



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 9, September 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.569



Natural Language Processing Development over Time and Future Scope

Asmita Fasale¹, Mansi Bamble²

U.G. Student, Department of Information Technology, B.K.Birla College of Arts, Science and Commerce,
Kalyan, India¹

U.G. Student, Department of Information Technology, B.K.Birla College of Arts, Science and Commerce,
Kalyan, India²

ABSTRACT: Natural language processing system is the computerized approach to analyze the text and voice/speech data and make computers understand those parameters the same as humans do. natural language processing combines the computational way of understanding linguistic or some rule-based models of human language. with statistical, machine learning, and deep learning models all those kinds of technology help to understand the human language. by processing the given data in the format of text or voice computer should understand the exact intent and sentiment of the speaker/writer called as Natural Language Processing. some of the great use of Natural Language Processing is the generating exact meaning of search queries, translating one language to another, extracting the core knowledge from the large paragraphs, and many more.

KEYWORDS: Natural Language Processing, python, machine learning, Linguistics, deep learning, Artificial intelligence, modeling, speech recognition

I. INTRODUCTION

text, and sentences ex-mail, SMS, reading material, email, websites, and many more. the scientific study of these languages is called Linguistics. with this type of data, humans must pose the computerized way to understand the language. the data for the language processing is available in text and voice format. natural language processing provides a method to understand those data meaning, intent, and sentiments. natural language processing is partially a branch of artificial intelligence with the ability of self learn or machine learning. it provides a way to interact with smart computers and smart humans. in the specific program, the computer analyzes, understands the natural human language data in large chunks. the core goal here is to make the computer understand the huge stack of data files or pages and understand those documents and extract relatively useful information from them.

In recent times natural language processing advances with the model architecture and model pretraining. transformer architecture provides help with building a higher level of the models and pretraining helps with the to use of a wide variety of tasks very efficiently. this is a library that takes natural language processing to next level. the library serves a wide number of communities and backed pre-trained models available for the community. We know that linguistics is the scientific study of the language with words, characters, sentences, sentiments, and phonetics. the term computer linguistic refers to the study of the fundamental study of language using computer tools. Today's generation's supercomputers can work with very powerful computational power. That is how we can work with large data sets and models.

In natural language processing, the increasing demand and efficiency of statistics also play an important role called statistical natural language processing. yet linguistic in large data set done by statistical approach gives very significant success in some areas but it has more room in future.

The paper is organized as follows. Section II describes the related work done on the natural language processing and related information impact sector and related. in section III fundamentally important Points are discussed in detail. analysis of each point with relevant information done in section III. After detection of the most important and analysis of all facts, figures, tables, statistics, justification of the analysis done in paper IV. Finally, Section V presents the conclusion.



II. RELATED WORK

The modern natural language processing has its old root, In 1950 scientist alan turing who is well known for his “Enigma Machine” published a article with the name Computing Machine and Inteligence. which now days called as a turing test, a task that involves the automated interpretation and understanding of natural language. which is called as starting days of natural language processing.

Second phase of natural language processing align between 1950s to 1990s. the term symbolic natural language processing accumulated by John Searle’s chinease room experiment in which some rule book computer was emulates to understand the natural language and to give answers based in book using natural language processing. The grogetown Experiment was conducted in 1954 to develop system to translate 60 russian sentences in plain english. but experiment was unable to give result in expected time period so because of funding stopped the project closed.

In the 1980s and early 1990s days, Focus sectors of the time included research on rule-based parsing with the development of HPSG self-developed grammar, two-level morphology, semantics with other language processing. till date all the natural language processing was happening with the predefined set of rules and regulations most of them were hand written.

After that, the era from the 1990s to 2010s the real revolution of natural language processing occurred with the introduction of a machine learning algorithm for processing with powerful computers and IBM research. after 2000 the growth of the internet additional data become available for use and research. and hence research was focused on unsupervised and semi-supervised learning.

And with present state representation learning and deep learning have become super popular and efficient for natural language processing. the result was very good and it helps with modeling of data, parsing of data, and many more. some areas of neural networks are text and speech recognition, Optical character recognition(OCR), speech segmentation, text to speech, speech to text, etc.

III. METHODOLOGY

Natural Language Processing :

A) Linguistic:

Any language is made up of some fundamental factors like words, characters, sentences, phrases, sentiment, semantics, and phonetics. The integration of all those fundamental points and studying them in a scientific manner is called Linguistic. the basic Linguistic and computational Linguistics are a key factors of natural language processing

1) Classical Linguistic contains the categorization and evaluation of the rules of the language but it resists the clean understanding of the mathematical aspects

2) with simplicity anyone who studied language can have linguistic but with being self language linguistic can make it out of scope.

3) mathematics is the heart of computation, mathematics about the language refers to mathematical linguistic. computational linguistic is the data-driven methodology of understanding language with mainstream approaches for computational linguistic.

4) The increasing amount of electronically stored data increase the strong contribution in natural language processing. in order to understand the natural language, it requires a great amount of knowledge in morphology, syntax, semantics, and pragmatics with general knowledge about the language with some understanding of variation over the geographical area.

B)Speech recognition with NLP:

1)Human language is so different in different geolocation at the same time there are so many ambiguities in languages. It is very difficult to write a dynamic program to understand this ambiguity to accurately detect the intent of the user.

2)Speech recognition is also known as speech to text. the main goal here is to convert the voice data into textual data with higher accuracy.

3)In natural conversation there are hardly any pauses between successive words and hence speech segmentation becomes the irreplaceable part of speech recognition.

4)even the same words are pronounced differently by different people with a different accent and hence speech recognition library has to contain all types of data to train the model and work efficiently even with a large chunk of data.

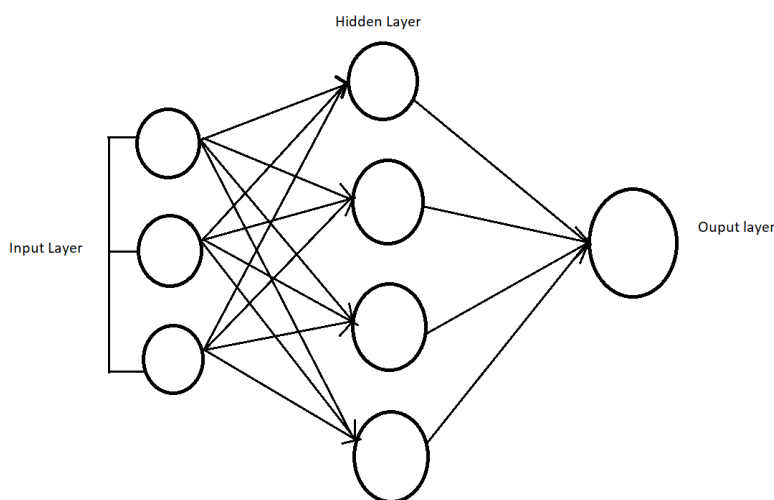
C) Active Machine Learning in NLP-

A)In natural language processing Machine Learning is the learning phase of the model to understand language linguistics. active machine learning is the type of machine learning in which learner has the control of the data for learning.

B)The controlled used by the learner data is used for machine learning. the learner can send data to the oracle to the experts then data can be labeled to make the learning phase easy. the main goal here is to train the classifier so well it will not need human intervention again.

C)Active learning applied in machine learning for natural language processing in wide areas of information extraction, named entity relation, part of speech tagging, segmentation, text categorization, and word sense disambiguation.

NEURAL NETWORK DIAGRAM



D)Deep learning for NLP-

Natural language processing is an integral part of artificial intelligence as it depends on taking input from the human either in text or voice format.

A)Deep learning is a wider area of machine learning containing artificial neural networks, based on that the learning can be 1)supervised 2)semi-supervised, or sometimes unsupervised.

B) Deep learning with architecture deep neural networks(dnn), deep belief networks(dbn), deep reinforcement learning(drl), recurrent neural networks(rnn) with a convolutional neural network(cnn), and they get used for computer vision tech, speech recognition, natural language processing(nlp), biometrics and medical, machine translation, image analysis, etc.

C) The word deep stated that there will be multiple layers in the architecture. deep learning is a modern way of artificial intelligence with efficiency, trainability, and understanding the structural part

E) NLP In Machine translation -

This is one greatest invention or we can say applications for humanity with natural language processing. google translate is the best example of NLP work in machine translation. machine translation refers to translate a given message from one language to another language with the intent and understanding the meaning of the text with tone. Day by day machine translation tools are making progress and companies working really hard towards it. still, there are some bugs in translation as it gets too tricky but it is improving with that too.

F) NLP In ChatBots and some Virtual Agent -

50% of companies are considering investing more money in chatbot because it serves so many advantages with the time range.

A. Corporate interest in chatbot has increased 5x times in the last 5 years(resource link given in reference part)

B. In each transaction chatbots are saving 0.70 dollars to service customers or handling customer

C. The predicted market capitalization is approx 1.3 billion dollars by the end of the year 2025

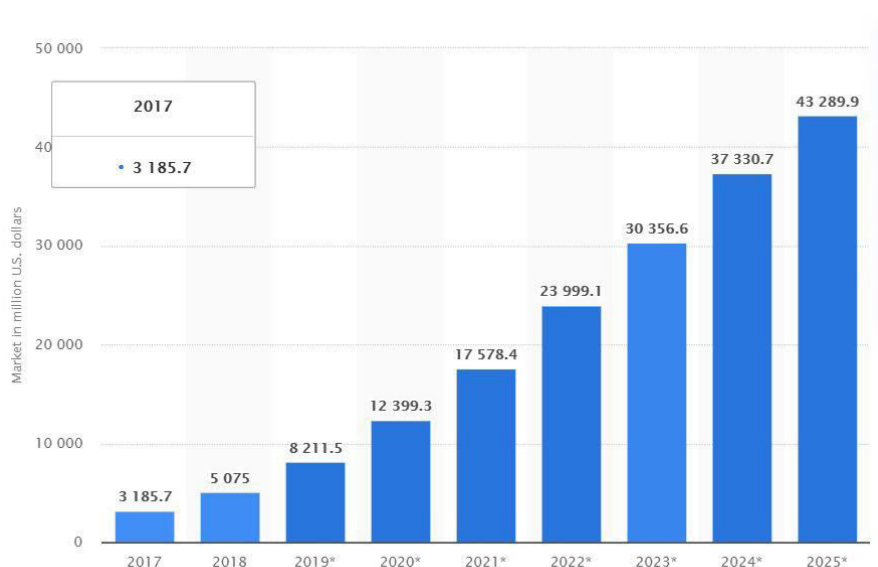
D. Some famous examples of Virtual agents are apple's Siri and amazon's Alexa

All chatbots and virtual agents use Natural language processing to understand customer queries and take appropriate action to fulfill customer's requirements. with machine learning and neural network as time passes they work even better. the next possible applicability of the virtual agent to run long chat sessions with humans which already in the working phase as Sophia.

IV. EXPERIMENTAL RESULTS

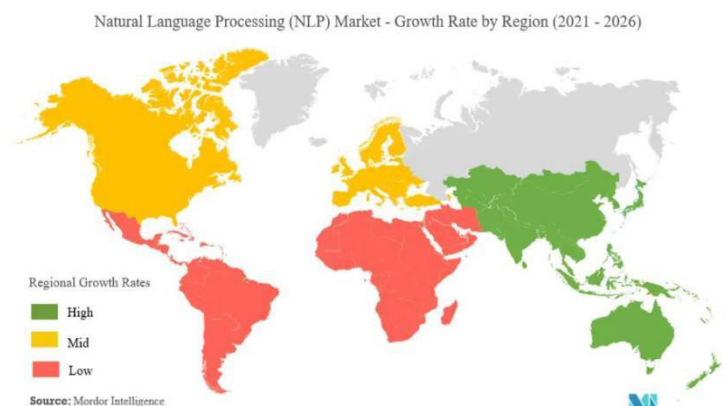
Figures shows the exploration of how false news is impacting social media and making health at risk

Health Volume: 38, Issue: 2, Pages: 143-149, First published: 12 July 2020, DOI: (10.1111/hir.12320)



(a) revenue generated using NLP

Following Figures (b) shows the revenue generated by world class company using natural language processing



(b) NLP growth rate by region world wide

Following Figures (c) shows the natural language processing growth rate world wide

(c) Layers of Internet of Things

V. CONCLUSION

We have seen how natural language processing is changing the world with deep learning and neural networks. the world trend says it is already big market like chatbots, virtual agents and language translator, etc. Linguistic of any language is very important and understanding linguistic had three major methods from past predefine rules, statistical model, neural respectively. deep-learning taking natural language processing to a new level and virtual agents are increasing the quality of the user experience. As time goes It will get better and better with powerful computational power computers from the future.

REFERENCES

- [1] "Transformers: State-of-the-Art Natural Language Processing" Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, Alexander M. Rush
- [2] <https://machinelearningmastery.com/natural-language-processing/>
- [3] Natural Language Processing | Elizabeth D. Liddy Syracuse University, liddy@syr.edu
- [4] <https://www.ibm.com/cloud/learn/natural-language-processing>
- [5] <https://www.mordorintelligence.com/industry-reports/natural-language-processing-market>
- [6] https://en.wikipedia.org/wiki/Natural_language_processing



**Impact Factor:
7.569**



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details