

Soham Mukherjee

Sxm180113

CS 4395

1/27/2023

Overview of NLP

NLP, or Natural Language Processing, concerns the intersection of linguistics and computer science. Much like how humans understand languages along with their idiosyncrasies, computers seek to understand languages through artificial intelligence. This doesn't include just memorizing sentences, but also understanding the relationships between words and sentences. NLP is concerned with being able to accurately extract information and ideas from given passages.

NLP is an application of AI, or artificial intelligence. AI is primarily concerned with simulating how humans comprehend new subjects, and works on simulating them through a number of methods. NLP delves more specifically into the human language, and deals with how computers can translate and more so, comprehend, human language. As a result, NLP can be seen as an application of AI in the field of linguistics.

Natural language understanding and Natural language generation both take in a linguistic dataset as input. However, their outputs differ – while Natural language understanding returns intended meanings as outputs, Natural language generation returns sentences or phrases as output given the syntax of its input.

Autocomplete is one example of NLP. For instance, when you're writing out an email to the team or you're writing a lengthy text, Autocomplete will try to fill in the rest of the sentence if it's a phrase that has been used before (ie; "per my <last email>" or "Hope you <are doing well>"). Another example is smart assistants; for instance, on ecommerce sites, if a user has an issue with their purchase, keywords like "refund" or "defective" allow the website's chatbot to correctly respond. A final example would be document analysis; many research papers often have text that may not be immediately understandable at first due to requiring a considerable amount of knowledge and groundwork prior to reading them. With NLP, the information can be properly organized and categorized to make it more palatable to the reader.

There are three main approaches to NLP: rule-based, statistical/probabilistic, and deep learning. Rule-based is the oldest of the three, and relies on preset rules using regular expressions and context-free grammars. This type of NLP is static, which makes it vulnerable to the dynamic nature of language itself since language continuously evolves. An example of rule-based NLP would be the Eliza chatbot in the 1960s.

The statistical/probabilistic approach involves counting words and associating their probabilities with synonyms that may mean the same thing. Decision trees are a good example of statistical/probabilistic approach, as by gathering the sentiment from given words, the probability of words that similarly match the sentiment will be higher than the rest.

Finally, there is deep learning. Deep learning approach involves using huge amounts of data, and various machine learning models, to make human-sounding interactions possible. An example would be transcribing audio to subtitles, which involves using a significant amount of data to determine how words sound to transcribe them accurately.

I'm not well versed in Artificial Intelligence, but I am an avid writer, both technically and personally. I'm always seeking to learn more about how to improve my writing and tools such as ChatGPT have helped me identify and fix problems that I didn't even know existed. Such powerful tools have opened my eyes to the exciting possibilities that AI has to offer, and NLP is something that is at the intersection of two of my favorite subjects – computer science and writing.

I would probably like to learn more about NLP for personal projects; I recently got into making Discord bots on Python and I'm in the process of creating a custom Discord bot for my writing server; for the MVP, the bot guides users on the Discord, and will provide lore info on characters/locations if prompted. However, my stretch goal would be for it to talk dynamically to users using NLP. That would be interesting.