Analyzing NLP Applications in a Real-World Case Study

Lauryn Davis
Dr. Sorio Boit
CIS 660-01
6 April 2024

I. Introduction to Articles Discussing NLP Applications

Within CIS-660, students are given the opportunity to learn about various data engineering topics, one of which being natural language processor (NLP) applications. This assignment aims to bridge the gap between the technical properties of NLP and how this can be used in real-world situations of concern. For this assignment, students were asked to use Google Scholar to obtain two recently published academic articles pertaining to NLP techniques.

The first article that was found relating to this topic was titled, "Natural Application Processes in Business", by Mohammed Bahja. The primary focus within this article was said to provide an overview of the beneficial applications in industrial sectors such as business, education, commerce, health-care, etc. It is said that with the growing access to large numbers of unstructured data, it's imperative for computer and data scientists to invent generative artificial intelligent algorithms that can handle human language and text. It is further discussed that linguistic semantics must be adequately incorporated into the conditional functions of various machine learning algorithms. A way that this is handled, is called "computational linguistics". This focuses on the statistical properties of speech and text. The article mentions, however, that many methods can utilize deep and machine learning to address complex situations regarding speech recognition. The methods of NLP that were discussed within this paper include text embedding, modeling and parsing, machine translation, neural network applications, and rule-based algorithms. With this in mind, the paper transitions into explaining how some of these methods can be utilized within business. The key findings from this were that commerce can benefit from sentiment analysis in understanding various key-word trends, educational institutions can utilize translation machinery to communicate in various languages, and researchers can find meaningful data and/or ideas from large bodies of work using text extraction NLP methodology [1].

In addition to the first article mentioned, another scholarly article that was found relating to NLP was titled, "Natural language processing in medicine: A review", by a group of researchers, one of which named Saskia Locke. The paper's primary focus is to explain the immense benefits that can be gained from using NLP applications within the healthcare field. For example, NLP methodology can aid in patient outcome predictions, hospital triage assignment systems, generative diagnostic models, and patient "Chat-box" implementation. NLP methods such as natural language generation/understanding, text extraction, tokenization, lemmatization, and word sense disambiguation can aid in the large volume of free text elements found within medical records and research papers. The paper goes on to discuss some key strengths of NLP applications. One strength is that linguistic analysis can aid in creating more user-friendly medical applications. Additionally, NLP's ability to discern subtleties within recorded speech can help people with cognitive and/or mental health issues such as depression or Alzheimer's. [2].

II. Case Study Analysis

A. Case Study Introduction and Text Data Explanation

The article that was chosen to provide an in-depth case-study analysis was the second article mentioned in Section I. There were so many noteworthy applications of NLP within a medical regard, that it was difficult to choose just one to focus on. The primary case that was decided to be focused upon is natural language processing of medical/electronic health records. It is said that this entity is difficult to interpret because of the large amounts of unstructured free-text embedded into these important records of patients' health. It is crucial that one is obtaining accurate results about diagnostic information, drugs that were administered, and reaction types [2].

B. NLP Techniques Utilized

Within this case study, many NLP techniques were mentioned. In addition to this, there was a comprehensive explanation as to what pre-processing tasks were utilized within this case study. Firstly, pre-processing NLP methods that were mentioned included tokenization, lemmitization, and word sense disambiguation. According to Data Engineering professor, Dr. Sorio Boit, tokenization is defined as, "The process of breaking text into individual words or tokens". Furthermore, lemmitization reduces words to their "dictionary form" [3]. After pre-processing, the two primary NLP properties that were mentioned included natural language understanding (NLU) and generation (NLG). NLU analyzes the linguistic properties of morphology, syntax/semantics, and phonology. NLG produces natural language text by selecting important content and organizing information. These two domains work

together to complete text summarization, machine translation, semantic analysis, and text categorization [2].

As discussed by Dr. Sorio Boit, text summarization aims to provide a concise overview of the comprehensive nature of text. Additionally, semantic and sentiment analysis aim to study the meaning of text, one example being the emotion within the piece. Furthermore, text categorization applies a label to text documents [3].

These four NLP techniques were implemented/intertwined within the medical record case-study by training a model with these ideals and producing a loss and accuracy score. It is mentioned that electronic health records depend on manual codes that impact doctors and clinicians decisions about a patient's care. This is said to be somewhat inefficient and dangerous if a code was inputted incorrectly. NLP, however, is said to improve this issue by creating a model that automates the process of summarizing unstructured text and potentially reduces the risk of human code interpretation error. The paper states that NLP techniques, such as those mentioned, transform the data and present it in a structured way. Additionally, in order to train an NLP model, outcomes must be defined in order to validate them. For example, a researcher within the paper named Rumshisky, trained their model to classify whether a patient would be readmitted into a hospital or not based on their medical records [2].

C. Impact and Effectiveness

This case study was quite beneficial in improving medical record interpretation and human error. Not only does it provide an automated approach to dealing with the vast amount of unstructured text data that is within the medical documentation, it furthers medical advancement for treating serious health issues such as cognitive, auditory, or mental issues. The paper also emphasizes that doctors utilize a terminology called "SNOMED-CT", that provides consistency when labeling patient outcomes and diseases. This terminology improves the accuracy of NLP models because there is reduced noise and subsequently reduces the loss on the model. This methodology enables intricate prediction capabilities, improved hospital triage, and adequately addressing patient concerns [2].

III. Discussion and Reflection

A. Impact and Implications

Overall, there is a profound positive impact on the use of natural language processing on medical/electronic health records. There can be streamlined and automated functionality within hospitals and other health care facilities that utilize the prediction, summarization, and semantic results of NLP models. Furthermore, models can become even more effective with the cohesive terminology being implemented by all doctors when writing medical notes about patients. A limitation that was discussed, however, is that doctors must still be able to write unique characteristics about a patient that may not fall under a specific term. Additionally, the NLP model may be bias and overfit based on the training data that is provided. The paper aims to continue addressing these concerns and develop further artificial intelligent machine learning models that are flexible in nature.

B. Future Applications

Based on the understanding gained from this paper, one could think about future applications of NLP methods within this regard. One application that could be addressed is the use of NLP on patient reviews. Hospitals could apply sentiment analysis and text summarization to gauge the attitudes and primary opinions of people who engaged in their facilities. The paper also mentioned that working with clinicians to explain the intricacies of machine learning NLP models will be imperative to the future developments/success rates of such medical concerns. Artificial intelligence must be interpretable, cohesive, and concise.

Not only can NLP be applied to the medical field, it can also be applied to many other human sectors of life. As mentioned in section I, NLP has proven to provide meaningful analysis on financial, educational, and technical domains. These techniques can be applied to many areas that deal with large amounts of text data such as social media posts, product reviews, and/or documents.

C. Reflection on Learning

Overall, this learning assignment allowed me to reflect on how NLP can be utilized within the real-world. There is a lot of technicality that must be upheld when creating NLP models and using them in an appropriate manner. I found the application of sentiment analysis and text extraction to be quite interesting. I think that there can be a lot of meaningful conversations surrounding the general opinion of a certain topic based on what these analyses are telling you. Furthermore, this assignment has enhanced my understanding of NLP's potential real-world implications by providing students with the opportunity to read academically published research papers and delve into the world of the researchers work.

References

[1] M. Bahja, "Natural application processing in business," *E-Business*, pp. 53 – 60, 2021.

[2] S. Locke, "Natural language processing in medicine: A review," *Medicine*, pp. 1 – 6, 2021.

[3] S. Boit, "Text analytics - week 13 content," 2024.