Rahul Das
RXD180029

Overview of NLP

Definition of Natural Language Processing (NLP): NLP is a subfield of artificial intelligence focused on implementing linguistics into software. The goal of NLP is to give computers the ability to process, understand, and generate natural sounding human language. This differs from the traditional rule-based language processing done by computers in that it uses machine learning models to understand and generate more natural language interactions by learning more abstract rules of conversation and language.

Natural Language Understanding (NLU) vs. Natural Language Generation (NLG): The two important aspects of NLP are NLU and NLG. NLU is focused on computers comprehending language, while NLG is focused on creating natural responses to prompts. They both require models to understand the nuances of language, but they have different goals.

Examples of NLP applications: One incredibly common application of NLP is in our phones' smart assistants. They are able to understand human prompts/questions and they are also able to generate realistic responses in almost real time. Our personal devices also use NLP for searches. Since humans don't always use properly formatted keyword-based searches, it is important to be able to understand our more abstract, nuanced search queries to find accurate results. Another one we use fairly often is for language translation. Many languages have significant nuance to their word and character choice, so it is important for the computer to properly understand the rules of the language.

The 3 Main Approaches to NLP:

1. The first ever approach of any sort of NLP was by implementing a rule-based system. This used a pattern matching system to figure out the meaning of inputs. This was a rudimentary NLP system but a good start. Basic chatbots, context free grammars, and other similar implementations utilized such rule-based systems. While using rules and pattern matching is a very intuitive approach and a reasonable start, some important flaws become apparent as you use these systems. Rule based approaches have no way of accounting for the nuances of human speech and context. They are much closer to matching responses/filling in the blanks instead of actually understanding the meanings of speech.
2. The next level of NLP is the use of probabilistic models, or machine learning. These systems use a corpus, or body, of existing data and use ML mathematical models to create relationships between features of the language. Using these models, the system infers meaning. This system is definitely a step up from the old-school rule-based systems, but it has its own limitations. The main issue is that probabilistic models require a corpus, which not only has a chance if overfitting data that isn't properly representative, but it also has more overhead without adaptability.
3. The current, and most advanced NLP is done with neural networks. The key difference between neural networks and probabilistic models is that they don't require the same

kind of pre-processed, feature engineered, data corpus as the ML approach. The neural network can learn and adapt, and additionally it has such a large corpus that it can more accurately cover the uncommon nuances of raw inputs, without requiring them to be engineered.

Personal interest in NLP: I am fascinated by linguistics and computing. NLP is a natural next step for me. I was incredibly interested by the idea of robots and other AIs interacting with humans as a regular part of our lives so I would love to explore the applications of NLP in those domains.