Overview of NLP

Definition:

Human-to-humann dialog is referred to as natural language. So it makes sense that Natural Language Processing, or NLP, is a field that involves transforming human speech into a format that computers are able to use. Language by itself means nothing to a machine that interprets things in simple 0's and 1's. Through NLP, computers can assign meaning and provide insight on something as complex as natural language.

AI vs. NLP:

Artificial Intelligence and Natural Language Processing are related fields. NLP, however, is a more specified term whereas AI is general. It can help to think of AI as an umbrella that encompasses fields such as NLP and Machine Learning. They are like branches. The NLP branch is specific to helping computers to work with human language. While there is some overlap between AI, NLP, and ML, all have different strengths and can be used independently within the same project to serve different purposes.

Natural Language Understanding and Generation:

There are two main steps associated with natural language: generation and understanding. Natural langauge generation means to create a response in some form of dialog. Natural language understanding means that all participants in the dialog know what was said. While these are the two main steps associates with natural language, there is a lot of human speech that goes beyond words- for instance, body language.

Uses of NLP:

Natural Langauge Processing has a lot of uses today. Many of these uses are encountered every day. Predictive text and autocorrect are two major applications of NLP. It can also be used to filter spam emails or in speech recognition. NLP can translate one langauge into anotheror be used to make a chatbot. There are tons of applications for NLP that help make our lives easier and help us communicate more effectively.

Main Approaches:

The oldest approach to Natural Langauge Processing is the rules-base approach. This form of NLP relies on things like regular expressions to form a sort of grammar. While this method is hard to scale up due to the complexity of human language, it is still used for relatively simple text processing problems.

Statistical and probabalistic approaches started to appear around the late 1980's. These methods are based on math and started out as simple word counting. Many classic machine learning algorithms fall into this category, but they have grown in complexity over the years. These models are used in predictive text as well as to help search engines find what you're

looking for. Statistical and probabilistic approaches do best when trained with moderate to large amounts of data

Increasing processing power brought us deep learning approaches to NLP. These algorithms include things like recurrent neural networks and LSTMs. The models are very advanced but require a lot of processing power to work with- processing power that not everyone has access to. This type of approach can be used to make machines sound and act more human like.

I

My interest in Natural Language Processing stems from a love of machine learning. I am part of the business intelligence team at work, and text data is something we haven't expored yet. I'm curious to see what we can learn from text data and how we learn it. The idea of transforming human language into a format that can be used by machines is really exciting, and I look forward to learning about how that works this semester!