

ZHUOWEI CHEN

◇ johnny.zhuowei.chen@gmail.com ◇ <https://johnnychanv.github.io>

EDUCATION

Guangdong University of Foreign Studies (GDUFS)

Bachelor of Engineering in Software Engineering

Sept 2021 - June 2025

GPA: 3.77/4.00, Ranking: 7/144

University of California, Berkeley (UCB)

Concurrent Enrollment Student

Aug 2023 - Jan 2024

GPA: 4.00/4.00

PUBLICATIONS

* represents equal contributions while † represents the corresponding author.

1. Lianxi Wang, **Zhuowei Chen**^{*†}, and Yujia Tian^{*}.
Enhancing Hindi Feature Representation Through Fusion of Dual-Script Word Embeddings.
Proceedings of the 31st International Conference on Computational Linguistics, 2024.
COLING 2024 (Long-paper, Main Conference)
2. **Zhuowei Chen**, Yujia Tian, Lianxi Wang[†], and Shengyi Jiang.
A Distantly-Supervised Relation Extraction Method Based on Selective Gate and Noise Correction.
China National Conference on Chinese Computational Linguistics, 2023.
CCL 2023 (Long-paper, Main Conference)
3. **Zhuowei Chen**^{*}, Huayu Huang^{*}, and Lianxi Wang[†].
LAKA: A Label-Aware and Knowledge-Augmented Framework for Multi-Label Text Classification.
Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, 2024.
ACL 2024 (Under review)

RESEARCH EXPERIENCE

Guangzhou Key Laboratory of Multilingual Intelligent Processing

Undergraduate Research Student

Guangzhou, China

Nov 2021 – Present

- Enhancing Hindi Representations via Fusion of Pre-trained Language Models. (COLING 2024)
 - Proposed a method to enhance feature representation for Hindi, by combining single-script features from Devanagari and Romanized Hindi Pre-trained Language Models (PLMs).
 - Designed four different fusion methods for dual-script word embedding fusion. Namely, concatenation, cross-attention, summation, and CNNs.
 - Conducted experiments on TC, NLI, POS Tagging, and NER tasks. In general, dual-script methods have advantages and outperform single-script representation methods.
 - This work excavates the potential of combining different scripts of one language, providing new ideas for the integration of PLMs. Also, it promotes the development of low-resource language models and promotes AI equity.
- Distantly Supervised Relation Extraction with Noise-Resistant Techniques. (CCL 2023)
 - Proposed a distantly supervised relation extraction (DSRE) method integrating noise-resistant selective gate and noise correction training framework.
 - Designed a three-stage training framework, in which the noise correction framework corrects noise labels during the training process.
 - Performed ablation study and experiments on two DSRE datasets. The results demonstrate that our method outperforms baselines and achieves a SOTA performance in the DSRE task.

- This project mitigates the problem that distantly supervised annotation generate a large amount of noisy data. The approach shows great potential for practical application in not only the field of DSRE but also in all information extraction tasks that face noise-label challenges.
- Multi-Label Text Classification with Knowledge Augmentation and Span Prediction.
 - Proposed a multi-label text classification (MLTC) method that casts multi-label prediction to span prediction, with a knowledge-augmentation module integrated.
 - Introduced span-prediction to MLTC, which enhanced interaction between and within labels and context. Also, integrated a knowledge-augmentation module that extracts external information with GNN, and augments the backbone model with external knowledge fusion layers.
 - Conducted visualization, ablation study, and experiments on three datasets. Our method outperforms the best benchmark by 1.32% in the macro F1-score. Comparative visualization further validates the effectiveness of our model.
 - This project not only advances the performance of MLTC but also reveals the significance of simultaneously considering label correlations and natural language understanding for MLTC.

AI Lab, Wisers Information Limited

NLP Research Intern

Hong Kong, China

Dec 2023 - Mar 2024

- Quantization of Hong Kong Tourism Popularity.
 - Predicted hotel prices, hotel occupancy, and number of arrivals in Hong Kong from analysis of the tendency of mainstream social media content.
 - Built sentiment and spam classification models from BERT-family PLMs with human annotated social media content.
 - Applied transformer for time series regression to predict the number of arrivals from statistical results of social media contents. Subsequently, used random forest model for predicting future hotel prices and occupancy levels.

SELECTED PROJECT

- Image-Text Interfacing with CLIP and Rational Speech Acts.
 - Used the CLIP model for image and caption retrieval, enhancing retrieval performance through parameter tuning and model optimization.
 - Developed a Rational Speech Acts inference procedure and applied it to the retrieval tasks, further improving retrieval effectiveness.

SELECTED HONORS

- First-class Scholarship GDUFS Academic Scholarship, 2023
- Gold Medal Cantonese College Computer Design Competition, 2022
- Silver Medal National College Student Mathematical Modelling Competition, 2023
- Silver Medal (Top 5%) National College Computer Design Competition, 2022

OTHER RELATED EXPERIENCE & SKILLS

- **Conference Participant.** Poster presentation on *CCL 2023*.
- **Teaching Assistance.** TA for Language Processing Technique.
- **Programming:** Python, Java, JS/HTML/CSS, C/C++, SQL, Golang.