ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 1, January 2020

Comprehensive Review on Advancements in Natural Language Processing

ASHALATHA P R

Senior Scale Lecturer, Government Polytechnic, K.R Pete, Karnataka, India

ABSTRACT: Natural language processing (NLP) developments have made it possible for robots to read and analyze human language with astounding precision, revolutionizing the field of text understanding. An overview of current advancements in NLP approaches and their effects on text comprehension are provided in this abstract. It examines significant developments in fields including named entity identification, sentiment analysis, semantic analysis, and question answering, highlighting the difficulties encountered and creative solutions put forth. To sum up, recent developments in natural language processing have raised the bar for text comprehension. Deep learning models and extensive pre-training have changed methods including semantic analysis, sentiment analysis, named entity identification, and question answering. These developments have produced text comprehension systems that are increasingly precise and complex. However, issues with prejudice, coreference resolution, and contextual comprehension still need to be resolved. The future of NLP for text understanding has considerable potential with continuing study and innovation, opening the door for increasingly sophisticated applications in numerous sectors.

KEYWORDS: Natural language processing (NLP), Deep learning models, analyze human language,

I. INTRODUCTION

NLP is built on semantic analysis, which aims to extract context and meaning from textual input. Recent developments have concentrated on deep learning techniques, such recurrent neural networks and transformers, which have greatly increased the accuracy of tasks like phrase parsing, word sense labeling, and semantic role labeling. These methods have made it possible for robots to comprehend the subtleties of language, resulting in more advanced text comprehension.

Sentiment analysis, a crucial component of NLP, is identifying the emotional undertone of text. Deep learning models' introduction has significantly advanced sentiment analysis. Modern performance on sentiment classification tasks has been attained using methods like convolutional neural networks and recurrent neural networks in conjunction with extensive pre-training. Sentiment-based applications, such as social media analysis, customer feedback analysis, and market trend prediction, have been made possible by this advancement.

Named Entity Recognition (NER) aims to identify and classify named entities in text, such as person names, locations, and organizations. Recent advancements in NER have leveraged deep learning models, such as Bidirectional Long Short-Term Memory networks (BiLSTMs) and Transformer-based architectures, to achieve remarkable accuracy. Finetuning these models with large-scale annotated datasets and incorporating external knowledge sources, such as pre-trained language models like BERT and GPT, have further enhanced NER performance.

Question answering (QA) systems have also witnessed remarkable progress in recent years. Traditional QA systems relied on rule-based approaches and limited question types. However, the introduction of neural network architectures, such as attention-based models and transformer-based models, has revolutionized QA systems. These models can comprehend complex questions, retrieve relevant information from large-scale document collections, and generate accurate answers. Applications like virtual assistants, information retrieval systems, and instructional platforms have profited immensely from this development.

Despite these developments, there are still problems with text interpretation. Understanding the context of utterances, particularly those that are vague or metaphorical, continues to be difficult. Research is still ongoing in the field of

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 1, January 2020

coreference and anaphora resolution tasks, where pronouns relate back to earlier entities. Furthermore, resolving bias and fairness problems in NLP models and datasets remains a major topic.

The capacity to comprehend and analyze enormous volumes of textual material has become more crucial in the age of the information explosion. A subfield of artificial intelligence called "Natural Language Processing" (NLP) aims to make it possible for computers to understand and utilise human language. NLP has made great strides throughout time, altering how we interact with and get insights from textual data. These developments have opened the path for improved text comprehension, allowing machines to comprehend the subtleties of spoken language and draw out important information.

II. LITERATURE REVIEW

Jiajun Zhang (2019) Natural language processing is a field of artificial intelligence and aims at designing computer algorithms to understand and process natural language as humans do. It becomes a necessity in the Internet age and big data era. From fundamental research to sophisticated applications, natural language processing includes many tasks, such as lexical analysis, syntactic and semantic parsing, discourse analysis, text classification, sentiment analysis, summarization, machine translation and question answering. In a long time, statistical models such as Naive Bayes (McCallum and Nigam et al., A comparison of event models for Naive Bayes text classification.

Hang Li (2018) Deep learning refers to machine learning technologies for learning and utilizing 'deep' artificial neural networks, such as deep neural networks (DNN), convolutional neural networks (CNN) and recurrent neural networks (RNN). Recently, deep learning has been successfully applied to natural language processing and significant progress has been made. This paper summarizes the recent advancement of deep learning for natural language processing and discusses its advantages and challenges.

Veronika Vincze, Richard Farkas (2014) Natural language processing (NLP) systems usually require a huge amount of textual data but the publication of such datasets is often hindered by privacy and data protection issues. Here, we discuss the questions of de-identification related to three NLP areas, namely, clinical NLP, NLP for social media and information extraction from resumes. We also illustrate how de-identification is related to named entity recognition and we argue that de-identification tools can be successfully built on named entity recognizers.

Laura Chiticariu et al (2013) The rise of "Big Data" analytics over unstructured text has led to renewed interest in information extraction (IE). We surveyed the landscape of IE technologies and identified a major disconnect between industry and academia: while rule-based IE dominates the commercial world, it is widely regarded as dead-end technology by the academia. We believe the disconnect stems from the way in which the two communities measure the benefits and costs of IE, as well as academia's perception that rulebased IE is devoid of research challenges. We make a case for the importance of rule-based IE to industry practitioners. We then lay out a research agenda in advancing the state-of-theart in rule-based IE systems which we believe has the potential to bridge the gap between academic research and industry practice.

Deep Learning Architectures for NLP

In the field of ordinary language dealing with (NLP), wide learning has become logically seen as giving supportive gadgets and strategies to supervising various language-related works out. This part reviews huge learning methodologies that have changed normal language taking care of (NLP, for example, change models, frontal cortex excess affiliations (RNN), and convolutional mental affiliations (CNN). Convolutional Associations (CNN) are renowned in standard language dealing with (NLP) in view of their ability to isolate social models and characteristics from printed information. CNNs have been changed by NLP by considering words or characters as a part of room and late undertakings have been made for PC. CNNs can segregate critical and specialist features from the data using convolutional channels, making them ideal for applications like visual assessment, message depiction, and analysis. Abundance Cerebrum Affiliations (RNNs) are another vital preparation in NLP, unquestionable for their ability to ponder relentless data and catch typical circumstances.

Pre-training Techniques

Setting up strategies recognize a fundamental part in working on the show and hypothesis limits of tremendous learning models for standard language making do (NLP) attempts. These strategies want to utilize a ton of unlabeled message

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 1, January 2020

data to present model limits effectively, engaging the model to get rich semantic depictions of words and clarifications. One of the most broadly used pre-getting ready frameworks is word embeddings, which address words as thick vector depictions pondering their proper use in gigantic text corpora. Models like Word2Vec, GloVe, and 'FastText' have shown the ampleness of word embeddings in getting semantic essentially indistinguishable qualities and connection between words, working with endeavors like semantic equivalence assessment and word relationship finish. Another conspicuous pre-preparing procedure is the utilization of language models, which are prepared to foresee the following word in a succession given the former setting. Language models, like the OpenAI GPT (Generative Pre-prepared Transformer) series and BERT (Bidirectional Encoder Portrayals from Transformers), have accomplished amazing execution across different NLP assignments by utilizing enormous scope pre-preparing on assorted text corpora.

Applications of Deep Learning in NLP

Profound learning methods have changed regular language handling (NLP) by empowering many uses across different spaces. One conspicuous application is feeling examination, which includes deciding the opinion or assessment communicated in a piece of text. Profound learning models, for example, convolutional brain organizations (CNNs) and repetitive brain organizations (RNNs), have exhibited exceptional execution in feeling examination errands by figuring out how to catch unpretentious context-oriented signs and semantic examples demonstrative of opinion extremity. These models are utilized in feeling examination applications across businesses, including web-based entertainment checking, client criticism examination, and brand notoriety the executives. Machine interpretation is another key application region where profound learning has taken huge steps in working on the exactness and familiarity of robotized interpretation frameworks. Transformer models, like Google's Transformer and OpenAI's GPT (Generative Pre-prepared Transformer), have upset machine interpretation by utilizing self-consideration systems to catch worldwide conditions inside input groupings.

Multimodal NLP

Multimodal natural language processing (NLP) includes pictures, videos, and audio that alter human language perception and processing. Due to the development of online multimedia content and smart devices with various sensors, NLP models that interpret and generate content across modalities are needed. Recently developed multimodal natural language processing systems interpret and produce content across modalities. These systems describe multimodal data using RNNs, CNNs, and Transformer models. Multimodal natural language processing models can better interpret multimodal content by capturing complex semantic links between textual and non-textual data utilising fusion and cross-modal attention.

Generation and Understanding of Natural Language

The process of generating natural language involves translating the information into easily readable language of human beings from the computerized databases. Natural language understanding involves change of text sections into much formal notations like the structure related to logistics in the first order that can be easily manipulated by the programs. Semantics recognition using feasible semantics that are obtained using natural language expressions that are normally disguised as organized notations present in natural language concepts are dealt by it.

III. METHODOLOGY

Data extraction was done to gather information on the uses of natural language processing (NLP) in healthcare after the pertinent literature was found. Two impartial reviewers extracted the data, and any discrepancies were settled through discussion and agreement. The title, authors, publication year, study design, NLP application, data source, and key conclusions were all taken out of each study. The data was then combined and examined to find recurring themes and trends in research. The study was undertaken to utilize a qualitative technique, where themes were uncovered through a process of inductive reasoning, allowing for the emergence of new insights from the data. The same two independent reviewers carried out the synthesis and analysis, and any differences were settled through discussion and consensus. The findings and discussion parts of this survey were created using the findings of the data synthesis and analysis. The process of data synthesis and analysis made it possible to find similar themes and patterns among studies, giving insights into the most cutting-edge NLP applications being used in healthcare today.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 1, January 2020

IV. RESULTS

Frequency of NLP Algorithms

A number of NLP techniques were applied prior to the use of ML algorithms, and these NLP techniques are described in Appendix A. NLP techniques are associated with various algorithms, which are defined in Appendix B. After registering the data in the SPSS v25 statistical software, it has been discovered that the Term frequency - inverse document frequency (TF-IDF) algorithm was present in 22 articles, 47.82%. The Word2Vec algorithm was used 15 times, 32.60%. Glove algorithm at 14, 30.43%, while Bag of words (BoW) was used in 11 studies, 23.91%, and N-Gram was used in six articles, 13.04%. On the other hand, seven studies, 15.21%, used three algorithms at the same time. Likewise, 12 articles, 26.09%, used two algorithms at the same time, while 26 articles used a single algorithm, 58.69%, in their research. The details of the NLP algorithms used are detailed in Appendix C.

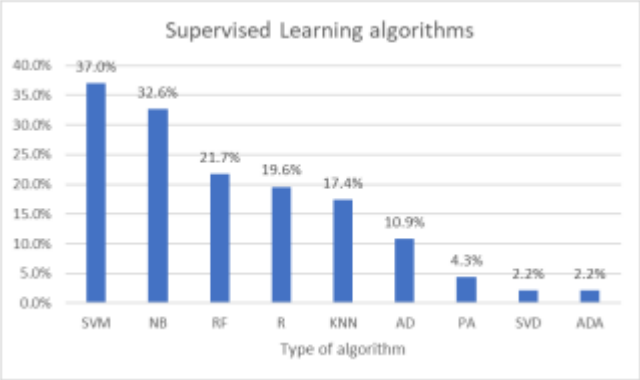

Fig. 1. Studies using S Algorithms.

Frequency of ML Algorithms

ML algorithms analyzed in this study are defined in Appendix D and grouped under supervised learning, unsupervised learning, and Deep Learning.

1) Frequency of supervised learning algorithms (S):

The Support Vector Machine (SVM) algorithm was used in 17 studies. The Naive Bayes (NB) algorithm was applied in 15 studies. The Radom Forest (RF) algorithm has 10 studies. R has 9 studies. K-NN has 8 studies. RF has 5 studies. The Passive aggressive (PA) algorithm was used in two studies. The AdaBoost (ADA) algorithm has 1 study like Singular Value Decomposition (SVD) algorithm, Fig. 1 shows the details.

2) Frequency of unsupervised learning algorithms (US):

The Latent Dirichlet Allocation (LDA) algorithm was used in three studies, 6.5%, while K-Means algorithm was used only in one study, 2.2%.

3) Frequency of deep learning algorithms (DL):

The Long Short Term Memory (LSTM) deep learning algorithm has 24 studies. Then, the Convolutional neural networks (CNN) algorithm has 12 studies. The Recurrent Neural Networks (RNN) algorithm has 9 studies. The Multilayer Perceptron (MLP) algorithm has 5 studies. The least used algorithms were Gating Circulation Unit (GRU) algorithm with four studies and Artificial neural network (ANN) with two studies. Fig. 2 shows the details.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 1, January 2020

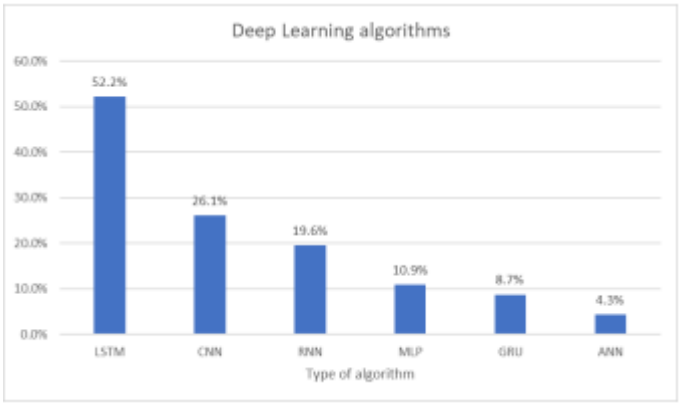

Fig. 2. Studies using DL Algorithms.

4) Studies with hybrid algorithms, Supervised (S),
Unsupervised (US) and Deep Learning (DL): Out of 46 articles, 10 (21.74%) use only S algorithms. In this regard, and use SVM. The author in use NB algorithm. On the other hand, the study of [51] uses the LDA algorithm. Three studies use S and US algorithms at the same time: use two S algorithms: SVM, NB and 1 NS: K-Means. use five S algorithms: SVM, NB, Regression (R), RF, KNN, and one US algorithm: LDA. The author in uses three algorithms: A supervised learning algorithm, the LDA, and two deep learning algorithms, GRU and CNN. Twenty-one studies, 45.65%, used only DL algorithms, in particular y. Eleven studies, 23.91%, used S and DL algorithms at the same time, in particular y. The details of the ML algorithms used by the 46 studies can be found in Appendix C.

V. CONCLUSION

Advancements in Natural Language Processing have propelled text understanding to new heights. Pretrained language models, contextual word embeddings, attention mechanisms, and multimodal learning techniques have revolutionized the field, enabling machines to comprehend human language more effectively. These developments have not only improved existing applications but have also opened doors to new possibilities in information retrieval, sentiment analysis, question-answering systems, and machine translation. However, challenges remain. Ethical considerations, bias mitigation, and responsible deployment of NLP technologies are critical areas that demand ongoing research and development. As NLP continues to evolve, it is crucial to strike a balance between technological advancements and societal impact, ensuring that text understanding technologies are transparent, fair, and accountable.

REFERENCES

1. Grégoire Mesnil, Yann Dauphin et al. “Using Recurrent Neural Networks for Slot Filling in Spoken Language Understanding” IEEE Press Piscataway, NJ, USA Volume 23 Issue 3, March 2015.
2. Gy. Szarvas, R. Farkas, and R. Busa-Fekete, “State-of-theart anonymisation of medical records using an iterative machine learning framework,” J Am Med Inform Assoc, vol. 14, no. 5, pp. 574–580, 2007. [Online]. Available: http://www.jamia.org/cgi/content/abstract/M2441v1.
3. Hang Li (2018) Deep learning for natural language processing: advantages and challenges, National Science Review, Volume 5, Issue 1, January, Pages 24–26, https://doi.org/10.1093/nsr/nwx110.
4. Jason Weston[2011], “Natural Language Processing (Almost) from Scratch”, Journal of Machine Learning Research , ISSN 1532-4435,Vol. 12 , Page 2493-2537.
5. Jiajun Zhang (2019) Deep Learning for Natural Language Processing, Deep Learning: Fundamentals, Theory and Applications (pp.111-138), DOI:10.1007/978-3-030-06073-2_5

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 1, January 2020

6. Laura Chiticariu et al (2013) Rule-based Information Extraction is Dead! Long Live Rule-based Information Extraction Systems, Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 827–832.
7. Matthew N. O. Sadiku, Yu Zhou, and Sarhan M. Musa, Natural Language Processing, International Journal of Advances in Scientific Research and Engineering (IJASRE), Volume 4, Issue 5, May – 2018.
8. Mausam, Michael Schmitz, Stephen Soderland, Robert Bart, and Oren Etzioni (2012) Open Language Learning for Information Extraction. In EMNLP-CoNLL, pages 523–534.
9. Rajkumar Singh Rathore[2016], "A Review on Natural Language Processing", International Journal of Engineering Development and Research, ISSN: 2321-9939, Volume 4, Issue 3
10. Vincze V, Farkas R. (2014) De-identification in natural language processing. In:2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE; p.1300–1303.