Tong Wang

Contact Information

100 West Squantum St, #215
Quincy, MA
02171 USA

Tel: (857) 203-1856
E-mail: tongwang0001@gmail.com
Homepage: `https://tongwang-umb.github.io/`

Interests

Deep Learning, Natural Language Processing, Machine Learning, Topic Models

Education

University of Massachusetts Boston, Boston, MA USA — Jan 2014 - Jan 2018

Ph.D., Computer Science, *Advisor: Dr. Ping Chen*
Thesis: Semantic Representation and Interpretation of Short Texts with Deep Learning

Northeastern University, Boston, MA USA — Sep 2011 - Jan 2013

M.S., Computer Systems Engineering

Huazhong Agricultural University, Wuhan, Hubei China — Sep 2006 - Jan 2010

B.S., Information and Computing Science

Work Experience

Amazon, Cambridge, MA — Feb 2018 - Current
Applied Scientist, Alexa AI-Natural Understanding

- Designed query understanding and entity ranking model using pre-trained BERT and entity graph embedding for implicit reference utterances in Alexa video domain.
- Built personalized Entity Resolution ranking model in Alexa music domain, which could help to select the most relevant entity based on user preference features.
- Incorporated entity features into natural language understanding models to improve interpretation ranking performance.

Disney Research, Pittsburgh, PA — May 2016 - Aug 2016
Lab Research Associate, Mentor: Dr. Albert Li

- Built a dataset for the narrative quality evaluation task by extracting stories and the upvotes from a social media website, Quora.
- Proposed several deep neural networks based on Attention CNN-LSTM that model the textual chunks in a story and their interrelations, which outperforms several strong baselines.

Shriver Center, UMass Medical School, Charlestown, MA USA — Jun 2015 - Aug 2015
Research Intern, Mentor: John Rochford

- Implemented lexical simplification system to replace complex words with their simpler synonyms, which shows a higher correlation with human assessment.

ioMosaic, Salem, NH USA — Jan 2013 - Dec 2013
Software Engineer

- Developed and contributed to the web application ioXpress using C# and ASP.NET.

Conference Publication

1. Ping Chen, Fei Wu and **Tong Wang**. A Semantic QA-Based Approach for Text Summarization Evaluation. 32nd AAAI Conference on Artificial Intelligence (AAAI 2018)
2. Jipeng Qiang, Yun Li, Yunhao Yuan, **Tong Wang**. Identifying the Number of Clusters in Short Text using Bayesian Nonparametric Model. The 29th IEEE International Conference on Tools with Artificial Intelligence, Boston, MA, USA, November 6-8, 2017. (ICTAI 2017)
3. Li B, Cardier B, **Wang T**, Metze F. Annotating High-Level Structures of Short Stories and Personal Anecdotes. The 11th Language Resources and Evaluation Conference (LREC). 2018.

4. **Tong Wang**, Ping Chen, Albert Li. Predicting the Quality of Short Narratives from Social Media. The 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia. (IJCAI 2017).
5. Jipeng Qiang, Ping Chen, **Tong Wang**, Xindong Wu. Topic Modeling over Short Texts by Incorporating Word Embeddings." The 21st Pacific-Asia Conference on Knowledge Discovery and Data Mining. (PAKDD 2017)
6. **Tong Wang**, Ping Chen, Kevin Amaral and Jipeng Qiang. An Experimental Study of LSTM Encoder-Decoder Model for Text Simplification. arXiv:1609.03663. (IJCAI-HLTIA 2016)
7. Jipeng Qiang, Ping Chen, Ding Wei, **Tong Wang**, Fei Xie, and Xindong Wu. Topic Discovery from Heterogeneous Texts, IEEE, The 28th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2016).
8. **Tong Wang**, Ping Chen, John Rochford and Jipeng Qiang. Text Simplification using Neural Machine Translation. Student Abstract. 30th AAAI Conference on Artificial Intelligence. (AAAI 2016)
9. **Tong Wang**, Vish Viswanath, and Ping Chen. Extended topic model for word dependency. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics. Vol. 2. 2015. (ACL 2015).

Journal Publication

1. Qiang, Jipeng, Ping Chen, Wei Ding, Tong Wang, Fei Xie, and Xindong Wu. "Heterogeneous-Length Text Topic Modeling for Reader-Aware Multi-Document Summarization." ACM Transactions on Knowledge Discovery from Data (TKDD) 13, no. 4 (2019): 42.
2. **Tong Wang**, Ping Chen and Dan Simovici. A New Evaluation Measure Using Compression Dissimilarity on Text Summarization. *Applied Intelligence* (2016): 1-8
3. Dan Simovici, Ping Chen, **Tong Wang** and Dan Pletea. Compression and Data Mining. *Journal of Communication*, 2015

Other Publication

1. **Tong Wang**, Han Wang, Feiyang Niu, Justin Flammia, Grace Deng, Thiago Mosqueiro, Huitian Lei, Bo Xiao, Yue Liu. Improving Search Relevance in Alexa Entity Resolution. *Amazon Machine Learning Conference* (AMLC 2019)
2. Thiago Mosqueiro, Huitian Lei, **Tong Wang**, Justin Flammia, Han Wang, Apoorva Balevalachilu, Yue Liu. Automated de-biasing for annotation-based component-independent metrics. *Amazon Machine Learning Conference* (AMLC 2019)

Service

Program Committee Member

- Empirical Methods in Natural Language Processing (EMNLP 2019)
- Annual Meeting of the Association for Computational Linguistics (ACL 2017, 2018, 2019)
- AAAI Conference on Artificial Intelligence (AAAI 2018)
- The International Conference on Computing, Networking and Communications (ICNC 2016, 2017, 2018)
- Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2016)
- The IEEE International Conference on Data Mining series (ICDM 2015) PhD Forum

Conference Reviewer

- AAAI Conference on Artificial Intelligence (AAAI 2018, 2019)
- Empirical Methods in Natural Language Processing (EMNLP 2019)
- Annual Meeting of the Association for Computational Linguistics (ACL 2017, 2018, 2019)

- The International Conference on Computing, Networking and Communications (ICNC 2016, 2017, 2018)
- SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2017)
- IEEE International Conference on Data Mining (ICDM 2016, 2017)
- International Conference on Connected Health: Applications, Systems and Engineering Technologies, 2017
- Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2016)
- The IEEE International Conference on Data Mining series (ICDM 2015) PhD Forum

Journal Reviewer

- Knowledge and Information Systems (KAIS)
- Social Network Analysis and Mining (SNAM)
- Communications of the Association for Information Systems (CAIS)
- IEEE Access

Talks

- NLU Reranker with ER features, Amazon NLU Conference — Dec 2019
- Entity-centric Language Understanding, Amazon Knowledge Conference — Oct 2019
- Science behind Alexa Entity Resolution, ML Seminar, Brandeis University. — Oct 2019
- Student Panelists: Text Simplification, Boston Accessibility Conference. — Oct 2017
- *Tricks from Deep Neural Networks*, CS697 Big Data Analytics, Umass Boston. — Nov 2016
- *Deep Learning in Natural Language Processing*, CS188SL-01 Science Gateway Seminar II, Umass Boston. — Apr 2016
- *Automated ICT Text Simplification for People with Cognitive Disability*, Boston Accessibility Conference. — Sep 2015

Honors and Awards

- Randall Malbone Scholarship Award, University of Masschusetts Boston (3%) — 2017
- Oracle Doctoral Research Fellowship Award — 2017
- 1st Grade Scholarship (3%), Huazhong Agricultural University — 2009
- National Scholarship (1%), Ministry of Education of China — 2007

Academic Projects

University of Massachusetts Boston, Boston, MA USA

Research Assistant — Jan 2014 - present

- Designed gated CNN-RNN neural network models for automatic narrative quality prediction, which achieves a good improvement over strong baselines. — Sep 2016 - May 2017
- Applied LSTM Encoder-Decoder model with global attention using English Wikipedia and Simple English Wikipedia for text simplification. — Jan 2016 - May 2016
- Applied topic model and opinion mining on social event "Chemical Spills in West Virginia". Proposed new topic model that could model interrelations between words, and incorporates with word embedding information. — Sep 2014 - May 2015
- Studied the effectiveness of compression algorithm on data mining and text summarization evaluation. — May 2014 - Sep 2015

Teaching Experience

Lecturer, CS310: Advanced Data Structures and Algorithms — May 2017 - July 2017

- Instructed a core course for students from Departments of Computer Science and Computer Engineering.
- Delivered three 90-minute presentations weekly, covering topics of basic data structures, as well as advanced algorithms.

Skills

- Programming Languages: Python, Java, C
- Machine Learning Tools: Pytorch, Keras, Theano, Tensorflow, Numpy, scikit-learn, NLTK, Stanford CoreNLP, OpenNMT
- Big Data: PySpark