

Overview of NLP

What is NLP?

Natural Language Processing enables machines to understand and interpret human language as it is spoken and written. Natural language processing equips computers to receive written or audio language, analyze it and interpret it in a way that a computer would understand. Similar to a human brain, machines can then process human language inputs.

Relationship between AI and NLP

NLP utilizes artificial intelligence (AI), machine learning (ML), and computational linguistics to interpret human language. AI is used to take inputs provided in its raw human form and convert them into a language that a machine understands. Natural language processing is a subset of AI. When we take AI with a focus on human linguistics, we get NLP.

Natural Language Understanding v/s Natural Language Generation

Natural language understanding and natural language generation are both subsets of natural language processing. Natural language understanding uses semantic and syntactic analysis of speech and text to determine the meaning of a sentence. The syntactic analysis relates to the grammatical structure of sentences. Semantic analysis works on making sense as to what is the meaning of the sentence.

Natural language understanding mimics the human ability to do such analysis. A human can distinguish between homonyms and homophones. Such sentiment analysis is often applied by companies to cluster positive and negative trends in customer comments on social media. This can further help companies to address issues in products at an early stage.

Natural Language Generation is the process of replicating the human ability to write. NLG consumes certain inputs and based on them produces human language text responses. Such text responses could be in English or other languages. It can be utilized to create summaries of documents without affecting the integrity of information within the document. It has evolved with the application of hidden Markov chains, recurrent neural networks, and transformers, enabling more dynamic text generation in real time.

Modern NLP applications

1. A recently widely used application of NLP is predictive texts. Nowadays, while we write emails or texts most applications attempt to interpret what we are writing and provide a prediction of the remaining words in the sentence we are trying. Predictive texts understand users' language quirks and learn and adapt accordingly. Another example of predictive text is autocorrect. When we type words in our google search, it automatically attempts to correct typing errors. Google searches also understand contexts and suggest what it thinks might be correct in the form of 'Did you mean "abc" instead of "xyz"?'

2. Today social media connects many of us from different parts of the globe. Many-a-times we see people commenting or posting in their native languages. NLP is applied in such cases to detect foreign languages on a page and allows the user to view translations of the same. Another such example of

language detection and translation is seen in google posts or results which are written in foreign languages.

3. NLP is used by companies to perform analysis on text from different platforms. Such analysis can help with understanding customer interactions like social media comments or brand mentions to get a sense of people's reactions to products. This can further be utilized to drive marketing campaigns for products and services by companies.

4. Another common application of NLP we see these days is digital phone calls. How many times have we called customer service numbers and interacted with bots who ask us questions, understand and interpret them to get our questions and concerns addressed? NLP systems are also used to schedule appointments, callbacks, etc. while a human touch is vital to resolve most intricate communication issues, NLP improves our lives as a digital assistant which can handle by automating most basic of human tasks.

5. A widely used application of NLP are the voice assistants like Alexa, Siri and Google Home. These smart devices understand and interpret what we ask and perform the tasks for us. The most mundane tasks like switching on lights and fans are now automated by our commands with these devices. These devices are now used for placing online orders and getting traffic updates via verbal input.

Approaches to NLP

There are 3 main approaches to NLP - The Rule-based approach, the Statistical and Probabilistic approach, and the Deep Learning approach.

1. Rule-based approach

This approach is based on pre-written rules which provide automatization to mundane tasks. A very good example of a rule-based approach is regular expressions which help with pattern matching. In a rule-based approach, there are predefined rules that determine the capabilities of a system which may cause limitations based on the rule creator's knowledge. Rule-based NLP also works well for systems that have a dictionary and look for spelling issues. Such systems would be restricted to detecting typos and won't be able to perform advanced tasks like suggesting alternate words or phrases. Rules are written by people who have a good grasp of the system and in-depth domain knowledge.

2. Statistical and Probabilistic approach

These are statistical methods for NLP that involve probabilistic modeling, and AI algorithms to solve problems. Unlike rule-based techniques, this technique utilizes ML to analyze text and look for patterns. Such approaches usually include training data, training a model based on parameters followed by fitting on test data. Such systems use features to perceive text apart from the corpus of training data. These features are different characteristics like word count, word frequency, etc. which help the system in understanding which words matter in the text. Word clouds are a very good example of the representation of prominent words. Word clouds are most commonly used to show live votes during conferences to visually represent frequent words with greater prominence.

3. Deep Learning approach

A deep learning network comprises several layers of neural networks. Neural networks simulate the way human brains work. Multiple interconnected layers perform various mathematical computations and input layers which receive data and communicate them. These are powerful mechanisms that take

raw input and work on learning important features. They constitute a very large corpus of training data. A pure neural network has 3 layers - an input layer, a hidden layer, and an output layer. Methods of traditional approaches are used to improve deep learning approaches. Deep learning can be used for performing a range of tasks for example analyzing data, converting speech to text, implementing facial recognition software, chatbots, weather prediction software, etc.

Personal interest in NLP

I have found natural language processing to be a very interesting subject and has numerous applications that we use multiple times each day in so many of our applications and home devices. It has become an integral part of our life with many more areas that could be explored further in the coming years. My interest is in sentiment analysis. By sentiment analysis, I mean understanding an overall emotion towards a certain topic. The major sentiment categories would be negative and positive. I can perceive a lot of applications of such sentiment analysis in varied fields. For example, it can be used by companies to get early feedback on product soft launches to get a quick idea of how their product is doing. Devise any alternate strategies to improve or market their products. Sentiment analysis can also be used for improvising marketing strategies. When we want to proceed with such analysis, one of the initial steps would be to use text data processing methods to eliminate gibberish data and retain useful information. Following this, the next step would be to analyze significant words that represent the sentiment of the reviewer. Strategies like TF-IDF (term frequency-inverse document frequency) can be used for feature extraction. Wherein, the key idea behind TF-IDF is that the words that occur more frequently in one document and less frequently in other documents should be given more importance as they are more important for classification. While NLP has endless domains to explore, I am particularly interested in the sentiment analysis domain of NLP.