XIAOBO GUO

Xiaobo.Guo.GR@dartmouth.edu | (267)353-5980 | github.com/xiaobo-guo

EDUCATION

Dartmouth College Hanover, NH, USA

PhD in Computer Science (Machine Learning, Deep Learning and NLP) *Sep. 2019 – Now*

Peking University Beijing, China

LLB. in Law *Sep. 2015 – June.2019*

BS in Computer Science and Technology *Sep. 2015 – June.2019*

WORK EXPERIENCE

Amazon People Experience and Technology Central Science New York, NY, USA

Applied Scientist Intern *June. 2023 – Now*

- Led efforts to scale the current **large-scale convolutional language model**, Hyena, from 125M to 3B parameters.
- Create a new task to evaluate model's ability for reasoning and in-context learning.
- Designing a new large-scale language model based on the **Attention Mechanisms** and **Fast Fourier Convolution Neural Network** to increase the performance by 30% in our new proposed task.

Amazon People Experience and Technology Central Science New York, NY, USA

Applied Scientist Intern *June. 2022 – Sep. 2022*

- Developed unique method for transforming traditional **summarization datasets** into **aspect-based summarization datasets**, facilitating more targeted data analysis.
- Engineered a comprehensive multi-step framework for summarizing employee feedback, leveraging **clustering** and **summarization** techniques for enhanced insight extraction.
- Enhanced summarization quality by **28%** through the introduction of a novel end-to-end **deep learning methodology**, effectively eliminating accumulated error.

RESEARCH PROJECTS

Dynamic Aspect-based Summarization *Sep. 2022 – June. 2023*

- Developed three unique datasets for **dynamic aspect-based summarization** tasks, filling a gap where no existing datasets were previously available.
- Proposed a **multi-objective learning framework**, which improved summarization quality by over **30%** relatively.

Reducing Length Bias for Summarization *Sep. 2022 – Apr. 2023*

- Demonstrated the pervasive existence of **length bias** via thorough metric and length analyses, strengthening the understanding of underlying biases in data.
- Implemented a **Bayesian network** to reduce length bias, enhancing the **correlation rate** between automatic metrics and human evaluation by over 40%, and thereby improving metrics quality.

Measuring Media Bias Through Masked Token Prediction *Aug. 2020 – June. 2022*

- Utilized **deep learning models** to extract the stance of individual media outlets.
- Developed innovative metrics to evaluate the performance of models for **media bias analysis.**
- Enhanced the **correlation rate** between models and human evaluations by a relative **5%**, improving the reliability and accuracy of model outputs.

Emotion-based Modeling of Mental Disorders on Social Media *Aug. 2020 – Feb. 2021*

- Formulated machine learning models including **Support Vector Machine**, **Logistic Regression**, and **Random Forests** to passively detect mental disorders, leveraging conversation data from Reddit.
- Utilized **deep learning models**, specifically **BERT**, to predict user emotions in individual posts.
- Surpassed existing model performance by achieving a **2%** improvement in both **F1-score** and **accuracy**.

Identification of State-Sponsored Propaganda on Twitter *Aug. 2019 – June. 2020*

- Created dataset for the state-sponsored propaganda identification task by leveraging the Twitter API.
- Utilized multimodal deep learning methods, specifically BERT and ResNet50, to tackle the task and achieved state-of-the-art performance.

PUBLICATION

- **Xiaobo Guo**, Weicheng Ma and Soroush Vosoughi "Capturing Topic Framing via Masked Language Modeling". Findings of the Association for Computational Linguistics: EMNLP 2022
- **Xiaobo Guo,** Weicheng Ma and Soroush Vosoughi "Measuring Media Bias via Masked Language Modeling". Proceedings of the International AAAI Conference on Web and Social Media (ICWSM), 2022
- **Xiaobo Guo** and Soroush Vosoughi "A Large-Scale Longitudinal Multimodal Dataset of State-Backed Information Operations on Twitter". Proceedings of the International AAAI Conference on Web and Social Media (ICWSM), 2022
- **Xiaobo Guo**, Yaojia Sun and Soroush Vosoughi "Emotion-based Modeling of Mental Disorders on Social Media". The 20th IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2021
- **Xiaobo Guo** and Soroush Vosoughi "Multi-modal Identification of State-Sponsored Propaganda on Social Media". International Conference on Pattern Recognition (ICPR), 2021

LANGUAGE AND TOOLS

- Language: Python3, C++, SQL
- Tools: Keras, NLTK, PyTorch, Scikit-learn, TensorFlow, Transformers, Numpy, Pandas