Karthik Srinivasan Muthusamy

2218930

2218930@brunel.ac.uk

NLP-based Clinical Decision Support and Text Classification in Medical Prescriptions

Ferdoos Hossein Nezhad

Abstract

An emerging trend in the field of interdisciplinary study known as natural language processing (NLP) is the use of deep learning techniques to solve problems involving the processing of human language. In their paper, Zhu et al. mention three applications of deep learning-based NLP techniques: healthcare-related NLP, deep learning's approach to handling EHR data characteristics, and the acknowledgement of deep learning's long road ahead in the healthcare revolution. (Zhu et al. 2019) In this dissertation, natural language processing is used to expedite the analysis of medical prescriptions, which in turn improves clinical decision-support systems. It is hoped that improved clinical decision making, and prescription interpretation will result from this. We use NLP to extract actionable data from prescription files, design a decision support system to aid in accurate diagnosis and treatment planning, develop and thoroughly test accurate prescription classification models. The project's primary focus is on data analysis, NLP algorithm implementation, and model development rather than system deployment and extensive clinical trials.

Natural language processing's (NLP) role in supporting clinical decision-making is systematically analysed; an NLP algorithm is developed to extract the relevant data; the algorithms' performance is evaluated. The overarching goal of this dissertation is to enhance clinical decision making through better prescription analysis using natural language processing techniques.

Background:

Qualitative methods are fundamental to numerous disciplines, including social sciences, history, education, and notably, healthcare. In the context of medical transcription, these methods are instrumental. The data in qualitative research, in this case, are transcribed medical records or patient narratives. Transcriptions are often created from recorded patient-doctor consultations, interviews, or medical reports, which are subsequently subjected to qualitative analysis. This approach yields comprehensive, nuanced insights into patients' health status, attitudes, experiences, and even patterns in their disease progression, supplementing the more clinical, quantitative data. By thoroughly analyzing these transcriptions, healthcare providers can derive a more holistic view of a patient's health, factoring in their lived experiences and perspectives. Moreover, qualitative analysis of medical transcripts can help to identify prevalent health behaviors, patient satisfaction, or areas where medical services may be lacking. In essence, it can enhance the depth and quality of patient-centered care and research. Indeed, for understanding the intricacies of patient experiences and the human side of healthcare, qualitative analysis of medical transcriptions can often provide a superior approach. (Leeson et al., 2019)

While Qualitative methods are basis for the healthcare, Clinical Decision Support (CDS) systems assist health professionals by providing accessible health-related information, significantly through Natural Language Processing (NLP). Demner-Fushman et al., (2009) in their paper states that NLP enables the extraction of data from narrative text, such as radiology reports and discharge summaries, enhancing all three primary components of CDS systems: patient data, decision rules/knowledge base, and patient-specific assessments/recommendations. CDS systems have improved practitioner performance in about 60% of reviewed cases, with proactive and automatic provision of decision support, and offering recommendations rather than assessments at the point of decision-making being key factors for success. Clinical decision support on patient health outcomes can be performed from free text with natural language processing techniques (Reyes-Ortiz et al., 2015). The literature

suggests a future need for reliable and high-quality NLP performance, as well as modular, flexible, and rapid systems. However, further research is required to gauge the effectiveness and adaptability of NLP systems for CDS and to establish evaluation methods for their impact on healthcare. (Demner et al., 2009)

In their seminal work, Al-Garadi et al. (2021) addressed the critical issue of prescription medication (PM) misuse and abuse, which has surged to the level of a national crisis in the United States. They recognized the potential of social media, particularly Twitter, as a resource for active monitoring of this issue.

Al-Garadi and colleagues extensively experimented with state-of-the-art bi-directional transformer-based language models like BERT, RoBERTa, XLNet, AlBERT, and DistilBERT, leveraging their ability to utilize tweet-level representations for transfer learning. In contrast to traditional machine learning and deep learning approaches, the focus of their work was on developing more sophisticated models capable of detecting and classifying instances of PM misuse from Twitter data.

The team proposed and evaluated fusion-based models, which showed significant improvement over traditional models. They reported an impressive F1-score (0.67 [95% CI: 0.64–0.69]) for these models, significantly outperforming the F1-score of traditional models (0.45 [95% CI: 0.42–0.48]). Additionally, Al-Garadi et al. (2021) noted the stability of these transformer-based models, which required less annotated data compared to other models, contributing to their effectiveness and efficiency.

Classification algorithm	Precision				Recall				F1-score				Accuracy (%)
	A	C	M	U	A	C	M	U	A	C	M	U	
BERT-2	0.6	0.79	0.86	0.91	0.61	0.77	0.86	0.85	0.61	0.78	0.86	0.88	79.85
RoBERTa	0.63	0.81	0.88	0.9	0.66	0.82	0.87	0.89	0.65	0.81	0.88	0.9	82.32
AlBERT	0.66	0.81	0.88	0.86	0.63	0.83	0.88	0.88	0.65	0.82	0.88	0.87	82.78
Proposed Fusion-1	0.6	0.84	0.91	0.78	0.76	0.81	0.84	0.93	0.67	0.82	0.87	0.85	82.22
Proposed Fusion-2	0.67	0.83	0.87	0.88	0.62	0.83	0.9	0.89	0.65	0.83	0.89	0.88	83.43
Proposed Fusion-3	0.56	0.83	0.9	0.75	0.73	0.8	0.83	0.92	0.64	0.82	0.86	0.82	80.92
Proposed Fusion-4	0.68	0.84	0.87	0.89	0.62	0.82	0.9	0.87	0.64	0.83	0.89	0.88	83.49

Table 1: Performances of transformer- and fusion-based model by Al-Garadi et al.

In conclusion, Al-Garadi et al. (2021) made a significant contribution to the literature, demonstrating the potential of transformer-based models for detecting self-reports of PM misuse on Twitter, while also outlining future research directions to further enhance these models.

Although transformer-based models have been found to yield higher accuracy, it is worth noting that the GloVe (Global Vectors for word representation) and Word2Vec methods played a pivotal role in laying the groundwork for the aforementioned transformer models. The success of Word2Vec and GloVe in capturing semantic relationships between words inspired researchers to explore more advanced techniques for word representations. The Transformer model, introduced in the paper "Attention Is All You Need" by Vaswani et al. in 2017, replaced the traditional recurrent neural network (RNN) and convolutional neural network (CNN) approaches in NLP tasks.

Model	Description	Pros	Cons
Word2Vec	A group of shallow two-layer neural networks, trained to reconstruct linguistic context of words.	Efficient and effective for creating word embeddings, reasonably handles semantic and syntactic word relationships.	Struggles with polysemy, does not consider sentence-level context.
GloVe	Combines global statistics of corpus (like LSA) with local context window methods (like Word2Vec).	Captures both global corpus-level information and local context, effective at capturing word relationships.	Similar to Word2Vec, struggles with polysemy, does not consider sentence-level context.
BERT	A transformer-based model that considers context in all directions, trained on a masked language model task.	Exceptionally good at understanding context, handles polysemy well by considering bidirectional context.	Requires significant computational resources, can be overkill for simpler tasks.
RoBERTa	A variant of BERT with optimized training approach and larger datasets.	Retains all benefits of BERT, improved performance due to optimized training.	Same as BERT, requires extensive computational resources.
AiBERT	This is the light version of BERT that has achieved new state-of-the-art results on several NLP benchmarks with fewer parameters compared with BERT-large	Likely shares the benefits of BERT and its variants, depending on the specific implementation and modifications.	Same as BERT, requires extensive computational resources due to its large structure

Table 2: Comparison of the Different NLP Models from legacy to New

Aim:

The aim of this project is to leverage Natural Language Processing (NLP) techniques to enhance clinical decision support systems and improve prescription analysis in medical settings.

Objectives:

To apply NLP techniques to extract relevant and meaningful information from medical prescriptions.
- To conduct a comprehensive analysis of existing research papers and literature on NLP techniques in clinical decision support and text classification of medical prescriptions.
- To develop NLP algorithms that effectively extract pertinent information from medical prescriptions.

- To develop a clinical decision support system using NLP algorithms that assists healthcare professionals in making accurate diagnoses and treatment decisions.
- To evaluate the performance and effectiveness of the developed models in improving prescription analysis.
- To generate an evaluation report comparing the performance of the developed models, utilizing appropriate metrics for assessing their effectiveness.

These objectives will guide the structure of the project and serve as measurable milestones towards achieving the broader aim of enhancing clinical decision support systems and advancing prescription analysis using NLP techniques.

Approach:

The proposed methodology is rooted in established literature and is designed to enhance clinical decision-making and prescription analysis using Natural Language Processing (NLP) techniques. As evidenced in previous research, NLP offers significant potential in healthcare, particularly in improving the efficiency and accuracy of Clinical Decision Support (CDS) systems (Demner-Fushman et al., 2009). The methodology encompasses several stages, each of which is critical to achieving the objectives outlined in the project proposal. Below, I provide an overview of each stage, justifying its inclusion based on established literature and practice.

Data Collection and Pre-processing:

The first stage involves collecting a corpus of medical prescriptions and pre-processing the data. Data collection is an essential first step in any research project (Creswell & Creswell, 2017). For the current project, this will be achieved using data extracted from MTSamples.com, which is also available on Kaggle. MTSamples.com is a rich resource that provides a wide collection of transcribed medical reports, encompassing various specialties and types of work. These transcripts offer an excellent dataset for our study as they encapsulate the intricacies of patient narratives and medical prescriptions. This data is open to the public for reference, and its utilization in this project complies with the website's sharing and distribution policies. After data collection, it will be cleaned and normalized to ensure it is in a suitable form for the subsequent analysis. This stage is crucial in guaranteeing the quality of the data, which will directly impact the effectiveness of the subsequent stages (Kuhn & Johnson, 2013).

NLP Technique Application:

The second stage involves applying NLP techniques, namely GloVe and Word2Vec, to the pre-processed data. These techniques will be used to generate word embeddings, which represent the words in the text as high-dimensional numeric vectors (Mikolov et al., 2013; Pennington et al., 2014). These techniques have been selected due to their ability to efficiently capture semantic relationships between words and their proven effectiveness in NLP tasks (Almeida et al., 2020).

Model Development:

The third stage involves the development of advanced machine learning models that will use the word embeddings generated in stage two to classify the medical prescriptions. The selected models, BERT, ALBERT, RoBERTa, and fusion models, have shown promising results in other NLP tasks, suggesting their potential utility for this project (Devlin et al., 2019; Al-Garadi

et al., 2021). The use of fusion models will also be explored, as these have been shown to provide enhanced performance in similar tasks (Al-Garadi et al., 2021).

Model Training and Evaluation:

The fourth stage involves training and evaluating the developed models. This stage is crucial for understanding the effectiveness of the models in classifying the medical prescriptions, and it will provide insights into their relative strengths and weaknesses. As suggested in previous research, the evaluation will be conducted using appropriate metrics such as accuracy, recall, precision, and F1-score (Sokolova & Lapalme, 2009).

Documentation and Reporting:

The final stage involves the production of a comprehensive report detailing the results of the project. Reporting is a critical component of the research process, allowing for the dissemination of the findings, providing insights for future research, and potentially informing healthcare practice.

By following this approach, the project aims to meet the objectives stated in the proposal, which are anchored in the overarching goal of enhancing clinical decision support and prescription analysis through the application of NLP techniques. As evidenced by the existing literature, this methodology presents a promising approach for improving healthcare outcomes, a goal at the core of this project.

Plan:

Gantt charts have been used in project management for over a century, but their popularity has not waned. When applied to projects, they excel at showcasing crucial data in a concise manner. Because of practical concerns and technological advancements, their earlier applications to more general production planning and control issues are now obsolete. Gantt charts continue to play a role because they offer a readily useful interface that enables users to define problems and accept solutions more easily, even though computing offers more robust techniques for modelling these problems. (Wennink et al., 2011)

For this Project we will use the Gantt chart and it is utilised throughout the entirety of the research process, beginning with the Planning stage, and continuing through monitoring the following Major tasks and Milestones.

1. Data Collection and Pre-processing
2. Literature Review
    Milestone: Completion of data collection, pre-processing and Literature Review
3. NLP Technique Application
    Milestone: Successful application of NLP techniques and generation of word embeddings
4. Model Development
    Milestone: Completion of model development (BERT, RoBERTa, AlBERT, and Fusion Models)
5. Model Training and Evaluation
    Milestone: Completion of model training and evaluation. Finalization of model selection based on the performance metrics.

6. Documentation and Reporting

Milestone: Finalization of the comprehensive report including all the analysis, insights, and potential future directions.

Figure 1: Gantt Chart
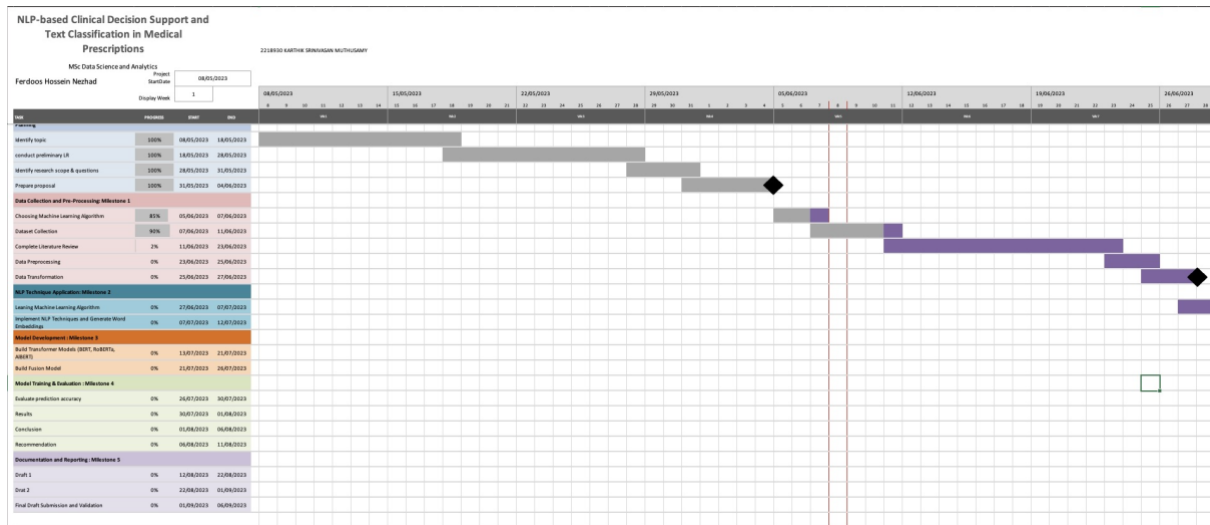
References:

- Al-Garadi, M.A. et al. (2021) Text classification models for the automatic detection of nonmedical prescription medication use from social media - BMC Medical Informatics and decision making, BioMed Central. Available at: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-021-01394-0#Sec3 (Accessed: 02 June 2023).
- Almeida Melonie , de et al. (2020) Deep Learning Enabled Semantic Communication Systems | IEEE journals ..., 20th International Conference on Advances in ICT for Emerging Regions (ICTer). Available at: https://ieeexplore.ieee.org/abstract/document/9398576 (Accessed: 06 June 2023).
- Demner-Fushman, D., W. Chapman, W. and J. McDonald, C. (2009) What can natural language processing do for clinical decision support?, Journal of Biomedical Informatics. Available at: https://doi.org/10.1016/j.jbi.2009.08.007 (Accessed: 28 May 2023).
- Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. (Accessed: 06 June 2023)
- J.W., Creswell. and J.D., C. (2017) Research Design Qualitative, quantitative, and mixed methods approaches. 4th edition, scientific research publishing. Available at: https://www.scirp.org/(S(351jmbntvnsjt1aadkozje))/reference/referencespapers.aspx?referenceid=2969274 (Accessed: 08 June 2023).
- Kuhn, M. and Johnson, K., 2013. Applied predictive modeling (Vol. 26, p. 13). New York: Springer (Accessed: 08 June 2023)
- Leeson, W. et al. (2019) 'Natural language processing (NLP) in qualitative public health research: A proof of concept study', International Journal of Qualitative Methods, 18, p. 160940691988702. doi:10.1177/1609406919887021 (Accessed: 06 June 2023).
- Mikolov, T., Chen, K., Corrado, G. and Dean, J., 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781. (Accessed: 06 June 2023).
- Pennington, J., Socher, R. and Manning, C.D., 2014, October. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543) (Accessed: 06 June 2023).
- Reyes-Ortiz, J.A., Gonzalez-Beltran, B.A. and Gallardo-Lopez, L. (2015) 'Clinical decision support systems: A survey of NLP-based approaches from unstructured data', 2015 26th International Workshop on Database and Expert Systems Applications (DEXA) [Preprint]. doi:10.1109/dexa.2015.47. (Accessed: 29 May 2023)
- Sokolova, M. and Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Information processing & management, 45(4), pp.427-437. (Accessed: 03 June 2023)
- Vaswani. et al,. (2017) Attention is all you need. Available at: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf (Accessed: 04 June 2023).
- Wennink, M. et al. (2011) Gantt charts: A centenary appreciation, European Journal of Operational Research. North-Holland. Available at:

https://www.sciencedirect.com/science/article/pii/S0377221702007695#aep-section-id20 (Accessed: May 28, 2023).

- Zhenzhong, L. et al. (2020) ALBERT: A LITE BERT FOR SELF-SUPERVISED LEARNING OF LANGUAGE REPRESENTATIONS, Published as a conference paper at ICLR 2020. Available at: https://arxiv.org/pdf/1909.11942.pdf (Accessed: 04 June 2023).
- Zhu, R., Tu, X. and Huang, J. (2019) 'Using deep learning based natural language processing techniques for clinical decision-making with ehrs', Studies in Big Data, pp. 257–295. doi:10.1007/978-3-030-33966-1_13 (Accessed: 06 June 2023).