Tongfei Chen

Contact Information

Office:	Center for Language and Speech Processing
	Johns Hopkins University
	Hackerman 321
Email:	tongfei@jhu.edu
Homepage:	https://cs.jhu.edu/~tongfei

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

2014 — **B.Sc. in Computer Science**
Peking University, Beijing, China
Department of Computer Science
Advisor: Prof. Junfeng Hu
Thesis: *Large-scale unsupervised word segmentation for Classical Chinese: Research & system*

Work Experience

Aug 2014 – present — **Research Assistant** (Ph.D. research)
Center for Language and Speech Processing, Johns Hopkins University, Baltimore, MD, USA
Advisor: Prof. Benjamin Van Durme
Topics: Information extraction; question answering; information retrieval; natural language inference; knowledge acquisition from text; scalable systems; approximate algorithms

May 2018 – Aug 2018 — **Applied Scientist Intern**
Amazon.com, Inc., Seattle, WA, USA
Host: Dr. Lambert Mathias
Topics: Amazon Alexa; dialogue context modeling; sequence transduction models

May 2017 – Aug 2017 — **Research Intern**
IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA
Hosts: Dr. Jiří Navrátil, Dr. Bing Xiang
Topics: Confidence scoring; model calibration; meta-models

Jun 2012 – Jun 2014 — **Research Assistant** (Undergraduate research)
Institute of Computational Linguistics, Peking University, Beijing, China
Advisor: Prof. Junfeng Hu
Topics: Word segmentation; graph & network analysis; ontology construction

Preprints and Working Papers

[P1] Yunmo Chen, **Tongfei Chen**, Seth Ebner, Benjamin Van Durme (2019): Reading the Manual: Event extraction as definition comprehension. *arXiv preprint arXiv:1912:01586.* [NLP]

Peer-Reviewed Papers

[1] **Tongfei Chen**, Yunmo Chen, Benjamin Van Durme (2020): Hierarchical entity typing via multi-level learning to rank. In *Proceedings of the Annual Conference of the Association of Computational Linguistics (ACL).* [NLP]

[2] **Tongfei Chen***, Zhengping Jiang*, Adam Poliak, Keisuke Sakaguchi, Benjamin Van Durme (2020): Uncertain natural language inference. In *Proceedings of the Annual Conference of the Association of Computational Linguistics (ACL).* [NLP]

[3] Yiming Wang, **Tongfei Chen**, Hainan Xu, Shuoyang Ding, Hang Lv, Yiwen Shao, Nanyun Peng, Xie Lei, Shinji Watanabe, Sanjeev Khudanpur (2019): ESPRESSO: A fast end-to-end neural speech recognition toolkit. In *Proceedings of the 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU).* [SPEECH]

[4] Arya D. McCarthy, **Tongfei Chen**, Rachel Rudinger, David W. Matula (2019): Metrics matter in community detection. In *Proceedings of the 8th International Conference on Complex Networks and their Applications (COMPLEX NETWORKS).* pp. 164–175. [GRAPH]

[5] Arya D. McCarthy, **Tongfei Chen**, Seth Ebner (2019): An exact No Free Lunch theorem for community detection. In *Proceedings of the 8th International Conference on Complex Networks and their Applications (COMPLEX NETWORKS).* pp. 176–187. [GRAPH]

[6] **Tongfei Chen**, Chetan Naik, Hua He, Pushpendre Rastogi, Lambert Mathias (2019): Improving long distance slot carryover in spoken dialogue systems. In *Proceedings of the First Workshop of NLP for Conversational AI (NLP4ConvAI@ACL).* pp. 96–105. [NLP]
[best paper award]

[7] Zhongyang Li, **Tongfei Chen**, Benjamin Van Durme (2019): Learning to rank for plausible plausibility. In *Proceedings of the Annual Conference of the Association of Computation Linguistics (ACL).* pp. 4818–4823. [NLP]

[8] Pushpendre Rastogi, Arpit Gupta, **Tongfei Chen**, Lambert Mathias (2019): Scaling multi-domain dialogue state tracking via query reformulation. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Industry Track (NAACL).* pp. 97–105. [NLP]

[9] J. Edward Hu, Huda Khayrallah, Ryan Culkin, Patrick Xia, **Tongfei Chen**, Matt Post, Benjamin Van Durme (2019): Improved lexically-constrained decoding for translation and monolingual rewriting. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).* pp. 839–850. [NLP]

[10] Yiming Wang, Xing Fan, I-Fan Chen, Yuzong Liu, **Tongfei Chen**, Björn Hoffmeister (2019): End-to-end anchored speech recognition. In *Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* pp. 7090–7094. [SPEECH]

[11] **Tongfei Chen**, Jiří Navrátil, Vijay Iyengar, Karthikeyan Shanmugam (2019): Confidence scoring using whitebox meta-models with linear classifier probes. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS); Proceedings of Machine Learning Research 89 (PMLR),* pp. 1467–1475. [ML]

[12] 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. [NLP]

[13] 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. [SPEECH] [NLP]

[14] 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. [NLP]

[15] **Tongfei Chen** (2017): Typesafe abstractions for tensor operations. In *Proceedings of the 8th ACM SIGPLAN International Symposium on Scala (SCALA@SPLASH)*. pp. 45–50. [PL]

[16] **Tongfei Chen**, Benjamin Van Durme (2017): Discriminative information retrieval for question asnwering sentence selection. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2 (EACL)*, pp. 719–725. [NLP]

[17] 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. [GRAPH]

[18] 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; LNCS 8801)*, pp. 23–33. [NLP]

[19] **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. [NLP]

[20] **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; LNCS 8202)*, pp. 25–35. [NLP]

System Descriptions

[S1] Yunmo Chen, Seth Ebner, **Tongfei Chen**, Patrick Xia, Elias Stengel-Eskin, J. Edward Hu, Nils Holzenberger, Ryan Culkin, Craig Harman, Max Thomas, Aaron Steven White, Kyle Rawlins, Benjamin Van Durme (2019): NIST TAC SM-KBP 2019 system description: JHU/UR framework. In *Proceedings of the Text Analysis Conference (TAC)*. [NLP]

[S2] Mozhi Zhang, Jordan Boyd-Graber, Michelle Yuan, C. Anton Rytting, Weiwei Yang, Philip Resnik, Ting Hua, Adam Poliak, Adam Teichert, **Tongfei Chen**, Xu Han, Linghao Jin, João Sedoc, Benjamin Van Durme (2019): LoReHLT19 System Description UMD-JHU. [NLP]

[S3] Patrick Xia, Elias Stengel-Eskin, **Tongfei Chen**, Seth Ebner, Nils Holzenberger, Ryan Culkin, Pushpendre Rastogi, Xutai Ma, Benjamin Van Durme (2018): NIST TAC SM-KBP 2018 system description: JHU/UR pipeline. In *Proceedings of the Text Analysis Conference (TAC)*. [NLP]

SELECTED PROJECTS

Jun 2016 – Aug 2016 | **SCALE 2016: Computer-Aided Discovery, Extraction and Translation**
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 keyphrase extraction and unsupervised ontology construction.

Presentation and Talks

Aug 1, 2019	Improving long distance slot carryover in spoken dialogue systems *NLP4ConvAI@ACL 2019: Best paper talk, Florence, Tuscany, Italy*
May 29, 2019	Uncertain natural language inference *DARPA LORELEI/AIDA site visit, JHU*
Mar 25, 2019	Uncertain natural language inference *Center for Language and Speech Processing Student Seminar, JHU*
Nov 16, 2018	Towards typesafe deep learning in Scala *Scale by the Bay 2018, San Francisco, CA, USA*
Mar 18, 2018	Towards typesafe deep learning in Scala *Northeast Scala Symposium 2018, Boston, MA, USA*
Oct 23, 2017	Typesafe abstractions for tensor operations *SCALA@SPLASH 2017, Vancouver, BC, Canada*
Feb 15, 2017	Discriminative information retrieval for knowledge discovery *DARPA DEFT/LORELEI site visit, JHU*
Oct 25, 2016	Discriminative information retrieval for knowledge discovery *Center for Language and Speech Processing Student Seminar, JHU*
Oct 17, 2013	HCI Chinese word segmentation: An adaptive Dirichlet process mixture model approach *IJCNLP 2013, Nagoya, Aichi, Japan*
Oct 11, 2013	A Kalman filter based HCI segmentation system for ancient Chinese texts *CCL 2013, Suzhou, Jiangsu, China*

Honors and Awards

Jul 2019	Best paper award (2/25) *The 1st Workshop on NLP for Conversational AI @ ACL 2019*
Jun 2014	Outstanding Undergraduate Thesis *Peking University*

Sep 2013	**Founder Inc. Scholarship**
	Peking University

Open Source

- Espresso: Fast end-to-end automatic speech recognition based on fairseq. (600+ stars on GitHub)
  https://github.com/freewym/espresso (contributor)

- Nexus: Experimental typesafe tensors and deep learning in Scala. (200+ stars on GitHub)
  https://github.com/ctongfei/nexus

- Progressbar: A terminal-based progress bar for JVM. (450+ stars on GitHub)
  https://github.com/ctongfei/progressbar

Service

- Reviewer / Program committee member:

  - AAAI 2020
  - AACL 2020
  - ACL 2020
  - CCL 2017
  - EMNLP 2020, 2019, 2018
  - KG4IR@SIGIR 2017
  - NAACL 2019
  - RCQA@AAAI 2020
  - TADGM@ICML 2018

- Secondary reviewer:

  - ACL 2019, 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 2018–2020, Johns Hopkins University

- North American Computational Linguistics Olympiad (NACLO) organizing committee 2016

Skills

- Programming languages:

  - Python
  - Scala
  - Java
  - C/C++
  - C#
  - Haskell

- Natural languages:

  - Mandarin Chinese (*native*)
    (Standard & Sichuanese)
  - 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
  - Typesetting: LaTeX