

Jia-Huei (Dylan) Ju

Research Center of Information Technology Innovation, Academia Sinica
128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan (R.O.C)
<https://dylanjuo.github.io> | <https://github.com/DylanJu>

+886-983766845
jhjoo@citi.sinica.edu.tw

EDUCATION

National Chengchi University

M.Sc. in Management Information Systems

Taipei, Taiwan
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

B.B.A in Information Management

Taoyuan, Taiwan
Sep. 2015 - June 2019

- Overall GPA: 3.09/4.0

RESEARCH EXPERIENCE

CFDA Lab, Academia Sinica

Research Assistant / Supervisor: Prof. Chuan-Ju Wang

Taipei, Taiwan
Oct. 2019 - Present

Representation Learning for Information Retrieval

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

A Multi-stage Pipeline for Conversational Search

- Constructed conversational information seeking (CIS) system with query rewriting, sparse/dense retrieval, and passage re-ranking models.
- Redesigned the pipeline with conversationally encoded query representation for retrieval and re-ranking.
- Led 4 members to build CIS systems and participate in CAsT contest of Text REtrieval Conference 2021 (TREC'21); team won 2nd and 7th Places in manual and automatic sessions.
- Led 2 members to build CIS systems and participate in CAsT contest of TREC'22; our results outperformed the official baseline system provided by TREC by 15%.
- Explored clarification question methods, and fine-tuned corpus-aware question generation models.

A Weakly Supervised Learning Framework for Conversational Search

- Constructed pseudo-labeling approach for conversational dense retrieval and re-ranking.
- Proposed methods improved supervised learning baseline method by 10% nDCG@3; had lower latency.
- Prepared a manuscript for next-year publication in SIGIR'23.

Aligning Cross-lingual Query Representations for Improving Multilingual Information Retrieval (CLIR)

- Constructed cross-lingual retrieval pipeline using Pyserini and Hugging Face.
- Fine-tuned T5 passage re-ranking model using bilingual query to align text in different languages.
- Led 3 members to build CLIR systems; team won 2nd Place in Chinese, 3rd Place in Russian, 3rd Place in Persian selected from 12 teams in CLIR contest of TREC'22.

Rationale Extraction for Discovering Signals in Financial Reports

- Developed BERT-based automatic word-level highlighting methods for streamlining financial report reviewing.
- Designed domain transfer learning and weakly-supervised learning to be applied in finance domain.
- Submitted a paper to ACL'23; a web-based system to EACL'23 demonstration track.

Dense Users Representation Learning for Personalized News Recommendation

- Designed co-training framework for collaborative filtering and dense retrieval models.
- Developed unified representation learning with users and text for personalized news recommendations.

Improving Interactive Conversational Search with Mixed-initiative Interactions

- Designed a corpus-aware clarification question generation models using pseudo-relevance feedback.

TEACHING EXPERIENCE

Management Information Systems Department, National Chengchi University

Teaching Assistant – Decision Science (master's-level course)

Taipei, Taiwan
Sep. 2020 - Feb. 2021

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

PUBLICATIONS

Peer-reviewed conference paper

- **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.

Non-peer reviewed paper

- **Jia-Huei Ju**, Chih-Ting Yeh, Cheng-Wei Lin, Chia-Ying Tsao, Jun-En Ding, Ming-Feng Tsai, and Chuan-Ju Wang. 2021. An Exploration Study of Multi-stage Conversational Passage Retrieval — Paraphrase Query Expansion and Multi-view Pointwise Ranking. In *Proceedings of the 28th Text REtrieval Conference (TREC'21)*.

PUBLICATIONS (UNDER REVIEW)

- **Jia-Huei Ju**, Yu-Shiang Huang, Cheng-Wei Lin, Che Lin, and Chuan-Ju Wang. 2023. A Multistage Pipeline for Discovering Rationales in Financial Reports. (submitted to *ACL*, under review).
- **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. (submitted to *EACL - System Demonstration Track*, under review) (^{*} indicates equal contributions).

ACADEMIA-INDUSTRY COLLABORATION PROJECTS

App Banner Recommendation Methods Based on Customers' Historical Behaviors Sep. 2022 - Jan. 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 a System Pipeline 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 recommender systems 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.

CONFERENCE PRESENTATIONS

The 44th ACM Conference on SIGIR July 2021

- Text-to-text Multi-view Learning for passage Re-ranking

The 29th Text REtrieval Conference Nov. 2021

- The Multi-stage Pipeline for Conversational Search

The 30th Text REtrieval Conference Nov. 2022

- The Conversationally Encoded Multi-stage Pipeline and Question Generation
- Cross-language Passage Re-ranking with Bilingual and Cross-lingual Query

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