

ZHAOMIN XIAO

zhaomin1995.github.io \diamond zhaominxiao@my.unt.edu

EDUCATION

University of North Texas

Ph.D. in Computer Science and Engineering

Sep 2019 - May 2023 (expected)

University of Pittsburgh

M.S. in Information Science

May 2019

Shenzhen University

B.S. in Mathematics & Applied Mathematics

July 2017

RESEARCH INTEREST

Natural Language Processing, Deep Learning, Spatiotemporal Knowledge Mining

RESEARCH EXPERIENCE

Extracting Spatiotemporal Knowledge from Twitter Timeline

Feb 2021 - Present

- Propose the task of determining if the author of the tweet was located in the mentioned location based on the Twitter timeline.
- Build neural models baselines. Experimental results indicate that the context-aware neural network yields better result than the model using only a single tweet.
- Error analyses show that temporal information and user information might bring further improvements.

Spatiotemporal Knowledge Mining

Aug 2019 - Feb 2021

- Propose a new task of determining if the author of the tweet was located in the mentioned location when the tweet was posted. Build a new corpus consisting of 6,540 tweets.
- Design interface to collect annotation on Amazon Mechanical Turk, identify unreliable annotators, and filter out low-quality annotations. The final inter-annotator agreement is over 0.60.
- Build neural model baselines. The experimental results show that the multi-modal model yields the best result. Error analyses indicate that the further improvement might be from location identification and advertisement filtering.

Temporal Ordering of Locations

Oct 2019 - Apr 2021

- Determine if the author of the travelogue is located in the mentioned location or not.
- Order the visited location in each document of the travelogue.
- Error analyses show that incorporating contextual information would be helpful in this task.

Native Language Identification

Aug 2018 - Dec 2018

- Determine the native languages of the authors of TOEFL essays based on POS tags, stop words, grammar errors.
- Build linear regression classifier. Parameters are tuned using GridSearch with 5-fold cross-validation. The classification results are visualized by using seaborn.

TECHNICAL SKILLS

**Computer Languages
Tools**

Python, Java, C, HTML, CSS, JavaScript
PyTorch, TensorFlow, nltk, spaCy, scikit-learn