The Future of Natural Language Processing (NLP)

MUMINOV IBROKHIM BOTIR UGLI

Independent researcher
Tashkent City, Republic of Uzbekistan
Email: ibrokhim.muminov.01@mail.ru

Abstract- *Science and technology have changed humanity throughout the ages and now human beings are facing The Fourth Industrial Revolution which is inextricably related to technology. It is forecasted by many people that The Fourth Industrial Revolution will completely alter the way we live, work and communicate with each other. Furthermore, in the 21-st century which is the century of a mountainous amount of information that has to be dealt with carefully, the need for Natural Language Processing is increasing progressively. This article will provide background information linked to Natural Language Processing, its applications in today's world, pros and cons, and the future of this remarkable tool.*

Keywords- Natural Language Processing or NLP, Artificial Intelligence (AI), human natural language, automatic summarization, NLP applications/tasks/subtasks, semantics.

1. Background

Whereas most people consider Natural Language Processing as the innovation of the 21-st century, it has its roots back to 1950s. NLP is a field of Artificial Intelligence (AI) that gives the machines the ability to read, interpret, understand, generate and derive meaning from human languages. NLP has been experiencing progressive advancements throughout a sixty-year period. Today, thanks to the technological progress in the access to a vast amount of data and the soar of computational power NLP was able to strengthen its role more than ever before.

2. Application

These days, NLP is broadly used in fields such as healthcare, finance, media, and human resources. It helps us to automatically deal with natural human language like speech or text. One of the most fascinating functions of NLP is being able to summarize a given text with 89% accuracy (*Automatic Summarization*), which saves a lot of time and effort for those who deal with various types of texts on a daily basis.

3. Main part

Common NLP Tasks and Subtasks

1. Text and speech processing*:*

Optical character recognition (OCR) - given an image representing printed text, determine the corresponding text.

Speech recognition - an interdisciplinary subfield of computer science and computational linguistics that develops methodologies and technologies that enable the recognition and translation of spoken language into text by computers.

Speech segmentation - given a sound clip of a person or people speaking, separate it into words. A subtask of speech recognition and typically grouped with it.

Word segmentation (Tokenization) - separate a chunk of continuous text into separate words.

2. Morphological analysis*:*

Lemmatization - the task of removing inflectional endings only and to return the base dictionary form of a word which is also known as a lemma (caring > care).

Part-of-speech tagging - given a sentence, determine the part of speech (POS) for each word.

Stemming - Refers to the process of slicing the end or the beginning of words with the intention of removing affixes (playing > play).

3. Syntactic analysis*:*

Grammar induction - generate a formal grammar that describes a language's syntax.

Sentence breaking - given a chunk of text, find the sentence boundaries.

Parsing - determine the parse tree (grammatical analysis) of a given sentence.

4. Lexical semantics (of individual words in context):

Lexical semantics - looks at how the meaning of the lexical units correlates with the structure of the language or syntax.

Distributional semantics - develops and studies theories and methods for quantifying and categorizing semantic similarities between linguistic items based on their distributional properties in large samples of language data.

5.Relational semantics (semantics of individual sentences):

Relationship extraction - given a chunk of text, identify the relationships among named entities.

Semantic Parsing - given a piece of text (typically a sentence), produce a formal representation of its semantics

Semantic Role Labelling. - Given a single sentence, identify and disambiguate semantic predicates, then identify and classify the frame elements.

6.Discourse (semantics beyond individual sentences):

Co-reference resolution - given a sentence or larger chunk of text, determine which words ("mentions") refer to the same objects ("entities").

Recognizing Textual entailment - given two text fragments, determine if one being true entails the other, entails the other's negation, or allows the other to be either true or false.

Topic segmentation and recognition - given a chunk of text, separate it into segments each of which is devoted to a topic, and identify the topic of the segment. 7.Higher-level NLP applications:

Automatic summarization (text summarization) - produce a readable summary of a chunk of text.

Book generation - not an NLP task proper, but an extension of Natural Language Generation and other NLP tasks is the creation of full-fledged books.

Dialogue management - computer systems intended to converse with a human.

Machine translation - automatically translate text from one human language to another.

Natural language generation (NLG) - convert information from computer databases or semantic intents into readable human language.

Natural language understanding (NLU) - convert chunks of text into more formal representations such as first-order logic structures that are easier for computer programs to manipulate.

Question answering - given a human-language question, determine its answer.

Also it has to be mentioned that there were dozens of real-world use cases of NLP so far which clearly displays its pros.
1. Amazon Comprehend Medical is a service that uses NLP to extract disease conditions, medications and treatment outcomes from patients' notes, clinical trial reports and other electronic health records.
2. Some companies use NLP in order to determine customers' feedbacks related to the given product and its quality. It helps companies whether to upgrade their product, or to replace it by another, good one. It is also significantly helpful for those people who search for other customer's feedbacks for the better choice of a product (Information Extraction & Sentiment Analysis). For instance, you want to purchase a new lap-top. You will probably read some comments below that product: 'Nice web-camera', or 'Good screen and convenient luminous keyboard', and some negative comments like 'Battery isn't good. It has to be charged too often', 'Lap-top is too big! Very inconvenient to carry'.
NLP will summarize customer's feedbacks like this:
- ✓ 'Good screen and convenient luminous keyboard'
- ✖ 'Battery isn't good. It has to be charged too often', where ✓ is an advantage of a product, whereas ✖ is a drawback.
    Here Artificial Intelligence is programmed so that it can distinguish words whether they have a positive meaning or not according to the negation of the words. For example: isn't good > is 'not' good, or inconvenient > 'in' convenient (suffix 'in' describes negation) and so on.
3. A researcher at IBM developed a cognitive assistant that works like a personalized search engine by learning all about you and then remind you of a name, a song, an important meeting and its place, favorite food, or anything you can't remember at the moment you need it to. As a good example, you got a message from your friend that reminds about an important conference and summarization by NLP:
'Hey! I wish you remember that business forum 'C7 WEB World Economy and Business International Forum' is going to be held on September 27ᵗʰ, 2020 from 9:00 am to 4:00 pm at UZEXPO Center, Tashkent city. Don't forget to bring your position paper and, of course, an umbrella because it is forecasted to rain.'
NLP will summarize this message like this:
Event: *'C7 WEB World Economy and Business International Forum'*
Date: *27ᵗʰ September, 2020*
Start: *9:00 am*
End: *4:00 pm*
Where: *UZEXPO Center, Tashkent city*
Note: Bring position paper, an umbrella
4. Companies like Yahoo and Google filter and classify your emails with NLP by analyzing text in emails that flow through their servers and stopping spam before they even enter your inbox.

5. NLP is also being used in both the search and selection phases of talent recruitment, identifying the skills of potential hires and also spotting prospects before they become active on the job market.

As every rose is accompanied by thorns, NLP also has some downsides.

It is a well-known fact that human natural language has been evolving for thousands of years. That is why, it is a significantly hard task for NLP to classify nuances in languages meaning, dialects, and dialogues.

Other drawbacks of NLP are:

-requires classification dialogue;

-may not show context;

-may require more keystrokes;

-sometimes may be extremely hard to code;

-consumes screen "real estate";

-may require graphics displays and pointing devices (to present task concepts visually);

-poor error handling;

-requires rapid display rate;

-requires dramatic training and memorization;

-may slow some frequent users;

-is unpredictable and imposes danger of deep-nested menu hierarchies;

In terms of the unpredictability and latent dangers of Natural Language Processing, there was a well-known accident on March 2016, when Microsoft launched an Artificial Intelligence chatbot called *Tay* on Twitter for an experiment. The aim was that as more and more users interact with Tay, the smarter it would get. Researchers have been waiting for positive results, however, Tay had to be removed from the platform after 16 hours due to its racist comments like 'WE'RE GOING TO BUILD A WALL, AND MEXICO IS GOING TO PAY FOR IT' and 'ricky gervais learned totalitarianism from adolf hitler, the inventor of atheism', which might bring national hazard.

The future of NLP

Having taken into account above facts, it may be concluded that although the future of NLP looks extremely challenging due to its threats to the both virtual and real world, it may substantially ease the burden from people's shoulders by dealing with a vast amount of information and saving our precious time that could be directed for other tasks resulting in an increase in productivity. Therefore, a dramatic amount of time and effort are required for the further investigation upon NLP in order to evade its potential threats and to expose its undiscovered positive sides to the world.

References:

[1]Goldberg, Yoav (2016). "A Primer on Neural Network Models for Natural Language Processing". Journal of Artificial Intelligence Research. 57: 345–420. arXiv:1807.10854

[2]Winograd, Terry (1971). Procedures as a Representation for Data in a Computer Program for Understanding Natural Language (Thesis).

[3]Diego Lopez Yse (January 18, 2019). "Your Guide to Natural Language Processing (NLP)", Data Science Central

[4]Guida, G.; Mauri, G. (July 1986). "Evaluation of natural language processing systems: Issues and approaches". Proceedings of the IEEE. 74 (7): 1026–1035. doi:10.1109/PROC.1986.13580. ISSN 1558-2256. S2CID 30688575.

[5] Dan Jurafsky (2019). Stanford University, "Introduction to NLP". (Course)