Xiang Deng

http://web.cse.ohio-state.edu/~deng.595

Contact Information

Department of Computer Science and Engineering
The Ohio State Univeristy

Email: deng.595@osu.edu
GitHub: xiang-deng

Research Interests

My research interests lie in NLP and data mining, with emphasis on knowledge discovery and utilization from heterogeneous sources. The aim is to *build AI-powered data systems that can assist with information acquisition and decision making for regular users as well as domain experts* in Digital Era. Specifically, my recent research focuses on the following directions: (1) **Large-scale pretraining and representation learning** for data from heterogeneous sources, such as plain text, web documents, knowledge graphs, and databases; both for general and domain-specific applications. (2) **Building natural language interfaces** (e.g., question answering, semantic parsing, and dialog systems) with varied data and services as backend.

Education

The Ohio State University, Columbus, OH — 2018 - 2023 (expected)

Studying towards Ph.D., Computer Science and Engineering

- Advisor: Prof. *Huan Sun*
- Major: Artificial Intelligence; Minor: Database, Graphics

University of Science and Technology of China, Hefei, China — 2014 - 2018

B.Eng. Computer Science, School of The Gifted Young

- The Talent Program in Computer and Information Science

Professional Experience

The Ohio State University — Aug 2018 - present
Graduate Research Associate
Supervisor: Prof. *Huan Sun*

- Prompting and reasoning with Large Language Models for solving complex tasks.
- Building dialog system that can assist users in accomplishing tasks. (OSU Tacobot team, ranked 3^{rd} place in the inaugural Alexa Prize TaskBot Challenge.)
- Textual and tabular data understanding via pre-training. (*VLDB*'21, *EMNLP*'21, *SIGMOD Research Highlight*'22. Collaboration with Google Research under Google Faculty Research Award.)
- Relation Extraction with extra signals from Web Tables. (*EMNLP*'19)
- Question answering and tabular query resolution with Knowledge Base.

Google — May 2022 - Aug 2022
Research Intern, Google Research NYC
Supervisor: *Vasilisa Bashlovkina**, *Feng Han*, *Simon Baumgartner*

- Financial market sentiment anlysis on social media content.

Amazon — May 2021 - Aug 2021
Applied Scientist Intern, Product Graph
Supervisor: *Prashant Shiralkar**, *Colin Lockard*, *Binxuan Huang*

- Robust and generalizable representation learning for semi-structured web pages. The resulting model brings significant improvement under zero-shot and few-shot settings, which greatly reduces human annotation efforts.

Microsoft Research May 2020 - Aug 2020
Research Intern, Natural Language Processing Group
Supervisor: *Matthew Richardson**, *Ahmed Awadallah*, *Christopher Meek*, *Oleksandr Polozov*

- Natural Language to SQL, with a focus on generalization ability in real-world applications. By weakly supervised pretraining using existing text-table parallel data on the web, we enhance the model's performance on value prediction and column selection, especially when the access to actual database content is limited at runtime. (*NAACL*'21)

Microsoft Research Asia Dec 2017 - May 2018
Research Intern, Natural Language Computing Group
Supervisor: *Lei Cui*

- News Recommendation and Summarization.

The Ohio State University July 2017 - Nov 2017
Visiting Student
Supervisor: Prof. *Huan Sun*

- Question Answering over Knowledge Base with Web Information.

University of Science and Technology of China Feb 2017 - June 2017
Undergraduate Research Dr. *Qi Liu*

- Representation Learning for Complex Network.

Publications

Bootstrapping a User-Centered Task-Oriented Dialogue System
Shijie Chen, Ziru Chen, **Xiang Deng**, Ashley Lewis, Lingbo Mo, Samuel Stevens, Zhen Wang, Xiang Yue, Tianshu Zhang, Yu Su, Huan Sun
Alexa Prize Proceedings, 2022
3^{rd} **place in the Alexa Prize TaskBot Challenge**

Shepherd Pre-trained Language Models to Develop a Train of Thought: An Iterative Prompting Approach
Boshi Wang, **Xiang Deng**, Huan Sun
arXiv preprint, 2022

DOM-LM: Learning Generalizable Representations for HTML Documents
Xiang Deng, Prashant Shiralkar, Colin Lockard, Binxuan Huang, Huan Sun
arXiv preprint, 2022

ReasonBERT: Pre-trained to Reason with Distant Supervision
Xiang Deng, Yu Su, Alyssa Lees, You Wu, Cong Yu, and Huan Sun
Conference on Empirical Methods in Natural Language Processing, (**EMNLP**), 2021

Structure-Grounded Pretraining for Text-to-SQL
Xiang Deng, Ahmed Hassan Awadallah, Christopher Meek, Oleksandr Polozov, Huan Sun, and Matthew Richardson
Annual Conference of the North American Chapter of the Association for Computational Linguistics, (**NAACL**), 2021

TURL: Table Understanding through Representation Learning
Xiang Deng, Huan Sun, Alyssa Lees, You Wu, and Cong Yu
International Conference on Very Large Data Bases, (**VLDB**), 2021
SIGMOD Research Highlight, 2022

Leveraging 2-hop Distant Supervision from Table Entity Pairs for Relation Extraction
Xiang Deng, Huan Sun
Conference on Empirical Methods in Natural Language Processing, (**EMNLP**), 2019

Easy-to-Hard: Leveraging Simple Questions for Complex Question Generation
Jie Zhao, **Xiang Deng**, Huan Sun
arXiv preprint, 2019

Automatic Table completion using Knowledge Base
Bortik Bandyopadhyay, **Xiang Deng**, Goonmeet Bajaj, Huan Sun, and Srinivasan Parthasarathy
arXiv preprint, 2019

Awards

3rd place in the Alexa Prize TaskBot Challenge, Amazon	2022
SIGMOD Research Highlight, *SIGMOD*	2022
Student Travel Award, *KDD 2019*	July 2019
Student Scholarship, *USTC*	2015, 2016, 2017
Freshman Scholarship, *USTC*	2014

Professional Service

Program Committee/Reviewer:
ACL ARR, SUKI 2022, NLP4Prog 2021; NLPCC 2020, 2021, 2022
Secondary/External Reviewer:
KDD 2020; NAACL 2019; KDD 2019

Teaching Experience

Syllabus of Digital Logic Lab, *USTC* Fall, 2016
Teaching Assistant

Programming Languages and Frameworks

Python, PyTorch, Tensorflow, Spark, C++, Java, SQL