Yu Su

✉ su.809@osu.edu | ⌂ http://cs.ucsb.edu/~ysu/ | ⌨ ysu1989

Research Interests

At the intersection of data mining, natural language processing, and machine learning, my research aims to develop a **AI-powered knowledge engine** to bridge the gap between human and data by innovations along the following dimensions:

- **Problems**: Knowledge Harvesting, Natural Language Interface (Semantic Parsing & Dialog System), Knowledge-based Reasoning
- **Methodology**: Deep Learning, Knowledge Representation (Primarily Knowledge Base), Human-Machine Collaboration, Transfer Learning
- **Applications**: Biomedicine, Healthcare, Internet of Things

Education

University of California, Santa Barbara   *Santa Barbara, CA, USA*
Ph.D. in Computer Science   *2012 - 2018*
- Thesis: *Towards Democratizing Data Science with Natural Language Interfaces*
- Committee: *Xifeng Yan (Chair), Ambuj Singh, Amr El Abbadi*

Tsinghua University   *Beijing, China*
B.Eng. in Computer Science and Technology   *2008 - 2012*
- Outstanding Graduate Award

Professional Experience

Microsoft Semantic Machines   *Berkeley, CA, USA*
Researcher   *Oct. 2018 -*

Department of Computer Science & Engineering, The Ohio State University   *Columbus, OH, USA*
Visiting Assistant Professor   *Aug. 2018 -*

Microsoft Research AI   *Redmond, WA, USA*
Research Intern   *Jun. 2016 - Sept. 2016, Jun. 2017 - Sept. 2017*

U.S. Army Research Laboratory (ARL)   *Adelphi, MD, USA*
Visiting Researcher   *May - Jun., Sept. - Nov. 2015, May 2016 - Jun. 2016*

IBM T.J. Watson Research Center   *Yorktown Heights, NY, USA*
Research Intern   *Jun. 2015 - Sept. 2015*

Department of Computer Science, University of California, Santa Barbara   *Santa Barbara, CA, USA*
Teaching and Research Assistant   *Sept. 2012 - Jun. 2018*

Professional Service

Co-organizer
- The First Workshop on Knowledge Base Construction, Reasoning and Mining, Los Angeles, 2018 (co-located with the ACM International Conference on Web Search and Data Mining).

Program Committee Member/Reviewer
- ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD**): 2018, 2019
- Annual Meeting of the Association for Computational Linguistics (**ACL**): 2017, 2018, 2019
- Conference on Empirical Methods in Natural Language Processing (**EMNLP**): 2017, 2018
- Annual Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL**): 2019

- The International World Wide Web Conference (**WWW**): 2018
- AAAI Conference on Artificial Intelligence (**AAAI**): 2019
- SIAM International Conference on Data Mining (**SDM**): 2019
- ACM International Conference on Information and Knowledge Management (**CIKM**): 2017
- Conference on Computational Natural Language Learning (**CoNLL**): 2018
- CCF Conference on Natural Language Processing and Chinese Computing (**NLPCC**): 2017, 2018
- IEEE Transactions on Neural Networks and Learning Systems (**TNNLS**)
- IEEE Transactions on Knowledge and Data Engineering (**TKDE**)
- ACM Transactions on Knowledge Discovery from Data (**TKDD**)

OUTREACH ACTIVITY

- March 2018. Introduce data science and AI to high school students from Laguna Blanca School, Santa Barbara, CA.
- March 2018. Faculty attendee in "Afternoon with Academia," UCSB chapter of Society of Asian Scientists and Engineers.

Invited Talks

[15] Towards Democratizing Data Science with AI-powered Knowledge Engines
Stanfard NLP Seminar. February 07, 2019. Stanford, CA, United States.

[14] Bridging the Gap between Human and Data with AI
Semantic Machines. August 13, 2018. Berkeley, CA, United States.

[13] Bridging the Gap between Human and Data with AI
Fudan University. July 24, 2018. Shanghai, China.

[12] Bridging the Gap between Human and Data with AI
Tsinghua University. June 21, 2018. Beijing, China.

[11] Bridging the Gap between Human and Data with AI
Fujitsu Laboratories of America. March 9, 2018. Sunnyvale, California, United States.

[10] Bridging the Gap between Human and Data with AI
College of Information and Computer Sciences, University of Massachusetts. February 15, 2018. Amherst, Massachusetts, United States.

[9] Bridging the Gap between Human and Data with AI
Dept. of Biomedical Informatics, OSU. January 26, 2018. Columbus, Ohio, United States.

[8] Bridging the Gap between Human and Data with AI
Dept. of Computer Science and Engineering, OSU. January 25, 2018. Columbus, Ohio, United States.

[7] On Building Natural Language Interface to Data and Services
Toutiao AI Lab. October 17, 2017. Beijing, China.

[6] On Building Natural Language Interface to Data and Services
Tsinghua University. October 16, 2017. Beijing, China.

[5] On Building Natural Language Interface to Data and Services
Fudan University. October 10, 2017. Shanghai, China.

[4] Building Natural Language Interfaces to Web APIs
Microsoft Research, MSR AI Seminar. July 25, 2017. Redmond, Washington, United States.

[3] Learning from Zero: Recent Advances in Bootstrapping Semantic Parsers using Crowdsourcing
USC/ISI. October 28, 2016. Los Angeles, California, United States.

[2] Exploiting Relevance Feedback in Knowledge Graph Search
IBM Research. August 15, 2015. Yorktown Heights, New York, United States.

[1] Search over Large-scale Knowledge Graphs
UCSB, IGERT Network Science Seminar. August 14, 2014. Santa Barbara, California, United States.

Grant Proposal Writing

[3] **ARL: Faceted Taxonomy and Search for Scientific Publications**
Wrote the white paper on deep learning for scientific taxonomy construction and evolution discovery.
Led a team of students and worked closely with Brian Sadler, Ryan Zelnio, Warren Greiff from ARL/DoD and
Jiawei Han's UIUC team.
PI: Jiawei Han, co-PI: Xifeng Yan. U.S. Army Research Laboratory (ARL). Grant: $1.5M, 2016-2018.

[2] **NSF III: Small: Knowledge Graph Query Processing and Benchmarking**
Significant contribution in writing the benchmarking and use feedback sections.
PI: Xifeng Yan. NSF IIS Award #1528175. Grant: $500K, 2015-2018.

[1] **NS-CTA IPAN I3/I5: Query Answering in Dynamic Knowledge Networks**
Significant contribution in writing white papers/tech reports and participating in project evaluations.
Visited and worked with Brian Sadler, Sue Kase, and Michelle Vanni at ARL multiple times.
PI: Xifeng Yan. ARL Network Science Collaborative Technology Alliance. Grant: ~$300K/year, 2014-2019.

Patents

[1] **Natural Language Interface to Web API (co-inventor)**
US Patent Pending 15/582,242
A framework for building natural language interface to web API via crowdsourcing, with an optimization
algorithm for crowdsourcing.

Teaching Experience

2016	**Guest Lecturer,** CS273: Data and Knowledge Bases (Graduate)	*UCSB*
2016	**Guest Lecturer,** CS291K: Advanced Data Mining (Graduate)	*UCSB*
2015	**Teaching Assistant,** CS290D: Advanced Data Mining (Graduate)	*UCSB*
2013	**Teaching Assistant,** CS32: Object-Oriented Design and Implementation (Undergraduate)	*UCSB*
2013	**Teaching Assistant,** CS176B: Network Computing (Undergraduate)	*UCSB*
2012	**Teaching Assistant,** CS176A: Introduction to Computer Communication Networks (Undergraduate)	*UCSB*

Honors & Awards

2018	Best Distinguished Graduate Student Lecture of UCSB CS Summit	*UCSB*
2016	SIAM International Conference on Data Mining Travel Award	*NSF*
2012	LINK Lab Fellowship	*UCSB*
2012	Outstanding Graduate Award (*top 10%*)	*Tsinghua University*
2009 - 2011	Merit-based Fellowships (*top 10%*)	*Tsinghua University*
2008	Outstanding Freshman Fellowship	*Tsinghua University*

Conference Tutorials

[2] **Scalable Construction and Querying of Massive Knowledge Bases**
Xiang Ren, Yu Su, Pedro Szekely, Xifeng Yan
The Web Conference (formerly **WWW**), 2018

[1] **Construction and Querying of Large-Scale Knowledge Bases**
Xiang Ren, Yu Su, Xifeng Yan
26th ACM International Conference on Information and Knowledge Management (**CIKM**), 2017

Publications

Refereed Conference Papers

[20] **How Large A Vocabulary Does Text Classification Need? A Variational Approach on Vocabulary Selection**
Wenhu Chen, Yu Su, Yilin Shen, Zhiyu Chen, Xifeng Yan and William Yang Wang
Annual Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL**), 2019

[19] **Learning to Compose Topic-Aware Mixture of Experts for Zero-Shot Video Captioning**
Xin Wang, Jiawei Wu, Da Zhang, Yu Su, William Yang Wang
AAAI Conference on Artificial Intelligence (**AAAI**), 2019

[18] **Concept Mining via Embedding**
Keqian Li, Hanwen Zha, Yu Su, Xifeng Yan
IEEE International Conference on Data Mining (**ICDM**), 2018

[17] **XL-NBT: A Cross-lingual Neural Belief Tracking Framework**
Wenhu Chen, Jianshu Chen, Yu Su, Xin Wang, Dong Yu, Xifeng Yan and William Yang Wang
23th Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2018

[16] **What It Takes to Achieve 100% Condition Accuracy on WikiSQL**
Semih Yavuz, Izzeddin Gur, Yu Su, Xifeng Yan
23th Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2018

[15] **DialSQL: Dialogue Based Structured Query Generation**
Izzeddin Gur, Semih Yavuz, Yu Su, Xifeng Yan
56th Annual Meeting of the Association for Computational Linguistics (**ACL**), 2018
(Acceptance rate: 258/1018=25.3%)

[14] **Natural Language Interfaces with Fine-Grained User Interaction: A Case Study on Web APIs**
Yu Su, Ahmed H. Awadallah, Miaosen Wang, Ryen White
41st International ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR**), 2018
(Acceptance rate: 86/409=21.0%)

[13] **Global Relation Embedding for Relation Extraction**
Yu Su*, Honglei Liu*, Semih Yavuz, Izzeddin Gur, Huan Sun, Xifeng Yan (*: equal contribution)
16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (**NAACL-HLT**), 2018
(Acceptance rate: 207/665=31.1%)

[12] **Unsupervised Neural Categorization for Scientific Publications**
Keqian Li, Hanwen Zha, Yu Su, Xifeng Yan
18th SIAM International Conference on Data Mining (**SDM**), 2018
(Acceptance rate: 87/375=23.2%)

[11] **Building Natural Language Interfaces to Web APIs**
Yu Su, Ahmed H. Awadallah, Patrick Pantel, Michael Gamon, Madian Khabsa, Mark Encarnacion
26th ACM International Conference on Information and Knowledge Management (**CIKM**), 2017
(Acceptance rate: 171/820=20.9%)

[10] **Cross-domain Semantic Parsing via Paraphrasing**
Yu Su, Xifeng Yan
22th Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2017
(Acceptance rate: 216/836=25.8%)

[9] **An End-to-End Deep Framework for Answer Triggering with a Novel Group-Level Objective**
Jie Zhao, Yu Su, Ziyu Guan, Huan Sun
22th Conference on Empirical Methods in Natural Language Processing (**EMNLP**), short, 2017
(Acceptance rate: 107/582=18.4%)

[8] **Recovering Question Answering Errors via Query Revision**
Semih Yavuz, Izzeddin Gur, Yu Su, Xifeng Yan
22th Conference on Empirical Methods in Natural Language Processing (**EMNLP**), short, 2017
(Acceptance rate: 107/582=18.4%)

[7] **On Generating Characteristic-rich Question Sets for QA Evaluation**
Yu Su, Huan Sun, Brian Sadler, Mudhakar Srivatsa, Izzeddin Gur, Zenghui Yan, Xifeng Yan
21th Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2016
(Acceptance rate: 177/687=25.8%)

[6] **Improving Semantic Parsing via Answer Type Inference**
Semih Yavuz, Izzeddin Gur, Yu Su, Mudhakar Srivatsa, Xifeng Yan
21th Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2016
(Acceptance rate: 177/687=25.8%)

[5] **A Fast Kernel for Attributed Graphs**
Yu Su, Fangqiu Han, Richard E. Harang, Xifeng Yan
16th SIAM International Conference on Data Mining (**SDM**), 2016
(Acceptance rate: 96/371=25.9%)

[4] **Table Cell Search for Question Answering**
Huan Sun, Hao Ma, Xiaodong He, Wen-Tau Yih, Yu Su, Xifeng Yan
25th International Conference on World Wide Web (**WWW**), 2016
(Acceptance rate: 115/727=15.8%)

[3] **Visual Graph Query Formulation and Exploration: A New Perspective on Information Retrieval at the Edge**
Sue Kase, Michelle Vanni, Joanne Knight, Yu Su, Xifeng Yan
SPIE 9851, Next-Generation Analyst IV (**SPIE Defense+Security**), 2016

[2] **Exploiting Relevance Feedback in Knowledge Graph Search**
Yu Su, Shengqi Yang, Huan Sun, Mudhakar Srivatsa, Sue Kase, Michelle Vanni, Xifeng Yan
21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD**), 2015
(Acceptance rate: 160/819=19.5%)

[1] **On the Validity of Geosocial Mobility Traces**
Zengbin Zhang, Lin Zhou, Xiaohan Zhao, Gang Wang, Yu Su, Miriam Metzger, Haitao Zheng, and Ben Y. Zhao
12th ACM Workshop on Hot Topics in Networks (**HotNets**), 2013
(Acceptance rate: 26/110=23.6%)