ZHAOMIN XIAO

zhaomin1995.github.io ◇ zhaominxiao@my.unt.edu ◇ (940)703-9518 ◇ linkedin.com/in/zhaominxiao/

EDUATION

University of North Texas *May 2023 (expected)*
Ph.D. in Computer Science and Engineering

University of Pittsburgh *May 2019*
M.S. in Information Science

Shenzhen University *July 2017*
B.S. in Mathematics & Applied Mathematics

RESEARCH INTEREST

Natural Language Processing, Machine Learning, Information Extraction, Multimodal Learning

WORK EXPERIENCE

Software Engineer Intern, Machine Learning May 2022 - Aug 2022
Meta (previously known as Facebook) *Bellevue, WA*

- Create a pipeline to normalize features and prepare the aggregated and anonymized data for the model trainer. Therefore, the pipeline can complete the requirements in Apple's new privacy policy.
- Build a privacy-preserving logistic regression model to predict if the user will click the advertisement with aggregated and anonymized input data, obtaining results that are close to state-of-the-art model.
- Collaborate with the product team and research team to evaluate the model with different parameter settings and privacy settings. The first prototype was developed based on initial experimental results.

PUBLICATION

1. **Zhaomin Xiao** and Eduardo Blanco. *Are People Located in the Places They Mention in Their Tweets? A Multimodal Approach.* In the Proceedings of **COLING 2022**.

RESEARCH EXPERIENCE

Temporal Information Extraction Aug 2021 - Present

- Propose a new task of determining how long Twitter users will stay at the geotagged location. Create a corpus that the quality is guaranteed by filtering out bad annotators and unreliable annotations.
- Corpus analysis shows that users will spend less than one month in the tagged places in most tweets.
- Build machine learning models and location-aware multimodal neural networks to tackle this problem. Experimental results show that both text and image are essential to make the right prediction.

Spatial Information Extraction Aug 2019 - Aug 2021

- Propose a new task of determining if Twitter users are located in the mentioned location when the tweet was posted. Create a corpus and build multi-modal models to tackle this problem.
- Create another corpus to include the history tweets. Build context-aware neural networks to leverage the contextual information. Error analyses show that incorporating temporal commonsense is beneficial.
- Design interfaces to collect annotations of two corpora on Amazon Mechanical Turk.

TECHNICAL SKILLS

Computer Languages	Python, Java, C, HTML, CSS, JavaScript, MySQL
API	PyTorch, TensorFlow, nltk, spaCy, scikit-learn, pandas
Tools	AWS, GCP, Git, Bitbucket, Unix, Linux, Hive