Comparative Study on Natural Language Processing

Lakshya Muralidhara, Ashwini Patil, Greeshma Murthy, Dr. S. Kavitha

Dayananda Sagar College of Art Science and Commerce, Bangalore, Karnataka, India

ABSTRACT

Natural language processing (NLP) is a branch of that helps computers understand, interpret and manipulate human language. NLP draws from many disciplines, including computer science and computational linguistics, in its quest to fill the gap between human communication and computer understanding. Many organizations use NLP techniques to optimize customer support, improves the efficiency of text analytics by easily finding the information they need, and enhance social media monitoring. For example, banks might implement NLP algorithms to optimize customer support; a large consumer products brand might combine natural language processing and semantic analysis to improve their knowledge management strategies and social media monitoring.

Keywords : Artificial Intelligence, Machine Learning, Linguistic, Ambiguous, Semantics

I. INTRODUCTION

Natural language processing (NLP) is the capability of a computer program to understand human language as it is spoken. NLP is a component of artificial intelligence (AI). The history of NLP generally started in the 1950's, although works can be found from the earlier periods. In 1950 Alan Turing published an article titled "COMPUTING MACH IN ERY A N D IN TE LL IGEN CE" , which is at present called the "TURING TEST" as a criterion of intelligence. . Starting in the late 1980s, however, there was a revolution in natural language processing with the introduction of machine learning algorithms for language processing Some of the earliest-used machine learning algorithms, such as decision trees, produced systems of hard if-then rules similar to existing hand-written rules. However, part-of-speech tagging introduced the use of hidden Markov models to natural language processing, and increasingly, research has focused on statistical models, which make soft, probabilistic decisions based on attaching real- valued weights to the features making up the input data. Some of the recent algorithms for Natural Language Processing Some of the recent algorithms for Natural Language Processing are (WOS) WORD SENSE DISAMBIGUATION TECHNIQUE, MACHINE TRANSLATION, SPEECH TAGGING AND RECOGNITION, and GENERATION OF NATURAL LANGUAGE.

Natural Language Processing is the driving force behind the following common applications: Google Translate, Word Processors such as Microsoft Word and Grammarly (to check grammatical accuracy), Interactive Voice Response (IVR) , Personal assistant applications such as OK Google, Siri, Cortana, and Alexa.

Here are some tools available for NLP: CoreNLP, NLTK, TextBlob, Gensim, spaCy.

There are three different levels of linguistic analysis done before performing NLP:

- ✓ Syntax – What part of given text is grammatically true.
- ✓ Semantics – What is the meaning of given text?
- ✓ Pragmatics – What is the purpose of the text?

NLP deal with different aspects of language such as: Phonology, Morphology.

Approaches of NLP for understanding semantic analysis:

- ✓ Distributional – It employs large-scale statistical tactics of Machine Learning and Deep Learning.
- ✓ Frame – Based – The sentences which are syntactically different but semantically same are represented inside data structure (frame) for the stereotyped situation.
- ✓ Theoretical – This approach is based on the idea
- ✓ that sentences refer to the real word (the sky is blue) and parts of the sentence can be combined to represent whole meaning.
- ✓ Interactive Learning – It involves pragmatic approach and user is responsible for teaching the computer to learn the language step by step in an interactive learning environment. The true success of NLP lies in the fact that humans deceive into believing that they are talking to humans instead of computers.

STEPS FOR NATURAL LANGUAGE PROCESSING

The user needs to import a file containing text written and then perform the following steps for natural language processing.

- ✓ Sentence segmentation – It identifies the start and end points of a given sentence. Usually start of a sentence uses capital letters or bullets or numbering whereas end of a sentence contains punctuation marks like '.' or '?'
- ✓ Tokenization – It recognizes words, numbers, and other punctuation marks and symbols.
- ✓ Stemming – It casts off the ending of the words.
- ✓ For example 'drinking' or 'drunk' is reduced to 'drink'.

- ✓ Part of speech (POS) tagging – It designates each sentence its part of speech tag such as assigning given word as adjective or verb or pronoun etc.
- ✓ Parsing – It is used to divide a given text into different categories. For example the first category can be to answer the part of a sentence and second category to modify another part of the sentence.
- ✓ Named Entity Recognition – It recognizes the entities or units such as people, place and time within the given document.
- ✓ Co-Reference resolution – It describes the relationship between the given word in a sentence with the preceding and the succeeding sentences with reference to the given word.

DIFFERENCE BETWEEN NLP AND TEXT MINING OR TEXT ANALYTICS

Natural language processing is responsible for understanding meaning and structure of given text. Text Mining or Text Analytics is a process of extracting hidden information inside text data through pattern recognition

NLP (Natural Language Processing)	Text Mining or Text Analytics
Automated Speech	Automated Grouping (n grams approach)
Automated Writing	Automated Classification (bag of words)
Automated Translation	Pattern Discovery

SOME KEY APPLICATION AREAS OF NLP

NLP has many other applications besides Big Data, Log Mining, Deep learning and Log Analysis. Despite of the fact that the term 'NLP' is not as popular as 'big data' 'machine learning' or 'artificial intelligence,' it's used very commonly on a day to day basis. Some of the other applications of NLP are:-

Automatic summarizer – When an input text is given, this gives us the summary by removing irrelevant points

Sentimental analysis – When an input is given, it analyses the given text to predict the tone of the text For example - whether the text conveys judgment, opinion, order, review or question.

Text classification – It is implemented to classify different journals, news stories according to their element. Multi-document classification is also possible. A popular example of text classification is spam detection in emails. This property can be used to detect the name of the author of the given journal based on the writing style.

Information Extraction – The process to extract specific information. One of the most common examples is when email extracts only from your messages and automatically adds the events to the calendar.

STUDY OF SOME OF THE RECENT NATURAL LANGUAGE PROCESSING ALGORITHMS

Machinery Translation Technique

Machine translation is the process of converting one natural language into another by preserving the meaning and producing a meaningful and fluent output. Machine Transition Techniques are based on different models:

1. Bilingual machine translation: A bilingual machine translation system is translates just a pair of languages and cannot be adapted to other languages.
2. Transfer-based machine translation: This translation model is based on three modules: Analysis module, Transfer module, Generation module.
3. Interlingual-based machine translation: This translation model is based on two main modules: Analysis module, Generation module
4. Memory-based machine translation: This translation model is based on the "translation memory." It is a corpus-based approach. The system just re-uses translations previously stored by the professional translator without really analyzing the source text and a dictionary (terminology support) is used to help the expert to translate the parts of texts that haven't been previously translated. This "new" translation concept frees the professional translator to attend to the finer

points of translation that require the judgment of an expert by offering a computer-assisted translation that automates repetitive tasks.
5. Statistical-based machine translation: This translation is a corpus-based approach. Statistical concepts are among the first techniques for machine translation. A few examples of Machine Translation techniques are

1. Word Alignment: - Word alignment is a primary crisis in statistical machine translation. The core of the task is to identify relations between words or phrase of two sentences articulated in different languages.
2. Data Matching: - Data matching, or in other words record linking, is the process of finding the matching pieces of information in large sets of data. The purpose can be to find entries that are related to the same subject or to detect duplicates in the database.

The main issues of Machine Translation are disambiguation, non-standard speech and named entities

II. MORPHOLOGICAL SPLITTING TECHNIQUE

Split Morphology is a hypothesis which requires definite information on the Derivation and Inflection that has separate components of grammar. Whereas Derivations are mostly handled by lexical rules and Inflections are handled by syntactic rules.

WORD SENSE DISAMBIGUATION

Word sense disambiguation is a concept where a word used any times in a particular context is given an appropriate or "sense" (meaning) to the corresponding sentence which is largely unaware in people. The feature of this context provides the evidences for classification. The research on this point has always been consistent on giving exact or precise results without any doubt. They researched a variety of techniques by using dictionary based methods and also the knowledge which was encoded in lexical recourses

to supervise machine learning methods providing a collection of manually sense – annotated examples. The research of unsupervised methods has also provided cluster of repeated words including word sense. But among all these learning approaches, supervised learning approach has been considered successful algorithm.

Applications are: WSD applications of language are used in technology, information retrieval,

lexicography, knowledge mining/acquisition and semantic interpretation, bioinformatics and the Semantic Web.

III. TAGGING AND RECOGNITION

POS tagging is the process of marking up a word in a body to a corresponding part of a speech tag, based on its context and definition. Part of Speech (hereby referred to as POS) Tags are useful for building parse trees, which are used in building NER's. POS Tagging is also essential for building lemmatizers which are used to reduce a word to its root form. Different types of POS Taggings are:

1. Lexical Based Method
2. Rule-BasedMethods
3. Probabilistic Methods
4. Deep Learning Methods

Speech recognition is the ability to recognize the spoken words by a electonic device. A microphone records a person's voice and the hardware converts the signal from analog sound waves to digital audio.The audio data is then processed by software, which interprets the sound as individual words.

GENERATION OF NATURAL LANGUAGE
Generative Adversarial Networks (GANs) have gathered a lot of attention from the computer vision community, yielding impressive results for image generation. Advances in the adversarial generation of natural language from noise however are not in proportion with the progress made in generating images, and still lag far behind likelihood based methods. In this paper, we take a step towards generating natural language with a GAN objective alone. We present quantitative results on generating sentences from context-free and probabilistic context-free grammars, and qualitative language modeling results.

Benefits of NLP

NLP hosts benefits such as:

✓ Enhanced accuracy and efficiency of documentation.
✓ Automatic summarizer
✓ Useful for personal assistants such as Alexa.
✓ Allows an organization to use chatbots for customer support.
✓ Permits sentimental analysis.

Challenges associated with NLP

NLP has not yet been wholly perfected. For example, semantic analysis can still be a challenge for NLP. Other difficulties include the fact that abstract use of language is typically tricky for programs to understand. For instance, NLP does not pick up sarcasm easily. These topics usually require the understanding of the words being used and the context in which the way they are being used. As another example, a sentence can change meaning depending on which word the speaker puts stress on. NLP is also challenged by the fact that language, and the way people use it, is continually changing.

IV. CONCLUSION

NLP provides a wide set of techniques and tools which can be applied in all the areas of life, by learning and

using the same in our everyday interactions, our life quality would highly improve. NLP techniques help us improving our communications, our goal reaching and the outcomes we receive from every interaction. They also allow us to overcome personal obstacles and psychological problems. NLP helps us by utilizing the tools and techniques we already have with us without being aware of it. NLP supposedly makes the job easier but still demands a human interference. People and the industry fear NLP would start a trend of job snatching which is true to a certain sense but it certainly cannot function the way it does without human inputs. The will to work

V. REFERENCES

[1].   https://searchbusinessanalytics-techtarget-com.cdn.ampproject.org/v/s/searchbusinessanalytics.techtarget.com/definition/natural-language-processing-NLP?amp_js_v=a2&amp_gsa=1&amp=1&usqp=mq331AQEKAFwAQ%3D%3D#aoh=15685236195886&referrer=https%3A%2F%2Fwww.google.c om&amp_tf=From%20%251%24s&ampshare=https%3A%2F%2Fsearchbusinessanalytics.techtar   get.com%2Fdefinition%2Fnatural-language-processing-NLP

[2].   https://ieeexplore.ieee.org/document/1587718/keywords

[3].   https://machinelearningmastery.com/natural-language-processing/

[4].   https://www.upwork.com/hiring/for-clients/artificial-intelligence-and-natural-language-processing-in-big-data/

[5].   https://www.researchgate.net/publication/243962849_Overview_of_machine_translation_techniques

[6].   https://en.m.wikipedia.org/wiki/Machine_translation

[7].   https://towardsdatascience.com/5-heroic- tools-for-natural-language-processing-7f3c1f8fc9f0