Saige Wright

Overview of NLP

Natural Language Processing (NLP) is a study of artificial intelligence specifically focused on human linguistics, using computer science tools, computers understand and analyze a large quantity of natural human language. NLP is a branch or subset of artificial intelligence (AI). Natural language understanding (NLU) and natural language generation (NLG) are both subsets of NLP. NLU focuses on the comprehension of human language to a computer us syntactic and sentiment tools to understand what a sentence meant and its context. While NLG gives the computer the capability to produce a human language-based response. Some modern NLP applications range from email filters and search results to chatbots and data analytics. There are three main approaches to NLP.

In the 1960s, the rules-based approach to NLP was the most predominant. The rules-based approach consisted of a given set of rules that the NLP system would follow and use to classify the language being analyzed. Unfortunately, this approach was only as good as the rules provided. An example of a rules-based approach in NLP would be spell check. In spell check, the rules would be considered the dictionary consisting of the correct spelling of each real word, along with rules regarding contractions and such being applied to correct the spelling of something.

In the 1980s, the statistical and probabilistic approach to NLP came to the forefront. This approach explored applying statistical and probabilistic algorithms in a machine-learning format consisting of quantitative approaches to linear algebra, automated language processing, probabilistic modeling, and information theory. An example of a statistical and probabilistic

approach in NLP is the learning that happens from a corpus of data. After reading in a corpus we can apply machine learning algorithms to find word and word-pair frequencies.

Finally, in the 2010s, the deep learning approach to NLP made its first appearance. The deep learning approach is comprised of applying various deep learning models to text. This approach improved and automated the text analytics functions and NLP features, essentially giving our data meaning. Through this deep learning process, the unstructured raw text could be turned into readable, classified, and useful data. An example of deep learning is language generation.

My personal interest in NLP stems from the final report I wrote in ECS 3390. I compared computer languages and human languages and then explored how learning each language affects the human brain. As I got deep into research I made my way to the development of human linguistics and thought about it from a computer perspective. Naturally, this class sparked my interest. Ideally, I would like to do NLP and machine learning in a more professional environment because I really enjoy giving data meaning.