

WEIYUAN WU

youngw@sfu.ca · wooya.me · github.com/dovahcrow

EDUCATION

Ph.D. Candidate in Database, Simon Fraser University, Canada Supervisor: Dr. Jiannan Wang	Sep. 2019 -
M.Sc. in Database, Simon Fraser University, Canada Thesis: "Enabling SQL-ML Explanation to Debug Training Data" Supervisor: Dr. Jiannan Wang	May 2017 - Sep. 2019
B.S. in Computing Science, UESTC, China	Sep. 2012 - July 2016

RESEARCH INTERESTS

Machine Learning Interpretation, SQL Explanation, Data Science Tooling

PROFESSIONAL EXPERIENCE

Project Leader <i>ConnectorX: The fastest way to load data into Python</i> Architecture design and implementation	Jan 2021 - github.com/sfu-db/connector-x
Project Leader <i>Dataprep: Data Preparation in Python</i> Project management and development	May 2019 - dataprep.ai
External Researcher <i>Vancity, Canada</i> Data augmentation, churn prediction and text mining	Jan 2018 - Sep. 2019
Data Scientist <i>Strikingly Inc., China</i> Churn prediction, anti spammer and data warehouse	June 2016 - May 2017

PUBLICATIONS

<i>Yejia Liu*, Weiyuan Wu*, Lampros Flokas, Jiannan Wang, Eugene Wu:</i> Enabling SQL-based Training Data Debugging for Federated Learning	VLDB 2022
<i>Brandon Lockhart, Jinglin Peng, Weiyuan Wu, Jiannan Wang, Eugene Wu:</i> Explaining Inference Queries with Bayesian Optimization	VLDB 2021
<i>Jinglin Peng*, Weiyuan Wu*, Brandon Lockhart, Song Bian, Jing Nathan Yan, Linghao Xu, Zhixuan Chi, Jeffrey Rzeszutarski, Jiannan Wang:</i> DataPrep.EDA: Task-Centric Exploratory Data Analysis for Statistical Modeling in Python	SIGMOD 2021
<i>Xiaoying Wang*, Changbo Qu*, Weiyuan Wu*, Jiannan Wang†, Qingqing Zhou:</i> Are We Ready For Learned Cardinality Estimation?	VLDB 2021
<i>Weiyuan Wu, Lampros Flokas, Eugene Wu, Jiannan Wang:</i> Complaint-driven Training Data Debugging for Query 2.0	SIGMOD 2020
<i>Weiyuan Wu, Lampros Flokas, Eugene Wu, Jiannan Wang:</i> Towards Complaint-driven ML Workflow Debugging	MLOps 2020, Demo

SKILLS

Frameworks: Tensorflow, Pandas, Numpy, Scikit-Learn, Dask
Programming Languages: Python/Cython, Rust, C++, Typescript, SQL
Platforms: Docker, AWS, Solana