

TAM, ZHI RUI (譚智睿)

zhirui09400@icloud.com

Currently in Hsinchu, Taiwan (Nationality : Malaysian)

Deep understanding in machine learning (ML) with major in natural language process. I also have hands on experience in integrating ML algorithm into system (web and desktop client)

Work Experience:

Lead Engineer
Pheme Pte Ltd

06/2019 - 05/2021

Setup a production service which provide global news feeds in program friendly form from ground up. This service includes crawling news from 70 countries in 18 different languages.

- Developed an in-house content extraction library (extractnet) which extracts content, headline, authors and date directly from website using a mix of machine learning approaches trained using in house data (~100k)
- Name extraction model using Conditional Random Field (CRF) trained on in house data

Master Student
National Yang Ming Chiao Tung University

09/2019 - Present

- Master research mainly focus in entity disambiguation where a mention text is linked to an existing entity. My work propose to use link prediction in knowledge graph as a regularisation method to tackle entities with tail heavy distribution.
- 03/2020 - 06/2020 : Worked on soft template based personalised title generation using bi-encoder decoder model with built in user recommendation model using negative matrix factorisation (NMF)
- 09/2019 - 03/2020 : Co work in story visualisation paper where we propose a new model that focus in high fidelity story characters and introduce a new metric called **Frechet Story Distance (FSD)** to evaluate the final performance.
- 06/2020 - Co work in a new regularisation method name Gradient Normalisation (GN) used in generative adversarial network (GAN). GN beats all existing state of the art methods in image and text generation.
- First author in 1 publication , Co-author in 2 studies

Intern Backend Developer
Perkd Taiwan/Singapore

02/2018 - 05/2018

Code optimisation in node-js framework loopback and write unittest for backend service.

- Prepare unittest and code refactoring

Research Assistant
EBIL

04/2017 - 01/2018

Design and developed a tool which collects, decode and visualise rice harvest heat map. Our decoding model is a neural network based model which segments and decode numeric values from harvester dashboard captured using android app written by myself. The decoded numeric values is recorded with differential GPS signal to create a harvest heat map overlay on google map. The tool is currently used by academic researcher to improve harvest optimisation

- Publish 1 study as first author

Academic Qualifications:

9/2021 **Master of Electrical and Computing Engineering**
National Yang Ming Chiao Tung University Taiwan
9/2019 **Bachelor of Electrical and Computing Engineering**
National Chiao Tung University Taiwan (GPA : 4.0 out of 5.0)

Academic Publication:

8/2020 **Character-Preserving Coherent Story Visualization [ECCV 2020]**
Second author, 26 percent acceptance rate
7/2021 **Gradient Normalization for Generative Adversarial Networks [ICCV 2021]**
Second author, 25.9 percent acceptance rate

Competition Experience

1st in 2018 全國智慧製造大數據分析競賽 hosted by Taiwan Ministry of Education

Reached 93% accuracy in classifying industry machinery failure using self supervised learning and contrastive learning to train neural network model.

Media coverage

1st in I'm the Best Coder! Challenge 2020 hosted by Shopee Taiwan (Student group)

Coded an algorithm to find whether a sets of items can be placed inside a given parcel size. Our team reached perfect 1.0 scores in final evaluation sets

Kaggle [Team Name : Two people is all you need]

Side Projects

Todayheadlines : News Aggregation Website

News clustering and headline extraction under limited computation budget using traditional graph algorithm. The graph algorithm performs faster than DBSCAN or KNN in terms of computation speed.

Unisearch : Machine Learning based Search Engine

Built upon vector based search design, Unisearch encodes text (9 different languages), image into a shared vector space which allows users to search image using text or vice versa. The search engine is able to scale search across millions of text and image using FAISS index.

Software and Skills:

Programming and scripting language:

Python and python packages in analytics (pandas, numpy, sklearn, seaborn, matplotlib, scipy) and web development (flask, sanic), C, C++, Java, Javascript, bash scripts, SQL, version control (Git), Tableau

Maintainer for python package : extractnet, fastlangid

Skills:

Statistical modeling (classification, regression, ranking), signal processing, computer vision, audio processing, natural language processing (NLP), generative modeling (GANs, GMM), web scraping (scrapy), large scale model training (distributed training using Pytorch, Slurm), web service maintenance (nginx, unicorn, sanic)