Nishanth Lam

CS 4395

Mazidi

9/4/2022

Overview of NLP

A) Natural language processing (NLP) is the a sub-field of computer science, linguistics, and statistics that is related to the study of how computers and human languages interact. More specifically, it relates to the field of learning how to program computers to process and analyze large amounts of natural language data to preform various tasks.

B) Natural language processing is generally understood as a branch of AI. A large natural language processing project will usually contain components from a wide range of topics. Sometimes, the component might be purely NLP, purely machine learning, and others in the general category of AI. Both fields are rapidly evolving at the same time, and generally include some intermixing of techniques.

C) In a normal human-to-human conversation, there are mainly two things that are going on in order to facilitate efficient communication. First is *natural language understanding*, which means that each party is able to understand what the other party is saying. The next part is *natural language generation*, otherwise known as the formation of spoken responses. Of course, there can be other smaller factors of conversation such as tone, gestures, etc. However natural language generation and understanding are the main parts.

D) Some examples of modern natural language processing applications include: automated assistants, news feeds, web searches, auto-correct, text summation, sentiment analysis, language translator, advertising, grammar checks, etc.

E) The basis behind most NLP is the ability to examine text. There are three main approaches to learning from words. Rules-based approaches, statistical and probabilistic approaches, and deep

learning.

Rules-based approaches are the oldest techniques in NLP. This approach generally consists of creating some framework of grammars, and rules that can be used to analyze if syntax is grammatically correct or generate grammatically correct syntax based on the rules. A famous rule-based approach was Eliza in the 1960's. Eliza was a chat bot that could mimic a conversation by echoing talking points back to the user through advanced use of regular expressions. When Eliza couldn't come up with a response, it would just reply with a generic canned response.

Statistical and probabilistic approaches started being developed in the 1980's. These approaches were based more on mathematics and statistics. Simple techniques like counting words and probabilities of words and sequences led to far more useful language models. These models also make use of machine learning algorithms such as Naive Bayes, Logistic Regression, SVM, and Decision Tree's. This type of approach is much more effective than the rules-based approach and can often outperform even more advanced approaches on smaller to moderate datasets. Some examples of these models include being able to translate text between different human languages and predictive text, such as search suggestions.

The deep learning approach evolved from neural networks when human amounts of data and computing power became available. This approach consists of making use of modified neural networks such as CNN's, LSTM's and RNN's. This approach generally requires petabytes of data accessing to increased processing power through GPU's and cloud computing. Deep learning aims to push the boundary of NLP by making human-sounding interactions with

computers possible. To achieve this, we to introduce ideas such as considering conversation context, remembering a user's past preferences, and much more.

F) I have a personal interest in natural language progression, because it seems like the natural intersection of all my interests. I find the fields of discrete math, scientific computing, and linguistics very interesting. Natural language processing is a field that makes use of all these topics on one way or another. I would like to learn more about NLP by first studying various algorithms, then being able to code them from scratch through examples projects. After this class, I hope to be able to employ various basic NLP techniques, as well as have a better idea on understanding NLP related literature, and have a starting ground for me to learn more on my own.