Ling Luo

E-mail: lingluo0415@gmail.com

Website: https://lingluodlut.github.io/   *Github*: https://github.com/lingluodlut

National Institutes of Health (NIH), Bethesda, USA

Education

- **Dalian University of Technology**   **Dalian, China**
  Ph.D., College of Computer Science and Technology   *09/2014-09/2019*
  Research Interests: Natural Language Processing, Biomedical Text Mining, Deep Learning

- **Xiamen University**   **Xiamen, China**
  M.S., School of Information Science and Engineering   *09/2011-06/2014*
  Research Interests: Natural Language Processing, Machine Translation

- **Xiamen University**   **Xiamen, China**
  B.S., School of Information Science and Engineering   *09/2007-06/2011*

Experience

- **National Institutes of Health (NIH)**   **Bethesda, USA**
  Postdoc Fellow, National Center for Biotechnology Information (NCBI)   *01/2020-present*

Scientific Research

- **Biomedical Text Classification**
  Automatic biomedical text classification from the exponentially growing biomedical literature is a fundamental step for large-scale biological database curation and has received much attention. The main contributions of my works can be summarized as follows:

  1) I presented a neural network ensemble approach to biomedical text classification. In this approach, several neural network models (i.e. LSTM (long-short term memory), CNN (convolutional neural network), LSTM-CNN, recurrent CNN and hierarchical LSTM) are used for document triage. Then the ensemble model is built by combining five models' results with a logistic regression classification to further improve the performance. I participated in the BioCreative VI precision medicine (PM) track, and our approach achieves an F-score of 69.04% in the BioCreative VI PM document triage task (The Second Place in 10 Teams).

  2) Recent years, biomedical text classification with neural networks has gained increasing attention, but domain knowledge has been rarely used in these methods. Aiming to exploit domain knowledge, I proposed a domain Knowledge-enriched Self-Attention Convolutional Neural Network (KeSACNN) approach to classify biomedical articles. In this approach, two

knowledge embeddings are proposed, and self-attention mechanism is used to directly capture the relationships of these feature embeddings. The experimental results show that KeSACNN can achieve better performance than the state-of-the-art methods on the BioCreative II and III Protein-Protein Interaction Article Classification corpora.

- **Biomedical Named Entity Recognition**

  Biomedical named entity recognition (BioNER) is an important task in the field of biomedical information extraction. Most popular BioNER methods are based on traditional machine learning and their performances are heavily dependent on the feature engineering. Recently, deep learning methods have been proposed and exhibit promising results for BioNER. The main contributions of my works can be summarized as follows:

  1) I presented a neural network approach to chemical and gene/protein entity recognition in patents. In the approach, a bidirectional long short-term memory with a conditional random field layer is employed to recognize biomedical entities from patents. To improve the performance, the effect of additional features (i.e., part of speech, chunking and named entity recognition features generated by the GENIA tagger) was explored for the neural network model. I participated in the BioCreative V.5 CEMP (chemical entity mention recognition) and GPRO (gene and protein related object recognition) tracks. In the official results, our best runs achieve the highest performances among all participating teams in both tracks.

  2) Most standard traditional machine learning methods and deep learning methods in practice treat NER as a sentence-level task. However, these sentence-level NER methods lead to the tagging inconsistency problem (i.e. the same mentions in a document are tagged with different labels). To mitigate the issues, I proposed a novel attention-based bidirectional Long Short-Term Memory with a conditional random field layer (Att-BiLSTM-CRF) approach for document-level biomedical NER. In this approach, Att-BiLSTM-CRF relies on a novel attention mechanism to capture similar entity attention at the document-level. This allows it to view the related tokens in different sentences of a document as a dependent tagging problem. Moreover, four different alternatives are considered to compute the score of attention matrix. The experimental results show that Att-BiLSTM-CRF achieves the state-of-the-art performances for BioNER on the BioCreative IV CHEMDNER corpus and the BioCreative V chemical-disease relation corpus.

- **Biomedical Relation Extraction**

  Biomedical relation extraction aims to uncover high-quality relations from biomedical literature with high accuracy and efficiency. Recently, my works focus on the biomedical relation extraction. Now, the manuscripts of these works are being prepared for submission. The main explorations of my works can be summarized as follows:

  1) Recently joint modeling methods of entity and relation exhibit more promising results than traditional pipelined methods in general domain. However, they are inappropriate to the biomedical domain due to numerous overlapping relations in biomedical text. To alleviate the problem, I proposed a neural network-based joint learning approach for biomedical entity and relation extraction. In this approach, a novel tagging scheme that takes into account overlapping relations is proposed to convert the joint extraction task to a tagging problem. Then an end-to-end neural network model is built to extract the entities and their relations directly with my extraction rules. Moreover, the contextualized ELMo representations

pre-trained on biomedical text are used to further improve the performance. Experimental results on biomedical corpora show that the method can significantly improve the performance of overlapping relation extraction and achieves the state-of-the-art performance.

2) The document-level biomedical relation extraction has received much attention. Most popular methods extract intra- and inter-sentential relations by building two relation classifiers, and merge them together to generate the document-level relations. However, these methods extract relations from single sentences or adjacent sentences without document-level information. Moreover, they suffer a lot from the problem of data noise. To address the above issues, I proposed a novel multi-instance learning based approach with multi-head attention for document-level biomedical relation extraction.

Participations in Research Projects

- NATIONAL LIBRARY OF MEDICINE (No. 1ZIALM091813-08): Named Entity Recognition and Relationship Extraction in Biomedicine. 2020-

- National Key Research and Development Program of China (No. 2016YFC0901902): Construction of Knowledgebase of Precision Medicine for Disease Studies. 2016-2020

- Natural Science Foundation of China (No. 61272373): Research on Implicit Knowledge Discover from Biomedical Literature. 2013-2016

Publications

1. **Ling Luo,** Shankai Yan, Po-Ting Lai, Daniel Veltri, Andrew Oler, Sandhya Xirasagar, Rajarshi Ghosh, Morgan Similuk, Peter N. Robinson, Zhiyong Lu. *PhenoTagger: A Hybrid Method for Phenotype Concept Recognition using Human Phenotype Ontology* [J]. Bioinformatics, 2021, 37(13): 1884-1890.

2. **Ling Luo,** Zhihao Yang, Mingyu Cao, Lei Wang, Yin Zhang, Hongfei Lin. *A neural network-based joint learning approach for biomedical entity and relation extraction from biomedical literature* [J]. Journal of Biomedical Informatics, 2020, 103: 103384.

3. **Ling Luo,** Zhihao Yang, Lei Wang, Yin Zhang, Hongfei Lin and Jian Wang. *KeSACNN: a protein-protein interaction article classification approach based on deep neural network* [J]. International Journal of Data Mining and Bioinformatics, 2019, 22(2): 131-148.

4. **Ling Luo**, Zhihao Yang, Pei Yang, Yin Zhang, Lei Wang, Hongfei Lin and Jian Wang. *An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition* [J]. Bioinformatics, 2018, 34(8): 1381-1388.

5. **Ling Luo**, Zhihao Yang, Pei Yang, Yin Zhang, Lei Wang, Jian Wang and Hongfei Lin. *A neural network approach to chemical and gene/protein entity recognition in patents* [J]. Journal of Cheminformatics, 2018, 10: 65.

6. **Ling Luo**, Zhihao Yang, Hongfei Lin and Jian Wang. *Document triage for identifying protein–*

protein interactions affected by mutations: a neural network ensemble approach [J]. Database-The Journal of Biological Databases and Curation, 2018, 2018(1): bay097.

7. **罗凌**, 杨志豪, 宋雅文, 李楠, 林鸿飞. *基于笔画 ELMo 和多任务学习的中文电子病历命名实体识别研究* [J]. 计算机学报, 2020, 43(10): 1943-1957.

8. **罗凌**, 陈毅东, 史晓东, 苏劲松. *基于复述技术的汉语成语翻译方法研究* [J]. 中文信息学报, 2015, 29(04): 166-174.

9. **Ling Luo**, Shankai Yan, Po-Ting Lai, Daniel Veltri, Andrew Oler, Sandhya Xirasagar, Rajarshi Ghosh, Morgan Similuk, Peter N. Robinson, Zhiyong Lu. *A Hybrid Method for Phenotype Concept Recognition using the Human Phenotype Ontology* [C]. Proceeding of 28th Conference on Intelligent Systems for Molecular Biology (ISMB 2020), Abstracts (oral and poster), July 2020, online.

10. **Ling Luo**, Zhihao Yang, Lei Wang, Yin Zhang, Hongfei Lin, Jian Wang, Liang Yang, Kan Xu and Yijia Zhang. *Protein-Protein Interaction Article Classification: A Knowledge-enriched Self-Attention Convolutional Neural Network Approach* [C]. Proceeding of 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2018, Regular Paper.

11. **Ling Luo**, Nan Li, Shuaichi Li, Zhihao Yang and Hongfei Lin. *DUTIR at the CCKS-2018 Task1: A Neural Network Ensemble Approach for Chinese Clinical Named Entity Recognition* [C]. Proceedings of the Evaluation Tasks at the China Conference on Knowledge Graph and Semantic Computing (CCKS-Tasks 2018), 2018.

12. **Ling Luo**, Zhihao Yang, Hongfei Lin and Jian Wang. *DUTIR at the BioCreative VI Precision Medicine Track: document triage for identifying PPIs affected by genetic mutations* [C]. Proceedings of the BioCreative VI Workshop, 2017: 120-123.

13. **Ling Luo**, Pei Yang, Zhihao Yang, Hongfei Lin and Jian Wang. *DUTIR at the BioCreative V.5.BeCalm Tasks: A BLSTM-CRF Approach for Biomedical Entity Recognition in Patents* [C]. Proceedings of the BioCreative V.5 Challenge Evaluation Workshop, 2017: 28-39.

14. **罗凌**, 杨志豪, 赵哲焕, 林鸿飞, 王健. *融合领域知识的深度神经网络生物医学文本分类* [C]. 第二十二届全国信息检索学术会议(CCIR 2016), 2016.

15. Qingyu Chen, Robert Leaman, Alexis Allot, **Ling Luo**, Chih-Hsuan Wei, Shankai Yan, Zhiyong Lu. *Artificial Intelligence in Action: Addressing the COVID-19 Pandemic with Natural Language Processing* [J]. Annual Review of Biomedical Data Science, 2021, Vol. 4: 313-339.

16. Alexis Allot, Kyubum Lee, Qingyu Chen, **Ling Luo,** Zhiyong Lu. *LitSuggest: A Web-based System for Literature Recommendation and Curation using Machine Learning* [J]. Nucleic Acids Research, 2021, 49(W1): W352–W358.

17. Zhiheng Li**,** Zhihao Yang, **Ling Luo**, Yang Xiang and Hongfei Lin. *Exploiting Adversarial Transfer Learning for Adverse Drug Reaction Detection from Texts* [J]. Journal of Biomedical Informatics, 2020, 106:103431.

18. Nan Li**,** Zhihao Yang, **Ling Luo**, Lei Wang, Yin Zhang, Hongfei Lin and Jian Wang. *KGHC: a knowledge graph for hepatocellular carcinoma* [J]. BMC Medical Informatics and Decision

Making, 2020, 20:135.

19. Zhiheng Li, Zhihao Yang, Yang Xiang, **Ling Luo**, Yuanyuan Sun, and Hongfei Lin. *Exploiting sequence labeling framework to extract document-level relations from biomedical texts* [J]. BMC bioinformatics, 2020, 21: 125.

20. Qingqing Li, Zhihao Yang, Zhehuan Zhao, **Ling Luo**, Zhiheng Li, Lei Wang, Yin Zhang, Hongfei Lin, Jian Wang, and Yijia Zhang. *HMNPPID—human malignant neoplasm protein–protein interaction database* [J]. Human genomics, 2019, 13: 44.

21. Wei Zheng, Hongfei Lin, **Ling Luo**, Zhehuan Zhao, Zhengguang Li, Yijia Zhang, Zhihao Yang and Jian Wang. *An attention-based effective neural model for drug-drug interactions extraction* [J]. BMC Bioinformatic, 2017, 18:445.

22. Zhehuan Zhao, Zhihao Yang, **Ling Luo**, Lei Wang, Yin Zhang, Hongfei Lin and Jian Wang. *Disease named entity recognition from biomedical literature using a novel convolutional neural network* [J]. BMC Medical Genomics, 2017, 10(5):73.

23. Zhehuan Zhao, Zhihao Yang, **Ling Luo**, Hongfei Lin and Jian Wang. *Drug drug interaction extraction from biomedical literature using syntax convolutional neural network* [J]. Bioinformatics, 2016, 32(22): 3444–3453.

24. Qingqing Li, Zhihao Yang, **Ling Luo**, Lei Wang, Yin Zhang, Hongfei Lin, Jian Wang, Liang Yang, Kan Xu and Yijia Zhang. *A multi-task learning based approach to biomedical entity relation extraction* [C]. Proceeding of 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2018: 680-682.

25. Zhehuan Zhao, Zhihao Yang, **Ling Luo**, Hongfei Lin, Jian Wang and Song Gao. *Deep neural network based protein-protein interaction extraction from biomedical literature* [C]. Proceeding of 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2015: 1156-1156.

26. 宋雅文, 杨志豪, **罗凌**, 王磊, 张音, 林鸿飞, 王健. *基于字符卷积神经网络的生物医学变异实体识别方法* [J]. 中文信息学报, 2021, 35(5): 63-69.

27. 丁泽源, 杨志豪, **罗凌**, 王磊, 张音, 林鸿飞, 王健. *基于深度学习的中文生物医学实体关系抽取系统* [J]. 中文信息学报, 2021, 35(5): 70-76.

Challenges

- The 2019 China Conference on Knowledge Graph and Semantic Computing (CCKS 2019) Challenge: Chinese Clinical Named Entity Recognition Task, The Third Place in 44 Teams   *07/2019*

- The 2018 China Conference on Knowledge Graph and Semantic Computing (CCKS 2018) Challenge: Chinese Clinical Named Entity Recognition Task, The Third Place in 69 Teams   *07/2018*

- BioCrative VI Precision Medicine Track: Document Triage Task, The   *10/2017*

Second Place in 10 Teams

- BioCreative V.5 Challenge: The CEMP (Chemical Entity Mention in Patents) Task, The First Place in 14 Teams *02/2017*

- BioCreative V.5 Challenge: The GPRO (Gene and Protein Related Object) Task, The First Place in 7 Teams *02/2017*

Awards & Honors

- IEEE BIBM 2018 Student Travel Award *12/2018*

- Best Student Paper Award, 22rd China Conference on Information Retrieval (CCIR 2016) *11/2016*

- Merit Student, Xiamen University *12/2013, 06/2011, 11/2008*

Academic Service

- Conference Reviewer: ACL(2021), EACL(2021), EMNLP(2020-2021), AACL(2020), ICHI (2020-2021), ISMB(2020)

- Journal Reviewer: Bioinformatics, JAMIA, Journal of Biomedical Informatics, BMC Bioinformatics, Bioinformatics Advances, Artificial Intelligence Review, PLOS ONE, IEEE Access, IJDMB,

Languages and Technologies

- Programming Languages: Python, Java, C
- Technologies: NumPy, Keras, scikit-learn, CoreNLP, Weka
- Natural Languages: Fluent in Chinese, advanced in English