Jheng Hong Yang

jheng-hong.yang@uwaterloo.ca | +1.647.561.9527

RESEARCH INTERESTS

Information Retrieval Recommender Systems Natural Language Processing

EDUCATION

NATIONAL CHIAO-TUNG UNIVERSITY
MSc in Electronics Engineering
Sept 2013 | Hsinchu, Taiwan
Overall GPA: 4.22 / 4.3

NATIONAL CHIAO-TUNG UNIVERSITY
BSc in Electrophysics
June 2011 | Hsinchu, Taiwan
Overal GPA: 3.86 / 4.3; Major GPA: 3.88 / 4.3

UNIVERSITY OF WATERLOO - DATA SYSTEM GROUPS | Ph.D. Student — Waterloo, Canada
Supervisor - **Prof. Jimmy Lin** — Sept 2020 - Present

Information Retrieval – Conversational Assistant

- Built a multi-stage conversational passage retrieval system comprised of a neural query re-writer, a dense-vector retriever, and a neural text re-ranker based on pretrained language models: **BERT** and **T5**
- Submitted top-ranked runs out of **15** teams, which are the **1st** places of both automatic systems and automatic systems using canonical contexts, on the leaderboards of Conversational Assistant Track (CAsT) of TREC 2020

Information Retrieval – *Ad hoc* Dense Retrieval

- Adopted knowledge distillation to improve a dense-vector retriever based on **BERT**, which achieves **0.364** MRR@10 in the MS MARCO passage retrieval task with **106** ms/query latency

Natural Language Understanding – Commonsense Reasoning

- Integrated a prompt pattern with a logit re-ranking method based on pretrained language models
- Submitted top-ranked runs on the leaderboards: WinoGrande, OpenbookQA, and ARC-Easy, hosted by Allen Institute for AI

RESEARCH EXPERIENCE

ACADEMIA SINICA - CFDA & CLIP LABS | Research Assistant — Taipei, Taiwan
Supervisor - **Prof. Chuan-Ju Wang** and **Prof. Ming-Feng Tsai** — Sept 2017 - May 2020

Information Retrieval – Conversational Assistant

- Led **4** team members to build a conversational-information-seeking system using **TensorFlow** and **Anserini**
- Created two conversation history tracking models with **BM25** and **BERT** that outperformed the median scores of **21** teams by **140%** in mAP@5 and **94.6%** in NDCG@5 in CAsT 2019

Computational Finance – Multiperiod Default Prediction

- Designed a recurrent neural model with a monotonic constrained module based on **TensorFlow**
- Improved risk prediction accuracy: area ratio by **16%** and normalized root mean squared error by **68%**, on the CRI corporate credit risk dataset that includes **1.6M time-series** instances of US public-listed firms

Recommender System – Graphical Collaborative Filtering

- Developed a hybrid algorithm to bridge the gap between **graphical** model and **factorization**-based model
- Programed the 2nd generation of CFDA & CLIP lab's recommender system benchmark based on **Python**
- Testified our model with a **billion**-instance dataset from a leading music streaming company in Asia: **KKBOX**

NUS-RMI CREDIT RESEARCH INITIATIVE | Development Team Intern — Singapore
Supervisor - **Prof. Jin-Chuan Duan** — July 2019 - Aug 2019

Computational Finance – Multiperiod Default Prediction

- Developed a neural forward intensity model, which is validated on two credit risk datasets that contain **600k** and **1.6M** instances of exchange-listed firms in India and US, respectively
- Hosted a **4-week** tutorial of neural networks and TensorFlow for CRI members

PROFESSIONAL EXPERIENCE

CATHAY FINANCIAL HOLDINGS | Data Analyst Intern — Taipei, Taiwan
Supervisor - **Tsai-Hsiang Hung** — Oct 2017 - Oct 2018

Customer Journey Analysis – Customer Churn Prediction

- Led **8** team members and structured a project of customer journey analysis
- Processed customers' behavior records, which contains **600 million** lines of unstructured data
- Developed a framework with to integrate the wide-and-deep neural model with graph embedding algorithms

TAIWAN SEMICONDUCTOR MANUFACTURING COMPANY | R&D Engineer — Hsinchu, Taiwan
Supervisor - **Hsiu-Chuan Shu** — Nov 2013 - Oct 2016
Simulation Program with Integrated Circuit Emphasis – Radio Frequency SPICE Model

- Coordinated **10** team members to deliver the **1st** generation RF model of **16nm** FinFET technology
- Implemented and validated an empirical flicker noise model with an in-house **C-based** extension library of SPICE
- Reduced **30%** cycle time of an advance transistor model parameter extraction flow using shell programming
- Prototyped an automatic electrical parameter analysis tool with **Python**
- Reduced **20%** model error rates by automatic verification script refinements

TALKS AND PRESENTATIONS

2021 The 44th ACM Conference on SIGIR
2020 The 28th International Conference on Computational Linguistics
2019 The 27th Text REtrieval Conference
2018 The 12th ACM Conference on RecSys

AWARDS

2018 The 12th ACM Conference on RecSys **Best short paper runner-up**
2010 National Chiao-Tung University **Academic achievement award**

PUBLICATIONS

[1] Sheng-Chieh Lin, Jheng-Hong Yang, and Jimmy Lin. Contextualized query embeddings for conversational search. In *Proc. EMNLP (to be appeared)*, 2021.

[2] Sheng-Chieh Lin, Jheng-Hong Yang, Rodrigo Nogueira, Ming-Feng Tsai, Chuan-Ju Wang, and Jimmy Lin. Multi-stage conversational passage retrieval: An approach to fusing term importance estimation and neural query rewriting. *ACM Trans. Inf. Syst.*, 39(4), 2021.

[3] Sheng-Chieh Lin, Jheng-Hong Yang, and Jimmy Lin. In-batch negatives for knowledge distillation with tightly-coupled teachers for dense retrieval. In *Proc. ACL RepL4NLP*, 2021.

[4] Jia-Huei Ju, Jheng-Hong Yang, and Chuan-Ju Wang. Text-to-text multi-view learning for passage re-ranking. In *Proc. SIGIR*, 2021.

[5] Sebastian Hofstätter, Sheng-Chieh Lin, Jheng-Hong Yang, Jimmy Lin, and Allan Hanbury. Efficiently teaching an effective dense retriever with balanced topic aware sampling. In *Proc. SIGIR*, 2021.

[6] Jimmy Lin, Xueguang Ma, Sheng-Chieh Lin, Jheng-Hong Yang, Ronak Pradeep, and Rodrigo Nogueira. Pyserini: A python toolkit for reproducible information retrieval research with sparse and dense representations. In *Proc. SIGIR*, 2021.

[7] Edwin Zhang, Sheng-Chieh Lin, Jheng-Hong Yang, Ronak Pradeep, Rodrigo Nogueira, and Jimmy Lin. Chatty Goose: A python framework for conversational search. In *Proc. SIGIR*, 2021.

[8] Jheng-Hong Yang, Sheng-Chieh Lin, Rodrigo Nogueira, Ming-Feng Tsai, Chuan-Ju Wang, and Jimmy Lin. Spotting text-to-text patterns for multiple-choice question answering. In *Proc. ICCL*, 2020.

[9] Sheng-Chieh Lin, Jheng-Hong Yang, Rodrigo Nogueira, Ming-Feng Tsai, Chuan-Ju Wang, and Jimmy Lin. Conversational question reformulation via sequence-to-sequence architectures and pretrained language models. *arXiv:2004.01909*, 2020.

[10] Sheng-Chieh Lin, Jheng-Hong Yang, Rodrigo Nogueira, Ming-Feng Tsai, Chuan-Ju Wang, and Jimmy Lin. Tttttackling winogrande schemas. *arXiv:2003.08380*, 2020.

[11] Jheng Hong Yang, Sheng Chieh Lin, Jimmy Lin, Chuan Ju Wang, and Ming Feng Tsai. Query and answer expansion from conversation history. In *Proc. TREC*, 2019.

[12] Jheng Hong Yang, Chih Ming Chen, Chuan Ju Wang, and Ming Feng Tsai. Hop-rec: High-order proximity for implicit recommendation. In *Proc. RecSys.*, pages 140–144, 2018.