250

[18] Han Xiao, Feng Wang, Jianfeng Yan, Jingyao Zheng in Dual Ask-Answer Network for Machine Reading Comprehen- sion.arXiv:1809.01997

[19] Chien-Sheng Wu, Steven Hoi, Richard Socher and Caiming Xiong in ”ToD-BERT: Pre-trained Natural Language Understanding for Task- Oriented Dialogues” arXiv:2004.06871 [cs.CL]

[20] Xiaodong Liu , Yu Wang , Jianshu Ji, Hao Cheng, Xueyun Zhu, Em- manuel Awa, Pengcheng He, Weizhu Chen, Hoifung Poon, Guihong Caoand Jianfeng Gao in ”The Microsoft Toolkit of Multi-Task Deep Neu- ral Networks for Natural Language Understanding”. arXiv:2002.07972 [cs.CL]

[21] Touseef Iqbala and Shaima Qureshia in ”The survey: Text generation models in deep learning”

[22] MingZhou,Nan Duan,ShujieLiu,Heung-Yeung Shum in ”Progress in Neural NLP: Modeling, Learning, and Reasoning”

[23] Lili Mou, Zhao Meng, Rui Yan, Ge Li, Yan Xu, Lu Zhang, Zhi Jin in ”How Transferable are Neural Networks in NLP Applications?”. arXiv:1603.06111 [cs.CL]

[24] Jonas Pfeiffer, Aishwarya Kamath, Andreas Rücklé, Kyunghyun Cho, Iryna Gurevych in ”AdapterFusion: Non-Destructive Task Composition for Transfer Learning”. arXiv:2005.00247 [cs.CL]

[25] Luigi Di Caro and Luigi Di Caro in ”Common-Sense Knowledge for Natural Language Understanding: Experiments in Unsupervised and Supervised Settings”.

[26] Hannah Rashkin, Maarten Sap, Emily Allaway, Noah A. Smith, Yejin Choi,”Event2Mind in Commonsense Inference on Events, Intents, and Reactions ”. arXiv:1805.06939 [cs.CL]

[27] Simon Ostermann, Sheng Zhang, Michael Roth. peter clark in ”Com- monsense Inference in Natural Language Processing (COIN) ”

[28] Hao Zhou, Tom Yang, Minlie Huang, Haizhou Zhao, Jingfang Xu, and Xiaoyan Zhu. 2018 in ” Commonsense knowledge aware conversation generation with graph attention. In IJCAI.”

[29] K. P. Scannell, “The Crúbadán project: Corpus building for under- resourced languages,” in Building and Exploring Web Corpora: Proceed- ings of the 3rd Web as Corpus Workshop, vol. 4, (Louvain-la-Neuve, Belgium), pp. 5–15, 2007.]

[30] Elena Klyachko, Alexey Sorokin, Natalia Krizhanovskaya, Andrew Krizhanovsky, Galina Ryazanskaya in ”LowResourceEval-2019: a shared task on morphological analysis for low-resource languages”. arXiv:2001.11285 [cs.CL]

[31] Daniel Grießhaber, Ngoc ThangVu, JohannesMaucher in ”Low-resource text classification using domain-adversarial learning”