A Review on Data Science: Concepts of Natural Language Processing

Pankaj Kumar Gupta[1], P. K. Tyagi[2], R. K. Agrawal[3]

1.Assistant Professor & Head, BCA Department, DPBS College, Anupshahr
Distt. BulandShahr (UP) India.
2.Professor & Head, Department of Statistics, DPBS College, Anupshahr
Distt. BulandShahr (UP) India.
3.Professor & Head, Department of Mathematics, DPBS College, Anupshahr
Distt. BulandShahr (UP) India.

Abstract-In this paper we will discuss about a challenging topic Natural Language Processing (NLP). NLP is included inside the Data Science but it is having very important role in machine Age. NLP is a research topic of data science which is the concept of analyzing human language data. This allows us to understand and make sense of text data. NLP techniques can design the predictive and interactive models from any language data. It is the technique how we get understood our language and feelings to machines, so that we can get done the desired things by the machines. A small discussion for Artificial Intelligence, Machine Learning, Deep Learning and Data Science is also done here because all the above concepts are connected to each other and they are necessary to express NLP.

Keywords-Artificial Intelligence, Machine Learning, Deep Learning, Data Science, Natural Language Processing.

Introduction

Natural Language Processing (NLP) is included inside the Data Science but it is having very important role in machine Age. NLP is a research topic of data science which is the concept of analyzing human language data. This allows us to understand and make sense of text data. NLP techniques can design the predictive and interactive models from any language data. It is the technique how we get understood our language and feelings to machines, so that we can get done the desired things by the machines. Applications of NLP are discussed here in details. A small discussion for Artificial Intelligence, Machine Learning, Deep Learning and Data Science is also done here because all the above concepts are connected to each other and they are necessary to express NLP.

Fig-1

Understanding the Language of human by the machines, processing that and after that taking the action according to what a human wants is known as NLP. In this way we can get done the things by the machines.
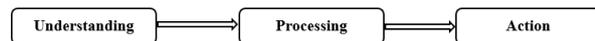

Fig-2

For example:

Let us take the example of Smart TV Remote, we want to search a Ramayana on youtube, there are two ways to do above:

1. Type the Ramayana
2. Press the voice button to on the mic of Remote and speak Ramayana.

In both ways you can see the Ramayana on your TV. But second way uses the Natural Language Processing, in that Remote (machine) understands the language of human, processes and then produces the desired output.
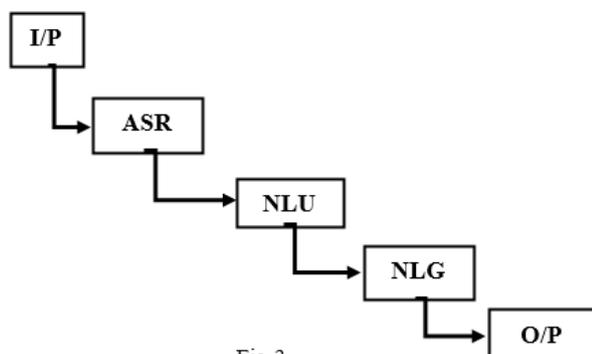
Process of NLP

Fig-3

Fig-4

I/P refers to Input which means you speak Ramayana in above example.

ASR refers to Automatic Speech Recognition means Remote (machine) recognizes your language and converts into text if required.

ASR is a subset of NLP. Automatic speech recognition or speech-to-text or computer speech recognition is a capability which provides a facility to a program to process human speech into a written format. While it is commonly confused with voice recognition, speech recognition focuses on the translation of speech from a verbal format to a text one whereas voice recognition just seeks to identify an individual user's voice.

NLU refers to Natural Language Understanding which means Remote (machine) understands the meaning of what you say to machine.

NLU is also a subset of NLP. It is an important branch of Artificial intelligence that uses computer software to understand input in the form of sentences using text or speech. NLU provides a interface between human & machine.

It is the comprehension of human language such as Hindi, English, French & Spanish for example that allows computers to understand commands without the formalized syntax of computer languages. NLU also allows machines to communicate with human in their own languages.

The main purpose of NLU is to design chat and voice enabled devices that can interact with the a group of human without any supervision. Various IT companies, Like Apple, Amazon, Microsoft, Google and startups have many NLU projects underway.

NLG refers to Natural Language Generation which means According to NLU Remote (machine) generates the output in desired mode what you say to machine.

NLG is also a subset of NLP. NLG allows machines to automatically generate outputs required by natural language text, which is a departure from traditional computer-generated text.

NLG is the software process of transforming data into natural language output using artificial intelligence.

Natural Language Generation enabled software is used to do this by using artificial intelligence models that are powered by machine learning and deep learning to turn numbers into natural language text or speech that humans can understand.

O/P refers to Output which means you are appreciating the Ramayana on TV in above example.

Since all above processing is done through Artificial Intelligence & Data Science, we are going to discuss about Artificial Intelligence(AI) and Data Science(DS) also.

NLP is a field of DS & AI that enables how machines & human interact.
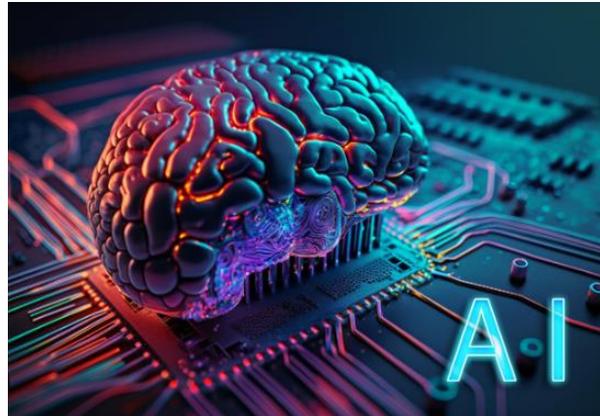
Artificial Intelligence

Fig-5

Artificial Intelligence enables the machine to think without intervention of human. Artificial intelligence is such a field, that combines two fields i.e. computer science and robust datasets so that problem can be solved easily. Machine learning(ML) and Deep learning(DL) are both the subsets of AI. These ML & DL uses AI algorithms which seek to create expert systems which make predictions or classifications for future use based on input data.

AI: Artificial Intelligence
ML: Machine Learning
DL: Deep Learning
DS: Data Science
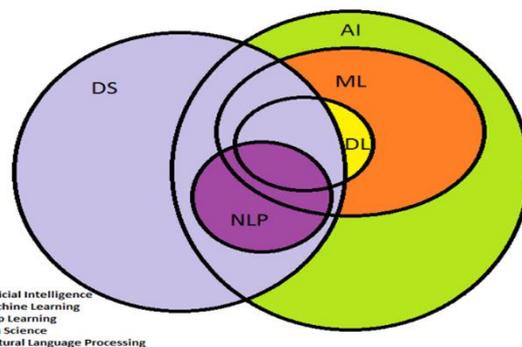NLP: Natural Language Processing

Fig-6

Machine Learning

As shown in above figure (Fig-6), Machine Learning is a subset of Artificial Intelligence. It provides statistical tools to explore & analyze the data. That means ML is a branch of AI and CS that focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. We have three different approaches / techniques:
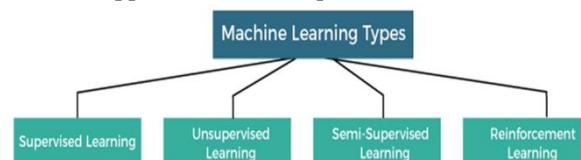

Fig-7

1. Supervised Machine Learning: it works on past data (i.e. Labelled Data).

2. Unsupervised Machine Learning: it works on data Clustering (i.e. grouping of data).
3. Semi supervised Machine Learning: it works on Labelled or data clustering both.
4. Reinforcement Learning: it works on to solve regression problems that are having a relationship between input and output variables.

Deep Learning

As shown in above figure (Fig-6), Deep Learning is a subset of Macine Learning which is again a subset of Artificial Intelligence. Idea behind the Deep Learning is *"Can we make the machine learn like how we are with the help of human brain".* Main idea of DL is *mimic the human brain.*
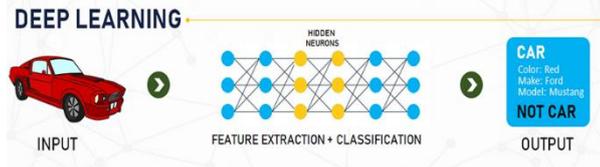

Fig-8

Whereas Machin learning works for same figure (Fig-8)

Fig-9

Deep Learning uses multi neural Network Architecture in which there are three architectures are used:
1. Artificial Neural Network (ANN)
2. Convolutional Neural Network (CNN)
3. Recurrent Neural Network (RNN)

Data Science

Data Science is an Application that is involved in all three above Applications (i.e. AI, ML, DL). Data Science uses some mathematical tools like:
1. Statistics
2. Probability
3. Linear Algebra
4. Differential Calculus

Data science is the study of various data of business organization to extract meaningful insights for business organization. It is an interdisciplinary approach which is used to combine the principles and practices using the fields of computer engineering, artificial intelligence, mathematics & statistics to analyze huge amounts of data. This analysis helps the data scientists to answer the questions like what did happen? why did it happen? When did it happen? How did it happen? what will be happened? How can it be changed? and what can be done with the results?

Working with Data Science
Working of Data Science can be divided into nine following categories:
•   Problem Understanding: The problem must be clear before working on its implementation. Data Scientist must know what to find out, which is crucial to get the right data for the perfect solution.
•   Collecting Right Data: When Data Scientist understands the problem, it's imperative to collect the right data to operate.
•   Data Analysis: After collecting right data scientist must work on wrangling of data, so that It can be fit into desired business. It is known as analysis of data. For this purpose, data scientist performs checking for NULL values, duplicate data, outliers and several other anomalies that are not falling under the desired business data convention. This is also termed as cleaning of data.
•   Data Visualization: when a data scientist cleans & preprocesses the data, it becomes necessary to visualize that data to determine the right features to use for his/her model.
•   Category Encoding: This step is applicable for those instances where the input features are categorical and needed to be transformed into numeric (0,1,2, etc.) to be used in our model as the machine cannot work with categories.
•   Selection of Model: Selecting the right model for a particular problem statement is essential as every model cannot fit in perfectly for every data set.
•   Right Metric Usage: Based on the business domain, the metric that would determine a model's perfectness should be selected.
•   Communicating the findings– The shareholders or businessman may not understand the technically Data Science models. So a data scientist must communicate the findings in simple terms to the business, developing measures to mitigate any foreseen risks.
•   Deployment of model: Once the model is built and the business is satisfied with the findings, the model could be deployed for production and used in the product.

After discussing AI, ML, DL & DS, we again are coming to the point that it is NLP. NLP is used by machines to understand, interpret, manipulate & analyse human's languages. It helps software designers/developers to organize knowledge for performing tasks such as automatic summarization, topic segmentation, translation, speech recognition, Named Entity Recognition (NER) and relationship extraction.

Steps in Natural Language Processing
For performing above tasks various steps are involved that are as follows:

Step 1: Breaking the Paragraph into Sentences

First any given paragraph of written language is broken into the strings. In English and other languages, break the sentences, where you see a full stop/period character. It is the first step for NLP. That means it breaks the paragraph into separate sentences.

Example: Let us take the following paragraph -

The Ramayana is a holy epic of Hindus. This epic expresses the character of Ram. This epic also shows Lord Ram as Purushottam Ram. We should bring the character of Lord Ram in our life.

Sentence Segment produces the following result:

1. *The Ramayana is a holy epic of Hindus.*
2. *This epic expresses the character of Ram.*
3. *This epic also shows Lord Ram as Purushottam Ram.*
4. *We should bring the character of Lord Ram in our life.*

Step 2: Word Tokenization

A tool is used to break the sentence into separate tokens or words, is known as Word Tokenizer.

Example:

DPBS College offers the Courses like BCA, BSc, BA, BCom and BEd etc.

Word Tokenizer breaks the sentence in following tokens as the following result:

"DPBS", "College", "offers", "the", "Courses", "like", "BCA", ",", " BSc", ",", " BA", ",", "BCom", "and", "BEd", "etc", "."

Step 3: Stemming

Stemming word can be used for the normalizing words into its fundamental or root form.

For example: deliberates, deliberated and deliberating, these all words are originated with a single root word "deliberate." The great problem with stemming is that sometimes it produces the root word which may not be having any meaning.

For Example: intelligently, intelligent and intelligence, these all words are originated with a single root word "intelligen." In English, there is no meaning for the word "intelligen".

Step 4: Lemmatization

Actually Lemmatization is also just like to the Stemming. It is used to collect & make group different inflected forms of the word, which is called Lemma. The great difference between Stemming and lemmatization is that it produces the root word, that must have a meaning.

For example: In lemmatization, the words intelligently, intelligent and intelligence has a root word intelligent, that has a dictionary meaning. The word "walk" may be seen as "walks," "walked" or "walking".

Step 5: Identifying Stop Words

In any language there are some stop words that must be identified. In such a way in English, we have a lot of words that are appeared very frequently like "and", "a", "the", and "is". Natural Language Processing pipelines will flag these words as stop words. Stop words might be filtered out before doing any statistical analysis.

Example:

He is a good boy.

You are the Best in Knowledge.

Step 6: Dependency Parsing

Dependency Parsing is used to check the relation among the words in the sentence. Dependency Parsing also refers to examine the dependencies among the words of a sentence for analyzing its syntactic and semantic structure. Depending upon this, a sentence may be broken into no. of tokens. This mechanism is based on the concept that there may be a direct link among every linguistic part of a sentence.

Step 7: Part of Speech(POS) tags

POS stands for parts of speech. Noun, verb, adverb, and Adjective are included in POS. Actually It indicates how a word functions with its meaning and also grammatically within any sentence. A word has one or more parts of speech based on the context in which it is used.

Example: This sentence is spoken by most of the people:

Google something on the Internet.

Here *Google* word is used as a verb, where as it is a proper noun.

Step 8: Named Entity Recognition (NER)

If we want to detect the named entities such as movie name, Location name, Person name and organization name, we use the concept of Named Entity Recognition (NER). Named entities are actually specific terms that can be used to represent real-world objects, like Location, organizations, dates and people. The named entity recognition NLP model gives the stress upon identifying and classifying these named entities to extract useful information from these text data.

For example: Our Honorable Prime minister Narendra modi introduced a Jan Dhan Khata yojna in India.

Ram Mandir is being constructed in Ayodhya which is located in Uttar Pradesh.

Step 9: Chunking

First of all, we collect the small individual piece of information and then group them into a bigger sentence. This concept of collecting and grouping is known as Chunking.

For example: For remembering the phone number, credit card number, aadhar number we use chunking strategy for numbers. In this strategy number is chunked into

small groups so that they can be remembered. Very common way to remember numbers is chucking the number in groups of three or four. Like:

1. any phone number may have 10 digits i.e. 3456542345 could be chunked
   as 345-654-2345.
2. any credit card number may have 12 digits i.e. 345667542345 could be chunked
   as 3456-6754-2345.
3. any aadhar number may have 16 digits i.e. 3456675423451234 could be chunked
   as 3456-6754-2345-1234.

Levels/Stages of NLP:
There are various 5 stages of Natural Language Processing Concept, That are discussed here:

1. Morphological Analysis
2. Syntax Analysis
3. Semantic Analysis
4. Pragmatic Analysis
5. Discourse Analysis

Fig - 10

1. Morphological Analysis
The very first stage of Natural Language Processing is the Morphological analysis. Morphology is actually a study of, how a word is formed and what is the structure of word in the sentence. Here we break the word in the tokens. Token is the smallest unit. So any word if it can be broken, it is broken into parts (tokens).
Like: There is a word "FAITHFULNESS".
It can be broken into three tokens (prefix, stem, and suffix)
1. *FAITH: is the prefix which refers to trust*
2. *FUL: is the stem which refers to filled with something*
3. *NESS: is the suffix which refers to a state of being*

Morphology is also having two meaning:
1. *Roots of the word: that means source (from where it is generated) of word.*
2. *Study of the word: that means full analysis of the word.*
Morphology is also termed as Lexical Analysis.

2. Syntax Analysis
Syntax Analysis is used to ensure for a given piece of text is in correct structure or not. It parses the sentence to check the grammar at the sentence level. Syntactic Analysis also known as parsing.
If it is possible, Part-of-speech is generated from the previous step, then an analyzer of syntax assigns Part-of-speech tags that are based on the sentence structure. For example:
Mango ate the Ram. (which is not correct)
Since in the real world, it does not make any sense, so this sentence will be rejected by the Syntactic analyzer.

Whereas the correct sentence is:
Ram ate the mango
Here:
1. Ram is Noun.(Noun Phrase)
2. Ate is Verb.(Verb Phrase)
3. The & Mango both are determinant. (Noun Phrase)
All the phase in proper way form the sentence.
Syntactic Analysis can be used for checking the word arrangements (noun, pronoun, verb, adverb subject etc.) and shows the relationship among the words.

3. Semantic Analysis
The semantic level of verbal processing deals with the determination of what a sentence really means by relating syntactic features and without ambiguity words with various definitions for any context. This level shows the appropriate interpretation of the sentence meaning, rather than the analysis at the level of individual phrases and words.
Semantic analysis is done for the representation of meaning for any sentence. It mainly gives the emphasis on the literal meaning of words, phrases, and sentences. Consider the sentence: "The mango earns money". Although the sentence is syntactically correct, it doesn't make sense because mango can't earn. Semantic analysis sees for meaning in the sentence which is given by you. It also combines words into phrases and sentences.

4. Pragmatic Analysis
Pragmatic is the fourth level of Natural Language Processing. It is used to discover the intended effect by applying a set of rules that can be characterized cooperative dialogues.
It is an important level of Natural Language Processing. Here it interprets the text given by a user using the data from the previous steps.
Let us consider a sentence, *"Ramesh loves his girlfriend and Mukesh does too"*.
This sentence is much confusing, since it has two meaning like:
1. *Ramesh loves his girlfriend and Mukesh also loves his(Mukesh's) girlfriend.*
2. *Ramesh loves his girlfriend and Mukesh also loves girlfriend of Ramesh.*
This is called as Pragmatic Analysis.
It deals with the use of real-world knowledge and understanding of how this impacts the meaning of what is being communicated. By analyzing the contextual dimension of the documents and queries, a more detailed representation is derived.
When the Information is retried, this level of Natural Language Processing primarily engages query processing and understanding by integrating the user's history and goals as well as the context upon which the

query is being made. Contexts may include time and location.

This level of analysis enables major breakthroughs in Information Retrieval as it facilitates the conversation between the Information Retrieval system and the users, allowing the elicitation of the purpose on which the data is being sought is planned to be used, thereby ensuring that the information retrieval system is fit for purpose.

5. Discourse Analysis

Discourse analysis compliments Pragmatic analysis in some places. Exact meaning of Discourse is "*A group of sentences which are related to each other*". The main difference between Discourse analysis and Pragmatic analysis is that: "*Pragmatic analysis deals with only one sentence whereas Discourse analysis deals with a group of sentences.*"

Discourse analysis works with the group of sentences that proceeds it and also interprets the meaning of the sentences that follow it. Discourse analysis works with the effect of a previous sentence on the current sentence. Like: "*Bhavyanshika is an intelligent student. She invests the maximum time in library.*"

Here, discourse analysis assigns "she" to refer to "Bhavyanshika".

Disclosure Analysis can be used to break A Research paper into following:
Title
Abstract
Introduction
Body
Results
Analysis
Conclusion
References

Ambiguity in Natural Language

In Natural Languages there may be ambiguity. Actually Ambiguity means the ability of having two or more meanings or may be understood in more than one way. Since Natural languages are always ambiguous, then computers may not be able to understand language the way that people do.

For Example:

Mukesh ate the bat.   Sentence 1

I don't even have my cricket bat yet Sentence 2

In the Sentence 1 bat is referred to as a bird where as in Sentence 2 bat is a playing tool.

NLP is used to develop the computational models of aspects of human language processing. Ambiguity may be defined in various levels of NLP. Levels of Ambiguity may be defined as follows:

1. Lexical Ambiguity
2. Syntactic Ambiguity
3. Semantic Ambiguity
4. Anaphoric Ambiguity

5. Pragmatic Ambiguity

Fig - 11

Lexical Ambiguity

If in two or more sentences one word may have two or more meaning then there will be a confusion to understand that word, this is known as Lexical ambiguity.

For example:

Pranav can play the cricket.   *Sentence 1*

Pranav may ask for the can.   *Sentence 2*

Here can word is used in both sentences and at both places same word has different meaning:

In Sentence 1 it is used as a verb where as in sentence 2 it is used as a noun.

In Sentence 1 it shows the ability of Pranav to play cricket where as in sentence 2 it shows the desire of Pranav for cold drink.

Syntactic Ambiguity

If one sentence has two or more meaning then there will be a confusion to understand that sentence, this is known as Syntactic ambiguity. It is also known as structural ambiguity, since we find different structure meaning for one sentence.

For example:

Nandini saw Pranav with the Binoculars.

This sentence is ambiguous; it may be defined in two ways that are as follows:

Nandini saw Pranav carrying Binoculars.   *Sentence 1*

Nandini saw Pranav through Binoculars.   *Sentence 2*

Semantic Ambiguity

Semantic ambiguity occurs if the meaning of the words can be misinterpreted. Whereas the syntax and the meanings of the individual words have been resolved, there may two or more ways of reading the sentence.

For example:

Nandini loves her mother and Bhavyanshika does too.

In this sentence meaning of every word is very clear although this sentence is ambiguous; it may be defined in two ways that are as follows:

Nandini loves her mother and Bhavyanshika also loves her mother.   *Sentence 1*

Nandini and Bhavyanshika both love Nandini's mother. *Sentence 2*

Nandini and Bhavyanshika may be real sisters and both love their mother *Sentence 3*

(i.e. Single Lady)

Semantic ambiguities still appear in front of human from the fact that generally a computer is not in a position to distinguishing what is logical from what is not. Semantic ambiguity appears if a sentence is formed with ambiguous words or phrases.

Anaphoric Ambiguity

Anaphoric ambiguity arises due to anaphora. Now anaphora means a noun is replaced by pronoun. When a

pronoun replaces a noun and this pronoun becomes a confusion, this becomes anaphoric ambiguity.

For example:

Pranav's School is in a Society. It is very dirty.

In this sentence the word "It", which is pronoun makes this sentence ambiguous; here "it" word may be used for two things either for school or for society.

Like:

Pranav's School is in a Society and School is very dirty. *Sentence 1*

Pranav's School is in a Society and Society is very dirty. *Sentence 2*

Pragmatic ambiguity

Pragmatic ambiguity arises due to the specificity of statement. If statement is not specific, and the it does not provide the information required to clarify the statement. That means information is missing, and must be inferred.

For example:

She hates you too.

This sentence can be explained like:

She hates you (just like you hate her) *Sentence 1*

She hates you (just like some other person does) *Sentence 2*

She hates you (as you hate some other person) *Sentence 3*

She hates you (as well as she hates me) *Sentence 4*

Above we have discussed about NLP and many things about NLP. Now we will discuss about the applications of NLP.

Applications of Natural Language Processing

Today's Era in the Age of Machines, Natural language processing has become very important part of Machine Learning. Here it is a discussion about the applications of Natural Language Processing.

ChatGPT

ChatGPT stands for Chat Generative Pre-trained Transformer. ChatGPT is an AI chatbot that uses NLP for creating dialogue of conversation just like human. The language model can respond to questions and compose various written content that includes various things like post of social media , essays, emails, articles and emails. ChatGPT works through its Generative Pre-trained Transformer, that uses some pre-defined algorithms to find what a user wants. ChatGPT used the GPT-3 vast language model, which is a neural network machine learning model and the 3rd generation of Generative Pre-trained Transformer. The transformer fetches from a significant amount of data to formulate a response.

Chatbots

One important application of NLP is Chatbot, it is a form of artificial intelligence which is codded in any language

to interact with humans in such a way that it speaks like a human itself. It also depends upon the complexity of the Chatbot, it may either just respond to specific keywords or it can even hold full conversations that make it tough to distinguish them from humans. It is created with the help of Machine Learning as well as Natural Language Processing, that means it understands the complexity of the English language and finds the actual meaning of the sentence and it can also learn from all conversations with human and has become better with time. Chatbot works in two simple steps. First, it identifies the meaning of the question that is asked and collects all the information from the user that may be needed to answer the question. Then it answers the question as required.

The AI chatbots that are in use, as follows:

- Sales and marketing: ChatSpot
- Leader: ChatGPT
- Personal coaching: Pi
- Runner-up: Google Bard
- Messaging: Personal AI
- Open source: HuggingChat
- Content writing: Jasper Chat, Chat by Copy.ai, ChatSonic, ZenoChat
- Building your own chatbot: Zapier AI Chatbot
- Up-and-coming: Claude, Khanmingo
- Searching the web: Microsoft Bing AI, Perplexity, YouChat, KoalaChat
- Tinkering: OpenAI playground, Poe, DeepAI Chat
- Coding auto-complete: GitHub Copilot, Amazon CodeWhisperer
- Fun: Character.AI
- Social media: Snapchat My AI

Voice or Virtual Assistants

Another important application of NLP is Voice Assistant that is very smart in use. Now a days voice assistants are in all rages, it may be Google Assistant, Siri or Alexa everybody uses one of these in making calls, surfing internet, scheduling meetings, placing reminders and also in setting alarms etc. These voice assistants made our life much easier. But how do they work? Voice assistants use a very complicated combination of natural language processing, natural language understanding, and also speech recognition, for understanding what a person is saying and then act on same. The real goal of voice assistant is to be used as a bridge between human and the internet. It provides all types of services based on just interaction of voice. However, a voice assistant is a little far from that goal seeing such as Siri, Siri still can't understand sometimes what you are saying.

The Voice or Virtual assistants that are in daily use, as follows:

Name	Amazon Alexa	Google Assistant	Siri
Type	Best for Device Compatibility	Best at Responding	Most Popular Mobile Voice Assistant
Voice-match technology	✓	✓	✓
'Delete recording' options	Voice and Settings	Settings	Settings (limited)
Instant translation technology	✓	✓	✓
IoT compatible brands	7,400+	1,000+	50+
Languages supported	3	4	21

Fig - 12

Autocomplete feature in Search Engines
Did you ever notice that how search engines tend to guess what anybody is typing and that search engine automatically completes your sentences?
For example, when someone types "movie" in Google Search bar of Google search engine, he or she may get further suggestions like:
- Moviesflix
- Movies4u
- Moviesmod
- Moviesmad
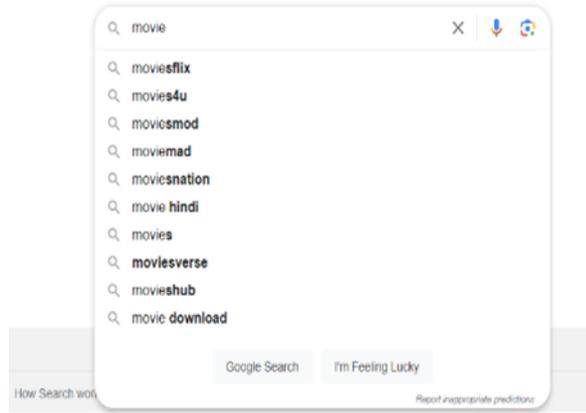- Moviesnation
- Movieshindi
- Movies
- Moviesverse etc.

Fig - 13

These all suggestions are provided using autocomplete feature that uses Natural Language Processing to guess what anybody wants to ask. Search engine uses its vast data sets to analyze what it's searchers are probably typing when they enter any particular word and then it suggests the various possibilities. Almost Search engines use Natural Language Processing for making the sense of those words and how these words are interconnected to form different sentences.

Sentiment Analysis
In today's era we see all the world is on social media. We also see various companies can use sentiment analysis to understand about the feeling of a user for a particular topic or product etc. Companies can use natural language processing, computational linguistics and text analysis etc. for understanding the common sentiment of the users for their products and services and find out about the sentiments whether it is good, bad or neutral. The Companies can use sentiment analysis tool in a lot of ways such as to find out the emotions of their audience, to understand product reviews, to gauge their brand sentiment etc. we see not only private companies but also governments use sentiment analysis to find popular opinion and also catch out any threats to the security of the nation.

Language Translator
There are many cases we want to translate a text from English to Hindi but we do not know about Hindi Language? This is possible to translate the same by using a tool Google Translate. If we say that: *Is it accurate?* The answer is: *It's not exactly 100% accurate*, although it is still a great tool to convert text from one language to another. All the Translation tools like Google Translate and others also use the sequence to sequence modeling, which is a technique in Natural Language Processing. This technique allows the algorithm to convert a sequence of words from one language to another which is in translation. Few years before the language translators used to use the Statistical machine translation

(SMT) which meant they analyzed millions of documents that were already translated from one language to another (English to Hindi in this case or vice versa) and then saw for the general patterns and basic vocabulary of that language. However, Sequence to sequence modeling was accurate as compared to this technique.

Checkers of Grammar & Spelling

When we write professional reports or even make some assignments for our lectures there is a very important thing which is called as Grammar and spelling becomes a very important factor. It makes the first and last impression of our task in others mind. If we have a major error in our report and assignments even, we may be fired or failed! This is why grammar and spell checkers software have become very important tool for any professional writer. These grammar & spell checkers may not only correct the grammatical errors and check spellings also. Same may also suggest better synonyms and improve the overall readability of our literature. Again a question arises how do they work? Answer is: they utilize natural language processing to provide the best possible piece of writing! The NLP algorithm uses millions of sentences to understand the correct format. So, they can suggest the correct tense, verb, a clear sentence structure & better synonym than what we have written. In the market various checkers are available some of the most popular grammar checkers that use NLP they include WhiteSmoke, Grammarly, & ProWritingAid, etc.

Filtering & Email Classifying

For professional communication we use Electronic mails (Email). It is now a days become very important. In real we see, all of us get hundreds of promotional Emails, it may be possible we don't want to read those emails. When we open our inbox, our emails are automatically divided into 3 categories like: Primary, Social, and Promotions that means we never have to open the Promotional Tab, But it is a question that how does this work? Email services use natural language processing for identifying the contents of each Email with text classification so that it can be put inside the corresponding tab. In real the method of such classification is not perfect since there are some Promotional newsletters appears in Primary tab, but it is better than nothing. In the advanced cases, various companies use specialty anti-virus software with natural language processing to scan the Emails and see if there are any patterns and phrases that may indicate a phishing attempt on the employees.

When classifying the emails, a natural language processing model reads the incoming email text and understands the meaning of it. After that a machine learning model learns what to do with the meaning that NLP understands which is based upon the way you train

this model, it intelligently sorts the emails into its categories.

Just assume that it is our first day in a mailroom (like post office). We already know how to read the envelopes (like NLP) but here we need to categorize which floor or office all the departments are in (like ML training). Once we understand that, we can organize and route all the incoming mail (like envelops).

REFERENCE

[1] https://www.upgrad.com
[2] https://www.study.com/
[3] https://www.corpnce.com
[4] https://www.techtarget.com
[5] https://www.ibm.com
[6] https://automationhero.ai
[7] https://www.techtarget.com
[8] https://byteiota.com
[9] https://www.javatpoint.com
[10] https://medium.com
[11] https://zapier.com
[12] https://www.zdnet.com
[13] http://www.cis.strath.ac.uk
[14] http://swdevelopers.tripod.com
[15] http://michealaxelsen.com
[16] http://www.coli.uni-saarland.de
[17] Pushpak Bhattacharyya, 'Natural Language Processing: A Perspective fromComputation in Presence of Ambiguity, ResourceConstraint and Multilinguality', CSI Journal of Computing | Vol. 1 • No. 2, 2012.
[18] Gobinda G Chowdhury, 'Natural Language Processing',
[19] TanveerSiddiqui, U S Tiwary, 'Natural Language Processing and Information Retrieval', Oxford University Press.
[20] Speech and Language Processing by Daniel Jurafsky and James H. Martin
[21] Deep Learning for Natural Language Processing by Stephan Raaijmakers
[22] Foundations of Statistical Natural Language Processing by Christopher D. Manning and Hinrich Schütze