YIFAN ZHANG

EMAIL: IVANBUAA@HOTMAIL.COM
MOBIL: 0452208587

EDUCATION BACKGROUNDS

Doctor of Philosophy, Computer Science (Data Science)
Queensland University of Technology
Feb. 2012-Feb. 2016
Brisbane, Australia

Master of Science, Control Science and Engineering (Software Safety)
Beihang University
Sept. 2008-Mar. 2011
Beijing, China

B.S., Spacecraft Design and Engineering (Reliability System Engineering)
Beihang University
Sept. 2004-Jun. 2008
Beijing, China

RECENT WORK EXPERIENCES

Jul. 2017– pres.

Postdoctoral Fellow
CSIRO

Helping agricultural industries to be more productive and providing more valuable knowledge through a new generation of decision tools.

As the machine learning researcher in CSIRO Future Science Platform, I am taking fully responsible for ML-based time series model design and development. My daily work include:

■ **Deep Learning**

- **Time series forecasting**
  - Designed DNN/RNN/CNN based deep learning models for water quality forecasting.
  - The forecasting models are based on Tensorflow/Keras/Pytorch.
  - High quality papers have been published for time series prediction.
  - The model is part of the real-time water quality monitoring system facing to sugercane industry and farmers in GBR.
- **Time series imputation**
  - Designed Attention seq2seq based deep learning models for sensor data imputation.
  - The imputation model has been widely tested on multiple IOT data systems to improve data quality.
  - The work has been published on premier IEEE IOT journal.
- **Time series estimation**
  - Designed transfer learning based deep learning models for sensor data estimation.
  - It provides the low cost way for real-time measurements, which is key in large scale environmental monitoring.
  - The work has been submitted to premier IEEE transactions.

■ **IOT System Development**

- **LoRa sensor development**
  - Build and test a LoRa monitoring unit with multiple soil moisture sensors.
  - The system are built both on Raspberry Pi and Mdot board.
  - Deployed in GBR to support both research and farming needs.
- **AWS based sensor data processing**
  - Developed real-time sensor data cleaning pipe line with AWS platform.
  - The data cleaning system has been tested on QLD/JCU IOT systems.

■ **Management & Coopration**

- **Supervise several student projects**
  - Queensland University Industry Project/ Data Science Capstone Project
  - CSIRO-China International Exchange Project
- **Coopration**
  - Work with remote sensor researchers in CSIRO for crop forecasting.
  - Work with James Cook University for sensor development and deployment.
  - Work with QLD Government for water quality real-time monitoring data analysis.

Jul. 2016– Jun. 2017

Big Data Specialist
SilverRail Tech

SilverRail provides the real-time rail travel searching and booking system. As the big data specialist, I am responsible for the following projects:

■ **Business data analysis**

- **Statistics analysis with TB-level dairy log files**
  - Python/ C# scripts for extracting data from different data sources: MySQL, Shared File system, AWS cloud, NOSQL database, …
  - Data analysis: SQL query and join from different data tables, statistical analysis such as curve-fitting, variance analysis…
  - Data visualisation: Kibana, Python, Excel
- **Data pipe line development**
  - Developing data pipe line to integrate log data from multiple data sources.
  - Clean and organise log information for further data analysis.

■ **Data querying and storage system developing**

- **Data storage system developing and integration**
  - Data storage API developing: Hbase, HDFS, DynamoDB, Redis, ElastiCache, S3
  - Java developing
  - Language and systems: Java, Git, AWS EC2, AWS EMR
- **NoSQL Database Deployment and Performance Tuning**
  - Row-key, Cache, Data compression, Load Balancing, Region Design
  - NOSQL data table design, experiences in time series and geo info data.

PATENTS

- A distributed computing system and related strategies for processing all-to-all comparison problems with large-scale data sets (CN103942235A) — May. 2013
- A pair-based distributed storage system and related method for big data (CN106445403A) — Apr. 2015

RECENT PUBLICATIONS

Machine Learning & IoT

Yi-Fan Zhang, Peter Thorburn, Wei Xiang and Peter Fitch, SSIM - A Deep Learning Approach for Recovering Missing Time Series Sensor Data. IEEE Internet of Things Journal, doi:10.1109/JIOT.2019.2909038, 2019

Yi-Fan Zhang, Peter Fitch, Peter Thorburn and Maria Vilas, Applying multi-layer artificial neural network and mutual information to the prediction of trends in dissolved oxygen. Frontiers in Environmental Science, doi: 10.3389/fenvs.2019.00046, 2019

Yi-Fan Zhang, Yu-Chu Tian, Wayne Kelly, Colin Fidge and Jing Gao, Application of Simulated Annealing to Data Distribution of All-to-All Comparison Problems in Homogeneous Systems, The 22nd International Conference on Neural Information Processing (ICONIP), 683-691

Cloud Computing

Yi-Fan Zhang, Yu-Chu Tian, Wayne Kelly and Colin Fidge, Scalable and Efficient Data Distribution for Distributed Computing of Large-Scale All-to-All Comparison Problems. Future Generation Computer Systems, volume 67, pages 152-162, 2017.

Yi-Fan Zhang, Yu-Chu Tian, Wayne Kelly and Colin Fidge, Data-Aware Task Scheduling for Large-Scale All-to-All Comparison Problems in Heterogeneous Systems. Journal of Parallel and Distributed Computing, volume 93, pages 87-101, 2016.