

# Context Helps Determine Spatial Knowledge from Tweets

Zhaomin Xiao, Yan Huang, Eduardo Blanco  
University of North Texas, University of Arizona



## Motivations

- Spatial information plays an important role in many applications, such as transportation planning and emergency management systems.
- Most previous work on spatial information with social media are on named entity recognition/disambiguation and location prediction.
- Location prediction, whose goal is to assign a location to a user, targets home location or real-time location.
- Contextual information has been proven in many tasks, such as hate speech detection and sentiment analysis. It can be user information, conversation, or history tweets.
- We are the first to tackle the problem of real-time location prediction with tweets, along with user's history tweets.

## Background

- AMT (Amazon Mechanical Turk) is a platform for annotation collection that allows people to publish and complete annotation tasks.
- MACE [1] is a tool that is designed to identify which annotators in AMT are trustworthy and predict the correct underlying annotations.
- BERT [2] is a pretrained language encoder, designed to generate a vector to capture the information contained in a sequence of text.
- LSTM [3] is a type of recurrent neural network which can capture long-term dependencies in sequential data.

## Objectives

- Construct a dataset of Twitter streams with spatial annotations. A Twitter stream consists of seven tweets posted chronologically.
- Build neural networks to predict the real-time locations of Twitter users using annotated Twitter streams.
- Conduct a qualitative analysis to provide insights into the errors made by the best-performing model.

## Methods

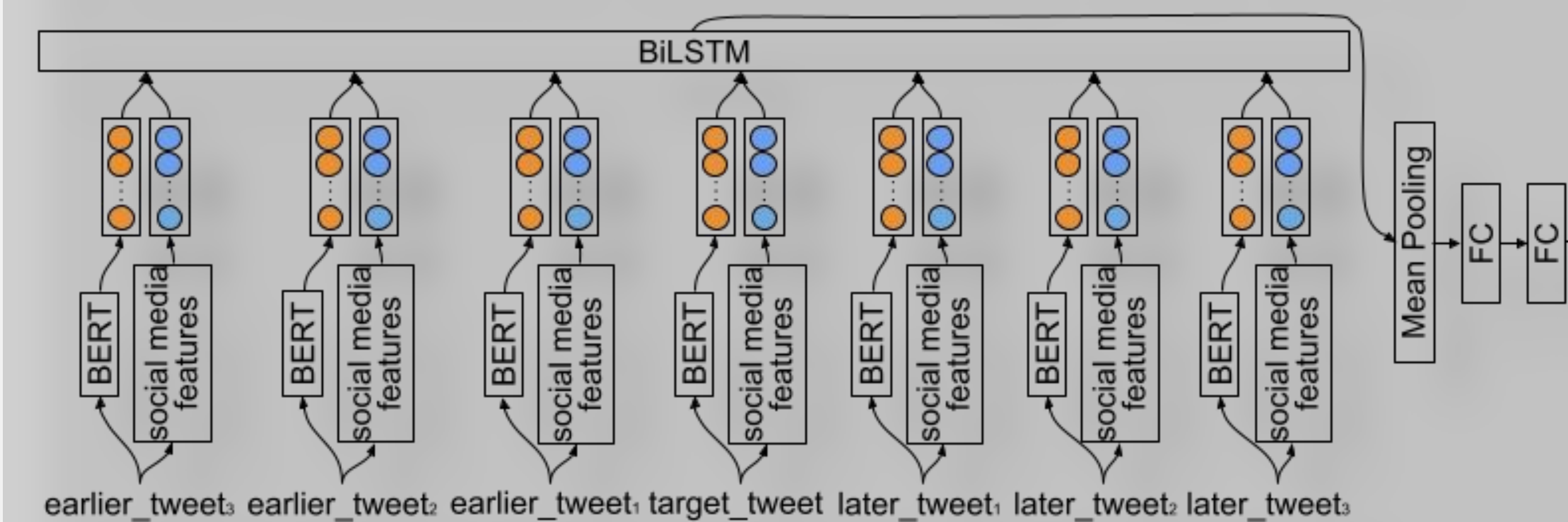
### Dataset Creation

Twitter streams → AMT → MACE → Our dataset

- Annotation question: Was the Twitter user at the mentioned location when the tweet was posted?
  - yes: The Twitter user was at the mentioned location.
  - no: I cannot tell if the Twitter user was at the mentioned location.

### Context-aware Neural Network

- Social media features: number of hashtags, emojis, URLs, etc.



## Results/Discussion

- Our dataset consists of 3,494 Twitter streams. 67.7% of them are annotated yes, and 32.3% are annotated no. Most annotations change depending on if we show annotators the context.

Earlier tweet: Sometimes it's just as good as it gets LOL! (9:14 PM - Jan 4, 2020)

Target tweet ( $t_0$ ): 6 Christmas tree recycling sites in Philadelphia open Mon-Fri., 8am-6pm. (4:23 PM - Jan 6, 2020)

Later tweet: And in South Philly, right outside my bedroom window, a big beautiful hawk! (12:01 PM - Jan 7, 2020)

Was the author in Philadelphia when  $t_0$  was published? Without context: no; With context: yes

Earlier tweet: The Mayor Pro Tem of Walnut, @AndrewForAsm55, introduced @PeteButtigieg at this event on the campus of @MtSAC (5:32 PM - Dec 20, 2019)

Target tweet ( $t_0$ ): As he hinted at earlier this week, @CoryBooker campaign announces Reno will be the first stop back on the trail (right on my neighborhood!) after Christmas (5:44 PM - Dec 20, 2019)

Later tweet: Spotted on the highway near San Gabriel in Southern California: @AndrewYang banner and American flag (6:48 PM - Dec 20, 2019)

Was the author in Reno when  $t_0$  was published? Without context: yes; With context: no

## Results/Discussion (Cont.)

### Results with neural networks

	no			yes			Weighted Average		
	P	R	F1	P	R	F1	P	R	F1
Majority baseline	0.00	0.00	0.00	0.68	1.00	0.81	0.46	0.68	0.55
Context-Unaware Network (target tweet)	0.00	0.00	0.00	0.68	1.00	0.81	0.46	0.68	0.55
Context-Aware Networks									
earlier + target tweets	0.00	0.00	0.00	0.68	1.00	0.81	0.46	0.68	0.55
target + later tweets	0.00	0.00	0.00	0.68	1.00	0.81	0.46	0.68	0.55
earlier + target + later tweets	0.44	0.28	0.35	0.71	0.83	0.76	0.62	0.65	0.63
without social media features	0.39	0.30	0.34	0.70	0.78	0.74	0.60	0.62	0.61

### Most common context-related errors

- Multiple named entities (e.g., Tom Hanks, Peter Liang) mislead the model.
- Lack of timestamps misleads the model.
  - Moving among Denver, San Diego, and Bahamas takes more than 5 hours (10:39 AM - 2:54 PM).

Multiple references to people and named entities (63%)

Earlier tweet: Do I feel sympathy for Tom Hanks? Knowing what I know about him and his family. (7:20 PM - Mar 13, 2020)

Target tweet ( $t_0$ ): These cops attacking Black college students in Miami on Spring Break. Some things never change. Assholes. (12:41 PM - Mar 14, 2020)

Later tweet: But you supported Peter Liang who killed an innocent Black person. Suck a dick. (1:13 PM - Mar 14, 2020)

Was the author in Miami when  $t_0$  was published? Ground Truth: no; Predicted: yes

Tweet timestamps are key (21%)

Earlier tweet: Like Denver airport's talking gargoyles? I don't like but it is better than nothing LOL (10:39 AM - Mar 1, 2019)

Target tweet ( $t_0$ ): They always posts this San Diego pic to get money in spring break. (12:30 PM - Mar 1, 2019)

Later tweet: Any news about Bahama's travel restrictions? really miss the ocean (2:54 PM - Mar 1, 2019)

Was the author in San Diego when  $t_0$  was published? Ground Truth: no; Predicted: yes

## References

- Hovy et al., Learning Whom to Trust with MACE (2013).
- Devlin et al., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (2019).
- Sepp Hochreiter and Jürgen Schmidhuber, Long Short-Term Memory (1997).