International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137

Volume – 07, Issue – 10, October 2022, PP – 01-15

A Comprehensive Analysis of Deep Learning's Impact on Natural Language Processing

*I. Venkata Dwaraka Srihith[1]
Alliance University

R. Varaprasad[2]
G. Pullaiah College of Engineering and Technology

Y. Rama Mohan[3]
G. Pulla Reddy Engineering College

T. Aditya Sai Srinivas[3]
G. Pulla Reddy Engineering College

Y. Sravanthi[4]
Independent Researcher

Abstract: The application of deep learning (DL) models has become significantly more widespread in recent years, which has facilitated significant development in the field of natural language processing (NLP). Through the enhancement of verbal interactions between humans and computers, NLP makes intelligent robots possible. The need for data-driven automation of semantic analysis has increased in response to recent advancements in processing capability as well as the appearance of enormous quantities of linguistic data. The enormous advancements made by deep learning techniques in the fields of computer vision, automatic speech recognition, and NLP have contributed to the widespread acceptance of data-driven initiatives. In this paper, NLP applications that can benefit from deep learning are classified. It highlights fundamental NLP tasks and applications, in addition to the ways in which deep learning might improve those jobs and applications. In our analysis, we compare various tactics to various modern models.

Index Terms: Deep Learning (DL), Natural Language Processing (NLP), Artificial Intelligence (AI)..

1. Introduction

NLP bridges the gap between computers and natural languages. Processing and comprehension of human languages through the use of computers is what natural language processing, or NLP, is all about [1]. Since the 1980s, the sector has become increasingly reliant on computing that is driven by data[2], [3]. Deep learning is an application of artificial neural networks (ANNs) that makes use of billions of trainable parameters. It is made possible by the growing processing power and parallelization of GPUs [4]. As a result of advances in data collection methods, enormous data sets are now easily accessible, which paves the way for the training of such deep structures[5][6]. Natural language processing is beneficial for comprehending data supplied by humans since it takes into account context-dependent data. Text mining and analysis are both improved by having a better understanding of the context of the data. The communication structures and patterns of humans can be altered through the use of NLP[7].

It is becoming overly reliant on data-driven methods to create new ways for natural language processing, which help in the design of models that are both more effective and more resilient[8]. One of the most alluring approaches to natural language processing, known as deep learning, is now doable as a result of recent developments in computing power and enormous volumes of data[9] and [10]. ANNs have been successfully utilised by NLP scholars and practitioners in recent years, beginning with[11]. In related domains such as computer vision [12] and speech recognition [13], deep learning[14], [15] has already demonstrated superior performance. The old approaches to natural language processing were upgraded to more data-driven methods as a result of these advances. Innovative methods are not only more likely to be successful but also easier to put into practise. The excellent standard of research conducted by IBM, which included the development and application of intricate statistical models, is responsible for a considerable proportion of the early achievements in the field of machine translation.

NLP techniques like part-of-speech tagging [16], named entity classification [17], and semantic role labelling [18] have been performed using deep neural networks. The vast majority of research on natural language processing (NLP) deep learning utilises either supervised or unsupervised learning.

Section 2 presents natural language processing (NLP), artificial intelligence (AI), and deep learning. The third section of this guide covers the basics of natural language processing. Section 4 focuses on the motivation. Section 5 covered the groundwork for natural language processing. Various types of datasets were covered in section 6. In section 7, we looked at some of the ways that NLP has been put to use. The most significant problems in natural language processing were covered in section 8. Section 9: NLP's Prospects Moving Forward and Final Thoughts.

2. Background

Because the ability to comprehend and produce natural language is indicative of intelligence, natural language processing (NLP) is included in the field of artificial intelligence (AI). Since deep learning is a tool for artificial intelligence, we have placed it in that category. Next, we will discuss the reasons why deep learning ought to be used to NLP.

i. Natural Language Processing

The interpretation of human languages through the use of computing models and procedures is what natural language processing, also known as computational linguistics, does. Software is produced by making use of these different solutions. Work in natural language processing (NLP) can be broken down into its fundamental domains and applications, although it isn't always easy to tell where specific challenges fit in. The core areas consist of morphological processing, which segments meaningful components of words and identifies the true parts of speech (POSs) of words as they are used; language modelling, which quantifies associations between naturally occurring words; syntactic processing, also known as parsing, which constructs sentence diagrams as possible precursors to semantic processing; and semantic processing, which derives meaning from words. Information extraction (including named entities and relationships), text translation, summarization, inferring responses to queries, document classification, and clustering are all examples of application fields. People frequently have to deal with one or more primary issues, and they can apply these ideas and strategies to handle difficulties that occur in the actual world. The processing of natural languages requires a combination of machine learning, statistical analysis, and probabilistic calculations. Methods such as naive Bayes, k-nearest neighbours, hidden Markov models, conditional random fields (CRFs), decision trees, random forests, and support vector machines have all been used in the past. In recent years, neural models have either completely replaced traditional methods or significantly improved upon them.

ii. AI and DL

There are "islands of success" in which artificial intelligence analyses significant amounts of data to accomplish operational goals (e.g., fraud detection). Improvements that affect the entire application are something that both scientists and customers want. This demands an understanding of the processes and procedures used by AI (e.g., algorithms). According to Ted Greenwald, artificial intelligence (AI) is defined as anything a computer can accomplish that humans used to do [19]. It is the goal of the powers of AI to expand the capabilities of IT beyond the basic functions of creating, communicating, and storing data to include the transformation of data into knowledge[20]. It is difficult for humans to process all of the data that is now available since the volume of data is growing at such a rapid rate. This leaves us with two choices: either we (1) discard the majority of the already available data or (2) we build artificial intelligence in order to convert the vast amounts of data into knowledge that we can use. AI and data are brought together through deep learning.

Learning a skill or procedure for a specific job by employing deep neural networks is what "deep learning" refers to. It is possible to do everything from straightforward categorization to in-depth reasoning. Deep learning[21] is a set of algorithms that can solve any problem, provided that a sufficient amount of data is provided. Deep learning is a method for solving problems that involves locating and analysing various data structures and attributes. There is a conflict between AI and deep learning. The goal of artificial intelligence (AI) is to eventually exceed human brains. Deep learning helps here.

iii. Neural Networks(NN) and DL

Neurons in neural networks[22] are responsible for processing inputs and producing outputs. Calculations involving weighted sums are carried out on the input data by every node that is a part of the output layers. These calculations are followed by the production of outputs through the use of fundamental nonlinear transformation techniques. Modifications to the weight are brought about by specific faults or losses that occur

at the output nodes. Stochastic gradient descent and back propagation are two error-correction methods used in modern networks[23]. Networks[24] can be differentiated from one another based on the number of tiers and the connectivity between nodes. In a feed forward neural network, the inputs that are sent to each node in a layer come from the nodes in the levels that are below it (FFNNs). There is not a universally accepted definition of what constitutes a DNN; nonetheless, most people agree that deep neural networks have multiple hidden layers[6].

1. **Multi Layer Perceptron (MLP):** There are three levels to an MLP (input, hidden, and output layers). A cluster of neurons that is responsible for converting information between layers is referred to as a layer. In MLP, layer neurons don't communicate. MLPs make advantage of nonlinear activation in their processes. A fully linked network is the consequence of every node in each layer having a connection to the following layer (Fig. 1). MLPs are forward-feeding neural networks. Functional neural networks (FNNs) are neural networks in which node connections don't form a cycle.

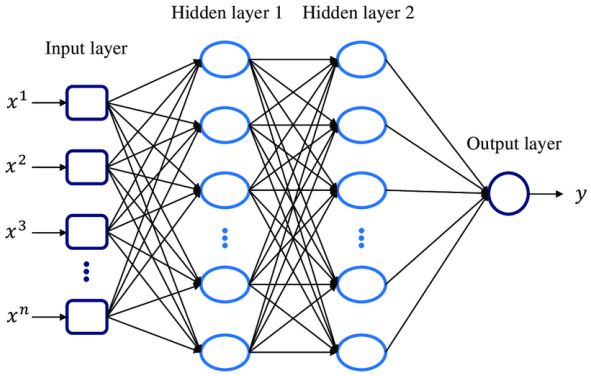

Fig.1 MLP Architecture

2. **Convolutional Neural Networks(CNN):** Fukashima's neocognitron is used to construct convolutional neural networks (CNNs)[25], [26][27]. Filters are utilised by CNNs in order to perform simultaneous data analysis[28]. CNNs are used in image, video, audio, and NLP applications [29]. It is not crucial necessary where traits exist; rather, what is important is whether or not they do exist. Pooling helps to cut down on the overall size of feature maps (the outputs of the convolutional filters). The loss of precision is mitigated by using smaller pools.
3. **Recurrent Neural Network (RNN):**The output of one feed-forward neural network (FNN) is fed into the next FNN in order to build a recurrent neural network (RNN). RNNs have three types of layers: input, hidden, and output, just like FNNs. In discrete time frames, input vector sequences are supplied one at a time; for instance, after inputting each batch of vectors, performing some operations, and updating the network weights, the next batch of vectors is supplied. This continues until all of the input vector sequences have been processed. Figure 3 demonstrates how we use the hidden layer's parameter values to produce predictions at each and every time step.
4. **Long Short-Term Memory Networks (LSTM):**The LSTM is a sophisticated RNN [30]. Information can be stored, forgotten, or exposed through the production of recursive nodes in LSTMs by connected neurons. The information stored in the memories of generic RNNs with single neurons feeding back to themselves is gradually lost with each cycle. There are times when it is important to recall things from a long time ago, but it is not important to remember recent events[31]. LSTM blocks enable individuals to remember critical information while simultaneously forgetting unimportant specifics. Standard LSTMs are outperformed by the gated recurrent unit (GRU), which is a simpler variation of the LSTM[32].
5. **Recursive Neural Networks:** Weight sharing is employed by recursive networks like CNN in order to minimise the need for training. Recursive networks, in contrast to CNN, distribute their weights in a vertical fashion (between layers). Because of this, modelling parse trees is much easier. Weights are able to be incorporated into a single tensor in recursive networks (or generalised matrix)[33].
6. **Generative Adversarial Networks (GAN):**In 1941, Good fellow was the first to market GANs. A neural network discriminator and a neural network generator are brought together to form a GAN (Fig. 5). Iterative processes are utilised during the training of the network. To begin, the network will provide an inaccurate sample. The discriminator network is responsible for determining whether or not a sample (such as an input image) comes from the genuine training data (data used to develop the model). The purpose of the generator is to trick the discriminator into believing that the bogus samples it produces are authentic. Different GANs,

including Sim GAN [34], Wasserstein GAN [35], info GAN [36], and DC GAN [37], have been introduced. Generated fake celebrity faces are almost as good as the real thing, thanks to one of the most complicated GAN implementations. GANs have showed astonishing results in a variety of settings, which has sparked curiosity in their potential use. In natural language processing, GANs are put to use to generate text[38].

3. Steps in NLP

Generally speaking, there are five basic steps:

a. Lexical analysis: is characterised by the process of recognising and studying the structure of individual words. Essentially, it involves breaking up the entire division of the text into passages, phrases, and individual words.
b. Syntactic analysis: often known as "parsing," entails conducting an examination of the words in the sentence to determine their grammatical correctness and then rearranging the words that serve as examples to illustrate the connection between the words.
c. Semantic analysis: This method determines the precise meaning of the text, often known as the dictionary meaning. The meaningfulness of the text is evaluated based on degree to which syntactic structures and objects in the task domain correspond to one another in a satisfactory manner.
d. Integration of discourse: the meaning of every given sentence is dependent on the meaning of the sentence that came before it and the sentence that immediately followed it.
e. Pragmatic analysis: this type of study entails deriving those characteristics of language that require knowledge of the real world. During this, what was said is rethought in light of what it was intended to mean all along.

4. Motivation

Deep learning applications[39] are dependent on feature representations, algorithmic structures, and computer architecture. A Structure for the Representation and Acquisition of Knowledge Surprisingly, the information that is considered significant for a task and the representation that produces effective results for data representation are not the same thing. Linguists believe that the study of semantics, grammatical structure, and context all play an important role in the understanding of sentiment. Studies that used the bag-of-words (BoW) method in the past have demonstrated satisfactory results[40]. The bag-of-words method [41] simply accounts for the frequency with which terms appear. BoW does not take into account word order or how words interact with one another and instead approaches each word individually. BoW ignores syntax yet delivers decent performance for syntax-dependent applications. This indicates that fundamental representations can perform just as well as, if not better than, with massive amounts of data. These findings lend credence to the effectiveness of deep learning methodologies and systems. Language modelling typically drives NLP discoveries. The objective of statistical language modelling is to provide a probabilistic definition of word sequences, which is a difficult task because of the dimension[42]this researcher contributed to neural network language modelling by developing a distributed word representation and a probability function for sequences..In contrast to computer vision, NLP research focuses on fundamental issues such as the representation of language through the use of statistical models. NLP programmes represent texts like documents. In order to accomplish this, one must first learn the features of the data or extract meaningful information from raw data. The tedious handcrafting of features is the starting point for traditional approaches, subsequently being followed by the creation of algorithms that can extract and make use of such traits. Deep supervised learning methods are data-driven and have the ability to construct accurate representations of the data being studied.

The presence of unlabeled data makes unsupervised feature learning an essential component of natural language processing. Learning high-dimensional data from unlabeled data using unsupervised feature learning reflects that data. Both K-means clustering and PCA come in handy at times. Deep learning and the use of unlabeled data are prerequisites for representation learning, which is the first step toward natural language processing (NLP). The vast majority of jobs in natural language processing rely on labelled data, whereas unstructured data motivates study. Because deep learning[15] can assist with natural language processing, it is essential to investigate its methods and structures while keeping NLP in mind.

5. Deep Learning In The Fundamentals Of NLP

Any type of computational linguistic system will inevitably face fundamental obstacles. In order to translate, summarise, or caption photos, you need to have a solid knowledge of the language that is being used. Language modelling, morphology, parsing, and semantic analysis are the components that make up this comprehension.

There are two schools of thought when it comes to language modelling. First, it determines word order. Words on their own have very little value; rather, they get all of their meaning from the way they interact with one another and other words. The study of morphology focuses on the construction of words. It does this through demonstrating tense, gender, and other linguistic concepts through the use of word roots, prefixes, suffixes, compounds, and other intra word techniques. The process of parsing a sentence involves determining which words modify which others and generating the constituent parts of sentences. The study of the meanings of words is referred to as semantics. It investigates the ways in which words are connected to one another and the ways in which they influence one another, in addition to the context and "common sense."

I. **Language Modeling:** The most essential component of natural language processing is known as language modelling. Language modelling is indispensable for NLP. The process of language modelling forecasts future words or linguistic features based on those that have come before[43]. Applications in which the user types can take advantage of this feature to provide predictive text entering. Because of its capacity to capture syntactic and semantic correlations between words or components in a linear neighbourhood, it is useful for both machine translation and text summarization. This is due to the fact that syntactic and semantic linkages can be found in linear neighbourhoods. Specifically, it is excellent for translating text automatically. These machines are able to generate phrases that have a human-like tone by using prediction.

   a. **Neural Language Modelling:** The statistical language models were incapable of dealing with synonyms or other types of terms that were not included in the training corpus. The neural language model has seen some levels of development with positive results[44]. Early on, members of the language modelling community created intricate models with the assistance of ANNs. Many of these models were reported by[45].
   b. **Analysis of Language Models:** It might be challenging to assess the effectiveness of neural network advancements in language modelling. Examining models of language is important to do regardless of the applications they have. There is no such thing as a perfect measuring tool[46].The inverse probability of a test set, normalised by the number of words in the set, is the perplexity of the test. When many vocabularies are utilised to train language models, there is reduced likelihood of major confusion occurring. We are fortunate to have access to a large number of data sets that can be used as benchmarks. These kinds of data collections include, for instance, the Penn Treebank (also known as PTB) and the Billion Word Benchmark.
   c. **CNN in Language Modeling:** In CNNs, the use of fully linked layers was substituted for pooling layers for the purpose of language modelling [38]. These layers, which reduce the dimensionality of feature maps in a manner analogous to that of pooling layers, Pooled layers lose this information, but fully linked layers keep part of the information identifying the position of such features. A multilayer perceptron convolutional neural network (MLPConv) equipped with tiny MLP filters[47], a multilayer CNN (ML-CNN) with stacked convolutional layers, as well as a hybrid of these networks referred to as COM, whose filter kernel sizes can vary (in this case, they were three and five). Stacking convolutional layers had a negative impact on language modelling, although MLPConv and COM were able to reduce complexity. The results were significantly improved by combining the kernel sizes of MLPConv and COM. Linguistic patterns such as "as..." as were among those that could be learned by the networks. The findings of this study showed that CNNs are capable of identifying long-term sentence dependencies. Although words that were close by were the most vital, words that were further away were still quite important.
   d. **Latest Issues and improvements:** Language modelling has evolved weekly since [48] and[49]. To capture backward context, [49]added bidirectionality to Embeddings from Language Models (ELMo). In addition to this, they collected vectorizations at a number of different layers. Empirical performance was improved when the same information was encoded in many ways.

II. Morphology

Morphology is the study of the roots, stems, prefixes, suffixes, and infixes of words. Affixes modify gender, number, person, etc. stems.

[50] models linguistic morphology. Modeling morphology with RvNN RvNN was then topped with a neural language model. [51] was used to segment the WordSim-353 model. One model used context, one didn't. The context-insensitive model over accounted for morphological structures. Even antonyms with the same stem were clustered. Context-sensitive models performed better, noting stem connections and other factors like "un." The new model outperformed earlier embedding models on various common data sets [52].

Many NLP tasks require a strong morphological analyzer. [53] studied how neural machine translation (NMT) models learn and exploit morphology. Several translation models from French, German, Czech, Arabic, or Hebrew were created. LSTM-based (some with attention mechanisms) or character-aware CNNs were trained on WIT3 [54]. Decoders were replaced with POS and morphological taggers, with encoder weights fixed to preserve internal representations. Encoder and decoder training effects were explored. The study found that focus decreases encoder performance but improves decoder performance. Character-aware models are better for learning morphology, and the output language influences encoder performance. Encoders create inferior representations of morphologically rich output languages.

Besides universal morphology, morphological embeddings could improve multi language processing. They could be used across cognate languages, which is useful when some are better resourced. Morphological structures may be important for interpreting specialist language like biological publications. Since natural language processing (NLP) often uses deep learning, a better way to handle morphological components could make models work better.

III.Parsing

Parsing studies how sentences' words and phrases link. Parsing might be by constituency or dependence [43]. Hierarchical parsing extracts phrasal elements from a sentence. Dependency parsing examines word relationships.

The most recent application of deep learning parsing has been in dependency parsing, which faces another significant challenge in terms of solution division. Graph-based parsing involves the construction of parse trees and the subsequent search for the appropriate one. The vast majority of graph-based approaches are generative models that make use of a formal grammar that is based on natural language [43]. Techniques that are based on transitions and generate one parse tree have gained greater popularity than graph-based ones in recent years. A buffer containing all of the words in the sentence is produced by transition-based dependency parsing, and a stack containing only the ROOT label is produced. Following that, arcs will be used to connect the top two words. Following the identification of dependencies, the words are popped. The operation continues until the buffer is empty and just the ROOT label remains. Each of the aforementioned activities is governed by three primary strategies. In arc-standard [55], Words' dependents are connected to a word before the word that the word depends on. Whether or not their own offspring are connected, arc-eager words are connected to their parents as rapidly as feasible. In swap-lazy [56], the arc-standard technique allows stack position shifting. Non projective edges can be graphed.

1) Early Neural Parsing:[57]RNNs equipped with probabilistic context-free grammars were utilised (PCFGs). [58]The neural model was the first to achieve state-of-the-art performance when it comes to parsing. This performance was attained on the PTB for both labelled and unlabeled attachment scores by utilising an inside-out recursive neural network. This network used two vector representations (an inner and an outer) to allow for top-down and bottom-up data flows. Using these two vector representations, the network was able to allow top-down and bottom-up data flows. Treebank [59] versus the Wall Street Journal portion of the PTB [54]), showing that neural models can generalise across domains. Stenetorp [60] introduced dependency embeddings. The RNN created a directed acyclic graph. This model's performance on the Wall Street Journal part of the CoNLL 2008 Shared Task data set [61] was within 2% of the best, but it had trouble remembering phrases from the beginning of a sentence near the end.
2) Transition-Based Dependency Parsing:[62] improved UAS and LAS on English and Chinese PTB data. They did this with a transition-based parser and a basic FFNN. So, they subverted the sparsity problem in statistical models. [63]'s work was improved by the addition of residual connections and a perceptron layer, which came after the soft max layer. By employing pretraining [64], in which possible data samples are transmitted to two different parsers, they were able to gain knowledge from a far larger number of examples than is typical.
3) Generative Dependency and Constituent Parsing:[65]paradigm for parsing and modelling that makes use of RNN grammars was proposed. The full text as well as the currently constructed parse tree are both required inputs for this top-down parser. This made it possible to take into consideration not just the local terms but also the complete sentence. This model performed exceptionally well when it came to generative parsing of English and modelling single sentences. It produced a performance that is almost on par with the best in Chinese generative parsing. Combining two parsers explicitly was preferred over using one to produce candidate trees and another to rank them. Two parsers were employed to choose and rerank candidates, providing cutting-edge results. Using three parsers improved this model. At the time, the best PTB results came from a group of eight models that used two parsers.

IV. Semantics

The interpretation of words, phrases, sentences, and documents is what the field of processing semantics is all about. Word embeddings like Word2Vec and GloVe, which are based on the distributional theory of meaning, confidently capture the meanings of individual words [66]. The processing of vectors that are equivalent to phrases, sentences, or other text components by a neural network results in the generation of a compositional semantic representation. In this part, research on semantic processing in the brain is broken down into two categories: determining the degree to which the meanings of two different pieces of text are comparable, and recording and transmitting the meanings of phrases.

1) **Semantic Comparison:** Check to see if two identical phrases, sentences, or documents that were determined by humans to have the same meaning may also be deemed to have the same meaning by computer programmes. [66] provided a comparison between two CNNs with regard to semantics. [67] had an effect on the initial model, which was called ARC-I. It was a Siamese network with two CNNs that traded weights with each other. In the second network, linkages made it possible to share information before the CNNs reached their final states. The method performed significantly better than earlier English and Chinese models.

   [68] developed a Bi-CNN-MI (MI for multigranular interaction features) using a CNN sentence model, a CNN interaction model, and a logistic regressor. Using dynamic CNNs, they changed a Siamese network. Each level's feature maps were compared, not just the top-level maps. They produced state-of-the-art MSRP results [69].

   [70] contrasted feature maps utilising a CNN's "similarity measurement layer," fully connected layer, and log-softmax output layer. Convolutional layers employed 1-to-4-pixel windows. The network was trained on MSRP, SICK, and MSRVID. First and third were state-of-the-art.

2) **Sentence Modeling:** Sentence modelling extends neural language modelling to vectorize sentence meaning. [71] model paragraphs or bigger texts in this way. Produced phrase representations using a dynamic convolutional neural network (DCNN). When applied to sentences with varied structures, dynamic pooling is able to identify features of varying types and lengths without resorting to padding. This led to both short-term and long-term dependence being discovered. The performance of DCNN was assessed using semantic tasks. When it came to predicting the sentiment of movie reviews [72] and Twitter [73], it outperformed every other comparison algorithm. It was among the best when it came to organising the TREC questions [74].

6. Datasets

Many researchers in the academic world make use of benchmark datasets like the ones that are described below. When it comes to machine learning, benchmarking is an examination of the pattern-learning capacity of different strategies and algorithms. The process of benchmarking involves comparing a newly developed method or practise to others that came before it. There are three kinds of benchmark datasets.

1) Experiment data from the real world

2) Synthetic data often mimics real-world trends. Synthetic data replaces genuine data. When more data is needed than is available or when privacy is very important, like in healthcare, these kinds of datasets are very important.

3) This type is demonstration and visualisation datasets. Typically, they're artificially manufactured; real-world data patterns aren't needed.

Data is needed to teach Deep Learning about pattern recognition. Model accuracy depends on data quality. Despite the effectiveness of universal language modelling techniques like BERT[75], they can only be used to pre-train models. After that, the model is trained using the task's data. To meet daily needs in machine fields like NLP, developing new datasets is vital.

Creating fresh datasets is difficult. The newly constructed dataset should be accurate, sufficient for evaluation, and good for training. "What is proper and accurate data?" is an application-based question. Data quality and quantity determine whether there is enough information.

Asking "what are we trying to do and what problem do we need to solve?" is a good place to start. "What kinds of data do we require, and how much of them?" Next, develop a training and examination plan. A model can learn how to identify input-output links through the use of a training data set. The test data set is used to evaluate the machine's intelligence by determining how well the trained model does on test samples that it has not previously seen. The next step is to organise the data in a way that is clear to experts in the field. After that comes the issue of access and ownership of the data. It's possible that you'll need specialised approval to distribute confidential or private information.

7. Deep learning NLP applications

The study of NLP is essential for gaining an understanding of how brain models function; however, from an engineering point of view, which prioritises the development of solutions that benefit humanity over philosophical and scientific inquiry, NLP is of no value in and of itself. Here are some current NLP approaches. These concerns only pertain to written language, not spoken communication. Speech processing [76] is distinct from natural language processing in that it calls for expertise in a variety of other areas, most notably acoustic processing.

A. Information Retrieval

IR systems help users locate the correct information at the right time [77]. Ranking documents according to their relevance scores for ad hoc retrieval tasks like those performed by a search engine is one of the more challenging aspects of IR. Deep learning algorithms match query strings to document texts to score relevancy. Thus, such models must produce representations of query-document interactions. Some representation-focused approaches generate good text representations using deep learning models and then match them [66]. Interaction-focused approaches first build local interactions directly and then use DNNs to learn how two texts match based on word interactions [78]. Finding how each word in the query links to segments of the document is important for matching a big document to a brief query.

B. Information Extraction

Extraction of information from text. System outputs vary, but extracted data and relationships are often stored in relational databases [79]. People often pull out named entities and their relationships, events and their participants, information about time, and fact tuples. Named entity recognition identifies proper nouns, dates, hours, prices, and product IDs. [11]used a multitasking technique but didn't provide results. A basic feed forward network with a fixed-size window surrounding each word was used. This makes it impossible to capture word relationships, presumably.

C. Event Extraction:

Event extraction identifies words or phrases that refer to events, including agents, objects, recipients, and times. Event extraction involves four subtasks: recognising event mentions (phrases that describe events), event triggers (typically verbs or gerunds), event arguments, and argument roles. According to [80], CNNs that adopt max-pooling may miss significant facts in phrases that relate to many occurrencestion involves four subtasks: recognising event mentions (phrases that describe events), event triggers (typically verbs or gerunds), event arguments, and argument roles.

D. Relationship Extraction:

Relationships are another key textual clue. Possessive, antonymous, or synonymous relationships, or natural, familial, or geographic.[81] employed a CNN to classify sentence associations. Using only two layers, a 3-pixel window, and 50-dimensional word embeddings, they achieved greater results than before. [82] employed a bidirectional LSTM and CNN to classify relationships and recognise entities. [83] utilised a copy-based attention-based GRU model. This network used a coverage method to ensure all critical information was extracted correctly. [84] used the BERT[75] model with supervised training on a set of biological data to find temporal relationships in clinical data.

E. Text Classification

One further classic use for natural language processing is the categorization of free-text texts into known categories. There are various applications for classification. When classifying sentences, [29] utilised a CNN that was pre-trained with word vectors. According to the findings of Kim's research, straightforward CNNs consisting of one convolutional layer followed by a dense layer featuring dropout and soft max output might produce excellent results with relatively minimal adjustment to the hyper parameters. CNN models performed better on four out of seven sentence classification tasks, including sentiment analysis and the classification of questions. In a later study, [85] shown that convolutional networks perform admirably when it comes to document classification.

F. Text Generation

NLP tasks necessitate human-like language. Texts can be converted from seq2seq format using summarization and machine translation. Other operations, such as the captioning of images and videos, turn data that is not textual into text form. Some exercises produce text without input data (or with only limited amounts

used as a topic or guide) (or with only small amounts used as a topic or guide). Examples include poetry, comedy, and short stories.

1) Poetry Generation: Poetry generation is perhaps the hardest subtask because the content must be given aesthetically, usually following a pattern. Recurrent models are conventional for literary tasks. Recurrent networks are great at learning internal language models, but they don't have any style and their output isn't organized.

[86] trained by utilising specific poets and controlling for style in Chinese poetry. With enough training data, they found sufficient results. [87] solved the structure problem by training the network on a single sort of poem to produce rhythmic poetry with a single structure. Human evaluators found the poetry to be lower quality than human-produced poems but indistinguishable.

2) Joke and Pun Generation: Deep learning joke and pun generation has received little attention. Using a tiny LSTM, [88] constructed homographic puns. The network used multiple-meaning terms to create ambiguous sentences but failed to make the puns funny. Most machine-generated jokes are classified as such by human examiners. The authors said that training on pun data alone isn't enough to generate decent puns. [89] trained an LSTM on Conan O'Brien jokes to generate jokes. Since many of these jokes are current, the network was trained on news stories. This gave the jokes context. [90] used a neural network to generate jokes, quotations, and tweets. Providing general language skills and non joke examples improved the funny quality.

G. Machine Translation

NLP's signature application is machine translation. Documents are translated using mathematical and algorithmic procedures. Effective translation is difficult, even for humans, as it necessitates knowledge of morphology, syntax, and semantics, as well as cultural sensitivity for both languages (and societies) [43]. [91] was the first to attempt NMT, although neural models had been employed for the comparable problem of transliteration [92]employed a feed forward network with seven-word inputs and outputs. Encoder–decoder models allow for sentence translation

A recent experiment [93] shown that a single neural network that is both basic and vast can be taught to convert up to 12 separate languages. The network is able to automatically discern the language of the source from which it is receiving data; all it needs is an input token in order to decide the language that will be used as the target. The model is able to produce mixed outputs when it is presented with a large number of language tokens, some of which may be in languages that are related to the selected languages but are not necessarily those languages. There is a chance that deep neural networks (DNNs) would be able to acquire universal representations for information regardless of the language that the information is expressed in, in addition to etymology and linkages between language families.

8. Major Challenges in NLP

Natural language is the most natural means of communication for humans, regardless of whether it is spoken, written, or typed. Given the prevalence of natural language use among humans, it is not inconceivable that it could be adapted for use in dialogue with artificial intelligence. The goal of natural language processing (NLP) is to develop technology that can use natural language as effectively as individuals. There is currently no software that can disambiguate words as well as a person can. NLP gives computers the ability to communicate in human-like ways by mimicking human speech and language. Even people who aren't computer programmers can easily and successfully communicate with computers. When this objective is accomplished, computers will have the capacity to comprehend, investigate, summarise, translate, and produce accurate human writing and language, which is the most natural form of human communication. The linguistic hierarchy is taken into account by natural language processing (NLP). According to the NLP definition, a phrase is a collection of words, and a sentence is a series of stages. As a result, goal computers will require capabilities related to natural language processing, which will present NLP systems with a significant challenge.

a. Interactive dialogue:

Since the 1980s, Natural Language Processing has placed a significant emphasis on research pertaining to dialogue. Through the use of interactive dialogue, people are able to manage computer systems, come to decisions using natural language, and solve problems using natural language. A few examples of applications include database access, command and control, production control, office automation, logistics, and computer-assisted instructions. Interactions between humans and machines need to be natural, uncomplicated, and support multiple modes. The scope of research on text-based dialogue has been broadened to include spoken dialogue on mobile devices for the purposes of information access and task-based applications. SDS have only met with

moderate levels of success in open domain interactions, in which users are free to talk about any subject. Other concerns with the SDS include locating and re-creating activities that involve natural human dialogue.

b. Machine Translation:

Machine translation transforms text from one natural language to another while maintaining its meaning and delivering grammatically correct content in the target language. Computers should be able to perceive numerous languages, output multiple languages, and translate across languages despite linguistic uncertainty. The ability to analyse and compose sentences in human languages, in addition to having knowledge and context comparable to that of a person, are all requirements for accurate translation. Computers must be able to comprehend numerous languages, output multiple languages, and translate across languages. To establish their intended meaning and develop a response, it is necessary to communicate and comprehend words and phrases. Science, diplomacy, international trade, and intelligence are all potential applications.

The vast majority of the NLP techniques and systems are only available for so-called "high resource languages" (HRLs), which include the likes of English, French, and German. There are millions of people who speak and write LRLs like Indonesian and Swahili, but these languages do not have access to the same resources or systems. The NLP community will make the creation of information and tools for tens of thousands, if not hundreds of thousands, of different languages one of its top priorities in the near future.

c. Sentence Generation:

It is not simple to develop models that generate grammatical, comprehensible statements.

d. Reading and Writing Text:

Reading and writing text is tough in NLP. By reading and understanding presented material, machines can become intelligent, integrate, and summarise information for people. This implies computers can analyse and comprehend data. Intelligence, logistics, office automation, and library management use it. Since the birth of the modern online world, we have access to a massive online archive of human-encoded material. Practically all scientific discoveries are written in human language.

Scientists can't keep up with new findings. As a result, machine reading is in demand for comprehending and summarising material and extracting facts and hypotheses. Machine reading must design question-answer systems so individuals can access knowledge bases and get answers.

9. Future of NLP

A. Invisible User Interface

When the user interface (UI) of a feature appears so natural, it is referred to as an invisible UI. Sliding doors are a real-world example; no button or signal is required, and the solution is concealed.

The desire for invisible user interfaces increases as software and the Internet grow more valuable. This level of intuition enables the user to complete their objective more quickly and easily, resulting in their satisfaction. An invisible user interface demands a direct connection between user and machine. NLP increases context comprehension of human language. Important for programmes with an invisible user interface, it learns how we speak and what we do.

B. Bots

Chatbots are able to interpret human input. This is achieved by combining machine learning with NLP. Siri, Clever bot, and Crotona are well-known conversation bots. In customer service, chatbots eliminate the waiting period by addressing consumers' questions and directing them to the appropriate resources. Quick, intelligent, and user-friendly chatbots are useful. Using NLP, chatbots interpret text or voice conversations.

The major issue of modern chatbots is their inability to comprehend human speech. They misunderstand or disregard what you type. Even more embarrassingly, chatbots that attempt to comprehend every word you say can fail to do so. Consequently, natural language processing research is mainly focused on chatbots (NLP).

C. Smart Search:

Additionally, NLP will improve search engines. The same capabilities that allow a Chat bot to interpret a customer's request may also enable "search while you chat."

10. Conclusion

NLP is currently seeing widespread adoption. The machine translators developed by Google and Microsoft can convert text from one language into dozens of others. In addition, a wide variety of devices can read voice commands and reply appropriately. These intricate applications, particularly when used in operational environments, stand as a testament to the remarkable progress that has been made in this area over the course of the past 60 years. There is no doubt that a lot of progress has been made in recent years. The astonishing progress that has been made in ANNs can undoubtedly be attributed to this latest development. These "old" machine learning frameworks have paved the way for remarkable progress, shattering previous performance records in a variety of tasks spanning a wide range of industries. Deep neural networks have improved models' "imperfect" natural language ability. The assessments of the models have shed light on a few general patterns that have emerged. In the past, advancements in the state of the art have been made thanks to contributions from convolutional models as well as recurrent models. Nevertheless, using stacks of attention-powered transformer units as encoders and decoders has shown better results in the NLP sector. These models have had significant pretraining on general language knowledge, but their training on specific tasks, whether unsupervised or supervised, has been very modest. Second, the most effective encoder-decoder links can be found in attention strategies that do not involve recurrences or convolutions. Third, encouraging networks to study a wide variety of characteristics improves the outcomes. Finally, although highly constructed networks typically produce the best possible results, there is no substitute for massive quantities of high-quality data. Pretraining on large generic corpora, however, does appear to be helpful in this regard. In light of this conclusive finding, it could be more beneficial to investigate additional methods for preparing people for training rather than developing highly specialised components for complex models.

References:

[1]. D. W. Otter, J. R. Medina, and J. K. Kalita, “A survey of the usages of deep learning for natural language processing,” *IEEE Trans. neural networks Learn. Syst.*, vol. 32, no. 2, pp. 604–624, 2020.

[2]. E. D. Liddy, “Natural language processing in Encyclopedia of Library and Information Science 2nd ed., New York: Marcel Decker.” Inc, 2001.

[3]. K. S. Jones, “Natural language processing: a historical review,” *Curr. issues Comput. Linguist. honour Don Walk.*, pp. 3–16, 1994.

[4]. I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[5]. Y. LeCun, Y. Bengio, G. Hinton, and others, “Deep learning. nature, 521 (7553), 436-444,” *Google Sch. Google Sch. Cross Ref Cross Ref*, 2015.

[6]. J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural networks*, vol. 61, pp. 85–117, 2015.

[7]. A. Torfi, R. A. Shirvani, Y. Keneshloo, N. Tavaf, and E. A. Fox, “Natural language processing advancements by deep learning: A survey,” *arXiv Prepr. arXiv2003.01200*, 2020.

[8]. X. Zhang, J. Zhao, and Y. LeCun, “Character-level convolutional networks for text classification,” *Adv. Neural Inf. Process. Syst.*, vol. 28, 2015.

[9]. K. Cho *et al.*, “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” *arXiv Prepr. arXiv1406.1078*, 2014.

[10]. R. Collobert and J. Weston, “A unified architecture for natural language processing: Deep neural networks with multitask learning,” in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 160–167.

[11]. R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, “Natural language processing (almost) from scratch,” *J. Mach. Learn. Res.*, vol. 12, no. ARTICLE, pp. 2493–2537, 2011.

[12]. A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, “Large-scale video classification with convolutional neural networks,” in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.

[13]. A. Graves and N. Jaitly, “Towards end-to-end speech recognition with recurrent neural networks,” in *International conference on machine learning*, 2014, pp. 1764–1772.

[14]. M. Gheisari, G. Wang, and M. Z. A. Bhuiyan, “A Survey on Deep Learning in Big Data,” *Proc. - 2017 IEEE Int. Conf. Comput. Sci. Eng. IEEE/IFIP Int. Conf. Embed. Ubiquitous Comput. CSE EUC 2017*, vol. 2, no. 06, pp. 173–180, 2017.

[15]. G. Mahalaxmi, T. Tirupal, T. Srinivas, and D. Raziya, “Categorization of Leaf Ailments Using Deep Learning Techniques: A Review.,” *IUP J. Telecommun.*, vol. 14, no. 1, 2022.

[16]. C. Dos Santos and B. Zadrozny, “Learning character-level representations for part-of-speech tagging,” in *International Conference on Machine Learning*, 2014, pp. 1818–1826.

[17]. G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural architectures for named entity recognition," *arXiv Prepr. arXiv1603.01360*, 2016.

[18]. J. Zhou and W. Xu, "End-to-end learning of semantic role labeling using recurrent neural networks," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2015, pp. 1127–1137.

[19]. T. Greenwald, "Artificial intelligence (A special report)--what exactly is artificial intelligence, anyway? Everybody's talking about AI these days. here's what all the fuss is about," *Wall Str. Journal, available https//proxy. cecybrary. com/login*, 2018.

[20]. U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of Big Data challenges and analytical methods," *J. Bus. Res.*, vol. 70, pp. 263–286, 2017.

[21]. T. A. S. Srinivas, A. S. Priya, and B. S. Priya, "A Comprehensive Survey of Big Data in the Age of AI," vol. 13, no. 06, pp. 922–931, 2022.

[22]. N. Sudhakar Yadav, S. Ramasubbareddy, and M. Ravikanth, "Neural Network-Based Activity Recognition System," in *Innovations in Computer Science and Engineering*, Springer, 2022, pp. 511–518.

[23]. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," 1985.

[24]. S. Sankar, S. Ramasubbareddy, A. K. Luhach, W. S. alnumay, and P. Chatterjee, "NCCLA: new caledonian crow learning algorithm based cluster head selection for Internet of Things in smart cities," *J. Ambient Intell. Humaniz. Comput.*, 2022.

[25]. A. Sarabu and A. K. Santra, "Distinct two-stream convolutional networks for human action recognition in videos using segment-based temporal modeling," *Data*, vol. 5, no. 4, p. 104, 2020.

[26]. A. Sarabu and A. K. Santra, "Human action recognition in videos using convolution long short-term memory network with spatio-temporal networks," *Emerg. Sci. J.*, vol. 5, no. 1, pp. 25–33, 2021.

[27]. Y. LeCun *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.

[28]. Y. LeCun, Y. Bengio, and others, "Convolutional networks for images, speech, and time series," *Handb. brain theory neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[29]. Y. Zhang and B. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," *arXiv Prepr. arXiv1510.03820*, 2015.

[30]. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[31]. R. Varaprasad, S. Ramasubbareddy, and K. Govinda, "Event Recommendation System Using Machine Learning Techniques," in *Innovations in Computer Science and Engineering*, Springer, 2022, pp. 627–634.

[32]. K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv Prepr. arXiv1409.1259*, 2014.

[33]. R. Socher, E. Huang, J. Pennin, C. D. Manning, and A. Ng, "Dynamic pooling and unfolding recursive autoencoders for paraphrase detection," *Adv. Neural Inf. Process. Syst.*, vol. 24, 2011.

[34]. A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2107–2116.

[35]. M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*, 2017, pp. 214–223.

[36]. X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," *Adv. Neural Inf. Process. Syst.*, vol. 29, 2016.

[37]. A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv Prepr. arXiv1511.06434*, 2015.

[38]. L. Yu, W. Zhang, J. Wang, and Y. Yu, "Seqgan: Sequence generative adversarial nets with policy gradient," in *Proceedings of the AAAI conference on artificial intelligence*, 2017, vol. 31, no. 1.

[39]. T. A. S. Srinivas, G. Mahalaxmi, R. Varaprasad, and D. Raziya, "A Comprehensive Survey of Techniques, Applications, and Challenges in Deep Learning: A Revolution in Machine Learning."

[40]. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," *arXiv Prepr. cs/0205070*, 2002.

[41]. Z. S. Harris, "Distributional Structure, WORD, vol. 10, no. 2--3." August, 1954.

[42]. Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model. journal of machine learning research, Vol. 3, No." Feb, 2003.

[43]. D. S. McNamara, L. Allen, S. Crossley, M. Dascalu, and C. A. Perret, "Natural language processing and learning analytics," *Handb. Learn. Anal.*, vol. 93, 2017.

[44]. Y. Bengio, R. Ducharme, and P. Vincent, "A neural probabilistic language model," *Adv. Neural Inf. Process. Syst.*, vol. 13, 2000.

[45]. W. De Mulder, S. Bethard, and M.-F. Moens, "A survey on the application of recurrent neural networks to statistical language modeling," *Comput. Speech \& Lang.*, vol. 30, no. 1, pp. 61–98, 2015.

[46]. R. Iyer, M. Ostendorf, and M. Meteer, "Analyzing and predicting language model improvements," in *1997 IEEE Workshop on Automatic Speech Recognition and Understanding Proceedings*, 1997, pp. 254–261.

[47]. M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv Prepr. arXiv1312.4400*, 2013.

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

[49]. J. Sarzynska-Wawer *et al.*, "Detecting formal thought disorder by deep contextualized word representations," *Psychiatry Res.*, vol. 304, p. 114135, 2021.

[50]. M.-T. Luong, R. Socher, and C. D. Manning, "Better word representations with recursive neural networks for morphology," in *Proceedings of the seventeenth conference on computational natural language learning*, 2013, pp. 104–113.

[51]. M. Creutz and K. Lagus, "Unsupervised models for morpheme segmentation and morphology learning," *ACM Trans. Speech Lang. Process.*, vol. 4, no. 1, pp. 1–34, 2007.

[52]. G. A. Miller and W. G. Charles, "Contextual correlates of semantic similarity," *Lang. Cogn. Process.*, vol. 6, no. 1, pp. 1–28, 1991.

[53]. Y. Belinkov, N. Durrani, F. Dalvi, H. Sajjad, and J. Glass, "What do neural machine translation models learn about morphology?," *arXiv Prepr. arXiv1704.03471*, 2017.

[54]. M. Cettolo, C. Girardi, and M. Federico, "Wit3: Web inventory of transcribed and translated talks," in *Conference of european association for machine translation*, 2012, pp. 261–268.

[55]. J. Nivre, "An efficient algorithm for projective dependency parsing," in *Proceedings of the eighth international conference on parsing technologies*, 2003, pp. 149–160.

[56]. J. Nivre, M. Kuhlmann, and J. Hall, "An improved oracle for dependency parsing with online reordering," in *Proceedings of the 11th International Conference on Parsing Technologies (IWPT'09)*, 2009, pp. 73–76.

[57]. R. Socher *et al.*, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proceedings of the 2013 conference on empirical methods in natural language processing*, 2013, pp. 1631–1642.

[58]. P. Le and W. Zuidema, "The inside-outside recursive neural network model for dependency parsing," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 729–739.

[59]. J. Judge, A. Cahill, and J. Van Genabith, "Questionbank: Creating a corpus of parse-annotated questions," 2006.

[60]. P. Stenetorp, "Transition-based Dependency Parsing Using Recursive Neural Networks," *Nips*, pp. 1–9, 2013.

[61]. M. Surdeanu, R. Johansson, A. Meyers, L. Màrquez, and J. Nivre, "The CoNLL 2008 shared task on joint parsing of syntactic and semantic dependencies," in *CoNLL 2008: Proceedings of the Twelfth Conference on Computational Natural Language Learning*, 2008, pp. 159–177.

[62]. D. Chen and C. D. Manning, "A fast and accurate dependency parser using neural networks," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 740–750.

[63]. D. Weiss, C. Alberti, M. Collins, and S. Petrov, "Structured training for neural network transition-based parsing," *arXiv Prepr. arXiv1506.06158*, 2015.

[64]. Z. Li, M. Zhang, and W. Chen, "Ambiguity-aware ensemble training for semi-supervised dependency parsing," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2014, pp. 457–467.

[65]. J. R. Brennan, C. Dyer, A. Kuncoro, and J. T. Hale, "Localizing syntactic predictions using recurrent neural network grammars," *Neuropsychologia*, vol. 146, p. 107479, 2020.

[66]. B. Hu, Z. Lu, H. Li, and Q. Chen, "Convolutional neural network architectures for matching natural language sentences," *Adv. Neural Inf. Process. Syst.*, vol. 27, 2014.

[67]. A. Bordes, X. Glorot, J. Weston, and Y. Bengio, "A semantic matching energy function for learning with multi-relational data," *Mach. Learn.*, vol. 94, no. 2, pp. 233–259, 2014.

[68]. H. T. Nguyen, P. H. Duong, and E. Cambria, "Learning short-text semantic similarity with word embeddings and external knowledge sources," *Knowledge-Based Syst.*, vol. 182, p. 104842, 2019.

[69]. W. Dolan, C. Quirk, C. Brockett, and B. Dolan, "Unsupervised construction of large paraphrase corpora: Exploiting massively parallel news sources," 2004.

[70]. H. He, K. Gimpel, and J. Lin, "Multi-perspective sentence similarity modeling with convolutional neural networks," in *Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015, pp. 1576–1586.

[71]. Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *International conference on machine learning*, 2014, pp. 1188–1196.

[72]. R. Socher, J. Bauer, C. D. Manning, and A. Y. Ng, "Parsing with compositional vector grammars," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2013, pp. 455–465.

[73]. A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," *CS224N Proj. report, Stanford*, vol. 1, no. 12, p. 2009, 2009.

[74]. X. Li and D. Roth, "Learning question classifiers," in *COLING 2002: The 19th International Conference on Computational Linguistics*, 2002.

[75]. C. Sun, L. Huang, and X. Qiu, "Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence," *arXiv Prepr. arXiv1903.09588*, 2019.

[76]. G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, 2012.

[77]. A. A. Gulhane, "A Study of Deep Learning and NLP," *J. homepage www. ijrpr. com ISSN*, vol. 2582, p. 7421.

[78]. Z. Lu and H. Li, "A deep architecture for matching short texts," *Adv. Neural Inf. Process. Syst.*, vol. 26, 2013.

[79]. J. Cowie and W. Lehnert, "Information extraction," *Commun. ACM*, vol. 39, no. 1, pp. 80–91, 1996.

[80]. Y. Chen, L. Xu, K. Liu, D. Zeng, and J. Zhao, "Event extraction via dynamic multi-pooling convolutional neural networks," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2015, pp. 167–176.

[81]. D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao, "Relation classification via convolutional deep neural network," in *Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers*, 2014, pp. 2335–2344.

[82]. S. Zheng *et al.*, "Joint entity and relation extraction based on a hybrid neural network," *Neurocomputing*, vol. 257, pp. 59–66, 2017.

[83]. Z. Tu, Z. Lu, Y. Liu, X. Liu, and H. Li, "Modeling coverage for neural machine translation," *arXiv Prepr. arXiv1601.04811*, 2016.

[84]. C. Lin, T. Miller, D. Dligach, S. Bethard, and G. Savova, "A BERT-based universal model for both within-and cross-sentence clinical temporal relation extraction," in *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, 2019, pp. 65–71.

[85]. A. Conneau, H. Schwenk, L. Barrault, and Y. Lecun, "Very deep convolutional networks for text classification," *arXiv Prepr. arXiv1606.01781*, 2016.

[86]. J. Wei, Q. Zhou, and Y. Cai, "Poet-based poetry generation: Controlling personal style with recurrent neural networks," in *2018 International Conference on Computing, Networking and Communications (ICNC)*, 2018, pp. 156–160.

[87]. J. Hopkins and D. Kiela, "Automatically generating rhythmic verse with neural networks," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2017, pp. 168–178.

[88]. Z. Yu, J. Tan, and X. Wan, "A neural approach to pun generation," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 1650–1660.

[89]. H. Ren and Q. Yang, "Neural joke generation," *Final Proj. Reports Course CS224n*, 2017.

[90]. B. Chippada and S. Saha, "Knowledge amalgam: Generating jokes and quotes together," *arXiv Prepr. arXiv1806.04387*, 2018.

[91]. H. Schwenk, "Continuous space translation models for phrase-based statistical machine translation," in *Proceedings of COLING 2012: Posters*, 2012, pp. 1071–1080.

[92]. T. Deselaers, S. Hasan, O. Bender, and H. Ney, "A deep learning approach to machine transliteration," in *Proceedings of the Fourth Workshop on Statistical Machine Translation*, 2009, pp. 233–241.

[93]. M. Johnson *et al.*, "Google's multilingual neural machine translation system: Enabling zero-shot translation," *Trans. Assoc. Comput. Linguist.*, vol. 5, pp. 339–351, 2017.

