Global Scientific JOURNALS

GSJ: Volume 9, Issue 7, July 2021, Online: ISSN 2320-9186
www.globalscientificjournal.com

Natural Language Processing Techniques and Problems in Artificial Intelligence

Alina Anjum
Department of Software Engineering
Sir Syed University of Engineering & Technology
Karachi, Pakistan

Kaneez Tahera Batool
Department of Software Engineering
Sir Syed University of Engineering & Technology
Karachi, Pakistan
K_t_batool@hotmail.com

Qamar Sultana
Department of Software Engineering
Sir Syed University of Engineering & Technology
Karachi, Pakistan
qqamarSultana@gmail.com

Abstract

As the world is growing towards digitalization, there is a need to introduce the computers to interact with Humans by understanding the Human Language & thus producing the Outputs based on the Language provided as an input to the system. These abilities have become the hot area of modern research & world is moving towards some success in order the computers are getting enabled to understand the meanings of the raw data provided as a language. Natural Language processing is a field of Artificial intelligence which is enabling the world to get succeeded in the communications between a Human and a computer, but still there are lacks in almost all the techniques that we still do not have the fine solution or application which can 100% guarantee the communication between the two entities i.e., system and the Human. In this paper we will be discussing various tools which we currently have for processing Linguistics and the common problems which still exists in those solutions.

Keywords - *Natural Language Processing, Linguistics Text Mining, Deep Learning, Machine Learning*

I. Introduction

As today we are living in a world of IT and digitalization, there is a vast need of the communication between a Human and a computer system. Natural language processing is continuously enabling the world to gain the success in introducing the computers with Linguistics (Linguistics are the set of words spoken by a Human). NLP is gaining much attention now a day and various research are still in progress to get the fine solution to resolve the problems of NLP. In this research we will be discussing the ways how NLP works and the recent techniques which are being used to acquire NLP and the problems which still exists in those solutions. We will also be discussing some of the current applications of NLP in this paper.

Deep Learning is one of the main NLP technologies used nowadays [2]. This is a function of AI where a computer can have the abilities of a Human neural functions i.e., brain functions of the human in recognizing objects, speech, decision making, translation of languages etc.

Machine learning is making the machine learn on its own like the human brain which keeps learning over the years. Machine needs huge amount of data to be trained upon and learn different features. During trained the algorithm adjusts its parameters using the datasets [7]. NLP and machine learning both are the focused functions of AI field & are working together to get the better results.

NLP also includes the recognition of text from images and then getting the information extracted from that text. The current market need is for filters that can detect spam emails, filters that in spite of the many innovative ways spammers use to create spam emails can stop them from entering inboxes [6] but it can become possible if we have a proper fine solution of extracting information from raw texts. There are certain methodologies have already developed to retrieve the text from the images but still there is a lacking in retrieving meaningful information from that text which again bring into focus the need of NLP & NLU.

We will discuss How NLP works what are the steps included till now in NLP Applications & what are the current applications exists to achieve the same and what are the loop wholes in those applications.

II. NATURAL LANGUAGE PROCESSING

NLP, the Hot function of AI came into existence for the ease of users in this fast-growing world which lets the Computer understand the Human Language. NLP works to understand the meaning of a certain Language. A Language is a combination of Symbols & Rules. Symbols are used to denote any kind of information whereas Symbols are abided together by certain Rules to make a formal Sentence which has some meanings. NLP Techniques focuses to understand the meanings of Symbols & tries to understand the Rules of a certain language to properly get the information from the raw text which is in the form of Sentence.

Natural Language Processing basically can be classified into two parts i.e., Natural Language Understanding and Natural Language Generation which evolves the task to understand and generate the text (Figure 1) [1].

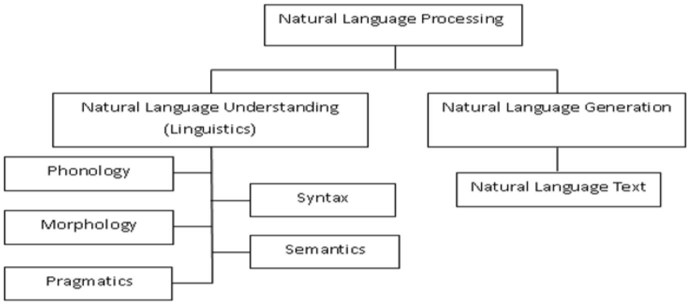
Natural Language Processing
Natural Language Understanding (Linguistics)
Natural Language Generation
Phonology
Morphology
Pragmatics
Syntax
Semantics
Natural Language Text

Figure 1. Broad Classification of NLP

A. *Natural Language Understanding*

NLU is concerned with Linguistics, which is the science of any Language, it comprises of the following five things:

Phonology: we refer this as the sound of the language.
Morphology: the word formation is called the Morphology,
Syntax: Sentence structure is called syntax.
Semantics & Pragmatics: referred as Understanding of the language which is the most important key concern now a days which also includes the emotions and the context of the meaning of the sentence.

B. *NLP Phases*

There are five main Phases currently being used to facilitate the NLP Applications:

1. Morphological Analysis/ lexical Analysis
   Processing words

2. Syntax Analysis/ Parsing
   Parsing of sentences (Parts of speech comes here) includes grammatical analysis of words in a sentence and their arrangement of words. Sentences such as "School goes to boy" are rejected by the English syntactic analyzer [3].

3. Semantic Analysis
   Distinguishing Nouns, verbs, adjectives
   It is the main concern of NLP as it is concerned with Sentiment analysis by reviewing and understanding the words, the use of Good words Expresses Positive Moods similarly the use of Bad words expresses the negative mood i.e., the person may or may not be in a good mood. This analysis expresses the current mood of the person.

4. Discourse integration
   Breaking down multiple sentences for example, the word "that" in the sentence "He wanted that" depends upon the prior discourse context.
   The meaning of any sentence depends on the meaning of the sentence from the beginning. It also gives the meaning of an immediate success sentence [3].

5. Pragmatic Analysis
   Understanding the context of the meaning

C. *NLP Current Challenges*

The current challenges of NLP Apps and techniques are:

1. Extract the context from the sentence.
2. Extract the sentiments and relating the meanings of the text based on the sentiments.
3. Extract the required meaning if there's ambiguous meaning of a certain text.

D. *Lacks in NLP*

NLP dictionaries works upon predefined vocabularies, NLP fails when a certain word has any ambiguous meanings i.e., a single word can have multiple meanings, NLP fails to grasp the required meaning of the word.

Natural language is formed by normal sentences what human says in various languages whereas NLP is comprised of idioms and some pre-defined structures which can not interpret the sentences which has different meanings in different moods or emotions. Also, there is a vast chance of redundancy.

Current systems have limited discourse capabilities that are almost ex-collusively handcrafted. Thus, current systems are limited to viewing interaction, translation, and writing text as processing a sequence of either isolated sentences or loosely related paragraphs. Consequently, the user must adapt to such limited discourse [8].

E. *Current Applications of NLP*

NLP is being used in the following APIs to facilitate applications which are facilitating the mining of texts and recognizing the texts from image at a large extent but still not completely successful because of the lack of the abilities to understand sentiments and ambiguities:

- IBM Watson API
- Chatbot API
- Speech to text API
- Sentiment Analysis API
- Translation API by SYSTRAN
- Text Analysis API by AYLIEN
- Cloud NLP API
- Google Cloud Natural Language API

F. *PREPROCESSING*

Like all machine learning tasks, language learning starts with problem definition and data collection.77 This initial phase is known as preprocessing [5]. Pre-Processing is the step just before the actual interpretation of the Language. The goal of preprocessing is to synthesize and process the words before it is parsed for NLP. This step is accomplished by using two methodologies, Text Corpus & Vector Space.

i. Text Corpora

NLP uses data in the form of a text corpus, which is a body of text commonly stored in various formats including SQL, CSV, TXT, or JSON [5]

ii. Vector Space

Vector space models represent words as real-valued vectors. The vector values are associated with abstract features [5]

F. *Tokenization*

Tokenization involves splitting of text documents, phrases, sentences, and words etc. into chunks.

For example: “Tokenization is the feature of Natural Language Processing” can be tokenized as [“Tokenization”, “is”, “the”, “feature”, “of”, “Natural”, “Language”, “Processing”] [7].

G. *Part of speech tagging (POS)*

Part of speech (POS) tagging is the process in which each word/ token in a document is tagged with its respective word class. The word classes include noun (NN), verb (VB), adjective (JJ), adverb (RB), conjunction (CC), preposition (PRP, TO, IN), etc. [10].

.

IV. CONCLUSION

In this paper we have briefly explained how NLP works & what are the components of NLP, the need of NLP. We have also discussed some existing techniques to acquire NLP.

ACKNOWLEDGMENT

This Research was supported by the faculty members of Sir Syed University of Engineering and Technology. We are thankful for their guidelines and making us to able to put efforts on this research.

REFERENCES

[1] Khurana, D., Koli, A., Khatter, K., & Singh, S. (2017). Natural language processing: State of the art, current trends and challenges. arXiv preprint arXiv:1708.05148.

[2] Robaldo, L., Villata, S., Wyner, A., & Grabmair, M. (2019). Introduction for artificial intelligence and law: special issue “natural language processing for legal texts”.

[3] Kupiyalova, A., Satybaldiyeva, R., & Aiaskarov, S. (2020, June). Semantic search using Natural Language Processing. In 2020 IEEE 22nd Conference on Business Informatics (CBI) (Vol. 2, pp. 96-100). IEEE.

[4] Spyns, P. (1996). Natural language processing in medicine: an overview. Methods of information in medicine, 35(4-5), 285-301.

[5] Haney, B. (2020). Applied Natural Language Processing for Law Practice. Brian S. Haney, Applied Natural Language Processing for Law Practice.

[6] Yaseen, Y. K., Abbas, A. K., & Sana, A. M. (2020). Image Spam Detection Using Machine Learning and Natural Language Processing. Journal of Southwest Jiaotong University, 55(2).

[7] Vangara, R. V. B., Vangara, S. P., & Thirupathur, V. K. (2020). A Survey on Natural Language Processing in context with Machine Learning.

[8] Joseph, S. R., Hlomani, H., Letsholo, K., Kaniwa, F., & Sedimo, K. (2016). Natural language processing: A review. Natural Language Processing: A Review, 6, 207-210.

[9] Alexopoulou, T., Michel, M., Murakami, A., & Meurers, D. (2017). Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing natural language processing techniques. Language Learning, 67(S1), 180-208.

[10]Vani, K., & Gupta, D. (2018). Unmasking text plagiarism using syntactic-semantic based natural language processing techniques: Comparisons, analysis, and challenges. Information Processing & Management, 54(3), 408-432.