An Extant of Natural Language Processing

Vanita

Assistant Professor, Panjab University Constituent College, Dharamkot (Moga)

Abstract

Natural Language Processing (NLP) is a Subfield of Artificial Intelligence (AI). This is a widely used technology for personal assistants that are used in various business fields/areas. This technology works on speech provided by the user breaks it down for proper understanding and process it accordingly. This is very recent and effective approach due to which it has a really high demand in today's market. Natural Language Processing is an upcoming field where already many transitions such as compatibility with smart devices and interactive talks with human have been made possible. Knowledge representation, logical reasoning and constraint satisfaction were the emphasis of AI applications in NLP. Here first it was applied to semantics and later to grammar. In the last decade, a significant change in NLP research has resulted in the widespread use of statistical approaches such as machine learning and data mining on a massive   scale. The need for automation is never ending courtesy of the amount of the work required to be done these days. NLP is very favorable, but aspect when it comes to automated applications. The applications of NLP have led it to be one of the most south-after methods of implementing machine leaning. NLP is a field that combines computer science, linguistics, and machine leaning to study how computers and humans communicate in natural language. The goal of NLP for computers is to be able to interpret and generate human language. This not only improves the efficiency of work done by humans but also helps in interacting with the machine. NLP bridges the gap of humans and electronic devices.

Natural Language Processing (NLP) is a subfield of Artificial Intelligence (AI) that deals with interaction between humans and computers in natural language. It involves the computational techniques to process and analyze language data, such as text and speech, with the goal of understanding the meaning behind the language.

NLP is used in wide range of applications, including machine translation, sentiment analysis, speech recognition, chatbots, and text classification. Some common techniques used in NLP include:

1. **Tokenization:** the process of breaking text into individual words or phrases.
2. **Part-of-speech tagging:** the process of labeling each word in a sentence with its grammatical part of speech.
3. **Named entity recognition:** the process of identifying and categorizing named entities, such as people, place, and organizations, in text.
4. **Sentiment analysis:** the process of determining the sentiment of piece of text, such as whether it is positive, negative, or neutral.
5. **Machine Translation:** the process of automatically translating text from one language to another.
6. **Text classification:** the process of categorizing text into predefined categories or topics.

Recent advances in deep leaning, particularly in the area of neural networks have led to significant improvements in the performance of NLP systems. Deep learning techniques such as Convolutional Neural Networks(CNNs) and Recurrent Neural Networks(RNNs) have been applied to tasks such as sentiment analysis and machine translation, achieving state-of-the-art results.

Overall, NLP is a rapidly evolving field that has the potential to revolutionize the way we interact with computer and the world around us.

Keywords: Natural Language Processing (NLP) tasks, Working of Natural Language Processing(NLP), Natural Language Generation(NLG), Natural Language Understanding (NLU), Technologies related to Natural Language Processing, Applications related to Natural Language Processing, Future Scope

What is Natural Language Processing?

Natural language processing (NLP) is a field of computer science and artificial intelligence that aims to make computers understand human language. NLP uses computational linguistics which is the study of how language works and various models based on statistics, machine leaning and deep leaning. These technologies allow computers to analyze and process text or voice data and to grasp their full meaning, including speaker's or writer's intentions and emotions.

NLP powers many applications that use language, such as text translation, voice recognition, text summarization and chatbots. You may have used some of these applications yourself, such as voice-operated GPS systems, digital assistants, speech-to-text software and customer service bots. NLP also helps business improve their efficiency, productivity and performance by simplifying complex tasks that involve language.

Human language is filled with ambiguities that make it incredibly difficult to write software that accurately determines the intended meaning of text and voice data. Homonyms, homophones, sarcasm, idioms, metaphors, grammar and usage expectations, variations in sentence structure- these just a few of irregularities of human language that take human years to learn, but that programmers must teach natural language- driven applications to recognize and understand accurately from the start, if those applications are going to be useful.

NLP Tasks

Several NLP tasks break down human text and voice data in the ways that help the computer make sense of what it's ingesting. Some of these tasks include the following.

- **Speech recognition,** also known as speech to text, is a challenging task that involves converting voice data into text data. This technology is essential for many applications that requires voice command and spoken responses. However people's speaking habits such as speaking quikly, slurring words, using different accents and incorrect grammar, make speech recognition even more challenging.

- **Part of speech tagging** also known as grammatical tagging, is a crucial process that determines the part of speech of a specific word or piece of text based on it's context and usage. For example it can identify 'make'as a verb in 'I can make a paper plane' and as a noun in ' What make of car do you own?'

- **Word sense disambiguation** is a semantic analysis process that select the most appropriate meaning of word with multiple meanings based on the given context. This process is helpful in distinguishing the meaning of verb 'make' in 'make the grade' (achieve) versus 'make a bet' (place).

- **Named entity recognition(NEM)** identifies useful entities or phrases, such as 'kentucky' as a location or 'Fred' as a person's name. Co-reference resolution is the task of identifying when two words refer to make the same entity such as determining that 'she' refers to 'Mery' . Sentiment

analysis is a process that attempts to extract subjective qualities, including attitudes, emotions, sarcasm, confusion and suspicion from text.

- **Natural Language Generation** is the opposite of speech recognition, as it involves putting structured information into human language. Overall, Understanding these processes is essential in building effective natural language processing systems.

Natural Language Processing

The goal of NLP is to develop algorithms and models that enable computers to understand, interpret, generate and manipulate human languages.NLP techniques are widely used in search engines, machine translation, sentiment analysis, question answering and many more. NLP research is an active field and recent advancements in deep learning have led to significant improvements in NLP performance. However, NLP is still a challenging field as it requires an understanding of both computational and linguistic principals.

Working of Natural Language Processing (NLP)

Working in natural language processing (NLP) typically involves using computational techniques to analyze and understand human language. This can include tasks such as language understanding, language generation and language interaction.

The field is divided into three different parts:

1. **Speech Recognition:** The translation of spoken language into text.
2. **Natural Language Understanding (NLU):** The computer's ability to understand what we say.
3. **Natural Language Generation (NLG):** The generation of natural language by a computer.

NLU and NLG are the key aspects depicting the working of NLP devices. These 2 aspects are very different from each other and are achieved from different methods.

Individuals working in NLP may have background in computer science, linguistics, or a related field. They may also have experience with programming languages such as python and C++ and be familiar with various NLP Libraries and frameworks such as NLTK, spaCy, OpenNLP.

Speech Recognition

- First, the computer must take natural language and convert it into machine-readable language. This is what speech recognition or speech-to-text does. This is first step of NLU.
- Hidden Markov Models (HMMs) are used in majority of voice recognition system nowadays. These are statistical models that use mathematical calculations to determine what you said in order to convert your speech to text.
- HMMs do this by listening to you talk, breaking it down into small units comparing it to prerecorded speech to figure out which phoneme you uttered in each unit. The program then examines the sequence of phonemes and uses statistical analysis to determine the most likely words and sentence you were speaking.

Natural Language Understanding (NLU):

The next and hardest step of NLP is the understanding part.

- First, the computer must comprehend the meaning of each word. It tries to figure out whether the word is noun or verb, whether it is in the past or present tense, and so on. This is called the part of

speech tagging.

- A lexicon and a set of grammatical rules are built into NLP systems. The most difficult part of NLP is understanding.
- The machine should be able to grasp what you said by conclusion of the process. There are several challenges in accomplishing this when considering problems such as words having several meanings or different words having similar meanings, but developers encode rules into their NLU systems and train them to learn to apply the rules correctly.

Natural Language Generation (NLG):

NLG is much simpler to accomplish. NLG converts a computer's machine-readable language into text and can also that text into audible speech using text-to-speech technology.

- First, the NLP system identifies what data should be converted to text. If you asked the computer a question about the weather, it most likely did an online search to find your answer, and from there it decides that the temperature, wind, humidity are the factors that should be read aloud to you.
- Then, it organize the structure of how it's going to say it. This is similar to NLU expect backward. NLG system can construct full sentences using a lexicon and a set of grammar rules.
- Finally, text-to-speech takes over. The text-to-speech engine uses a prosody model to evaluate the text and identify breaks, duration and pitch. The engine then combines all the recorded phonemes into one cohesive string of speech using a speech database.

Some common roles of NLP (Natural Language Processing) include:

- NLP engineer: designing and implementing NLP systems and models.
- NLP researcher: conducting research on NLP techniques and algorithms.
- ML engineer: Designing and deployment of various machine leaning models including NLP.
- NLP data scientist: analyzing and interpreting NLP data.
- NLP consultant: providing expertise in NLP to organizations and businesses.

Working in NLP can be both challenging and rewarding as it requires a good understanding of both computational and linguistic principles. NLP is fast-paced and rapidly changing field, so it is important for individuals working NLP to say up-to-date with the latest developments and advancements.

Technologies related to Natural Language Processing

There are variety of technologies related to Natural Language Processing (NLP) that are used to analyze and understand human language. Some of most common include:

1. **Machine Learning:** NLP relies heavily on Machine Learning techniques such as such as supervised and unsupervised leaning, deep learning and reinforcement leaning to train models to understand and generate human language.
2. **Natural Language Toolkits:** and other libraries NLTK is a popular open-source library in python that provides tools for NLP tasks such as tokenization, stemming and part-of-speech tagging. Other popular libraries includes paCy, OenNLP and coreNLP.
3. **Parsers:** parsers are used to analyze the syntactic structure of the sentences, such as dependency parsing and constituency parsing.
4. **Text-to-speech and speech-to-text systems:** TTS systems convert written text into spoken words, while SST systems convert spoken words into written text.

5. **Named Entity Recognition (NER) Systems:** NER systems identify and extract named entities such as people, places and organizations from text.
6. **Sentiment analysis:** A technique to understand the emotions or opinions expressed in a piece of text by using various techniques like lexicon based, Machine-leaning Based and Deep Learning-based methods.
7. **Machine Translation:** NLP is used for language translation from one language to another through computer.
8. **Chatbots:** NLP is used for chatbots that communicate with other chatbots or human through auditory or textual methods.
9. **AI Software:** NLP is used in question answering software for knowledge representation, analytical reasoning as well as information retrieval.

Applications of Natural Language Processing(NLP):
- **Spam filters:** one of the most irritating things about email is spam. Gmail uses NLP(Natural Language Processing) to discern which emails are legitimate and which are spam . these spam filters look at the text in all the emails you receive and try to figure out what it means to see if it's spam or not.
- **Algorithmic trading:** Algorithmic trading is used for predicting stock market conditions. Using NLP, this technology examines headlines about companies and stocks and attempts to comprehend their meaning in order to determine if you should buy , sell or hold certain stocks.
- **Questions Answering:** NLP can be seen in action by using Google Search or Siri Services. A Major use of NLP is to make search engines understand the meaning of what we are asking and generate natural language in return to give us the answer.
- **Summarizing Information:** On the internet there is lot of information and lot of it comes in the form of long documents or articles. NLP is used to decipher the meaning of data and then provides shorter summaries of the data so that humans can comprehend it more quickly.

Future scope:
- **Bots:** chatbots assist clients to get to the point quickly by answering inquiries and referring them to relevant resources and products at any time of the day and night. To be effective, chatbots must be fast, smart and easy to use, to accomplish this, chatbots employ NLP to understand language, usually over text or voice recognition interactions.
- **Supporting Invisible UI:** Almost every connection we have with machines involves human communication, both spoken and written. Amazon's Echo is only one illustration of the trend towards putting human's in closer contact with technology in future. The concept of an invisible or zero user interface will rely on direct communication between the user and machine, whether by voice, text or a combination of the two. NLP helps to make this a real world thing.
- **Smart Search:** NLP future also includes improved search, Something we have been discussing at Expert System for a long time. Smarter search allows a chatbot to understand a customer's request can enable "search like you talk" functionally rather than focusing on keywords or topics. Google recently announced that NLP capabilities have been added to Google Drive, allowing users to search for documents and content using natural language.

References:

1. Aaron Elkiss, Siwei Shen, Anthony Fader, Gunes¸ Erkan, ¨ David States, and Dragomir R. Radev. 2008. Blind men and elephants: what do citation summaries tell us about a research article? Journal of the American Society for Information Science and Technology, 59(1):51–62

2. Mark T. Joseph and Dragomir R. Radev. 2007. Citation analysis, centrality, and the acl anthology. Technical Report CSE-TR-535-07, University of Michigan. Department of Electrical Engineering and Computer Science.

3. Min-Yen Kan and Yee Fan Tan. 2008. Record matching in digital library metadata. Communications of the ACM (CACM), 51(2):91–94.

4. Min-Yen Kan. 2007. SlideSeer: A digital library of aligned document and presentation pairs. In Proceedings of the Joint Conference on Digital Libraries (JCDL '07), pages 81–90, Vancouver, Canada, June.

5. Steve Lawrence, C. Lee Giles, and Kurt Bollacker. 1999 Digital libraries and autonomous citation indexing. IEEE Computer, 32(6):67–71.

6. Ariel Schwartz, Anna Divoli, and Marti Hearst. 2007. Multiple alignment of citation sentences with conditional random fields and posterior decoding. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), pages 847–857