

Neal Gandhi

NDG190002

1/23/2023

CS4395.001

NLP History & Overview

Natural Language Processing (NLP) is a branch of computer science that involves computational understanding and interpretation of human language. Types of understanding can involve sentiment, translation, summarization, and can extend as far as being able to have full regular conversations with an artificial intelligence. NLP is considered a subfield of AI that mainly deals with language models, and the interaction between computers and human languages. Its main focus is to create algorithms and models that only apply to generation, understanding, and interpretation of the human language. On the other hand, AI is a much broader term that incorporates more than just computational understanding of the human language. Other fields can include, but are not limited to, computer vision, robotics, and certain areas of machine learning.

Natural Language Processing can be broken up into two main categories: Natural Language Understanding and Natural Language Generation. NLU can be described as the computational ability to analyze text for tone and other non-verbal cues based on the language being used. A key example of this would be sentiment analysis, which is used to determine the overall sentiment of a phrase. NLG is the process of computationally creating phrases and sentences that read as if they were written by a human, with proper grammar and syntax. Generation of natural sounding language is the basis of several chat bots that use NLP. The biggest application of NLP currently is ChatGPT, an AI chatbot that has near humanistic levels of conversational skills, and ResumeWorded, an NLP model that optimizes student's resumes by improving word choice and shifting the overall tone to a more professional standing.

There are three main techniques when approaching NLP; a rules-based approach, a statistical and probabilistic approach, and deep learning, each with increasing sophistication thanks to the advancements in computing power.

The rules-based approach was the earliest approach to NLP, and used basic syntax and grammar rules to generate natural sounding language. These rules could also be used to check if

the output sentences were grammatically correct, creating a rough version of NLP that has the ability to generate human language. This version of NLP has its limits due to the complex nature of the grammatical rules that language follows, it becomes difficult to encapsulate all of the right rules; however, its simplicity allows it to solve many lower-level problems without the need to train a neural network.

Statistical and probabilistic approaches were developed on the mathematical approaches to text analysis in the late 1980s, which began to use probability to predict the next strings of words that might be input. This approach falls more in line with the typical approach to machine learning algorithms, which are used to solve NLP problems. The most expected usage of this kind of approach would be search bar word prediction and, more recently, autofill within different email providers.

The last approach, which is the most computationally complex of the three, is the deep learning approach. This approach only became viable due to the increase in processing power provided by advancing GPU technology and the availability of cloud computing. This usually involves the more complex machine learning algorithms, such as LSTMs, RNNs, CNNs, and extensions of the classic neural network model. These require a large amount of data to be able to train, but when extensively trained, are able to produce near human-like text.

I personally think NLP has a lot of applications involving customer support and advancing search engines to be able to correctly predict and return data that is relevant to the search terms. It is a subfield of AI that has always interested me, as I enjoy the study of linguistics and NLP perfectly encompasses it with computer science. I would love to learn more on it's application to build predictive search bars, or even chatbots to integrate into the numerous website projects I create, to take each to the next level.