RESEARCH STATEMENT

Wei Lu

Introduction

My long-term research goal is to provide a systematic and unified understanding of various natural language processing (NLP) problems, uncovering the underlying connections between different tasks and models proposed for tackling them. I believe that pursuing a unified approach is not only academically fulfilling due to its elegance but also offers practical convenience in identifying suitable solutions to various NLP challenges. To achieve this, I have focused my research on developing principled mathematical and algorithmic solutions to fundamental NLP problems, with a specific emphasis on *natural language semantics*.

Specifically, my contributions have been concentrated in the following major areas:

1. ***Word-level semantics***: We proposed several principled mathematical approaches for learning word and graph representations, leveraging network-level, node/word-level, and cross-domain information.
2. ***Entity-level semantics***: We developed efficient models and algorithms based on novel hypergraph designs capable of recognizing complex entities, such as nested entities.
3. ***Sentence-level semantics***: We created robust algorithms for semantic parsing that work across different languages and proposed models for language generation from semantic representations in the form of trees, forests, and graphs.
4. ***Document-level semantics***: We addressed fundamental and practical issues related to using neural architectures for modeling semantics within text at the document level.
5. ***Feature-level semantics***: We established new understandings on how neural networks capture semantically meaningful features through mathematical analyses.

To advance our goal of providing a unified framework for natural language processing, and recognizing that modern language models offer a potential unified solution to NLP problems, we have also pioneered efforts in developing and researching ***small language models***. This is a step towards better understanding how modern language models capture natural language semantics.

Background

Many of the problems within the field of NLP, such as chunking, tagging, word segmentation, named entity recognition, relation extraction, constituency parsing, dependency parsing, semantic parsing, and fine-grained sentiment analysis, are essentially *structured prediction* problems. These problems involve predicting a structured output from a given input. Recognizing this commonality motivated me to pursue a unified approach to solving a variety of NLP challenges.

Over years of research focused on designing models for different structured prediction problems, I concluded that it was indeed possible to create a unified approach for addressing a wide range of these tasks. Specifically, in 2014, I discovered connections between different structured prediction models and developed a unified framework based on a specific type of directed hypergraphs. This novel framework integrates various existing structured prediction models, including classic ones such as linear-chain conditional random fields, structural support vector machines, and their latent variable, semi-Markov, and neural variants. It also encompasses tree-structured prediction models, such as *parsing conditional random fields* for syntactic tree structures and the latent variable variant we proposed called *hybrid trees* [Lu et al., 2008, Lu, 2014] for semantic parsing. Our research group has built various models for different applications based on this framework over the years [Lu, 2014, Lu and Roth, 2015, Muis and Lu, 2016b,a, Li and Lu, 2017, Jie et al., 2017, Susanto and Lu, 2017b, Zou and Lu, 2018, 2019, Li and Lu, 2019, Xu et al., 2020].

In 2017, we introduced our hypergraph-based framework for structured prediction to the community through a tutorial at EMNLP [Lu, 2017]. This framework forms the foundation for much of our work, which can be categorized into five major sub-themes under the overarching theme of *natural language semantics*: semantics at the word, entity, sentence, document, and feature levels.

With the recent development of language models as a promising unified tool for solving NLP problems, we have shifted our research priorities in this emerging direction. Over the past 1-2 years, we have embarked on a journey to develop our own language models from scratch Zhang et al. [2024], as our first step towards systematically understanding their behaviors and fundamental properties. This endeavor aims to further our pursuit of a comprehensive understanding of NLP problems and the models that solve them, continuing our goal of building a unified approach centered on natural language semantics.

Research Contributions

Word-level Semantics

We proposed several models for learning representations of words and graphs (where the nodes of the graph can be words). One such model is the GraRep model [Cao et al., 2015], a principled general approach for learning graph representations. It learns a vector representation for each vertex in a weighted undirected graph. Compared to existing approaches, our model captures richer global structural information associated with the graph. Additionally, we mathematically demonstrated the connections between our approach and several previous classic approaches for learning graph and word representations. We also proposed another model based on neural networks for learning graph representations [Cao et al., 2016]. While these models focus on learning representations based on graph information, we subsequently explored the problem of exploiting sub-word level information for learning word representations. We designed a model that uses convolutional neural networks to exploit sub-word level information for learning word representations for English [Cao and Lu, 2017], and another model that exploits sub-word (stroke) level information for learning Chinese word representations [Cao et al., 2018]. Furthermore, we proposed models for learning word representations in a cross-domain setup [Yang et al., 2017], based on our regularization-based domain adaptation framework [Lu et al., 2016].

Based on Google Scholar, our GraRep paper [Cao et al., 2015] has been recognized as the **Most Cited CIKM Paper** among all papers published in CIKM over two consecutive five-year spans (2014-2019 and 2015-2020).[1]

Entity-level Semantics

We proposed a novel hypergraph representation for named entity or mention recognition [Lu and Roth, 2015]. This model was built based on our hypergraph-based structured prediction framework. To the best of our knowledge, it was the first hypergraph-based approach for this important information extraction task. One advantage of our approach is its ability to capture overlapping entity mentions. While previous research efforts on handling nested named entities exist, we have demonstrated that our approach is effective and offers lower inference time complexity – it is linear in the number of words and linear in the number of semantic types (such as ORG, LOC, etc.). We further proposed several novel models, including a model that can recognize discontinuous and overlapping entities [Muis and Lu, 2016a, Wang and Lu, 2019], a model based on a multigraph representation that labels gaps between words for extracting overlapping entities [Muis and Lu, 2017], a neural transition-based approach for efficiently recognizing nested entities, and another neural approach based on a novel segmental hypergraph representation for recognizing overlapping mentions [Wang and Lu, 2018]. Leading NLP labs from Cornell and CMU also investigated this task and made extensive comparisons with our models [Katiyar and Cardie, 2018, Shibuya and Hovy, 2020]. Specifically, the Cornell NLP group used a hypergraph representation similar to ours with "exactly the same" expressiveness.

We have also conducted other work related to entity recognition. Specifically, we introduced a neural approach to cross-domain named entity recognition, where we used neural adaptation layers to adapt an existing named entity model trained on one domain for a new domain [Lin and Lu, 2018]. Additionally, we investigated how to effectively utilize dependency information for improved named entity recognition. For example, within a semi-Markov conditional random fields framework, we performed provably efficient named entity recognition [Jie et al., 2017], and explored ways to incorporate dependency information into the recognition process involving an LSTM [Jie and Lu, 2019]. We also proposed new models for better handling of incomplete annotations for named entity recognition [Jie et al., 2019], and examined other fundamental problems such as relation extraction [Guo et al., 2019a, Nan et al., 2020] and joint entity and relation extraction [Wang and Lu, 2020]. Recently, we successfully established a connection between the task of math word problem solving and complex relation extraction Jie et al. [2022].

Together with Alibaba and other collaborators, we participated in an international competition on "Complex Named Entity Recognition" and achieved the best performance on 10 out of 13 tasks. The resulting paper [Wang et al., 2022] also received the **Best System Paper Award** at SemEval 2022.[2].

Sentence-level Semantics

Semantic parsing is the task of parsing a natural language sentence into its semantic representation. In 2008, we introduced the *hybrid tree* semantic parser to the community [Lu et al., 2008]. The parser assumes there exists a latent joint representation of words and semantics for a given sentence-semantics pair, allowing for efficient inference based on such a representation. The original parser was a generative model, but we made several extensions to enhance its effectiveness. Specifically, we first introduced the discriminative version of the hybrid tree parser [Lu, 2014]. Next, in 2017, we incorporated a neural component into the hybrid tree model, resulting in the neural hybrid tree semantic parser [Susanto and Lu, 2017b]. In 2018, we proposed an alternative assumption on the latent representation for

[1] See these links: `https://scholar.google.com/citations?hl=en&vq=eng_databasesinformationsystems&view_op=list_hcore&venue=V-IMg2OTpU8J.2019` (2014–2019), and `https://scholar.google.com/citations?hl=en&vq=eng_databasesinformationsystems&view_op=list_hcore&venue=V-IMg2OTpU8J.2020` (2015–2020).

[2] See this link: `https://semeval.github.io/SemEval2022/awards`

connecting words and semantics and developed a novel dependency-based hybrid tree model [Jie and Lu, 2018]. This model has proven to be more robust and particularly effective when dealing with languages such as Indonesian, and it currently achieves state-of-the-art results on benchmark datasets across several languages.

We also looked into the problem of parsing sentences into abstract meaning representations (AMR) using a transition-based model [Guo and Lu, 2018]. Though it employs a transition-based approach to structured prediction, the parser learns to construct a derivation process that maps natural language words to basic semantic units. This process, which is inspired by the approach adopted by our hybrid tree semantic parsers, also yields a joint representation that contains both the words and semantics. Besides monolingual semantic parsing, we also looked into the task of multilingual semantic parsing under various different setups [Susanto and Lu, 2017a, Zou and Lu, 2018].

We also worked on the natural language generation task, which can be regarded as the inverse of the semantic parsing task. We first proposed a natural language generation model by inverting our hybrid tree semantic parser [Lu et al., 2009]. We then extended the hybrid tree model to support forest-structured semantic representations. This extension allowed us to build a model capable of generating natural language sentences from typed lambda calculus expressions [Lu and Ng, 2011].

Additionally, we examined the problem of generating natural language sentences from graph-structured semantic representations (such as AMR and dependency structures) and proposed novel language generation models based on densely connected graph convolutional networks [Guo et al., 2019b]. This approach modifies the conventional graph convolutional neural network architecture used in the literature by introducing novel dense connections, which were shown to be effective in alleviating the issue of training graph convolutional neural networks with many layers. The model yielded significant improvements over previous models on the language generation task while requiring significantly fewer model parameters.

In 2011, the best paper committee of EMNLP, comprising leading researchers, unanimously selected our EMNLP 2011 paper [Lu and Ng, 2011] for the only **Best Paper Award** at that conference. This was also the first time that a best paper award at a top NLP conference (ACL, EMNLP, NAACL) was given to a paper from Asia.[3]

Document-level Semantics

While semantic parsing focuses on understanding the semantics of individual sentences, comprehending document-level semantics presents additional challenges. The encoder-decoder architecture has proven effective for handling document-level tasks such as translation and summarization, which involve capturing and modeling document-level semantics. However, these models have known limitations, one of which is *exposure bias*. This bias arises from the mismatch between different assumptions made during training and testing, and it is particularly significant in document-level translation involving long input and output sequences.

To address this issue, we first identified the limitations of the prior effort known as *scheduled sampling* and then proposed a new training approach that dynamically adapts the loss during the training process to better simulate test-time scenarios [Li and Lu, 2021]. Our analyses demonstrate that while scheduled sampling can simulate *test-time inputs*, our proposed approach goes a step further to better simulate *test-time behavior*, thereby alleviating the exposure bias issue more effectively. This work was selected for a **Long Oral Presentation** at ICML 2021, placing it in the top 3.0% of submissions.

Besides tackling the fundamental challenges with encoder-decoder models, we also examined the practical aspects of creating effective mechanisms for transforming large amounts of text into their semantic representations using such an architecture. Specifically, we explored the problem of grounding document meanings in visible semantic forms.

Leveraging our unique position as a design university, we focused on human-recognizable visual designs as our target semantic forms, making the end product particularly valuable for designers who transform descriptions into visual representations. However, input descriptions can be arbitrarily long and complex, with text interactions across sentence boundaries. Additional constraints, such as ensuring the final visual representation respects certain prior conditions (e.g., boundary information in floor plans), also exist. To tackle this, we proposed a novel way of utilizing a pre-trained encoder-decoder model, formulating visual designs using formal languages and achieving superior results compared to existing approaches from the vision community. Our work on generating floor plans from natural language descriptions [Leng et al., 2023] has shown a promising direction towards building robust document-level language understanding systems. Besides introducing a new task to our field, this work also contributed a novel dataset, enriching research opportunities in document-level semantics.

Together with our collaborators, including a design researcher, our work on design generation from document-level text instructions [Leng et al., 2023] received the **Area Chair's Award** at ACL 2023 within the "Resources & Evaluation" area. Additionally, it received other Best Paper/Outstanding Paper nominations.[4]

[3]See this link: `https://aclweb.org/aclwiki/Best_paper_awards`
[4]See this link: `https://2023.aclweb.org/program/best_papers/`

Feature-level Semantics

We have conducted a series of studies focused on model explainability, particularly concerning neural network and deep learning models for NLP. Our goal is to understand the internal mechanisms of neural models at a low level (e.g., feature level) and offer semantically meaningful interpretations. In our first study, we specifically examined how the attention mechanism works in text classification Sun and Lu [2020]. Unlike some existing efforts, we analyzed the behavior of the attention mechanism mathematically. We found that to understand the significance of each word token as captured by the attention mechanism, we should examine the global *attention score* rather than the commonly examined local *attention weight*. We demonstrated why attention may become less "interpretable" in certain scenarios and investigated why the model may still perform well despite this. Specifically, we showed how the *attention score* and another quantity, the *polarity score*, jointly impact overall classification performance.

Next, we sought ways to interpret recurrent neural networks (RNNs) Sun and Lu [2022]. Through mathematical analyses, we discovered salient components within RNN cells that can capture information interpretable as classic n-gram information. We empirically verified the significance of such information through sequence modeling tasks. This reveals the connection between non-Markovian sequential models, such as RNNs, and Markovian models, such as n-gram models, suggesting that the former might be able to capture semantically interpretable features that are well understood and well captured by the latter, more classical models.

Building on these efforts, we took a step further to investigate how different types of neural networks (e.g., convolutional networks, attention-based networks, and recurrent networks) capture semantically meaningful features through mathematical analyses, using neural tangent kernels as a tool Sun et al. [2023]. This work reveals how feature-level semantics are captured by different neural models, which we believe is a useful step towards not only understanding these models but also designing new ones. Additionally, this work provides explanations for why certain design choices, such as specific activation functions, are better suited for modern large language models.

Our most recent work on understanding how neural networks capture semantic features Sun et al. [2023] was recognized with an **Outstanding Paper Award** at EMNLP 2023.[5]

Small Language Models

As our goal has been to provide a unified understanding for NLP problems, we recognize that a unified approach can be achieved with modern language models. This has led us to shift our research priorities recently, motivating us to embark on the journey of building our own language models first, so as to better investigate them next.

We have made some pioneering contributions. Specifically, we introduced TinyLlama [Zhang et al., 2024], a small language model with 1.1 billion parameters that is pre-trained from scratch with around 3 trillion tokens. TinyLlama is the world's first known effort originating from a university research lab aimed at pre-training an effective small-scaled language model from scratch. We embarked on this as a live, open research project, which has excited the entire community and attracted significant worldwide attention. Empirically, TinyLlama outperforms OPT-1.3B [Zhang et al., 2022] and Pythia-1.4B [Biderman et al., 2023] on various commonsense reasoning and problem-solving benchmarks in a zero/few-shot setting.

Although our original goal in developing TinyLlama was to advance our group's research agenda, we believe its practical impact extends well beyond that. Scientifically, TinyLlama aims to provide in-depth knowledge about language models in an easily accessible way, offering a way for us and the community to democratize language model research. From a social perspective, we believe that a free, compact, eco-friendly, yet capable model like TinyLlama has the potential to make AI assistance more accessible to society. Its compact size makes it suitable for deployment on small devices, further broadening its accessibility and utility while prioritizing privacy and sustainability. By open-sourcing TinyLlama, we hope to promote transparency and enable further research into making language models more reliable, unbiased, and ethically aligned before wide-scale deployment.

Immediately after its release, TinyLlama [Zhang et al., 2024] achieved the **Global Top Spot**[6] on the trending list of models on Hugging Face (top 1 out of over 450,000 models) for about a week. To date, it has been downloaded over 1 million times, received over 7,000 GitHub stars, and inspired the creation of over 2,000 variants by the community.

Conclusion

In summary, my past 18 years of research have been dedicated to advancing the field of NLP by providing unified understandings of tasks and models through focused work on various aspects of natural language semantics. Throughout this ongoing journey, we have achieved significant milestones. As we embrace the rise of new paradigms, particularly language models, we strive to make pioneering advancements with a specific focus on small language models. Our original goal of establishing a unified understanding of NLP through natural language semantics remains central to our efforts. We eagerly anticipate further contributions to our field, making a significant impact during this exciting new era of language models.

[5] See this link: `https://2023.emnlp.org/program/best_papers/`

[6] See this link: `https://web.archive.org/web/20240105200138/https://huggingface.co/models`

References

Stella Biderman, Hailey Schoelkopf, Quentin Gregory Anthony, Herbie Bradley, Kyle O'Brien, Eric Hallahan, Mohammad Aflah Khan, Shivanshu Purohit, USVSN Sai Prashanth, Edward Raff, et al. Pythia: A suite for analyzing large language models across training and scaling. In *Proceedings of International Conference on Machine Learning (ICML)*, 2023.

Shaosheng Cao and Wei Lu. Improving word embeddings with convolutional feature learning and subword information. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, 2017.

Shaosheng Cao, Wei Lu, and Qiongkai Xu. GraRep: Learning graph representations with global structural information. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management (CIKM)*, 2015.
(Most-cited CIKM Paper (twice: 2014–2019, 2015–2020), according to Google Scholar).

Shaosheng Cao, Wei Lu, and Qiongkai Xu. Deep neural networks for learning graph representations. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, 2016.

Shaosheng Cao, Wei Lu, Jun Zhou, and Xiaolong Li. cw2vec: Learning chinese word embeddings with stroke n-gram information. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, 2018.

Zhijiang Guo and Wei Lu. Better transition-based AMR parsing with a refined search space. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018.

Zhijiang Guo, Yan Zhang, and Wei Lu. Attention guided graph convolutional networks for relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019a.

Zhijiang Guo, Yan Zhang, Zhiyang Teng, and Wei Lu. Densely connected graph convolutional networks for graph-to-sequence learning. *Transactions of the Association for Computational Linguistics (TACL)*, 7, 2019b.

Zhanming Jie and Wei Lu. Dependency-based hybrid trees for semantic parsing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018.

Zhanming Jie and Wei Lu. Dependency-guided LSTM-CRF for named entity recognition. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019.

Zhanming Jie, Aldrian Obaja Muis, and Wei Lu. Efficient dependency-guided named entity recognition. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, 2017.

Zhanming Jie, Pengjun Xie, Wei Lu, Ruixue Ding, and Linlin Li. Better modeling of incomplete annotations for named entity recognition. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, 2019.

Zhanming Jie, Jierui Li, and Wei Lu. Learning to reason deductively: Math word problem solving as complex relation extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2022.

Arzoo Katiyar and Claire Cardie. Nested named entity recognition revisited. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, 2018.

Sicong Leng, Yang Zhou, Mohammed Haroon Dupty, Wee Sun Lee, Sam Conrad Joyce, and Wei Lu. Tell2design: A dataset for language-guided floor plan generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2023.
(Area Chair Award: Top 0.42%=1/236 within "Resources & Evaluation" area).

Hao Li and Wei Lu. Learning latent sentiment scopes for entity-level sentiment analysis. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, 2017.

Hao Li and Wei Lu. Learning explicit and implicit structures for targeted sentiment analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019.

Haoran Li and Wei Lu. Mixed cross entropy loss for neural machine translation. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2021.
(Long oral presentation: Top 3.0%=166/5513).

Bill Yuchen Lin and Wei Lu. Neural adaptation layers for cross-domain named entity recognition. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018.

Wei Lu. Semantic parsing with relaxed hybrid trees. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014.

Wei Lu. A unified framework for structured prediction: From theory to practice. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP): Tutorial Abstracts*, 2017.

Wei Lu and Hwee Tou Ng. A probabilistic forest-to-string model for language generation from typed lambda calculus expressions. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2011.
(Best Paper Award: Top 0.16%=1/628).

Wei Lu and Dan Roth. Joint mention extraction and classification with mention hypergraphs. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2015.

Wei Lu, Hwee Tou Ng, Wee Sun Lee, and Luke Zettlemoyer. A generative model for parsing natural language to meaning representations. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2008.

Wei Lu, Hwee Tou Ng, and Wee Sun Lee. Natural language generation with tree conditional random fields. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2009.

Wei Lu, Hai Leong Chieu, and Jonathan Löfgren. A general regularization framework for domain adaptation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2016. short paper.

Aldrian Obaja Muis and Wei Lu. Learning to recognize discontiguous entities. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2016a.

Aldrian Obaja Muis and Wei Lu. Weak semi-Markov CRFs for noun phrase chunking in informal text. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, 2016b. short paper.

Aldrian Obaja Muis and Wei Lu. Labeling gaps between words: Recognizing overlapping mentions with mention separators. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2017.

Guoshun Nan, Zhijiang Guo, Ivan Sekulic, and Wei Lu. Reasoning with latent structure refinement for document-level relation extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020.

Takashi Shibuya and Eduard Hovy. Nested named entity recognition via second-best sequence learning and decoding. *Transactions of the Association for Computational Linguistics*, 8:605–620, 2020.

Xiaobing Sun and Wei Lu. Understanding attention for text classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020.

Xiaobing Sun and Wei Lu. Implicit *n*-grams induced by recurrence. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, 2022.

Xiaobing Sun, Jiaxi Li, and Wei Lu. Unraveling feature extraction mechanisms in neural networks. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023.
(Outstanding Paper Award: Top 0.59%=29/4909).

Raymond Hendy Susanto and Wei Lu. Neural architectures for multilingual semantic parsing. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2017a. short paper.

Raymond Hendy Susanto and Wei Lu. Semantic parsing with neural hybrid trees. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, 2017b.

Bailin Wang and Wei Lu. Neural segmental hypergraphs for overlapping mention recognition. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018.

Bailin Wang and Wei Lu. Combining spans into entities: A neural two-stage approach for recognizing discontiguous entities. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019.

Jue Wang and Wei Lu. Two are better than one: Joint entity and relation extraction with table-sequence encoders. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020.

Xinyu Wang, Yongliang Shen, Jiong Cai, Tao Wang, Xiaobin Wang, Pengjun Xie, Fei Huang, Weiming Lu, Yueting Zhuang, Kewei Tu, Wei Lu, and Yong Jiang. DAMO-NLP at SemEval-2022 task 11: A knowledge-based system for multilingual named entity recognition. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval)*, 2022.
(Best System Paper Award: Top 0.45%=1/221).

Lu Xu, Hao Li, Wei Lu, and Lidong Bing. Position-aware tagging for aspect sentiment triplet extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020.

Wei Yang, Wei Lu, and Vincent Zheng. A simple regularization-based algorithm for learning cross-domain word embeddings. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2017. short paper.

Peiyuan Zhang, Guangtao Zeng, Tianduo Wang, and Wei Lu. Tinyllama: An open-source small language model. *arXiv preprint arXiv:2401.02385*, 2024.
(TinyLlama was globally trending #1 (out of over 450K models) on Hugging Face for about 1 week after its release.).

Susan Zhang, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan, Mona Diab, Xian Li, Xi Victoria Lin, et al. Opt: Open pre-trained transformer language models. *arXiv preprint arXiv:2205.01068*, 2022.

Yanyan Zou and Wei Lu. Learning cross-lingual distributed logical representations for semantic parsing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2018. short paper.

Yanyan Zou and Wei Lu. Text2Math: End-to-end parsing text into math expressions. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019.