Rosary College of Commerce & Arts
Navelim, Salcete-Goa.

Department of Computer Applications

BITS N BYTES

Vol : **XIX** **JULY 2022** Issue : **2**

Principal's Message

Natural Language Processing (NLP) is the ability of a computer program to understand human language as it is spoken and written. It's a technology that many people use daily and has been around for years, but is often taken for granted. It uses artificial intelligence to take real world input, process it, and make sense of it in a way a computer can understand.

Natural Language Processing can be used in business to track, facilitate, and analyze thousands of customer interactions in order to improve their product or service. In education NLP help students to improve their reading and writing skills. The algorithms used in NLP can quickly identify problems in a student reading ability and can provide real –time, automatic feedback on how to improve.

The current issue of the news-letter 'Bits N Bytes' published by the Department of Computer Applications has interesting contributions from faculty on Natural Language Processing and its importance in business, education and other sectors. I am sure 'Bits N Bytes' will make an interesting reading.

Congratulations to the Department of Computer Applications and the editorial board for the hard work and contributions.

- **Dr. Helic M. Barretto**
Acting Principal

Introduction to Natural Language Processing

-Asst. Prof. Ramkrishna Reddy

Natural language processing (NLP) is a branch of artificial intelligence that helps computers understand, interpret and manipulate human language. NLP draws from many disciplines, including computer science and computational linguistics, in its pursuit to fill the gap between human communication and computer understanding.

Natural language processing (NLP) describes the interaction between human language and computers. It enables machines to recognize characters, words and sentences, then apply meaning and understanding to that information. This helps machines to understand language as humans do.

NLP Analyzes Unstructured Text Data

For natural language processing to help machines understand human language, it must go through speech recognition, natural language understanding and machine translation. It is an iterative process comprised of several layers of text analysis, including:

Morphological Level: Morphemes are the smallest units of meaning within words, and this level deals with morphemes in their role as the parts that make up a word.

Lexical Level: This level of speech analysis examines how the parts of words (morphemes) combine to make words and how slight differences can dramatically change the meaning of the final word.

Syntactic Level: This level focuses on the text at the sentence level. Syntax revolves around the idea that, in most languages, the meaning of a sentence is dependent on word order and dependency

Semantic Level: Semantics focuses on how the context of words within a sentence helps determine their meaning on an individual level.

Discourse Level: Discourse reveals how sentences relate to one another. Sentence order and arrangement can affect the meaning of the sentences.

Pragmatic Level: Pragmatic analysis bases the meaning of words or sentences on situational awareness and world knowledge. Basically, what meaning is most likely and would make the most sense.

NLP Uses AI to Process Language

Text analysis is only part of the NLP process. For machines to truly understand words in context, they need to be able to disambiguate language at a human-like level. The level at which you are able to disambiguate language depends on your approach to artificial intelligence. The different approaches are as follows:

Symbolic Approach: The symbolic approach to NLP is based on human-developed rules and lexicons. In other words, the basis behind this approach is generally accepted rules of speech within a given language that are materialized and recorded by linguistic experts for computer systems to follow.

Statistical Approach: The statistical approach to NLP is based on observable and recurring examples of linguistic phenomena. Models based on statistics recognize recurring themes through mathematical analysis of large text corpora. By identifying trends in large samples of text, the computer system can develop its own linguistic rules to use when analyzing future input and/or generating language output.

Hybrid Approach: The hybrid approach to NLP combines the best capabilities of symbolic and statistical approaches. You can leverage hybrid AI in a variety of ways depending on your needs. For example, existing symbolic rules can provide base knowledge for a machine learning model to learn from. On the other hand, machine learning can generate symbolic rules for humans to validate and then use to train a model.

NLP Extracts and Classifies Text

Natural language processing automates the extraction of information from unstructured documents, emails and even social media posts so that it can be labelled and categorized in an enterprise knowledge management system for future reference or analysis. This saves organizations time and money by not having to manually process each document by hand.

NLP can improve the functionality of many different systems such as search engines, question answering systems, translation systems, text summarization tools, and machine translation tools. It also plays a role in image captioning and text-to-speech synthesis, which converts written text into audio files.

NLP has grown beyond simple sentence analysis into other areas such as text classification and summarization. These techniques are widely used across various industries, including healthcare, financial services, and media companies. Here are four ways text classification can be used:

Named entity recognition: Named entity recognition is a subfield of information extraction that deals with identifying names of entities (e.g., people, places, organizations, etc.) in text. Named entity recognition helps machines understand core subject matter and relates to question answering and textual entailment.

Sentiment analysis: Sentiment analysis is based on NLP and computational linguistics. You can use sentiment analysis software to determine the tone or attitude someone expresses through their words. This tool uses an algorithm that analyzes text for positive or negative words and phrases.

Data annotation: Data annotation is the addition of metadata to data sets. In NLP, this refers to labelling data with attributes that can be used to train machine learning models. Data annotation is crucial for NLP because it allows machines to understand the content and structure of documents which, in turn, helps them make sense of unstructured texts such as emails, tweets and other online content.

Document classification: Document classification is the process of categorizing documents into groups based on features you extract from their content. You can do this via a symbolic approach which may already have pre-established taxonomies, or you can use machine learning — supervised or unsupervised — to build out your own.

Applications of Natural Language Processing

-Asst. Prof. Sonia Mashal

Chatbots

Chatbots are a form of artificial intelligence that are programmed to interact with humans in such a way that they sound like humans themselves. Depending on the complexity of the chatbots, they can either just respond to specific keywords or they can even hold full conversations that make it tough to distinguish them from humans. Chatbots are created using Natural Language Processing and Machine Learning, which means that they understand the complexities of the English language and find the actual meaning of the sentence and they also learn from their conversations with humans and become better with time.

Auto Correct and Auto Prediction

There are many softwares available nowadays that check grammar and spelling of the text we type and save us from embarrassing spelling and grammatical mistakes in our emails, texts or other documents. NLP plays an important role in these softwares and functions. This is one of the most widely used applications of NLP. These softwares offer a lot of features like suggesting synonyms, correcting

grammar and spellings, rephrasing sentences and giving clarity to the document and can even predict the tone of the sentence that might be implied by the user.

Auto prediction is also a feature developed through NLP where the computer suggests automatic prediction of the text we have started typing. This saves time of the user and makes the job easier for them

Voice Assistants

Nowadays almost everyone uses voice assistants to make calls, place reminders, schedule meetings, set alarms, surf the internet, etc. These voice assistants have made life much easier. But how do they work? They use a complex combination of speech recognition, natural language understanding, and natural language processing to understand what humans are saying and then act on it. The long term goal of voice assistants is to become a bridge between humans and the internet and provide all manner of services based on just voice interaction.

Language Translator

Social Media has brought the entire world together but with unity comes challenges like language barrier. With different translating softwares that work individually or are integrated within other applications, this hurdle has been easily defeated.

Many of the translation tools use Sequence to sequence modelling that is a technique in Natural Language Processing. It allows the algorithm to convert a sequence of words from one language to another which is translation. Earlier, language translators used Statistical machine translation (SMT) which meant they analyzed millions of documents that were already translated from one language to another (English to Hindi in this case) and then looked for the common patterns and basic vocabulary of the language. However, this method was not that accurate as compared to Sequence to sequence modelling.

Sentiment Analysis

Companies can use sentiment analysis in a lot of ways such as to find out the emotions of their target audience, to understand product reviews, to gauge their brand sentiment, etc. And not just private companies, even governments use sentiment analysis to find popular opinion and also catch out any threats to the security of the nation.

Email Classification and Filtering

Emails are still the most important method for professional communication. Our emails are automatically divided into 3 sections namely, Primary, Social, and Promotions which means we never have to open the Promotional section! But how

does this work? Email services use natural language processing to identify the contents of each Email with text classification so that it can be put in the correct section.

Speech Recognition

Speech Recognition is a technology that enables the computer to convert voice input data to machine readable format. There are a lot of fields where speech recognition is used like, virtual assistants, adding speech-to-text, translating speech, sending emails etc. It is used in search engines where the user can voice out the name of their search requirements and get the desired result, making our work easier than typing out the entire command.

Advertisement to Targeted Audience

If you ever search any product or object in any shopping site, you would often see ads of those products and other related products on other sites. This type of targeted online advertising is done with the help of NLP and it is known as Targeted Advertising. Through NLP, keywords that are searched by the user are matched with the keywords of the product ad. If they are similar, the user gets an advertisement. This process is called keyword matching. This has been highly beneficial to many companies and saved them a lot of investment as the ads are only shown to customers who are actually interested in the product, which is determined on the basis of their online activity.

Social Media Analytics

Social Media is an integral part of everyone's life nowadays and many people use it to post their thoughts about different businesses and products.

The companies can understand their market position and get their customer reviews by analyzing the data. But due to the huge number of users, the information could be very hard to grasp. It is not convenient to go through millions of comments and updates just to get the insights.

Recruitment

In this competitive world, big and small companies are on the receiving end of thousands of resumes from different candidates. It has become a tough job for the HR team to go through all the resumes and select the best candidate for one single position. NLP has made the job easier by filtering through all the resumes and shortlisting the candidates by different techniques like information extraction and name entity recognition. It goes through different attributes like Location, skills, education etc. and selects candidates who meet the requirements of the company closely.