The potential of artificial intelligence to promote grit

Natural Language Processing

Natural Language Processing (NLP) is a discipline that aims to develop automatic methods for understanding text in human language. This is an expanding area of study and today plays a role of absolute centrality, due to the ever-increasing computerization of activities in today's society. With the advent of the Internet, in fact, it is enormously increasing the availability of information, messages and textual content that is necessary to analyze. Natural Language Processing allows this process to be automated, as well as with many other activities related to text in natural language: the classification of documents, the translation from one language to another, the extraction of information and so on. It is therefore clear that NLP constitutes an extremely useful resource, with wide applications.

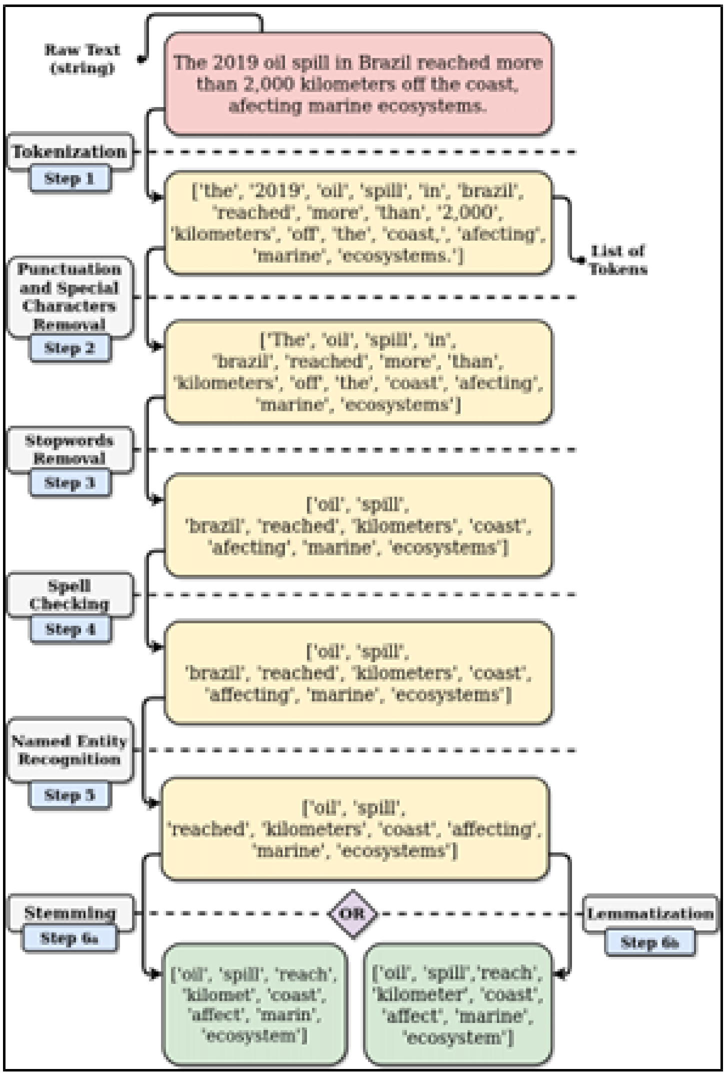
Natural Language Processing is one of the most difficult tasks for a computer, due to the ambiguities inherent to human language: many words have different meanings based on the context and often the sentences must be interpreted metaphorically rather than literal. Furthermore, very often the text to be analyzed come from the web, where grammatical errors, improper use of punctuation and the presence of special characters or emoticons or emojis are quite frequent: this only increases the difficulty of the analysis of natural language.

NLP has two main tasks:

- **Natural Language Understanding** (NLU), which focuses, as suggested by the name, on semantic analysis and on understanding the meaning of a text in natural language.

- **Natural Language Generation** (NLG), which instead deals with the generation of natural language by the machine.

Text pre-processing is an important first step that needs to be performed. It consists of various tasks such as: *tokenization* and *splitting* (that is subdividing the text in groups of words called chunks), elimination of punctuations and special characters, elimination of *stop-words*, *named entity recognition* (NER), and finally *stemming* or *lemmatization*. This process is summarized in the figure below.



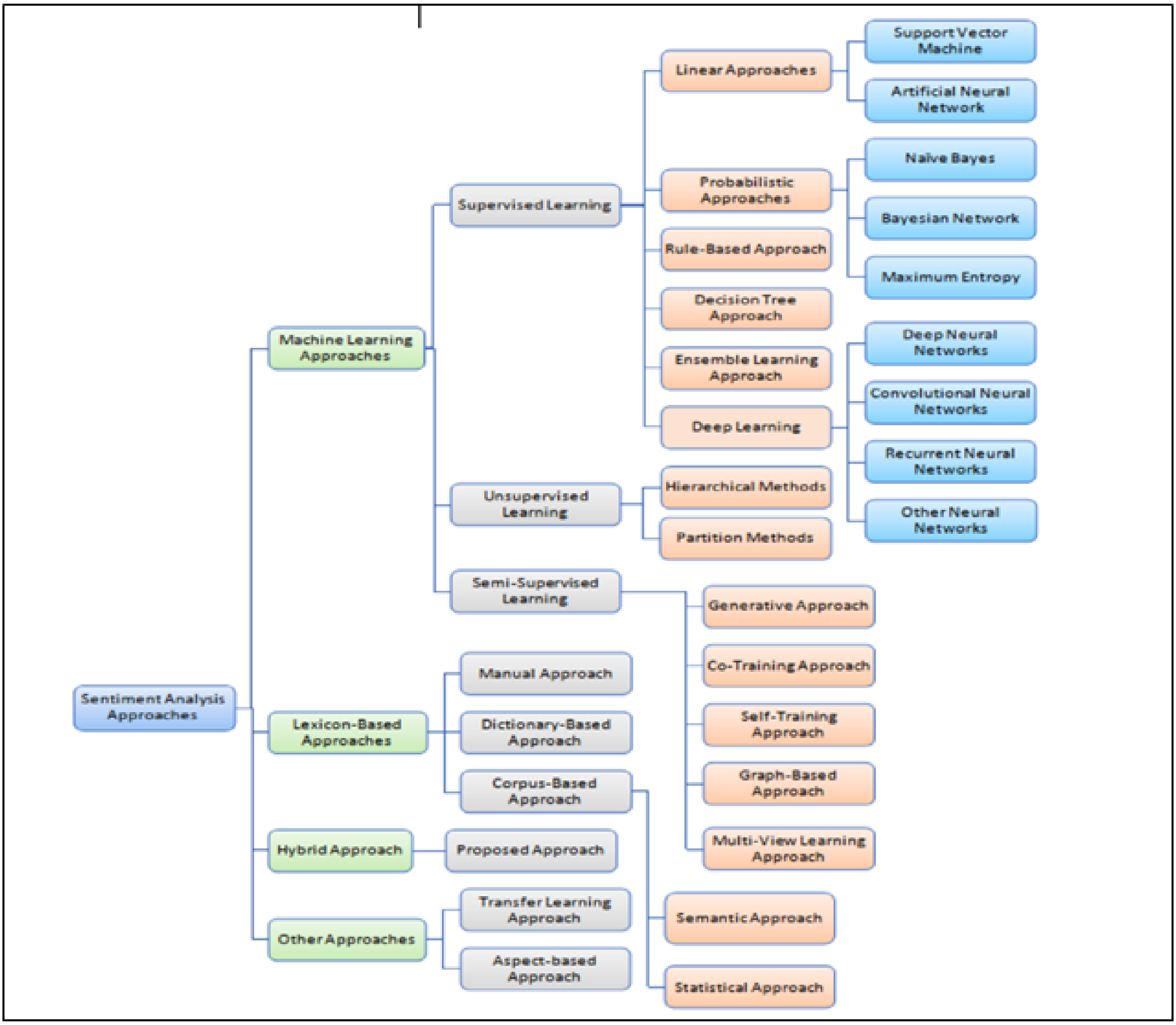
Source: de Oliveira, N.R., Pisa, P.S., Lopez, M.A., de Medeiros, D.S.V. and Mattos, D.M., 2021. Identifying fake news on social networks based on Natural Language Processing: trends and challenges. Information, 12(1), p.38.

There are many methods and techniques currently used for the fundamental task of text segmentation: among these, an example of great relevance is the Hidden Markov Models. They represent extensions of Markov stochastic processes, in which the sequence of events is not directly knowable. Hidden Markov models were first described in the second half of the 1960s, in statistical studies carried out by the American mathematician Leonard E. Baum.

Sentiment Analysis (SA, also known as opinion mining), is a field in NLP which focuses on identifying and interpreting emotions and opinions expressed in digital texts. The main goal is to analyze whether a certain text expresses a positive, negative, or neutral feeling.

If the labels associated with a text are only positive or negative, we are in the context of binary classification. Otherwise, if labels can assume a greater number of values to better describe the rating scale, we have a multi-class classification problem.

Sentiment Analysis can be seen as an application of text classification, which deals with assigning one or more categories to a given document. The primary task is the categorization of a text. Sentiment Analysis focuses exclusively on reading sentiment and tone emotional expression. Figure below shows different approaches for Sentiment Analysis.



Source: Birjali, M., Kasri, M. and Beni-Hssane, A., 2021. A comprehensive survey on Sentiment Analysis: Approaches, challenges and trends. Knowledge-Based Systems, 226, p.107134.

Current trends in NLP

In recent years, the field of Natural Language Processing (NLP) has been transformed by the advent of Deep Learning, which has significantly enhanced the capabilities of Sentiment Analysis. Neural networks, particularly Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), have become key tools in understanding and interpreting human language. RNNs are adept at handling sequences, making them ideal for analyzing the flow and progression of sentiments in texts. CNNs, on the other hand, are effective in extracting features from textual data, aiding in the identification of key sentiment indicators.

These advanced techniques have found applications in various domains, such as monitoring social media sentiment, analyzing customer feedback, and conducting market research. They have enabled businesses and organizations to gain deeper insights into public opinion, customer satisfaction, and market trends.

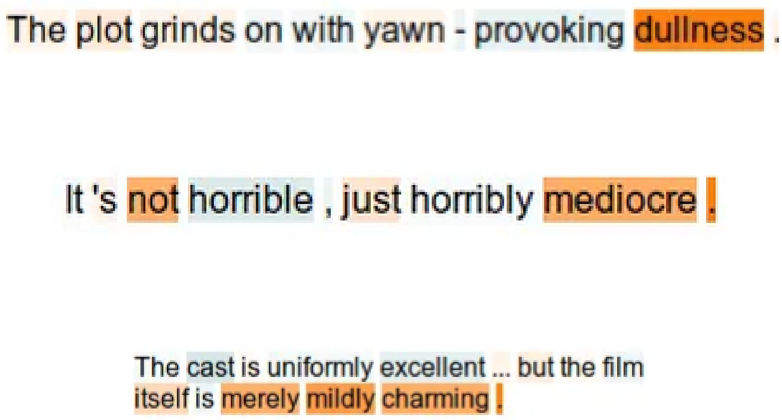
Moreover, recently neural networks based on models like BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) have enabled bigger jumps in NLU and NLG advancements. These models have significantly advanced the state-of-the-art by offering deeper contextual understanding and generating more coherent, context-aware text outputs. They have revolutionized how machines understand the nuances and complexities of human language by analyzing the context in which words appear. This has profound implications for tasks like Sentiment Analysis, where understanding context is crucial for accurately interpreting the tone and emotion behind text.

Furthermore, the use of GPT models has greatly enhanced the accuracy of Sentiment Analysis. GPT models include the attention mechanisms, which allow models to focus on specific parts of the text that are more relevant for determining sentiment, like how human attention works. For instance, in a product review, the model can learn to pay more attention to adjectives and adverbs, which often carry more sentiment weight. In practice, these advancements have enabled more nuanced and sophisticated Sentiment Analysis applications. For example, in customer service, AI systems can now better understand and respond to customer queries by detecting not just the content but also the sentiment behind the queries. In the media and entertainment industry, Sentiment Analysis is used to gauge audience reactions to movies, TV shows, and advertisements, providing valuable feedback to creators and marketers.

An emerging trend in NLP is the focus on model **interpretability** and **explainability**. As NLP models grow more complex, understanding how they make decisions becomes crucial, especially in sensitive applications like Sentiment Analysis. Efforts are being made to make these models more transparent and understandable, which is crucial for trust and reliability in real-world applications.

Explainability in Sentiment Analysis is pivotal, as it enhances the transparency and trustworthiness of AI models. This concept revolves around elucidating how and why a model arrives at a specific sentiment conclusion from a given text. As Sentiment Analysis models, particularly those based on deep neural networks, become more intricate and less transparent, ensuring their decisions are interpretable becomes essential. The need for explainability is particularly acute in domains where Sentiment Analysis has significant implications, such as in analyzing customer feedback for product improvement, monitoring social media for public sentiment, or even in clinical settings where patient sentiment can inform treatment approaches.

Techniques for enhancing explainability include feature visualization and model-agnostic methods. Feature visualization focuses on identifying key words or phrases within a text that heavily influence the model’s sentiment prediction. By highlighting these features, users can visually trace the reasoning behind the model's classification. Model-agnostic methods like LIME and SHAP offer a broader approach, providing insights into how each feature in the dataset influences the model's prediction, regardless of the model's complexity. These techniques help in demystifying the decision-making process of even the most complex models, making their outputs more comprehensible. Figure below shows some examples of Sentiment Analysis, where major focus words are highlighted with LIME explainability approach.



In real-world applications, explainability can be invaluable. For instance, in customer service, understanding why certain products receive negative feedback can guide improvements. In social media monitoring, explainable Sentiment Analysis can unravel the key drivers behind public opinions or reactions, offering businesses and policymakers insights into the populace's mood and preferences.

Co-funded by the European Union

DISCLAIMER

The European Commission support for the production of this publication does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein. Project Number 2022-1-IT02-KA220-HED-000085944

Find out more on: **www.gritproject.eu**