Ching-Chuan Chen
Data Scientist / Data Engineer

+886-966-676-326
zw12356@gmail.com
celestial0230
chingchuan-chen.github.io

I am currently working as a senior data scientist in Trend Micro which is a global well-known cybersecurity company. My main mission is to analyze the network behaviors of IoT devices.

Over the past three years, I design several systems of quality control for wafer manufacturing. The following summarised my major achievements:

- Process over 6 terabyte data from multiple sources per day for the quality control systems.
- Improve yield of product 10% by automatically discovering the dog tools.
- Reduce 30% out-of-control rate with the regression model.
- Reduce 500% scrapped products with a homemade change points detection algorithm.
- Achieve a 93.5 percent accuracy rate in the defect recognization via a homemade CNN model.

Skills

Advanced	High-Intermediate	Intermediate		Basic
R	SQL	Shell Script	Scala	C++
Statistics	Python	Spark	Hive	C#
Supervised learning	MATLAB	MongoDB	Docker	Java
Data Visualization	Clustering	Ansible	MPI	AWS

Experience

JAN. 2019 – PRESENT
Senior Software Engineer / Trend Micro Inc., Taipei, Taiwan

- Automate the process of a report to save at least 50% time of the analysts.

SEPT. 2018 – JAN. 2019
Senior Engineer / Taiwan Semiconductor Manufacturing Company, Taichung, Taiwan

JULY 2016 – AUG. 2018
Engineer / Taiwan Semiconductor Manufacturing Company, Taichung, Taiwan

- Process over 6 terabyte data from multiple sources per day for the quality control systems.
- Improve yield of product 10% by automatically discovering the dog tools.
- Reduce 30% out-of-control rate via the regression model.
- Reduce 500% scrapped products via a homemade change points detection algorithm.
- Achieve a 93.5 percent accuracy rate in the defect recognization via a homemade CNN model.
- Decrease 50% debugging time of the data scientists with the docker which is written by me for providing a constant developing environment.
- Decrease the implementation time of the machine learning jobs by 20 times with a Hive cluster which I build in 5 months along with an ETL engine and monitoring.
- Increase the quality of the program and the productivity of the data scientists by training and sharing libraries which are designed and developed by me. It decreases the errors reported by users 50% and the developing time by 20%.

SEPT. 2015 – JUNE 2016
Research Assistant / Academia Sinica, Taipei, Taiwan
- ❖ Impute the missing values in the functional data with 40% lower RMSE than other methods.
- ❖ Automate the research data downloading to save 50% working time.
- ❖ Realize the algorithms of the travel time estimation in the journals.
- ❖ Provide a system for in-time data visualization via R shiny and d3.js.

Education

06 2014   GPA: 4.0 / 4.0
Master in Statistics / National Cheng Kung University, Tainan, Taiwan
- ▪ Thesis: A Classification Approach Based on Density Ratio Estimation with Subspace Projection / Advisor: Ray-Bing Chen
- ▪ I get 95 points in the statistical methods, the generalized linear model, and statistical data mining classes, 92 points in the linear model class. I am confident in building models and inference of models.
- ▪ I get through the advanced probability theory class which is opened for Ph. D. students.

06 2012   GPA: 3.5 / 4.0
Double Degree of Bachelor in Economics and Statistics / National Cheng Kung University, Tainan, Taiwan
- ▪ With an advanced plan and hard work, I earn 175 credits for 2 majors within 4 years.

Journal

- • Ping-Yang Chen, Ching-Chuan Chen, Chun-Hao Yang, Sheng-Mao Chang, and Kuo-Jung Lee. "milr: Multiple-Instance Logistic Regression with Lasso Penalty." R Journal (2017) 9:1, pages 446-457. https://journal.r-project.org/archive/2017/RJ-2017-013/index.html

Packages

RcppBlaze
I mimic the RcppArmadillo and RcppEigen to write this package to pack the Blaze for the R users. Blaze is an open-source, high-performance C++ math library for dense and sparse arithmetic. With its state-of-the-art Smart Expression Template implementation 'Blaze' combines the elegance and ease of use of a domain-specific language with HPC-grade performance, making it one of the most intuitive and fastest C++ math libraries available.

milr
I write all C++ code for the performance and most R code matching the operations in R stat package. The 'milr' package focuses on the predictive model for the multiple instance data set with binary outcomes and performs the maximum likelihood estimation with the Expectation-Maximization algorithm under the framework of logistic regression. Moreover, the LASSO penalty is attached to the likelihood function for simultaneous parameter estimation and variable selection.