Material named entity recognition (MNER) for knowledge-driven materials using deep learning approach

Saef Ullah Miah M.[a]*; Sulaiman, Junaida*[a, b]

[a] Faculty of Computing, College of Computing and Applied Sciences, Universiti Malaysia Pahang, Pahang, Pekan, 26600, Malaysia

[b] Center for Data Science and Artificial Intelligence (Data Science Center), Universiti Malaysia Pahang, Pahang, Pekan, 26600, Malaysia

ABSTRACT

The scientific literature contains an abundance of cutting-edge knowledge in the field of materials science, as well as useful data (e.g., numerical values from experimental results, properties, and structure of materials). To speed up the identification of new materials, these data are essential for data-driven machine learning (ML) and deep learning (DL) techniques. Due to the large and growing amount of publications, it is difficult for humans to manually retrieve and retain this knowledge. In this context, we investigate a deep neural network model based on Bi-LSTM to retrieve knowledge from published scientific articles. The proposed deep neural network-based model achieves an F1 score of 9 ~ 7 % for the Material Named Entity Recognition (MNER) task. The study addresses motivation, relevant work, methodology, hyperparameters, and overall performance evaluation. The analysis provides insight into the results of the experiment and points to future directions for current research.

KEYWORDS

Bi-LSTM; EDLC; Material named entity recognition; Materials science; Named entity recognition

REFERENCES

[1]Miah, M.S.U., Sulaiman, J., Sarwar, T.B., Naseer, A., Ashraf, F., Zamli, K.Z., Jose, R.: Sentence boundary extraction from scienti_c literature of electric double layer capacitor domain: Tools and techniques. Applied Sciences 12(3), 1352 (2022)

[2]Nadeau, D., Sekine, S.: A survey of named entity recognition and classification. Lingvisticae Investigationes 30(1), 3{26 (2007)

[3]Goyal, A., Gupta, V., Kumar, M.: Recent named entity recognition and classification techniques: a systematic review. Computer Science Review 29, 21{43 (2018)

[4]Jose, R., Ramakrishna, S.: Materials 4.0: Materials big data enabled materials discovery. Applied Materials Today 10, 127{132 (2018)

[5]Lowe, D.M., Sayle, R.A.: Leadmine: a grammar and dictionary driven approach to entity recognition. Journal of cheminformatics 7(1), 1{9 (2015)