Zicun Cong

Burnaby, B.C., Canada, V5A 0A8
+1 (604) 7825610 congzicun@gmail.com minzc.github.io github.com/Minzc

Education

Simon Fraser University **Jan. 2017 – July 2022**

Ph.D. in Computing Science, GPA: 4.0/4.33, Supervisor: Prof. Jian Pei

Thesis: "Towards Trustworthy Data Analytics: Algorithmic Tools for Interpretability and Fairness"

Work Experience

Fortinet Inc. **Sep. 2016 – Present**

Staff Software Engineer *Burnaby, Canada*

- Developed machine learning and data mining models in Python using Pandas, NumPy, Scikit-Learn, and PyTorch to detect malicious traffics protecting millions of endpoints and achieved an improvement up to 20% in accuracy
- Developed a system to detect malware binaries by hierarchical clustering and local feature selection on malware behavior logs with billions of features and achieve a 15% accuracy improvement over the legacy model
- Collaborated with cyberthreat experts to do feature engineering for malicious traffic and malware binary identification
- Applied PySpark, Hadoop, and HBaase to process and store billions of network traffic and malware behaviour logs
- Delivered a NLP service in Python to automatically extract open source threat intelligence from unstructured text

Simon Fraser University. **Jan. 2017 – July. 2022**

Research Assistant *Burnaby, Canada*

- Researched interpretable data analytic problems and proposed two post-hoc interpretation methods to interpret deep neural networks and statistical hypothesis tests with fidelity guarantees
- Researched scalable data pricing models and proposed an approach that not only guarantees the exactness of Shapley value computation, but also achieves faster computation by orders of magnitudes than the baselines

LinkedIn. **May. 2021 – Dec. 2021**

Machine Learning Engineer Intern *Remote*

- Implemented state-of-the-art bias mitigation approaches to debias Graph Neural Networks
- Researched the problem of unfairness in graph neural networks and proposed a sampling framework to improve fairness in graph neural networks with reinforcement learning
- Prepared and submitted a research paper

Datastory. **July 2013 – Sep. 2014**

Software Engineer *Guangzhou, China*

- Developed distributed computation infrastructures for large-scale data acquisition, storage, and analysis
- Applied Hadoop and Apache Storm to do batch and realtime analysis on billions of social network accounts

Publications

- **Zicun Cong**, Baoxu Shi, Shan Li, Jaewon Yang, Jian Pei, Qi He. "FairSample: Training Fair and Accurate Graph Neural Networks Efficiently." Under review.
- Xuan Luo, Jian Pei, **Zicun Cong**, Cheng Xu. "On Shapley Value in Data Assemblage Under Independent Utility." Proc. VLDB Endow. 15, 11 (2022), 2761–2773.
- **Zicun Cong**, Xuan Luo, Jian Pei, Feida Zhu, Yong Zhang. "Data Pricing in Machine Learning Pipelines." Knowledge and Information Systems (KAIS), 2022.
- Jian Pei, Feida Zhu, **Zicun Cong**, Xuan Luo, Huiwen Liu, Xin Mu. "Data Pricing and Data Asset Governance in the AI Era." In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD '21). Association for Computing Machinery, New York, NY, USA, 4058–4059.
- **Zicun Cong**, Lingyang Chu, Yu Yang, Jian Pei. "Comprehensible counterfactual explanation on Kolmogorov-Smirnov test." Proc. VLDB Endow. 14, 9 (May 2021), 1583–1596.
- **Zicun Cong**, Lingyang Chu, Lanjun Wang, Xia Hu, and Jian Pei. "Exact and Consistent Interpretation of Piecewise Linear Models Hidden behind APIs: A Closed Form Solution." In 2020 IEEE 36th International Conference on Data Engineering (ICDE), pp. 613-624. IEEE, 2020.

Technical Skills

Artificial Intelligence: Deep Learning, Machine Learning, Data Mining
Languages and Frameworks: Python, Java, C++, SQL, PyTorch, TensorFlow, Sklearn
Cloud and Distributed Computing: Hadoop, Spark, HBase, Phoenix db, Redis, Mysql, Zookeeper, Apache Storm