Ghaida Alshiddi

CS 4395

Assignment 1

January 28th, 2023

Overview of NLP

Natural Language Processing is a way for programs and machines to understand and respond to human languages. NLP is software that analyzes human languages (either text or speech) to understand the request and produce a result.

Artificial intelligence involves many parts, and one of them is Natural Language Processing, as the program or machine attempts to understand a sentence from the user and provide a response to their request.

There are two main components in Natural Language Processing: natural language understanding and natural language generation. Natural language understanding means when the program or the user tries to understand and process the others' request/response. Natural language generation describes the process of responding to the user.

Some examples of NLP applications:

- Chatbots (including Amazon Alexa and Google Siri)
- Google Translate
- Auto response
- Predicting user inputs

There are three approaches to Natural Language Processing: Rules-based, statistical and probabilistic approaches, and deep learning. The first approach is rules-based, where programs rely mainly on regular expression and context-free grammar to understand the user input and produce a response. NLPs that use regular expressions tend to minimize words, get the roots, and compare them to an existing word root list. After the NLP compares the user word root with its existing list of roots, it provides a solution to the user. However, as mentioned in class, the technique is only sometimes correct as it might be words that end

with s to get their roots. Hence, this technique would not work for words that end with s, such as Dallas. The rules-based NLPs that use CFG can check the user's grammar and syntax to see if it is correct or not, and if it is not, the program will provide them with a solution with better grammar and syntax.

The second approach to NLPs is statistical and probabilistic approaches. This approach uses some machine learning algorithms and techniques to provide solutions and predictions. For example, NLPs can use regression models to make predictions for user input. The statistical approach could allow the NLP to learn from the data. The usage of statistical and probabilistic approaches increased significantly with the growth of data. Hence, this allows the NLP to produce better results. Application examples of this approach are translating languages and making predictions.

The third approach of NLP is deep learning. And deep learning also involves algorithms from machine learning, such as neural networks. However, because the deep learning approach is considered new, there are not many applications that use this approach, as deep learning requires a very large space from the computer and hard drive. Thus, applications that use deep learning technology tend to use an appropriate dataset that would not require large memory. So, because the applications do not use a large dataset, the applications are required to use the first and second approaches along with the deep learning approach to provide users with better results.

I am very interested in Natural Language Processing, and I would like to learn more about it. I have become interested in NLP since childhood, especially when I used to search on Google, and it would make predictions of what I was going to type. I would like to develop applications that use NLP, regardless if they are for professional or personal reasons. I would like to learn more about the statistical and probabilistic approaches to NLP and how to implement that, as I am very interested to learn about the mathematical part of it. I developed several projects that use regression and classification models to make predictions and test and train data. Hence, I would like to learn more about how I can use these models to make NLP applications.