Visualization and Data Analytics Research Center (VIDA)
370 Jay Street
Brooklyn, NY, 11201
ky630@nyu.edu | keyang0923@gmail.com
Webpage : keyang0923.github.io
+1-215-900-5854

Ke Yang

PhD Candidate, Computer Science, New York University

Education

Ph.D. in Computer Science, New York University, New York, U.S.A
Advisor: Julia Stoyanovich *January, 2019 - January, 2021 (expected)*
GPA 4.0/4.0

Ph.D. in Computer Science, Drexel University, Philadelphia, U.S.A
Advisor: Julia Stoyanovich *September, 2015 - December, 2018*
GPA 3.98/4.0

M.S.E in Computer Science, Beijing Technology and Business University, Beijing, China
Advisors: Qian Mo and Zhongming Han *September, 2012 - June, 2015*
Graduated summa cum laude, GPA 3.51/4.0

B.Eng. in Software Engineering, Beijing Technology and Business University, Beijing, China
Advisor: Qian Mo *September, 2008 - June, 2012*
Graduated summa cum laude, GPA 3.45/4.0

Research Interests

Responsible data management, non-discriminatory machine learning, and data ethics
My research focuses on several quality aspects of algorithmic outcome and how to quantify and ensure them in different scenarios. The quality aspects include fairness, diversity, interpretability, and stability of algorithmic outcome. I develop algorithms and systems to ensure high-quanlity outcome outputted from a data-driven system.
Current project focuses on quantifying fairness in rankings and interpretability of algorithmic outcome.

Research Projects

Data, Responsibly
Supervisor: Prof. Julia Stoyanovich *December, 2015 - Present*

- Proposed fairness measures in ranking-related applications and deveoped models to mitigate bias against minorities such as women in technical community and African-American people in risk assesment. More details in dataresponsibly.github.io.
- Proposed algorithms to ensure diverse outcomes for online set selection tasks and to reduce intersectional effect caused by diversity requirements on multiple sensitive attributes in the task of set selection.
- Help users to interpret algorithmic outcome for ranking-related applications and developed an online tool - RankingFacts.

Detect spammers in online shopping website
Supervisors: Prof. Zhongming Han and Qian Mo *May, 2013 - May, 2015*

- Extracted demographic and behavioral features from large group of spammers in online shopping websites, such as Taobao.com and JD.com.
- Proposed a graph model modeling users interactions among group members to efficiently detect the large spamming groups in online shopping website.
- Published a survey paper in Chinese Journal of Software, one of the Chinese premier peer-reviewed journals in Computer Science.
- Published a paper in Chinese Journal of Computers, one of the Chinese premier peer-reviewed journals in Computer Science.
- Published a China invention patent for above methodology.

Detect and predict hot topics in social network through analyzing human behavior pattern

Supervisor: Prof. Zhongming Han *December, 2012 - June, 2013*

- Extracted demographic and behavioral features from Sina Weibo spamming users.
- Proposed a new probabilistic graphical model that out-performs the baseline SVM model.
- Published a paper in Journal of Embedded System.

Publications

Ke Yang, Vasilis Gkatzelis, and Julia Stoyanovich, "Balanced ranking with diversity constraints", in *Proceedings of the 28th International Joint Conference on Artificial Intelligence* (IJCAI 2019), Macao, China, DOI:10.24963/ijcai.2019/836

Ke Yang, Julia Stoyanovich, Abolfazl Asudeh, Bill Howe, HV Jagadish, and Gerome Miklau, "A Nutritional Label for Rankings" (demo) to appear in *Proceedings of the 2018 ACM International Conference on Management of Data* (SIGMOD 2018), Houston, USA

Julia Stoyanovich, **Ke Yang**, HV Jagadish, "Online Set Selection with Fairness and Diversity Constraints" in *Proceedings of the 21th International Conference on Extending Database Technology* (EDBT 2018), Vienna, Austria, DOI:10.5441/002/edbt.2018.22

Ke Yang, Julia Stoyanovich, "Measuring Fairness in Ranked Outputs" in *Proceedings of the 29th International Conference on Scientific and Statistical Database Management* (SSDBM 2017), Chicago, USA, DOI:10.1145/3085504.3085526

Zhongming Han, **Ke Yang**, Xusheng Tan, "Analyzing Spectrum Features of Weight User Relation Graph to Identify Large Spammer Groups in Online Shopping Websites" in *Chinese Journal of Computers*, 40(4): 939-954 (2017) (in Chinese), DOI:10.11897/SP.J.1016.2017.00939

Zhongming Han, **Ke Yang**, Fengmin Xu, Dagao Duan, "Probabilistic Graphical Model for Detecting Spammers in Microblog Websites" in *Journal of Embedded System*, 8(1): 12-23 (2016), DOI:10.1504/IJES.2016.073747

Qian Mo, **Ke Yang**, "Overview of Web spammer detection" in *Journal of Software*, 25(7): 1505-1526 (2014) (in Chinese), DOI:10.13328/j.cnki.jos.004617

Kai Wang, **Ke Yang**, "A method and application system of detecting web spammers" in *China Invention Patent*, (2015) (in Chinese), No. CN201510012860

Teaching Experience

Teaching Assistant at Drexel University
- CS500 Fundamentals of Databases *January, 2017 - March, 2017*
- CS461 Database Systems *March, 2017 - May, 2017*

Work Experience

Research intern at AT&T Lab Research (New York Office) *June, 2019 - August, 2019*

Supervisors: Emily Dodwell, Ritwik Mitra, and Balachander Krishnamurthy
Worked on a project to insure fairness and transparency for internal AT&T machine learning projects. This project aims to ensure that the outcome of a machine learning project meets reasonable guidelines for fairness and does not include illegal bias and discrimination. During the internship, we developed a tool for project managers and data scientists to check whether there are bias-related issues through the whole pipeline of machine learning.

Research engineer at Elite & Resource (start-up company) *November, 2014 - August, 2015*
Supervisor: Peng Sun
Worked on a project to analyze historical flood disaster data and to predict a specific characteristic that can help with the prevention of future flood. In this project, we proposed a model to predict the water flow rate of small watershed torrents that can be used in the prevention of future flood. Our model is integrated as a core component of national watershed data management system.

Software Projects

Ranking Facts: full-stack development using Python, Django, JavaScript, and JQuery.
Data Synthesizer: front-end development using JavaScript and JQuery.
Researcher Questionnaire: database design and front-end development using PHP, Drupal, and MySql.
DataResponsibly: website development using Jekyll.

Extra Interests

Start-up company co-founder (*since September 2018*): work as technical consultant for a start-up company that aims to preserve cultural heritage through collecting and presenting information about heritage-preserving services and products.
Hobbies: an amateur marathon runner and an enthusiastic language learner.