The foregoing description of the exemplary embodiments of the disclosure has been presented only for the purposes of illustration and description and is not intended to be exhaustive or to limit the disclosure to the precise forms disclosed. Many modifications and variations are possible in light of the above teaching.

The embodiments were chosen and described in order to explain the principles of the disclosure and their practical application so as to enable others skilled in the art to utilize the disclosure and various embodiments and with various modifications as are suited to the particular use contemplated. Alternative embodiments will become apparent to those skilled in the art to which the present disclosure pertains without departing from its spirit and scope. Accordingly, the scope of the present disclosure is defined by the appended claims rather than the foregoing description and the exemplary embodiments described therein.

Listing of References (incorporated herein by reference in their entirety):

1. Christoph Alt, Marc Hubner, and Leonhard Hennig, Improving relation extraction by pre-trained language representations, 2019, arXiv:1906.03088.

2. Iz Beltagy, Kyle Lo, and Arman Cohan, SciBERT: a pretrained language model for scientific text, 2019, arXiv:1903.10676.

3. Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou, A walk-based model on entity graphs for relation extraction, 2018, arXiv:1902.07023.

4. Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou, Connecting the dots: document-level neural relation extraction with edge-oriented graphs, 2019, arXiv:1909.00228.

5. Junyong Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014, arXiv:1412.3555.

6. Kevin Clark, Urvashi Khandelwal, Omer Levy, Christopher D. Manning, What does BERT look at? An analysis of BERT's attention, 2019, arXiv:1906.04341.

7. Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, 2019,

arXiv:1810.04805.

8. Rong-En Fan, and Chih-Jen Lin, A study on threshold selection for multi-label classification, 2007, Semantic Scholar.

9. Priya Goyal, Piotr Dollár, Ross Girshick, Pieter Noordhuis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, Kaiming He, Accurate, large minibatch SGD: training ImageNet in 1 hour, 2017, arXiv:1706.02677.

10. Zhijiang Guo, Yan Zhang, and Wei Lu, Attention guided graph convolutional networks for relation extraction, 2019, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 241-251.

11. Pankaj Gupta, Subburam Rajaram, Hinrich Schütze, Bernt Andrassy, Thomas Runkler, Neural relation extraction within and across sentence boundaries, 2019, arXiv:1810.05102.

12. Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano, Stan Szpakowicz, SemEval-2010 Task 8: multi-way classification of semantic relations between pairs of nominals, 2010, Proceedings of the 5th International Workshop on Semantic Evaluation, 33-38.

13. John Hewitt, Christopher D. Manning, A structural probe for finding syntax in word representations, NAACL-HLT, 2019, 4129–4138.

14. Sepp Hochreiter and Jurgen Schmidhuber, Long short-term memory. Neural Computation, 1997, 9(8): 1735–1780.

15. Robin Jia, Cliff Wong, and Hoifung Poon, Document-level N-ary relation extraction with multiscale representation learning, 2019, arXiv:1904.02347.

16. Urvashi Khandelwal, He He, Peng Qi, and Dan Jurafsky, Sharp nearby, fuzzy far away: how neural language models use context, 2018, arXiv:1805.04623.

17. Thomas N. Kipf and Max Welling, Semi-supervised classification with graph convolutional networks, 2017, arXiv:1609.02907.

18. Jiao Li, Yueping Sun, Robin J. Johnson, Daniela Sciaky, Chih-Hsuan Wei, Robert Leaman, Allan Peter Davis, Carolyn J. Mattingly, Thomas C. Wiegers, and Zhiyong Lu, BioCreative V CDR task corpus: a resource for chemical disease relation extraction,

Database, 2016, 1-10.

19. Yann LeCun, Yoshua Bengio & Geoffrey Hinton, Deep learning, Nature, 2015, 521:436–444.

20. Xiaodan Liang, Xiaohui Shen, Jiashi Feng, Liang Lin, and Shuicheng Yan, Semantic object parsing with graph LSTM, 2016, arXiv:1603.07063.

21. Yang Liu, and Mirella Lapata, Learning structured text representations, Transactions of the Association for Computational Linguistics, 2018, 6: 63–75.

22. Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov, RoBERTa: a robustly optimized BERT pretraining approach, 2019, arXiv:1907.11692.

23. Ilya Loshchilov, and Frank Hutter, Decoupled weight decay regularization, 2019, ICLR 2019 Conference.

24. Aditya K. Menon, Ankit Singh Rawat, Sashank Reddi, and Sanjiv Kumar, Multilabel reductions: what is my loss optimising? 2019, NeurIPS2019.

25. Paulius Micikevicius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Ginsburg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, Hao Wu, Mixed precision training, 2018, arXiv:1710.03740.

26. Makoto Miwa, Mohit Bansal, End-to-End Relation Extraction using LSTMs on sequences and tree structures, 2016, arXiv:1601.00770.

27. Guoshun Nan, Zhijiang Guo, Ivan Sekulic, Wei Lu, Reasoning with latent structure refinement for document-level relation extraction, 2020, arXiv:2005.06312.

28. Dat Quoc Nguyen, and Karin Verspoor, Convolutional neural networks for chemical-disease relation extraction are improved with character-based word embeddings, 2018, arXiv:1805.10586.

29. Nanyun Peng, Hoifung Poon, Chris Quirk, Kristina Toutanova, and Wen-tau Yih, Cross-sentence N-ary relation extraction with graph LSTMs, 2017, arXiv:1708:03743.

30. Chris Quirk, and Hoifung Poon, Distant supervision for relation extraction beyond the sentence boundary, 2017, arXiv:1609.04873.

31. Sashank J. Reddi, Satyen Kale, Felix Yu, Dan Holtmann-Rice, Jiecao Chen, and Sanjiv Kumar, Stochastic negative mining for learning with large output spaces, 2019,

arXiv:1810.07076.

32. Mike Schuster, and Kuldip K. Paliwal, Bidirectional recurrent neural networks, IEEE Transactions on Signal Processing, 1997, 45(11): 2673-2681.

33. Peng Shi, and Jimmy Lin, Simple BERT models for relation extraction and semantic role labeling, 2019, arXiv:1904.05255.

34. Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski, Matching the blanks: distributional similarity for relation learning, 2019, arXiv:1906.03158.

35. Linfeng Song, Yue Zhang, Zhiguo Wang, and Daniel Gildea, N-ary relation extraction using graph state LSTM, EMNLP, 2018, 2226-2235.

36. Hengzhu Tang, Yanan Cao, Zhenyu Zhang, Jiangxia Cao, Fang Fang, Shi Wang, and Pengfei Yin, HIN: hierarchical inference network for document-level relation extraction, Advances in Knowledge Discovery and Data Mining, 2020, 12084: 197–209.

37. Yun Tang, Jing Huang, Guangtao Wang, Xiaodong He, Bowen Zhou, Orthogonal relation transforms with graph context modeling for knowledge graph embedding, ACL, 2020, 2713-2722.

38. Ian Tenney, Dipanjan Das, and Ellie Pavlick, BERT rediscovers the classical NLP pipeline, ACL, 2019, 4593-4601.

39. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin, Attention is all you need, 2017, arXiv:1706.03762.

40. Patrick Verga, Emma Strubell, and Andrew McCallum, Simultaneously self-attending to all mentions for full-abstract biological relation extraction, NAACL-HLT, 2018, 872-884.

41. Jesse Vig, and Yonatan Belinkov, Analyzing the structure of attention in a transformer language model, 2019, arXiv:1906.04284.

42. Hong Wang, Christfried Focke, Rob Sylvester, Nilesh Mishra, and William Wang, Fine-tune BERT for DocRED with two-step process, 2019, arXiv:1909.11898.

43. Haoyu Wang, Ming Tan, Mo Yu, Shiyu Chang, Dakuo Wang, Kun Xu, Xiaoxiao Guo, and Saloni Potdar, Extracting multiple-relations in one-pass with pre-trained

transformers, 2019, arXiv:1902.01030.

44. Linlin Wang, Zhu Cao, Gerard de Melo, and Zhiyuan Liu, Relation classification via multi-Level attention CNNs, ACL, 2016, 1298-1307.

45. Felix Wu, Tianyi Zhang, Amauri Holanda de Souza Jr, Christopher Fifty, Tao Yu, Kilian Q Weinberger, Simplifying graph convolutional networks, 2019a, arXiv:1902.07153.

46. Ye Wu, Ruibang Luo, Henry C.M. Leung, Hing-Fung Ting, and Tak-Wah Lam, RENET: a deep learning approach for extracting gene-disease associations from literature, RECOMB, 2019b, 272-284.

47. Yuan Yao, Deming Ye, Peng Li, Xu Han, Yankai Lin, Zhenghao Liu, Zhiyuan Liu, Lixin Huang, Jie Zhou, and Maosong Sun, DocRED: a large-scale document-level relation extraction dataset, ACL, 2019, 764-777.

48. Deming Ye, Yankai Lin, Jiaju Du, Zhenghao Liu, Peng Li, Maosong Sun, Zhiyuan Liu, Coreferential reasoning learning for language representation, 2020, arXiv:2004.06870.

49. Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao, Relation classification via convolutional deep neural network, COLING, 2014, 2335-2344.

50. Yuhao Zhang, Peng Qi, and Christopher D. Manning, Graph convolution over pruned dependency trees improves relation extraction, EMNLP, 2018, 2205-2215.

51. Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D. Manning, Position-aware attention and supervised data improve slot filling, EMNLP, 2017, 35-45.

52. Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu, ERNIE: enhanced language representation with informative entities, 2019, arXiv:1905.07129.

53. Heliang Zheng, Jianlong Fu, Zheng-Jun Zha, and Jiebo Luo, Learning deep bilinear transformation for fine-grained image representation, 2019, arXiv:1911.03621.