WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Applications of Natural Language Processing

Ujjawal Agarwal*[*1], Aman Raj Jain*[*2], Mahesh Tiwari*[*1]

[*1,2]*Student, Department Of Computer Science, NPGC, Lucknow, UP, India.*
[*3]*Assistant Professor, Department Of Computer Science, NPGC, UP, India.*

A B S T R A C T

Natural Language Processing (NLP) is a rapidly developing approach that is utilised to create many sorts of Artificial Intelligence (AI) that we see today, and it will continue to be a top focus for current and future work on cognitive applications. We discussed some of the most useful applications of NLP in this paper. Our goal is to develop a theoretical analysis of numerous domains where NLP can play a significant role and, through its automation approaches, alter the entire scenario. It's a popular topic that's pulling everyone into investing in it. A sophisticated and extensive understanding of NLP and its field is used to complete these applications. This article begins with a description of natural language processing (NLP) trends and components, then moves on to the application and evolution of NLP, as well as its issues. Text to Speech generator, Liver cancer prediction, IOT-enabled smart home, Word Extraction, and Sentence Generator are among the features we've added. NLP has changed the way we interact with computers and will continue to do so. AI will be the underpinning force for the change from data-driven to intelligence-driven efforts as AI technologies adapt and enhance communication technology in the years ahead.

Keywords: Natural Language Processing, Artificial intelligence, Automation, Internet of Things

1. INTRODUCTION

Natural Language Processing (NLP) is a research and application field that examines how computers can comprehend and use natural language teaching or speech in order to complete meaningful tasks. Common language processing is a hypothetically defined set of computer approaches that address at least one level of writing of regularly occurring semantic probes in order to complete the handling of human-like language for assignment or application classification. The goal of NLP is to do certain tasks. Program outlines, talk tests, machine interpretation, and other bulk assignments are used in NLP. NLP enables computers to communicate with people in their own language and to assess other language-related initiatives.

With no limitations and a calm, unbiased mind, current PCs can notice more language-based information than individuals. Computerization will be necessary to separate text and discourse information as a whole product, given the large amount of unstructured data that is transferred every day from healthcare records to internet media. Human language is distinct from that of surprise. We communicate in a variety of ways, both verbally and in writing. Despite the lack of accents, each language has a unique punctuation system and linguistic structural laws, phrases, and tongues. We frequently speak or collect words incorrectly or eliminate utterances when we compose. We have regional accents and mumble, stumble, and learn words in colloquial idioms when we talk. Currently, administered and unplanned learning, particularly intensive learning, are extensively employed to demonstrate human language, as well as a necessity for syntactic and semantic understanding and spatial mastery, which are essentially the same thing. These artificial intelligences do not exist. NLP is vital because it helps with language ambiguity correction and, in some cases, adds considerable numerical design to data, akin to downstream applications, discourse acknowledgments, or text checks.

2. APPLICATIONS OF NLP

2.1. Sentence Composition

When language constraints are combined with the model's outputs, meaningful and cohesive phrases can be produced. Conversation systems, translation software, and visual captioning are only some of the applications. When an auto-regressive model creates phrases from left to right using Beam search, it's

* *Corresponding author.* Tel.:916307455595
E-mail address: agarwalujjwal90@gmail.com

tough to instill lexical limitations. This is a difficult circumstance that BFGAN, a new algorithmic framework, will tackle. This system employs a reverse producer and a forward producer to build coherent sentences based on lexical constraints, and a discriminator to integrate the backward and forward phrases using signal rewards. Aside from BFGAN, a variety of training tactics ensure that the training process is more reliable and productive. BFGAN will improve over time, and other flaws will be resolved. When a lexicon constraint is a set of terms that are willing to emerge in the conclusion, it is called lexically restricted sentence generation. This is now the trendiest topic in the field of natural language processing.

RNN has lately emerged as a major player in NLP, with notable outcomes in tasks such as neural networks, tech blog development, textual summarization, table-to-text creation, and effective text output. Auto-regressive models are used to produce left to right sentences using Beam Search (BS). Generating sentences with lexical constraints is a tough procedure. If we replace the arbitrary word in the output, the sentence's fluency will suffer. If further information about the term is provided, there is no guarantee that the desired phrase will appear in the outputs. In (B/F-LM), lexically limited phrases are constructed using reverse and forward linguistic models that work together. The reverse linguistic model generates the first half of the phrase, with a lexical constraint as the input. The forward linguistic model then uses the first half-sentence to construct the whole sentence. Maximum likelihood estimation (MLE) objectives are used to train these two models. The outputs were inconsistent and incoherent when the reverse language model developed the first half and the forward language model created the rear half. This problem emerged because both language models were trained independently; they should be trained simultaneously and have access to each other's output. A unique approach termed reverse-Forward Generative Adversarial Network (BFGAN) was created to overcome the challenges of lexically constrained languages. The three components are the reverse producer, forward producer, and classifier. The two generators cooperate to deceive the discriminator. For coherence, the forward generator is given the dynamic attention mechanism. From beginning to end, there is a recursion, and the attention function's scope expands. During inference, the forward producer can focus on the reverse producer's first split. BFGAN solves problems and enhances the performance of lexically constrained languages. Algorithms are used to make the model easier to use while still providing stability. Work on numerous lexical constraints will be a part of future projects.

2.2. Word Extraction

The field of NLP is expanding its horizons as it grows quicker. Because of NLP, word extraction is also possible. From the given words, we can develop and gather associated words. This took the place of manual word gathering. This application employs a method for autonomously extracting linked words. It has two stages: extraction of association words and creation of a machine association network. The reading comprehension algorithm is mostly used for automatic word extraction. Being associated with someone or something, and receiving a cue response, involves receiving a signal message in response to some activity or behaviour. This method of word association exercises is both time-consuming and costly. Various experiments are carried out in quite different environmental and experimental circumstances than in real life. Because the cue response word association is ineffective and inaccurate, research is still ongoing in the NLP field. NLP associative words are discovered using the (RC) Reading Comprehension method. A machine association network is one in which words are generated by a machine, whereas a human association network is one in which words are humanly collected. Words are extracted using the attention model. The attention mechanism is utilised to figure out how cue-response words are related. Associated terms are extracted from plain texts, and a network is constructed. NLP is capable of linguistically and semantically identifying, interpreting, and evaluating the link between words. The reading comprehension algorithm (RC) maps the cue word to the response word . To some extent, the human connection pattern is stable. As a result, we anticipate that the machine association network will be accurate to some extent even if the attention entity remains constant. The attention mechanism is recognised for producing reasonably permanent cognitive patterns, regardless of the attention algorithm used. This leads us to believe that the task is completed using a neural network-based framework that accumulates related phrases from text format, as well as a Reading Comprehension algorithm and an attention mechanism.

2.3. NLP pipeline for liver cancer prediction

Despite the rise of the NLP profession, EMR processing jobs have remained uncommon and difficult to come by due to a lack of information, datasets, and semantical elements, particularly in radiology reports. The NLP pipeline is being utilised to extract features from laboratory reports in this study. Cancer risk can be anticipated as well. The random forest has the best performance and accuracy for predicting liver cancer. Clinical tests and other predictive tasks can potentially benefit from the NLP pipeline. Electronic medical records, or EMRs, are valuable assets. Radiology reports and cancer predictions are the emphasis of this application. The feature extraction NLP process could be used in a variety of clinical setting and disease speculation activities. Electronic medical reports are important in the field of research for better care quality and coordination. In today's digital era, machine learning-based techniques play a crucial role in data analysis and are useful in domains such as medical decision-making, disease treatment, and administration. Medical photographs and lab data are the primary means of communication between radiographers who scan photos and physicians who write the final control study. NLP employs mathematical and linguistic techniques to extract data from unstructured data, which is subsequently transformed into massive datasets. The benefits of feature extraction based on natural language processing are numerous. As a result, diagnostic monitoring, cohort construction, performance evaluation, and clinical aid in radiology have all benefited from NLP-based feature extraction. In a clinical setting, NLP is utilised to aid the extraction of characteristics, entities, and relations from clinical texts, yet free text is found to be more relevant and natural. Named entity recognition is a series of labelling activities (NER). It's a time-consuming operation to retrieve data from EMRs. In order to improve NER's effectiveness, deep neural networks were recently introduced. Tasks such as drug-related investigations, disease studies, and feature extraction are implemented in these tasks. The feature selecting tool Lasso is added. Furthermore, these test reports employ lasso feature selection and logistic regression with a binomial distribution. Both the patient and the country's regime are affected by liver cancer. When the scenario and health conditions get out of hand, many people are diagnosed with fatal liver cancer. As a result, early detection of liver cancer should be possible. Because there is a lack of corpus and datasets for EMR processing of radiology reports, we must manually create the corpus for our study. The radiological experts collect a lexicon, which is then annotated based on the linguistic limitations and clinical experience of the experts. This vocabulary considers words that are clinically relevant. The pipeline could be used as a reference in similar Electronic medical applications because we just used a small percentage of radiological data to develop the lexicon. The pipeline can extract radiological features, and the diagnosis model can provide physicians with guidance. The experiment is being carried out to see how a neural network-based method can assist radiologists in making diagnosis judgments, particularly those with less expertise. The diagnosing scheme's capability has vastly increased. The pipeline is also demonstrating high-performance accuracy in the diagnosis of liver cancer. This article describes an NLP pipeline for detecting liver cancer. A deep learning lexicon model is included in this work to improve the performance of the NER BiLSTM-CRF; the model provides accurate results with both NER and liver cancer prediction. The proposed NLP method could be applied to the construction of lexicons for more diseases and medical texts.

2.4. Interactive text 2 speech

Language is one of the most efficient means of communication, but it can also be a source of ambiguity and misinterpretation. A person may make ambiguous statements that can be interpreted in a number of different ways]. Despite the ambiguity in human commands and limits, Interactive Text to Pickup can complete the work without linguistic interpretation or object recognition. Object Recognition is a computer vision method that recognises and determines objects in images or videos, and can be used to count the objects and trace their precise location. Linguistic interpretation is the process of analysing a string of symbols, whether in natural language, code generators, or database systems, using the rules of a formal grammar, syntax analysis, or syntactical analysis. By developing an end-to-end Text to pickup network, we eliminate the need for pre-processing, as well as language interpretation and object recognition. The prepared Text to pickup network can manage the task flexibly when an input instruction is complicated or multiple objects are placed in different ways. The remedy to ambiguity is interaction. It is a system that comprises of a question generator and a text 2 pickup network. In addition, the Text to Pickup network generates a position heat map. A position heat map is a two-dimensional data visualisation approach that displays the magnitude of a phenomenon as colour. It's a two-dimensional distribution that includes a heat map of uncertainties as well as values showing certainty in the target object's position. The user's initial language instructions will be vague, thus the question generating network will decide which question to ask. Determining the question to be asked is a difficult task because the procedure is then continued based on this. The inquiry must be pertinent, to-the-point, and well-defined, and it must not duplicate the previously provided information. The answer to the additional question is then attached to the original command and sent back to the Text to pickup network; in other words, it's a visual question answering system (VQA). It's important to note that creating a query to reduce the ambiguity of an image-related language set of instructions and obtaining a response to a specific image-related inquiry are two different tasks.] This network uses the Baxter robot to demonstrate how it works, and a dataset is produced based on the images obtained by the Baxter robot for training and testing. The camera is mounted on the robot's arm. Each image is made up of three to six blocks of delicate colour. The goal of repairing the camera was to reacquaint the robot with the actual world.

For example, the network might ask "This one," and if the answer is yes, the simulator will say yes; if the answer is no, the experiment will be judged a failure. If the command is "pick up the red one," the network responds with "red one?" to make the experiment more rigorous and realistic. In response to a human language command, we propose the Interactive Text to Pickup (IT2P) network for retrieving the desired object. When a command is delivered in an unclear language, the IT2P network engages with a user interface to clarify the ambiguity. By learning the specified language command, the recommended network can accurately anticipate the position of the targeted item as well as the uncertainty associated with the expected target location. By providing a question that is relevant to the situation, the proposed IT2P network has been shown to be capable of efficiently connecting with people. A Baxter robot, as well as collaboration between a real robot and a person, have been used to test the proposed network. We believe that by asking questions based on assessing uncertainty, the proposed system may successfully communicate with humans, allowing for more straightforward human-robot collaboration.

2.5. Smart Home implementation

In this current era, the goal of the modern Internet of Things (IOT) is to provide services based on fully trained systems that are designed with the user's convenience and accuracy in mind. Various studies into smart settings, such as smart houses, grids, and industrial IOT environments, are also underway. Smart houses that are built on IOT technologies have IOT characteristics that go beyond standard network technology. IOT-based smart home systems, for example, require a variety of IOT technologies to build an expert system in the surrounding area that requires continuous management, or to support device and monitoring services and create environments that are suitable for the user through communications between devices to enhance domestic lifestyle services. In order to provide services that are tailored for the user, real-time sensor data is used as an integral component of an IOT ecosystem. The number of sensors required to collect data in an IoT environment, as well as the processing mechanism for the data collected via a sensor network. When we talk about a traditional home system, the user modifies the environment manually by managing devices linked to the same network via a mobile device, which results in data waste and power loss when tasks are conducted without regard for the user's individual features. Modern IoT-based home systems, on the other hand, employ a technology that automatically controls devices depending on threshold levels set by the user. Multiple MJoin operators are used to efficiently processing sensor data (stream data) in an IOT setting throughout this research. The Support Vector Machine classification algorithm, which was used to identify and minimise the data to enable structured storage management, was also employed as a global shared query strategy for the optimised query. Sensing technology monitors changes in the environment; interface technology executes or links bound options through people, objects, and services; and network infrastructure technology provides networks between sensors and services are the three technologies that make up IOT environments. These applied sciences offer a variety of contexts, including remote control that adapts to the user's needs and automatic controls that recognise persons and deliver personalised services. In the case of an IoT environment, join queries are required to process extensive data collected from not just one, but numerous sensors. Operators based on hash tables, windows, and both hash tables and windows are examples of join operators. Because the traditional home system has difficulties in terms of single-task execution that result in various losses, it is necessary to conduct numerous jobs. The IoT environment is configured with additional process data and job priorities, ensuring that high-priority tasks are processed first. The data is reduced and classified using the SVM classification technique, and then saved to TinyDB. The server separates real-time and non-real-time data. The user is given real-time tasks, which are then saved in the MainDB. Non-real-time data is saved in the MainDB and made available to the administrator upon request. The processor board is an Arduino (Uno) module, with five sensors acquiring streams (temperature, humidity, gas, vibration, and recognition) for sensor data processing. All of these data come from the same source. As a result, they are merged into a single packet and sent. Each packet adds to the amount of traffic and energy used. As a result, a query is processed in a single packet. The system requires data generated by the sensors and user task commands to check the frequency and timing of the resultant data in order to govern the smooth sensor flow. Based on an examination of utilisation rates by month, week, and day, sensors with low usage rates are switched to standby or dropped. Users establish the sensor threshold value range using applications, and a device can be operated by checking sensor data. When completing tasks, the database range values are loaded and referred to. Temperature, humidity, gas, recognition, and vibration sensors are among the sensors in the home. Temperature sensors were designed to work in tandem with gas sensors to address potentially dangerous circumstances. Tasks completed with several sensors are also digitized and kept in the database. The system provides a convenient IOT environment by completing activities within which the events connected with the tasks are appropriate for items or the environment, based on sensor data. The user examines the data in the application to regulate the sensors based on the sensor data range selected by the user and the database data. Low-usage sensors can be put into standby mode, and the status of a sensor can be altered to make it operative again. Furthermore, task priorities are defined such that the number of concurrent processes can be lowered by suspending existing tasks when a higher-priority activity happens, reducing unnecessary and wasteful electricity consumption while offering intelligent services to the user.

3. CONCLUSION

Natural Language Processing is currently one of the most popular subjects in computer science. Companies are putting a lot of money into research in this subject. To pursue a profession in this field, everyone is seeking to grasp Natural Language Processing and its applications. It must be included into the operations of any company in some form. We read about a few NLP applications in this paper, but there are many more on the list.

REFERENCES

[1] BFGAN: Backward and Forward Generative Adversarial Networks for Lexically Constrained Sentence Generation Dayiheng Liu, Jie Fu, Qian Qu, and Jiancheng Lv

[2] Y. Wu et al., “Google’s neural machine translation system: Bridging the gap between human and machine translation,” 2016, arXiv preprint arXiv:1609.08144.

[3] L. Dong, S. Huang, F.Wei, M. Lapata, M. Zhou, and K. Xu, “Learning to generate product reviews from attributes,” in Proc. Annu. Meeting Assoc. Comput. Linguistics, 2017, pp. 623– 632.

[4] A. See, P. J. Liu, and C. D. Manning, “Get to the point: Summarization with pointer-generator networks,” in Proc. Annu. Meeting Assoc. Comput. Linguistics, 2017, pp. 1073–1083.

[5] T. Liu, K. Wang, L. Sha, B. Chang, and Z. Sui, “Table-to-text generation by structure-aware seq2seq learning,” in Proc. AAAI Conf. Artif. Intell., 2018, pp. 4881–4888.

[6] S. Ghosh, M. Chollet, E. Laksana, L.-P.Morency, and S. Scherer, “Affect- LM: A neural language model for customizable affective text generation,” in Proc. Annu. Meeting Assoc. Comput. Linguistics, 2017, pp. 634–642.

[7] F. J. Och and H. Ney, “The alignment template approach to statistical machine translation,” Comput. Linguistics, vol. 30. pp. 417–449, 2004.

[8] T.-H. Wen et al., “Stochastic language generation in dialogue using recurrent neural networks with convolutional sentence reranking,” in Proc. Annu. Meeting Special InterestGroup Discourse Dialogue, 2015, pp. 275– 284.

[9] J. Yin, X. Jiang, Z. Lu, L. Shang, H. Li, and X. Li, “Neural generative question answering,” in Proc. Int. Joint Conf. Artif. Intell., 2016, pp. 2972– 2978.

[10] A Natural Language Processing Framework for Assessing Hospital Readmissions for Patients With COPD Ankur Agarwal, Christopher Baechle, Ravi Behara, and Xingquan Zhu

[11] D. Ferrucci and A. Lally, “UIMA: An architectural approach to unstructured information processing in the corporate research environment,” Nat. Lang. Eng., vol. 10, no. 3/4, pp. 327–348, 2004.