Kai-Chou, Yang

As a Kaggle Competition Master and a **winner of international data science challenges**, I am experienced in machine learning, deep learning and related frameworks such as PyTorch.

My research focuses on **natural language processing (NLP)**, where I have released 11 open-source projects such as MianBot **(700+★ on Github)** and presented certain academic papers on top conferences like ACL, AAAI, CIKM, **and** WSDM.

International Awards

For the following achievements, I am the **first author** as well as the **team leader**.

2nd Place, CIKM Cup: Cross-lingual Short-text Matching Challenge

- Proposed two densely-connected architectures, CPRNN and DACNN, for sentence pair modeling.
- Fused semantic features from different levels to create diversity intra-models.
- The solution has been oral presented on **CIKM 2018 in Turin, Italy.**

3rd Place, WSDM Cup: Fake News Classification Challenge

- Implemented various NLI networks like **ESIM** and injected world knowledge using **BERT**.
- Proposed a disagreement-aware model based on the single-word attention.
- The paper has been oral presented on **WSDM 2019 in Melbourne, Australia.**

4th place, Google AI: Gendered Pronoun Resolution Competition

- Leveraged the information redundancy from **BERT** and extracted features from the optimal layer.
- Proposed a **multi-heads Siamese semantic scorer** for answer selection.
- The paper has been published on **ACL 2019 in Florence, Italy.**

Kaggle Competition Master, Ranks top 0.2% (233/114,366)

- Top 1% (4/838), Gendered Pronoun Resolution Competition.
- Top 1% (27/4,550), Toxic Comment Classification Challenge.
- Top 3% (30/1,449), CareerCon 2019 - Help Navigate Robots.
- Top 4% (103/3,165), Jigsaw Unintended Bias in Toxicity Classification.
- Top 10% (384/3,946), TalkingData AdTracking Fraud Detection Challenge.

Education

Master in Department of Computer Science, NCKU

GPA: 4.30

- Honorary member of the Phi Tau Phi Scholastic Honor Society. (Ranked 1st among all graduates.)
- As a teaching assistant for Introduction to Data Science, Data Mining and Discrete Mathematics.
- As a speaker / teaching assistant for introduction lectures of machine learning.

Bachelor in Department of Computer Science, NCKU
GPA: 3.92

- Academic excellence awards 2016.
- Academic excellence awards 2015.
- Honorable mention on the graduation exhibition.
- Research assistant on a question answering system project for the Ministry of Science and Technology.

Projects

I list some of my project experiences. You can refer to my Github for the other interesting ideas.

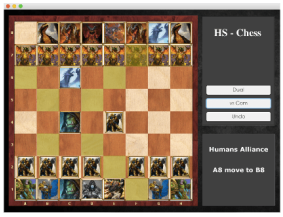
Mianbot

- Got 700+ stars and 200+ forks on Github.
- Implemented the hierarchical keywords matching using **word2vec**.
- Implemented the IR-based searching module to support chit-chat.
- Allow user to define customized scenarios with JSON.
- The extracted QA pairs were released in PTT-Gossiping-Dataset, a widely-used Chinese chit-chat corpus.

NCKU Smart-Life LineBot

- A Linebot that helps solve trivial matters such as restaurant recommendation.
- The dialogue system is based on LUIS for intent classification.
- The backend was built with Django / Flask (new version) and host on Heroku.
- The backend is connected with Line server using web API.

Hearthstone Chess



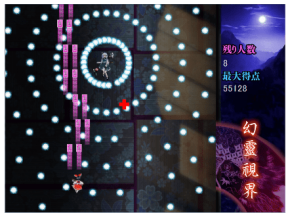
- Chess game with the hearthstone skin.
- Implemented with Java and Swing.
- Built the AI using Game Tree Searching with heuristic weights.
- Optimize the DFS-searching with several pruning methods.

NCKU Online Judge

- a Lightweight Online Judge System, built from scratch using PHP.
- Implemented the member system with profile pages that records solved tasks.
- Support some custom settings like favorite problems and personal information.

- The judge-node was using an open source project EasySandBox .

Fantasy Invision



- Touhou style shoot-em-ups game implemented with C#
- Implemented a boss, several enemies with different barrages.
- Implemented the score-board and the power-up system.

Knowledge & Skills

- Machine Learning
  - Classification, Regression, Generation, Clustering, Validation, Regularization, Feature Engineering, Transfer Learning.
- Deep Learning
  - ANN, RNN, CNN, Attention, Transformer, GAN, VAE, Seq2Seq Learning, Adversarial Learning.
- Natural Language Processing
  - Natural language understanding / Natural language inference.
  - Sentence Pair Modeling.
  - Text Classification / Regression.
  - Language Model.
  - Distributed word representation (word2vec / GloVe / FastText).
  - Deep contextual representation (ELMO / BERT / GPT-2 / XLNet / Roberta / ERNIE 2.0).
- ML Related Framework Experience
  - numpy, pandas, sklearn, keras, PyTorch, pytorch-transformers.