

Overview of NLP

NLP in my own words: Natural language processing, or NLP for short, is the practice of writing code that is able to process natural languages. Diving into that definition a bit more, “natural languages” are those languages spoken by human beings such as English or Spanish, as opposed to formal languages such as Java or Python. “Process” is a bit more difficult to define, at least for me. I would like to use words such as “understand” or “interpret” but those words may give off the wrong impression. NLP code may or may not fully simulate AI who can understand English like a human. Rather, it is code that is able to take in data (in this case, text) and be able to extract useful information from that data (such as understanding the sentiment)

Relationship between AI and NLP: AI and NLP are closely linked together. Besides the first approach toward NLP of a rules-based method back in the 1960s, every approach to NLP has involved AI to some extent. More specifically, NLP has used the techniques of machine learning to be able to scale up to the complexity of real human languages

Difference Between Natural Language Understanding and Generation: Natural Language Understanding is being able to take in text and extract information out of it like a human would such as being able to understand if sarcasm is present. Natural Language Generation is being able to create text that is as similar to regular human speech as possible

Some examples of NLP Applications: Siri, Google Assistance, Alexa, voice recognition software in general, and automatic closed captioning for those with hearing disabilities are all a few examples of NLP applications today

Three Main Approaches to NLP

The first approach taken towards NLP was the rules-based approach. As the name would suggest, the rules-based approach was focused on having a set of rules the code would follow, with the desire being that the more rules were added, the more realistic and human-like the NLP would become. Some examples include basic rulesets like changing plural words into singular or lowercasing everything. Spell-check is a common example in which words that aren't on a certain word list are considered misspelled and we may even write code to automatically correct the misspelled word into the closest correct spelling on the word list. While easy to get one's head around, the rules-based approach just simply isn't feasible. With just how wildly complex human language is, the set of rules that could accurately capture all the various nuances and so on would be far too big to expect any human or set of humans to write.

The second set of approaches taken towards NLP were the statistical and probabilistic approaches. In these approaches, examples learn from a body of data and text. Text was analyzed to reveal the frequency of occurrence for various words and phrases. This seemingly basic piece of information can be leveraged to do predictive text and better translations between languages. The downside of this approach compared to the rule-based approach is that statistical and probabilistic approaches require a decent amount of data to learn from as well as good processing power.

The third set of approaches used most recently is deep learning. With the advent of more and more data and processing power, algorithms involving deep learning have become more common-place. Deep learning has allowed for advancement in language translation, generation, understanding and much more. Some example of deep learning algorithms involve recurrent neural networks and convolutional neural networks. However, one of the major problems with deep learning as of late is the hype surrounding it.

Personal Interest In NLP

I'm a believer in the Lord Jesus Christ. As such, I've always been interested in language as a whole and specially, ancient Hebrew and Koine Greek, the two major languages the Bible was written in (aside for a little bit in the Old Testament written in Aramaic). Learning of the origin and spreading out of the language always involves really fascinating history which is always fun to read. To see that the next advance in mankind's relationship to language is through computers is exciting, to say the least. This past summer, I worked as an intern for Fidelity and one of our potential projects has a spam-recognition bot. We decided to go for a different project but that was something that really piqued my interest so I'm hoping to learn the techniques to make a spam-recognition bot on my own.