

Huan Sun

My research interests lie in data mining and machine learning, with emphasis on *text mining and understanding*, *question answering*, *network analysis*, and *human behavior understanding*. My work develops advanced computational frameworks that enable users to effectively query big data, collaborate with each other, and acquire knowledge for decision making in various domains such as *healthcare*, *education*, and *business*.

699 Dreese Lab
2015 Neil Ave
Columbus, OH 43210
sun.397@osu.edu
<http://web.cse.ohio-state.edu/~huansun/>

Education	Postdoc 01-06/2016	Computer Science and Engineering University of Washington Advisor: Prof. Daniel S. Weld	
	Ph.D. 2010-2015	Computer Science University of California, Santa Barbara (UCSB) Advisor: Prof. Xifeng Yan <i>Thesis: Mining Disparate Sources for Question Answering</i> <i>2016 SIGKDD Ph.D. Dissertation Runner-Up Award</i>	
	B.S. 2006-2010	Electronic Engineering and Information Science University of Science and Technology of China (USTC) <i>Excellent Graduation Thesis Award (top 5%)</i>	
Honors & Awards		SIGKDD Ph.D. Dissertation Runner-Up Award Rising Star in EECS, MIT Outstanding Dissertation Award, Computer Science Department, UCSB Ph.D. Progress Award, Computer Science Department, UCSB Regents' Special Fellowship, UCSB Excellent Graduation Thesis Award (top 5%), USTC Guanghua Education Scholarship (first-class), USTC National Scholarship (top 2%), USTC	2016 2015 2015 2014 2010, 2014 2010 2009 2007, 2008
Professional Experience		Assistant Professor, The Ohio State University (OSU) Consultant, Fujitsu Laboratories of America Consultant, Rulai, Inc. (A virtual assistant startup) Visiting Scholar, Baidu Research, Sunnyvale Summer Research Intern, Microsoft Research, Redmond Summer Research Intern, IBM Watson Research Center Research Intern, Microsoft Research Asia	09/2016 - 05/2017 - 03-08/2017 09-12/2015 2014, 2015 2013 2010
Grants		[1] Huan Sun (Single PI) . National Science Foundation (NSF), "III: Small: Towards Resolving Ad-hoc Concept Queries with Table Answers via Multi-source Data Mining," 08/2018-07/2021. \$499,001 . [2] Huan Sun (PI) . Patient-Centered Outcomes Research Institute (PCORI), "Unlocking Clinical Text in EMR by Query Refinement Using Both Knowledge Bases and Word Embedding," 01/2018-12/2020. \$1,060,000 (total); \$483,001 (my part) . In collaboration with Nationwide Children's Hospital (PI: Dr. Simon Lin, Primary Contact). [3] Huan Sun (Single PI) . Army Research Office (ARO), "Advancing Human and Machine Question Answering via Human-Machine Collaboration," 09/2017-09/2020.	

\$498,526.

- [4] **Huan Sun (Single PI)**. Fujitsu Lab of America, “Information Extraction for IoT Device APIs Based on Developer Documentation,” 05/2017-05/2018. **\$50K**.

Publications
(*: My advisee)

After joining OSU:

- [1] Jayavardhan Reddy Peddamail*, Ziyu Yao, Zhen Wang, **Huan Sun**, “A Comprehensive Study of StaQC for Deep Code Summarization”
SIGKDD 2018 Deep Learning Day.
- [2] Ziyu Yao*, Daniel S. Weld, Wei-Peng Chen, **Huan Sun**, “StaQC: A Systematically Mined Question-Code Dataset from Stack Overflow”
The Web Conference (former WWW Conference) 2018.
- [3] Jie Zhao*, Yu Su, Ziyu Guan, **Huan Sun**, “An End-to-End Deep Framework for Answer Triggering with a Novel Group-Level Objective”
The 22th Conference on Empirical Methods in Natural Language Processing (EMNLP) 2017.
- [4] Yu Su⁺, Honglei Liu⁺, Semih Yavuz, Izzeddin Gur, Huan Sun, Xifeng Yan. “Global Relation Embedding for Relation Extraction”
In Proc. of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT) 2018. (+: Equal contribution)
- [5] Yaliang Li, Nan Du, Chaochun Liu, Yusheng Xie, Wei Fan, Qi Li, Jing Gao, **Huan Sun**, “Reliable Medical Diagnosis from Crowdsourcing: Discover Trustworthy Answers from Non-Experts”
The 10th ACM International Conference on Web Search and Data Mining (WSDM) 2017.
- [6] Chaochun Liu, **Huan Sun**, Nan Du, Shulong Tan, Hongliang Fei, Wei Fan, Tao Yang, Hao Wu, Yaliang Li, Chenwei Zhang, “Augmented LSTM Framework to Construct Medical Self-diagnosis Android”
IEEE 16th International Conference on Data Mining (ICDM) 2016.
- [7] Yu Su, **Huan Sun**, Brian Sadler, Mudhakar Srivatsa, Izzeddin Gur, Zenghui Yan, Xifeng Yan, “On Generating Characteristic-rich Question Sets for QA Evaluation”
The 21st Conference on Empirical Methods in Natural Language Processing (EMNLP) 2016.

Before joining OSU:

- [8] **Huan Sun**, Hao Ma, Xiaodong He, Wen-tau Yih, Yu Su, Xifeng Yan, “Table Cell Search for Question Answering”
The 25th International Conference on World Wide Web (WWW) 2016.
- [9] Yang Li, Shulong Tan, **Huan Sun**, Jiawei Han, Dan Roth, Xifeng Yan, “Entity Disambiguation with Linkless Knowledge Bases”
The 25th International Conference on World Wide Web (WWW) 2016.
- [10] Fangqiu Han, Shulong Tan, **Huan Sun**, Xifeng Yan, Mudhakar Srivatsa, Deng Cai, “Distributed Representations of Expertise”
SIAM International Conference on Data Mining (SDM), 2016.
- [11] **Huan Sun**, Hao Ma, Wen-tau Yih, Chen-Tse Tsai, Jingjing Liu, Ming-Wei Chang, “Open Domain Question Answering via Semantic Enrichment”
The 24th International Conference on World Wide Web (WWW) 2015.

- [12] Yu Su, Shengqi Yang, **Huan Sun**, Mudhakar Srivatsa, Sue Kase, Michelle Vanni, Xifeng Yan, “Exploiting Relevance Feedback in Knowledge Graph Search”
The 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD) 2015.
- [13] Ziyu Guan, Shengqi Yang, **Huan Sun**, Mudhakar Srivatsa, Xifeng Yan, “Fine-Grained Knowledge Sharing in Collaborative Environments”
IEEE Transactions on Knowledge and Data Engineering (TKDE) 2015.
- [14] **Huan Sun**, Mudhakar Srivatsa, Shulong Tan, Yang Li, Lance Kaplan, Shu Tao, Xifeng Yan, “Analyzing Expert Behaviors in Collaborative Networks”
The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD) 2014.
- [15] Feida Zhu, **Huan Sun**, Xifeng Yan, “Network Mining and Analysis for Social Applications”
The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD) 2014, Tutorial.
- [16] Shengqi Yang, Yinghui Wu, **Huan Sun**, Xifeng Yan, “Schemaless and Structureless Graph Querying”
The 40th International Conference on Very Large Data Bases (VLDB) 2014.
- [17] Shengqi Yang, Yanan Xie, Yinghui Wu, Tianyu Wu, **Huan Sun**, Jian Wu, Xifeng Yan, “SLQ: A User-friendly Graph Querying System”
ACM International Conference on Management of Data (SIGMOD) 2014, Demo.
- [18] **Huan Sun**, Mudhakar Srivatsa, Lance Kaplan, Xifeng Yan, “Analyzing Expert Behaviors in Collaborative Networks”
International School and Conference on Network Science (NetSci) 2014.
- [19] Nan Li, **Huan Sun**, Kyle Chipman, Jemin George, Xifeng Yan, “A Probabilistic Approach to Uncovering Attributed Graph Anomalies”
SIAM International Conference on Data Mining (SDM) 2014.
- [20] **Huan Sun**, Alex Morales, Xifeng Yan, “Synthetic Review Spamming and Immunization”
The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD) 2013.
- [21] Shulong Tan, Yang Li, **Huan Sun**, Ziyu Guan, Xifeng Yan, Jiajun Bu, Chun Chen, Xiaofei He, “Interpreting the Public Sentiment Variations on Twitter”
IEEE Transactions on Knowledge and Data Engineering (TKDE) 2014.
- [22] **Huan Sun**, Gengxin Miao, Xifeng Yan, “Noise-Resistant Bicluster Recognition”
IEEE 13th International Conference on Data Mining (ICDM) 2013.
- [23] Alex Morales, **Huan Sun**, Xifeng Yan, “Synthetic Review Spamming and Immunization”
The 22nd International Conference on World Wide Web (WWW) 2013, poster.
- [24] **Huan Sun**, Gengxin Miao, Yu S. Huang, Xifeng Yan, “Noise-Resistant Bicluster Recognition”
The 17th Annual International Conference on Research in Computational Molecular Biology (RECOMB) 2013, poster.

Students Supervised/Mentored

Current Students:

- Ziyu Yao, Ph.D. student, Fall 2016-Now.
 - One paper published in WWW 2018
 - 2017 Summer Internship at Fujitsu Laboratories of America

- 2018 Summer Internship at Microsoft Research, Redmond (Reinforcement Learning + NLP group)
- Jie Zhao, Ph.D. student, Fall 2016-Now.
 - One paper published in EMNLP 2017.
 - 2017 Summer Internship at Rulai (a startup for building AI chatbots for enterprise).
- Zhen Wang, Ph.D. student, Fall 2017-Now.
 - One paper under review.
- Jayavardhan Reddy Peddamail, Master student, Spring 2018-Now.
 - One paper accepted in KDD 2018 Deep Learning Day.

OSU Alumni:

- Dushyanta Dhyani, Master student, Graduated in May 2018.
 - Joined Amazon (working on AWS + Deep Learning)

Before joining OSU:

- Alex Morales, Master student, 2012-2013, UCSB. Collaborator of my work published in SIGKDD 2013, “Synthetic Review Spamming and Defense.” Now a Ph.D. student at Univ. of Illinois at Urbana Champaign (5-year fellowship).
- Sang Nguyen, Undergraduate student from Oxnard College, Summer 2012, UCSB. “Perceptron for Classification: Analysis and Application.”

Teaching Experience

At OSU:

- Human-Machine Collaboration Mechanisms in Data Mining, Database, and AI (CSE5249, Graduate Level)
Instructor, Spring 2018, OSU, [Course Website](#).
- Introduction to Data Mining (CSE5243, Mixed Undergraduate & Graduate Level)
Instructor, Autumn 2017, OSU, [Course Website](#).
- Natural Language Question Answering (CSE5539, Graduate Level)
Instructor, Autumn 2016, OSU, [Course Website](#).

Before joining OSU:

- Artificial Intelligence (CS165A)
Guest lecturer, Winter 2012, UCSB
- Special Topics in CS: Information Management (CS290D)
Guest lecturer, Spring 2014, UCSB

**Invited Talks
&
Tutorials**

- UMass Amherst, 02/2018
“The Quest for Knowledge: Quest. Answering beyond Knowledge Bases and Texts”
- Rice University, 02/2018
“The Quest for Knowledge: Quest. Answering beyond Knowledge Bases and Texts”
- University of Illinois, Chicago, 01/2018
“The Quest for Knowledge: Quest. Answering beyond Knowledge Bases and Texts”
- Penn State University, 01/2018
“The Quest for Knowledge: Quest. Answering beyond Knowledge Bases and Texts”
- Fujitsu Laboratories of America, 06/2017
“Mining Big Data to Build Intelligent Systems”
- University of Science and Technology of China, Hefei, 01/2017
“Mining Disparate Sources for Question Answering”
- Baidu Research, Sunnyvale, 09/2015
“Network Centric Information Search and Question Answering”
- University of Washington, 08/2015
“Collaborative and Intelligent Query Resolution”
- Army Research Lab, Silver Spring, 05/2015
“Knowledge Enhanced Information Search and Text Understanding”
- The Ohio State University, 03/2015
“Collaborative and Intelligent Query Resolution”
- Simon Fraser University, 03/2015
“Collaborative and Intelligent Query Resolution”
- ShanghaiTech University, China, 09/2014
“Data Mining for General Problem Solving: Methodology and Analysis”
- Tutorial @ SIGKDD, 08/2014
“Network Mining and Analysis for Social Applications”

**Outreach
Activities**

- Hosting two high school summer interns, 06-08/2018
 - Sounak Dey: Thomas Worthington High School
 - Jacob Mayhew: St. Charles Preparatory School
- Served as Judge in AI Club hackathon at Ohio State, 04/2018
Event theme: Weekend-long challenge on AI related projects
Participants: *Undergraduate/graduate students.*
- Enrolled in the 2018 Metro Early College High School Internship program, 10/2017
Program goal: *To host high school interns in Summer 2018.*
- Served as Judge in ASA DataFest at Ohio State, 05/2017
Event theme: Weekend-long challenge on data science related projects
Participants: *Undergraduate students from 5 colleges.*
- Invited speaker at the OSU Innovation Summit, 05/2017
“Mining Big Data to Build Intelligent Assistants Who Can Speak Natural Language”
Event theme: Innovation and disruption in the next digital age
Audience: *Entrepreneurial-minded leaders, academics & students.*
- Invited speaker at High School I/O (the first high school hackathon), 04/2017
“How to Build Intelligent Assistants Who Can Speak Natural Language”
Audience: *High school students from Central Ohio.*
- Attendee of the 2017 Ohio Celebration of Women in Computing (OCWiC), 02/2017
Event theme: *Engaging and empowering Ohio women in computing.*

- Invited speaker at Center of Science and Industry (COSI), 10/2016
“Data Mining and To Be a Data Miner”
Audience: *Underrepresented high school students from Central Ohio.*

**Academic
Services**

Chair Positions:

- 2019: SDM (publicity co-chair)

Review Editor:

- 2018: Frontiers in Big Data and AI (Journal, Data Mining and Management Section)

PC Member:

- 2018: ACL, ICDM, CoNLL
- 2017: SIGKDD, ACL, EMNLP, WWW, CIKM (senior PC), IJCNLP, IJCAI BOOM workshop
- 2016: SIGKDD, EMNLP, ICDM, CIKM, BigMine, IJCAI BOOM workshop

Journal Reviewer:

- ACM Transactions on Knowledge Discovery from Data (TKDD), since 2016
- IEEE Transactions on Knowledge and Data Engineering (TKDE), since 2015

Conference External Reviewer:

- WWW 2015, ICDM (2012-2013), ICDE 2011