Natural Language Processing for Enterprise Applications

A. Bharathi[1]

Abstract:

Researchers are concentrating on more efficient communication technologies that can emulate human interactions and comprehend natural languages and human emotions as a result of people's growing reliance on computer-assisted systems. Unstructured data, which is deemed useless, has increased due to information overload in every industry, including business, healthcare, education, etc. In this context, natural language processing (NLP) is one of the efficient technologies that may be used along with more sophisticated technologies, such as machine learning (ML), artificial intelligence (AI), and deep learning, to enhance the interpretation and processing of natural language. In addition to improving human-computer interaction, this can also enable massive amounts of useless and unstructured data to be analysed and formatted in numerous industrial applications. This will produce significant results that can improve decision-making and hence increase operational effectiveness. This paper introduces the idea of NLP, its background, and its current state while also going through examples of its use in various industrial fields.

Keywords: Natural Language Processing, Artificial Intelligence, Machine Learning.

[1]Advanced Technology Group (CoE), Renault Nissan and Technology Business Centre India, Chennai

1. INTRODUCTION:

Businesses are working to create a data-driven business, where all corporate decisions are based on data and not always on team members' intuition. They are on a mission to gather useful data and successfully use it to automate procedures and save time. In the financial services sector, a lot of data is produced. The methods used to collect, evaluate, analyse, and store this data have a significant impact on its value. To make use of data's possibilities and create new business prospects, financial firms have begun implementing cutting-edge technology and solutions [1]. Robotic Process Automation (RPA), AI, ML, and big data analytics are examples of technologies that assist in monitoring vast amounts of generated data, spotting trends, and making predictions using those patterns. AI's NLP subcategory seeks to overcome the complexity in processing the natural language by deciphering and synthesising human language. NLP applications use a variety of computational approaches to analyse and represent texts, for a variety of purposes. Research in the area of computational linguistics, which is concerned with analysing spoken and written language from a computer standpoint, is also the foundation of NLP[8].

The technological advancements that surround us have produced and processed massive databases with both structured and unstructured data including audio-visual data, textual data, and other data [4]. The necessity for an effective method that translates these massive volumes of information while upholding users' rights has increased due to the rapid development of technology and the expanding volumes of data. One of the objectives of AI, which is described as "a system that can rationally solve difficult issues or take actions to achieve its goals in the diverse contexts it meets in the real world,"(source) is the conversion of information into knowledge. Applications that use AI are widespread, and this is especially true of narrow AI applications that concentrate on a certain set of issues and tasks. While organised data can be processed automatically, unstructured data, such as audio recordings or textual information is far more difficult to process and to derive insights from, by employing algorithms[12]. This

difficulty is exacerbated by the fact that textual sources contain much of the pertinent information that needs to be processed. The business sector in general, and the capital market in particular, heavily relies on textual sources, and NLP systems can improve some text-based tasks like financial statement analysis, news feed processing, and automated client contacts (e.g.,chat-bots) [2].

2. ELEMENTS OF NLP TECHNOLOGY:

Human language has a complicated structure and understanding it necessitates a series of hierarchical processing tasks, including understanding individual words and determining the relationships between individual words and phrases[9]. Because of this, NLP technologies execute complex processes to evaluate natural text. This section provides an overview of each phase, from the most detailed to the most general. We will also go through the most typical data sources that NLP systems employ. Note that some technical details have been simplified because this paper is intended as an introduction to NLP technology.

2.1 Text Pre-processing

These processes analyse the text down to the level of individual sentences and words.

- **Breaking down into sentences:** Textual sources often consist of vast collections of files that are broken down into sentences and paragraphs. Therefore, segmenting the material into sentences is a fundamental step in the study of these sources. The ability to examine each sentence separately and assign the resulting meanings to the paragraph and text levels is made feasible by this split.

- **Sentence tokenization:** Tokens, the basic units that NLP systems process, are then separated into each sentence. A building block (or token) is typically specified as a single word, though this definition may vary depending on the application.

- **Part-of-speech tagging:** Each word or token in the phrase is connected to a specific part of speech. As an illustration, the token "*Aya*" will be tagged as the sentence's subject and the token "*bought*" as the verb in the sentence "*Aya bought a share from Boaz*"." A corpus that acts as the benchmark database for many NLP systems contains several samples of typical part-of-speech tagged sentences.

- **Lemmatization:** Lemma is the canonical form, dictionary form, or citation form of a set of word forms. Words are inflected or modified in most languages for different grammatical functions (e.g., verbs are inflected for time, and nouns are inflected for numbers). Words are transformed into their common base forms, which are derived from inflectional forms, to derive meanings from them uniformly.

- **Recognising and eliminating stop words:** Connecting words like "with" and "on" provide relationships and links between the numerous lemmas in a phrase. As a result, these stop words are marked differently from syntactical sentence components such as the subject and object.

- **Analysis of dependencies in the sentence:** Based on the elements in a sentence, a syntactical structure is assigned to it. The relationships between the words are defined by the syntactical framework.

- **Entity extraction:** At this stage, the names of people, locations, and businesses in the text as well as their exact locations are classified as entities. It should be noted that even an object can appear in various forms in the text.

- **Coreference resolution:** Different terms may refer to the same thing in a natural language document (for example, "The share climbed by 5%. This year, the prices hit all-time highs.) This process discovers all the references that are connected to the same entity.

- **Information extraction:** Using a variety of techniques, this

task extracts structured information (data and context) from a text.

2.2. NLP tools

The collection of higher-level NLP applications that focuses on extracting meaning[10] from a text include the tools and tasks that are listed below, which come after text pre-processing:

- **Topic-modelling:** Based on the frequency of specific terms in the text, this tool determines the themes covered in the text. Additionally, topics are found by contrasting the frequency of phrases in the target text and other texts. For instance, even when a text merely contains terms that are related to politics and economics, such as elections, democrat, speaker of the house, or budget, the system still recognises that the content is about politics and economics.

- **Sentiment analysis:** This method categorises texts according to the sentiment or opinion they reflect (e.g., positive, negative, or neutral) , as well as the strength of those feelings. For instance, automated sentiment analysis across a huge number of texts enables one to comprehend public opinion on a particular problem in real-time, particularly when texts are gathered from social media, on how contentious a topic is, or how much attention it garners.

- **Relational extraction:** This tool shows how two things are related.

a. Relative size (e.g., Company X has a bigger market cap than Company Y) is one meaning that a connection may express.

b. Affinity - For example, Share X and Bonds Y belong to Company Z.

c. Ranking - (e.g., Shares Y and X are ranked first and second, respectively, in Index Z in terms of the market cap).

- **Text classification and clustering:** Text classification uses predetermined thematic categories to determine the topic of a text given a huge collection of texts whose content is unknown, such as an uncatalogued database of public inquiries. Without pre-classifying probable topics, a comparable operation called clustering also finds relationships between texts whose content is unknown.

3. STRATEGIES TO IMPLEMENT NLP:

These NLP components are developed using a range of ML methods, which can be divided into two types of learning methods: supervised learning and unsupervised learning[11]. These are briefly described below.

3.1 Supervised Learning: Given a set of input data and output labels, supervised learning creates a model that is used to produce a certain output. This kind of model effectively picks up knowledge *by example*. We utilize part-of-speech tagging to demonstrate how supervised learning is used in NLP technology. A collection of sentences, the input data, and part-of-speech tags make up the set of input and output data needed to create a part-of-speech tagging model which are the desired outputs for the input data. The sample tags will have undergone human expert preparation and verification. Based on the provided sample input and output data, the model will automatically learn to tag inputs during supervised learning.

3.2 Unsupervised Learning: Unsupervised learning relies solely on a predetermined set of untagged inputs, from which the model infers relations and patterns. We use text clustering as an example of an NLP application based on unsupervised learning . A database of unknowable documents could make up the given set of inputs. The model builds links between documents based on similarities based on several statistical factors using unsupervised learning (e.g., similar frequencies of specific words).

3. BUSINESS DIMENSION USES OF NLP:

Figure 1: Applications in Business Domains [Source: Wikipedia]

Figure 1 shows the relationship between text-object devices that inform customers about the company's products and enable them to learn and engage with them. The components in the image represent Industry 4.0[3] technological choices that today employ NLP to complete many business tasks. The following apps have emerged as the best in the business world today as a result of improvements in technology and the digitisation of corporate operations. The applications for NLP processing highlight the significance of NLP for commercial companies. .

Figure 1 explains the most common applications used by businesses to advance their goals and how they are suited for the business[5]. Virtual support manages commonly asked questions, and smart communication tools including audio and video conferencing. Google Translate is mostly used for smart contracts. These choices make it easier for customers—including those with special needs or disabilities—to interact with a company's products[6]. The most popular applications used now for commercial purposes are shown in Figure 1. Modern company operations are striving for improved

choices to please clients in pace with technological advancements. With this approach, company operations are attuned to the needs of customers.

4.1 **Text Summarisation:** The NLP application simplifies and synthesises text that is often found in search results, into synopses on databases and indexes for academic study. Natural language generation (NLG) and semantics are used in text summarisation and categorisation, which contribute significant textual material and conclusions for documents.

4.2. **Sentimental Analysis:** Sentiment analysis is an NLP component used with contemporary business application systems that reveals the hidden meaning of consumer comments and communications regarding commercial goods and services, in open-source data. Nowadays, social media excerpts, posts, and reviews are crucial tools for businesses to learn more about their services and goods.

4.3. **Chatbots and Virtual Assistants:** Virtual agents and chatbots are increasingly very common and frequently the initial point of contact for new consumers. Amazon Alexa has a speech recognition system, command functions, and NLG that offers the business suitable and beneficial comments. Modern virtual assistants pick up contextual clues from human inquiries and utilise it to respond to client requests with even better information.

4.4. **Machine Translation:** Google Translate is one of the most used NLP tools. The capacity to switch out one language for another is provided by NLP as machine translation. This is one of the largest NLP technologies and has contributed to a friendlier society. Language barriers are eliminated by the availability of the internet. To capture tone and emotions, NLP is currently used in conjunction with semantic systems. Moreover, AI is getting better and better at attaching feelings and emotions to a variety of NLP objects. Currently, methods are used to record tone, emotions, and thoughts. .

4.5. **_Spam detection:_** The application of natural language processing (NLP) has advanced with a contemporary twist where dangers in text and unhealthy contextual data can be identified. The receiver also thinks the text is genuine. But, an NLP application detects the genuineness of an email in the recipient's inbox. .

4.6 **_Frequently Asked Questions (FAQs):_** FAQs include questions and answers specifically customized to the company's goods and services. The FAQs are one of the NLP applications for business insiders as well as potential customers. Nowadays, most businesses use FAQ pages to measure website visitors and for the future use of potential and future consumers.

5. CONCLUSION:

This paper focuses on the key methods and technologies NLP uses to successfully learn and comprehend natural language. Enhancing communication between humans and machines across different formats can efficiently address the problems with translation and transliteration. Based on the assessment, it is clear its inclusion in different industry-based solutions has grown recently. It is used to range from chatbots on e-commerce platforms to streamline consumer interactions and to make them more engaging to simplifying complex business processes like forecasting and decision-making in stock market trading. Moreover, when NLP is combined with cutting-edge technologies like ML, AI, and deep learning, the results can be more accurate than when used with conventional techniques. Its use in artificial intelligence, robotics, and other cutting-edge systems, however, is not well studied. In conclusion, NLP approaches have enormous potential in the fields of robotics and business intelligence in the future,s given their effectiveness in enhancing the accuracy of data analysis and processing of natural language.

REFERENCES:

Perazzoli, S.; de Santana Neto, J.P.; de Menezes, M.J.M.B. Systematic analysis of constellation-based techniques by using Natural Language Processing. Technol. Forecast. Soc. Chang. 2022, 179, 121674.

Io, H.N.; Lee, C.B. Chatbots and conversational agents: A bibliometric analysis. In Proceedings of the 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, 10–13 December 2017; pp. 215–219.

Singhal, N. An empirical investigation of Industry 4.0 preparedness in India. Vision 2021, 25, 300–311.

Zhou, K.; Liu, T.; Zhou, L. Industry 4.0: Towards future industrial opportunities and challenges. In Proceedings of the IEEE 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Zhangjiajie, China, 15–17 August 2015; pp. 2147–2152.

Bittencourt, V.L.; Alves, A.C.; Leão, C.P. Lean Thinking contributions for Industry 4.0: A systematic literature review. IFACPapersOnLine 2019, 52, 904–909.

Teixeira, J.E.; Tavares-Lehmann, A.T.C. Industry 4.0 in the European union: Policies and national strategies. Technol. Forecast. Soc. Chang. 2022, 180, 121664.

Pilc, M.; Wo ´zniak-J ˛echorek, B.; Wo ´zniak, K.; Pi ˛atek, D. Industry 4.0 in the Messages Published by Employers and Trade Unions in France, Germany, Poland, and the UK. In Digitalization and Firm Performance; Palgrave Macmillan: Cham, Switzerland, 2022; pp. 157–188.

Chen, Q.; Leaman, R.; Allot, A.; Luo, L.; Wei, C.H.; Yan, S.; Lu, Z. Artificial intelligence (AI) in action: Addressing the COVID-19 pandemic with natural language processing (NLP). 2020

Tanniru, M.R.; Agarwal, N.; Sokan, A.; Hariri, S. An agile digital platform to support population health—A case study of a digital platform to support patients with delirium using IoT, NLP, and AI. Int. J. Environ. Res. Public Health 2021, 18, 5686.

Mohapatra, K.; Nayak, M.; Nayak, A.K. Techniques Behind Smart Home Automation System Using NLP and IoT. In Green Technology for Smart City and Society; Springer: Singapore, 2021; pp. 105–115.

Akerkar, R. Artificial Intelligence for Business; Springer: Berlin/Heidelberg, Germany, 2019.

Wamba-Taguimdje, S.L.; Wamba, S.F.; Kamdjoug JR, K.; Wanko, C.E.T. Influence of artificial intelligence (AI) on firm performance: The business value of AI-based transformation projects. Bus. Process Manag. J. 2020, 26, 1893–1924.