FUBAO WU

Address: Easton, PA   +1 4138018291   fubaowu@gmail.com   windhaunting.github.io

CORE COMPETENCIES

- **Quantitative Skills:** Machine Learning, Deep Neural Networks, Generative AI, LLM, Reinforcement Learning, Statistics, Relational Database, Graph Database
- **Programming Languages**: Python, Scala, Java, C, C++, C#, SQL, Html, CSS, Matlab
- **Systems/Frameworks:** Spark, Airflow, Docker, Vertex AI, MLflow, BigQuery, Pytorch, Tensorflow, Scikit-learn, Hadoop, MySQL
- **Cloud Services:** Google Cloud Platform, Amazon Web Services
- **Soft skills:** Quick learner, Great Communication, Enjoying Collaboration, High Persistence and Curiosity.

PROFESSIONAL EXPERIENCE

Senior Machine Learning Engineer, Inventory Data Science Team, Walmart, Hoboken, NJ 02/2022 - present

- Leading the development of 3 projects of Sport Fan shop product retagging, and product title optimization, catalog attribute validation and accuracy, by prompting and finetuning through OpenAI, Gemini, LLaMA models, facilitating the merchant and e-commerce product management and enhancing customer experience.
- Developed a real-time machine learning model for order pickability prediction and improving the existing model for F1-score from 0.51 to 0.67.
- Developed an end-to-end pickability prediction pipeline at a daily batch level and rolled it out in several departments in 5000+ stores to reduce labor costs by 20% and improve nilpick rate by 3%.
- Collaborated with other data science and business teams to improve item buffer optimization and other use case analyses.

Research Assistant, Electrical and Computer Engineering, UMass Amherst   09/2014 - 12/2021

- Dissertation title: scalable data analytics for relational databases, graphs and videos. Published and submitted 3 first-authored papers at peer-reviewed conferences/journals.
- Developed a real-time video analytics system of object detection and tracking through movement estimation with a machine learning-based configuration adaptation on Pytorch, achieving average 6% higher accuracy and 12x speed-up compared with state-of-the-art approaches.
- Improved the graph query quality and efficiency with hierarchical inheritance relations and developed a distributed algorithm on Spark GraphX.
- Designed a meta path-based personalized query algorithm to find drugs in drug graph databases, improving accuracy by 8% compared to baselines.

Applied Research Intern, Bentley Systems, Watertown, CT   05/2018 - 08/2018

- Developed a diamond mining system fault detection system by devising statistics and classification methods on large multivariate time series data with 122 attributes and hundreds of millions of records.
- Achieved a fault detection test accuracy of 92% on a large dataset with millions of records and developed an integration tool with C#.
- Supported and improved team processes daily such as development of presentations and training materials.
- Integrated large unstructured to structured data sets into a big data system and provided data exploration, analysis, and visualization.

Applied Research Intern, Cisco Systems, San Jose, CA   06/2016 - 09/2016

- Yielded 93% accuracy of database matching for large network management databases through similarity score and classification methods over 600GB data.
- Assisted engineers for faster troubleshooting by remodeling product questions and solutions to incidents on graph data with clustering techniques in a distributed Spark environment.
- Designed, built and supported prescriptive models to enable self-service analytical capability for key technical stakeholders.
- Improved large-scale reporting systems that allow market and product managers to tailor, measure, and optimize campaigns.

Software Development Engineer, Huawei, Shanghai, China   07/2011-07/2013
- Enhanced reliability and functionality for large 2.5G/3G unified serving nodes, reducing the incident rate by about 30% through redesigning modules and structures.
- Designed new security protocols based on IPSec and authentication for 4G network security, improving reliability performance by about 8%.
- Fixed approximately 4 online network service incidents per week through diving deep into the code and collaborating with multiple testing departments.
- Conducted the SQL database design and implementation widely used for the security department, and maintained the database system.

OTHER PROJECTS

Efficient and Accurate Human Pose Estimation for Video Streams   (Python & Pytorch)
- Proposed a scheduling algorithm to dynamically select different DNN models for real-time human pose estimation.
- Implemented the pose estimation system, reducing computation resources by 48% with 2% accuracy loss compared to the most expensive model.

Group Fairness for Learning Representation on Variational Autoencoder   (Python & Pytorch)
- Investigated fairness and bias on a representation algorithm based on an variational autoencoder
- Explored the group fairness of the autoencoder on attributes such as gender, age, and race on three datasets, showing the different levels of biases on different attributes.

Facial Emotion Recognition In Images and Videos   (Python & Tensorflow)
- Proposed a LSTM model based on a pretrained CNN features to recognize 7 types of human facial emotions.
- Improved the accuracy by about 10% in different facial emotions compared to w/o pretrained CNN features on two common datasets.

DATA SCIENCE VOLUNTEERING

Data Scientist, Datakind, Remote   02/2022 - 09/2022
- Worked on visualizing the housing insecurity and created an interactive map displaying statistics of houses in Florida Counties.
- Built dashboards to automate the reports and prediction model for new client retention of Grameen America organization
- Worked on the black wealth gaps analysis for homeownership and home mortgage analysis.

Machine Learning Engineer, Omdena, Remote   08/2021 - 11/2021
- Constructing an end-to-end personal language recommender system for bilingual children with the Poikilingo customer.

EDUCATION

Ph.D. in Electrical and Computer Engineering   02/2022
University of Massachusetts Amherst, MA

M.Sc. in Computer Science   02/2021
University of Massachusetts Amherst, MA

M.Eng. in Electrical and Computer Engineering
University of Science and Technology of China (USTC), Hefei, China

B.Eng. in Electrical and Computer Engineering
Northeastern University (NEU), Shenyang, China