Xindi Wang

Email: xindi.w1993@gmail.com; Website: http://www.wangxindi.org

EDUCATION

Ph.D in Network Science, Northeastern University, Boston, MA   2015-2021
- GPA: 4.0/4.0; Advisor: Prof. Albert-László Barabási and Prof. Tina Eliassi-Rad
- Coursework: Data Mining Techniques, Computational Statistics, Bayesian & Network Statistics, Complex Networks and Application, Dynamic Processes on Complex Networks, Network Science Data

B.Eng, University of Electronic Science and Technology of China (UESTC), Chengdu, China   2011-2015
- Electronic Engineering and Computer Science. GPA: 3.93/4.0

SKILLS

- Machine Learning: Natural Language Processing, Information Retrieval, Deep Learning, Graph Mining
- Data mining and analysis, Statistics, Experience with large scale data, Extensive experience with graph data
- Programming Languages: Python (primer language), R, C, Matlab

EXPERIENCE

Applied Scientist II, **Alexa AI**, **Amazon,** Boston, MA   October 2021 - November 2022
- Projects: Ranker-based Entity Exploration Model for Entity Resolution, Graph-based Data Augmentation for Entity Resolution
- Skills involved: Natural Language Processing, Deep Learning, Information Retrieval, Graph Mining, Tensorflow

Applied Scientist Intern, **Alexa AI**, **Amazon,** Boston, MA   Fall 2020
- Project: Cross-query ranker on ASR N-best
- Skills involved: Natural Language Processing, Deep Learning, Information Retrieval, Python

Applied Scientist Intern, **Search**, **Amazon**, Boston, MA   Summer 2019
- Project: Entity linking on customer reviews and queries
- Skills involved: Natural Language Processing, Information Retrieval, Spark, Python

Research Assistant, **CCNR**, **Northeastern University**, Boston, MA   2015-2021
- Projects: Success of Books and Authors, Gender Inequality in Visual Art, Fairness in Machine Learning
- Skills involved: Machine Learning, Data Mining, Algorithm Design, Graph Mining, Statistics, Python

Computational Social Science Summer School, **University of Southern California**, Los Angeles, CA   Summer 2018
- Project on Cyberbullying in Instagram and won the "Best Project Award"
- Skills involved: Natural Language Processing, Sentiment Analysis, Python

Complex Systems Summer School, **Santa Fe Institute**, Santa Fe, NM   Summer 2018
- A highly selective four week intensive summer school with lectures and project
- Projects: Understanding Music with Higher Order Network, Singapore City Data Analysis
- Skills involved: Network Analysis, Data Mining, Data Processing, Machine Learning, Python, Gephi

SELECTED PROJECTS

Ranker-based Entity Exploration Model for Entity Resolution
- Lead, design and develop a ranker-based entity exploration model for entity resolution
- Applied the model on a use case with more than 500K weekly traffic. Through online A/B test and offline analysis, demonstrate an improvement of 5.04% comparing to the current production system

Graph-based Data Augmentation for Entity Resolution
- Design and develop graph-based data augmentation method for entity resolution for a use-cases with more than 500K traffic weekly to improve robustness to upstream ASR (Automatic Speech Recognition) error.
- Experiment achieves 5.19% improvement on accuracy overall, and 27.86% improvement on harder cases.

Cross-query Ranker on ASR N-best for Entity Resolution

- Develop a machine learning ranker to leverage results from upstream ASR (Automatic Speech Recognition) to make the Entity Resolution result robust to ASR errors
- Experiment the ranker on two use-cases with 100k examples and 1 million examples separately, and achieved about 10% gain in accuracy.

Entity linking on Customer Reviews and Queries
- Using natural language processing and learning to rank method, developed an entity linking system using wikipedia data on customer queries and reviews.
- Designed evaluation method on both wikipedia data and collected Mechanical Turk labeled data, and achieved about 20% improvement comparing to baseline.

Success of Books and Authors
- Using various datasource, utilizing machine learning and data mining techniques to understand how books and authors become successful (more than 20,000 books involved)
- Developed a machine learning algorithm *Learning to Place* for heavy-tailed attribute data prediction to predict the book sales prior to its publication, outperformed traditional algorithm such as linear regression by about 20% for high-selling books

Quantifying Systemic Gender Inequality in Visual Art
- Using artist exhibition and auction data, revealing gender inequality and explaining causes of gender inequality in the art world with statistical method and machine learning

Information Access Equality on Generative Models of Complex Networks
- Developed network generative models to understand information access equality of nodes with simulation. Revealed key features of the network that would promote or harm equality.

PUBLICATIONS AND TALKS

- Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information access equality on generative models of complex networks." *Applied Network Science* 7, no. 1 (2022): 1-20.
- Marton, Rebecca M., Xindi Wang, Albert-László Barabási, and John Ioannidis. "Science, advocacy, and quackery in nutritional books: an analysis of conflicting advice and purported claims of nutritional best-sellers." *Palgrave Communications* 6, no. 1 (2020): 1-6.
- Wang, Xindi, Burcu Yucesoy, Onur Varol, Tina Eliassi-Rad, and Albert-László Barabási. "Success in books: predicting book sales before publication." *EPJ Data Science* 8, no. 1 (2019): 1-20.
- Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "L2P: an algorithm for estimating heavy-tailed outcomes." *arXiv preprint arXiv:1908.04628* (2019).
- Yucesoy, Burcu, Xindi Wang, Junming Huang, and Albert-László Barabási. "Success in books: a big data approach to bestsellers." *EPJ Data Science* 7 (2018): 1-25.
- Wang, Xindi, Alex J. Gates, and Albert-László Barabási, "Quantifying systemic gender inequality in visual art." *Nature Communications* (Under review).

TALKS

- Quantifying data bias in U.S. justice system with affinity networks, *International Conference on Network Science*, 2019
- Learning to place objects, *International Conference on Complex Networks*, 2018
- Success of books and authors, *The Central Winter Conference on Network Science*, 2018

HONORS AND AWARDS

- Best Project Award at USC/ISI 2018 Computational Social Science Summer School — 2018
- Outstanding Student of University of Electronic Science and Technology of China (UESTC) — 2014
  - *Top 10 out of 5500 senior students of UESTC*
- National Scholarship (Highest scholarship given by Ministry of Education of China, top 0.1%) — 2014
- Honorable Mention of Interdisciplinary Contest in Modeling (ICM), sponsored by SIAM, NSA and INFORMS — 2014
- Tang Lixin Scholarship — 2013
  - *Top 50 out of 25000 undergraduate and graduate students of UESTC*
- First Prize in Contemporary Undergraduate Mathematical Contest in Modeling of China — 2013