

Zhuyun Dai

CONTACT INFORMATION	5000 Forbes Avenue Pittsburgh, PA 15213 Homepage: http://www.cs.cmu.edu/~zhuyund/	412-960-0176 zhuyund@cs.cmu.edu
RESEARCH INTERESTS	Information Retrieval, Natural Language Processing, and Machine Learning. My main focus is to develop deep learning approaches to improve language understanding in today's large-scale information retrieval systems. Broader research topics include large-scale distributed IR, multi-modal systems, and conversational AI.	
EDUCATION	Carnegie Mellon University , Pittsburgh, PA, USA Ph.D., Language and Information Technologies, <i>Expected</i> : Summer 2020 M.S., Language Technologies, December 2016 <ul style="list-style-type: none">• Advisor: Prof. Jamie Callan Peking University , Beijing, China B.S., Computer Science, June 2014	
EXPERIENCES	Graduate Research Assistant Language Technologies Institute, Carnegie Mellon University Advisor: Prof. Jamie Callan <ul style="list-style-type: none">• My Ph.D. dissertation research develops neural network solutions to improve language understanding in today's information retrieval (IR) systems, with a focus on relevance ranking ([3, 5, 8, 10, 11, 13]) and document understanding ([1, 2, 4]). My systems achieved strong performances in high-competitive academic benchmarks such as NTCIR13 (1st place, 2017) and Microsoft MS MARCO (2nd place from Sep 2019 - Jan 2020, currently 3rd place).• More broadly, I have developed data-driven approaches to improve large-scale distributed IR systems' efficiency ([9, 12, 14, 15]). Currently, I am also studying tabular data summarization for conversational AI systems.	Sept 2014 to present
	Ph.D. Intern at Facebook Search Team Facebook, Seattle Advisor: Linjun Yang, Engineering Manager, Research Scientist We developed <i>deep collective ranking</i> models for Facebook Search, a novel neural network for jointly ranking multiple search candidates. Our models were deployed to the production system.	May 2018 to Aug 2018
	Undergraduate Research Assistant Search Engine and Web Mining Group, Peking University Advisor: Hongfei Yan, Ph.D. User-generated reviews on online E-Commerce platforms provide valuable feedback about the good/bad aspects of products. We proposed a novel, unsupervised topic modeling approach that automatically discovers aspects and their corresponding sentiments without using manual annotations. <i>2014 Excellent Graduation Thesis Award, Peking University</i>	June 2013 to May 2014
	Undergraduate Research Intern Language Technologies Institute, Carnegie Mellon University Advisor: Prof. Alexander G. Hauptmann We developed a cognitive assistive system to help patients use asthma inhalers. Incorrect use of inhalers significantly weakens the medical effects. Our system uses multi-modal signals from audio, RGB video, and depth video to detect user errors, and provides alerts and guidance to users.	Jul 2013 to Sept 2013

1. **Z. Dai** and J. Callan. “Context-aware document term weighting for Ad-Hoc search” To appear in *The Web Conference (WWW)*. 2020.
2. **Z. Dai** and J. Callan. “An evaluation of weakly-supervised DeepCT in the TREC 2019 Deep Learning Track” To appear in *TREC*. 2020.
3. R. Padaki, **Z. Dai**, and J. Callan. “Rethinking query expansion for BERT reranking” (Short Paper). To appear in the *42nd European Conference in Information Retrieval (ECIR)*. 2020.
4. **Z. Dai** and J. Callan. “Context-aware sentence / passage term importance estimation for first stage retrieval”. arXiv preprint arXiv:1910.10687. 2019.
5. **Z. Dai** and J. Callan. “Deeper text understanding for IR with contextual neural language modeling” (Short Paper). In *Proceedings of the 42nd International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR)*. 2019.
6. **Z. Dai***, Z. Fan*, H. R. Mohammad, and J. Callan, “Local matching networks for engineering diagram search”. In *Proceedings of The Web Conference (WWW)*. 2019.
7. S. Medina, **Z. Dai**, Y. Gao, “Where is this? Video geolocation based on neural network features”. *arXiv preprint arXiv:1810.09068*. 2018
8. M. A. Pyreddy, V. Ramaseshan, N. N. Joshi, **Z. Dai**, C. Xiong, J. Callan and Z. Liu. “Consistency and variation in Kernel Neural Ranking Model” (short paper). In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR)*. 2018.
9. **Z. Dai** and J. Callan. “Inverted list caching for topical index shards”. In *Proceedings of the 40th European Conference on Information Retrieval (ECIR)*. 2018.
10. **Z. Dai**, C. Xiong, J. Callan, and Z. Liu. “Convolutional neural networks for soft-matching n-grams in ad-hoc search.” In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (WSDM)*. 2018.
11. **Z. Dai**, C. Xiong, and J. Callan. “An evaluation of the Kernel Based Neural Ranking Model in NTCIR13 WWW”. In *Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies*. 2017.
12. **Z. Dai**, Y. Kim and J. Callan. “Learning to rank resources” (Short Paper). In *Proceedings of the 40th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR)*. 2017.
13. C. Xiong*, **Z. Dai***, J. Callan, Z. Liu, and R. Power. “End-to-end neural ad-hoc ranking with kernel pooling.” In *Proceedings of the 40th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR)*. 2017.
14. **Z. Dai**, C. Xiong, and J. Callan. “Query-biased partitioning for selective search.” In *Proceedings of the 25th ACM Conference on Information and Knowledge Management (CIKM)*. 2016.
15. **Z. Dai**, Y. Kim, and J. Callan. “How random decisions affect selective distributed search” (Short Paper). In *Proceedings of the 38th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR)*. 2015.
16. Y. Xu, Y. Yue, **Z. Dai** and X. Wang. “A management model for SDN-based data center networks.” In *Proceedings of Computer Communications Workshops (INFOCOM WKSHPS)*. 2014.

PROFESSIONAL ACTIVITIES Academic Conference Program Committee Member
 SIGIR 2020, AAAI 2020, ECIR 2020, EMNLP 2019, SIGIR 2019, WWW 2019,
 ECIR 2019, ECIR 2018, CIKM 2017

Academic Journal Reviewer
 ACM Transactions on Information Systems 2019, 2020
 Information Processing and Management 2019, 2020

Teaching Assistant Fall 2016 and Fall 2017
 CMU 11642 - Search Engines (Instructor: Prof. Jamie Callan)

AWARDS

- 2nd Place in the 2019 CMU McGinnis Venture Competition 2019
- Honorable Mention, LTI Student Research Symposium 2016
- Outstanding Graduating Student of Beijing 2014
- Excellent Graduation Thesis Award, Peking University 2014
- The Google Anita Borg Memorial Scholarship 2013
- National Scholarship 2012
- Student Travel Award: SIGIR 2015, CIKM 2016, SIGIR 2017, TheWebConf 2019,
 SIGIR 2019

TECHNICAL SKILLS

- Programming: Python, C++, C, Java, SQL
- Tools: TensorFlow, PyTorch, Keras, AllenNLP, NLTK, Spacy, XGBoost, LibSVM,
 SVMlight, Numpy, SciPy, Scikit-learn

Last Updated: February 18, 2020