ZHAOMIN XIAO

zhaomin1995.github.io ◇ zhaominxiao@my.unt.edu

EDUATION

University of North Texas *Sep 2019 - Present*
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, Deep Learning, Spatio-temporal Knowledge Mining

RESEARCH EXPERIENCE

Extracting Spatial-temporal Knowledge from Twitter Timeline Feb 2021 - Sep 2021

- Propose the task of determining whether the author of the tweet was located in the mentioned location, based on the Twitter timeline.
- Construct multiple neural networks to solve the task. Experimental results indicate that the context-aware neural network which incorporates Twitter timeline yields the best result.
- Error analyses show that the timestamp and user information can be used as additional information to further improve the model performance.

Spatial-temporal Knowledge Mining Aug 2019 - Feb 2021

- Propose a new task of determining whether 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. The annotation quality is guaranteed by ensuring the Krippendorf's alpha is over 0.60.
- Construct several neural models to build baselines for this task. The experimental results show that the multi-modal model yields the best result.
- Error analyses indicate that the usage of advanced tools for location identification and advertisement filtering in data preprocessing has potential to strengthen neural models.

Temporal Ordering of Locations Oct 2019 - Apr 2021

- Determine whether 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 useful for this task.

Native Language Identification Aug 2018 - Dec 2018

- Determine the native languages of the authors of TOEFL essays.
- Use nltk to extract lots of linguistic features, such as POS tags, stop words, and punctuation marks.
- Construct linear regression classifier and the parameters were tuned using GridSearch with 5-fold cross-validation. The classification results are visualized by using seaborn.

TECHNICAL SKILLS

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