



PHAN CHAU THANG

✉ phanchauthang02@gmail.com  github.com/phanchauthang  0963530772

Education

University of Information Technology

Data Science

Ho Chi Minh City

Dec. 2020 - Present

Experience and Project

University of Information Technology

Student

Ho Chi Minh city

Dec. 2020 - Present

- A Novel Incremental Learning Approach Based on Deep Recurrent Neural Networks for Multi-domain Sentiment Analysis
 - * Proposed an incremental learning training algorithm adaptable to a wide variety of real-world online multi-domain sentiment analysis challenges and capable of continuous learning.
 - * The proposed algorithm reduces training time on multi-domain sentiment analysis compared to traditional deep recurrent neural networks.
- ViCGCN: Graph Convolutional Network with Contextualized Language Models for Social Media Mining in Vietnamese
 - * Proposed a novel approach for text classification on Vietnamese social media using a joint training of large-scale language pre-trained language model PhoBERT and Graph Convolutional Networks modules.
 - * Achieved state-of-the-art results on Vietnamese social media benchmarks, including Vietnamese emotion recognition, Vietnamese constructive and toxic analysis, and Vietnamese sentiment analysis.
 - * Addressed the challenge of imbalanced and noisy data in benchmark social media dataset.

UIT Data Science Society

Head of Department of Training and Research

Ho Chi Minh city

Aug. 2023 - Present

- Developed enrollment projects, plans, methods of promotion, and admission counseling.
- Developed training plans for members.

UIT@NLP GROUP

NLP Research Engineer

Ho Chi Minh city

Dec. 2022 - Present

- Research and development Language Model for Vietnamese social media text processing and combine Language model with Graph neural network.
- Authored 1 accepted papers in the field of Natural language processing at Empirical Methods in Natural Language Processing (EMNLP).
- Authored 1 under review papers in the field of Natural language processing: ISI Q1 Journal.

Publications

PLM4ViSMT: A Pre-Trained Language Model for Vietnamese Social Media

Text Processing

*The 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP) - Rank A**

First author

- Presented PLM4ViSMT, the first pre-trained language model for Vietnamese social media texts.
- Achieved SOTA performances on a range of Vietnamese downstream social media tasks.
- Analyzed experimental results on the masking rate, examining social media characteristics, and implementing feature-based extraction for task-specific models.

Link Prediction for Wikipedia Articles as a Natural Language Inference Task

The 2023 Data Science and Advanced Analytics (DSAA) - Rank B

First author

- Our system achieve top-3 in private test in the contest.
- Proposed system for problem Link Prediction for Wikipedia Articles.

NomNaOCR: The First Dataset for Optical Character Recognition on Han-Nom Script

2022 International Conference on Computing and Communication Technologies (RIVF)

Co author

- Introduce the NomNaOCR dataset for the old Hán-Nôm script based on 3 tremendous and valuable historical works of Vietnam.

An Empirical Study of Time Series Models for Small-Scale Temperature Forecasting in Ho Chi Minh City

2022 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT) - UnRank

Co author

- Introduced a new temperature and related properties dataset in Ho Chi Minh City from 2000 to 2021.
- Built a small-scale temperature forecasting system applying state-of-the-art time series models.

Technical Skills

Programming Skills: Python (proficient) , C++, SQL (prior experience)

Developer Tools: GIT (prior experience)

Technologies/Frameworks: Pytorch (proficient), HuggingFace (proficient)