Characterizing Anti-Asian Rhetoric During The COVID-19 Pandemic: A Sentiment Analysis Case Study on Twitter

Ramya Tekumalla*[1], Zia Baig*[2], Michelle Pan*[3], Luis Alberto Robles Hernandez[1], Michael Wang[4], Juan M. Banda[1]

[1] Department of Computer Science, Georgia State University, [2] Faculty of Science, University of Waterloo, [3] Department of Electrical Engineering & Computer Sciences, University of California, Berkeley, [4] Coronavirus Visualization Team
{rtekumalla1,jbanda}@gsu.edu, z6baig@uwaterloo.ca, michellepan@berkeley.edu, lrobleshernandez1@student.gsu.edu, mwang@understandcovid.org,

Abstract

The COVID-19 pandemic has shown a measurable increase in the usage of sinophobic comments or terms on online social media platforms. In the United States, Asian Americans have been primarily targeted by violence and hate speech stemming from negative sentiments about the origins of the novel SARS-CoV-2 virus. While most published research focuses on extracting these sentiments from social media data, it does not connect the specific news events during the pandemic with changes in negative sentiment on social media platforms. In this work we combine and enhance publicly available resources with our own manually annotated set of tweets to create machine learning classification models to characterize the sinophobic behavior. We then applied our classifier to a pre-filtered longitudinal dataset spanning two years of pandemic related tweets and overlay our findings with relevant news events.

Introduction

The COVID-19 pandemic has brought on a wave of sinophobia, resulting in an increase in violence and hate speech targeted towards Asian Americans. In particular, reported anti-China rhetoric has included scapegoating of the Chinese, anti-immigrant nationalism, use of the "Chinese Virus" term, and use of orientalist depictions of Chinese culture (AAP 2021).

Much of this anti-Asian rhetoric has been spread online through social media, but its effect has extended to physical violence as well. Stop AAPI Hate details the anti-Asian hate incident reports they have received since March 2020. Of the 2583 incidents reported, 70% involved verbal harassment, 9% were physical assaults, and 8% involved a potential civil rights violation (AAP 2021).

This xenophobia has been exacerbated by the language and word choice used when reporting on the pandemic. News media, political figures, and government agencies have used language implying blame on China for the pandemic, such as "China virus" and "kung flu." On March 16, 2020, senior members of the United States government, including former President Donald Trump, accused Beijing of failing to warn other countries of the outbreak sooner and referred to the situation as a result of the "Chinese virus" (Karalis Noel 2020).

The first quarter of 2021 saw a surge in violence and hate crimes against Asian Americans. The Center for the Study of Hate and Extremism (CSHE) at Cal State University San Bernardino conducted an analysis using police department statistics (Levin 2020). The analysis spanned across 16 cities and counties in the United States. The CSHE observed that the number of reported hate crimes against Asians had increased by 164% since last year.

Media outlets have covered a number of violent attacks on Asian Americans. These have included a Filipino-American who was cut with a box cutter, an elderly Thai immigrant who died after being pushed to the ground, and a Chinese woman who was set on fire. In one night, eight individuals were killed in a shooting spree that spanned three Asian spas in Atlanta (BBC News 2021). According to advocates and activists, such incidents are fueled by the anti-Asian rhetoric that started from the COVID-19 pandemic a year ago.

There is an increasing level of concern within Asian American communities about their safety and inclusion within the United States. The Pew Research Center found that 81% of Asian American adults said that violence against them and their communities is increasing in the country, and 32% of the adults said that they feared someone might threaten or harm them. Approximately 20% of those who were surveyed cited former President Donald Trump's comments as a contributing factor to the surge in violence. Specifically, there was mention of the President's rhetoric about the pandemic originating from China or his labeling of the coronavirus as the "Chinese flu" or "kung flu" (Yam 2021).

In this work, we aim to analyze anti-Asian sentiment expressed in COVID-19-related tweets. Similar to other approaches (He et al. 2021; Vidgen et al. 2020), we finetune a transformer language model to identify hate, counter-hate, and neutral tweets. We then deploy our model on a filtered set of tweets and calculate the frequency of hateful tweets over time alongside related news reports.

Over the years, researchers have been able to track and quantify racism in social media platforms like Facebook (Rauch and Schanz 2013), Reddit (Yang and Counts 2018),

*Authors contributed equally

Copyright © 2022, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

4chan (Hine et al. 2017), and Twitter (Chaudhry 2015). While previous works study racism in general or in individual country-specific topics, the COVID-19 pandemic has provided a common ground of an event that affects all countries, hence creating the perfect storm to evaluate racism across timezones in a common topic. With the COVID-19 pandemic exacerbating xenophobia due to the origin of the virus, several other researchers have turned to social media to examine trends found in Twitter about negative tweets as the pandemic developed (Pei and Mehta 2020). Others have turned to sentiment analysis to quantify the emotions behind cyber racism (Dubey 2020). With the advantage of sites like Reddit that provide structured groups (Zhang et al. 2020) and Twitter that use certain hashtags to tag specific content (Lyu et al. 2020), large-scale analyses during the pandemic have become the norm. Web communities like 4chan have been used to identify slurs and racial terms in order to fully understand the sinophobic behavior of their users (Schild et al. 2020). These terms are highly correlated with what traditional studies have found (Gover, Harper, and Langton 2020). Some researchers have focused on the granularity of these hate and counterhate tweets and how they spread across social networks (He et al. 2021), while others have tried to identify the demographics and account characteristics of the users producing such posts (Lyu et al. 2020). The traditional way of using Twitter for such analyses relies on manual curation of sets of tweets, as He et al. (He et al. 2021) have demonstrated. Our contributions in this paper are the following: 1) we present an additional non-overlapping manually curated dataset for hate, counterhate, and neutral COVID-19 tweets. 2) we combine our datasets with the manually annotated datasets of (He et al. 2021) and (Vidgen et al. 2020) to create a larger training set for machine learning models. 3) we fine tuned state-of-the-art Twitter specific language models: BERTweet (Nguyen, Vu, and Tuan Nguyen 2020) and Covid Twitter BERT (CT-BERT) (Müller, Salathé, and Kummervold 2020), achieving state-of-the-art performance results in separating sinophobic tweets. 4) we applied the fine-tuned model to a pre-filtered dataset of COVID-19 tweets, covering over two full years, making it the largest longitudinal evaluation of sinophobic tweets to date.

Methods

Our methodology is split into several tasks: sinophobic slurs identification, dataset selection, manual curation, model building, filtering and classification of COVID-19 tweets.

Sinophobic slurs identification

As a starting point, we conducted a slur frequency analysis on a limited initial set of 28,651 randomly selected tweets from the Internet Archive Twitter stream grab collection (Internet 2011) using a custom dictionary of 45 key terms. These key terms encompass slurs that were used against people from China or of Chinese descent and have been gathered from Hatebase (Hatebase.Org 2022) and the Wikipedia list of slurs (Wikipedia contributors 2022).This initial discovery set allowed us to refine and enhance our selection terms to a list of 151, based on additional frequent terms and lexical variants, listed in Table 1.

baby muncher, baby-muncher, bamboo coon, bucket-head, bugland, bugman, bugmen, chankoro, chiegro, chigger, china devil, china-devil, chinaman, chinamen, chinazi, chinee, chinegro, chinese devil, ching chong, ching-chong, chingchong, chinig, chinina, chink, chink a billies, chink a billy, chinkerbell, chinki, chinksta, chino, chonkies, chonky, choo-choo, chork, chunkies, chunky, cina, cokin, coolie, crabrangook, dink, dinkladies, dinklady, dog eater, dog-eater , dog-muncher, dwo, egghead, el chino, fingernail rancher, fishhead, forty fiver, forty-fiver, gink, goloid, gong fei, gook, gook eye, gook eyed, gookemon, gookette, gookie, gooklet, gooky eyes, honger, honkies, honky, hun, insectoid, jaundy boy, jaunta, jek, jin, jjanggae, johk sing, lemonhead, momo, nink, noodle nigger, nooger, pancake face, pancake-face, panface, panhead, pastel de flango, ping pang, ping-pang, pizdaglaz, plate tosser, pointy head, pointy-head, rice nigger, round eye, roundeye, schlitzauge, sideways cooter, sideways pussies, sideways pussy, sideways vagina, slant eye, slant-eye, slantey eye'd, slantey eyed, slantey-eye'd, slantey-eyed, slants, slanty eyed, slanty-eyed, slope head, slope-head, slopehead, sloper, slopey, slopies, slopy, socket face, socket-face, spink, squint nigger, table face, table-face, tai chink, tai-chink, tanka, tape head, tape-head, thin eye, thin eyed, thin-eye, thin-eyed, ting tong, ting-tong, touch of the tar brush, twinkie, whoriental, wog, woggle, yellow cab, yellow devil, yellow invader, yellow man, yellow monkey, yellow on the outside, yellow woman, yellowman, yellowwoman, yigger, zip, zipper, zipperhead, zippohead

Table 1: Sinophobic slurs

Data Selection

We utilized a large longitudinal COVID-19 tweets dataset in addition to previously curated and annotated sets of tweets. Note that we only selected manually annotated tweets for our model building process as we are focusing on quality rather than quantity.

Twitter COVID-19 dataset As one of the largest COVID-19 chatter datasets available (Banda et al. 2021), we used version 95, which includes over 1.28 billion tweets from January 2020 to January of 2022. We started by removing all retweets, leaving us with 328,851,757 tweets to utilize. Focusing on only English speaking tweets that have Asian and sinophobic expressions for practical and better precision purposes, we filtered this larger set into a total of 7,311,677 tweets which we will use for our manual curation and downstream classification tasks.

Manually curated publicly available sinophobia datasets During the development of this work, other authors created and released manually curated datasets of their own.

The COVID-HATE dataset by He et al. (He et al. 2021) released several million sinophobic tweets, however, only

2,290 were manually curated: 429 hate tweets, 1,344 neutral tweets and 517 counter-hate tweets.

In Vidgen et al. (Vidgen et al. 2020) the authors use experts to annotate (and release) 20,000 tweets with five different classes: Hostility against East Asia, Criticism of East Asia, Counter speech, Meta-discussions of East Asian prejudice, and a neutral class. We combined the hostility and criticism labels into the hate category, a total of 5,331 tweets and used the 116 counter speech annotated tweets.

Note that we selected these two datasets to aggregate them to the non-overlapping tweets from our curated dataset described in the following section. Researchers have theorized (He et al. 2021) that aggregating the available datasets would yield better machine learning models, a claim we prove in this work.

Manual Data Curation

Due to the ambiguity in the expression of sentiments, tools like VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hutto and Gilbert 2014) are useful for general measurements but not fine-grained topics, such as sinophobia. Hence the need to manually curate sets of tweets.

For our manually curated dataset, we labeled the sentiment of a randomly selected set of 10,000 extracted from (Banda et al. 2021) between January 27th 2020, and October 31st 2020. Tweets were labeled as hate if they contained racial slurs, scapegoating, or anti-Asian rhetoric. Tweets were labeled as counter-hate if they criticized sinophobic language and actions. All other tweets were considered neutral. Our initial curation identified 3,797 neutral tweets, 961 hate-related tweets, and 658 counter-hate tweets. In Table 2, we present the calculated pairwise Cohen kappa statistic (κ) (Cohen 1960). If both annotators are in complete agreement then κ = 1, in the worst case, if there is no agreement among the annotators (other than what would be expected by chance) then $\kappa \leq 0$. We resolved disagreements by having an external reviewer adjudicate.

Category	Labels	κ
Hate	961	0.88
Neutral	3,797	0.92
Counter-hate	32	0.95

Table 2: Annotator Agreement

In order to increase the classification power of any machine learning model, as mentioned in the previous section we added additional publicly available manually curated data from He et al. 2020 (He et al. 2021) and Vidgen et al. (Vidgen et al. 2020). The final dataset is described in Table 3.

Model Building

In order to capture the fine-grained nuances of the sinophonic sentiment in tweets, we evaluate ten different machine learning models. Ranging from classical models: Logistic Regression (LR) (Cox 1958), Multi-nomial Naive-Bayes (NB) (Kibriya et al. 2004), Decision Trees (DT) (Quinlan 1986), Random Forest (RF) (Breiman 2001), and Support Vector Machine (SVM) (Cortes and Vapnik 1995) with a linear kernel, to deep learning models like biLSTM (Schuster and Paliwal 1997), and several different fine-tuned flavors of BERT-based transformers (BERT (Devlin et al. 2019), RoBERTa (Liu et al. 2019)), including Twitter-specific ones (BERTweet (Nguyen, Vu, and Tuan Nguyen 2020) and Covid-Twitter-BERT (Müller, Salathé, and Kummervold 2020)), an improvement over previous works in the space (He et al. 2021; Vidgen et al. 2020). All text preprocessing was carried out using the python utilities called Social Media Mining Toolkit (Tekumalla and Banda 2020).

Dataset	Hate	Neutral	Counter
Our dataset	961	3,797	32
Curated (He et al. 2021)	670	908	358
Curated (Vidgen et al. 2020)	5,331	0	116
Total	**6,721**	**5,141**	**665**

Table 3: Aggregated dataset statistics

For reproducibility purposes, we provide the model training and fine-tuning setup used. We used the Scikit-learn (Pedregosa et al. 2011) library for the classical models and the TF-IDF vectorizer was used to convert raw tweet text to TF-IDF features and return the document-term matrix which is sent to the model. The transformer models were implemented using Simple Transformers (Rajapakse 2019) and in built method is utilized to convert the text to features. The only requirement for the transformers library is to have a dataframe which consists of only text and label which is converted to features and then sent to model for training. The LSTM model was implemented using Keras framework (Kowsari et al. 2019). We tokenized the text and then created sequences of tokenized words and padded the sequences to form a text. Further we used a word2Vec model to identify the embeddings and assigned the index to words of the text which was used to build the model. We used Twitter Word2vec Embeddings (Godin 2019) since we utilized Twitter data in this paper. All experiments used a 90% training and 10% testing split.

Model	Hyper Parameters
SVM - Linear SVC	default
Logistic Regression (LR)	Max_iter = 1000
Decision tree (DT)	Max_features = auto criterion = entropy max_depth = 150
Random Forest (RF)	Max_features = auto criterion = entropy;
Multinom. Naive Bayes (NB)	default

Table 4: Classical ML models parameters

Additionally, for comparison purposes we took the best performing models from He et al. (He et al. 2021) and Vidgen et al. (Vidgen et al. 2020).

Results

With the high class imbalance from the training dataset we present several analyses to determine the best performing

Model	Pre-trained model	Configurations
BERT	bert-large-uncased	24-layer, 1024-hidden, 16-heads, 340M parameters
RoBERTa	roberta-large	24-layer, 1024-hidden, 16-heads, 355M parameters
BERTweet	bertweet-base	12-layer, 768-hidden, 12- heads,135M parameters
COVID-Twitter-BERT	COVID-Twitter-BERT v2	24-layer, 1024-hidden, 16-heads, trained on 1.2B samples
LSTM	N/A	Adam Optimizer, bidirectional, max sequence length= 280, dropout=0.2, softmax activation function, Glove Embedding model

Table 5: Deep Learning models parameters

model for all three classes: hate, neutral, counter-hate.

Model Comparison

Due to the small changes in performance between machine learning models, we will present Tables 6 and 7 for the precision and recall results.

Model	Hate	Neutral	Counter
Logistic Regression	0.9051	0.8154	0.8333
Decision Tree	0.7477	0.6348	0.4462
Naive Bayes	0.8590	0.8220	0.9667
Random Forest	0.6862	0.8323	**1.0000**
SVM	0.8942	0.8374	0.8936
LSTM	0.8837	0.8250	0.6667
BERT	0.9022	0.8831	0.7083
RoBERTa	**0.9256**	0.8913	0.7467
BERTweet	0.8799	0.9034	0.7746
CT-BERT	0.9233	**0.9124**	0.7887

Table 6: Precision metric for trained models

One of the main goals of the models built is that they can separate the hate and counter-hate classes efficiently. By only looking at precision (Table 6), we get misleading insights, such as the random forest model having a 100% precision for the counter-hate class, but being the worst for the hate class. However, there are other more balanced models with leveled performance across all classes like the SVM and logistic regression models. We investigated the recall metric to make sure that the models are performing in a balanced way when distinguishing between the three classes.

In the recall table we see that while most models perform well at separating the hate and neutral classes, only a handful (the BERT-based) models perform well for the counter-hate class. This is due to the high class imbalance in the dataset (nearly 10 to 1 for between hate/neutral to counter-hate. We also see that when random forest has a 100% precision for the counter-hate class, it has an abysmal 9% recall.

Model	Hate	Neutral	Counter
Logistic Regression	0.8795	**0.8852**	0.5224
Decision Tree	0.7277	0.6595	0.4328
Naive Bayes	**0.9568**	0.5019	0.0896
Random Forest	0.8884	0.8444	0.4328
SVM	0.8929	0.8716	0.6269
LSTM	0.8929	0.8346	0.5373
BERT	0.9196	0.8521	0.7612
RoBERTa	0.9256	0.8774	**0.8358**
BERTweet	0.9375	0.8191	0.8209
CT-BERT	0.9494	0.8716	**0.8358**

Table 7: Recall metric for trained models

In order to select the best-performing model to classify our tweets, we use weighted precision, recall, and F1-score to compare the models. This will measure the average performance metrics across classes and with the addition of the F1-score we can confidently select the best model.

Model	Precision	Recall	F1-score
Logistic Regression	0.864	0.863	0.861
Decision Tree	0.685	0.684	0.684
Naive Bayes	0.763	0.724	0.694
Random Forest	0.850	0.846	0.842
SVM	0.871	0.870	0.869
LSTM	0.848	0.850	0.849
BERT	0.884	0.883	0.884
RoBERTa	0.902	0.901	0.901
BERTweet	0.884	0.883	0.882
CT-BERT	**0.912**	**0.911**	**0.911**

Table 8: Performance metrics for trained models

Based on the weighted metrics, the CT-BERT model performs the best with a precision of 0.912, a recall of 0.911, and a F1-score of 0.911. While not completely unexpected, it makes sense that a transformer model built using Twitter data outperform more general-purpose BERT models. One interesting finding was that the logistic regression and SVM classifiers performed as well as they did. However, these models probably do not generalize as they are only trained on the data available in the dataset. While the BERT models are trained on large amounts of data and fine tuned with the dataset data and classes, allowing for better generalization.

In order to evaluate the performance metrics of our fine-tuned model in comparison to the models built by He et al. (He et al. 2021) and Vidgen et al. (Vidgen et al. 2020), we list their weighted precision, recall and F1-score on table 9.

Note that while the model from He et al. (He et al. 2021) features the same classes as our dataset, the model from Vidgen et al. (Vidgen et al. 2020) has two additional classes,

Model	Precision	Recall	F1-score
He et al. (BERT)	*0.830*	*0.834*	*0.832*
Vidgen et al. (RoBERTa)	*0.850*	*0.830*	*0.830*
CT-BERT	**0.912**	**0.911**	**0.911**

Table 9: Performance metrics of best models

some that we have collapsed to make our comparison. We show that our fine tuned CT-BERT model improved precision by 7 points, while it improved recall and F1-score by nearly 8 points. This is important considering that we will be applying this model to a set of several million tweets. Having a theoretical error rate of 9 out of 100 tweets is considerably better than 15 or 17 out of 100 of the other two models respectively. We compare the performance of the same classes between He et al., Vidgen et al., and our model to showcase that we have higher accuracy for the hate and counter-hate classes, which leads to higher yield of relevant tweets when classifying at scale. This comparison is presented to elucidate the utility of reusing, combining, and enhancing datasets for fine-grained tasks such as this one.

Longitudinal Analysis of Sinophobic Tweets

Now that we have determined the best classifier model, we will classify the previously filtered set of 7,311,677 tweets (minus the 10,000 tweets we annotated). Note that we decided to pre-filter our longitudinal dataset in order to focus on identifying the sentiment of COVID-19 discourse tweets that have already been identified to have sinophobic slurs or mentions of China. As shown by He et al. (He et al. 2021), if such a classifier is applied to a large dataset, it will yield more than 98% neutral tweets. This means that over 200 million tweet classification operations were made that did not yield interesting tweets. In our compute environment, applying a classifier to a set of one million tweets takes around 65 minutes of GPU compute time. Hence we set up our evaluation to have a higher yield and use less resources.

Class	Total	Percentage
Hate	106,593	1.46%
Neutral	7,190,143	98.47%
Counter-hate	4,941	0.07%

Table 10: Classification results statistics

While we still have a very low yield of hate (1.46%) and counter-hate (0.07%) classes, we are less likely to have additional false positives due to the pre-filtering step. We now turn our analysis to evaluate how the proportion of hateful tweets change over time from January 1st, 2020 to December 31st, 2021. We calculate proportion, rather than total count of hate tweets, in order to plot an absolute measure per day and avoid biasing our figures to days with a higher count of tweets.

In Table 11 and Figure 1 we demonstrate each date when a peak occurs, and relate it to significant events that occurred on that day. These events consist of CNN and CBS news headlines, official government statements, as well as statements made by former President Donald Trump. These news events are relevant since several researchers observed the impact created. When the U.S included race in coronavirus vaccine plans, Schmidt et al. researched on laws and ethics to prioritize racial minorities for Covid-19 vaccine (Schmidt, Gostin, and Williams 2020). Similarly other researchers observed that though Anti-Asian violence surged in the U.S since Covid-19, it didn't start with Covid-19 (Gover, Harper, and Langton 2020).

Discussion

The results show a high proportion of hate tweets posted late January 2020 and in the month of February 2020. This is due to the fact that the start of the pandemic involved a lot of uncertainty. Many major political figures initially used terms such as “china virus”, “wuflu”, and “kung flu”. Some individuals blamed Chinese citizens for the spread of the Coronavirus, which was followed by a reported increase in Anti-Asian hate crimes.

Major events during the COVID-19 pandemic were seen to influence the sentiment on social media platforms, particularly Twitter. Most notably, a large negative count was observed late January to early February 2020, when the WHO declared the virus a Public Health Emergency of International Concern. Further negative sentiment was observed around March 2020 when the Trump administration insisted the U.N. named China as the origin of the coronavirus, and right-wing influencers were led to believe that Dr. Anthony Fauci was working with Hillary Clinton to undermine Donald Trump. Events during the COVID-19 pandemic, influenced by politics and policy making, show significant online activity related to sinophobic behavior. This behavior becomes more evident when leaders in the community demonstrate any such traits or ideas that may have an adverse effect on Asian populations. It is interesting to note that over the duration of the pandemic the overall negative sentiment towards Asian populations has decreased in the last few months of 2020, despite case/hospitalization due to the Omicron variant. While some smaller flare ups can be identified in 2021, these are not as large as during the first year, probably due to pandemic fatigue setting in on English speaking communities.

Design limitations include the difficult to detect sarcasm or the use of rhetorical devices using classifiers, particularly in short social media messages. However, this work demonstrates that fine-grained models can find relevant tweets as they correlate with important news events.

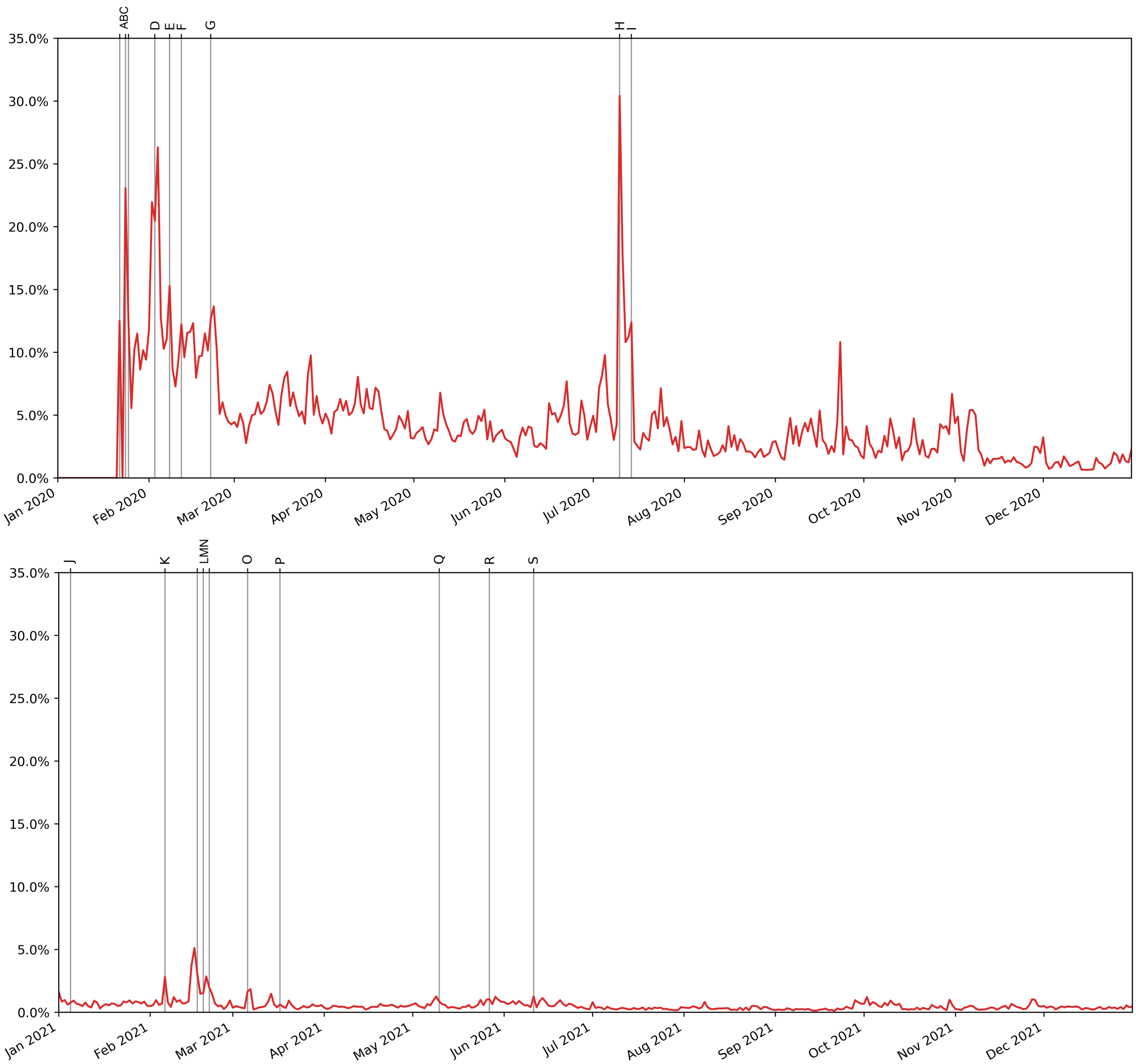

Figure 1. Proportion of negative tweets between 2020 and 2021

Date	ID	Description	URL
1/22/20	A	Chinese authorities quarantine the entire city of Wuhan to prevent the further spread of the novel coronavirus as the death toll rises to 17.	https://www.theguardian.com/world/2020/jan/22/coronavirus-china-measures-rein-spread-mutate-disease-death-toll
1/24/20	B	China expanded a travel lockdown in central China essentially penning in more than 35 million residents	https://www.nytimes.com/2020/01/23/world/asia/china-coronavirus-outbreak.html
1/25/20	C	U.S. government actively works to evacuate American citizens by air from the epidemic-stricken Chinese city of Wuhan	https://www.wsj.com/articles/u-s-plans-to-evacuate-citizens-from-epidemic-stricken-chinese-city-11579951256?mod=hp_lead_pos2
1/31/20	D	WHO declared the virus a Public Health Emergency of International Concern	https://www.euro.who.int/en/health-topics/health-emergencies/international-health-regulations/news/news/2020/2/2019-ncov-outbreak-is-an-emergency-of-international-concern
2/8/20	E	one Japanese and one American – die from coronavirus in Wuhan	https://www.nbcnews.com/news/world/coronavirus-updates-u-s-japanese-citizens-die-wuhan-global-deaths-n1132951
2/12/20	F	The United States Postal Service suspends time guarantees for all shipments to China and Hong Kong.	https://www.scmp.com/news/china/article/3050103/coronavirus-us-postal-service-suspends-items-destined-china-and-hong
2/22/20	G	South Korea confirms that 229 more people have contracted the coronavirus in the country.	https://www.bbc.com/news/world-asia-51596665
7/10/20	H	U.S. Considers Race in Coronavirus vaccine Plans	https://www.nytimes.com/issue/todaysheadlines/2020/07/10/todays-headlines
7/14/20	I	Moderna is expected to start a late stage clinical trial at 87 study locations in the United States for its COVID-19 vaccine on July 27	https://www.reuters.com/article/us-health-coronavirus-moderna-phase3-idUSKCN24F231
01/05/21	J	WHO Chief 'Disappointed' China Hasn't Allowed Researchers into Wuhan	https://www.voanews.com/a/covid-19-pandemic_who-chief-disappointed-china-hasnt-allowed-researchers-wuhan/6200364.html
2/06/21	K	Li Wenliang: 'Wuhan whistleblower' remembered one year on. Dr Li had tried to warn fellow medics of a disease that looked like Sars - another deadly coronavirus.	https://www.bbc.com/news/world-asia-55963896
2/17/21	L	COVID fallout: Biden wants to stop, not stoke, racism against Asian Americans like me. Trump fueled aggression against Asian groups with phrases like 'Chinese virus.' Words matter, especially those used by the leaders of the free world.	https://www.usatoday.com/story/opinion/voices/2021/02/17/covid-fallout-racism-against-asian-americans-column/6761944002/
2/19/21	M	Covid-19 Was Spreading in China Before First Confirmed Cases, Fresh Evidence Suggests	https://www.wsj.com/articles/covid-19-was-spreading-in-china-before-first-confirmed-cases-fresh-evidence-suggests-11613730600
2/21/21	N	WHO panel to recommend 'deeper' study of early Covid-19 clues.	https://www.cnn.com/2021/02/21/china/who-covid-19-origins-intl/index.html
3/06/21	O	Anti-Asian violence has surged in the U.S. since Covid-19. But it didn't start there.	https://www.nbcnews.com/news/asian-america/anti-asian-violence-has-surged-u-s-covid-19-it-n1259858
3/17/21	P	How the WHO's Hunt for Covid's Origins Stumbled in China. A team of scientists hoped a mission to Wuhan would provide some clarity about the virus's origins.	https://www.wsj.com/articles/who-china-hunt-covid-origins-11616004512
5/10/21	Q	'We were scared': Asian-owned small businesses devastated by double whammy of Covid and hate crime	https://www.cnbc.com/2021/05/10/covid-and-racism-are-devastating-for-many-asian-owned-small-businesses.html
5/27/21	R	New Report: Asian Americans Face Unprecedented Mental Health Concerns Due to the COVID-19 Pandemic and Anti-Asian Hate	https://aapaonline.org/2021/05/new-report-asian-americans-face-unprecedented-mental-health-concerns-due-to-the-covid-19-pandemic-and-anti-asian-hate/
6/11/21	S	China brands COVID-19 lab-leak theory as 'absurd,' Blinken urges transparency	https://www.reuters.com/world/china-us-top-diplomats-hold-phone-call-chinese-state-media-2021-06-11/

Table 11: News events tied to increases of sinophobic tweets

Conclusions

Since the beginning of 2020, there has been an observed increase in the usage of sinophobic comments or terms on online social media platforms. This online behavior has translated to action - verbal or physical - in the real world. In the United States, Asian Americans have primarily been targeted by violence and hate speech stemming from negative sentiments about the origins of the novel SARS-CoV-2 virus. The development of technology has helped to identify the correlations between real-world events and the anti-Asian rhetoric on social media. Behavioral insights can be derived to assess the sentiment of the general public. The COVID-19 pandemic has spotlighted many critical aspects of society. Unfortunately, during this unprecedented time, discrimination and racism against Asian and Chinese communities have ascended to staggering heights. The use of negatively charged comments against Asian Americans has perpetuated a harmful rhetoric that has fuelled hate crimes and violence against such communities. The pandemic has shown that individuals tend to place blame on often uncontrollable events or external groups (the 'others'). Many tend to minimize the similarities and emphasize the differences. The coronavirus pandemic represents an opportunity to reevaluate how individuals interact with each other over social media and how negatively charged rhetoric can influence the public perception of certain communities. This works adds additional manually curated data, and a state-of-the-art performant classification model to the sinophobia detection literature. For reproducibility and future analysis, our annotated dataset, our fine-tuned model, and all code to fuse datasets, train models, and make predictions on new tweets, is available for download in the following repository: https://10.5281/zenodo.6523152.

Acknowledgments

Coronavirus Visualization Team: Lucas Chu, Valerie Aguilar Dellisanti, Anson Yu, Nima Pourjafar, Vansh Tibrewal. This work was supported from a grant by the National Institute on Aging (3P30AG059307-02S1).

References

2021. STOP AAPI HATE. https://stopaapihate.org/. Accessed: 2021-1-25.

Banda, J. M.; Tekumalla, R.; Wang, G.; Yu, J.; Liu, T.; Ding, Y.; Artemova, E.; Tutubalina, E.; and Chowell, G. 2021. A Large-Scale COVID-19 Twitter Chatter Dataset for Open Scientific Research—An International Collaboration. *Epidemiologia* 2(3): 315–324.

BBC News. 2021. Covid 'hate crimes' against Asian Americans on rise. *BBC* .

Breiman, L. 2001. Random Forests. *Mach. Learn.* 45(1): 5–32.

Chaudhry, I. 2015. #Hashtagging hate: Using Twitter to track racism online. *First Monday* .

Cohen, J. 1960. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* 20(1): 37–46.

Cortes, C.; and Vapnik, V. 1995. Support-vector networks. *Machine learning* 20(3): 273–297.

Cox, D. R. 1958. The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)* 20(2): 215–232.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proc. of the 2019 Conference of the North American Chapter of the ACL: Human Language Technologies, Volume 1*, 4171–4186. Association for Computational Linguistics.

Dubey, A. D. 2020. The Resurgence of Cyber Racism During the COVID-19 Pandemic and its Aftereffects. *JMIR Public Health Surveill* 6(4): e19833.

Godin, F. 2019. *Improving and Interpreting Neural Networks for Word-Level Prediction Tasks in Natural Language Processing*. Ph.D. thesis, Ghent University.

Gover, A. R.; Harper, S. B.; and Langton, L. 2020. Anti-Asian Hate Crime During the COVID-19 Pandemic: Exploring the Reproduction of Inequality. *Am. J. Crim. Justice* 1–21.

Hatebase.Org. 2022. Hatebase:Search Results. shorturl.at/bjxF9. Accessed: 2022-4-8.

He, B.; Ziems, C.; Soni, S.; Ramakrishnan, N.; Yang, D.; and Kumar, S. 2021. Racism is a virus: anti-asian hate and counterspeech in social media during the COVID-19 crisis. In *Proc. of the 2021 IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining*, ASONAM '21, 90–94. ACM.

Hine, G.; Onaolapo, J.; De Cristofaro, E.; and et al. 2017. Kek, Cucks, and God Emperor Trump: A Meas. Study of 4chan's Politically Incorrect Forum and Its Effects on the Web. *ICWSM* 11(1).

Hutto, C. J.; and Gilbert, E. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.

Internet, A. 2011. Archive Team:The Twitter Stream Grab : Free Web : Free Download, Borrow and Streaming : Internet Archive. shorturl.at/jsvM9. Accessed: 2022-4-8.

Karalis Noel, T. 2020. Conflating culture with COVID-19: Xenophobic repercussions of a global pandemic. *Soc Sci Humanit Open* 2(1): 100044.

Kibriya, A. M.; Frank, E.; Pfahringer, B.; and Holmes, G. 2004. Multinomial naive bayes for text categorization revisited. In *Proc. of the 17th Australian joint conference on Advances in AI*, AI'04, 488–499. Springer-Verlag.

Kowsari, K.; Jafari Meimandi, K.; Heidarysafa, M.; Mendu, S.; Barnes, L.; and Brown, D. 2019. Text Classification Algorithms: A Survey. *Information* 10(4): 150.

Levin, B. 2020. Report to the Nation: Anti-Asian Prejudice & Hate Crime Center for the Study of Hate & Extremism. Technical report, Center for the Study of Hate & Extremism.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach .

Lyu; Chen; Wang; and Luo. 2020. Sense and Sensibility: Characterizing Social Media Users Regarding the Use of Controversial Terms for COVID-19. *IEEE Transactions on Big Data* 1–1.

Müller, M.; Salathé, M.; and Kummervold, P. E. 2020. COVID-Twitter-BERT: A Natural Language Processing Model to Analyse COVID-19 Content on Twitter .

Nguyen, D. Q.; Vu, T.; and Tuan Nguyen, A. 2020. BERTweet: A pre-trained language model for English Tweets. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language*

Processing: System Demonstrations, 9–14. Online: Association for Computational Linguistics.

Pedregosa, F.; Varoquaux, G.; Gramfort; and et al. 2011. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12(null): 2825–2830.

Pei, X.; and Mehta, D. 2020. #Coronavirus or #Chinesevirus?!: Understanding the negative sentiment reflected in Tweets with racist hashtags across the development of COVID-19 .

Quinlan, J. R. 1986. Induction of Decision Trees. *Mach. Learn.* 1(1): 81–106.

Rajapakse, T. 2019. simpletransformers: Transformers for Classification, NER, QA, Language Modelling, Language Generation, T5, Multi-Modal, and Conversational AI .

Rauch, S. M.; and Schanz, K. 2013. Frequency and purpose of Facebook use and the acceptance of prejudiced and egalitarian messages. *Comput. Human Behav.* 29(3): 610–615.

Schild, L.; Ling, C.; Blackburn, J.; Stringhini; and et al. 2020. "Go eat a bat, Chang!": An Early Look on the Emergence of Sinophobic Behavior on Web Communities in the Face of COVID-19 .

Schmidt, H.; Gostin, L. O.; and Williams, M. A. 2020. Is It Lawful and Ethical to Prioritize Racial Minorities for COVID-19 Vaccines? *JAMA: the journal of the American Medical Association* 324: 2023–2024.

Schuster, M.; and Paliwal, K. K. 1997. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* 45(11): 2673–2681.

Tekumalla, R.; and Banda, J. M. 2020. Social Media Mining Toolkit (SMMT). *Genomics Inform.* 18(2): e16.

Vidgen, B.; Hale, S.; Guest, E.; Margetts, H.; Broniatowski, D.; Waseem, Z.; Botelho, A.; Hall, M.; and Tromble, R. 2020. Detecting East Asian Prejudice on Social Media. In *Proc. of the Fourth Workshop on Online Abuse and Harms*, 162–172. Online: ACL.

Wikipedia contributors. 2022. List of ethnic slurs. shorturl.at/ioFSY. Accessed: 4-4-2022.

Yam, K. 2021. One-third of Asian Americans fear threats, physical attacks and most say violence against them is rising. shorturl.at/akzDN. Accessed: 2021-9-2.

Yang, D.; and Counts, S. 2018. Understanding Self-Narration of Personally Experienced Racism on Reddit. *ICWSM* 12(1).

Zhang, J. S.; Keegan, B. C.; Lv, Q.; and Tan, C. 2020. A Tale of Two Communities: Characterizing Reddit Response to COVID-19 through /r/China_Flu and /r/Coronavirus .