NLP

Introduction

The field of natural language processing (NLP) is at the convergence of computer science, linguistics, and machine learning. NLP focuses on making computers comprehend and generate human language, whereas the field focuses on communication between computers and humans in natural language. NLP techniques are used in a variety of applications, including voice assistants such as Amazon's Alexa and Apple's Siri, as well as machine translation and text filtering.

Recent improvements in machine learning, particularly deep learning approaches, have had a significant impact on natural language processing (NLP). There are three sections to the field:

- Speech recognition - Converting spoken words into text using software.
- Natural Language Understanding—The computer's capacity to comprehend what we say.
- Natural Language Generation—The era of natural speech by a computer.

Syntactic and Semantic Analysis

Natural language can be understood using a variety of methods, the two most important of which are syntax and semantic analysis. But what exactly makes a sentence valid in the first place? Syntax and semantics.

The difference between syntax and semantics is how a text's grammatical structure communicates the text's meaning. However, even if a sentence is grammatically correct, it may not be conceptually correct. Even if "cows flow supremely" is technically correct (subject — verb, adverb), it makes no sense linguistically.

Syntactic Analysis

A formal grammar's rules are used to analyse natural language in syntactic analysis, which is also known as syntax analysis or parsing. Individual words are not subject to grammatical rules; only categories and groups of words are. Semantic structure is assigned to text through syntactic analysis.

If the subject and predicate are noun phrases, the sentence has a subject and a predicate of verb phrases. Consider the sentence that follows: It went away because of the dog (the noun

phrase)." You'll notice that each noun phrase has a matching verb phrase. A sentence can be syntactically correct, but it still may not make any sense.

Semantic Analysis

This subconscious process relies on our innate intuition and familiarity with the rules of language to help us decipher what someone has said. To put it another way, meaning and context play a huge role in how we comprehend language. The approach to computers, on the other hand, must be different. Language uses the term "semantic," which means "connected to meaning or logic."

Semantic analysis is the act of figuring out what words, symbols, and sentence structures signify and how they're supposed to be understood. This enables computers to interpret natural language to some extent in the same manner that they do in the human brain. There hasn't been enough work done on semantic analysis, which is a tough element of NLP.

For example, speech recognition technology has advanced to the point that it almost always works flawlessly, but natural language understanding still lags far behind. In general, your phone understands what you say, but it frequently is unable to act on that understanding. There are also technologies out there that lead you to believe they comprehend the meaning of a text when they don't in reality. There are several approaches to determine what a document is "about," including keyword analysis, statistical analysis, or even pure machine learning. These approaches are stifled by the fact that they ignore the true underlying significance.

Now let's have a look at some of the most widely utilised NLP methods nowadays. It is important to note how some of these subtasks are interwoven with one another and only serve to help solve the larger difficulties.

Parsing

What does the term "parsing" mean? To parse a sentence is to "resolve a sentence into its component pieces and identify their syntactic responsibilities," according to the dictionary definition.

That's pretty much perfect, but it's missing a few key details. A parse refers to the computer's formal breakdown of a sentence into its constituent parts, resulting in a parse tree that shows their syntactic relationship to one another in visual form. This can be utilised for further processing and understanding.

Speech Natural Language Processing

Single words have letters above them indicating the parts of speech (noun, verb and determiner). The next step up is to organise words into phrases using hierarchies. "The thief" and "robbed the apartment" are noun phrases and verb phrases, respectively, that when combined make a sentence that is one level higher than the previous one.
How do nouns and verbs differ? Noun phrases are a group of words that include a noun as well as one or more adjectives, verbs, or adverbs. Putting nouns together with words that relate to them is the goal.

Due to the structure of its representation, a parse tree also tells us about the grammatical relationships between the words. Take "the thief," for example. He is the topic of "robbed."

It has a structure of "robbed" with a "V" above it and "VP" above that, with a "S" linking it to the subject ("the thief"), which has a "NP" above it. That's what I mean by structure. A subject-verb relationship might be thought of as a kind of template for other kinds of relationships.

Stemming

Stemming is an NLP pre-processing and efficiency strategy derived from morphology and information retrieval. The dictionary describes it as having its "originate in or be caused by"

Stemming, in its simplest form, is the process of condensing words down to their root stem. The part of a word left over after all prefixes and suffixes have been removed is referred to as the stem. "Touched," for example, has the stem "touch." Because of this, we might say that "touch" is the stem of "touching," and so on.

You could be asking yourself, "What's the point of the stem?" Although numerous spellings may be used, the stem and meaning remain the same.As an illustration:

I was taking a ride in the car.

I was riding in the car.

The meaning and usage of the word are the same in these two statements.

If you can, try to visualise all of the English words in your vocabulary with all of their various affixes. To store them all, you'd need a massive database with many terms that have the same meaning in actuality. Focusing solely on the word's stem is a simple solution to this problem. In 1979, the Porter stemming method was developed, and it's still effective today.

Text Segmentation

It is a method in Natural Language Processing (NLP) whereby text is segmented into meaningful units like words, phrases, and various subtopics. Text is usually broken down into its constituent words, which is a challenging operation to accomplish depending on the language being studied. Again, this is a result of the intricacy of human language. The only exception to this rule is the term "icebox," which belongs together but is separated by a space in English. The issue is that "ice-box" is also a common misspelling.

Named Entity Recognition

The goal of named entity recognition (NER) is to find and classify the "named entities" in a text into predefined categories. Categories can include everything from names of people, organisations, and places to dollar amounts and percentages. These are just a few examples.

Before NER, Martin, for instance, purchased 300 SAP shares in 2016.
Within one year of NER, [Martin]Person purchased 300 shares of the [SAP]Organization.

Relationship Extraction

Relationship extraction examines the semantic relationships among the listed entities in NER. Knowing who is married to whom or working for which corporation, for example, could be part of this. For each sort of relationship, a machine learning model can be developed to solve this issue.

Sentiment Analysis

Sentiment analysis seeks to ascertain a speaker's or writer's attitude toward a document, interaction, or event (i.e. the sentiment). Text needs to be understood in order to forecast the underlying intent, hence this is a natural language processing challenge. Positive, negative, and neutral sentiment are the most common sentiment classifications.

You might want to forecast a customer's attitude and opinion about a product using sentiment analysis, for example. Review, polls, documents, and much more all benefit from sentiment analysis.