CONGCONG WANG

https://wangcongcong123.github.io/

University College Dublin Belfield, Dublin 4, Ireland.

wangcongcongcc@gmail.com

ABSTRACT

Recently, I completed my PhD degree at the School of Computer Science at University College Dublin (UCD) under the guidance of Dr. David Lillis. At the moment, I am working as a research assistant at UCD, focusing on hierarchical multi-label classification for lengthy documents. In addition to my research, I have a keen interest in creating and open sourcing projects that I find challenging to obtain online.

EDUCATION

University College Dublin, Ireland *September 2018 - February 2023*
Ph.D. with Scholarship from the School of Computer Science PG Grant
School of Computer Science and Informatics (CSI)
Thesis: **Coping with low data availability for social media crisis message categorisation**

University College Dublin, Ireland *September 2017 - July 2018*
4th Year of B.Sc. (Hons.) Software Engineering, CSI **First Class Honors**
Relevant Coursework: Data Analytics, Machine Learning, Computing Performance Evaluation, Contemporary Software Development, Operating Systems, Computer Network, Computer Security.
Thesis: **Evaluating Recommender Systems (A Grade)**

Beijing University of Technology, China *September 2014 - July 2017*
1st - 3rd Years of B.Sc. (Hons.) Software Engineering, Beijing-Dublin International College (BDIC)
Relevant Coursework: Object-oriented Programming, Software Project Development, Computer Vision, Distributed System, Database Management, Web Development, Data Structures and Algorithms

The Second Middle School of Hengnan County, Hunan, China *September 2012 - June 2014*
Senior High School study

Chejiang Middle School of Hengnan County, Hunan, China *September 2009 - June 2012*
Junior High School study

WORK EXPERIENCE

Research Assistant *January 2023 - September 2023*
University College Dublin, Ireland
The research is on hierarchical multi-label classification for legal documents. This research opportunity comes with a hosting agreement, and while my current residence status in Ireland is stamp 2, it is transferable to stamp 1G or 1.

Head of Demonstrators *September 2022 - December 2022*
University College Dublin, Ireland
I was the head of demonstrators on the Introduction to Artificial Intelligence course, responsible for managing demonstrators in assisting and assessing more than 130 students.

AIOPs Research Intern, Part-time *March 2021 - September 2021*

AISafe, Huawei Ireland Research Centre, Ireland

My major work included constructing multilingual datasets of user comments via generative models with human in the loop and exploring classification models for detecting harmful comments. Other work included **AIBoard** - a leaderboard platform for hosting various NLP tasks, **AISchedule** - a tool for monitoring and scheduling GPUs of multiple remotes, **AIDemo** - a web application for various models inference.

Teaching Assistant   *September 2015 - June 2017*

Beijing-Dublin International College, China

I was the seminar leader on the Information Retrieval course, responsible for support and assessment of 80 students.

Programming Instructor (Python)   *September 2015 - June 2016*

YiXinXian (startup) Technology Education Company, China

I was a Python instructor for one year, my duties included writing tutorials and assessing students. I adopted the use of funny storytelling to explain concepts to students.

PUBLICATIONS

Congcong Wang. Coping with low data availability for social media crisis message categorisation. PhD Thesis, 2023.

Congcong Wang, David Lillis. ISA: Iterative Self-controlled Text Augmentation for Text Classifications in Low-data Regimes. Experiments done, 2023.

Congcong Wang, Gonzalo Fiz Pontiveros, Steven Derby, Tri Kurniawan Wijaya. STA: Self-controlled Text Augmentation for Improving Text Classifications. Submitted and under review, 2023.

Congcong Wang, Paul Nulty, David Lillis. Using Pseudo-Labelled Data for Zero-Shot Text Classification. In Proceedings of the 27th International Conference on Natural Language & Information Systems (NLDB 2022), Valencia, Spain, June 2022.

Congcong Wang, David Lillis. UCD-CS at TREC 2021 Incident Streams Track. In Proceedings of the Thirty Text REtrieval Conference (TREC 2021), Virtual, 2022.

Congcong Wang, Paul Nulty, and David Lillis. Crisis Domain Adaptation Using Sequence-to-sequence Transformers. In Proceedings of the 18th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2021), Virtual, 2021.

Congcong Wang, Paul Nulty, and David Lillis. Transformer-based Multi-task Learning for Disaster Tweet Categorisation. In Proceedings of the 18th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2021), Virtual, 2021.

Congcong Wang, David Lillis. Multi-task transfer learning for finding actionable information from crisis-related messages on social media. In Proceedings of the Twenty-Ninth Text REtrieval Conference (TREC 2020), Virtual, 2021.

Congcong Wang, and David Lillis. UCD-CS at W-NUT 2020 Shared Task-3: A Text to Text Approach for COVID-19 Event Extraction on Social Media. In Proceedings of the 2020 EMNLP Workshop W-NUT: The 6th Workshop on Noisy User-generated Text, November 19, 2020.

Congcong Wang, Paul Nulty, and David Lillis. A Comparative Study on Word Embeddings in Deep Learning for Text Classification. In Proceedings of 4th International Conference on Natural Language Processing and Information Retrieval (NLPIR 2020), Seoul, Korea.

Congcong Wang, David Lillis. Classification for Crisis-Related Tweets Leveraging Word Embeddings and Data Augmentation. In Proceedings of the Twenty-Eighth Text REtrieval Conference (TREC 2019), Gaithersburg, MD, 2020.

SKILLS

Research	Natural Language Processing (NLP), Weak Supervision Learning, Deep Learning
General	Python (preferred), Java, Web Application, Streamlit, Linux, DevOps
Languages	Chinese (First), English (Fluent)

RESEARCH PROJECTS

TREC Incident Streams Participation *January 2019 - November 2021*
National Institute of Standards and Technology, U.S.
We developed systems iteratively for finding crisis-related actionable tweets and estimating their criticality using different computational linguistic techniques. In the latest edition, our approaches **hit the top performance** compared to other participating groups.

Covid19-search *Mar, 2020*
Work from home, Dublin
A retrieval system for searching insights from covid-related academic articles.

Research Assistant *December 2016 - June 2017*
China Center for Information Industry Development, China
This is a group project with my peers. We developed a library indexing and borrowing system. This was used for book indexing construction and research on Chinese-based book items retrieval techniques. This project finally won **2nd place prize for the China Software Cup**.

TALKS

Presenting our work using pseudo-labelled data for zero-shot text classification, published to the International Conference on Natural Language & Information Systems (NLDB 2022), Spain, June 2022.

Presenting our crisis domain adaptation work published to the International Conference on Information Systems for Crisis Response and Management (ISCRAM 2021), Virtual, May 2021.

Presenting our evolved systems in multi-task crisis streams categorisation at the Twenty-Ninth Text REtrieval Conference (TREC 2020), Virtual, December 2020 and December 2021.

Presenting our study on word embeddings published to International Conference on Natural Language Processing and Information Retrieval (NLPIR 2020), Virtual, December 2020.

OPEN SOURCE PROJECTS

TTT: A package for fine-tuning HuggingFace's Transformers with TPUs (code release), written in Tensorflow2.0, Sep, 2020.

Autocoder: A basic and simple tool for code auto completion (blog and source code releases), Jun, 2020.

Transection: A tool for fine-tuning sequence-to-sequence Transformers for English to Chinese Translation (blog and source code releases). Oct, 2020.

Data-Explorer: A lightweight web application for data explanatory analysis (code release), Aug, 2020.

Feeder-Bot: A tool for automating updates of RSS feeds (e.g., arxiv) to users' receiver platforms (e.g., Twitter) (code release), Mar, 2020.

MISC.

EMNLP Reviewer. Contributor for `flops-counter.pytorch` and `x-transformers`.