MJRET
OPEN ACCESS
Multidisciplinary Journal of Research in Engineering and Technology
www.mjret.in
eISSN:2348-6953

Evolution of Natural Language Processing : Describing basics of NLP to non-basic users

Ms. Shivani Kamble, Ms. Flavia Gonsalves,
MET Institute of Computer Science,
Bandra West, Mumbai, India

Abstract: *Natural language processing (NLP) is a major area of artificial intelligence research, which in its turn serves as a field of application and interaction of a number of other traditional AI areas. Until recently, the focus in AI applications in NLP was on knowledge representation, logical reasoning, and constraint satisfaction - first applied to semantics and later to the grammar. In the last decade, a dramatic shift in the NLP research has led to the prevalence of very large-scale applications of statistical methods, such as machine learning and data mining. Naturally, this also opened the way to the learning and optimization methods that constitute the core of modern AI, most notably genetic algorithms and neural networks. In this paper we give an overview of the current trends in NLP and discuss the possible applications of traditional AI techniques and their combination in this fascinating area.*

Keywords: *Natural language processing, Artificial Intelligence, Machine Learning.*

1. INTRODUCTION

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.

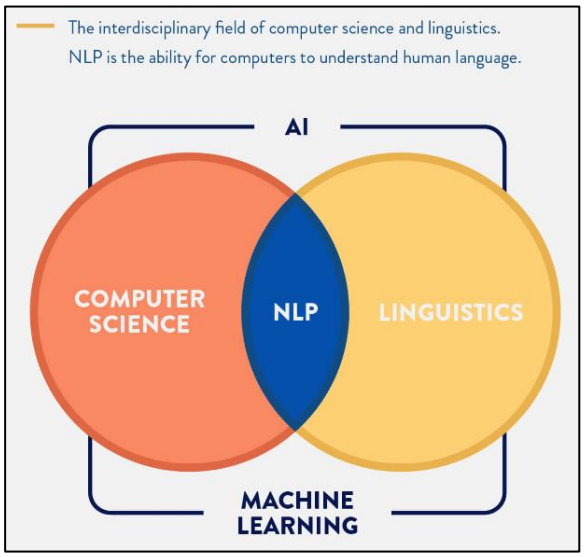

Fig.1: (source: @clevertap.com)

Natural Language Processing, or NLP for short, is broadly defined as the automatic manipulation of natural language, like speech and text, by software. The study of natural language processing has been around for more than 50 years and grew out of the field of linguistics with the rise of computers. Natural language refers to the way we, humans, communicate with each other. Namely, speech and text. We are surrounded by text. Think about how much text you see each day:

- Signs
- Menus
- Email
- SMS
- Web Pages
- and so much more…

The list is endless. Now think about speech. We may speak to each other, as a species, more than we write. It may even be easier to learn to speak than to write. Voice and text are how we communicate with each other. Given the importance of this type of data, we must have methods to understand and reason about natural language, just like we do for other types of data.

2. EVOLUTION OF NLP

While natural language processing isn't a new science, the technology is rapidly advancing thanks to an increased interest in human-to-machine communications, plus an availability of big data, powerful computing and enhanced algorithms.

As a human, you may speak and write in English, Spanish or Chinese. But a computer's native language – known as machine code or machine language – is largely incomprehensible to most people. At your device's lowest levels, communication occurs not with words but through millions of zeros and ones that produce logical actions.

Indeed, programmers used punch cards to communicate with the first computers 70 years ago. This manual and arduous process was understood by a relatively small number of people. Now you can say, "Alexa, I like this song," and a device playing music in your home will lower the volume and reply, "OK. Rating saved," in a humanlike voice. Then it adapts its algorithm to play that song – and others like it – the next time you listen to that music station.

Let's take a closer look at that interaction. Your device activated when it heard you speak, understood the unspoken intent in the comment, executed an action and provided feedback in a well-formed English sentence, all in the space of about five seconds. The complete interaction was made possible by NLP, along with other AI elements such as machine learning and deep learning.

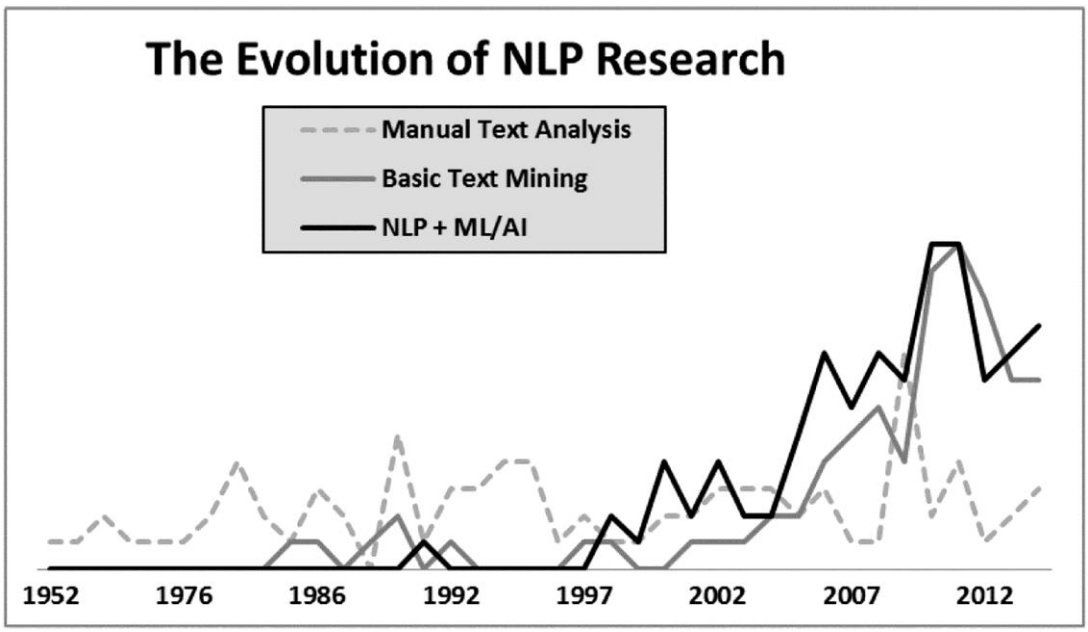

Fig.2: (source: @onlinelibrary.wiley.com)

3. WHY IS NLP IMPORTANT?

Large volumes of textual data:

Natural language processing helps computers communicate with humans in their own language and scales other language-related tasks. For example, NLP makes it possible for computers to read text, hear speech, interpret it, measure sentiment and determine which parts are important.

Today's machines can analyse more language-based data than humans, without fatigue and in a consistent, unbiased way. Considering the staggering amount of unstructured data that's generated every day, from medical records to social media, automation will be critical to fully analyse text and speech data efficiently.

Structuring a highly unstructured data source:

Human language is astoundingly complex and diverse. We express ourselves in infinite ways, both verbally and in writing. Not only are there hundreds of languages and dialects, but within each language is a unique set of grammar and syntax rules, terms and slang. When we write, we often misspell or abbreviate words, or omit punctuation. When we speak, we have regional accents, and we mumble, stutter and borrow terms from other languages.

While supervised and unsupervised learning, and specifically deep learning, are now widely used for modelling human language, there's also a need for syntactic and semantic understanding and domain expertise that are not necessarily present in these machine learning approaches. NLP is important because it helps resolve ambiguity in language and adds useful numeric structure to the data for many downstream applications, such as speech recognition or text analytics.

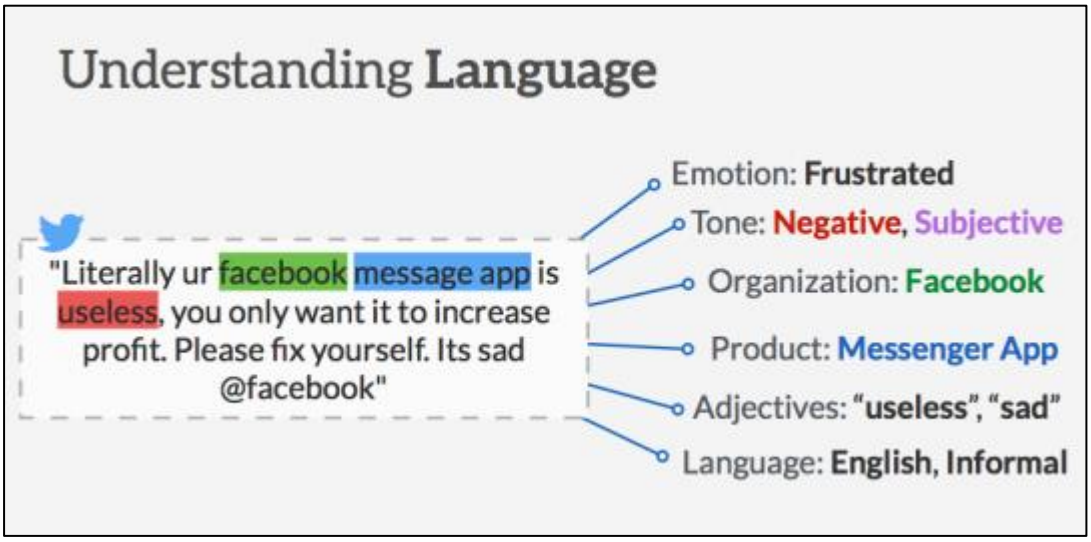

Fig.3: (source: @devopedia.org)

4. HOW DOES NLP WORK?

Breaking down the elemental pieces of language:

Natural language processing includes many different techniques for interpreting human language, ranging from statistical and machine learning methods to rules-based and algorithmic approaches. We need a broad array of approaches because the text- and voice-based data varies widely, as do the practical applications.

Basic NLP tasks include tokenization and parsing, lemmatization/stemming, part-of-speech tagging, language detection and identification of semantic relationships. If you ever diagramed sentences in grade school, you've done these tasks manually before.

In general terms, NLP tasks break down language into shorter, elemental pieces, try to understand relationships between the pieces and explore how the pieces work together to create meaning.
These underlying tasks are often used in higher-level NLP capabilities, such as:

- **Content categorization**. A linguistic-based document summary, including search and indexing, content alerts and duplication detection.
- **Topic discovery and modelling.** Accurately capture the meaning and themes in text collections, and apply advanced analytics to text, like optimization and forecasting.
- **Contextual extraction.** Automatically pull structured information from text-based sources.
- **Sentiment analysis.** Identifying the mood or subjective opinions within large amounts of text, including average sentiment and opinion mining.
- **Speech-to-text and text-to-speech conversion.** Transforming voice commands into written text, and vice versa.

- **Document summarization.** Automatically generating synopses of large bodies of text.
- **Machine translation.** Automatic translation of text or speech from one language to another.

In all these cases, the overarching goal is to take raw language input and use linguistics and algorithms to transform or enrich the text in such a way that it delivers greater value.

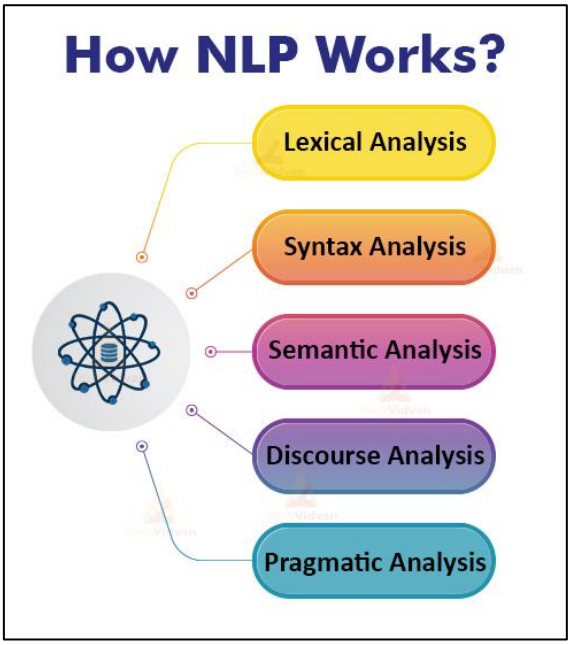

Fig.4: (source: @techvidvan.com)

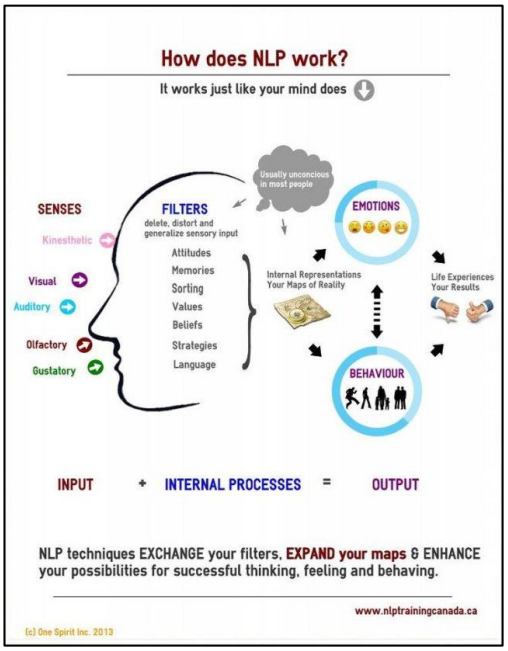

Fig.5: (source: @pinterest.es)

5. NLP MODELS

How computers make sense of textual data

NLP and text analytics:

Natural language processing goes hand in hand with text analysis, which counts, groups and categorizes words to extract structure and meaning from large volumes of content. Text analytics is used to explore textual content and derive new variables from raw text that may be visualized, filtered, or used as inputs to predictive models or other statistical methods.

NLP and text analytics are used together for many applications, including:

- Investigative discovery. Identify patterns and clues in emails or written reports to help detect and solve crimes.
- Subject-matter expertise. Classify content into meaningful topics so you can take action and discover trends.
- Social media analytics. Track awareness and sentiment about specific topics and identify key influencers.

Everyday NLP examples :

There are many common and practical applications of NLP in our everyday lives. Beyond conversing with virtual assistants like Alexa or Siri, here are a few more examples:

- Sentiment Analysis :

Owner of a hat & t-shirt company gets daily calls from customers about their reviews regarding the hats and t-shirts. He answers and listens to each and every call. Here he has two options :

1) To make a note of each and every call and keep writing all the feedback.
2) Using NLP to distinguish the calls into positive and negative feedbacks.
   (This can be done by providing a like/dislike button on the website or feedback page.)

Fig.6: (source: @storybydata.com)

- Topic Modelling :

Suppose you are working at a law suite and you have a client (bank/financial organisation). You have to find out who is making fraud transactions. But there are 2000-3000 emails, it is impossible for you to read all of it. Here the organisation can classify emails as personal, project, money($$), etc. So that it becomes easy to recognise the fraud transactions. This is where NLP can be used.

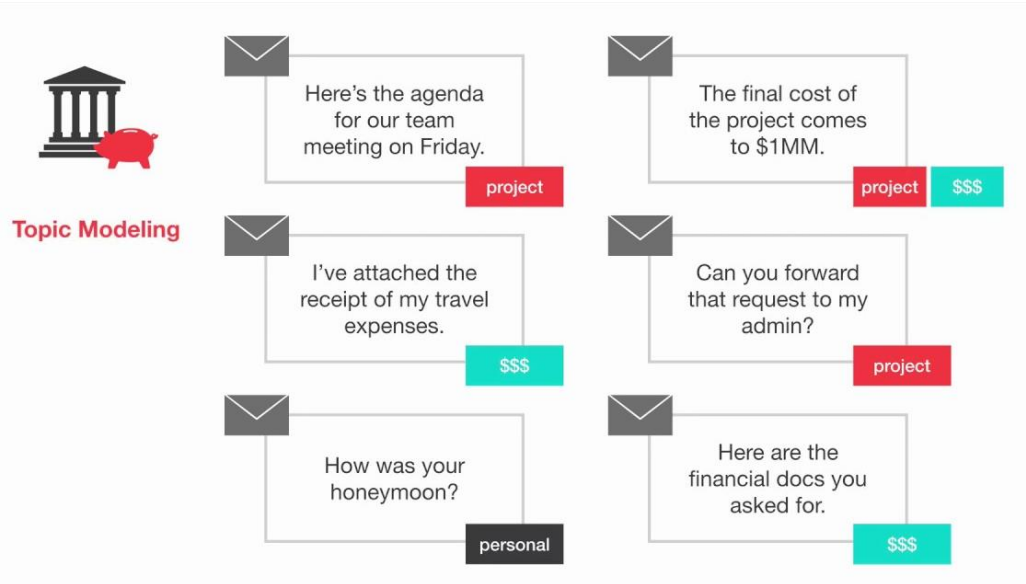

Fig.7: (source: @medium.com)

- Text Generation :

Suppose you are an author, and an organisation has hired you to write inspirational quotes for 365 days. You write quotes for 11 months and for the last 1 month you go out of ideas. Here you can create a text generator using python which will take the input from all your past quotes (data) and generate a new quote based on that.

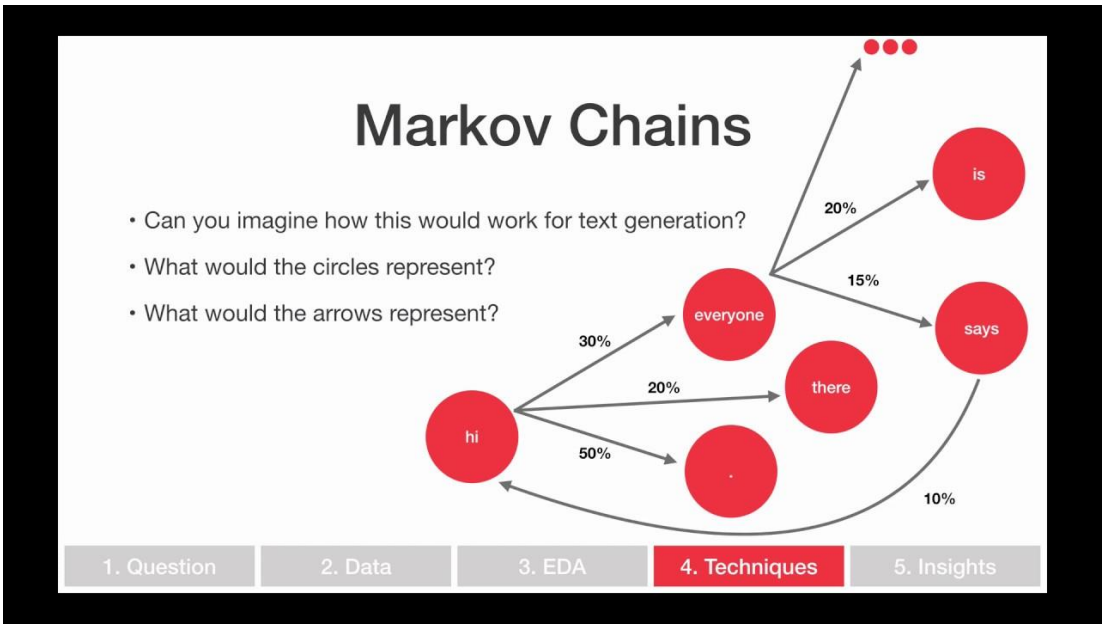

Fig.8: (source: @youtube.com)

6. FUTURE ENHANCEMENT

A subfield of NLP called natural language understanding (NLU) has begun to rise in popularity because of its potential in cognitive and AI applications. NLU goes beyond the structural understanding of language to interpret intent, resolve context and word ambiguity, and even generate well-formed human language on its own. NLU algorithms must tackle the extremely complex problem of semantic interpretation – that is, understanding the intended meaning of spoken or written language, with all the subtleties, context and inferences that we humans are able to comprehend.

The evolution of NLP toward NLU has a lot of important implications for businesses and consumers alike. Imagine the power of an algorithm that can understand the meaning and nuance of human language in many contexts, from medicine to law to the classroom. As the volumes of unstructured information continue to grow exponentially, we will benefit from computers' tireless ability to help us make sense of it all.

Experts believe that some of NLP's next steps will be huge, centring around the move from structured data (databases) to unstructured data (text), as well as an increased ability to "understand" humans as they speak normally.

At the moment NLP is battling to detect nuances in language meaning, whether due to lack of context, spelling errors or dialectal differences.

Microsoft learnt from its own experience and some months later released Zo, its second-generation English-language chatbot that won't be caught making the same mistakes as its predecessor. Zo uses a combination of innovative approaches to recognize and generate conversation, and other companies are exploring with bots that can remember details specific to an individual conversation.

7. CONCLUSION

As NLP continues to make "data" more user-friendly and conversational, more and more mainstream users will adopt NLP-driven Data Platforms. In a way, NLP will remove the current barriers to entry for Big Data BI. Someday, business users may engage in BI tasks through "conversational" interactions with smart assistants or chatbots. The "conversational platform" will encourage lot of reticent users to attempt advanced BI. This growing trend is exploring.

Although the future looks extremely challenging and full of threats for NLP, the discipline is developing at a very fast pace (probably like never before) and we are likely to reach a level of advancement in the coming years that will make complex applications look possible.

REFERENCES

[1]. *https://www.google.com/*

[2]. *https://www.youtube.com/*

[3]. *https://www.kdnuggets.com/*

[4]. *https://ieeexplore.ieee.org/*