A Review Paper on Natural Language Processing

Pruthviraj K. L., Rachith M R, Rahul, Punya N

Third Semester Student Scholar, Department of Computer Science Engineering,
Alva's Institute of Engineering and Technology, Moodabidire, Dakshina Kannada, Karnataka, India

Abstract: *This study examines natural language processing (NLP) from the late 1940s to the present, with the goal of identifying trends that reflect concerns about various difficulties or the development of various approaches to solving these problems and constructing systems as a whole. In this we will also know about Natural Language Processing is the practice of teaching machines to understand and interpret conversational inputs from humans. NPL based on machine learning can be used to establish communication channels between humans and machines.*

Keywords: Natural Language Processing

I. INTRODUCTION

Don Walker, Jane Robinson were discussing about how they got started in NLP research at the 1987 ACL Coference. Fred Thomson told them he started in 1954, and others, including Martin Kay, started in the 1950s as well. Work in the subject has tended to focus on one problem at a time, sometimes because solving problem X depends on solving problem Y and some other times simply because problem Y appears to be more tractable than solving problem X . It's comforting to suppose that NLP research, like scientific research in general, is consolidating, and while this may be more faith than substance, we can surely perform NLP now that we couldn't do in the 1950s. The fact that we are being lured by the fact that we are being seduced by the fact that we are While advances in computing technology have made intellectual advances in understanding how NLP is performed, the improvements in technology have also eliminated some of the problems that previously plagued us. But more importantly, good technology means that when you get back to a long-standing problem, it's not always as difficult as it used to be. For those who have been around like Don for a long time, old ideas can be seen re-emerging in new guises, such as lexicalist approaches to NLP and MT in particular. But the new costumes are better made, of better materials, and more pleasing: so the research doesn't go so much in circles as on a spiral, albeit a fairly shallow one. Looking back at the history of NLP, I see four phases, each with their own concerns and styles. Don has moved in one way or another and like all of us to some degree Time to the current beat. But it is noteworthy that in the push he made for linguistic resources, like corpus collections, he not only significantly promoted what I have called the data-bashing decade that is now with us, but also returned to what was a major concern in the first period of NLP research: building the powerful and comprehensive dictionaries that serious NLP applications, like MT, need. I define the first phase of work in NLP as lasting from the late 1940s to the late 1960s, the second from the late 60s to the late 70s and the third to the late 80s, while we are in a clear fourth phase now.

II. HISTORICAL REVIEW (LATE 1940S TO LATE 1960S)

The first phase of work focuses on machine translation (MT). Following the investigations of Booth and Richens and Weaver's influential translation memo published in 1949 (Locke and Booth, 1955), NLP research officially began in the 1950s. The 1954 IBM-Georgetown Demonstration demonstrated automatic translation from Russian to English, a very rudimentary form and limited experimentation. MT (Mechanical Translation), the predecessor of Computational Linguistics, also started publishing in 1954. The first international conference on machine translation was held in 1952, the second in 1956 (the year of the first conference on artificial intelligence); at the important Washington International Conference on Scientific Information in 1958, language processing was linked to information retrieval, For example in the use of thesaurus; Minsky draws attention to artificial intelligence; Luhn provides automatic abstracts (actually excerpts) for a conference papers. The Teddingt on International Conference on Machine Translation and Applied Language Analysis in 1961 was probably the high point of this first phase. From morphology, syntax, semantics, interpretation and generation, and formal theory to hardware.

This first phase was a period of enthusiasm and optimism. Not only because the people engaged attacked very difficult NLP tasks and ran into syntactic and semantic processing issues, and language diversity issues with all their might. They were trying to use a new tool, the computer, for non-numeric data processing purposes when the data processing itself was not well established. It's important to remember how primitive the available computing resources were. This was the era of punch cards and batch processing. There were no suitable high-level languages and almost all programming was done in assembler. Access to machines was often restricted; They had very limited disk space and were extremely slow. Plath (1967) reports processing speeds of about 7 minutes for long sentence analysis, even with the most advanced algorithms and on the best machines available at the time. To save space and time, a lot of programming effort was expended on bitpacking. It is remarkable how much has been done with such meager resources, for example in building grammars and lexicons: some of the grammars and dictionaries of the early 1960s were very extensive even by today's standards. Research during this period was completely international, with significant activity in the USSR, as well as in the USA and Europe, as well as in Japan. US grant funding increased after Sputnik 1, but work began earlier. The dominant languages were Russian and English, but other languages were involved, including Chinese.

II. RESEARCH AREA IN NATURAL PROCESSING LANGUAGE

It mainly consists of following specific research areas:

1. Machine Translation.
2. Information Retrieval.
3. Text Categorizatoin.
4. Text Summarization.
5. Sentiment Analysis and Opinion Mining.

2.1 Machine Translation

Machine translation implies that it is about translating, while NLP, or Natural Language Processing, implies that it is about language. This topic is about language translation. Machine translation (MT) is the process of automatically translating one natural language into another.
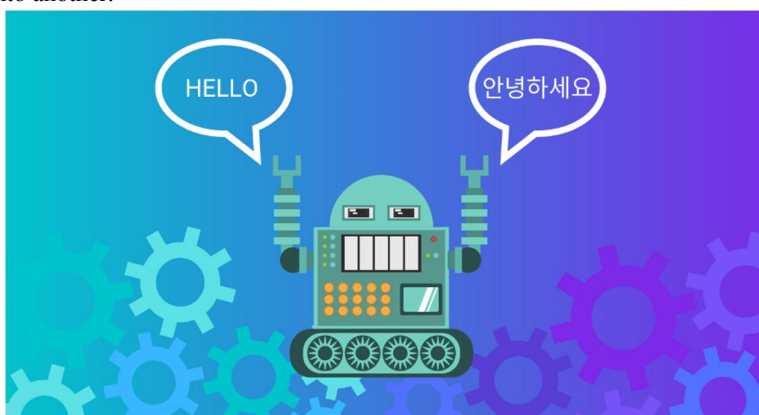

2.2 Information Retrieval

Advanced natural language processing techniques are employed to improve the effectiveness of term-based document retrieval in this information retrieval system. Our system's backbone is a typical statistical engine that creates inverted index files from pre-processed texts before searching and ranking them in response to user queries.
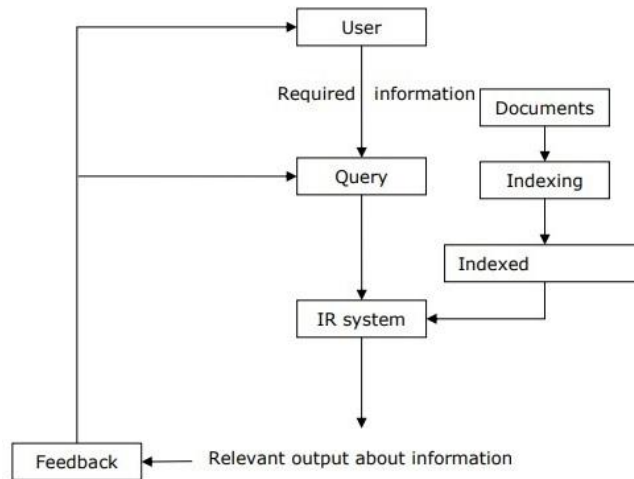
2.3 Text Categoration

The practise of categorising text into ordered groupings is known as text classification, sometimes known as text tagging or text categorization. Text classifiers use Natural Language Processing (NLP) to automatically analyse text and assign pre-defined tags or categories depending on its content. It is mostly organised in a step-by-step format to make the procedure easier.
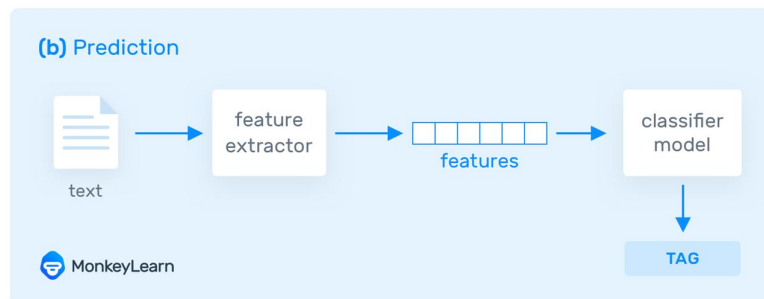

2.4 Text Summarization

Automatic Text Summarization is a technique in which a computer software shortens lengthy texts and provides summaries to convey the desired content. It is a prevalent challenge in machine learning and natural language processing (NLP). It will become shorter and have points.

2.5 Sentiment Analysis and Opinion Mining

The task of identifying, extracting, and classifying opinions about anything is known as Opinion Mining (OM) or Sentiment Analysis (SA). It is a sort of natural language processing (NLP) used to track public opinion on a certain law, policy, or marketing campaign.
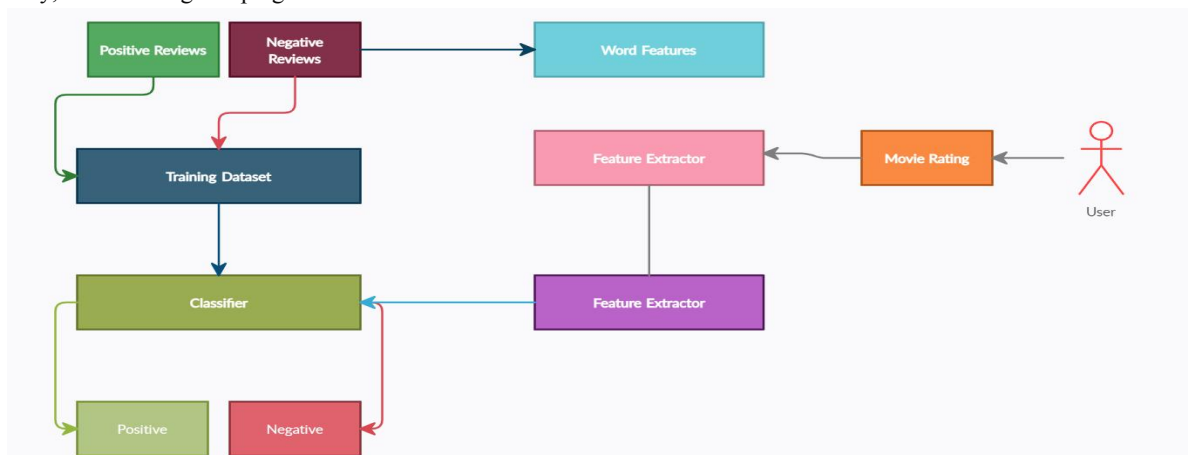
2.6 Natural Language Processing in Machine Learning

Artificial Intelligence study includes natural language processing as a subfield. It is mostly concerned with the creation of models and protocols. Natural Language Processing (NLP) is a type of AI that allows machines to understand and interpret human language rather than just read it. NLP is a branch of machine learning that involves a computer's ability to comprehend, interpret, manipulate, and possibly synthesis human language.
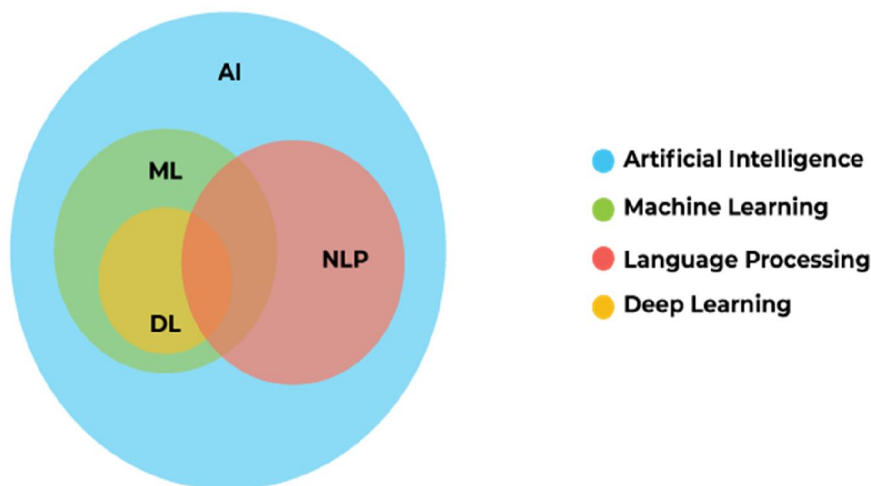
III. RELATIONSHIP BETWEEN AL, ML, AND DL

- Machine Learning (ML) is commonly used along with AI but it is a subset of AI.
- ML refers to an AI system that can self-learn based on the algorithm.
- Systems that get smarter and smarter over time without human intervention is ML.
- Deep Learning (DL) is a machine learning (ML) applied to large data sets.

3.1 Top NPL Algorithms

- Lemmatization and Stemming.
- Keyword Extraction.
- Topic Modelling.
- Knowledge Graphs.
- Named Entity Recognition.
- Words Cloud.
- Machine Translation.
- Sentiment Analysis. Sentiment analysis is the most common method used by NLP algorithms.

3.2 Advantage of NPL Include

- When properly applied, NLP is less expensive and time-consuming than hiring a person.
- NLP can also assist companies in providing quicker customer care response times. Customers get rapid replies to their questions no matter what time of day it is or what day of the week it is.
- Developers can use pre-trained machine learning models to help with a variety of NLP applications, making them simple to build.

3.3 Disadvantage of NPL Include

- Training can take a long time. It can take weeks to get a high level of performance if a new model is constructed without the help of a pre-trained model.
- Another drawback of NLP is that machine learning isn't always accurate. There is always the chance of inaccuracies in predictions and outcomes, which must be considered.

IV. CONCLUSION

- The practice of teaching machines to read and interpret human conversational inputs is known as natural language processing.
- NLP based on Machine Learning can be used to construct human-machine communication channels.
- NLP has already shown effective in a variety of disciplines, despite the fact that it is still evolving.
- Various NLP implementations can assist organizations and individuals in saving time, increasing efficiency, and improving customer satisfaction.

REFERENCES

[1]. L. R. Bahl, P. F. Brown, P. V. de Souza, and R. L. Mercer, "A tree based statistical language model for natural language speech recognition," in Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 37, Issue 7, (Yorktown Heights, NY,USA), pp. 1001–1008, July 1989.

[2]. P . Clarkson and R. Rosenfeld, "Statistical language modeling using the cmu-cambridge toolkit," in Proceedings EUROSPEECH (N. F.G. Kokkinakis and E. Dermatas, eds.), vol. 1, (Rhodes, Greece), pp. 2707–2710, September 1997.

[3]. J. Tejedor, R. Garca, M. Fernndez, F. J. LpezColino, F. Perdrix, J. A. Macas, R. M. Gil, M. Oliva, D. Moya, J. Cols, , and P. Castells, "Ontology-based retrieval of human speech," in Database and Expert Systems Applications, 2007. DEXA '07. 18th International Conference on, (Regensburg, Germany), pp. 485– 489, September 2007.

[4]. J. R. Bellegarda, "Statistical language model adaptation: Review and perspectives," vol. 42, no. 1, pp. 93–108, 2004.

[5]. Y.-Y. Wang, M. Mahajan, and X. Huang, "A unified context-free grammar and n-gram model for spoken language processing," in IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. III, (Istanbul, Turkey), pp. 1639–1642, Institute of Electrical and Electronics Engineers, Inc., 2000