

# Linfeng SONG

---

CURRENT POSITIONS	PhD Student	University of Rochester
CONTACT INFORMATION	2403 Wegmans Hall, Univ. of Rochester Rochester, NY, 14626	lsong10@cs.rochester.edu <a href="http://cs.rochester.edu/u/lsong10">http://cs.rochester.edu/u/lsong10</a> Cell: (585) 622-8056
EDUCATION	Computer Science Department, University of Rochester, New York, America	
	Ph.D., Computer Science (graduating in May 2019)	<i>Sep. 2014 - present</i>
	Advisor: Prof. Daniel Gildea	
	Institute of Computing Technology, Chinese Academy of Science, Beijing, China	
	M. S., Computer Science	<i>Sep. 2010 - Jul. 2014</i>
	Advisor: Prof. Qun Liu	
	Northeastern University, Shenyang, China	
	B. Eng., Software Engineering	<i>Sep. 2006 - Jul. 2010</i>
	Senior member of university ACM/ICPC club	
INTERNSHIP EXPERIENCES	IBM Thomas J. Watson Research Center	<i>June 2018 - August 2018</i>
	Advised by Zhiguo Wang, Mo Yu and Radu Florian	
	We studied the problem of multi-hop question answering (MHQA). Comparing with SQuAD, where local information is enough for answering questions, MHQA requires integrating multiple pieces of evidence to correctly answer a question. We introduced a graph-based approach, where a graph is first created to connect related evidence, then evidence integration is performed on the graph using our recent graph recurrent network. This work is available at <a href="https://arxiv.org/abs/1809.02040">https://arxiv.org/abs/1809.02040</a> .	
	IBM Thomas J. Watson Research Center	<i>May 2017 - July 2017</i>
	Advised by Zhiguo Wang and Wael Hamza	
	We cast both tasks of natural question generation (NQG) and generative question answering (QA) into one unified process, where a query $\mathbf{q}$ is first matched against a passage $\mathbf{p}$ before generating the output $\mathbf{o}$ word-by-word. To tackle the tasks, we introduced a model following the encoder-decoder framework. The encoder adopts multi-perspective context matching (MPCM) to match $\mathbf{q}$ against $\mathbf{p}$ , and the decoder is an LSTM with copy mechanism generating $\mathbf{o}$ . A policy gradient reinforcement learning algorithm is adopted for tackling the exposure bias problem, resulted from sequence learning with cross-entropy loss. A preprint version of our paper is available at <a href="https://arxiv.org/abs/1709.01058">https://arxiv.org/abs/1709.01058</a>	
	Bosch American Research Center, Palo Alto, CA	<i>May 2016 - July 2016</i>
	Advised by Lin Zhao	
	The project is about natural question generation and parsing: 1) The generation task is to generate questions from a domain-specific knowledge base (KB). Since there is	

no annotated data (such as human generated questions for triples within the KB) for conducting supervised training, we tried to develop an approach that requires minimal human effort. In particular, a small-scale template is first handcrafted based on each predicate in the KB, then a set of initial questions is generated by applying the template on the KB, finally the initial question set is augmented by the related queries on the Web. 2) The parsing task is basically named entity recognition and linking for questions. For the entity recognition part a sequential labeler is developed, and for the entity linking part a collective entity linker is implemented.

IBM Thomas J. Watson Research Center

May 2015 - August 2015

Advised by Zhiguo Wang, Haitao Mi and Abraham Ittycheriah

It is about “dynamic” word embedding, where word embeddings are dynamically calculated based on its context. Our approach produces vectors of high variety and does not outperform our baseline on the widely-used SCWS benchmark.

Singapore University of Design and Technology

November 2013 - February 2014

Advised by professor Yue Zhang

We extends the partial tree linearization model of Zhang (2013) by jointly learning both linearization and morphological realization. This work has been published at AAAI 2014. In another project, we model machine translation with target syntax as a syntactic linearization problem. My contribution is developing a phrase extractor and running experiments. This paper is has been published at EMNLP 2014.

Alibaba Inc. Hangzhou, China

December 2011 - June 2012

Advised by Feng Lin

Developed an effective implementation of the “Maximum entropy based phrase re-ordering model for statistical machine translation” for Alibaba.

#### RECENT PUBLICATIONS

- Linfeng Song, Yue Zhang, Zhiguo Wang and Daniel Gildea. 2018. *N*-ary Relation Extraction using Graph State LSTM. To appear in the *Proceedings of EMNLP 2018*.
- Linfeng Song, Yue Zhang, Zhiguo Wang and Daniel Gildea. 2018. A Graph-to-Sequence Model for AMR-to-Text Generation. In the *Proceedings of ACL 2018*.
- Yue Zhang, Qi Liu and Linfeng Song. 2018. Sentence-State LSTM for Text Representation In the *Proceedings of ACL 2018*.
- Xiaochang Peng, Linfeng Song, Daniel Gildea, Giorgio Satta and Hao Zhang. 2018. Sequence-to-sequence Models for Cache Transition Systems. In the *Proceedings of ACL 2018*.
- Linfeng Song, Zhiguo Wang, Wael Hamza, Yue Zhang and Daniel Gildea. Leveraging Context Information for Natural Question Generation. In the *Proceedings of NAACL 2018*.
- Linfeng Song, Xiaochang Peng, Yue Zhang, Zhiguo Wang and Daniel Gildea. 2017. AMR-to-text Generation with Synchronous Node Replacement Grammar. In *Proceedings of ACL 2017*.
- Linfeng Song, Yue Zhang, Xiaochang Peng, Zhiguo Wang and Daniel Gildea. 2016. AMR-to-text generation as a Traveling Salesman Problem. In *Proceedings of EMNLP 2016*.

- Xiaochang Peng, Linfeng Song and Daniel Gildea. 2015. A Synchronous Hyperedge Replacement Grammar based approach for AMR parsing. In *Proceedings of CoNLL 2015*, Beijing, China,
- Linfeng Song, Yue Zhang, Kai Song and Qun Liu. 2014. Joint Morphological Generation and Syntactic Linearization. In *Proceedings of AAAI 2014*, Quebec City, Canada, July 27-31, 2014.
- Yue Zhang, Kai Song, Linfeng Song, Jingbo Zhu and Qun Liu. 2014. Syntactic SMT Using a Discriminative Text Generation Model. In *Proceedings of EMNLP 2014*, Doha, Qatar, 2014.

SELECTED AWARDS

- Huawei Scholarship for Graduate Student, *2012*
- Second Prize of ACM/ICPC China Northeastern Programming Contest, *2008*

*Last updated on September 17th, 2018.*