IJARSCT

ISSN (Online) 2581-9429

IJARSCT

Impact Factor: 5.731

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Volume 10, Issue 1, October 2021

Machine Learning (ML) for Natural Language Processing (NLP)

Jaydeep Pandya
Student, M.Sc., Department of Information technology
Sir Sitaram and Lady Shantabai Patkar College of Arts and Science, Mumbai, India
Jaydeepkumarpandya178@gmail.com

Abstract: *Over the past few years, Artificial Intelligence and its allied fields, subfields, such as Machine Learning, Deep Learning and Natural Language Processing, have seen unprecedented growth in research and development. Researchers have gained a new sense of excitement due to cheaper computing devices and the variety of applications available in these fields. In today's world, I believe it is safe to say that artificial intelligence and its subfields have been positively affecting a large number of industries. Machine learning and deep learning don't just improve the efficiency of businesses, they have also had a significant impact on several subfields of artificial intelligence, including computer vision and natural language processing. Natural Language Processing is the ability for computers to understand human languages, which is a difficult task, and where learning methodologies have played a very important role in proper analysis. This paper highlights the important role played by learning techniques in improving natural language processing efficiency.*

Keywords: Machine Learning, Deep Learning, Natural Language Processing, Artificial Intelligence, and Word Sense Disambiguation.

I. Introduction

In recent years, Artificial Intelligence subfields such as Machine Learning and Natural Language Processing have gained prominence. Taking advantage of ML and NLP plays a crucial role in making an artificial agent an artificial 'intelligent' agent. The advancement in Natural Language Processing has enabled AI systems to better understand the environment and act on it in a user-friendly manner [1].

A system that understands and processes natural languages, on the other hand, is called Natural Language Processing. There is no human language other than the language of 0's and 1's that the computer understands [2]. The computing system was able to understand English and Hindi by using Natural Language Processing. Because of its user friendliness, Natural Language Processing has seen wide-scale adaptation in recent years. Every electronic appliance, from air conditioners and ovens to ceiling fans and light bulbs can be controlled remotely, from music to lights to air conditioners can be done using your voice, thus making these electronic items smart...!! A NLP system enables all of this. Even though NLP has simplified interaction with complex electronics, in order for this to happen there is a lot of processing going on behind the scenes [1]. The processing of the language has greatly benefited from machine learning.

The adoption of Machine Learning techniques allows an Artificially Intelligent System to process the received information and predict its actions more accurately. By using machine learning, the system can learn from past experiences. A general algorithm performs a fixed set of operations according to what it has been programmed to do, and it is not capable of solving unknown problems. Moreover, most real-world problems involve many unknown variables, making traditional algorithms ineffective. Throughout this process, machine learning plays a significant role [2]. Machine learning algorithms are far better equipped to solve unknown problems with the help of past examples.

Spam mail detection is one of the classic examples given. The process of detecting and classifying whether a message is legitimate or spam involves many unknowns. Spam filters can be circumvented in many ways. It is extremely difficult, if not impossible, to hardcode each and every feature and variable in a traditional algorithm. A machine

DOI: 10.48175/568

learning algorithm on the other hand, has the capability to learn and form a general rule in such an environment. It is possible that the linguistic knowledge is ambiguous or contains ambiguities. Various NLP tasks such as POS, NER, SBD, word sense disambiguation, and word segmentation are used to resolve ambiguity in the discovered linguistic knowledge. Models of machine learning play a vital role in the resolution of ambiguities as well as in the capture of all linguistic knowledge [3]. Literature-based, advance NLP algorithms that are based on statistical machine learning while others are purely supervised. In the past, all NLP tasks were carried out using various rules-based approaches, where large sets of rules were compiled manually. In contrast to most previous language processing endeavors, machine learning uses a different paradigm. Various ML techniques have been extensively investigated in the literature for various NLP tasks. Algorithms for machine learning can be parametric, nonparametric, or kernel-based [3]. In ML-based approaches, ML algorithms are trained in the training phase on enough pretagged data to generate model data, after that the model data are used in the testing phase to test new data. The focus of research, however, has shifted to stochastic machine learning. In such a model, to each input feature, a real -valued weight is attached, which generates soft probabilistic decisions [3]. The benefit of such models is that: these models can represent a relation quality in different dimensions.

Computers can become intelligent and make decisions with artificial intelligence. Machine learning and Natural language processing (NLP) enable Artificial Intelligence on the platform as well as enabling the platform to reach machine learning goals more effectively [1]. As a result, machines are expecting more from us. In many fields, machine learning has been used. Furthermore, natural language processing is used in a wide range of fields, including search auto correction, auto complete, language translators, social media monitoring, chatbots, target advertising, grammar checkers, email filtering, voice assistance for recommendations, as well as other areas of research [3].

II. Natural Language Processing And Machine Learning

The field of Natural Language Processing and Machine Learning is a subfield of artificial intelligence that has gained wide popularity and has been widely adapted in many industries in the recent years. Using machine learning, computers are able to solve problems that weren't explicitly programmed for them. Many real-world applications make it virtually impossible to explicitly program or develop an algorithm that can anticipate all possible input types and solve the problem at hand [4]. Deep Learning and Machine Learning help computer systems learn from the data and make conclusions on their own. This helps the computer system solve problems that it might not have encountered before or improve its response based on its past experiences over time. Machine learning and deep learning are popular techniques because of their ability to learn from millions of documents, and they can be applied in a variety of domains such as healthcare, transportation, customer service, etc. Hence, machine learning has been gaining popularity in recent years for this reason [4].

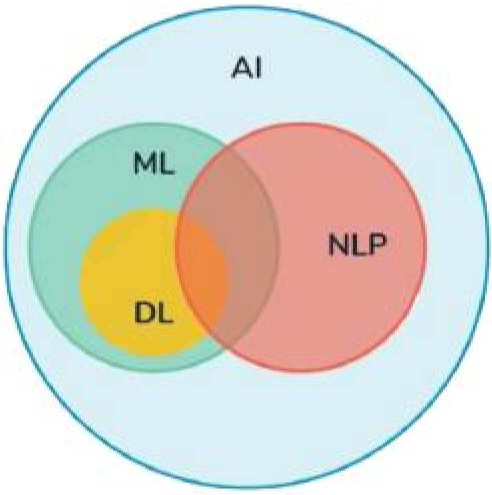

Figure 1: AI, ML, NLP

Machine learning and deep learning have become popular because of the availability of cheap, high-performance computing devices and huge amounts of data that can be exploited to gain new insights. Machine learning is also popular because of the wide range of applications it can be applied to. [5] Machine learning has been adapted in many fields, such as manufacturing, healthcare, transportation, automobiles, e-commerce, insurance, customer service, and energy, among others. Machine learning and deep learning have contributed significantly to the intelligence of computer systems. However, leveraging these intelligent systems is a difficult task. In addition, only a few machine learning engineers were able to use or communicate with these intelligent machine learning systems. This ability to communicate with intelligent systems in human languages has been made possible by Natural Language Processing, in a sense, bringing these systems to the mass market [4].

It is a field of study that involves a variety of disciplines computer science and linguistics. Natural language processing is another important subfield of artificial intelligence. Aims to provide the computer with the ability to understanding human spoken or written language. The task of understanding a natural language sentence is difficult. Due to the inherent ambiguity of languages, this task becomes even more challenging. Ambiguity refers to the fact that the same words have different meanings depending on the context , such as the word 'Bank' can mean either a financial institution or a slope down a body of water. As a result, computers have a harder time understanding natural language [6].

Processing and analysis of human language are carried out in multiple stages. A morphological analysis, a syntactic analysis, a semantic analysis, a discourse analysis, or a pragmatic analysis can take place at each stage. This level of analysis determines the machine's ability to process and understand natural language. Natural language processing has evolved into separate fields of study by itself as well. Natural language processing has become popular due to systems like Siri, Alexa, and Google Assistant. In addition to these benefits, businesses can also save millions of dollars by using natural language processing systems. There are many reasons why billions of dollars are being spent on research and development in natural language processing.

Natural language processing has several important applications which contribute to its popularity. In addition to Sentiment Analysis and Machine Translation, Chatbots or Conversational Agents, Text Classification, and Information Retrieval are some applications of natural language processing. A number of artificial intelligence subfields have also become more efficient and accurate by collaborating one with another. For example, natural language processing has played a major role in the success of robotics, and machine learning has contributed to the success of computer vision, another subfield of artificial intelligence [7]. The field of NLP still faces a lot of challenges (such as research has been conducted on computer interfaces). It has opened up many opportunities for people with interests in robotics, automation, and digital technologies transformation.

Similarly, machine learning has played a very important role. Plays a crucial role in natural language processing. The case may be Natural language processing stages or types of analysis Among the many applications of natural language processing, Deep learning and machine learning have played a key role in. They play a crucial role in improving their efficiency and accuracy. Thus, this paper attempts to provide a review and highlight this important issue [4]. Machine learning and deep learning techniques play an important role Enhancing the efficiency of natural language processing.

III. Natural Language Processing And Learning Techniques

Machines cannot understand human language; they can only understand the language of 0's and 1's. Processing tasks are needed in order to make a machine understand human language. You will need to identify the words and sentences for a stream of characters, check whether the sentence identification matches the rules of the language, extract or comprehend the meaning of the given sentence, and try to make sense of it. In this case, you need to identify what the given sentence means and identify the intended meaning. We refer to these five tasks as morphological analysis, syntactic analysis, semantic analysis, discourse analysis, and pragmatic analysis.

It has been observed that machine learning techniques like Naive Bayes, Support Vector Machines, Decision Trees, Random Forests and deep learning techniques like Recurrent Neural Networks and Convolution Neural Networks have made a very positive contribution in almost all fields Figure 2 shows the stages of natural language processing. It is important to note that these stages need to be performed sequentially one after another for better processing of natural language.

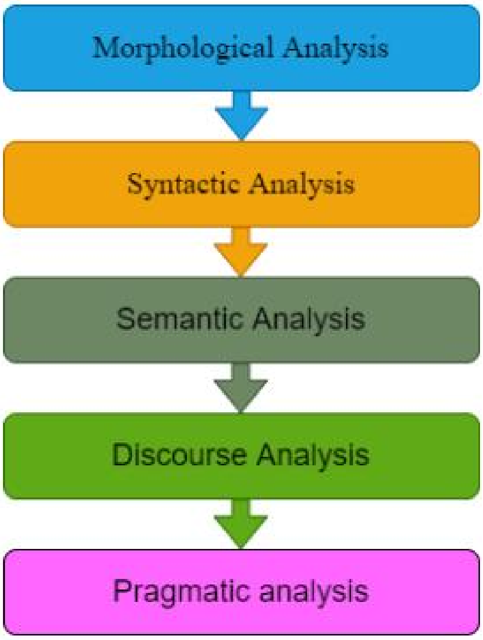

Figure 2: Stages in Natural Language

3.1 Morphological Analysis

In natural language processing, the first step is to identify the words and sentences that are being processed. This process is called tokenization. These words are often used to contain affixes which might confuse the machine. So, a A process called stemming is carried out in order to remove these affixes. Tokenization and Stemming is carried out at the morphological stage in natural language processing.

Tokenization is a key task at a morphological level. In recent years, machine learning approaches have been studied in order to determine how they can improve tokenization efficiency [7]. This computing system receives data as 0s and 1s. The following 0s and 1s Using the ASCII code, 1s can be converted into alphabets. When a machine receives a sentence or paragraph, it receives a bunch of characters. At the level of morphological Analyze. Identifying words and sentences is the first step. Tokenization is the process of identifying. Support Vector Machines and Recurrent Neural Networks are among the Machine Learning and Deep Learning algorithms that have been employed for tokenization. The machine collects words and sentences once tokenization is completed. The majority of sentences formed contain affixes [8]. The affixes complicate the problem for the machines because it is virtually impossible to compile a dictionary of words with all their possible affixes.

So, the next task at the morphological analysis level is removing these affixes. By stemming or lemmatizing, these affixes can be removed. For stemming, algorithms like the random forest and decision tree have proven quite successful [1].

3.2 Syntactic Analysis

The next step in natural language processing is to determine whether the given sentence follows the grammar rules of a language. This is done by first categorizing each word. Syntactic parsers can use this information to check grammar rules. Several machine learning and deep learning algorithms have been successfully implemented for this task, such as the random forest and recurrent neural network. Syntactic parsers have also been implemented using machine learning algorithms, such as K- nearest neighbor [9]. In NLP to check if the given sentences follow the rules of a language, there are four basic rules in english language which mentioned below.

- A complete sentence requires a subject and a verb and expresses a complete thought.
- Separate ideas generally require separate sentences.
- English word order follows the subject-verb-object sequence(Ram(S) plays(V) Cricket(O)).
- A dependent clause contains a subject and a verb [9].

There are many rules in the language similar to these. The following sentences should be followed in order to qualify and pass Syntactic analysis. The checking and analysis of An important aspect of the sentence is its syntax The 'right' meaning of the given sentence. The rules of a language are usually elaborately defined and there are no updates in these rules hence, there are well established rule-based parsers which can parse the sentence and check the syntax of a sentence. Because of this, learning techniques with has not made a major contribution in terms of syntactic analysis [7].

3.3 Semantic Analysis

After the syntax has been confirmed and verified in the syntactic analysis stage, the next step is to understand the meaning of the sentence, from the sentences themselves. Using word and word-meaning dictionaries this task of understanding the meaning of the sentence is achieved. This step is the semantic analysis step in the processing of natural language. A language by nature is ambiguous, and this is especially true for word meanings. In any language there will be many words which will have different meanings or will give different sense. There are many words which exhibit different meanings under different circumstances. Generally, these meanings are Determined by the context.
For example, the word ‘Bank’ can mean a financial institution or a storage of some kind or a riverside. Identifying which meaning to take becomes difficult for a machine. So, even While using a word and word-meaning dictionaries it becomes difficult to identify the actual sense of a particular word in a Given context, this is called Word Sense Ambiguity. Hence, disambiguation of this word sense is a very important task which needs to be performed while performing semantic analysis [9].

Word Sense Disambiguation is an important research problem which is being researched even now because of the different factors which need to be considered while performing the task. Word sense disambiguation has been considered as a classification problem and machine learning and deep learning techniques have been employed in order to solve this ambiguity in the word meanings. Understanding the importance of word sense disambiguation many research Works have tried to tackle this issue for many different languages [2].

3.4 Discourse Analysis

There are cases where sentences can start with pronouns or the sentences may refer to a subject or an object which is not present in the present sentence. For example, consider the two sentences “Ram is a good by. He likes to play Cricket.” If the second sentence is analyzed independently it would make no sense at all because of the presence of the pronoun ‘He’. The second sentence will only make sense when it is analyzed together with the first sentence. Here, he refers to the person ‘Ram’. This is discourse analysis [10].

There some tougher examples and cases which have to be handled by the system while performing discourse analysis. For example, consider the given two sentences “Ram went to Shyam’s shop to check out new Cricket bats. In the example given, the words 'He' and 'it' can refer to many different things in the second sentence, even if the first sentence is taken in to consideration as well.. ‘He’, can refer to either ‘Ram’ or ‘Shyam’ and ‘it’ can refer to either the ‘shop’ or the ‘Cricket Bat’ [10].

In such cases it becomes difficult to analyze the sentence and extract the actual meaning of the sentence. For such cases, it is not enough to have only the knowledge of the previous sentences some like of methodology is required in order to resolve such reference problems. Resolving such cases which is carried out in discourse analysis is called as Reference Resolution. As mentioned earlier, reference resolution is a very active area of research by itself [9].

3.5. Pragmatic Analysis

Finally, there are many cases where the written meaning and the actual intended meaning may be completely different. In such cases the meaning of the sentences understood by semantic analysis is not sufficient. So, pragmatic analysis is carried out in order to identify the intended meaning of a given sentence. For example, consider the sentence “The soldier fought like a lion.” The literal meaning does not make any real sense at all. The actual meaning of the sentence is that the soldier fought very ferociously, which is the intended meaning. The given examples showcases the necessity of perform pragmatic analysis.

One of the most integrating task of natural language processing which has baffled and has kept engaged many researchers has been the automatic detection of sarcasm by the machine. Sarcasm is a satirical remark which generally intend to give the opposite meaning to what has been said. Automatic sarcasm detection is one of the most classic examples of pragmatic analysis.

Consider the example, “I slipped and fell in my bathroom today morning. What a perfect start to a morning.” The semantic meaning of the second sentence is a positive sentiment. Whereas, the intended meaning is exactly the opposite. It is actually an expression of aghast [11]. Pragmatic analysis tries to find out intended meaning from such sentences. Pragmatic analysis in general and automatic sarcasm detection has been one of the very widely researched topic in natural language processing. Multitudes of research work has been carried out by many different researchers for accurate sarcasm detection. Many different machine learning and deep learning approaches have been experimented with in order accurately identify sarcasm. Machine Learning algorithms like K-Nearest Neighbour, Support Vector Machines, and Random Forest algorithm [11] has been employed for sarcasm detection on twitter data. In addition to these machine learning algorithm being used independently, machine learning techniques have also been used in conjunction with rule based systems for getting better detection accuracy.

IV. Application Of NLP

As with the processing task of natural language machine learning and deep learning algorithms have played a very important role in almost all of the applications of natural language processing. In recent times there has been a renewed research interest in these fields because of the ease with which machines learning and deep learning algorithms can be implemented, especially deep learning. To achieve good accuracy, almost all deep learning techniques have been experimented with, such as Deep Neural Networks, Autoencoders, Restricted Boltzmann Machines, Recurrent Neural Networks, and Convolutional Neural Networks [1].

There has been extensive research done on these applications of a recurrent neural network, as well as its variants, Long Short Term Memory, Gated Recurrent Units as well as Convolutional Neural Networks. We will look at some of these applications of Natural Language Processing where deep learning techniques have been very useful.

4.1 Sentiment Analysis

Sentiment Analysis analyzes user opinions or sentiments about a specific product. Sentiment Analysis plays an increasingly important role in Customer Relationship Management. A single negative can make a difference. A negative opinion can be disastrous for a product. Deep learning has become increasingly popular in recent years to analyze sentiment. An interesting fact to note here is that new deep learning techniques have been quipped especially for analysis of sentiments, which is the level of research that is being conducted for sentiment analysis using deep learning.

4.2 Chatbot Systems

The purpose of chatbots is to interact with users in a conversational manner. Text or voice can be used for this conversation. The popularity of personal assistants like Amazon's Alexa and Google Assistant has paved the way for chatbot systems and demonstrated just how easy it can be to interact with users.

It may sound easy, but developing a chatbot that can replace a human agent is extremely challenging. This requires both Natural Language Understanding and Natural Language Generation. A chatbot system can be developed easily using frameworks like Google's Dialog Flow, IBM's Watson AI, and Amazon's Alexa AI. All of these frameworks utilize complex and proprietary deep learning architectures.

4.3 Question Answering Systems

The name implies that a question answering system attempts to answer users' questions. In recent times, the division between dialog and question-answer systems has become a thinner one. The task of answering questions is generally handled by chatbots, and it's true that most often they're automated. It works both ways, too. Thus, the research works that aim to develop a chatbot system will, in general, result in developing a question-answering system within it as well. The three components of a question answering system are: Question Processing, Information Retrievel, and Answer Processing [2].

Throughout all three of these components, machine learning and deep learning techniques have played a crucial role. In particular, Question Processing has attracted a lot of research. Understanding the question is extremely important for better retrieval of answers. Question processing is viewed as a classification problem, and many research works have investigated deep learning techniques to improve question classification.

4.4 Machine Translation

The purpose of a machine translation system is to translate a text from one language to another with minimal or no human intervention. Google Translate is one of the best examples of a machine translation system. The use of a translation system that translates word for word is not sufficient since the construction of a sentence may differ from one language to another. For example, English uses the Subject-Verb-Object format for sentence construction, while Hindi uses SubjectObject-Verb. Besides these, there are many other rules that need to be followed. A machine translation project is made difficult by all of these factors [1].

Machine translation has been extensively explored via recurrent neural networks, as well as their variants, Long Short Term Memory, and Gated Recurrent Unit, which can work in both directions. In order to translate properly, these neural networks must be able to hold onto contextual information. Convolution neural networks have also been tested with varying degrees of success.

The application of Machine Learning and Deep Learning techniques to the field of Natural Language Processing is therefore being extensively studied. Obviously, these learning techniques contribute to almost all aspects of natural language processing and to its applications. Each of the different natural language tasks and each of the different applications of natural language processing is a different field of research. Currently, Machine Learning and Deep Learning are being researched extensively in all these fields of research with a high rate of success. In conclusion, Natural Language Processing and its applications have benefited greatly from Machine Learning and Deep Learning.

V. Conclusion

Machine Learning and natural language Processing are vital subfields of AI which have again gained prominence within the last few decades. Machine Learning and its specialization, Deep Learning techniques have revolutionized many sectors of business. As a matter of fact, these Learning techniques have positively contributed to many Different technologies as well, including Natural Language Processing. There are five important tasks in Natural Language Processing, performing which enables the machine or a computing device to understand human language. The accuracy with which these tasks are performed determines the level of natural language understanding by the machine. The paper showcases how machine learning and deep learning techniques have been successfully employed in performing most of these tasks and they have produced very good results as well Having said that, there is still ample scope for research in increasing the efficiency with which these tasks of natural language processing can be performed. This also warrants further study and research in natural language processing. Having highlighted the important role played by the learning techniques in natural language processing it is expected that by further experimenting with different learning techniques the efficiency and accuracy of natural language processing can be improved.

REFERENCES

[1]. Tatwadarshi P. Nagarhalli, Dr. Vinod Vaze, and Dr. N. K. Rana, "Impact of Machine Learning in Naturaln 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-istokenization-nlp/.

[3]. Wahab Khan, Ali Daud, Jamal A. Nasir, Tehmina Amjad," A survey on the state-of-the-art machine learning models in the context of NLP",Kuwait J. Sci. 43 (4) pp. 95-113, 2016.

[4]. E. Alpaydın, "Introduction to Machine Learning", 2nd Edition, The MIT Press, 2010.

[5]. J. Brownlee, "Machine Learning is Popular Right Now", Available at: https://machinelearningmastery.com/machinelearning-is-popular/.

[6]. P. P. Shinde and S. Shah, "A Review of Machine Learning and Deep Learning Applications", IEEE Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1-6.

[7]. Edited by Alexander Clark, Chris Fox, and Shalom Lappin, "The Handbook of Computational Linguistics and Natural Language Processing", Blackwell Publishing Ltd., 2010.

[8]. Edited by Alexander Clark, Chris Fox, and Shalom Lappin, "The Handbook of Computational Linguistics and Natural Language Processing", Blackwell Publishing Ltd., 2010.

[9]. Tatwadarshi P. Nagarhalli,Dr. Vinod Vaze,Dr. N. K. Rana,"Impact of Machine Learning in Natural Language Processing: A Review", Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV 2021).IEEE Xplore Part Number: CFP21ONG-ART; 978-0-7381-1183-4.

[10]. J. Wu and W. Ma, "A Deep Learning Framework for Coreference Resolution Based on Convolutional Neural Network", IEEE 11th International Conference on Semantic Computing (ICSC), 2017, pp. 61-64.

[11]. N. Pawar and S. Bhingarkar, "Machine Learning based Sarcasm Detection on Twitter Data", IEEE 5th International Conference on Communication and Electronics Systems (ICCES), 2020, pp. 957-961.

DOI: 10.48175/568