ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

4 Conclusion and future work

We compared different approaches for pre-training entity and entity-pair embeddings to improve relation extraction performance in the biomedical domain. We have shown that (1) incorporation of these embeddings into neural models helps in achieving better performance, (2) using rich features as context (instead of using the surrounding words, i.e. the normal word2vec approach) leads to better results; (3) using pair embeddings with/without entity embeddings leads to better results compared to using entity embeddings alone. Our best model achieves an F-score of 77.19, improving the best previous result by +0.73pp over a strong baseline, and setting a new state-of-the-art for the task. As future work, we aim to investigate the effect of entity and entity-pair embeddings on other biomedical relation extraction data sets.

5 Acknowledgements

We would like to thank Dr. Sampo Pyysalo for his invaluable recommendations and Dr. Jari Björne for his help in running TEES software. The research was partly funded by the Finnish Cultural Foundation and Academy of Finland (315376).

References

[1] Martin Krallinger, Obdulia Rabal, Saber A Akhondi, Martín Pérez-Pérez, Jesus Santamaría, et al. Overview of the BioCreative VI chemical-protein interaction track. In *Proceedings of the sixth BioCreative challenge evaluation workshop*, volume 1, pages 141–146, Bethesda, MD, USA, 2017.

[2] Yifan Peng, Anthony Rios, Ramakanth Kavuluru, and Zhiyong Lu. Extracting chemical-protein relations with ensembles of SVM and deep learning models. *Database*, 2018, 07 2018.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[4] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 09 2019.

[5] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates Inc., 2013.

[6] Omer Levy and Yoav Goldberg. Dependency-based word embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 302–308, 2014.

[7] Jari Björne. *Biomedical Event Extraction with Machine Learning*. PhD thesis, TUCS Dissertations, 2014.

[8] Kai Hakala, Suwisa Kaewphan, Tapio Salakoski, and Filip Ginter. Syntactic analyses and named entity recognition for PubMed and PubMed Central –up-to-the-minute. In *Proceedings of the 2016 Workshop on Biomedical Natural Language Processing*, pages 102–107. Association for Computational Linguistics, 2016.