Jia-Huei (Dylan) Ju

Research Center of Information Technology Innovation, Academia Sinica +886-983766845
128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan (R.O.C) jhjoo@citi.sinica.edu.tw
https://dylanjoo.github.io | https://github.com/DylanJoo

EDUCATION

National Chengchi University Taipei, Taiwan
M.Sc. in Management Information Systems Sep. 2019 - Sep. 2021

- Overall GPA: 4.17/4.3
- Thesis: High-Dimensional VAR for Retail Marketing and Sales Performance Analysis (advisors: Prof. Hao-Chun Chuang and Prof. Yen-Chun Chou)

National Central University Taoyuan, Taiwan
B.B.A in Information Management Sep. 2015 - June 2019

- Overall GPA: 3.09/4.0

RESEARCH EXPERIENCE

The CFDA Lab, Academia Sinica Taipei, Taiwan
Research Assistant | Supervisor: Prof. Chuan-Ju Wang Oct. 2019 - Present

Neural Information Retrieval

- Designed text-to-text multi-view learning framework for passage re-ranking; the multi-view model using 770M parameters outperformed single-view model using 3B parameters.
- Published a paper and presented the work at SIGIR'21.

Weakly-supervised Learning Methods for Conversational Search

- Constructed a view ensemble pseudo-labeling approach for training conversational passage re-ranking model.
- Proposed methods improved supervised learning baseline method by 10% nDCG@3 with lower latency.
- Published a paper and presented the work at SIGIR'23.

Semi-supervised Rationale Extraction Pipeline for Unconvering Signals in Financial Reports

- Defined a new task for signal highlighting as well as the performance measurements for rationale extraction.
- Designed a weakly-supervised learning methods for finance domain transfer.
- Published papers to EACL'23 and ACL'23.

Domain-adaptive Dense Retrieval with Passage-centric Signals from Relevance-aware Question Generation

- Constructed instruction-tuned question generators for fine-tuning domain-adaptive dense retrievers.

Improving Multi-modal Representation for Product Retrieval

- Developed hybrid retrieval pipelines for Text REtrieval Conference 2023 (TREC'23) Product Search Track.
- Led 10 members to build the retrieval system; the system outperforms baselines by 15%.
- Designed a multimodal representation learning framework via text-image alignment and contrastive learning.

Developing Context-aware Retrieval for Interactive Search and Personalized Search.

- Built CIS systems and won 2nd/7th Places in manual/automatic sessions in CAsT of TREC'22;
- Developed clarification question generation methods based on retrieved provenances and collaborated it with open-domain conversational QA models for mixed-initiative response generation for proactive search.
- Simulated user feedback for fine-tuning and constructed statement-aware retrievers and readers based on Fusion-in-decoders for personalized search.

Collaborating Cross-lingual Query for Improving Cross-lingual Information Retrieval (CLIR)

- Constructed T5 passage re-ranking model with bilingual query for NeuCLIR track of TREC'23.
- Built CLIR systems and won 2nd, 3rd, 3rd Place (in 12 teams) in Chinese, Russian, Persian in NeuCLIR.

TEACHING EXPERIENCE

Management Information Systems Department, National Chengchi University Taipei, Taiwan
Teaching Assistant – Decision Science (master's-level course) Sep. 2020 - Feb. 2021

- Offered small group supervision on machine learning, statistics, and R/Python programming.

PUBLICATIONS

- **Jia-Huei Ju,** Sheng-Chieh Lin, Ming-Feng Tsai, and Chuan-Ju Wang. 2023. Improving Conversational Passage Re-ranking with View Ensemble. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval **(SIGIR'23)***, pages 2077–2081.
- **Jia-Huei Ju,** Yu-Shiang Huang, Cheng-Wei Lin, Che Lin, and Chuan-Ju Wang. 2023. A Compare-and-contrast Multistage Pipeline for Uncovering Financial Signals in Financial Reports. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics **(ACL'23)** (Volume 1: Long Papers)*, pages 14307–14321.
- **Jia-Huei Ju***, Ta-Wei Huang*, Yu-Shiang Huang, Cheng-Wei Lin, Yi-Shyuan Chiang, and Chuan-Ju Wang. 2023. FISH: A Financial Interactive System for Signal Highlighting. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics **(EACL'23)**: System Demonstrations*, pages 50–56. *(* indicates equal contributions)*.
- **Jia-Huei Ju**, Jheng-Hong Yang, and Chuan-Ju Wang. 2021. Text-to-Text Multi-view Learning for Passage Re-ranking. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval **(SIGIR'21),*** pages 1803–1807.

ACADEMIA-INDUSTRY COLLABORATION PROJECTS

App Banner Recommendation Methods Based on Customers' Historical Behaviors Sep. 2022 - Jul. 2023

- Collaborated with AI R&D Center, E.SUN Commercial Bank.
- Led 6 members to develop recommendation methods for selecting mobile banner ads to display.
- Developed NeuralCF models with clustering and boosted baseline system by 12%.
- Constructed multi-relation recommender systems to integrate heterogeneous user-item relations.

Analyzing Customer Characteristics using E-invoice Purchasing History May 2022 - Oct. 2022

- Collaborated with eCloud Mobile Corporation.
- Designed a product name rewriting system for over 100,000 products in Chinese; system could identify same products with different names, which further benefited downstream customer analysis.

Redesigning System Pipelines for Mutual Fund Recommendation Oct. 2021 - Sep. 2022

- Collaborated with AI R&D Center, E.SUN Commercial Bank.
- Led 4 members to redesign system pipeline for mutual funds recommendation.
- Constructed graph convolutional network RecSys and boosted original system's accuracy by 40%.
- Designed content-based neural network recommender systems for solving cold-start users' issues.

Estimating the Effects of Cross-Product Promotions on Sales Demand Dec. 2019 - Dec. 2020

- Collaborated with Industrial Technology Research Institute
- Designed optimization algorithms for time-series estimation to capture cross-product promotion effects.
- Conducted Monte-Carlo simulation experiments for empirical evaluation.

SKILLS

- Languages: Mandarin (native), English (fluent)
- Programming: Python, R, JAVA, C++, HTML/CSS
- Machine learning: Hugging Face, PyTorch, JAX, TensorFlow
- Other: Linux, Google Cloud Platform & Cloud TPU