Evolving Trends in Conversational Systems with Natural Language Processing

V. Sriguru
Engineer
Altran Technologies India Private Ltd
Coimbatore

D. Francis Xavier Christopher
Director- School of Computer Studies
Rathnavel Subramaniam College of Arts & Science
Sulur, Coimbatore

Abstract- **Today, with digitization of everything, 80% of the data being created is unstructured. Audio, video, our social footprints, the data generated from conversations between customer service reps, tons of legal documents, and texts processed in financial sectors are examples of unstructured data stored in Big Data. Organizations are turning to natural language processing (NLP) technology to derive understanding from the myriad unstructured data available online, in call logs, and in other sources. In NLP, chatbots and intelligent automation are on the rise, enterprises should look at NLP infused chatbots to drive cost saving, operational efficiencies, and enhanced customer experiences throughout their businesses.**

***Keywords*— Machine Learning, Artificial Intelligence, Machine Language, Chatbot, Syntactic Analysis, Semantic Analysis**

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. Natural Language Processing (NLP) is all about leveraging tools, techniques and algorithms to process and understand natural language-based data, which is usually unstructured like text, speech and so on. If you walk to an intersection of computational linguistics, artificial intelligence (https://livebook.manning.com/#!/book/natural-language-processing-in-action/chapter-1/v-10/185), and computer science, you are more than likely to see Natural Language Processing (NLP) there as well.

2. OVERVIEW

With the rise of voice interfaces and chat-bots, NLP is one of the most important technologies of the information age a crucial part of artificial intelligence. Fully understanding and representing the meaning of language is an extremely difficult goal. Why? Because human language is quite special. What's special about human language? A few things actually: Human language is a system specifically constructed to convey the speaker/writer's meaning. It's not just an environmental signal but a deliberate communication. Besides, it uses an encoding that little kids can learn quickly; it also changes. Human language is mostly a discrete/ symbolic/ categorical signaling system, presumably because of greater signaling reliability. The categorical symbols of a language can be encoded as a signal for communication in several ways: sound, gesture, writing, images, etc. human language is capable of being any of those. Human languages are ambiguous (unlike programming and other formal languages); thus, there is a high level of complexity in representing, learning, and using linguistic / situational / contextual / word / visual knowledge towards the human language.

3. EVOLUTION

NLP began in the 1950s as machine translation (MT). These early MT efforts were intended to aid in code-breaking during World War II. Developers hoped MT would translate Russian into English, but results were unsuccessful. Although the translations were not successful, these early stages of MT were necessary stepping stones on the way

to more sophisticated technologies. Developed in the 1960s, ELIZA and SHRDLU are two successful tokens of early NLP. SHRDLU was primarily a language program that allowed user interaction with a block world using English terms. A user could ask the program to move or manipulate the blocks, and the computer would respond. 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 (https://www.forbes.com/sites/forbestechcouncil/2018/11/06/the-evolution-of-natural-language-processing-and-its-impact-on-ai/#128260781119).

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.

4. COMPONENTS

(i) Natural Language Understanding (NLU): Mapping the given input in natural language into useful representations and analyzing different aspects of the language.

(ii) Natural Language Generation (NLG): It is the process of producing meaningful phrases and sentences in the form of natural language from some internal representation. It involves,

Text planning − It includes retrieving the relevant content from knowledge base.

Sentence planning − It includes choosing required words, forming meaningful phrases, setting tone of the sentence.

Text Realization − It is mapping sentence plan into sentence structure.

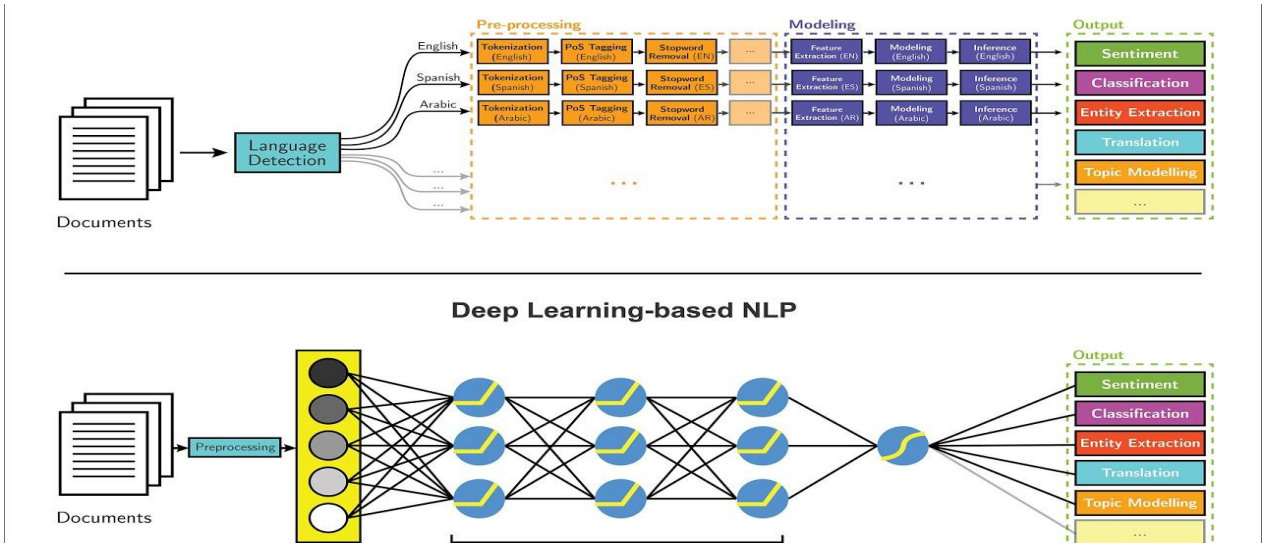

Figure 1. Why study NLP

There's a fast-growing collection of useful applications derived from this field of study. They range from simple to complex. Below are a few of them:

- Spell Checking, Keyword Search, Finding Synonyms.

- Extracting information from websites such as: product price, dates, location, people, or company names.
- Classifying: reading level of school texts, positive/negative sentiment of longer documents.
- Machine Translation.
- Spoken Dialog Systems.
- Complex Question Answering.

Indeed, these applications have been used abundantly in industry: from search (written and spoken) to online advertisement matching;from automated / assisted translation to sentiment analysis for marketing or finance/trading; and from speech recognition to chatbots/dialog agents(automating customer support, controlling devices, ordering goods).
Deep Learning: Most of these NLP technologies are powered by Deep Learning in Figure 1 a subfield of machine learning. Deep Learning only started to gain momentum again at the beginning of this decade, mainly due to these circumstances: Larger amounts of training data and Faster machines and multicore CPU/GPUs.

New models and algorithms with advanced capabilities and improved performance: More flexible learning of intermediate representations, more effective end-to-end joint system learning, more effective learning methods for using contexts and transferring between tasks, as well as better regularization and optimization methods.

Most machine learning methods work well because of human-designed representations and input features, along with weight optimization to best make a final prediction. On the other hand, in deep learning, representation learning attempts to automatically learn good features or representations from raw inputs. Manually designed features in machine learning are often over-specified, incomplete, and take a long time to design and validate. In contrast, deep learning's learned features are easy to adapt and fast to learn.

Deep Learning provides a very flexible, universal, and learnable framework for representing the world, for both visual and linguistic information. Initially, it resulted in breakthroughs in fields such as speech recognition and computer vision. Recently, deep learning approaches have obtained very high performance across many different NLP tasks. These models can often be trained with a single end-to-end model and do not require traditional, task-specific feature engineering.

5. STEPS

The steps are described in the figure 2,

- Lexical Analysis: It involves identifying and analyzing the structure of words. Lexicon of a language means the collection of words and phrases in a language. Lexical analysis is dividing the whole chunk of txt into paragraphs, sentences, and words.

- Syntactic Analysis (Parsing): It involves analysis of words in the sentence for grammar and arranging words in a manner that shows the relationship among the words. The sentence such as "The school goes to boy" is rejected by English syntactic analyzer.

- Semantic Analysis: It draws the exact meaning or the dictionary meaning from the text. The text is checked for meaningfulness. It is done by mapping syntactic structures and objects in the task domain. The semantic analyzer disregards sentence such as "hot ice-cream".

- Discourse Integration: The meaning of any sentence depends upon the meaning of the sentence just before it. In addition, it also brings about the meaning of immediately succeeding sentence.

- Pragmatic Analysis: During this, what was said is re-interpreted on what it actually meant. It involves

deriving those aspects of language which require real world knowledge.

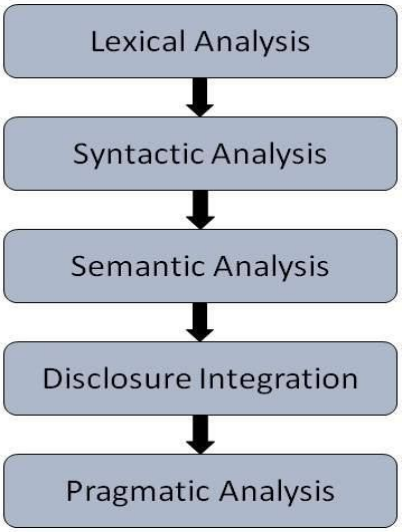

Figure 2. Steps

6. NLP TODAY

NLP has come a long way from MT, which would be laughable to us now, given the vast array of technology available. Today, families around the world are welcoming virtual additions like Alexa into their homes. In fact,47.3 million U.S. adults now own a smart speaker, a platform wholly dependent on NLP for survival by intaking a user's commands and applying algorithms to decipher language and formulate responses (https://www.lexalytics.com/ lexablog/ machine-learning-vs-natural-language-processing-part-1).

Chatbots are another implementation of NLP on the rise. They rely on NLP technology to formulate applicable responses to customer questions by analyzing the language typed into the text fields. Chatbots not only streamline incoming FAQs but also allow customers to access new information or be rerouted to relevant pages almost instantaneously, providing a value proposition on both ends of the communication. In a recent Oracle survey, 80% of respondents said they already used or planned to use chatbots by 2020 for consumer-facing products.

In fact, Google recently announced expansions to its Cloud AutoML platform, including NLP and translation, while Hearst already benefits from the technology and its ability to organize international and domestic content automatically. Among the industries impacted by AI-based communications, talent acquisition is highly susceptible to significant disruption given its innate people-centric and communicative nature. NLP plays an important role in increasing accuracy in candidate matching from large talent pools.

NLP also aids in guiding applicants using chatbots, simplifying scheduling, making accessible job descriptions, intuitive resume pairing and more. In an AI-driven world, it's not surprising that nearly all industries are impacted by NLP.

7. A CHATBOT NATURAL LANGUAGE PIPELINE

The NLP pipeline required to build a dialog engine, or chatbot, is similar to the pipeline required to build a question answering system described in *Taming Text.* Each of the four processing stages can contain one or more processing algorithms working in parallel or in series (see Figure 3).

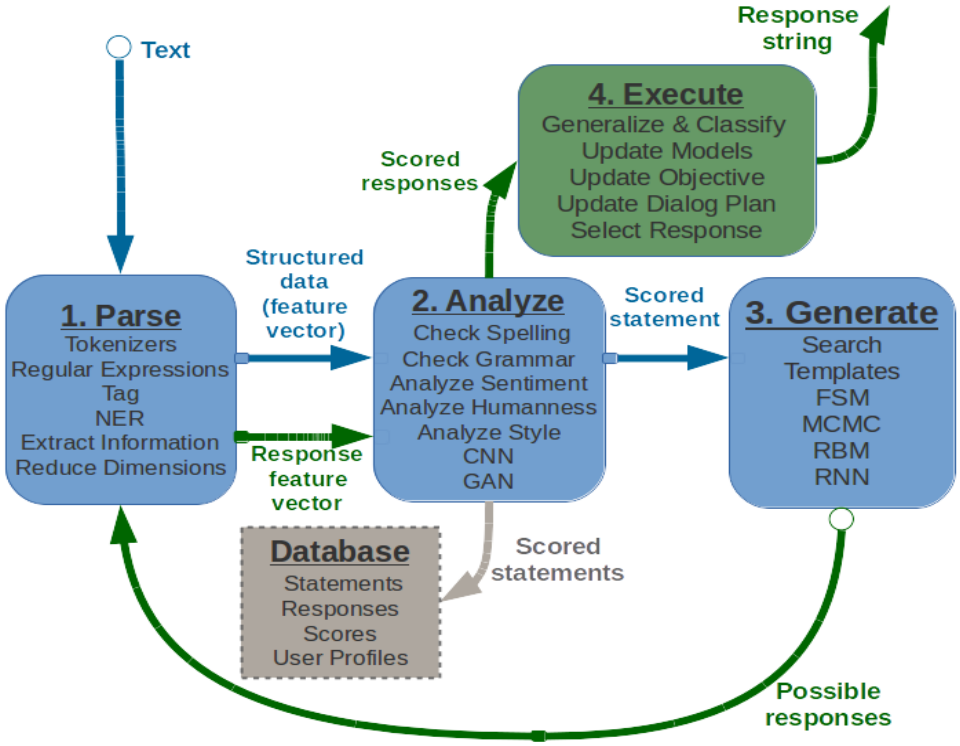

Figure 3. A chatbot requires four kinds of processing as well as a database to maintain a memory of past statements and responses

- **Parse:** Extract features, structured numerical data, from natural language text.
- **Analyze :** Generate and combine features by scoring text for sentiment, grammaticality, semantics.
- **Generate :** Compose possible responses using templates, search, or language models.
- **Execute :** Plan statements based on conversation history and objectives, and select the next response

Each of these four stages can be implemented using one or more of the algorithms listed within the corresponding boxes in the block diagram. We show you how to use Python to accomplish near state-of-the-art performance for each of these processing steps. And we show you several alternative approaches to implementing these five subsystems. Most chatbots will contain elements of all five of these subsystems (the four processing stages as well as the database). But many applications require only simple algorithms for many of these steps. Some chatbots are better at answering factual questions, and others are better at generating lengthy, complex, convincingly human responses. Each of these capabilities require different approaches; we show you techniques for both.

In addition, deep learning and data-driven programming (machine learning, or probabilistic language modeling) have rapidly diversified the possible applications for NLP and chatbots. This data-driven approach allows ever greater sophistication for an NLP pipeline by providing it with greater and greater amounts of data in the domain you want to apply it to. And when a new machine learning approach is discovered that makes even better use of this data, with more efficient model generalization or regularization, then large jumps in capability are possible.

The NLP pipeline for a chatbot shown in figure 3 contains all the building blocks for most of the NLP applications that we described at the start of this chapter. As in *Taming Text*, we break out our pipeline into four main subsystems or stages. In addition we've explicitly called out a database to record data required for each of these stages and persist their configuration and training sets over time. This can enable batch or online retraining of each of the stages as the chatbot interacts with the world. In addition we've shown a "feedback loop" on our generated text responses so that our responses can be processed using the same algorithms used to process the user statements. The response "scores" or features can then be combined in an objective function to evaluate and select the best possible response, depending on the chatbot's plan or goals for the dialog (https://heartbeat.fritz.ai/the-7-nlp-techniques-that-will-change-how-you-communicate-in-the-future-part-i-f0114b2f0497). This book is focused on configuring this NLP pipeline for a chatbot, but you may also be able to see the analogy to the NLP problem of text retrieval or "search," perhaps the most common NLP application. And our chatbot pipeline is certainly appropriate for the question answering application that was the focus of *Taming Text*.

The application of this pipeline to financial forecasting or business analytics may not be so obvious. But imagine the features generated by the analysis portion of your pipeline. These features of your analysis or feature generation can be optimized for your finance or business prediction. That way they can help you incorporate natural language data into a machine learning pipeline for forecasting. Despite focusing on building a chatbot, this book gives you the tools you need for a broad range of NLP applications, from search to financial forecasting.

One processing element in figure that is not typically employed in search, forecasting, or question answering systems is natural language **generation**. For chatbots this is their central feature. Nonetheless, the text generation step is often incorporated into a search engine NLP application and can give such an engine a large competitive advantage. The ability to consolidate or summarize search results is a winning feature for many popular search engines (DuckDuckGo, Bing, and Google). And you can imagine how valuable it is for a financial forecasting engine to be able to generate statements, tweets, or entire articles based on the business-actionable events it detects in natural language streams from social media networks and news feeds.

8. THE FUTURE FOCUS OF NLP

Until now, we have interacted with computers in ways they understand, adapting to their language, rather than having them adapt to ours. But within the past few decades, they have begun to learn our language. Before long, business transforming, life-changing information will not only be discovered merely by talking with a chatbot but will impact many other parts of our lives.

While it's certainly amazing to contemplate what's to come, voice-activated searches and commands can still be frustratingly hit-or-miss. When they work the way the user would like, it's because the question can be answered with highly structured data. The challenge is identifying which specific structured data to use, how to gather it and how much of it is necessary to understand and answer a human's question. NLP aims to eventually dominate human-to-machine interaction to the point where talking to a machine is as easy as talking to a human. NLP will continue to harness unstructured data and make it more meaningful to a machine. IDC recently forecasted that "the amount of analysed data 'touched' by cognitive systems will grow by a factor of 100 to 1.4 ZB by 2025," impacting thousands of industries and companies around the globe. Robotics, health care, financial services, connected auto and smart homes are just a handful of the sectors that will continue to be advanced by NLP.

Joining forces with big data, NLP will play a critical role in deriving business intelligence from raw business data, including product data, sales and marketing data, customer support, brand reputation and the current talent pool of an enterprise. This means NLP will be the key to shifting many legacy companies from data-driven to intelligence-driven platforms, helping humanity quickly get the insights they need to make decisions.

NLP has changed the way we interact with computers and machines. What started out as complicated, handwritten

formulas is now a streamlined set of algorithms powered by AI. NLP technologies will be the underlying force for transformation from data-driven to intelligence-driven endeavors, as they shape and improve communication technology in the years to come.

9. CONCLUSION

Therefore, Natural Language Processing takes a very important role in new machine human interfaces and conversational systems. When we look at some of the products that are based on technologies with NLP we can see that they are very advanced but very useful. But there are many limitations, requiring improvements and development of NLP oriented systems. For example, language we speak is highly ambiguous. This makes it very difficult understand and analyze. Also, with so many languages spoken all over the world it is very difficult to design a system that is 100 % accurate. These problems get more complicated when we think of different people speaking the same language with different styles. Therefore, most of research on speech recognition is more concentrated on their areas. Information retrieval can be improved to give very accurate results for various searches. This will involve intelligence to find and sort all the results. So, such intelligent systems are being experimented right now are we will be able to see improved applications of NLP in the near future.

REFERENCES

https://livebook.manning.com/#!/book/natural-language-processing-in-action/chapter-1/v-10/185

https://www.forbes.com/sites/forbestechcouncil/2018/11/06/the-evolution-of-natural-language-processing-and-its-impact-on-ai/#128260781119

https://www.lexalytics.com/lexablog/machine-learning-vs-natural-language-processing-part-1

https://heartbeat.fritz.ai/the-7-nlp-techniques-that-will-change-how-you-communicate-in-the-future-part-i-f0114b2f0497