

Zhiying Xu

Cambridge, MA | zhiyingxu@g.harvard.edu | xuzhiying9510.github.io | (+1)857-253-8842

5+ years' experience in the intersection area of machine learning and systems:

- ML for traffic engineering systems (paper published with Microsoft in SIGCOMM'23)
- Accelerate optimization problems with neural network inference
- Transformers training and inference on network traces

EXPERIENCE

Research Assistant, **Harvard University**, Cambridge, MA 09/2018 — current

- Learning-accelerated Optimization for Resource Allocation Problems
 - Work on a scalable learning-augmented solution for resource allocation problems
 - Decouple resource and demand constraints with ADMM and then decompose into subproblems
 - Accelerate Euclidean projections in subproblems with neural network inference
- Modeling Network Dynamics with Generative Traffic Pre-training
 - Work on modeling network dynamics from ISP's network traces
 - Leverage LSTM to capture time-series network behaviors; Updated to transformer model later
 - Fine-tuning with survival analysis to make early predictions of various network anomalies
 - LSTM version in MLSys'20 workshop and CoNext'22; Transformer version under submission

Software Engineering Intern, **Google**, Sunnyvale, CA 07/2023 — 09/2023

- Learning-based Productivity Metrics for Heterogeneous Workloads
 - Work on productivity metrics for Spanner platforms under heterogeneous workloads
 - Automatically extract key features from Spanner event logs with machine learning
 - Build an interpretable model to normalize heterogeneous workloads for productivity comparison

Research Intern, **Microsoft Research**, Redmond, WA 06/2022 — 09/2022

- Learning-Accelerated Optimization of WAN Traffic Engineering
 - Work on scalable traffic engineering in large wide-area networks
 - Model traffic engineering with neural network inference to unlock massive parallelism on GPUs
 - Enforce capacity constraints on the neural networks' coarse solution using parallelizable ADMM
 - Accepted in SIGCOMM'23

EDUCATION

Harvard University Cambridge, MA
Ph.D. in Computer Science, Advisor: Minlan Yu 09/2018 — current

- Papers published in top system conferences. Click Google Scholar for details.

Shanghai Jiao Tong University Shanghai, China
B.S. in Information Engineering, GPA: 3.85/4.00, Rank: 2/169 09/2014 — 06/2018

SKILLS

Programming Python, C/C++, Verilog, VHDL
Tools PyTorch, TensorFlow, FPGA, MATLAB, LaTeX