Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-8071-1_P235-cd

NLP Advances in Risk Analysis Context: Application of Quantum Computing

July Bias Macedo[a], Isis Didier Lins[a] and Márcio José das Chagas Moura[a]

[a]*Center for Risk Analysis and Environmental Modeling, Department of Industrial Engineering, Federal University of Pernambuco, Brazil. E-mail: july.bias@ufpe.br, isis.lins@ufpe.br, marcio.cmoura@ufpe.br*

The models used for facilitating Risk Analysis have undergone notable changes in recent times. This can be attributed to the significant progress made in computing capabilities, as well as the breakthroughs achieved in the field of artificial intelligence. These developments have allowed for the effective extraction of information from complex and unstructured datasets. A particularly useful tool in this regard is Natural Language Processing (NLP), which facilitates the extraction, organization, and classification of information from textual sources. This enables the identification of patterns and trends in an automated fashion. The emergence of Quantum Computing (QC) over the last fifty years has made it one of the most significant fields in computational science. Although fully scalable QC remains elusive, the availability of intermediate noisy scalable QC devices has enabled the realization of near-term computation on quantum devices using quantum algorithms. QC can (i) solve classically intractable problems and (ii) problems that, while tractable, are classically infeasible. Given the significant computational resources required by state-of-the-art NLP models to extract meaning from text, exploring methods to address such problems using QC would be valuable. The primary aim of this study is to develop a model based on Quantum Natural Language Processing (QNLP) to classify whether human error caused aviation accidents, based on the accident narrative. The research utilizes a database of accident investigation reports conducted by the National Transportation Safety Board (NTSB) spanning from 1982 to 2022.

Keywords: Quantum Natural Language Processing; Quantum Computing; Aviation Accidents; Accident Investigation Reports.

1. Introduction

RA is critical for guiding investments to prevent and mitigate risk events, particularly in the aviation industry where accidents often result in loss of life. The aviation industry collects vast amounts of data from various sources, such as written accident reports. The internal investigation and reporting of accidents are aimed at preventing similar incidents from occurring in the future. However, due to the sheer volume of reports produced, a complete human review is impractical [1].

There has been an increase in the number of NLP approaches to RA, mainly due to breakthroughs in the field of artificial intelligence and advances in data processing and recording [2]. However, as far as the authors are concerned, no QNLP approaches have been proposed in the context of RA. Thus, the main objective of this research is to develop a QNLP-based model that can extract knowledge from textual data in aviation accident reports that can be useful to improve safety culture.

2. Natural Language Processing

In general, algorithms need to convert raw text inputs into numerical representations (word embeddings) through a process known as language modeling, which forms the basis for knowledge distillation [3]. Thus, several modelling approaches have been designed, such as Deep Learning-based models. However, training such models has a high computational cost.

In recent years, quantum computing has become increasingly popular due to its potential to significantly enhance computational power. This technology harnesses the properties of quantum mechanics, including superposition and entanglement, to manipulate and represent data. By allowing for the simultaneous evaluation of multiple hypotheses, quantum computers offer a significant advantage over classical computers.

Quantum embeddings, which capture the semantic structure of text data, are created by utilizing the principles of quantum mechanics. Initially, text data is transformed into a vector space using classical techniques. Subsequently, a

quantum circuit applies a series of unitary operations to the vector representation to map it to a quantum state [4].

In this paper, we developed a QNLP model that can effectively classify aviation accidents as either caused by human error or not using binary classification. The experimental methodology and results are presented in the following section.

3. Experiment

The proposed model was applied using a public database comprising accident investigation reports conducted by the National Transportation Safety Board (NTSB) between 1982 and 2022 [5].

First, the accident descriptions were preprocessed by applying two operations: (i) stop word filtering and (ii) lowercasing. All descriptions, written in English, were processed using Python string methods and functions from the NLTK library. Aviation accidents caused by human error were randomly selected by searching for the expression 'pilot's failure' in the accident descriptions. In other words, accidents with the 'pilot's failure' expression were categorized as caused by human error, while all others were considered due to non-human error.

Three datasets were constructed, comprising (i) 32 training and 9 test samples, (ii) 80 training and 21 test samples, and (iii) 160 training and 41 test samples, and used to train and evaluate the performance of the QNLP classifiers. The Lambeq pipeline [6] was adopted to convert sentences to quantum circuits, and develop a QNLP model for determining whether an aviation accident was caused by human error based on the accident narrative. The models were trained using the 'binary cross-entropy' loss function, with a 'sigmoid' activation function applied in the final layer. 'Adam' was chosen as the optimizer, and a learning rate of 10^{-3} and batch size of 4 were employed in training the classifiers.

The QNLP models demonstrated accuracy rates on test data of 55.56% (i), 57.14% (ii), and 65.85% (iii). Notably, the training duration for (i) was less than one minute, while the training for (iii) took about five minutes. Additionally, we ensured that the accident narratives were as realistic as possible without oversimplification, further strengthening the validity of the model. These promising results demonstrate the potential of QNLP in identifying human error as the cause of aviation accidents.

4. Conclusion

Quantum embeddings have demonstrated potential in a range of natural language processing tasks, including sentiment analysis, text classification, and machine translation. Although achieving fully scalable quantum computing remains elusive, the world's major technology companies, including Google, IBM, Microsoft, and Amazon, invest billions of dollars in quantum computing research and development and provide limited access to quantum computers for public use. Consequently, developments in quantum computing hold significant promise for enhancing the efficiency of NLP-based models.

Acknowledgement

The authors thank CNPq (nº 409701/2022-0, 310892/2022-8, 305696/2018-1) and FACEPE for the financial support through research grants. This study was financed in part by CAPES – Finance Code 001.

References

[1] Perboli G, et al. Natural Language Processing for the Identification of Human Factors in aviation accidents causes An Application to the SHEL Methodology. Expert Syst Appl 2021;186:115694.

[2] Puranik TG, et al. Towards online prediction of safety-critical landing metrics in aviation using supervised machine learning. Transp Res Part C Emerg Technol 2020;120:102819.

[3] Bianchi B, et al. Human and computer estimations of the Predictability of words in written language. Sci Rep 2020;10:1–11.

[4] Meichanetzidis K et al. Quantum natural language processing on near-term quantum computers. Electron Proc Theor Comput Sci EPTCS 2021;340:213–29.

[5] NTSB. Accident Data. Download Census US Civ Aviat Accid 2022. https://data.ntsb.gov/avdata (accessed August 1, 2022).

[6] Kartsaklis D et al. lambeq: An Efficient High-Level Python Library for Quantum NLP. ArXiv Prepr 2021;arXiv:2110:1–19.