Overview of NLP

Natural Language Processing is a field of artificial intelligence that aims to enable computers to read, understand, and use human language in a meaningful way, including analysis and text generation. It employs techniques from rules-based processing to deep learning. NLP is considered a subfield of AI, being one of the many fields contained within it. Natural language understanding is the ability of a computer to receive text and understand its full meaning, not just the meaning of individual words. Natural language generation, on the other hand, is the ability to create language based on given data or context, such as what is understood through NLU. The two are different and together compose NLP. Modern examples of NLP applications range from spell check and autocomplete to chatbots, topic and sentiment analysis, and machine translation.

The first approach to NLP was a rules-based one. This involves creating a set of rules that the computer uses to process text. This does not require large amounts of preexisting data and has applications like spell check, context-free grammars, and regular expressions. This method is effective in more straightforward use cases where efficiency is desired. The problem is that this approach has limited scope and cannot deal with the complex and ambiguous nature of human language.

The second approach is traditional machine learning, using statistical and probabilistic methods. This includes supervised and unsupervised models like regression, classification, and clustering. This is the known machine learning approach of training a model with specific features on data with parameters and then fitting the model on test data. It is good for things like semantic filling, next word, and classification. This allows for more advanced processing and can improve over time, but requires good pre-existing data due to its statistical nature and processing power.

The third approach to NLP is using deep learning. This is an advanced form of machine learning employing neural networks. This method is able to learn to understand language at a high level and has provided superior results in understanding, generation, and translation. It is best for more complicated tasks like speech recognition and machine translation. It is the most popular and current method but requires a very large amount of data and lots of processing power.

I am interested in NLP because I am interested in AI and linguistics. I want to learn more about how it is actually done so I can both apply it in any personal projects and understand if I want to pursue it further. I would like to be able to actually be able to do some practical NLP by the end of this course.