Reg Gonzalez
CS 4395.001

Overview of NLP

Define NLP in your own words

Natural language processing (NLP) is a subsect of computer science that uses artificial intelligence, machine learning, and other techniques in order to create and develop algorithms to process and analyze human language.

Describe the relationship between AI and NLP

Natural language processing and artificial intelligence are connected in the sense that NLP uses techniques employed by AI. Similar to machine learning, NLP is a branch of AI. You can use machine learning and artificial intelligence techniques in order to work on projects that involve natural language processing.

Comparing & contrasting natural language understanding/natural language generation

Natural language understanding and natural language generation is a cyclical process. Natural language generation means producing responses (speaking or writing), while natural language understanding means interpreting and understanding what someone else has said/written. Humans engage in natural language generation/understanding all the time when we interact with each other in conversations: someone will say something; the other person hears it, understand and interpret it; and respond accordingly.

Examples of modern NLP applications

- Predictive text (e.g., spell check, autocorrect, or autocomplete in search engine results)
- Machine language translation (e.g., Google translate, SpanishDict, etc.)
- Email filters (e.g., filtering spam messages vs. legitimate ones)
- Virtual assistants (e.g., Siri, Alexa, etc.)
- Chatbots

3 main approaches to NLP

Rules-based approaches:

Out of the three different apporaches to natural language processing, rules-based approaches are the oldest ones. Like the name suggests, this approach uses a variety of rules in order to parse and make sense of human language input. These rules may include identifying the nouns in a sentence, identifying verbs (including tenses and conjugations), analyzing whether or not there's a pattern, etc. It's important to note that the reason why there are more approaches than just rules-based is because this type of approach does not scale well. Language is complex, so simply

Overview of NLP

---

having a list of rules is not going to fully encapsulate the nuances of human language. Examples of rules-based approaches include context-free grammars and regular expressions.

Statistical & probabilistic approaches:
This type of approach came after rules-based approaches, first coming into use in the late 1980s. Like the name suggests, this type of approach was more mathematical-based, as opposed to having various lists of rules. Statistical & probabilistic approaches were used to create better language models. These models help to find the nuances in language that most rule-based approaches would have a difficult time finding. For example, if someone wanted to translate the phrase “big brother” to another language, a model using statistical and probabilistic approaches may be able to determine that “big brother” should be translated to something like “older brother” and not “larger brother.” Examples of statistical and probabilistic apporaches can be found in many fundamental machine learning algorithms including Logistic Regression, Decision Trees, Naive Bayes, etc.

Deep learning:
Compared to rules-based and statistical/probabilistic approaches, deep learning is the newest approach to natural language processing. The evolution of neural network techniques is where deep learning came from; it also works very well with large amounts of data. Examples of deep learning techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are used a lot for NLP. More research and experimentation still needs to be done regarding deep learning and natural language processing, but the hope is that machines could understand and generate responses that are more human. This involves more complex processing of sentiment analysis, understanding the context of a particular conversation, having the machine “remember” previous interactions, etc.

Personal interest in NLP

NLP is a relatively new interest of mine. I’m currently a senior at UTD, and when I started out my undergraduate degree in CS, I didn’t have a solid idea of what I wanted to do specifically. However, over time I became more interested in topics like artificial intelligence and machine learning, which is what I think I’ll focus on moving forward. I had taken both CS 4365 (Artificial Intelligence) and CS 4375 (Intro to Machine Learning) last semester, the latter of which I took with Professor Mazidi. I enjoyed both classes a lot, but particularly Intro to Machine Learning because I really enjoyed the way Professor Mazidi taught the class. She made the concepts easy to understand and the assignments were great ways for us to practice all the different algorithms we had to learn. So, when I found out that she was teaching a ML-adjacent class (CS 4395), I wanted to take it. I think it also helps that language and writing are subjects

Reg Gonzalez
CS 4395.001

Overview of NLP

---

that I enjoy a lot too. Ever since middle/high school, English has always been one of my favorite subjects. I love writing and learning about language on a technical level. That combined with my interest in computer science, AI, and machine learning made me very interested in taking this course. Finally, I would like to learn more about NLP for both personal projects and professional application. Personally, besides my love and interest for writing/language and computer science, I'm taking this class to see if focusing on NLP *specifically* moving forward, as a career, is right for me.