TONG XIANG

xtong@vt.edu ◇ LinkedIn ◇ Github ◇ Google Scholar

EDUCATION

Virginia Tech *Aug. 2021 - May 2023 (Expected)*
Master of Engineering in Computer Science

Georgetown University *Aug. 2019 - May 2021*
Master of Science in Computer Science
Thesis: A Predictive and Interpretable Model for Toxic Content Classification

Wuhan University *Sep. 2015 - Jun. 2019*
Bachelor of Engineering in Computer Science

PUBLICATIONS

* equal contribution

1 **Tong Xiang**, Eugene Yang, Sean MacAvaney, and Nazli Goharian. 2021. ToxCCIn: Toxic Content Classification with Interpretability, In *Proceedings of the Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA@EACL).*

2 Jiayuan Ding*, **Tong Xiang***, Zijing Ou*, Ruihui Zhao, Yefeng Zheng, and Bang Liu. 2022. Tell Me What to Read: Literature Review Made Simple with Automatic Reading Path Generation, In *International Conference on Data Engineering (ICDE).*

3 Jacob Daniel Beel, **Tong Xiang**, Sandeep Soni, and Diyi Yang. 2022. Linguistic Characterization of Contentious Topics Online: Case Studies on Controversy in Abortion, Climate Change, and Gun Control, In *International AAAI Conference on Web and Social Media (ICWSM).*

4 Sajad Sotudeh*, **Tong Xiang***, Hao-Ren Yao, Sean MacAvaney, Eugene Yang, Nazli Goharian, and Ophir Frieder. 2020. GUIR at SemEval-2020 Task 12: Domain-Tuned Contextualized Models for Offensive Language Detection, In *Proceedings of the International Workshop on Semantic Evaluation (SemEval@COLING).*

SELECTED PROJECTS

Toxic Language Identification and Analysis in Social Media *Dec. 2019 - Apr. 2021*
Georgetown University, Information Retrieval Lab Supervised by Prof. Nazli Goharian

· Proposed a two step domain-tuning method to improve model performance, by first further pre-training the contextualized language model on domain corpus and then fine-tuning it on task-specific data
· Ranked 4/82 in sub-task A, 6/41 in sub-task B, and 11/37 in sub-task C, in SemEval 2020 task 12
· System description paper [4] accepted to **[SemEval 2020]**
· Proposed a sample yet efficient assumption to improve interpretability for neural models
· Incorporated the assumption with transformer-based models using multi-task learning framework
· Long paper [1] accepted to and orally presented at **[WASSA 2021]**

TEACHING

Graduate Teaching Assistant
Virginia Tech, Department of Computer Science

· CS-2104: Intro to Problem Solving in CS *Aug. 2021 - Dec. 2021*
· CS-5824: Advanced Machine Learning *Jan. 2021 - May 2022*

Teaching Assistant
Georgetown University, Department of Computer Science

· COSC-576: Introduction to Deep Learning *Jan 2020 - May 2020*
· COSC-051: Computer Science I *Feb. 2021 - May 2021*

SKILLS

Programming languages: Python, C++, HTML/CSS, JavaScript, SQL, Java
Packages/Frameworks: Pytorch, Scikit-learn, Transformers, Numpy, Pandas, Matplotlib, NLTK, Git, LaTeX, spaCy, XGBoost, React, Django, Keras, Gensim, SciPy, Scrapy
Databases: MySQL, MongoDB