International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT228429

A Study on The Role of Machine Learning in Natural Language Processing

Anuj Kumar Dwivedi, Mani Dwivedi
Assistant Professor, Ajay Kumar Garg Engineering College, Ghaziabad, India

Article Info

Publication Issue :

Volume 8, Issue 4
July-August-2022

Page Number : 193-198

Article History
Accepted: 10 July 2022
Published: 30 July 2022

ABSTRACT

Natural Language Processing (NLP) is the part of Artificial Intelligence that helps computers communicate with humans in their own language and scales other language-related tasks. NLP makes it possible for computers to read text, hear speech, interpret it, measure sentiment, and determine which parts are important.

Combined with machine learning algorithms, NLP generates systems that learn to perform tasks on their own and get better through experience. NLP-powered tools can help you to classify social media posts by sentiment, or extract named entities from business emails, among many other things. The role of machine learning in the NLP is clearly defined in this paper with its application areas.

Keywords : Security, Encryption, Blockchain

I. INTRODUCTION

Machine Learning and Natural Language Processing are important subfields of Artificial Intelligence that have gained prominence in recent times. The goal of NLP is to build systems that can make sense of the text and automatically perform tasks like translation, spell check, or topic classification. Machine Learning and Natural Language Processing play an awfully important part in making a synthetic agent into an artificial 'intelligent' agent. An Artificially Intelligent system can accept better information from the environment and might act on the environment in a user-friendly manner because of the advancement of Language Processing.

Similarly, an Artificially Intelligent System can process the received information and perform better predictions for its actions due to the adoption of Machine Learning techniques.

Machine Learning gives the system the power to find out past experiences and examples. General algorithms perform a hard and fast set of executions in line with what it's been programmed to try, and they don't possess the flexibility to resolve unknown problems. And, in the real world, most of the issues faced contain many unknown variables, making the standard algorithms very less effective. This is often where machine learning comes to the fore. With the assistance of past examples, a machine learning algorithm is much better equipped to handle such unknown problems.

Some of the classic examples given include spam mail detection. To detect and classify if a mail could be a legitimate one or spam includes many unknowns. There are some ways within which spam filters are

often evaded. For a standard algorithm to figure, every feature and variable should be hardcoded, which is extremely difficult, if at all possible[3]. Whereas, a machine learning algorithm is going to be able to add such an environment because of its ability to learn and form a general rule.

Deep Learning could be a specialization of machine learning algorithms, the Artificial Neural Network. In recent times it has been observed that deep learning techniques are widely adopted and have produced good results still. The flexibility provided by the deep learning techniques in deciding upon the architecture is one of the important reasons for the success of these techniques. Deep learning techniques are at the forefront of machine learning techniques used for research in natural language processing.

Natural Language Processing, on the opposite hand, is the ability of a system to understand and process human languages. Automatic data processing only understands the language of 0's and 1's, it does not understand human languages like English or Hindi. Natural Language Processing gave the system the flexibility to understand English or the Hindi language.

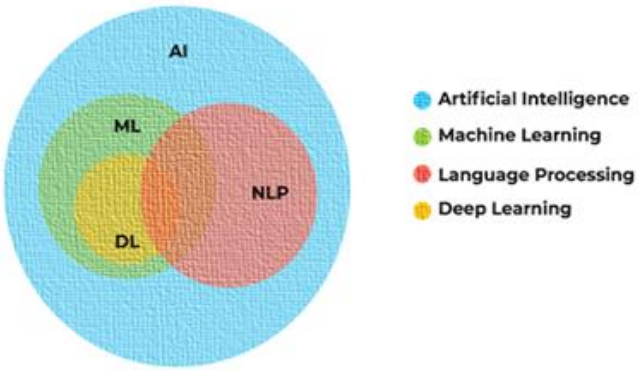

Figure 1. Domain of Artificial Intelligence

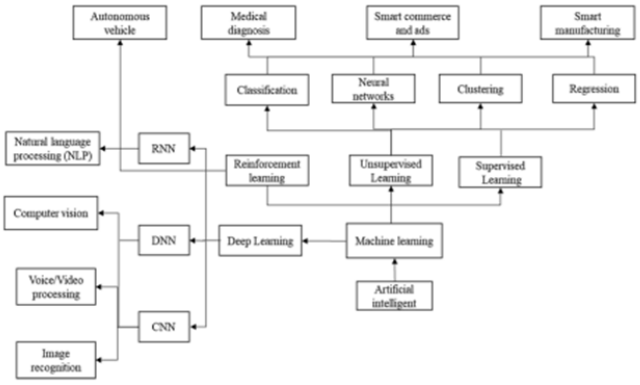

Figure 2. Domain ontology of machine learning and deep learning methodologies

Natural Language Processing has seen large-scale adaptation in recent times because due to the extent of user-friendliness it brings to the table. From choosing your choice of music to controlling your electronic appliances like Air conditioners, and ovens, of course even the ceiling fans and light bulbs, everything and anything can now be done using your voice, thus making these electronic items smart. This can be all possible due to Natural Language Processing[1].

Even as NLP has made it easier for the users to interact with complex electronics, on the other hand, there's plenty of processing happening behind the scenes which makes this interaction possible. Machine learning has played a really important role in this processing of the language[5].

Apart from playing a role in the proper processing of natural language Machine Learning has played a constructive role in important applications of natural language processing as well. Important NLP applications like Sentiment Analysis, Chatbot Systems, Question Answering Systems, Information Retrieval Systems, Machine Translation, and Email Classification, among others have all included machine learning techniques for better working.

II. NATURAL LANGUAGE PROCESSING

Natural Language Processing broadly refers to the study and development of computer systems that will interpret speech and text as humans naturally speak and form it. Human communication is frustratingly vague at times; we all use colloquialisms, and abbreviations, and don't often bother to correct misspellings. These inconsistencies make computer analysis of natural language difficult at the best. But within the last decade, both NLP techniques and machine learning algorithms have progressed immeasurably.

There are three aspects to any given chunk of text:

2.1 Semantic Information

Semantic information is the specific meaning of a personal word. A phrase like "the bat flew through the air" can have multiple meanings reckoning on the definition of the bat: winged mammal, wooden stick, or something else entirely? Knowing the relevant definition is important for understanding the meaning of a sentence.

Another example: "Billy hit the ball over the house." Because the reader, may assume that the ball in question may be a baseball, how can we know? The ball might be a volleyball, a ball, or even a bocci ball. We assume baseball because they're the sort of balls most frequently "hit" in such a way, but without natural language machine learning, a computer wouldn't know to create the connection.

2.2 Syntax Information

The second key component of text is sentence or syntax, called syntax information. Take the sentence, "Sarah joined the group already with some search experience." Who exactly has the search experience here? Sarah, or the group? Depending on how you read it, the sentence has very different concerning Sarah's abilities meanings.

2.3 Context Information

Finally, we must understand the context that a word, phrase, or sentence appears. What is the concept being discussed? If an individual says that something is "sick", are they talking about healthcare or video games? The implication of "sick" is usually positive when mentioned in an exceeding context of gaming, but nearly always negative when discussing healthcare.

Natural Language Processing in real life:

- Information Retrieval (Google finds relevant and similar results).
- Information Extraction (Gmail structures events from emails).
- Machine Translation (Google Translate translates language from one language to another).
- Sentiment Analysis (Hater News gives us the sentiment of the user).
- Spam Filter (Gmail filters spam emails separately).
- Auto-Predict (Google Search predicts user search results).
- Auto-Correct (Google Keyboard and Grammarly correct words otherwise spelled wrong).
- Speech Recognition (Google WebSpeech or Vocalware).
- Question Answering (IBM Watson's answers to a query).
- Natural Language Generation (Generation of text from image or video data.)

III. ROLE OF MACHINE LEARNING IN NATURAL LANGUAGE PROCESSING

Processing of natural language so that the machine can understand the natural language involves many steps. These steps include Morphological Analysis, Syntactic Analysis, Semantic Analysis, Discourse Analysis, and Pragmatic Analysis, generally, these analysis tasks are applied serially[1]. Machine Learning acts as important value addition in the majority of these processes in some form or the opposite. Let us discuss these steps in detail[2].

3.1 Morphological Analysis:

As we all know that the information received by the computing system is within the type of 0s and 1s. These 0s and 1s can be converted into alphabets using the ASCII code. So, it can be said that a machine receives a bunch of characters when a sentence or a paragraph has been provided to that. At the stage of morphological analysis, the primary task is to detect the words and therefore the sentences. This identification is named tokenization. Many different Machine Learning and Deep Learning algorithms are employed for tokenization including Support Vector Machines and Recurrent Neural networks.

Once the tokenization is complete the machine has with it a bunch of words and sentences. Most of the sentences which are formed contain affixes. These affixes complicate the matter for the machines as having a sense dictionary containing all the words with all its possible affixes is sort of impossible. So, the subsequent task that the morphological analysis level is removing these affixes. These affixes are often removed either using stemming or lemmatization. Machine Learning algorithms just like the random forest and decision tree are quite successful in performing the task of stemming[3].

3.2 Syntactic Analysis

The next task in natural language processing is to test whether the given sentence follows the grammar rule of a language. To try to do this the words are first tagged with their part of speech. This helps the syntactic parsers in checking the grammar rules. Machine learning and Deep learning algorithms are just like the random forest and therefore the recurrent neural network has been successfully used and implemented for this task. Machine learning algorithms like K- nearest neighbor are used for implementing syntactic parsers as well.

3.3 Semantic Analysis

At this level, the word meanings are identified using word-meaning dictionaries. The difficulty encountered here is, that the identical word might need different meanings according to the context of the sentence. For example, the word 'Bank' might mean a Blood Bank or a Financial Bank, or maybe a River Bank / Shore, this creates ambiguity. So, removing this ambiguity is one of the important tasks at this level of natural language processing called Word Sense Disambiguation.

Word sense disambiguation is one of the classical classification problems which are researched with different levels of success. Machine learning just like the random forest, gradient boosting decision trees are successfully employed. But, in recent times it is the learning algorithms just like the recurrent neural network, long short short-terms based recurrent neural network, gated recurrent unit but-based rent neural network, and convolution neural network is researched and have produced very great results.

3.4 Discourse Analysis

There are instances where pronouns are used or certain subjects/objects are remarked, which are outside of the present preview of the analysis. In such cases, the semantic analysis won't be ready to give proper aspiring to the sentence. This is often another classical problem of reference resolution which has been tackled by machine learning and deep learning algorithms.

3.5 Pragmatic Analysis

Many a time sentences convey a deeper meaning than what the words can describe. That is, the machine has got to discard the acceptance understood after semantic analysis and capture the intended or implied meaning. It's easier said than done. For several years now this can be of natural language process has intrigued researchers. One of the classic samples of pragmatic analysis is sarcasm detection.

Many, of course, the various machine learning and deep learning algorithms have been employed with varied success for performing sarcasm detection for performing pragmatic analysis normally.

IV. MACHINE LEARNING IN THE APPLICATIONS OF NLP

As with the processing task of the natural language machine learning and deep learning algorithms have played an awfully important role in most of the applications of natural language processing. Hence, most of the deep learning techniques including, Deep Neural networks, Autoencoders, Restricted Boltzmann machines, Recurrent Neural networks, and Convolution Neural networks are experimented with to urge good accuracy within the different applications of Natural Language Processing. A number of these applications of Natural Language Processing where the deep learning techniques have had an extremely positive role to play are sentiment analysis, chatbot systems, question answering systems, information retrieval systems, and machine translation.

4.1 Sentiment Analysis

Sentiment Analysis strives to investigate the user opinions or sentiments on a specific product. Sentiment analysis has become a really important part of Customer Relationship Management. Even one negative opinion will be disastrous for the product. Recent times have seen greater use of deep learning techniques for sentiment analysis. A noteworthy fact to notice here is that new deep learning techniques are quipped especially for the analysis of sentiments that's the extent of research that's being conducted for sentiment analysis using deep learning.

4.2 Chatbot Systems

Chatbot systems are conversational agents or dialog systems that attempt to engage the user during a conversation. This conversation may be through voice or text. Personal assistants like Amazon's Alexa and Google Assistant have popularized the chatbot systems and have also showcased the amount of ease through which user interactions are often allotted. As easy as it may sound, the event of a true chatbot system that can replace an individual's agent is an exceptionally difficult task. Which requires Natural Language Understanding and Natural Language Generation.

Recent frameworks like Google's DialogFlow, IBM's Watson AI, and Amazon's Alexa AI provide a simple way of developing a chatbot system. And, these frameworks employ complex and proprietary deep learning architectures.

4.3 Question Answering Systems

As the name suggests, a question-answering system is a system that tries to answer a user's questions. Recent times have seen the skinny line separating a dialog system and a question answering system getting blurred and most of the time a chatbot system performs the question answering task and it's true the opposite way round likewise. So, the research works which pledge to develop a chatbot system will, in all told probability, be developing a question-answering system within it similarly.

A question-answering system has three important components, Question Processing, Information Retrieval, and Answer Processing. Machine Learning and Deep Learning techniques have played an important role altogether in three components. Especially, Question Processing has attracted quite a few research. The thought here is that understanding the question is extremely important for better answer retrieval. The question processing task is taken as a classification problem and lots of research works have experimented with deep learning techniques for better question classification

4.4 Information Retrieval Systems

Information Retrieval is another important application of Natural Language Processing that tries to retrieve relevant information. Information retrieval systems act because of the backbone of the systems like chatbot systems and question-answering systems.

The most basic way of retrieving any information is using the frequency method where the frequency of keywords determines if a selected data is retrieved or not. But, smart systems process the desired query similarly because they present large data to retrieve only the relevant information. This process is administrated using deep learning techniques.

4.5 Machine Translation

A machine translation system is striving to translate a text from one language to a different one with minimum or no human intervention. Applications like Google Translate are among the simplest samples of the machine translation system.

Having a translation system that translates word to word isn't enough because the construction of a sentence might vary from one language to a different one. For example, English follows the Subject-Verb-Object format whereas Hindi follows Subject -Object-Verb form for sentence construction. Apart from this, there are many various rules which need require to be followed. All these things make the task of machine translation difficult.

V. CONCLUSION

So, it may be observed that Machine Learning and Deep Learning techniques are being extensively researched for their employment within the field of Natural Language Processing. These learning techniques are playing a crucial role in the majority of the processing of natural language tasks yet as in most applications of natural language processing. All the various processing of natural language tasks and therefore the different applications of natural language processing are different fields of research by themselves. Machine learning might be a good solution for analyzing text data. It's vital purely rules-based text analytics is dead-end. But it is not enough to use a single type of machine learning model. Certain aspects of machine learning are very subjective. We need to tune or train your system to match your perspective. In conclusion, it can be said that Machine Learning and Deep Learning techniques have been playing a very positive role in Natural Language Processing and its applications.

VI. REFERENCES

[1]. Tatwadarshi P. Nagarhalli, Dr. Vinod Vaze, and Dr. N. K. Rana, "Impact of Machine Learning in Natural Language Processing: A Review", Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV 2021), 2021.

[2]. Aravind Pai, What is Tokenization in NLP? Here's All You Need To Know. Available at: https://www.analyticsvidhya.com/blog/2020/05/what-is-tokenization-nlp/.

[3]. F. Jia, Y. Lei, J. Lin, X. Zhou, N. Lu, Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, Mech. Syst. Sig. Process. 72 (2016) 303–315.4.

[4]. D.H. Milanez, L.I.L. de Faria, R.M. do Amaral, J.A.R. Gregolin, Claim-based patent indicators: a novel approach to analyze patent content and monitor technological advances, World Patent Info. 50 (2017) 64–72.

[5]. Poli, M. Healy, A. Kameas, Theory and Applications of Ontology: Computer Applications, Springer, Berlin, 2010.

Cite this article as :

Anuj Kumar Dwivedi, Mani Dwivedi , "A Study on The Role of Machine Learning in Natural Language Processing", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 4, pp. 192-198, July-August 2022. Available at doi : https://doi.org/10.32628/CSEIT228429

Journal URL : https://ijsrcseit.com/CSEIT228429