Kinsey Mellon

Overview of NLP

NLP, short for natural language processing, is a broad term that includes many different things. In general NLP uses algorithms to listen/recognize human language, identify different words or phrases in different text formats and many other tasks. NLP is a subfield of linguistics, computer science and Artificial Intelligence. Much like machine learning, NLP branches off from Artificial Intelligence and many times general AI techniques can be used in tandem with NLP. There are many modern-day examples of NLP all around us including OK Google, spam email filters, and sentiment analysis.

There are two general processes going on in human-to-human dialogue in NLP, these are natural language understanding and natural language generation. Natural language understanding means that everyone involved in the conversation or interaction understand what is being said, while natural language generation is simply the development of what is being said. In conversations these two processes hold different purposes but continually feed off each other.

There are three main approaches to NLP that we will go over in this course, which are rules-based, statistical and probabilistic, and deep learning. Starting off with the rules-based approach it includes the oldest techniques in NLP. This approach dominated the field during the 1960's and follows a list of rules to analyze the data. Some examples of the rules-based approach are context-free grammar which can generate syntactically correct sentences or check the grammar of sentences, spell check, and the Eliza chatbot which tried to copy a talk therapist. The big problem with this approach is that it couldn't match the complexity of human language.

The second group of techniques are the statistical and probabilistic approaches. These became popular in the mid to late 1980's and involves approaching language from a

mathematical view. Language models can be created by finding probabilities of words and sequences and can be helpful in machine translation systems and for predicting text. Also, the common machine learning algorithms can be classified under this approach since they learn by statistical and probabilistic means. The major issue with this approach is that a lot of data and good processing power is needed.

The final approach discussed is deep learning. This approach includes algorithms such as recurrent neural networks, convolution neural networks, and LSTMs. The more readily available data and higher processing power allowed for better results in language translation, generation, and understanding, with the hopes to eventually make human-sounding interactions by taking into account more than what the previous two approaches do. Some problems with the deep learning approach are the over hype going on in the field right now surrounding it and the even bigger need for data and processing than the previous approach.

It's hard to say when my personal interest in NLP started but to summarize, being a Cognitive science major, I've always been intrigued by all the cognitive processes our brains can do, especially the language aspects. After studying the neuroscience behind our language processes, it only seemed natural to move into the other areas I'm interested in which include computer science and Artificial Intelligence and NLP fit perfectly to combine all those areas. I would love to learn more about NLP and gain a better, more complete understanding of what it is and the different things that can be done with it since at the moment I only have a general overview.