CS8691-ARTIFICIAL INTELLIGENCE

Natural Language Processing

- Natural Language Processing is the technology used to aid computers to understand the human's natural language.
- It's not an easy task teaching machines to understand how we communicate.
- Leand Romaf, an experienced software engineer who is passionate at teaching people how artificial intelligence systems work, says that "in recent years, there have been significant breakthroughs in empowering computers to understand language just as we do."
- This article will give a simple introduction to Natural Language Processing and how it can be achieved.

What is Natural Language Processing?

- Natural Language Processing, usually shortened as NLP, is a branch of artificial intelligence that deals with the interaction between computers and humans using the natural language.
- The ultimate objective of NLP is to read, decipher, understand, and make sense of the human languages in a manner that is valuable.
- Most NLP techniques rely on machine learning to derive meaning from human languages.

Trending AI Articles:

1. Cheat Sheets for AI, Neural Networks, Machine Learning, Deep Learning & Big Data
2. Data Science Simplified Part 1: Principles and Process
3. Getting Started with Building Realtime API Infrastructure
4. AI & NLP Workshop

In fact, a typical interaction between humans and machines using Natural Language Processing could go as follows:

1. A human talks to the machine
2. The machine captures the audio
3. Audio to text conversion takes place
4. Processing of the text's data
5. Data to audio conversion takes place
6. The machine responds to the human by playing the audio file

What is NLP used for?

Natural Language Processing is the driving force behind the following common applications:

- Language translation applications such as Google Translate
- Word Processors such as Microsoft Word and Grammarly that employ NLP to check grammatical accuracy of texts.
- Interactive Voice Response (IVR) applications used in call centers to respond to certain users' requests.
- Personal assistant applications such as OK Google, Siri, Cortana, and Alexa.

Why is NLP difficult?

- Natural Language processing is considered a difficult problem in computer science. It's the nature of the human language that makes NLP difficult.
- The rules that dictate the passing of information using natural languages are not easy for computers to understand.
- Some of these rules can be high-leveled and abstract; for example, when someone uses a sarcastic remark to pass information.
- On the other hand, some of these rules can be low-levelled; for example, using the character "s" to signify the plurality of items.
- Comprehensively understanding the human language requires understanding both the words and how the concepts are connected to deliver the intended message.
- While humans can easily master a language, the ambiguity and imprecise characteristics of the natural languages are what make NLP difficult for machines to implement.

How does Natural Language Processing Works?

- NLP entails applying algorithms to identify and extract the natural language rules such that the unstructured language data is converted into a form that computers can understand.
- When the text has been provided, the computer will utilize algorithms to extract meaning associated with every sentence and collect the essential data from them.
- Sometimes, the computer may fail to understand the meaning of a sentence well, leading to obscure results.
- For example, a humorous incident occurred in the 1950s during the translation of some words between the English and the Russian languages.
- Here is the biblical sentence that required translation:

  "The spirit is willing, but the flesh is weak."
- Here is the result when the sentence was translated to Russian and back to English:

  The vodka is good, but the meat is rotten."

What are the techniques used in NLP?

- Syntactic analysis and semantic analysis are the main techniques used to complete Natural Language Processing tasks.
- Here is a description on how they can be used.

1. Syntax

- Syntax refers to the arrangement of words in a sentence such that they make grammatical sense.
- In NLP, syntactic analysis is used to assess how the natural language aligns with the grammatical rules.
- Computer algorithms are used to apply grammatical rules to a group of words and derive meaning from them.

Here are some syntax techniques that can be used:

- **Lemmatization**: It entails reducing the various inflected forms of a word into a single form for easy analysis.
- **Morphological segmentation**: It involves dividing words into individual units called morphemes.
- **Word segmentation**: It involves dividing a large piece of continuous text into distinct units.
- **Part-of-speech tagging**: It involves identifying the part of speech for every word.
- **Parsing**: It involves undertaking grammatical analysis for the provided sentence.

- **Sentence breaking**: It involves placing sentence boundaries on a large piece of text.
- **Stemming**: It involves cutting the inflected words to their root form.

2. Semantics

Semantics refers to the meaning that is conveyed by a text. Semantic analysis is one of the difficult aspects of Natural Language Processing that has not been fully resolved yet.

It involves applying computer algorithms to understand the meaning and interpretation of words and how sentences are structured.

Here are some techniques in semantic analysis:

- **Named entity recognition (NER):** It involves determining the parts of a text that can be identified and categorized into preset groups. Examples of such groups include names of people and names of places.
- **Word sense disambiguation:** It involves giving meaning to a word based on the context.
- **Natural language generation**: It involves using databases to derive semantic intentions and convert them into human language.

Wrapping up

Natural Language Processing plays a critical role in supporting machine-human interactions.

As more research is being carried in this field, we expect to see more breakthroughs that will make machines smarter at recognizing and understanding the human language.