Study of Manifestation of Civil Unrest on Twitter

Abhinav Chinta*, Jingyu Zhang*, Alexandra DeLucia, Mark Dredze
Center for Language and Speech Processing
Johns Hopkins University

Anna L Buczak
Applied Physics Laboratory
Johns Hopkins University

Introduction

Motivation

Twitter is commonly used for civil unrest detection and forecasting tasks. However, there is a lack of work in evaluating *how civil unrest manifests on Twitter across countries and events*.

Research Goals

- Give in-depth case studies for two large-scale civil unrest events on Twitter that occurred in countries with (1) high Twitter usage and (2) low Twitter usage
- Build n-gram-based event detection models and use explainability tools (SHAP) to identify important features it learned from a new Twitter dataset: **Global Civil Unrest on Twitter (G-CUT)**

Findings

- While there is event signal during the events, there is little signal leading up to the events
- The event detection models were able to find words indicative of civil unrest that generalized across countries
- Location-bias exists as indicated by the discovery of the event detection models' reliance on location-specific keywords that do not generalize well

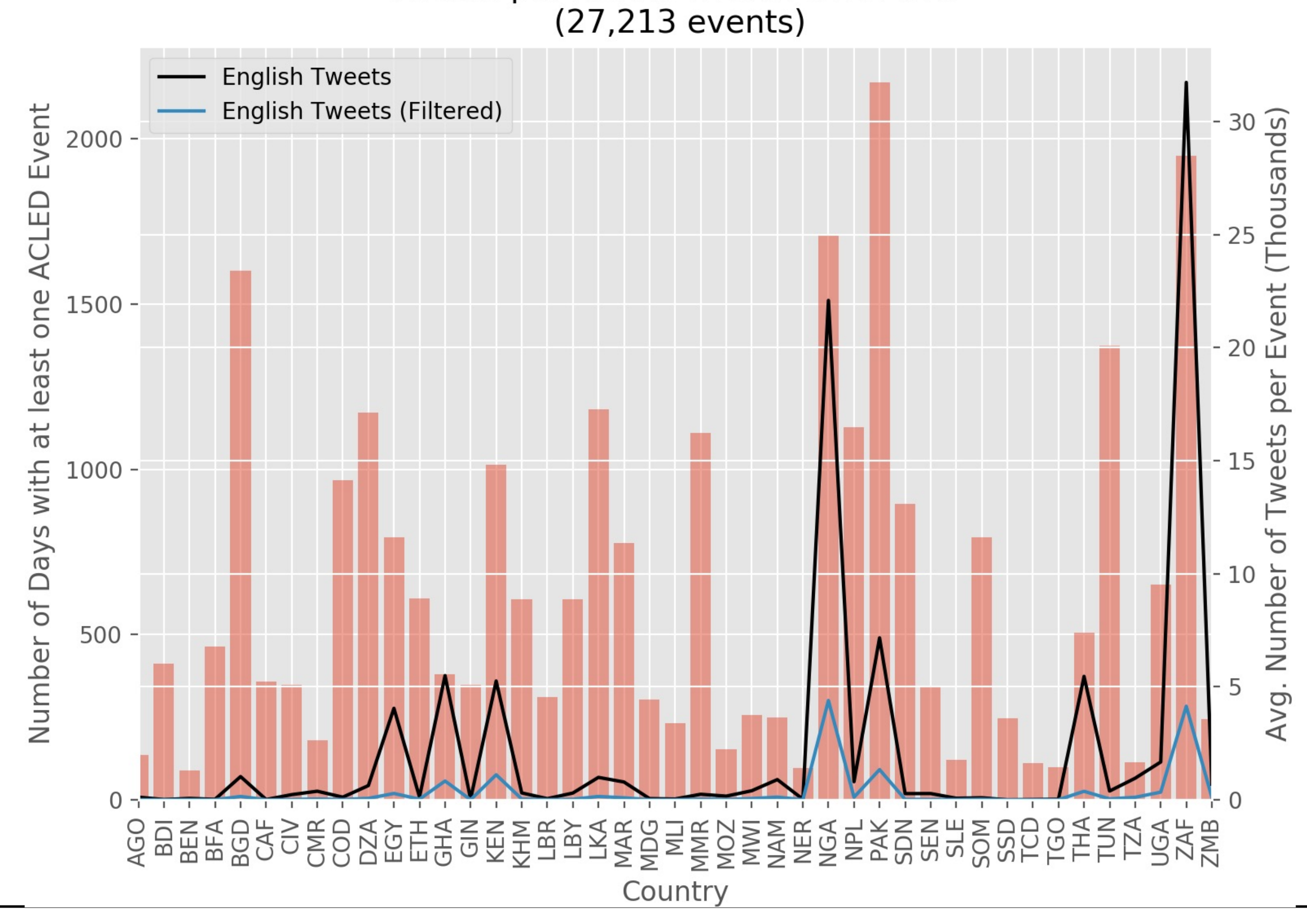
Global Civil Unrest on Twitter (G-CUT)

Dataset Creation

- Tweets from 2014-2019 collected from Twitter streaming API
- Filtered by geolocation to include 42 countries in Africa, the Middle East, and Southeast Asia
- Restricted to English tweets as identified by langid[1]
- Filtered tweets using the BERTweet civil unrest tweet classifier in Sech et al.[2]
- Labelled events with Riots and Protests from the Armed Conflict Location & Event Data Project (ACLED)[3] , a manually curated database of civil unrest events, as the ground truth data

Dataset Overview

Tweets per ACLED Event 2014-2019
(27,213 events)



Case Studies

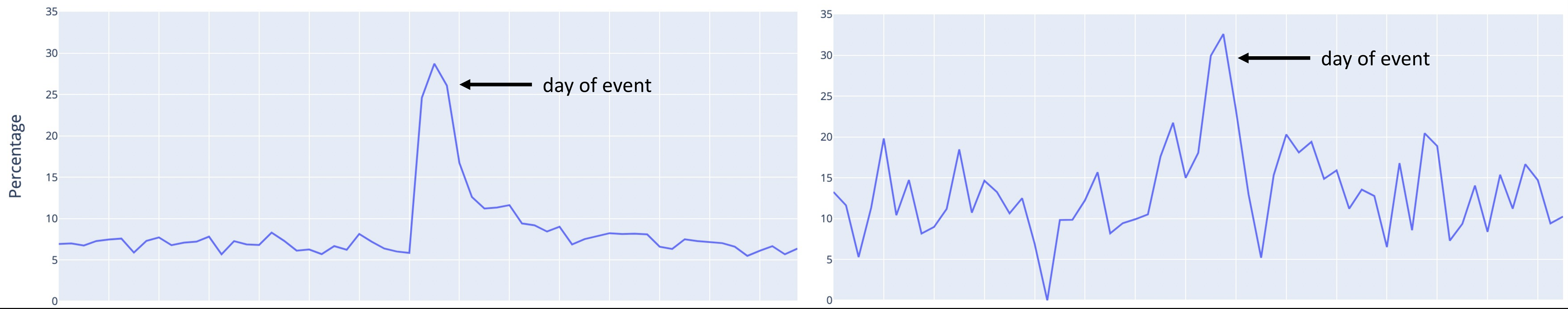
Framework

We compare events in two countries (South Africa and Ethiopia) and ask:
(1) is the event discussed on Twitter?
(2) is there any noticeable buildup in Twitter activity prior to the event?
(3) who is talking about the event?

Findings

- Both events have significant spike in tweets discussing civil unrest on the day of the event, but not before the event, regardless of Twitter user base size
- The South Africa event had event-specific hashtags but Ethiopia only had location-specific hashtags

Civil unrest-related tweets during the week of the
2019 Johannesburg Riots in South Africa (left) and the 2018 Burayu Massacre Protests in Ethiopia (right)



Indicators of Civil Unrest

Motivation

- The identified relevant tweets and hashtags from case studies are event specific (e.g. , **#etv** vs **#shutdownsouthafrica**)
- Want to discover trends/indicators of civil unrest across many countries and events

Indicator Extraction

- Used Shapley Additive explanation (SHAP)[4] to produce robust and consistent feature importance values on sparse count-based features
- Pick features (words) with largest SHAP value magnitudes as the "indicators"

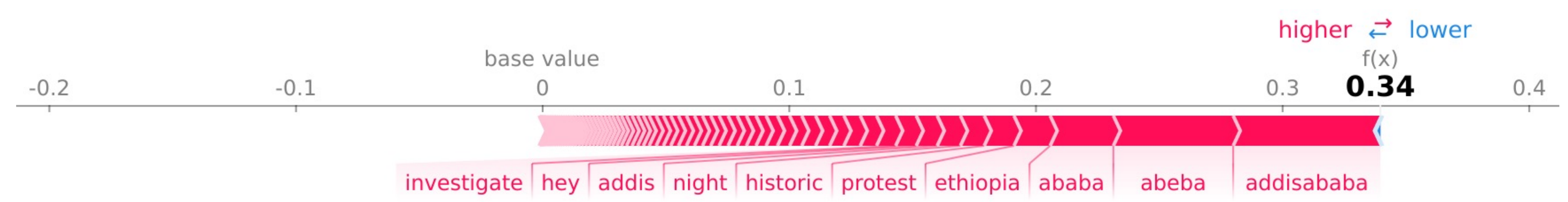
Experimental Setup

- Formulated the civil unrest event detection task as a binary classification problem to predict whether an event occurred in a particular country on a particular day
- The ground truth "positive" is days in a country with at least one event occurred
- Extracted n-gram token counts and aggregated by country and day
- Used a random forest model since it is both simple enough to be interpretable, less prone to overfitting, and powerful enough to capture relevant features from data

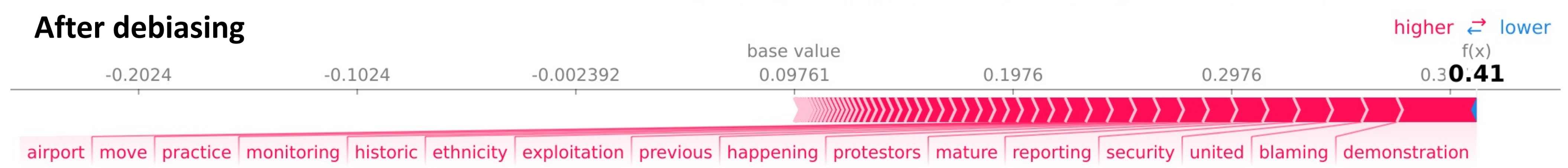
Country Debiasing

- Country and city names appear extensively because tweets discussing civil unrest are frequently location related
- Introduced "country debiasing" i.e., removing location-specific tokens from the data to encourage model generalizability across different geolocations
- Improved generalizability of most important features

Before debiasing



After debiasing



	F1	Precision	Recall
N-gram	0.45	0.58	0.37
N-gram (biased)	0.50	0.61	0.42
Random	0.29	0.32	0.27

Test scores for civil unrest detection task. Ground truth labels are from ACLED Riots and Protests.

Acknowledgements

The authors thank Arya McCarthy, Keziah Naggita, Sougol Gheissi, Rachel Wicks, Elizabeth Salesky, and the anonymous reviewers for their feedback. This work relates to Department of Navy award N00014-19-1-2316 issued by the Office of Naval Research. The United States Government has a royalty-free license throughout the world in all copyrightable material contained herein. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Office of Naval Research.

References

1. Marco Lui and Timothy Baldwin. 2012. langid.py: An Off-the-shelf Language Identification Tool. In Proceedings of the ACL 2012 System Demonstrations, pages 25–30, Jeju Island, Korea. Association for Computational Linguistics
2. Justin Sech, Alexandra DeLucia, Anna L. Buczak, and Mark Dredze. 2020. Civil Unrest on Twitter (CUT): A Dataset of Tweets to Support Research on Civil Unrest. In Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020), pages 215–221, Online. Association for Computational Linguistics.
3. Clionadh Raleigh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. 2010. Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature. Journal of Peace search. Publisher: SAGE PublicationsSage UK: London, England
4. Scott M. Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 30.