

Education

- **Johns Hopkins University** *Aug. 2014 – present*
Ph.D. in Computer Science *Baltimore, MD, USA*
Center for Language and Speech Processing
Advisor: Prof. Benjamin Van Durme
- **Peking University** *Sep. 2010 – Jul. 2014*
B.Sc. in Computer Science *Beijing, China*
Department of Computer Science
Advisor: Prof. Junfeng Hu
Thesis: *Large-scale unsupervised word segmentation for Classical Chinese: Research & system.*

Research Interests

I design new methods to tackle various problems arisen from knowledge discovery in NLP: question answering, information extraction, entity search etc., with an emphasis on scalability to large corpora, while drawing inspirations from related fields such as information retrieval, machine learning, and computational semantics.

Research & Work Experience

- **Research Assistant (Doctoral research)** *Aug. 2014 – present*
Center for Language and Speech Processing, Johns Hopkins University, Baltimore, MD, USA
Conducted research on information extraction and question answering under the supervision of Prof. Benjamin Van Durme.
- **Applied Scientist Intern** *May. 2018 – Aug. 2018*
Amazon.com, Inc., Seattle, WA, USA
Conducted research on modeling in dialog systems under the supervision of Dr. Lambert Mathias in the Amazon Alexa group.
- **Research Intern** *May. 2017 – Aug. 2017*
IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA
Conducted research on question answering, model calibration and confidence scoring under the supervision of Dr. Jiří Navrátil, Dr. Vijay Iyengar, and Dr. Bing Xiang.
- **Research Assistant** *Jun. 2012 – Jul. 2014*
Institute of Computational Linguistics, Peking University, Beijing, China
Conducted research on Chinese word segmentation, named-entity recognition and social network analysis under the supervision of Prof. Junfeng Hu.

Publications

PREPRINTS

- [P1] **Tongfei Chen**, Jiří Navrátil, Vijay Iyengar, Karthikeyan Shanmugam (2018): Confidence scoring using whitebox meta-models with linear classifier probes. *ArXiv:1805.05396*.

PEER-REVIEWED PAPERS

- [9] Rashmi Sankepally, **Tongfei Chen**, Benjamin Van Durme, Douglas W. Oard (2018): A test collection for coreferent mention retrieval. In *Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 1209-1212.
- [8] Hainan Xu, **Tongfei Chen**, Dongji Gao, Yiming Wang, Ke Li, Nagendra Goel, Yishay Carmiel, Daniel Povey, Sanjeev Khudanpur (2018): A pruned RNNLM lattice-rescoring algorithm for automatic speech recognition. In *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.
- [7] Benjamin Van Durme, Tom Lippincott, Kevin Duh, Deana Burchfield, Adam Poliak, Cash Costello, Tim Finin, Scott Miller, James Mayfield, Philipp Koehn, Craig Harman, Dawn Lawrie, Chandler May, Max Thomas, Annabelle Carrell, Julianne Chaloux, **Tongfei Chen**, Alex Comerford, Mark Dredze, Benjamin Glass, Shudong Hao, Patrick Martin, Pushpendre Rastogi, Rashmi Sankepally, Travis Wolfe, Ying-Ying Tran, Ted Zhang (2017): CADET: Computer Assisted Discovery Extraction and Translation. In *Proceedings of the 8th International Joint Conference on Natural Language Processing, System Demonstrations (IJCNLP Demo)*, pp. 5-8.
- [6] **Tongfei Chen** (2017): Typesafe abstractions for tensor operations. In *Proceedings of the 8th ACM SIGPLAN International Symposium on Scala (SCALA)*, pp. 45-50.
- [5] **Tongfei Chen**, Benjamin Van Durme (2017): Discriminative information retrieval for question answering sentence selection. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2 (EACL)*, pp. 719-725.
- [4] Junhao Zhang, **Tongfei Chen**, Junfeng Hu (2015): On the relationship between Gaussian stochastic blockmodels and label propagation algorithms. *Journal of Statistical Mechanics: Theory and Experiment (J. Stat. Mech.)*. 2015(3), P03009.
- [3] Ni Sun, **Tongfei Chen**, Liumingjing Xiao, Junfeng Hu (2014): Diachronic deviation features in continuous space word representations. In *Proceedings of the 13th China National Conference on Computational Linguistics (CCL; LNAI 8801)*, pp. 23-33.
- [2] **Tongfei Chen**, Xiaojun Zou, Weimeng Zhu, Junfeng Hu (2013): Human-computer interactive Chinese word segmentation: An adaptive Dirichlet process mixture model approach. In *Proceedings of the 6th International Joint Conference on Natural Language Processing (IJCNLP)*, pp. 1278-1284.
- [1] **Tongfei Chen**, Weimeng Zhu, Xueqiang Lv, Junfeng Hu (2013): A Kalman filter based human-computer interactive segmentation system for ancient Chinese texts. In *Proceedings of the 12th China National Conference on Computational Linguistics (CCL; LNAI 8202)*, pp. 25-35.

Selected Projects

- **SCALE 2016: Computer-Aided Discovery, Extraction and Translation** *Jun. 2016 – Aug. 2016*
Participated in the 2016 Summer Camp for Applied Language Exploration (SCALE) workshop at the Human Language Technology Center of Excellence (HLTCOE) at Johns Hopkins University. Completed a system for user-customizable trainable cross-lingual information retrieval.
- **SCALE 2015: Chinese Entity Discovery and Linking** *Jun. 2015 – Jul. 2015*
Participated in the 2015 Summer Camp for Applied Language Exploration (SCALE) workshop at the Human Language Technology Center of Excellence (HLTCOE) at Johns Hopkins University. Worked on entity linking and coreference resolution on Chinese data.
- **Reviewer Assignment System for Funding Applications** *Sep. 2013 – Oct. 2013*
Collaborated in the development of an intelligent reviewer assignment system for the National Science Foundation of China (NSFC). Utilized techniques such as recommendation systems, graph-based keyphrase extraction and unsupervised ontology construction.

Teaching Experience

- **Teaching Assistant** *Sept. 2016 – Dec. 2016*
EN.600.365: Knowledge Discovery from Text, Johns Hopkins University
Graded assignments; gave guest lectures; and aided students in their course projects.
- **Guest Lecturer** *Nov. 2013*
048-30530: Introduction to Computing (Honor Track), Peking University
Gave lectures on introductory Bayesian statistics and machine learning.

Presentations and Talks

- Towards typesafe deep learning in Scala *Mar. 18, 2018*
Northeast Scala Symposium 2018, Boston, MA, USA
- Typesafe abstractions for tensor operations *Oct. 23, 2017*
SCALA 2017, Vancouver, BC, Canada
- Discriminative information retrieval for knowledge discovery *Feb. 15, 2017*
DARPA DEFT/LORELEI site visit, HLTCOE, JHU
- Discriminative information retrieval for knowledge discovery *Oct. 25, 2016*
Center for Language and Speech Processing Student Seminar, JHU
- Human-computer interactive Chinese word segmentation: An adaptive Dirichlet process mixture model approach *Oct. 17, 2013*
IJCNLP 2013, Nagoya, Aichi, Japan
- A Kalman filter based human-computer interactive segmentation system for ancient Chinese texts. *Oct. 11, 2013*
CCL 2013, Suzhou, Jiangsu, China

Services

- Program committee member:
 - CCL 2017: China National Conference on Computational Linguistics
 - KG4IR 2017, 2018: Workshop on Knowledge Graphs and Semantics for Text Retrieval and Analysis
 - TADGM 2018: Theoretical Foundations and Applications of Deep Generative Models
- Reviewer:
 - CCL 2017
 - EMNLP 2018
 - KG4IR 2017
 - TADGM 2018
- Secondary reviewer:
 - ACL 2014, 2015, 2017, 2018
 - ACL Demo Track 2017
 - EACL 2017
 - EMNLP 2014, 2017
 - IJCNLP 2017
 - NAACL 2015
 - TAACL 2015, 2017
 - WWW 2015
- PhD recruitment committee 2018, Johns Hopkins University
- North American Computational Linguistics Olympiad (NACLO) organizing committee 2016, Johns Hopkins University

Honors and Awards

- Outstanding Undergraduate Thesis, Peking University. *Jun. 2014*
- Founder Inc. Scholarship, Peking University. *Sep. 2013*

Skills

- Programming languages: Scala, Java, Python, C/C++, C#, Haskell
- Natural languages: Mandarin Chinese (*native*), English
- Libraries and tools:
 - Deep learning: PyTorch, TensorFlow
 - Information retrieval: Lucene
 - Data serialization: Thrift
 - Data visualization: Gephi
 - Distributed computing: Spark
 - Scala ecosystem: Cats, Shapeless