25th International Congress on Modelling and Simulation, Darwin, NT, Australia, 9 to 14 July 2023
mssanz.org.au/modsim2023

NLP-based predictive model for identifying and presenting construction accident scenarios

Seung-Hyeon Shin [a], **Jeong-Hun Won** [a,b,c], **Gyeongyun Kim** [b] and **Hyeon-Ji Jeong** [c]

[a] *Department of Big Data, Chungbuk National University, Cheongju, Republic of Korea*
[b] *Department of Safety Engineering, Chungbuk National University, Cheongju, Republic of Korea*
[c] *Department of Disaster Prevention Engineering, Chungbuk National University, Cheongju, Republic of Korea*
Email: jhwon@chungbuk.ac.kr

Abstract: This study presents an NLP-based predictive model for identifying and presenting similar construction accident scenarios with the aim of enhancing safety and health management. The construction industry is recognized as one of the most dangerous industries worldwide with a high fatality rate. In Korea, work-related fatalities in the construction sector are 3.09 times higher on average than those in the manufacturing industry from 2013 to 2022 (Ministry of Employment and Labor in Korea, 2023). To reduce accidents, stakeholder involvement in safety and health (S&H) management systems is crucial. Despite government interventions, accident prevention is limited due to stakeholders' constrained S&H management capabilities. One of the most effective ways to improve stakeholders' S&H management capabilities is to make them aware of accidents that may occur during construction work, and the hazards that cause them. Therefore, this study proposes a model that presents accident scenarios at construction sites to stakeholders.

To address this issue, an accident prediction model was developed for the construction industry using natural language processing (NLP) techniques. The model was trained on 18,000 accident cases collected by the Korea Authority of Land and Infrastructure Safety and contained variables such as accident type, time, weather, work type, and number of fatalities and injuries. The model was used the BERT model. The model performance was qualitatively evaluated by construction S&H experts who assessed its interpretability, relevance, and applicability in real-world construction scenarios. The S&H experts' feedback helped refine the model to ensure its effectiveness in predicting accidents and presenting similar scenarios. Stakeholders input construction site information, such as process rate, construction period, cost, weather, and work details, in natural language. The model retrieves accident information from similar sites based on embedded data and calculates a similarity score using cosine similarity. This score helps stakeholders to evaluate the relevance of accident scenarios to their sites. The model also provides site information at the time of an accident, thereby offering a valuable context. By examining the similarity score and on-site information, stakeholders can assess the likelihood of a proposed accident occurring at their construction site, thereby enabling them to develop targeted measures to prevent potential accidents and enhance S&H management in the construction industry. This study contributes to the improvement of the S&H management capabilities of construction stakeholders and proposes a novel approach to improving stakeholder awareness of potential accidents and hazards. Despite these promising results, there are limitations to be addressed in future research. The model's performance could be improved by incorporating additional variables and data sources and expanding its applicability to a wider range of construction activities and industries.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation, Korea, under project BK21 FOUR; and a Human Resources Development of the Korea Institute of Energy Technology Evaluation and Planning grant funded by the Korean government (No. 20224000000070).

REFERENCES

Oliaee A.H., Das S., Liu J., Rahman M.A. (2023). Using Bidirectional Encoder Representations from Transformers (BERT) to classify traffic crash severity types. Nat. Lang. Process 3, pp.100007.

Pinto, A., Nunes, I.L., Ribeiro, R.A. (2011). Occupational risk assessment in construction industry – Overview and reflection. Saf. Sci. 49(5), pp. 616–624.

Keywords: *Natural language processing, construction accident scenarios, BERT model*