DONGYU ZHANG

Worcester, Massachusetts dzhang5@wpi.edu

SKILLS

Deep Research Acumen in natural language processing, representation learning, multi-task learning, and weakly-supervised learning, evidenced by publications in top-tier conferences and journals.
Robust Technical Proficiency in deep learning and machine learning, demonstrated in applications like fraudulent transaction detection, foodborne illness detection, explainable text categorization, clinical notes forecasting, and transaction sequence representation.
Programming & Analytics Mastery, encompassing tools such as Python, R, Java, SQL, and Hadoop.

EDUCATION

Worcester Polytechnic Institute (WPI), Worcester, MA Aug. 2018 – Expected May. 2023
Ph.D. in **Data Science**, GPA: 4.000/4.000

The University of Texas at Dallas (UTDallas), Richardson, TX Aug. 2016 – May. 2018
M. Science in **Business Analytics**, GPA: 3.917/4.000

Beijing Forestry University (BJFU), Beijing, China Sep. 2012 – Jun. 2016
B. Economics in **Statistics**, GPA: 89.13/100
B. Science in **Computer Science & Technology** (minor)

RESEARCH EXPERIENCES

Fact: Innovative Big Data Analytics Technology for Microbiological Risk Mitigation Assuring Fresh Produce Safety, WPI & University of Illinois Urbana-Champaign Sep. 2020 – Present
Advisor: Prof. Elke Rundensteiner & Prof. Hao Feng

- Created Tweet-FID, the first public dataset for multi-task foodborne illness detection.
- Employed LLMs, using the Chain of Thought method, to generate labels for unlabeled tweet data.
- Developed a multi-task framework to extract information related to foodborne illness incident.
- Mentored a team of 6 in the Major Qualifying Project annually.

Explainable Text Classification with Limited Human Guidance, WPI Jun. 2020 – Feb. 2021
Advisor: Prof. Elke Rundensteiner

- Introduced the novel problem of text classification with limited human attention supervision.
- Devised a transformer-based multi-task architecture, HELAS, to address this novel problem.

Time-Aware Network for Clinical Notes Series Prediction, WPI Sep. 2019 – May. 2020
Advisor: Prof. Elke Rundensteiner

- Crafted a transformer-based method to classify a patient's health state based on clinical note series.

INTERSHIP EXPERIENCE

Research Scientist Intern, Visa Inc. May. 2022 – Aug. 2022

- Developed a hierarchical transformer approach for transaction sequence representation learning.
- Utilized this method for pretraining and finetuning on fraudulent transaction detection task.

Data Analyst Intern, Shanghai PnR Data Service Co., Ltd. May. 2017 – July. 2017

- Optimized a method for posts' themes classification and replies' emotion tendency analysis.
- Refined bank card character extraction algorithm.
- Developed a method for information extraction from National ID card.

PUBLICATIONS

CoLafier: Collaborative Noisy Label Purifier With Local Intrinsic Dimensionality Guidance.
Dongyu Zhang, Ruofan Hu, Elke Rundensteiner. *Proceedings of the 2024 SIAM International Conference on Data Mining (**SDM**)*. 2024

UCE-FID: Using Large Unlabeled, Medium Crowdsourced-Labeled, and Small Expert-Labeled Tweets for Foodborne Illness Detection.
Ruofan Hu, **Dongyu Zhang**, Dandan Tao, Huayi Zhang, Hao Feng, Elke Rundensteiner. *2023 IEEE International Conference on Big Data (**Big Data**)*. 2023

Epidemiological Data Mining for Assisting with Foodborne Outbreak Investigation.
Dandan Tao, **Dongyu Zhang**, Ruofan Hu, Elke Rundensteiner, Hao Feng. ***Foods***. 2023, 12(20):3825.

FATA-Trans: Field and Time-Aware Transformer for Sequential Tabular Data.
Dongyu Zhang, Liang Wang, Xin Dai, Shubham Jain, Junpeng Wang, Yujie Fan, Chin-Chia Michael Yeh, Yan Zheng, Zhongfang Zhuang, Wei Zhang. *Proceedings of the 32nd ACM International Conference on Information & Knowledge Management (**CIKM**)*. 2023.

A Novel Foodborne Illness Detection and Web Application Tool Based on Social Media.
Dandan Tao, Ruofan Hu, **Dongyu Zhang**, Jasmine Laber, Anne Lapsley, Timothy Kwan, Liam Rathke, Elke Rundensteiner, Hao Feng. ***Foods***. 2023, 12(14):2769.

TWEET-FID: An Annotated Dataset for Multiple Foodborne Illness Detection Tasks.
Ruofan Hu, **Dongyu Zhang**, Dandan Tao, Thomas Hartvigsen, Hao Feng, Elke Rundensteiner. *Proceedings of the Thirteenth Language Resources and Evaluation Conference (**LREC**)*. 2022, 6212-6222.

Human-like Explanation for Text Classification with Limited Attention Supervision.
Dongyu Zhang, Cansu Sen, Jidapa Thadajarassiri, Thomas Hartvigsen, Xiangnan Kong, Elke Rundensteiner. *2021 IEEE International Conference on Big Data (**Big Data**)*. 2021, 957-967.

Crowdsourcing and machine learning approaches for extracting entities indicating potential

foodborne outbreaks from social media.
Dandan Tao, **Dongyu Zhang**, Ruofan Hu, Elke Rundensteiner, Hao Feng. ***Scientific Reports***, 2021, 11: 21678.

Time-Aware Transformer-based Network for Clinical Notes Series Prediction.
Dongyu Zhang, Jidapa Thadajarassiri, Cansu Sen, Elke Rundensteiner. *Proceedings of the 5th Machine Learning for Healthcare Conference (**MLHC**)*, PMLR, 2020, 126: 566-588.

An Empirical Analysis on the Willingness to Purchase Social Services and Its Influencing Factors in State - owned Forest Farms.
Yizhen li, **Dongyu Zhang**, Hui Zhang, Huiyi Shi. ***Consume Guide***, 2016, 05: 45.

The discussion on the application prospect of crowdfunding in the realm of forestry.
Chuanyuan Liu, Xuanyuan Yin, **Dongyu Zhang**. ***Modern Business Trade Industry***, 2015, 09: 104-108.

PRESENTATIONS

Crowdsourcing and Machine Learning Approaches for Extracting Information Indicating Potential Foodborne Outbreak from Social Media.
Dandan Tao, **Dongyu Zhang**, Ruofan Hu, Elke Rundensteiner, Hao Feng. ***International Association for Food Protection Annual Meeting (Virtual)***, 2021