Nathan Tompkins

Overview of NLP

Often times, we want to convert a sentence into a readable format or deconstruct it. This is where Natural Language Processing comes in. NLP provides techniques to take a given sentence and turn it into a format that a computer can understand. This is often combined with Artificial Intelligence to try and understand the sentiment of the sentence, predict additional sentences, and add on to the sentence in ways that usually only a human could do. Natural Language Understanding is the ability to take in a sentence and either understand the meaning, or understand the sentiment of the sentence. Natural Language Generation is the ability to generate sentences, whether they are based on a previous sentence or just context. One example of a program using NLP is ChatGPT. This program takes in text from a user and makes a surprisingly accurate response that can provide very useful information. Another is midjourney. It takes user input, uses NLP to understand the sentence, and creates a picture from the text.

There are three approaches to NLP; The first is the Rules-Based Approach. This is the first technique used to NLP. It is a very simple approach, in that it uses a set of rules to make and understand sentences. It is not completely accurate and can use default statements that use regular expressions to respond.

The second approach to NLP is probabilistic approaches which makes a bit more sense than the prior method. Instead of the sentences being understood through strict rules, the probabilistic approach determines what is *most likely*. This is what most classical machine learning algorithms use and is generally better on smaller datasets.

The final approach to NLP is Deep Learning. With the last technique being used for smaller datasets, we use Deep Learning for larger datasets. This has been used for techniques like Convolutional Neural Networks. This technique is usually pretty expensive and requires lots of hardware and/or cloud computing. It is not unnatural to use all three methods when working on a project using NLP.

One reason I am very interested in NLP is that I am a Data Science major. Data scientists use a lot of AI and methods such as RNN and CNN. The thing that really interests me though is text vs image classification. I have learned that text classification can actually be harder than image classification. This is because there is a lot of nuance and sarcasm in languages; as well as idioms. Recognizing visual patterns is easier for a computer than recognizing subtleties in language. I hope to learn all the techniques used to classify text.