Ziyu Li

zli17nl@gmail.com +31 630983914 ziyuli_li

ziyuli.me zLizy

Personal Statement

Committed and dedicated, my passion lies in exploring the fusion of database and machine learning (DBML) realms. With a solid background in research and practical experience, I am deeply engaged in advancing impactful research endeavors. My current focus revolves around (meta)data management and representation, aiming to optimize machine learning workflows for heightened efficiency and effectiveness.

Experience

2024.2 - Now	**Postdoc researcher**	TU Delft	Delft, Netherlands
	› Web Information Systems (WIS) group		
2019 - 2024	**PhD researcher**	TU Delft	Delft, Netherlands
	› Web Information Systems (WIS) group		
2018.5 - 2018.8	**Data analyst intern**	Corbion	Amsterdam, Netherlands
2016.7 - 2016.9	**Android developer intern**	South China University of Technology	
	› Communication & Computer Network Lab of Guangdong		

Education

2019 - 2024	**PhD (Computer Science)**	**TU Delft**	Delft, Netherlands
	Research direction: Databases, Machine Learning		
	Supervisors: Prof. Alessandro Bozzon, Dr. Asterios Katsifodimos		
	Dissertation: On the Utility of Metadata to Optimize ML Workflows		
2017 - 2019	**Master (Data Intrinsic Track)**	**TU Delft**	Delft, Netherlands
	Thesis topic: Graph, Network embedding		
	Thesis: Diffusion Based Temporal Network Embedding for Link Prediction		
2013 - 2017	**Bachelor (Network Engineering)**	**South China University of Technology**	
	Guangzhou, China		

Projects

› **TransferGraph**: a novel framework that reformulates model selection as a **graph learning problem**. TransferGraph constructs a graph using extensive **metadata** extracted from models and datasets, while capturing their inherent relationships. We accumulated more than 100 real datasets, both images and texts, and over 300 open-source models for CV and NLP tasks.
Library: pytorch-geometric, transformers

- **Macaroni**: a tool designed and implemented to demonstrate the metamodel that represents the metadata of models and datasets in structured and queryable manner. Macaroni supports metadata search as well as automatic model evaluation, in total more than 800 model-dataset pairs are evaluated. **Toolkits: MongoDB, transformers**
- **ML inference query optimization**: we exploit metadata to optimize for ML inference queries, where predicates in the queries are ML models. User may impose constraints on multiple aspects, such as execution cost and accuracy performance. We formulate the problem as a Mixed Integer Programming (MIP) and the subsequent development of a MIP-based optimizer.

Competences & Languages

Data Analysis	R, Pandas, Numpy, Matplotlib, ggplot2, Keras, Scikit-learn
Graph	pytorch-geometric, NetworkX **Optimization** Gurobi
Tools	SSH, Git, Docker **Programming** Python, R
Framework/Library	PyTorch, TensorFlow, transformers
ML/DL	Computer Vision, Natural Language Processing, RAG
Language	English (fluent); Chinese (mother language); Dutch (A2)

Selected Publications

Conference

- **Ziyu Li**, Alessandro Bozzon, Rihan Hai, and et al. "Model Selection with Model Zoo via Graph Learning" *In 2024 IEEE 40th International Conference on Data Engineering (ICDE)*, IEEE, 2024.
- **Ziyu Li**, Mariette Schönfeld, Asterios Katsifodimos, and et al. "Optimizing ML Inference Queries Under Constraints." *In International Conference on Web Engineering*, pp. 51-66. Cham: Springer Nature Switzerland, 2023.
- **Ziyu Li**, Mariette Schönfeld, and Asterios Katsifodimos. "Optimizing machine learning inference queries for multiple objectives." *In 2023 IEEE 39th International Conference on Data Engineering Workshops (ICDEW)*, pp. 74-78. IEEE, 2023.

Journal

- **Ziyu Li**, Wenbo Sun, Rihan Hai, and et al. "Amalur: the convergence between data integration and machine learning." *IEEE Transactions on Knowledge and Data Engineering* (2024).
- **Ziyu Li**, Marco Brambilla, and Alessandro Bozzon, et al. "Metadata Representations for Queryable Repositories of Machine Learning Models." *IEEE Access* (2023).
- Xiuxiu Zhan, **Ziyu Li**, Masuda N., et al. "Susceptible-infected-spreading-based network embedding in static and temporal networks"[J]. EPJ Data Science, 2020, 9(1): 30.

Activities

2023.9 - 2024.5	Co-chair with Prof. Dr. Carsten Binnig and Dr. Gerardo Vitagliano at 3rd DBML workshop at International Conference on Data Engineering (ICDE)
2023.5	Invited as reviewer for IEEE Transactions on Parallel and Distributed Systems
2020 - 2024	Coached in total seven master students