

TONGFEI CHEN

CONTACT INFORMATION

ADDRESS: Hackerman 321, 3400 N Charles St, Baltimore, MD 21218, USA
PHONE: +1 (443) 813-5763
EMAIL: tongfei@jhu.edu; i@tongfei.me
HOMEPAGE: <http://cs.jhu.edu/~tongfei>; <http://tongfei.me>

EDUCATION

2020 (expected)	Ph.D. in Computer Science Johns Hopkins University , Baltimore, MD, USA Center for Language and Speech Processing Department of Computer Science Advisor: Prof. Benjamin Van Durme
Jun 2014	B.Sc. in Computer Science Peking University , Beijing, China Department of Computer Science Advisor: Prof. Junfeng Hu Thesis: <i>Large-scale unsupervised word segmentation for Classical Chinese: Research & system</i>

WORK EXPERIENCE

Aug 2014 - Current	Research Assistant (Ph.D. research) <i>Center for Language and Speech Processing, Johns Hopkins University, Baltimore, MD, USA</i> Advisor: Prof. Benjamin Van Durme Topics: Question answering; information retrieval; knowledge acquisition from text; scalable systems; approximate algorithms
May 2018 - Aug 2018	Applied Scientist Intern <i>Amazon.com, Inc., Seattle, WA, USA</i> Host: Dr. Lambert Mathias Topics: Amazon Alexa; dialog context modeling; dialog state tracking; sequence transduction models
May 2017 - Aug 2017	Research Intern <i>IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA</i> Host: Dr. Jiří Navrátil, Dr. Bing Xiang Topics: Confidence scoring; model calibration; meta-models
Jun 2012 - Jul 2014	Research Assistant (Undergraduate research) <i>Institute of Computational Linguistics, Peking University, Beijing, China</i> Advisor: Prof. Junfeng Hu Topics: Word segmentation; network analysis; ontology construction

PUBLICATIONS

- [1] **Tongfei Chen**, Jiří Navrátil, Vijay Iyengar, Karthikeyan Shanmugam (2018): Confidence scoring using whitebox meta-models with linear classifier probes. *ArXiv: 1805.05396*.
- [2] 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.
- [3] 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)*, pp. 5929-5933.
- [4] 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.
- [5] **Tongfei Chen** (2017): Typesafe abstractions for tensor operations. In *Proceedings of the 8th ACM SIGPLAN International Symposium on Scala (SCALA)*. pp. 45-50.
- [6] **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.
- [7] 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.
- [8] 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)*, pp. 23-33.
- [9] **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.
- [10] **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)*, pp. 25-35.

SELECTED PROJECTS

Jun 2016 | **SCALE 2016: Computer-Aided Discovery, Extraction and Translation**
- Aug 2016 | *Human Language Technology Center of Excellence, Johns Hopkins University*

- 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.
- Jun 2015 - Jul 2015 | **SCALE 2015: Chinese Entity Discovery and Linking**
Human Language Technology Center of Excellence, Johns Hopkins University
 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.
- Sep 2013 - Oct 2013 | **Reviewer Assignment System for Funding Applications**
Institute of Computational Linguistics, Peking University
 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 key-phrase extraction and unsupervised ontology construction.

TEACHING EXPERIENCE

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

PRESENTATIONS AND TALKS

- Nov 16, 2018 | Towards typesafe deep learning in Scala
Scale by the Bay 2018, San Francisco, CA
- Mar 18, 2018 | Towards typesafe deep learning in Scala
Northeast Scala Symposium 2018, Boston, MA
- Oct 23, 2017 | Typesafe abstractions for tensor operations
SCALA 2017, Vancouver, BC, Canada
- Feb 15, 2017 | Discriminative information retrieval for knowledge discovery
DARPA DEFT/LORELEI site visit, HLTCOE, JHU
- Oct 25, 2016 | Discriminative information retrieval for knowledge discovery
Center for Language and Speech Processing Student Seminar, JHU
- Oct 17, 2013 | Human-computer interactive Chinese word segmentation: An adaptive Dirichlet process mixture model approach
IJCNLP 2013, Nagoya, Aichi, Japan
- Oct 11, 2013 | A Kalman filter based human-computer interactive segmentation system for ancient Chinese texts
CCL 2013, Suzhou, Jiangsu, China

HONORS AND AWARDS

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

SERVICE

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

SKILLS

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