NLP Review: Architectures, Techniques, Applications and Challenges

Ankit Sirmorya
ML Engineering Manager, Amazon

Sowmyashree Ramesh Kumar
Senior Data Scientist, Microsoft

Mehul Vishal Sadh
University of Texas at Arlington

ABSTRACT

The natural language processing (NLP) field entails the application of a broad range of computational approaches to the automatic analysis and representation of human language. It's a field of artificial intelligence in which computers analyze, understand, and derive meaning information from human language in a smart and useful way. A large percentage of NLP applications are used to organize and structure knowledge in order to perform tasks such as automatic summarization, translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, and topic segmentation. The paper goes through different NLP architectures that can perform such tasks. Various architectures have been discussed in detail, such as CNN, RNN, LSTM, and GRU. Additionally, we cover NLP Techniques such as Morphological Analysis, Semantic Analysis, Sentiment Analysis, Keyword Extraction, Stemming, and Lemmatization. There are also several limitations of this methodology. Almost every industry uses NLP. NLP plays a major role in many fields like Health care, Information Retrieval, and Web mining. We finally gave a brief review on different NLP topics and future research.

Keywords

Deep Learning, Machine Learning

1. INTRODUCTION

Natural language processing (NLP) refers to computer systems that analyze, attempt to understand, or produce one or more human languages, such as English, Japanese, Italian, or Russian[1]. It is a computerized approach to analyzing text that is based on both a set of theories and a set of technologies.[2] suggests a definition of Natural Language Processing as "A theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications." Naturally occurring texts can be of any language, mode, genre, etc. The texts can be oral or written. The only requirement is that they be in a language used by humans to communicate with one another. Also, the text being analyzed should not be specifically constructed for the purpose of the analysis, but rather that the text be gathered from actual usage. [2]

2. HISTORY OF NLP

The development of Natural Language Processing is divided into 4 phases.[3] The late 1940s to 1960s, 1960s to 1970s, 1970s to 1980s, and 1980s onwards. The work of the first phase was focused on machine translation (MT). . Following a few early birds, including Booth and Richens' investigations and Weaver's influential memorandum on translation of 1949 (Locke and Booth, 1955), research on NLP began in earnest in the 1950s. Automatic translation from Russian to English, in a very rudimentary form and limited experiment, was exhibited in the film-Georgetown Demonstration of 1954. [3]The Georgetown experiment in 1954 involved the fully automatic translation of more than sixty Russian sentences into English. The authors claimed that within three or five years, machine translation would be a solved problem.[4]It is extremely difficult to define how we would ever know that a system actually "understands" language. For testing whether a system appears to understand language by successfully performing its task. The Turing test (q.v.), proposed by Turing (1950) (reprinted in Computers and Thought (1963)), has been the classical model.[1]. In this test, the system must be indistinguishable from a human when both answer arbitrary interrogation by a human over a terminal. Some notably successful natural language processing systems developed in the 1960s were SHRDLU, a natural language system working in restricted "block worlds" with restricted vocabularies, and ELIZA, a simulation of a "Rogerian psychotherapist" written by Joseph Weizenbaum between 1964 and 1966.[5]During the 1970s, many programmers began to write "conceptual ontologies", which structured real-world information into computer-understandable data. Examples are MARGIE (Schank, 1975), SAM (Cullingford, 1978), PAM (Wilensky, 1978), TaleSpin (Meehan, 1976), QUALM (Lehnert, 1977), Politics (Carbonell, 1979), and Plot Units (Lehnert 1981). During this time, the first many chatterbots were written.[5]

3. NLP ARCHITECTURES

Natural language processing (NLP) has benefited greatly from the resurgence of deep neural networks (DNNs), due to their high performance with less need for engineered features. There are two main DNN architectures: convolutional neural network (CNN)[31] and recurrent neural network (RNN) [32]. In order to overcome the limitations of the basic RNN, gating mechanisms have been developed, resulting in two types of RNN: long short-term memory (LSTM)and gated recurrent units (GRU).[33]

3.1 Convolution Neural Network(CNN)

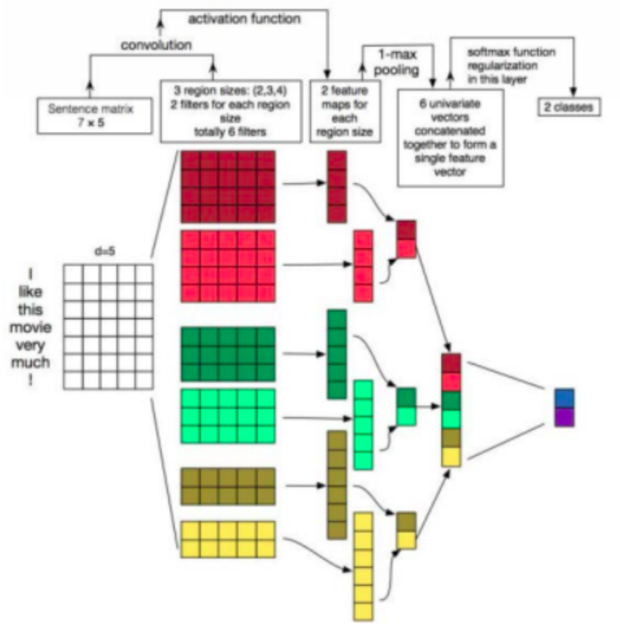

Fig. 1. CNN model

In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analyzing visual imagery. Most recently, however, Convolutional Neural Networks have also found prevalence in tackling problems associated with NLP tasks like Sentence Classification, Text Classification, Sentiment Analysis, Text Summarization, Machine Translation, and Answer Relations. The use of CNNs for sentence modeling traces back to Collobert and Weston [29]. This work used multi-task learning to output multiple predictions for NLP tasks such as POS tags, chunks, named-entity tags, semantic roles, semantically-similar words, and a language model. The input layer of the CNN in NLP has a sequence x with n entries (mostly n-words). The input then goes through the Convolution layer and then finally through the Max Pooling layer. Contextual information in a sentence is significant to disambiguate the meaning of words. For instance, the following two sentences "I slept deeply at night." and "I slept deeply halfway through this movie." have the same phrase "slept deeply" which should be recognized differently in terms of sentimental meaning. While CNN's hierarchical layers can deal with low-to-high level information, it has limitations in capturing the contextual meaning of words within the whole sentence because its architecture relies on the feed-forward hierarchical path, particularly in the inference stage[34]. CNN-based framework to solve a large number of NLP tasks. CNN has shown the power of computer vision tasks and people are beginning to believe CNN's performance. CNN can extract salient N-Grams features from the input sentence and create a sentence representation task for the latent information semantics. [54]
As shown in the Figure 1[53], a text vector is passed as input to the (7x5) sentence matrix. The convolution is performed at the input vector level. Filters are applied to arbitrary word size windows to produce new features. These feature vectors are then downsampled using the pooling layer. In the last second layer, 6 vectors are concatenated to form a fully connected vector. This is then applied to a softmax function to finally create an output layer of 2 classes.CNNs can also be used in classification tasks. In literature [55], Kim used CNN in a sentence structure for the classification task, including emotional and subjective classification. After a specific task is trained, the randomly initialized convolution kernel becomes an N-Grams feature detector that is specific to the target task.

3.2 Recurrent neural networks (RNN)

A Recurrent Neural Network(RNN) is a deep learning model that not only depends on the immediately previous value but the whole sequence before the current stage.RNN is particularly well suited for solving problems in which the sequence is of greater importance than the individual elements themselves.RNNs are ideal for solving problems where the sequence is more important than the individual items themselves. An RNN is essentially a fully connected neural network that contains a refactoring of some of its layers into a loop. That loop is typically an iteration over the addition or concatenation of two inputs, matrix multiplication, and a nonlinear function.[35] RNN is mostly used in the context of speech processing and natural language processing [36, 37]. RNNs use sequential data in their network. The embedded structure in the data sequence provides valuable information, so this feature is essential to a range of applications. Understanding the context of a sentence, for example, is crucial to understanding the meaning of a particular word within it. Convolutional Neural Networks (CNNs) offer advantages in selecting good features and Long Short-Term Memory (LSTM) networks have proven good abilities to learn sequential data.[48]Among the text usages, RNNs perform well at Sequence labeling, Natural Language Processing (NLP) text classification, Natural Language Processing (NLP) text generation.[52]

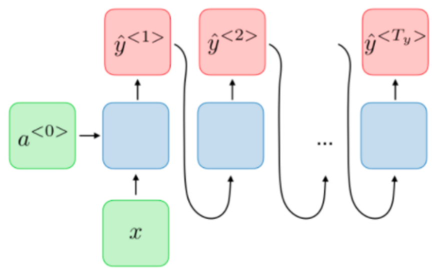

Fig. 2. RNN model

Figure 2[53] represents a One to Many RNN, it takes one input and can produce multiple outputs. It is widely used in tasks such as writing an essay on a given topic or generating a poem giving a category. As the number of layers of RNNs increases, the loss landscape can become impossible to train, this is the vanishing gradient problem. To solve this problem a Gated Recurrent Unit (GRU) or a Long Term Short Term Memory (LSTM) network can be used.[52]

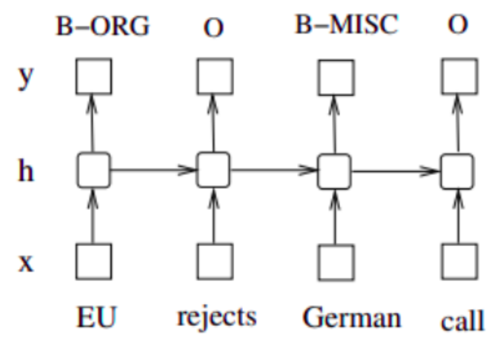

Fig. 3. RNN model for a text document

RNN can also be used in a named entity recognition system. Figure 3[70] depicts a model that has an input x and produces an output y. In the named entity tagging context, x represents input features and y represents tags. Figure 4 illustrates a named entity recognition system in which each word is tagged with other (O) or one of four entity types: Person (PER), Location (LOC), Organization (ORG), and Miscellaneous (MISC). The sentence "EU rejects German call to boycott British lamb" is tagged as B-ORG O B-MISC O O O B-MISC O O, where B-, I- tags indicate beginning and intermediate positions of entities [57]. An input layer represents features at time t. It might be a one-hot encoding for word features, a dense vector feature, or a sparse feature. An input layer is dimensionally equivalent to the feature size. An output layer represents a probability distribution over labels at time t [57].

3.3 Long Short Term Memory(LSTM)

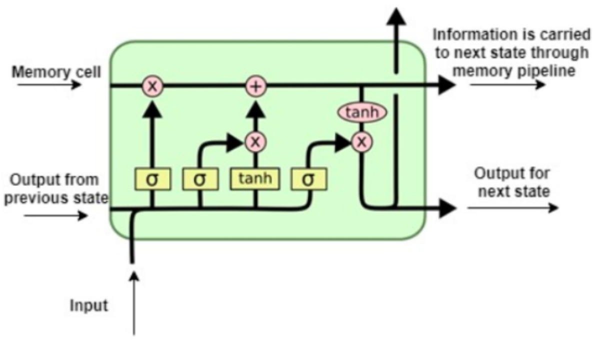

Fig. 4. LSTM cell

Natural Language Processing (NLP) is an important subfield within Machine Learning, and various deep learning architectures and preprocessing techniques have led to many improvements. Long-short-term memory (LSTM) is the most well-known architecture for time series and textual data. Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. LSTMs are often used for time-series data as they are equipped with an internal memory that can remember important data from previous time steps. They can also decide which data to ignore or forget when the model determines they are not material to its output.[47] It can process not only single data points (such as images) but also entire sequences of data (such as speech or video). For example, LSTM is applicable to tasks such as unsegmented, connected handwriting recognition, speech recognition, and anomaly detection in network traffic or IDSs (intrusion detection systems).[38]. LSTM networks attempt to improve on the methods of naive RNN models, but we know that LSTM structures still have many limitations. They do not work well enough when the number of words, or the variability in word count between texts, is high. [47]

Figure 4[71] shows a simple LSTM unit. It has a memory cell at the top which helps to carry the information from a particular time instance to the next time instance in an efficient manner. So, it is able to remember a lot of information from previous states when compared to RNN and overcomes the vanishing gradient problem. Information might be added or removed from the memory cell with the help of valves.[56]LSTM network is fed by input data from the current time instance and output of hidden layers from the previous time instance. These two data passes through various activation functions and valves in the network before reaching the output.[56]

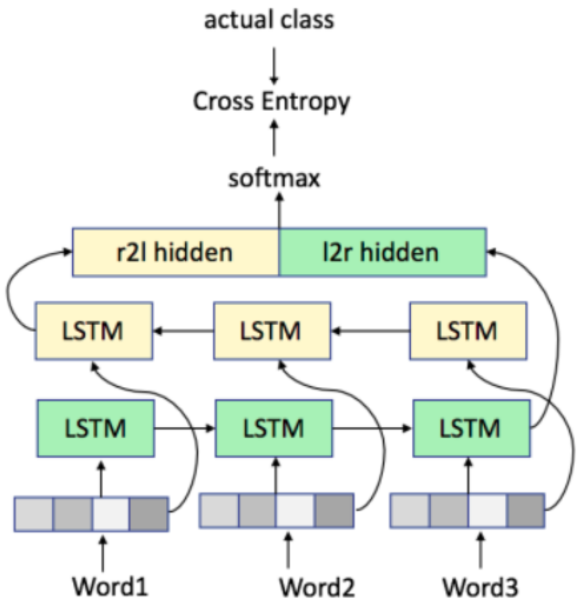

Fig. 5. Bidirectional LSTM model

Figure 5[58] shows a bidirectional LSTM Model for text classification.Each word is transformed into its dense embedding representation and is then being fed into two LSTM networks: 1) left-to-right LSTM network, and 2) right-to-left LSTM network. The outputs of the two networks (the LSTM hidden states) are further concatenated and the classification is performed. When applying this approach to text, a model learns from target word's prefix (left-to-right) and suffix (right-to left) contexts and performs the final classification based on the jointly learned representation of this context.[58]

3.4 Gated Recurrent Unit (GRU)

Traditional RNN architectures suffer from the problem of vanishing and exploding gradients [51], rendering optimization difficult and prohibiting them from learning long-term dependencies. GRU is one of the solutions presented for the problem. GRU or Gated recurrent unit is an advancement of the standard RNN i.e recurrent neural network. It was introduced by Kyunghyun Cho et al[49] in the year 2014. The input and output structure of GRU is similar to ordinary RNN, but its internal structure is similar to LSTM[50]

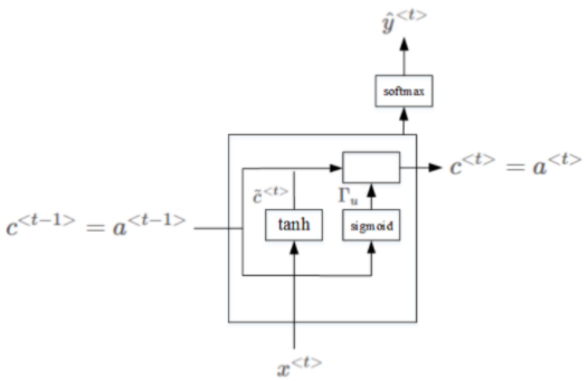

Fig. 6. CNN model

Figure 6[53] represents a GRU Cell.GRU consists of an additional memory unit commonly referred to as an update gate or a reset gate. Apart from the usual neural unit with sigmoid function and softmax for output, it contains an additional unit with tanh as an activation function. Tanh is used since its output can be both positive and negative hence can be used for both scaling up and down. The output from this unit is then combined with the activation input to update the value of the memory cell.[53]

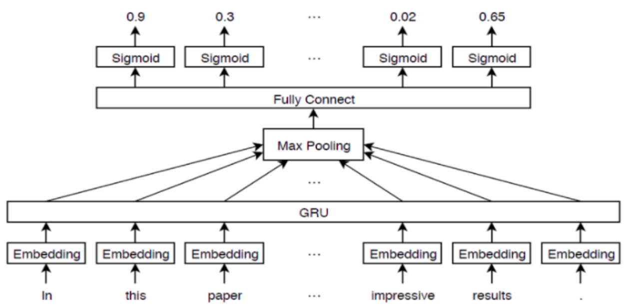

Fig. 7. GRU cell model

Figure 7[59] shows how to classify text data that has multiple independent labels. For classification tasks where there can be multiple independent labels for each observation—for example, tags on a scientific article—you can train a deep learning model to predict probabilities for each independent class. To enable a network to learn multilabel classification targets, you can optimize the loss of each class independently using binary cross-entropy loss [59]. The input is a word embedding that maps a sequence of words to a sequence of numeric vectors. Then a GRU operation is applied that learns dependencies between the embedding vectors. Followed by a max pooling operation that reduces a sequence of feature vectors to a single feature vector. A fully connected layer that maps the features to the binary outputs. Is put on top of it and finally a sigmoid operation for learning the binary cross entropy loss between the outputs and the target labels is present at the top and it produces the probability of each word to be present in the particular class. This probabilities are independent and thus may not sum to 1.[59]

4. NLP TECHNIQUES

4.1 Morphological Analysis

Natural Language Processing includes morphological analysis among its initial methods. It is the process of identifying morphemes(the smallest lexical element of a word) from which a given word is constructed, for example, “foxes” is derived from “fox”. Morphology is mainly useful for identifying the parts of speech in a sentence and words that interact together.[21] Morphology is a systematic description of words in a natural language. It describes relationships between lexical forms and words’ surface forms. Words are analyzed into their surface form, which is their spoken or graphical form, and then lexical form, which is the analysis of their lemmas (also known as their dictionary forms) and their grammatical descriptions.[21]

4.2 Lexical semantic analysis

Natural language processing (NLP) is a popular application for artificial intelligence, helping machines to communicate in human languages. Among various tasks, sentiment analysis is the basis of many NLP problems. This involves creating a model that learns the semantic value of each word and classifying a text according to the overall mood valence.[47] Lexical semantics is a form of Semantic Analysis. It examines the meanings of words individually. Lexical semantics is concerned with inherent aspects of word meaning and the semantic relations between words, as well as the ways in which word meaning is related to syntactic structure. It includes words, sub-words, affixes (sub-units), compound words, and phrases also. Lexical items include all the words, sub-words, etc. As a result, we can define lexical semantics as the relationship between lexical items, the meaning of sentences, and the syntax of sentences.[22] Lexical Semantic Analysis involves the classification of lexical items like words, sub-words, affixes, followed by the decomposition of lexical items like words, sub-words, affixes. Differences, as well as similarities between various lexical-semantic structures, are also analyzed. The aim of lexical analysis is in Data Cleaning and Feature Extraction with the help of techniques such as Stemming, Lemmatization, and correction of Spellings.

4.3 Syntactic Analysis

Syntactic analysis or parsing or syntax analysis is the third phase of NLP. The purpose of this phrase is to draw exact meaning, or you can say dictionary meaning from the text. Syntax analysis checks the text for meaningfulness compared to the rules of formal grammar. For example, a sentence like “hot ice cream” would be rejected by the semantic analyzer.[23] Syntactic Analysis assigns a semantic structure to the text, i,e, gives a logical meaning to the text. It applies grammar rules to a group of words or a whole sentence to provide it with meaning. The rules are not applicable to an individual word. Semantic Analysis aims at finding the role played by a word in a sentence. Interpret the relationship between words, Interpret the grammatical structure of sentences. Syntactic analysis [24] gets at the structure or grammar of the sentences. Processing a sentence syntactically involves determining the subject and predicate and the place of nouns, verbs, pronouns, etc. Given a lexicon telling the computer the part of speech for a word, the computer would be able to just read through the input sentence word by word and in the end, produce a structural description[25]

4.4 Pragmatic Analysis

An analysis of the pragmatics of a text is part of the process of extracting information from it. More specifically, it is the part of the course that deals with analyzing a structured set of text and finding what the meaning is. Pragmatics is "the analysis of the real meaning of an utterance in a human language, by disambiguating and contextualizing the utterance". This is accomplished by identifying ambiguities encountered by the system and resolving them using one or more types of disambiguation techniques. [21]Pragmatic analysis helps users to discover this intended effect by applying a set of rules that characterize cooperative dialogues, for e.g., "close the window?" should be interpreted as a request instead of an order.

4.5 Sentiment Analysis

Sentiment analysis or opinion mining is the computational study of people's opinions, sentiments, emotions, appraisals, and attitudes towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes [60].t. Sentiment analysis provides companies with a means to estimate the extent of product acceptance and to determine strategies to improve product quality. It also facilitates policy makers or politicians to analyze public sentiments with respect to policies, public services or political issues.[61]The architecture of Sentiment analysis has 5 major steps, a corpus, a document processing unit, document analysis unit,lexicon and linguistic resources and finally sentiment scores for entities and aspects. The input to the system is a corpus of documents in any format (PDF, HTML, XML, Word, among others). The documents in this corpus are converted to text and are preprocessed using a variety of linguistic tools such as stemming, tokenization, part of speech tagging, entity extraction, and relation extraction. The system may also utilize a set of lexicons and linguistic resources. The main component of the system is the document analysis module, which utilizes the linguistic resources to annotate the pre-processed documents with sentiment annotations. The annotations may be attached to whole documents (for document-based sentiment), to individual sentences (for sentence based sentiment) or to specific aspects of entities (for aspect-based sentiment). These annotations are the output of the system and they may be presented to the user using a variety of visualization tools. [62]

4.6 Stemming and Lemmatization

Stemming is a procedure to reduce all words with the same stem to a common form whereas lemmatization removes inflectional endings and returns the base or dictionary form of a word.[63]Stemming has been the most widely applied morphological technique for information retrieval. With stemming, the searcher does not need to worry about the correct truncation point of search keys. Stemming also reduces the total number of distinct index entries [64]. Hull [65] also found that stemming is always useful with short queries. With short queries and short documents, a derivational stemmer is most useful, but with longer ones the derivational stemmer brings in more non relevant documents. Lemmatization is another normalization technique: for each inflected word form in a document or request, its basic form, the lemma, is identified. The benefits of lemmatization are the same as in stemming.[64] In addition, when basic word forms are used, the searcher may match an exact search key to an exact index key. Such accuracy is not possible with truncated, ambiguous stems. Homographic word forms cause ambiguity (and precision) problems – this may also occur with inflectional word forms [66].

4.7 Keyword Extraction

Keywords, which we define as a sequence of one or more words, provide a compact representation of a document's content.Keywords are widely used to define queries within information retrieval (IR) systems as they are easy to define, revise, remember, and share[67]. The automatic keyword extraction is treated as a supervised machine learning task, an approach first proposed by Turney [68]. Treating the automatic keyword extraction as a supervised machine learning task means that a classifier is trained by using documents with known keywords. The trained model is subsequently applied to documents for which no keywords are assigned: each defined term from these documents is classified either as a keyword or a non-keyword; or—if a probabilistic model is used—the probability of the defined term being a keyword is given.[69]

5. LIMITATIONS OF DEEP LEARNING AND THEIR MITIGATIONS

5.1 Ambiguity

For decades, ambiguity has been a challenging issue for NLP researchers. In spite of some results on resolving ambiguity problems that have been obtained, a number of important research problems have not been solved yet [19]. The input presented by the user can be ambiguous in its tone and the intentions of the input. What is said can differ in meaning and connotation. . Ambiguity has been a critical issue for human-computer interaction because of its pervasiveness in everyday life, yet its emergent nature challenges the role of design.[17]Several types of ambiguity have been identified. These include structural, syntactical, form class, word sense, and local ambiguity. Structural Ambiguity occurs when one can devise multiple parse trees for a statement. Syntactic ambiguity is a grammatical ambiguity of a whole sentence that occurs in the sub-part of a sentence. It is a grammatical construct and results from the difficulty of applying universal grammatical laws to a sentence structure.[17]

5.2 Colloquialisms and slang

Informal phrases, expressions, idioms, and culture-specific lingo present a number of problems for NLP – especially for models intended for broad use. Because as a formal language, colloquialisms may have no "dictionary definition" at all, and these expressions may even have different meanings in different geographical areas. Furthermore, cultural slang is constantly morphing and expanding, so new words pop up every day.[20] Humans can understand the connotation and sarcasm in the speech of a person, but that is not the case with machines. Machines fail to understand sarcasm and can take the literal meaning of a particular sentence.

6. APPLICATIONS OF NATURAL LANGUAGE PROCESSING

6.1 Health Care

The health care system has a large number of documents and medical reports. The data stored is unstructured data and thus Natural Language Processing plays an important role in processing these documents. Electronic Medical Records (EMR) are the e-version of these records and use NLP for preprocessing. Although research in the medical informatics field only recently gained popularity, researchers have applied NLP techniques to medical vocabulary for some time.[6]The Specialist system is being developed at the National Library of Medicine (NLM) [12, 8-11] and is intended to function as an information-extraction tool for biomedical knowledge bases, and the Medline abstracts in particular. The Linguistic String Project-Medical Language Processor (LSP-MLP) of New York University is the first (and up till now the longest-lasting) large-scale project about NLP in medicine [6]. The LSP-MLP aims at enabling physicians to extract and summarize sign/symptom information, drug dosage, and response data, to identify possible side effects of medications, and to highlight or flag data items [13]

6.2 Industrial Applications

Natural Language Processing has been used in various domains. For instance, the first search engine allowing one to search a database in natural language, SMART, appeared in the early 1960s [14], as did the first systems able to converse with humans, the most well-known example being ELIZA, which simulated a "Rogerian psychotherapist" [15]. Applications such as "semantic" enterprise search engines, document categorizers, speech recognizers, and – last but not least – conversational agents, also known as virtual assistants or "chatbots" have paved their way in different industrial applications.[16]

6.3 Question Answering Models

With the aid of Natural Language Processing, one can construct a question answering system that, when given a question, can analyze the text and present an appropriate answer. Question and answering (QA) is a system that is able to automatically understand and build answers to human user questions which are posed in natural language.[17] Nowadays there is multiple software that uses Natural Language Processing for answering their queries. The NLP model is able to extract information from the text, analyze the question and think for a suitable answer. The QA system is expected to allow users to ask questions in an "everyday language". Answers may be long or short, they may be lists or narrative. They may vary with the intended use and intended user [18]To answer the question, the system must be able to analyze the question first. Questions can be distinguished by their type; factual, opinion, or summary. If the question is in the context of ongoing interaction, the answer must be consulted with the online resource. Normally, the answer is presented in some kind of form [18].

6.4 Information Retrieval

Information is an abstract and ethereal entity. Providing timely access to relevant information has always been difficult and since the explosion in the volume of information especially in recent decades, effective access to information has now become a critical task. In order for human beings to communicate information to each other and to record it, we use either formal artificial languages like computer programming languages or mathematical logic, or more commonly, we use natural language.[41]Many Natural Language Processing (NLP) techniques, including stemming, part-of-speech tagging, compound recognition, de-compounding, chunking, word sense disambiguation, and others, have been used in Information Retrieval (IR).[40]The impact of NLP on information retrieval tasks has largely been one of promise rather than substance. The relationship between NLP and information management based on content has not been quite as symbiotic compared to NLP and Machine Translation. The content-based manipulation operations we refer to include indexing and retrieval, categorization, classification, filtering, and so on. If we broadly define information retrieval to be the retrieval of textual information based on its content then we see that NLP tools and techniques do not have very much impact on our current generation of information retrieval systems. [39]The simplest applications of NLP to information retrieval have been at the word level by indexing based on some normalized or derived form of individual words occurring in the input.[41]

6.5 Web Mining

With the huge amount of information available online, the World Wide Web is a fertile area for data mining research. Web mining research is at the crossroads of research from several research communities, such as database, information retrieval, and within AI, especially the sub-areas of machine learning and natural language processing.[42]Some have claimed that resource or document discovery (IR) on the Web is an instance of Web (content) mining and others associate Web mining with intelligent Ilk. Actually II1. is the automatic retrieval of all relevant documents while at the same time retrieving as few of the non-relevant documents as possible [43]. Information Extraction(IE) has the goal of transforming a collection of documents, usually with the help of an IR system, into information that is more readily digested and analyzed [44]Building IE systems manually is not feasible and scalable for such a dynamic and diverse medium such as Web contents [46]. Due to the nature of the Web, most IE systems focus on specific Web sites to extract. Others use machine learning or data mining techniques to learn the extraction patterns or rules for Web documents semi-automatically or automatically [45]

7. CONCLUSION

Natural Language Processing is a relatively new field but it has seen various evolutions in recent times. NLP employs computational techniques for the purpose of learning, understanding, and producing human language content. Early computational approaches to language research focused on automating the analysis of the linguistic structure of language and developing basic technologies such as machine translation, speech recognition, and speech synthesis.[27]. The abundant volume of natural language text in the connected world, though having a large content of knowledge, is becoming increasingly difficult to disseminate by a human to discover the knowledge in it, specifically within any given time limits. The automated NLP is aimed to do this job effectively and with accuracy like a human does (for a limited amount of text).[26].
The paper summarizes common NLP sub-problems in this extensive area, as well as the historical development of NLP. It also explains a few of the many applications of NLP such as Health Care, Industrial Applications, and the Question Answering Model. Many

of the medical e-documents have been processed using NLP software. Doctors are using NLP-powered chatbots to serve patients efficiently. The paper also goes through the industrial Applications of NLP that provide an automated voice assistant and Document Processor. Many of the Education Fields also are powered by an NLP Question Answering Model.

Along with its applications, there are certain limitations that still need attention. Many NLP models are not capable of understanding human sarcasm and can take the literal meaning of the sentences. They create a set of rules for understanding a language and strictly follow them without considering the overall meaning of the sentences. Many NLP models are trained in a few of the popular languages and are unable to understand other languages. Many of the models are so complex that they cannot be integrated into a device with low computing power.

After providing a brief description of common machine-learning approaches that are being used for diverse NLP sub-problems, we discuss how modern NLP architectures are designed. Certain NLP techniques have become very popular since their advent. Techniques such as morphological analysis and lexical analysis are integral to the development of any NLP application. Syntactic Analysis, Pragmatic Analysis, Information retrieval are the other few techniques that are part of NLP modeling. We have tried to cover all of the aspects of NLP. But there is still scope for further research.

8. REFERENCES

(1) James F. Allen. 2003. Natural language processing. Encyclopedia of Computer Science. John Wiley and Sons Ltd., GBR, 1218–1222.

(2) Elizabeth D. Liddy: Natural Language Processing

(3) Jones, K. S. (1994). Natural Language Processing: A Historical Review. Current Issues in Computational Linguistics: In Honour of Don Walker, 3–16.

(4) Hutchins, J. (2005). ”The history of machine translation in a nutshell” (PDF).

(5) Wikipedia, “Natural Language Processing”, NLP.

(6) Spyns, Peter. ”Natural language processing in medicine: an overview.” Methods of information in medicine 35.04/05 (1996): 285-301.

(7) McCray AT, Sponsler J, Brylawski B, Browne A. The Role of Lexical Knowledge in Biomedical Text Understanding. In: SCAMC 87. IEEE Computer Society Press, 1987: 103-7.

(8) McCray AT. Natural language processing for intelligent information retrieval. In: Nagel J, Smith W, eds.IEEE 1991. Orlando, 1991: 1160-1.

(9) McCray AT. Extending a Natural Language Parser with UMLS knowledge. In: Clayton P, ed. SCAMC 91. McGraw Hill, 1991: 194-8.

(10) McCray AT, Srinivasan S, Browne A. Lexical Methods for Managing Variation in Biomedical Terminologies. In: SCAMC 94. 1994: 235-9.

(11) McCray AT, Razi A. The UMLS Knowledge Source Server. In: Greenes R, Peterson H, Protti D, eds. MEDINFO 95. Edmonton: Healthcare Computing Communications, 1995: 144-7.

(12) McCray AT, Nelson S. The Representation of Meaning in the UMLS. Meth Inform Med 1995; 34: 193-201.

(13) Lyman M, Sager N, Friedman C, Chi E. Computer-structured Narrative in Ambulatory Care: Its Use in Longitudinal Review of Clinical Data. In: SCAMC 85. IEEE, 1985: 82-6

(14) Salton G, Wong A, Yang CS (1975) A vector space model for automatic indexing. Commun ACM 18(11):613. doi:10.1145/361219.361220

(15) Weizenbaum J (1976) Computer power and human reason: from judgment to calculation. W H Freeman and Company, San Francisco, ISBN 0-7167-0463-3

(16) Quarteroni, Silvia. ”Natural language processing for industry.” Informatik-Spektrum 41.2 (2018): 105-112.

(17) Jusoh, Shaidah. ”A STUDY ON NLP APPLICATIONS AND AMBIGUITY PROBLEMS.” Journal of Theoretical Applied Information Technology 96.6 (2018).

(18) Hirschman, L., Gaizauskas, R. (2001). Natural language question answering: the view from here. Natural Language Engineering, 7, 275–300.

(19) Alfawareh, H.M. Jusoh, S. (2011). Resolving ambiguous entity through context knowledge and fuzzy approach. International Journal on Computer Science and Engineering (IJCSE), 3 (1), 410 – 422.

(20) Inés Roldós: “Major Challenges of Natural Language Processing (NLP)”, monkeylearn.com/blog/natural-language-processing-challenges

(21) Reshamwala, Alpa Mishra, Dhirendra Pawar, Prajakta. (2013). REVIEW ON NATURAL LANGUAGE PROCESSING. IRACST – Engineering Science and Technology: An International Journal (ESTIJ). 3. 113-116.

(22) “Language Processing - Semantic Analysis”, semantic$_{a}nalysis$

(23) Natural Language Processing - Syntactic Analysis, syntatic$_{a}nalysis$

(24) Roxana Girju, “Introduction to Syntactic Parsing”, 2004

(25) Mallamma V. Redd, Hanumanthappa M.,“Semantical and Syntactical Analysis of NLP ”

(26) Chowdhary, KR1442. ”Natural language processing.” Fundamentals of artificial intelligence (2020): 603-649.

(27) Hirschberg, Julia, and Christopher D. Manning. ”Advances in natural language processing.” Science 349.6245 (2015): 261-266.

(28) Akash Shastri, “3 neural network architectures you need to know for NLP!”, https://towardsdatascience.com/3-neural-network-architectures-you-need-to-know-for-nlp-5660f11281be

(29) R. Collobert and J. Weston, “A unified architecture for natural language processing: Deep neural networks with multitask learning,” in Proc. 25th Int. Conf. Machine Learning, 2008, pp. 160–167

(30) Yin, Wenpeng, et al. ”Comparative study of CNN and RNN for natural language processing.” arXiv preprint arXiv:1702.01923 (2017).

(31) LeCun, Yann, et al. ”Gradient-based learning applied to document recognition.” Proceedings of the IEEE 86.11 (1998): 2278-2324.

(32) Elman, Jeffrey L. ”Finding structure in time.” Cognitive science 14.2 (1990): 179-211.

(33) Hochreiter, Sepp, and Jürgen Schmidhuber. ”Long short-term memory.” Neural computation 9.8 (1997): 1735-1780.

(34) J. Shin, Y. Kim, S. Yoon and K. Jung, ”Contextual-CNN: A Novel Architecture Capturing Unified Meaning for Sentence Classification,” 2018 IEEE International Conference on Big

Data and Smart Computing (BigComp), 2018, pp. 491-494, doi: 10.1109/BigComp.2018.00079.

(35) Christopher Thomas: “ Recurrent Neural Networks and Natural Language Processing.” , https://towardsdatascience.com/recurrent-neural-networks-and-natural-language-processing-73af640c2aa1

(36) Batur Dinler Ö, Aydin N. An optimal feature parameter set based on gated recurrent unit recurrent neural net- works for speech segment detection. Appl Sci. 2020;10(4):1273.

(37) Jagannatha AN, Yu H. Structured prediction models for RNN based sequence labeling in clinical text. In: Proceed- ings of the conference on empirical methods in natural language processing. conference on empirical methods in natural language processing, vol. 2016, NIH Public Access; 2016. p. 856.

(38) Wikipedia, “Long short term memory”, LSTM

(39) Smeaton, Alan F. ”Using NLP or NLP resources for information retrieval tasks.” Natural language information retrieval. Springer, Dordrecht, 1999. 99-111.

(40) Brants, Thorsten. ”Natural Language Processing in Information Retrieval.” CLIN 111 (2003).

(41) Smeaton, Alan F. ”Progress in the application of natural language processing to information retrieval tasks.” The computer journal 35.3 (1992): 268-278.

(42) Kosala, Raymond, and Hendrik Blockeel. ”Web mining research: A survey.” ACM Sigkdd Explorations Newsletter 2.1 (2000): 1-15.

(43) van Rijsbergen. Information Retrieval. Butterworths, 1979.

(44) J. Cowie and W. Lehnert. Information extraction. Communications of the ACM, 39(1):80-91, 1996.

(45) Kushmerick. Gleaning the web. IBBE Intelligent Systems, 14(2):20-22, 1999.

(46) I. Muslea, S. Minton, and C. Knoblock. Wrapper induction for semistructured, web-based information sources. In Proceedings of the Conference on Automatic Learning and Discovery CONALD-98, 1999

(47) Marzban, Reza, and Christopher Crick. ”Lifting Sequence Length Limitations of NLP Models using Autoencoders.” ICPRAM. 2021.

(48) Alayba, Abdulaziz M., et al. ”A combined CNN and LSTM model for arabic sentiment analysis.” International cross-domain conference for machine learning and knowledge extraction. Springer, Cham, 2018.

(49) Cho, Kyunghyun, et al. ”Learning phrase representations using RNN encoder-decoder for statistical machine translation.” arXiv preprint arXiv:1406.1078 (2014).

(50) Yang, Shudong, Xueying Yu, and Ying Zhou. ”Lstm and gru neural network performance comparison study: Taking yelp review dataset as an example.” 2020 International workshop on electronic communication and artificial intelligence (IWECAI). IEEE, 2020.

(51) Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. Neural Networks, 5(2):157–166, 1994.

(52) Christopher Thomas: “Recurrent Neural Networks and Natural Language Processing”, RNN

(53) vibhor nigam”: “Natural Language Processing: From Basics to using RNN and LSTM”,https://medium.com/analytics-vidhya/natural-language-processing-from-basics-to-using-rnn-and-lstm-ef6779e4ae66

(54) W. Wang and J. Gang, ”Application of Convolutional Neural Network in Natural Language Processing,” 2018 International Conference on Information Systems and Computer Aided Education (ICISCAE), 2018, pp. 64-70, doi: 10.1109/ICISCAE.2018.8666928.

(55) Y. Kim, ”Convolutional neural networks for sentence classification.” ArXiv preprint arXiv: 1408.5882, 2014

(56) Santhosh Kumar T: “Natural Language Processing – Sentiment Analysis using LSTM, https://www.analyticsvidhya.com/blog/2021/06/natural-language-processing-sentiment-analysis-using-lstm/

(57) Huang, Zhiheng, Wei Xu, and Kai Yu. ”Bidirectional LSTM-CRF models for sequence tagging.” arXiv preprint arXiv:1508.01991 (2015).

(58) Makarenkov, Victor Rokach, Lior Shapira, Bracha. (2019). Choosing the Right Word: Using Bidirectional LSTM Tagger for Writing Support Systems.

(59) Mathworks,“Multilabel Text Classification Using Deep Learning”, mathworks.com/help/deeplearning/ug/multilabel-text-classification-using-deep-learning.html

(60) Liu, Li, et al. ”Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis.” The Lancet 385.9966 (2015): 430-440.

(61) Prabowo, Rudy, and Mike Thelwall. ”Sentiment analysis: A combined approach.” Journal of Informetrics 3.2 (2009): 143-157.

(62) Feldman, Ronen. ”Techniques and applications for sentiment analysis.” Communications of the ACM 56.4 (2013): 82-89.

(63) Balakrishnan, Vimala, and Ethel Lloyd-Yemoh. ”Stemming and lemmatization: a comparison of retrieval performances.” (2014): 174-179.

(64) Korenius, Tuomo, et al. ”Stemming and lemmatization in the clustering of finnish text documents.” Proceedings of the thirteenth ACM international conference on Information and knowledge management. 2004.

(65) Hull, D. Stemming algorithms: a case study for detailed evaluation. Journal of the American Society for Information Science, 47, 1 (1996), 70-84

(66) Alkula, R. From plain character strings to meaningful words: Producing better full text databases for inflectional and compounding languages with morphological analysis software. Information Retrieval, 4, (2001), 195-208.

(67) Rose, Stuart, et al. ”Automatic keyword extraction from individual documents.” Text mining: applications and theory 1 (2010): 1-20.

(68) Turney, Peter. (2000). Learning Algorithms for Keyphrase Extraction. Inf. Retr.. 2. 303-336. 10.1023/A:1009976227802.

(69) Hulth, Anette. ”Improved automatic keyword extraction given more linguistic knowledge.” Proceedings of the 2003 conference on Empirical methods in natural language processing. 2003.

(70) Rahul Agarwal: ”NLP Learning Series: Part 3 - Attention, CNN and what not for Text Classification”, RNN

(71) ”Understanding LSTM Networks” LSTM