Zhengjie Miao

D339 LSRC Building, 308 Research Drive, Durham, NC 27708

✉ zjmiao@cs.duke.edu 📞 +1 (919) 660-6594 ⧉ http://www.cs.duke.edu/~zjmiao

EDUCATION		
	Duke University, Durham, NC, U.S.A	
	▪ Ph.D. in Computer Science	Aug 2017 – Present
	• Advisor: Prof. Sudeepa Roy	
	• Research Initial Project: Explaining Wrong Queries Using Small Examples	
	• GPA: 3.90 / 4.00	
	Columbia University, New York, NY, U.S.A	
	▪ M.S. in Computer Science	Aug 2015 – Dec 2016
	• GPA: 4.00 / 4.00	
	Peking University, Beijing, P.R. China	
	▪ B.S. in Computer Science and Technology	Sep 2011 – Jul 2015
	• GPA: 3.50 / 4.00	

RESEARCH INTERESTS

Databases, Data Provenance, Natural Language Processing, Visual Analytics

AWARDS & HONORS

- ▪ 2020 PhD Fellowship Finalist, Microsoft Research — Nov 2019
- ▪ Outstanding Ph.D. Research Initiation Project Award, Department of Computer Science, Duke University — Sep 2019
- ▪ VLDB Travel Grant Recipient — Aug 2019
- ▪ SIGMOD Travel Award — Jun 2019, Jun 2018
- ▪ 7th Place in ACM/ICPC 2015 Greater New York Regional — Nov 2015
- ▪ Award for Excellent Detailed Analysis, VAST Challenge Mini-Challenge 1 — Nov 2014
- ▪ Award for Excellent Comprehensive Visual Analysis System, VAST Challenge Mini-Challenge 2 — Nov 2014
- ▪ Award for Excellent Collaborative Streaming Analysis, VAST Challenge Mini-Challenge 3 — Nov 2014
- ▪ The May Fourth Scholarship, Peking University — May 2013
- ▪ Silver medal in ACM/ICPC 2012 Asia Regional Contest in Tianjin — Oct 2012
- ▪ First Prize in National Olympiad in Informatics in Hunan Province — Nov 2009

PUBLICATIONS

I-Rex: An Interactive Relational Query Explainer for SQL Link to 🗎
Zhengjie Miao, Tiangang Chen, Alexander Bendeck, Kevin Day, Sudeepa Roy, Jun Yang
in *Proceedings of the VLDB Endowment (PVLDB), Vol 13, Demonstration Track*, Aug 2020

Snippext: Semi-supervised Opinion Mining with Augmented Data Link to 🗎
Zhengjie Miao, Yuliang Li, Xiaolan Wang, Wang-Chiew Tan
in *The Web Conference (WWW) 2020, Oral Presentation*, Apr 2020

CAPE: Explaining Outliers by Counterbalancing Link to 🗎
Zhengjie Miao[*], Qitian Zeng[*], Chenjie Li, Boris Glavic, Oliver Kennedy, and Sudeepa Roy
in *Proceedings of the VLDB Endowment (PVLDB), Vol 12, Demonstration Track*, Aug 2019
([*] denotes equal contribution)

LensXPlain: Visualizing and Explaining Contributing Subsets for Aggregate Query Answers Link to 📄

Zhengjie Miao, Andrew Lee, and Sudeepa Roy

in *Proceedings of the VLDB Endowment (PVLDB), Vol 12, Demonstration Track*, Aug 2019

Explaining Wrong Queries Using Small Examples Link to 📄

Zhengjie Miao, Sudeepa Roy, and Jun Yang

in *ACM SIGMOD International Conference on Management of Data (SIGMOD)*, Jun 2019

Going Beyond Provenance: Explaining Query Answers with Pattern-based Counterbalances Link to 📄

Zhengjie Miao*, Qitian Zeng*, Boris Glavic, and Sudeepa Roy

in *ACM SIGMOD International Conference on Management of Data (SIGMOD)*, Jun 2019

(* denotes equal contribution)

RATest: Explaining Wrong Relational Queries Using Small Examples Link to 📄

Zhengjie Miao, Sudeepa Roy, and Jun Yang

in *ACM SIGMOD International Conference on Management of Data (SIGMOD), Demonstration Track*, Jun 2019

Combining Design and Performance in a Data Visualization Management System Link to 📄

Eugene Wu, Fotis Psallidas, Zhengjie Miao, Haoci Zhang, Laura Rettig, Yifan Wu, Thibault Sellam

in *Conference on Innovative Data Systems Research (CIDR)*, Jan 2017

RESEARCH EXPERIENCE

Database Research Group, Duke University   Aug 2017 – present

Research Assistant, advised by Prof. Sudeepa Roy and Prof. Jun Yang

- **Helping Novices Learn and Debug Relational Queries**   ⧉[Project website]
  - Designed and implemented web-based debugging tools for Relational Algebra and SQL, which find a small counterexample for two input queries where the input queries return different results, and allow syntax-consistent tracing for the query execution.
  - Designed and implemented algorithms to find the smallest counterexample using data provenance and SMT solvers
  - Designed algorithms for generating general explanations on the semantic difference of the queries
  - Mentored a group of graduate and undegraduate students on designing and implementing features of our tools

- **Explaining Surprising Query Answers Using Patterns**
  - Designed the framework that provides explanations for surprising outcomes of an aggregate query by finding patterns and outliers in the data
  - Formalized the concept of aggregate regression patterns and the definition of counterbalancing explanations using aggregate regression patterns
  - Designed and implemented the explanation generating algorithm

Megagon Labs   May 2020 – Aug 2020

Research Intern, supervised by Dr. Yuliang Li and Dr. Wang-Chiew Tan

- Designed and implemented a meta-learned data augmentation framework for sequence classification tasks (text classification, entity matching, error detetion, etc.) based on pre-trained language models
- Proposed the optmization that enables the model to learn how to choose and combine augmented data
- Research paper submitted to a top-tier conference

Megagon Labs   May 2019 – Aug 2019

Research Intern, supervised by Dr. Yuliang Li and Dr. Wang-Chiew Tan

- Studied problems on data augmentation and semi-supervised learning for aspect-based sentiment analysis
- Designed and implemented MixDA, a new data augmentation operator for text data

Database Research Group, Columbia University   Sep 2015 – Dec 2016

Research Assistant, advised by Prof. Eugene Wu

- **Data Visualization Management System**
  - Designed an aggressive prefetching system to improve the response time of networked data visualizations, by allowing the server to push anticipated data to the client
  - Built a predictive model for user mouse interations on web-browsers and implemented a JavaScript library for the model

Visualization and Visual Analytics Group, Peking University   Apr 2014 – May 2015

Research Assistant, advised by Prof. Xiaoru Yuan

- **Visual Analytics in Ancient Chinese Literature**
  - Researched topic on text visualization and collaborative sense making
  - Designed and implemented the interface for viewing and annotating entities and events in the literature, and visualized the storylines using D3.js and Python
- **Store Explorer: A visual analysis tool for heterogeneous text data**
  - Designed and implemented a flexible filter and a visual analytics interface for text dataset
  - Built up the Story Explorer system for analyzing social network and behavior log information

WORK EXPERIENCE	**Yahoo Beijing R&D Center**, Beijing, P.R.China *Technical Intern, MUX Group*	Jul 2014 – Oct 2014

TEACHING EXPERIENCE

Teaching Assistant, Introduction to Database Systems (Duke CompSci 316)   Spring 2019
- Assisted in writing and grading the assignments and projects; deployed our RATest tool

Teaching Assistant, Everything Data (Duke CompSci 216)   Spring 2018
- Assisted in writing and grading the assignments, labs, and projects.

SERVICES
- External Reviewer, ACM Transactions on Database Systems (TODS)
- Student Mentor of CS+: CompSci Projects Beyond the Classroom, Duke University   2020
- Student Volunteer, ACM SIGMOD   2020
- Reproducibility Committee Member, pVLDB   2019