labeled OpenIE data from semantic role labeling, making supervised neural sequence labeling possible for OpenIE. Our representation of scientific statement sentences is different including modeling attributes and condition tuples.

Sequence Labeling: Neural networks have been applied to sequence labeling tasks with more promising performance than traditional statistical methods. Neural encoder-decoder model is one of the paradigms [13, 27, 30]. Ma *et al.* [13] combined CNNs-encoded character-level and word-level representations into BiLSTM to model contextual information of each word following a CRF to decode labels for the whole sentence. Yang *et al.* [27] leveraged continuous representations of KBs to enhance LSTM for sequence labeling. Zheng *et al.* [30] proposed BiLSTM-LSTMd which contains a BiLSTM as encoder and a new LSTM-decoder (LSTMd) as decoder. BiLSTM-LSTMd outperformed previous models on entity and relation extraction tasks [30]. We use BiLSTM-LSTMd for the sequence labeling module in our approach.

6 CONCLUSIONS

In this work, we proposed a novel representation of SciKG with the role of condition. The SciKG had three layers: concept/attribute nodes, fact/condition tuples and statement nodes. Inspired by a recent work that considers open information extraction as a sequence labeling task, we proposed a semi-supervised Multi-Input Multi-Output (MIMO) sequence labeling model that learned complex dependencies between the sequence tags from multiple signals to generate output sequences for fact/condition tuples. Experiments showed that our model outperformed existing methods.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable comments and insightful suggestions. This work was supported in part by NSF Grants IIS-1447795 and CNS-1622914, and National Natural Science Foundation of China (NSFC) No. 61632011 and No. 61772156.

REFERENCES

[1] Gabor Angeli, Melvin Jose Johnson Premkumar, and Christopher D. Manning. 2015. Leveraging Linguistic Structure For Open Domain Information Extraction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL)*. 344–354.

[2] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2D knowledge graph embeddings. In *32nd AAAI Conference on Artificial Intelligence (AAAI)*. 1811–1818.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*.

[4] Kata Gábor, Davide Buscaldi, Anne-Kathrin Schumann, Behrang QasemiZadeh, Haifa Zargayouna, and Thierry Charnois. 2018. Semeval-2018 Task 7: Semantic relation extraction and classification in scientific papers. In *Proceedings of The 12th International Workshop on Semantic Evaluation (SemEval)*. 679–688.

[5] Sonal Gupta and Christopher Manning. 2011. Analyzing the Dynamics of Research by Extracting Key Aspects of Scientific Papers. In *Proceedings of 5th International Joint Conference on Natural Language Processing (IJCNLP)*. 1–9.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. 770–778.

[7] Jeremy Howard and Sebastian Ruder. 2018. Universal language model fine-tuning for text classification. In *Proceedings of the 56rd Annual Meeting of the Association for Computational Linguistics (ACL)*. 328–339.

[8] Meng Jiang, Jingbo Shang, Taylor Cassidy, Xiang Ren, Lance M Kaplan, Timothy P Hanratty, and Jiawei Han. 2017. Metapad: Meta pattern discovery from massive text corpora. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM, 877–886.

[9] Matthieu Labeau, Kevin Löser, and Alexandre Allauzen. 2015. Non-lexical neural architecture for fine-grained POS tagging. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 232–237.

[10] John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the Eighteenth International Conference on Machine Learning (ICML)*. 282–289.

[11] Qi Li, Meng Jiang, Xikun Zhang, Meng Qu, Timothy P Hanratty, Jing Gao, and Jiawei Han. 2018. Truepie: Discovering reliable patterns in pattern-based information extraction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 1675–1684.

[12] Yi Luan, Luheng He, Mari Ostendorf, and Hannaneh Hajishirzi. 2018. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 3219–3232.

[13] Xuezhe Ma and Eduard Hovy. 2016. End-to-end sequence labeling via bi-directional lstm-cnns-crf. In *Proceedings of the 54rd Annual Meeting of the Association for Computational Linguistics (ACL)*. 1064–1074.

[14] David L Miller. 1947. The Nature of Scientific Statements. *Philosophy of Science* 14, 3 (1947), 219–223.

[15] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1532–1543.

[16] Xiang Ren, Jiaming Shen, Meng Qu, Xuan Wang, Zeqiu Wu, Qi Zhu, Meng Jiang, Fangbo Tao, Saurabh Sinha, David Liem, et al. 2017. Life-iNet: A Structured Network-Based Knowledge Exploration and Analytics System for Life Sciences. *Proceedings of the 55rd Annual Meeting of the Association for Computational Linguistics (ACL)* (2017), 55–60.

[17] Michael Schmitz, Robert Bart, Stephen Soderland, Oren Etzioni, et al. 2012. Open language learning for information extraction. In *Proceedings of the 2012 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 523–534.

[18] Jingbo Shang, Jialu Liu, Meng Jiang, Xiang Ren, Clare R Voss, and Jiawei Han. 2018. Automated phrase mining from massive text corpora. *IEEE Transactions on Knowledge and Data Engineering (TKDE)* 30, 10 (2018), 1825–1837.

[19] Gabriel Stanovsky, Julian Michael, Luke Zettlemoyer, and Ido Dagan. 2018. Supervised open information extraction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*. 885–895.

[20] Martin Sundermeyer, Ralf Schlüter, and Hermann Ney. 2012. LSTM neural networks for language modeling. In *Thirteenth annual conference of the international speech communication association (ISCA)*.

[21] Victor N Tomilin, Alena L Cherezova, Yuri A Negulyaev, and Svetlana B Semenova. 2016. TRPV5/V6 channels mediate Ca2+ influx in jurkat T cells under the control of extracellular pH. *Journal of cellular biochemistry* 117, 1 (2016), 197–206.

[22] Ioannis Tsochantaridis, Thorsten Joachims, Thomas Hofmann, and Yasemin Altun. 2005. Large margin methods for structured and interdependent output variables. *Journal of machine learning research* 6, Sep (2005), 1453–1484.

[23] Xueying Wang, Haiqiao Zhang, Qi Li, Yiyu Shi, and Meng Jiang. 2019. A Novel Unsupervised Approach for Precise Temporal Slot Filling from Incomplete and Noisy Temporal Contexts. In *The World Wide Web Conference*.

[24] Xuan Wang, Yu Zhang, Qi Li, Yinyin Chen, and Jiawei Han. 2018. Open Information Extraction with Meta-pattern Discovery in Biomedical Literature. In *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB)*. 291–300.

[25] Wenhan Xiong, Thien Hoang, and William Yang Wang. 2017. Deeppath: A Reinforcement Learning Method for Knowledge Graph Reasoning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 564–573.

[26] Mohamed Yahya, Steven Whang, Rahul Gupta, and Alon Halevy. 2014. Renoun: Fact extraction for nominal attributes. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 325–335.

[27] Bishan Yang and Tom Mitchell. 2017. Leveraging knowledge bases in lstms for improving machine reading. In *Proceedings of the 55rd Annual Meeting of the Association for Computational Linguistics (ACL)*. 1436–1446.

[28] Wenhao Yu, Zongze Li, Qingkai Zeng, and Meng Jiang. 2019. Tablepedia: Automating PDF Table Reading in an Experimental Evidence Exploration and Analytic System. In *The World Wide Web Conference*.

[29] Yuyu Zhang, Hanjun Dai, Zornitsa Kozareva, Alexander J Smola, and Le Song. 2018. Variational reasoning for question answering with knowledge graph. In *32nd AAAI Conference on Artificial Intelligence (AAAI)*. 6069–6076.

[30] Suncong Zheng, Feng Wang, Hongyun Bao, Yuexing Hao, Peng Zhou, and Bo Xu. 2017. Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme. In *Proceedings of the 55rd Annual Meeting of the Association for Computational Linguistics (ACL)*. 1227–1236.