Ikhlaq Ahmad

Dr. Mazidi

CS 4395

Jan 28, 2023

Overview of Natural Language Processing (NLP)

A computer program's ability to comprehend natural language, or human language as it is spoken and written, is known as natural language processing (NLP). It is a part of the umbrella term artificial intelligence (AI). NLP has been around for decades and has roots in linguistics. It has numerous applications in a range of industries, including robotics, aviation, and Medical.

Natural language processing is a subset of Artificial intelligence (AI). AI is a broad term for technologies that can mimic human intelligence. AI includes programs that imitate cognitive processes like problem-solving and learning from examples. This includes a wide range of applications, including predictive systems and self-driving cars. With NLP, computers can understand spoken or written language and carry out tasks like subject classification, keyword extraction, and translation.

But machine learning is needed to automate these procedures and provide reliable responses. Applying algorithms to teach machines how to automatically learn from experience and get better over time—without being explicitly programmed—is known as machine learning. There are numerous applications of NLP like Smart speakers e.g., Google Assistant, Siri, Alexa etc.; virtual chatbots e.g., ChatGPT by Open.ai; and Predictive Text in search engines e.g., google.com.

The earliest approaches to NLP methods are rule-based. It has been tried and tested and has a track record of success. When rules are applied to text, they can reveal a great deal of

information. Consider what can be discovered about a piece of arbitrary text by looking for nouns, verb tenses, or Python code patterns. Examples of rule-based methods to NLP include regular expressions and context-free grammars.

Another approach to NLP is via machine learning. Probabilistic modeling, likelihood maximization, and linear classifiers are some examples of "traditional" machine learning techniques. Training data, in this case a corpus with markup feature engineering, word type, surrounding words, capitalized, plural, etc., are characteristics of traditional machine learning methodologies. A model is trained on parameters, then it is fitted to test data (typical of machine learning systems in general) Finding the most likely terms, the next word, the best category, etc., is a characteristic of inference (using the model to test the data).

Finally, another approach to NLP is via neural networks. While there are several similarities to "conventional" machine learning; however, there are some variations in it. Typically, feature engineering is not done because networks will "learn" key features. Instead, a very huge training corpus of neural networks are given streams of unprocessed raw parameters without engineered features and are fed into neural networks. Recurrent neural networks (RNNs) and convolutional neural networks (CNNs) are specific neural networks used in natural language processing (NLP).

Personally, I would love to work on natural language processing as a part of robot programming. Ideally, the robot will be able to perform functions with only voice commands. Another project of NLP would be to reduce or block phishing email by recognizing certain words like "click on the link to claim money/price" or "Enter your bank information". The specific program would be able to learn the pattern of sentences and recognize it for restrictive purposes.