An Overview of Natural Language Processing

Scott Lorance

What is natural language processing?

Natural language processing is the field of study that has to do with how computers interact with human languages. The way computers communicate is vastly different from the way humans do, so creating software that can bridge the gap is a very non-trivial task.

How are artificial intelligence and natural language processing related?

Natural language processing is a subset of the pursuit of artificial intelligence. AI is a general term, which can be applied to any number of tasks where a computer makes decisions, this includes things like generating images, or deciding whether to approve or deny a loan. One of the subsets of these tasks is natural language processing, It is arguable however that some basic forms of NLP do not qualify as AI, such as regular expression.

NLP: understanding vs generation

NLP can broadly be split into two separate categories, generation and understanding. As the names would imply, natural language generation is when the machine creates sentences in a human language, and understanding is when a machine extracts the meaning from a human language sentence it receives. NLU deals with helping a computer understand the rules of human sentence structure, semantics, and implications. NLG deals with helping a computer generate a sentence that follows those rules.

Applications of NLP

Natural language processing has many applications, one of the most common of which is voice recognition. This doesn't fall under either NLU or NLG because the computer needs only transcribe the sounds, and text to speech similarly also fits under the same umbrella but in reverse.

More modern and technical in the field of NLP is the evolution of chatbots, like chat-gpt3, which utilizes deep learning to replicate human writing patterns on structures even large than sentences, allowing it to maintain coherency even from paragraph to paragraph.

On the NLU side, an important industry development is summarization features. AI is learning how to analyze large documents, such as code bases, and extract the most important parts into more easily digestible formats without sacrificing accuracy.

Three approaches to NLP

Historically there have been three different methods used in natural language processing, each with different strengths and weaknesses.

The oldest and most computationally cheap approach is the rules based approach. This approach utilizes regular expressions and a dictionary of rules to automate simple tasks, like finding

emails, suggesting spelling corrections, and the like. It however cannot adapt to any situations it does not have a rule for, and thus has issues dealing with tasks that may have rules that complex, situational or hard to define.

Statistical is the next method, and rather than rely on a programmer written set of rules, the program learns those rules for itself. Data is first prepared into a more machine digestible format, annotated features, and then the rules can be deduced by the system.

The most modern and cutting-edge method is deep learning, which is where there is a large neural network which is fed raw data and learns in a way similar to a human. It generates its own features, rather than relying on a set given to it by the programmers.

Why I chose NLP

NLP is closely intertwined with the field of AI at large. Harnessing the full potential of AI to help with human endeavors requires that the AI we make be able to communicate with us in a way that both parties can understand. While my primary interest in AI is more focused on the technical aspects, NLP is a closely related field that I will likely have to interact with on the job if I work with AI of any other sort. The best way to learn is hands on, and I believe this class can deepen my understanding of AI as a whole.

https://skl180002.github.io/NLP-Spring-2023-SKL180002/