Xiang (Emily) Zuo

220 Darryl Dr., Campbell, CA 95008, USA
+1 (408) 872-2860 | xiangzuo2012@gmail.com | www.linkedin.com/in/xiangzuo

Summary

Extensive industrial machine learning analytic and engineering experience. Strong backgrounds of architect design in machine learning/deep learning enterprise products. Skilled in **natural language processing (NLP), large-scale systems, and database**. Technical leadership on managing multiple projects and defining technical direction of emerging AI market.

Education

Ph.D. in Computer Science, University of South Florida, Florida (USF) 2010 - 2016
M.S. in Computer Science, Huazhong University of Science and Technology (HUST), China 2006 - 2009
B.S. in Computer Science, Huazhong University of Science and Technology (HUST), China 2000 - 2004

Work Experience

Teradata Santa Clara, CA
Staff Machine Learning Engineer March 2019 - Now

- Lead a team to develop a customer-facing **TeraAnalyticChat** that enables AI-based **conversations (Q&A)** and **SQL query generation** in database system, empowering users to easily use in-database analytic functions, all via natural language (Azure Open AI and Cognitive Search)
- Build a Teradata **ChatGPT-like** model by **fine-tuning** the **OpenLLaMA** with synthetically-generated data using retrieval augmented generation (RAG) and a search engine, combined with public and employee-created data. **Large language model (LLM)** techniques like **qLora, Adapter, PEFT** are used to enhance tuning and memory efficiency (PyTorch, Huggingface)
- Develop COVID-19 Q&A system by using **BERT**, intent classification, arima and elastic search by Tensorflow
- Lead (researched, designed, and developed) the project of GLMNet and NaiveBayes, ZTest, FTest, Chisquare, ANOVA, KNN, VectorDistance, SMOTE on **MapReduce** framework of Teradata platform (C/C++)
- Coach new hires and serve as a core interviewer for technical hiring

Teradata Santa Clara, CA
Senior Machine Learning Engineer Aug. 2016 - Feb. 2019

- Design and develop a suit of distributed machine-learning algorithms (e.g., XGBoost, Adaboost, Random Forest, TF-IDF, NaiveBayesTextClassifier, and Hashing Encodings) on MapReduce framework of Teradata Vantage (JAVA)
- Provide strong support for marketing, account, and professional service teams on solving all kinds of customers' issues
- Serve as **POC site manager** for enterprise customers including Verizon and AT&T

Distributed System & Social Computing Lab University of South Florida
Research Assistant Aug. 2010 - May. 2016
Design and implement algorithms and services for large-scale **social network** systems. Skills: distributed systems, characterization of social and peer-to-peer networks, **graph theory, data security**, and **cloud computing**

San Diego Supercomputer Center University of California San Diego
Research Intern Summer, 2014
Work on topics including big data computing, applied machine learning and statistical analysis of data using R

Human Computer Interaction Lab Nokia Research Center
Research Intern May - Aug. 2012
Implement a recommendation system based on user common contacts, similar interests, and physical encounters.

Skills

Machine Learning Tools and Libraries: **PyTorch, Tensorflow, HuggingFace, LangChain, Scikit-learn**
Programming Languages: Python, Java, C/C++, R, UNIX Shell
DBMS: Teradata Database, PostgreSQL
Application: MapReduce, GPU CUDA, ggplot
Tools: Vim, Bash, Docker, Kubernetes, Git

Honors Awards & Patents

"An Open-Source Alternative to ChatGPT", Teradata Hack Day **2nd** Prize (2nd among 83) June 2023
"Hyperparameter Tuner for Vantage Analytics", Teradata Hack Day **1st** Prize (Top 1 among 60) June 2021
"Hyper-Segment IDR" , **Teradata Patent** Dec 2020
"COVID-19 Q&A system based on **BERT**", Finalist in Teradata Hack Day (5 among 50) May 2020