Henry Kim HTK180000
CS 4395.001 Human Language Technology
Dr. Mazidi

Overview of NLP

Link to Portfolio Page: https://6henrykim.github.io/NLP-Portfolio/

a. Define NLP in your own words

One way to describe the field of natural language processing is the creation and use of "algorithms that allow computers to process human language" [1]. I would personally define NLP as a branch of computer science that attempts to give machines the capability to communicate with humans in an intuitive manner through language. I feel that the primary goal of NLP is to reduce the work done by humans in translating thoughts into computable data and have that work be done by the machine itself. For example, instead of a human filling out a course evaluation form and trying to quantify scores for all the categories, the human could write a single review and sentiment analysis could be performed to gauge satisfaction with various elements of the course.

b. Describe the relationship between AI and NLP

One view of artificial intelligence is that it is the field of designing agents that perform rational actions according to some performance measure [2]. On one hand, NLP can be viewed as a subfield of AI that considers linguistic output to be rational if it is intuitively understood by humans. On the other hand, language is often closely associated with thoughts, which are hard to observe in a computer. Alan Turing proposed one way to observe a machine's intelligence would be to have a human interrogator try to determine whether they were communicating with a human or machine, and that an intelligent machine would be able to pass as human [3]. For a machine to pass this test, it would need strong NLP capabilities although it is debatable if passing this test would demonstrate true intelligence.

c. Write a sentence or two comparing and contrasting natural language understanding and natural language generation

Natural language understanding is the process of converting linguistic stimuli (like spoken words or visual text) into an internal model of a concept or the world (picturing a furry, four-legged creature when encountering the word "dog"). Natural language generation is the process of converting an internal concept into a linguistic expression that can convey this idea to someone who uses the same language.

d. List some examples of modern NLP applications

Today many people use voice assistants like Siri or Alexa that use NLP to understand people's requests and generate appropriate responses. Other forms of NLP are the spellchecking and grammar checking features built into some text processors. NLP is also used in predictive text suggestions such as those on mobile devices.

e. Write 3 paragraphs describing each of the 3 main approaches to NLP, and list examples of each approach

The three main approaches to NLP are rules-based approaches, statistical and probabilistic approaches, and deep learning [1]. Rules-based approaches are techniques that require the programmer to define a set of rules by hand. These approaches are limited due to the vast amount of variation that occurs in naturally evolving languages. A simple index search program might use a rules-based approach by only showing results with letters or words that match the user's input. Another rules-based approach to a search program might include list of programmer-defined synonyms that can bring up results that use synonyms of words in the query.

Statistical and probabilistic approaches use machine learning methods such as Naïve Bayes, logistic regression, or support vector machines and require training on data. Instead of a programmer defining rules for the machine, the machine learns patterns from the data. For example, Naïve Bayes can be used to detect spam emails by converting each email into a bag of words (counting the occurrence of each word) and calculating the conditional likelihood of the email being spam or not based on whether it contains words similar to other emails identified as spam.

Deep learning is another machine learning approach based off large neural networks and includes algorithms like recurrent neural networks and long short-term memory techniques. These neural networks require a lot of processing power and large datasets to train. Deep learning has been used to create AI chatbots like IBM's Watson Assistant [4] to automate customer service tasks. Deep learning may also be combined with the other two techniques.

f. Write a paragraph describing your personal interest in NLP and whether/how you would like to learn more about NLP for personal projects and/or professional application

I am personally interested in NLP since it reflects our understanding of how our own human minds process language. Deep learning, which has been one of the major recent advancements in NLP, was inspired by theories about neural connections in the brain. Like human children, deep neural networks also need exposure to language to develop the ability to use it. However, our understanding of how human language works and how it relates to our ability to reason is still not complete, and I am interested to see how NLP with computers can improve our understanding of ourselves. In terms of professional application, I will be working at Epic, a company that develops medical software, after graduation. In the future, I might have the opportunity to implement NLP features in an app that allows doctors to speak with their computers instead of filling out spreadsheets when documenting patient visits.

References

[1] K. Mazidi, *Exploring NLP with Python,* 2019.

[2] P. Russell, *Artificial Intelligence : A Modern Approach*, global edition. S.L.: Pearson Education Limited, 2021.

[3] "Turing test," in *Encyclopedia Britannica.* [online document], Dec. 29, 2022. Available: https://www.britannica.com/technology/Turing-test.

[4] "AI chatbot that's easy to use," *ibm.com*, [Online]. Available: https://www.ibm.com/products/watson-assistant/artificial-intelligence. [Accessed Jan. 23, 2023]