Sense and Sensibility: Characterizing Social Media Users Regarding the Use of Controversial Terms for COVID-19

Hanjia Lyu, Long Chen, Yu Wang, and Jiebo Luo, *Fellow, IEEE*

Abstract—With the world-wide development of 2019 novel coronavirus, although WHO has officially announced the disease as COVID-19, one controversial term - "Chinese Virus" is still being used by a great number of people. In the meantime, global online media coverage about COVID-19-related racial attacks increases steadily, most of which are anti-Chinese or anti-Asian. As this pandemic becomes increasingly severe, more people start to talk about it on social media platforms such as Twitter. When they refer to COVID-19, there are mainly two ways: using controversial terms like "Chinese Virus" or "Wuhan Virus", or using non-controversial terms like "Coronavirus". In this article, we attempt to characterize the Twitter users who use controversial terms and those who use non-controversial terms. We use the Tweepy API to retrieve 17 million related tweets and the information of their authors. We find the significant differences between these two groups of Twitter users across their demographics, user-level features like the number of followers, political following status, as well as their geo-locations. Moreover, we apply classification models to predict Twitter users who are more likely to use controversial terms. To our best knowledge, this is the first large-scale social media-based study to characterize users with respect to their usage of controversial terms during a major crisis.

Index Terms—Classification, controversial term, COVID-19, social media, Twitter, user characterization

1 Introduction

The COVID-19 viral disease was officially declared a pandemic by the World Health Organization (WHO) on March 11. On April 13, WHO reported that 213 countries, areas and territories were impacted by the virus, totaling 1,773,084 confirmed cases and 111,652 confirmed death worldwide.[1] This disease has undeniably impacted the daily operations of the society. McKibbin and Fernando provided the estimated overall GDP loss caused by COVID-19 in seven scenarios, with the estimated loss range between 283 billion USD to 9,170 billion USD [1]. However, the economy is not the only aspect impacted by COVID-19. When COVID-19 was first spreading in the mainland of China, Lin found that a mutual discrimination was developed within the Asian societies [2]. With the world-wide development of COVID-19, the global online media coverage of the term "Chinese Flu" took off around March 18.[2] Fig. 1 shows the timeline of the density of global online media coverage using the term "Chinese Flu" and the global online media coverage of COVID-19-related racial attacks. Around February 2, there was a peak of the online media coverage of COVID-19-related racial attacks following the peak of online media coverage using "Chinese Flu". Zheng *et al*. found that some media coverage of COVID-19 has a negative impact on Chinese travellers' mental health by labeling the outbreak as "Chinese virus pandemonium" [3]. The online media coverage of COVID-19-related racial attacks is still increasing steadily as of April 2020.[3] Not only the online media coverage but also the usage of the term "Chinese Virus" or "Chinese Flu" is trending on social media platforms such as Twitter. On March 16, even the president of United States, Donald Trump posted a tweet calling COVID-19 "Chinese Virus".[4] Despite the defense by President Donald Trump that calling coronavirus the "Chinese Virus" is not racist,[5] racism and discrimination against Asian-Americans has surged in the US.[6]

Matamoros-Fernandez proposed the concept "platformed racism" as a new form of racism derived from the culture of social media platforms in 2017 and argued that it evoked platforms as amplifiers and manufacturers of racist discourse [4]. It is crucial for governments, social media platforms and individuals to understand such phenomena during this pandemic

1. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/
2. https://blog.gdeltproject.org/is-it-coronavirus-or-covid-19-or-chinese-flu-the-naming-of-a-pandemic/
3. https://blog.gdeltproject.org/online-media-coverage-of-covid-19-related-racial-attacks-increasing-steadily/
4. https://twitter.com/realdonaldtrump/status/1239685852093169664
5. https://www.cnbc.com/2020/03/18/coronavirus-criticism-trump-defends-saying-chinese-virus.html
6. https://www.washingtonpost.com/technology/2020/04/08/coronavirus-spreads-so-does-online-racism-targeting-asians-new-research-shows/

- *H. Lyu is with the Goergen Institute for Data Science, University of Rochester, Rochester, NY 14627, USA. E-mail: hlyu5@ur.rochester.edu.*
- *L. Chen and J. Luo are with the Department of Computer Science, University of Rochester, Rochester, NY 14627, USA. E-mail: lchen62@u.rochester.edu, jluo@cs.rochester.edu.*
- *Y. Wang is with the Department of Political Science, University of Rochester, Rochester, NY 14627, USA. E-mail: ywang176@ur.rochester.edu.*

Manuscript received 17 Apr. 2020; revised 12 May 2020; accepted 12 May 2020. Date of publication 21 May 2020; date of current version 3 Dec. 2021. (Corresponding author: Hanjia Lyu.)
Digital Object Identifier no. 10.1109/TBDATA.2020.2996401

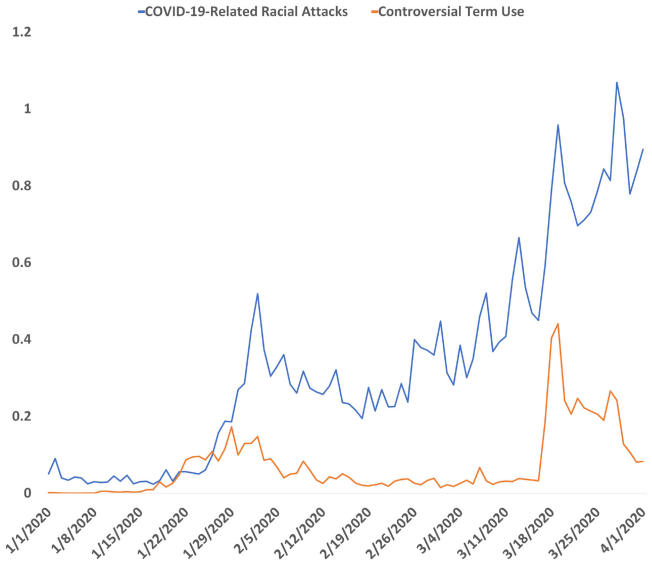


Fig. 1. Density of online media coverage with the controversial term and COVID-19 related racial attacks.

or similar crisis. In addition, the uses of controversial terms associated with COVID-19 could be associated with hate speeches as they are likely xenophobic [30]. Hate speeches reflect the expressions of conflicts within and across societies or nationalities [31]. On online social media platforms, hate speeches can spread extremely fast, even cross-platform, and can stay online for a rather long time [31], [32]. They are also itinerant, meaning that despite forcefully removed by the platforms, they can find expression elsewhere on the Internet and even offline [31]. The impact of hate speeches is not to be under-estimated in the sense of both developing racism in the population and damaging the relationship between societies and nations. Therefore, it is crucial to detect and contain the spreading of hate speeches in the early stage to prevent exponential growth of such a mindset among Internet users.

In this study, we attempt to characterize Twitter users who use controversial terms associated with COVID-19 (e.g., "Chinese Virus") and those who use non-controversial terms (e.g., "Corona Virus") by analyzing the population bias across demographics, user-level features, political following statuses, and geo-locations. Such findings and insights can be vital for policy-making on the global fight against COVID-19. In addition, we attempt to train classification models for predicting the usage of controversial terms associated with COVID-19, with features crawled and generated from Twitter data. The models can be used in social media platforms to monitor discriminating posts and prevent them from evolving into serious racism-charged hate speeches.

2 Related Work

Research has been conducted to analyze social media users at a general level [9]. Mislove *et al.* studied the demographics of general Twitter users including their geo-location, gender and race [5]. Sloan *et al.* derived the characteristics of age, occupation and social class from Twitter user meta-data [6]. Corbett *et al.* and Chang *et al.* also tried to study the demographics of general Facebook users [7], [8]. Pennacchiotti and Popescu used a machine learning approach to classify Twitter users according to their user profile, tweeting behavior, content of tweets and network structure [10]. With the connection between social media and people's lives getting increasingly closer, studies have been conducted at a more specific level. Gong *et al.* studied the silent users in social media communities and found that user generated content can be used for profiling silent users [11]. Paul *et al.* were the first to utilize network analysis to characterize the Twitter verified user network [12]. Users consciously deleting tweets were observed by Bhattacharya *et al.* [28]. Cavazos-Rehg *et al.* studied the followers and tweets of a marijuana-focused Twitter handle [29]. Moreover, attention has also been paid to users with specific activities. Ribeiro *et al.* focused on detecting and characterizing hateful users in terms of their activity patterns, word usage and the network structure [13]. Olteanu, Weber and Gatica-Perez studied the #BlackLivesMatter movement and hashtag on Twitter to quantify the population biases across user types and dempographics [14]. Badawy *et al.* analyzed the digital traces of political manipulation related to 2016 Russian interference in terms of Twitter users' geo-location and their political ideology [15]. Wang *et al.* compared the Twitter followers of the major US presidential candidates [16], [17], [18] and further inferred the topic preferences of the followers [19]. In addition to analyzing individual users, many studies were conducted on communities in social media platforms [20], [21], [22], user behavior [23], [24], [25], and the content that users publish [26], [27], [28], [29].

In our research, we focus on demographics, user-level features, political following status, and the geo-locations of the twitter users using controversial terms and the users using non-controversial terms. To the best of our knowledge, this is the first large-scale social media-based study to characterize users with respect to their usage of controversial terms during a major crisis.

3 Data Collection and Preprocessing

The related tweets (Twitter posts) were collected using the Tweepy API. We initialized an experimental round to collect posters for 24 hours for both controversial and non-controversial keywords, with chinese virus, china virus and wuhan virus as controversial keywords, and corona, covid19 and #Corona as non-controversial keywords. We then selected the most frequent keywords and hashtags in the controversial dataset and the non-controversial dataset for a refined keyword list. As a result, the controversial keywords consist of chinese virus and #ChineseVirus and non-controversial keywords included corona, covid-19, covid19, coronavirus, #Corona, #Covid_19 and #coronavirus.[7] The keywords were then used to crawl tweets to construct a dataset of controversial tweets (CD) and a dataset of non-controversial ones (ND) simultaneously for a four-day period from March 23-26, 2020. Some users post tweets containing the controversial terms to express their disagreement with this usage. We sampled 200 tweets from the dataset and manually labeled the tweets that expressed the disagreement with the usage of controversial terms. 15.5 percent of the tweets actually show the disagreement. Table 1 shows examples of the tweets that express

7. In Tweepy query, capitalization of non-hashtag keywords does not matter.

TABLE 1
Examples Found via Keyword Search for the Use of Controversial Terms

Tweets that Express Disagreement
Then stop calling it the "Chinese" virus!! You knew about the Coronavirus for a very long time, but did nothing and say "it snuck up on us"...
RT @USER: Maybe stop calling it the Chinese Virus then .....
@USER @USER Please stop calling it the Chinese virus.
Tweets that Do not Express Disagreement
@USER It is the Chinese virus! They are 100% at fault!
IT'S A CHINESE VIRUS...!!! Pass It On
Zero case for Chinese Virus? Anyone still believe in #HASHTAG China with their blank record?? #HASHTAG

disagreement and the tweets that do not. In the end, 1,125,285 tweets were collected for CD and 16,320,176 for ND. We created four pairs of CD-ND datasets as follows.

3.1 Baseline Datasets

We built the baseline datasets with basic attributes that can be used in all subsequent subsets. In the Baseline Datasets, 7 user-level features were either collected or computed, including followers_count, friends_count, statuses_count, favorites_count, listed_count, account_length (the number of months since the account was created) and verified status (verified users are the accounts that are independently authenticated by the platform[8] and are considered influential). Next, entries with missing values were removed. In the end, 7 features were computed for 1,125,176 tweets in CD and 1,599,013 tweets in ND. CD and ND were quite balanced with random sampling for convenience in classification process. In addition, since our analysis were performed with proportion tests, the randomly sampling process still maintained representation for the entire dataset.

Since our paper focuses on user-level features, and most of which remained unchanged for a user throughout the 4-day data collection period, we removed duplicate users in CD and ND, respectively, to reduce duplicate entries in our datasets. However, we did not remove duplicate users that appeared in both CD and ND, as the number of such users were unsubstantial (8.19 percent of the total users) and such "swing users" (users who used both controversial and non-controversial terms) were a potential type of users that we definitely should not exclude. In the end, there are 593,233 distinct users in CD and 490,168 distinct users in ND.

3.2 Demographic Datasets

We intend to investigate user-level features of the Twitter users with the demographic datasets, which were built upon the baseline datasets. Since Twitter does not provide sufficient demographic information in the crawled data, we applied Face++ API[9] to obtain inferred age and gender information by analyzing users' profile images. Profile images with multiple faces were excluded. We also found that some profile images were not real-person images and many URLs were invalid. Such data were considered noise and removed from the demographic datasets. Table 2 shows the counts of images with one intelligible face, multiple faces, zero intelligible face and invalid URL. Images with only one intelligible face were retained to form the demographic datasets.

TABLE 2
Composition of Profile Images

	Controversial	Non-Controversial
One Intelligible Face	47,011	109,718
Multiple Faces	5,596	11,393
Zero Intelligible Face	54,539	96,218
Invalid URL	264,894	187,379
Total	372,040	404,708

Politically related attributes by tagging users were also added if they follow the Twitter accounts of top political leaders who are or were pursuing nomination for the 2020 presidential general election. Five Democratic presidential candidates (Joe Biden, Michael Bloomberg, Bernie Sanders, Elizabeth Warren and Pete Buttigieg) and the incumbent Republican President (Donald Trump) were included in the analysis. We crawled Twitter followers' IDs for the political figures to determine the respective political following statuses.[10]

In the end, the Demographic Datasets consist of 15 features (7 features from the Baseline Datasets and the 8 aforementioned new features), with 47,011 distinct users in CD and 109,718 distinct users in ND.

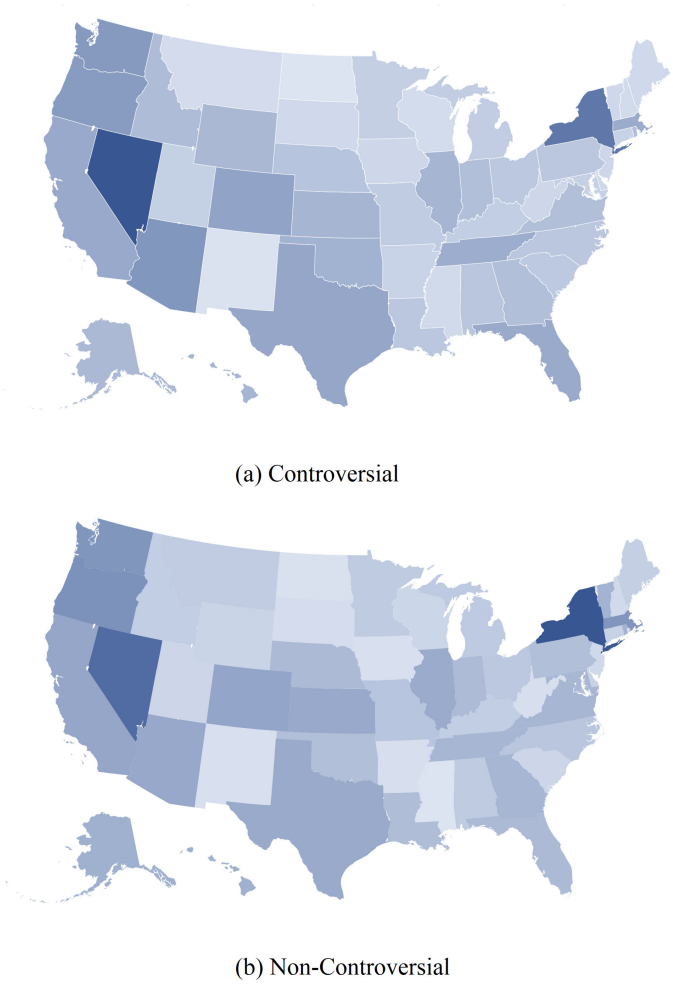
3.3 Geo-Location Datasets

We also intend to investigate the potential impact of the type of communities where users live in on their uses of controversial terms associated with COVID-19. Therefore, Geo-location Datasets were built upon the Baseline Datasets, in a similar fashion as the Demographic Datasets.

Observing that only a very limited number of tweets contain self-reported locations (1.2 percent of crawled data), we instead use the user profile location as the source of geo-location, which has a substantially higher percentage of entries in the crawled datasets (16.2 percent of crawled data). We collected posts with user profile location entries and then removed entries that are clearly noise (e.g., "Freedom land", "Moon" and "Mars"), non-US locations and unspecific locations (ones that only report country or state). In the end, there are 14,817 users for CD and 41,118 users for ND in the Geo-location Datasets.

At the state level, we observe little difference in the distribution of tweets between CD and ND, as shown in Fig. 2. Therefore, a more detailed analysis of geo-location information is required. Detailed information about locations was collected with python package uszipcode. Geo-location

8. https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/user-object
9. https://www.faceplusplus.com
10. Due to limitation of Twitter API, only about half of Donald Trump's follower ID was crawled.



(a) Controversial

(b) Non-Controversial

Fig. 2. Distribution of a) controversial and b) non-controversial tweets in the US, by state, and normalized by the state population. No significant differences can be observed. It is interesting to note that New York and Nevada have the highest number of COVID-19 related tweets per capita.

data in the datasets were processed to find the exact location at zip-code level. Based on population density of a zip code area, we then classified locations into urban (3,000+ persons per square mile), suburban (1,000 3,000 persons per square mile) or rural (less than 1,000 persons per square mile).[11]

3.4 The Aggregate Datasets

Datasets with both demographic and geo-location features were created. These datasets contain complete attributes that were analyzed in our study, while trading off with the relatively small size, with 5,772 for CD and 12,403 for ND. These datasets can be used in a classification model to compare feature importance among all attributes.

4 Characterizing Users Using Different Terms

We perform statistical analysis with the generated datasets in order to investigate and compare patterns of features in both CD and ND for demographic, user-level, political and location-related attributes.

11. https://greatdata.com/product/urban-vs-rural

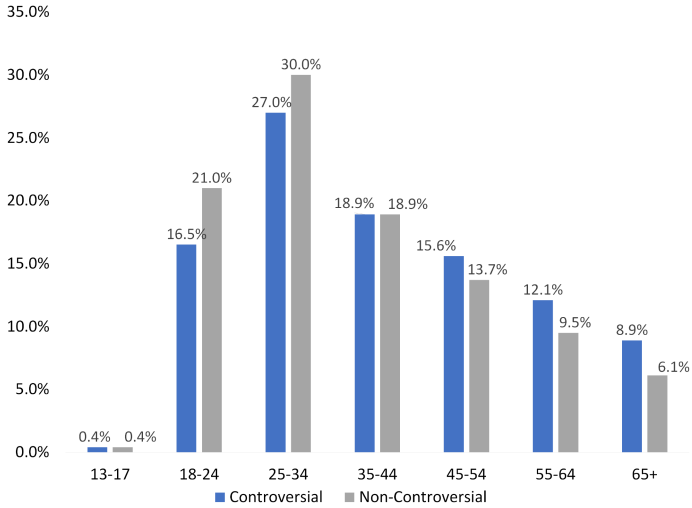


Fig. 3. Age Distribution among users of controversial terms and users of non-controversial terms.

TABLE 3
Gender Distribution

User Type	Controversial	Non-Controversial
Male	61.0%	56.2%
Female	39.0%	43.8%

4.1 Demographic Analysis

4.1.1 Age

Fig. 3 is the distribution of age. We separate the age range into seven bins similar to most social media analytic tools. In both CD and ND, the 25-34 bin comprises the biggest part, which is consistent with the age distribution of general Twitter users.[12] After performing the goodness-of-fit test, we find that the age distributions in these two groups are statistically different ($p < 0.0001$). The Twitter users in ND tend to be younger. More than half of the non-controversial group are the people under 35 years old. In ND, there are 21.0 percent users in the 18-24 bin while that proportion is only 16.5 percent in CD. Users that are older than 45 years old are more likely to use controversial terms.

4.1.2 Gender

Table 3 shows the gender distribution of each group. In both groups, there are more male. There are 61.0 percent male users in CD and 56.2 percent male users in ND. As of January 2020, 62 percent Twitter users are male, and 38 percent are female.[13] This observation suggests that the gender distribution of CD is not different from the overall Twitter users. Furthermore, We perform the proportion z-test and there is sufficient evidence to conclude that gender distributions in these two groups are statistically different ($p < 0.0001$). There are relatively more female users in ND which account for 46.1 percent, however there are only 38.0 percent female users in CD.

12. https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/

13. https://www.statista.com/statistics/828092/distribution-of-users-on-twitter-worldwide-gender/

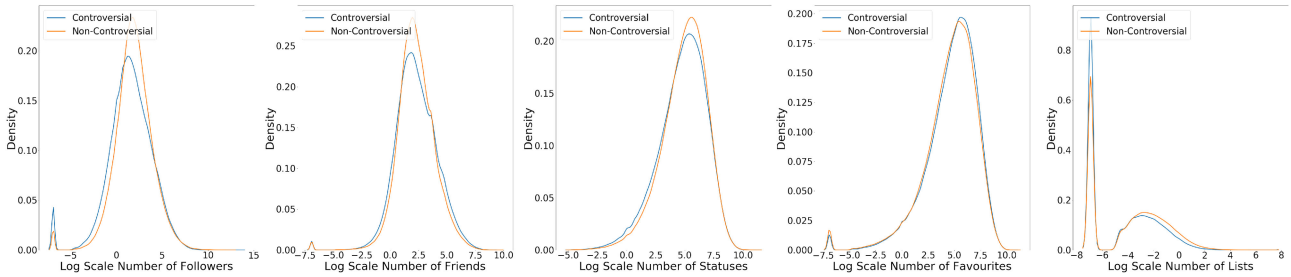


Fig. 4. Density plots (log-scale) of the normalized numbers of followers, friends, statuses, favourites, and listed.

4.2 User-Level Features

In this subsection, we attempt to find insights into 7 user-level features including the number of followers, friends, statuses, favourites, listed, and the number of months since the user account was created, and the verified status. The number of statuses retrieved using Tweepy is the count of the tweets (including retweets) posted by the user. The number of listed is the number of public lists that this user is a member of.

To better analyze the number of followers, friends, statuses, favorites and listed, we normalize them by the number of months since the user's account was created. Given the domain range of these five attributes is large and to better observe the distribution, we first add 0.001 to all these values to avoid zero probability and take the logarithm of them. Fig. 4 shows the density plots (in log scale) of the normalized numbers of followers, friend, statuses, favourites, and listed. Since the distribution is not close to a normal distribution, we apply the Mann-Whitney rank test on these five attributes. There is strong evidence ($p < 0.0001$) in all these five attributes to conclude that the respective medians of these features of CD are not equal to the ones of ND. Table 4 shows the medians of the five normalized features in each group. Users using non-controversial terms tend to have a larger social capital which means they have relatively more followers, friends and post more tweets. This suggests that users in ND normally have a larger size of audience and are more active in posting tweets. One hypothesis for this is that users with a larger audience and more experienced with using Twitter are more cautious when posting in Twitter, which means they pay more attention to the choice of words.

TABLE 4
Medians of Numbers of Followers, Friends, Statuses, Favourites, and Listed

Features	Controversial	Non-Controversial
# of Followers	227	360
# of Friends	413	494
# of Statuses	6,617	9,241
# of Favourites	7,681	7860.5
# of Listed	1	2

TABLE 5
Distribution of Verified Users

User Type	Controversial	Non-Controversial
Verified Users	0.6%	2.0%
Non-Verified Users	99.4%	98.0%

Twitter users are found to be more cautious in publishing and republishing tweets, and also more cautious in sharing among friends [36]. Although the medians of favourites and listed memberships of ND are also higher than those of CD, the differences are not large.

In addition, the users using controversial terms tend to have newer accounts. By applying the Mann-Whitney rank test, the number of months since account was created by users in CD is statistically smaller ($p < 0.0001$). The median number of months of CD is 63 and that of ND is 74. The difference is almost one year, which is significant. This observation is similar to the findings that hateful users have newer accounts [13]. Since using controversial terms does not necessarily correspond to delivering hateful speech, we hypothesize that newer accounts could indicate less experience using Twitter and thus being less cautious when posting tweets.

According to the latest statistics report, there were 0.05 percent verified users overall.[14] Although there are more non-verified users in each group, the proportions of the verified users in CD and ND are both higher than that of the overall Twitter population. After performing a proportion z-test, we find that the proportion of verified users in ND is greater than that in CD ($p < 0.0001$). Table 5 is the distribution of verified users. This indicates that verified users are more likely to use non-controversial terms. Existing work shows the tweets posted by verified users are more likely to be perceived as trustworthy [33] which leads to our hypothesis that verified users are more cautious publishing tweets since their tweets have more credibility partially due to their conscious efforts.

4.3 Political Following Status

To model the political following status, we record if the user is a follower of Joe Biden, Bernie Sanders, Pete Buttigieg, Michael Bloomberg, Elizabeth Warren, or Donald Trump. There are 64 different combinations of following behaviors among all the 1,083,401 Twitter users in our dataset. Fig. 5 is the pie chart of political following status, where we only show the proportions of the combinations greater than 1.0 percent. By performing the goodness-of-fit test, we find statistical difference with respect to the proportions ($p < 0.0001$). In both groups, most of users do not follow any of these people. This group constitutes 70.6 percent in ND and 63.4 percent in CD, respectively. The second biggest group of users in both CD and ND corresponds to users who only follow Donald Trump. The proportion of users in CD only following Donald Trump

14. https://www.digitaltrends.com/social-media/journalists-celebrities-athletes-make-up-most-of-twitters-verified-list/

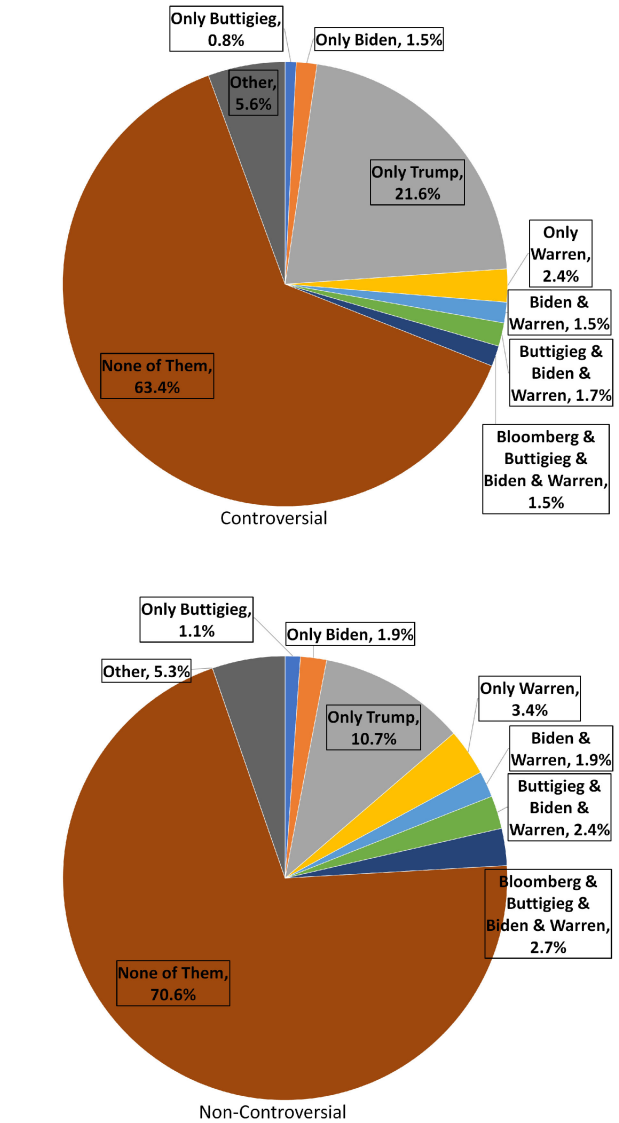


Fig. 5. Proportion of political following status.

is 21.6 percent which is higher than that in ND. This indicates that users only following President Trump are more likely to use controversial terms. This observation resonates with the findings by Gupta, Joshi, and Kumaraguru [20] that during a crisis, the leader represents the topics and opinions of the community. Another noticeable finding is the proportions of users that follow the members of the Democratic Party in ND are all higher than those in CD. This suggests that users following the members of the Democratic Party are more likely to use non-controversial terms.

4.4 Geo-Location

Geo-location is potentially another important factor in the use of controversial terms. As shown in Table 6, the proportion of controversial tweets by users from rural or suburban regions is significantly higher than that of non-controversial tweets ($p < 0.0001$). The result suggests that Twitter users in rural or suburban regions are more likely to use controversial terms when discussing COVID-19, where as Twitter users in urban areas tend to use non-controversial ones in the same discussion. There could be a number of reasons behind this.

TABLE 6
Tweet Percentages in Urban, Suburban and Rural Areas

	Controversial	Non-Controversial
Urban	56.14%	62.48%
Suburban	17.48%	15.58%
Rural	26.38%	21.94%

One explanation is the difference in political views between metropolitan and non-metropolitan regions. Especially after the 2016 presidential election, political polarization in a geographical sense has been increasing dramatically. Scala *et al.* pointed out that in the 2016 election, most urban regions supported Democrats, whereas most rural regions supported Republicans [34]. The study also showed that variations in voting patterns and political attitudes exist along a continuum, meaning that the Democratic-Republican supporter ratio gradually changes from metropolitan to rural areas [34]. With our finding in the previous section that Republican-politician Twitter followers have a higher tendency to use controversial terms, the difference in usage geographically could be explained. However, Scala *et al.* also pointed out that suburban areas were still heavily Democratic in the 2016 election [34]. Thus we cannot fully explain the overwhelming controversial usage in suburban regions from a pure political perspective.

Another explanation is that more urbanized regions are more affected by COVID-19. Rocklöv *et al.* discovered in a recent research that the contract rate of the virus is proportional to population density, resulting in significantly higher basic reproduction number[15] and thus more infections in urban regions than in rural regions [35]. Therefore, higher reporting of infections cases can be found in urban regions, contributing to the higher use of non-controversial terms associated with COVID-19. On the contrary, rural regions have relatively lower percentage of infected population, thus discussions about the nature and the origin of COVID-19 could be more prevalent, and during such discussions, relatively more users chose to use controversial terms. This could partially explain the higher use of controversial terms in suburban regions.

5 Modeling Users of Controversial Terms

One goal of this paper is to predict Twitter users who employ controversial terms in the discussion of COVID-19 using demographic, political and post-level features that we crawled and generated in the previous sections. Therefore, various regression and classification models were applied. In total, six models were deployed: Logistic Regression (Logit), Random Forest (RF), Support Vector Machine (SVM), Stochastic Gradient Descent Classification (SGD) and Multi-layer Perceptron (MLP) from the Python sklearn package, and XGBoost Classification (XGB) from the Python xgboost package. For SVM, a linear kernel was used. For SGD, the Huber loss was used to optimize the model. For MLP, three hidden layers were used with sizes of 150, 100 and 50 nodes respectively, the Rectified Linear Unit (ReLU) was used as activation for all layers, and the Adam Optimizer was used for optimization. The models

15. Basic reproduction number (R0) is the expected number of cases directly generated by one case in a population where all individuals are susceptible to infection.

TABLE 7
Classification Metrics for the Four Pairs of Datasets

Metric	Baseline		
	Logit	XGB	RF
Accuracy	0.625	0.628	0.645
Precision	0.604	0.594	0.606
Recall	0.268	0.315	0.396
F1	0.371	0.411	0.479
AUC-ROC	0.618	0.647	**0.669**

(a) Baseline (N_{CD}=593,233 , N_{ND}=490,168)

Metric	Demographics					
	Logit	XGB	RF	SVM	MLP	SGD
Accuracy	0.584	0.691	0.743	0.510	0.680	0.627
Precision	0.604	0.698	0.735	0.474	0.663	0.665
Recall	0.296	0.589	0.698	0.544	0.631	0.390
F1	0.397	0.639	0.716	0.507	0.647	0.491
AUC-ROC	0.633	0.783	**0.833**	0.699	0.714	0.783

(b) Demographics (N_{CD}=47,011, N_{ND}=109,718)

Metric	Geo-location					
	Logit	XGB	RF	SVM	MLP	SGD
Accuracy	0.611	0.651	0.811	0.543	0.615	0.658
Precision	0.695	0.699	0.757	0.508	0.713	0.702
Recall	0.305	0.436	0.875	0.499	0.504	0.456
F1	0.422	0.537	0.812	0.504	0.410	0.553
AUC-ROC	0.652	0.696	**0.874**	0.696	0.636	0.713

(c) Geo-location (N_{CD}=14,817, N_{ND}=41,118)

Metric	Demographics + Geo					
	Logit	XGB	RF	SVM	MLP	SGD
Accuracy	0.628	0.592	0.644	0.591	0.685	0.658
Precision	0.425	0.402	0.439	0.328	0.495	0.476
Recall	0.524	0.622	0.497	0.293	0.264	0.241
F1	0.469	0.489	0.467	0.310	0.344	0.319
AUC-ROC	**0.624**	0.545	0.604	0.510	0.571	0.560

(d) Aggregate (N_{CD}=5,772, N_{ND}=12,403)

Best AUC-ROC scores are highlighted in each dataset pair.

were applied to three datasets as described below. All datasets were split into training and testing sets with a ratio of 90:10. Since hyperparameter tuning was done in an empirical fashion, no development set was used. The data split was unchanged for models in each experiment. In addition, non-binary data were normalized using min-max normalization.

5.1 Classification on Baseline Datasets

7 attributes were collected in the baseline datasets for CD and ND. For classification, we removed the favorites count and listed count since they were found to be trivial in our preliminary analysis. In the end, 5 features (follower count, friend count, status count, account length and verified status) were included in the training dataset.

5.2 Classification on Demographic Datasets

With the demographic datasets, user-level (political following, age and gender) features were included. We converted age into dummy variables, with cutoffs as reported in Section 4. Gender was also converted into dummy variables *female* and *male*. In the end, 18 features were prepared as inputs for the machine learning models. In total, 21 variables were created.

5.3 Classification on Geo-Location Datasets

In addition to the 5 features in the baseline datasets, geolocation classification (urban, suburban and rural) was included in geo-location classification, resulting in 8 features in the datasets for geo-location classification.

5.4 Classification on Aggregate Datasets

For the aggregate datasets, all of the 24 aforementioned features in the previous three datasets were included. We aimed to see comparisons of importance and impact of features on the use of controversial terms.

5.5 Results and Evaluations

The classification metrics of the datasets are shown in Table 7. In general, Random Forest classifiers resulted in the best results in terms of the AUC-ROC measures. The best AUC-ROC score achieved in baseline dataset classification is 0.669, that in demographic dataset classification is 0.833, that in geo-location dataset classification is 0.874 and that in aggregate dataset classification is 0.624. The AUC-ROC scores in demographic and geo-location classification are rather high, showing strong signals in demographic and geo-location related features. The AUC-ROC score in the baseline datasets is understandably low since only 5 features were included. Model performance in the aggregate datasets is also relatively low. A possible explanation is the small training data size compared to the demographic and geo-location datasets.

The impact of features on predicting the use of controversial terms can be observed using logistic regression coefficients, as shown in Fig. 6. All coefficients were tested with 5 percent significance level. The coefficient of *suburban* is insignificant (P = 0.092). This could be due to the transitional property of this kind of regions between rural and urban regions, representing a mixed population present in suburbs and thus making prediction less reliable. The coefficient of other features are significant.

The most significant contribution to predicting the use of controversial terms is Trump_Follower, while the followings of democratic party politicians contributes negatively. Male users tend to use controversial terms, while female users tend to use non-controversial terms indicated by the higher absolute coefficient. Urban users tend to use non-controversial terms, which suburban and rural users have higher uses of controversial ones. Verified users tend to use non-controversial terms. Accounts with longer history tend to use non-controversial terms. Users under 45 years old tend to use non-controversial terms, especially for those between 18 to 35, while users older than 45 years old tend to use controversial terms. In addition, user-level attributes (following count, friends count and statuses count) have very little impact in the model. These findings are in principle comparable to the analysis in previous sections.

6 Conclusion and Future Work

We have analyzed 1,083,401 Twitter users in terms of their usage of terms referring to COVID-19. We find significant differences between the users using controversial terms and users

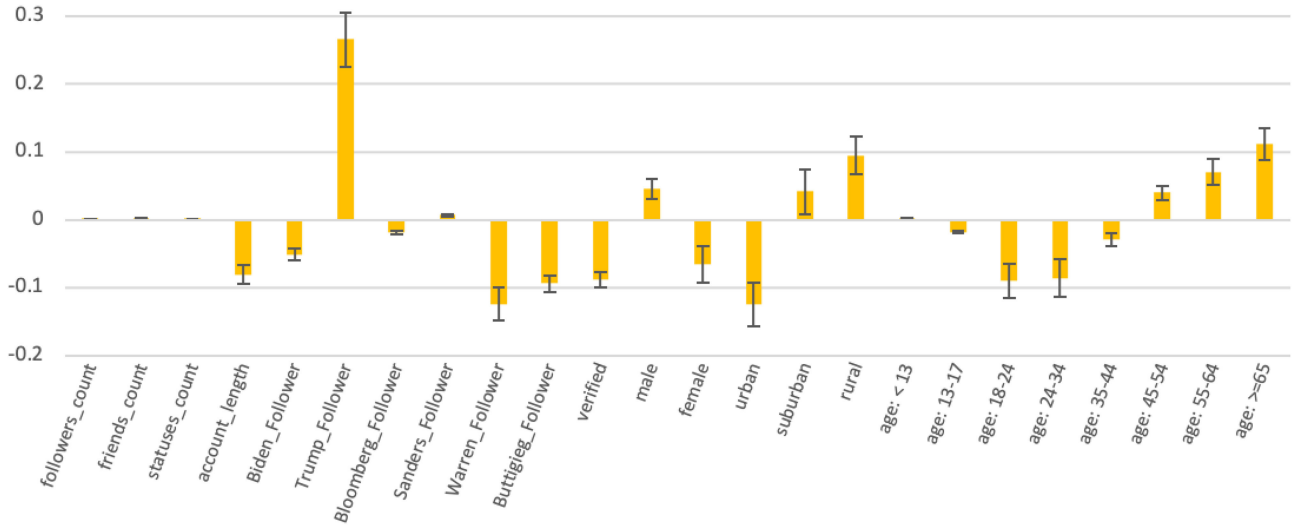


Fig. 6. Logistic regression coefficients for the aggregate dataset. Users who follows Trump, users in rural and suburban regions, male users and users aged over 45 years old tend to use controversial terms, whereas users who follow most top Democratic presidential candidates, users in urban regions, female users and users aged between 18 to 44 tend to use non-controversial terms. No significant association between account-level attributes (follower, friend and status count) and the use of controversial terms were found. All coefficients were tested with a 5 percent significance level. The coefficient of $suburban$ is insignificant ($P = 0.092$). The coefficients of other features are significant.

using non-controversial terms across demographics, user-level features, political following status, and geo-location. Young people tend to use non-controversial terms to refer to COVID-19. Although in both CD and ND groups, male users constitute a higher proportion, the proportion of female users in the ND group is higher than that in the CD group. In terms of user-level features, users in the ND group have a larger social capital which means they have more followers, friends, and statuses. In addition, the proportion of non-verified users is higher in both CD and ND groups, while the proportion of verified users in the ND group is higher than that of the CD group. There are more users following Donald Trump in the CD group than in the ND groups. The proportion of users in the ND group following the members of the Democratic Party is higher. There is no sufficient evidence to conclude that there is difference in terms of which state the user lives in. However, we find users living in rural or suburban areas are more likely to use the controversial terms than users living in urban areas. Furthermore, we apply several classification algorithms to predict the users with a higher probability of using the controversial terms. Random Forest produces the highest AUC-ROC score of 0.874 in geo-location classification and 0.833 in demographics classification.

Since high accuracy is achieved in both demographics and geo-location classification, future studies may collect larger datasets for aggregate classification for better model performance. In addition, this research mainly focuses on user-level attributes in understanding the characteristics of users who use controversial terms associated with COVID-19. Analysis from other perspectives, such as textual mining for the Twitter posts associated with controversial term uses, can be performed to gain more insights of the use of controversial terms and better understanding of those who use them on social media.

References

[1] W. J. McKibbin and R. Fernando, "The global macroeconomic impacts of COVID-19: Seven scenarios," CAMA Working Paper No. 19/2020, 2020. [Online]. AVailable: https://ssrn.com/abstract=3547729

[2] C.-Y. Lin *et al.*, "Social reaction toward the 2019 novel coronavirus (COVID-19)," *Soc. Health Behav.* vol. 3, no. 1, 2020, Art. no. 1.

[3] Y. Zheng, E. Goh, and J. Wen, "The effects of misleading media reports about COVID-19 on Chinese ourists-mental health: A perspective article," *Int. J. Tourism Hospitality Res.*, vol. 31, pp. 337–340, 2020.

[4] A. Matamoros-Fern ández, "Platformed racism: The mediation and circulation of an Australian racebased controversy on Twitter, Facebook and YouTube," *Inf. Commun. Soc.*, vol. 20, no. 6, pp. 930–946, 2017.

[5] A. Mislove *et al.*, "Understanding the demographics of Twitter users," in *Proc. 5th Int. AAAI Conf. Weblogs Social Media*, 2011.

[6] L. Sloan *et al.*, "Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user metadata," *PloS One*, vol. 10, no. 3, 2015, Art. no. e0115545.

[7] Founder, "Facebook demographics and statistics report 2010–145% growth in 1 Year," [Online]. Available: https://isl.co/2010/01/facebook-demographics-and-statistics-report-2010–145-growth-in-1-year/

[8] J. Chang *et al.*, "ePluribus: Ethnicity on social networks," in *Proc. 4th Int. AAAI Conf. Weblogs Soc. Media*, 2010, pp. 18-25.

[9] X. Jin, A. Gallagher, J. Han, and J. Luo, "Wisdom of social multimedia: Using flickr for prediction and forecast," in *Proc. 18th ACM Int. Conf. Multimedia*, 2010, pp. 1235–1244.

[10] M. Pennacchiotti and A.-M. Popescu, "A machine learning approach to Twitter user classification," in *Proc. 5th Int. AAAI Conf. Weblogs Soc. Media*, 2011, pp. 281–288.

[11] W. Gong, E.-P. Lim, and F. Zhu, "Characterizing silent users in social media communities," in *Proc. 9th Int. AAAI Conf. Web Soc. Media*, 2015.

[12] I. Paul, A. Khattar, P. Kumaraguru, M. Gupta, and S. Chopra, "Elites tweet? Characterizing the Twitter verified user network," in *Proc. IEEE 35th Int. Conf. Data Eng. Workshops*, 2019, pp. 278–285.

[13] M. H. Ribeiro *et al.*, "Characterizing and detecting hateful users on Twitter," in *Proc. 12th Int. AAAI Conf. Web Soc. Media*, 2018, pp. 676–679.

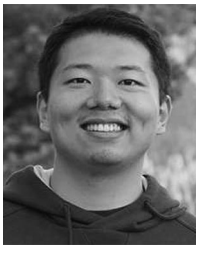
[14] A. Olteanu, I. Weber, and D. Gatica-Perez, "Characterizing the demographics behind the#blacklivesmatter movement," in *Proc. AAAI Spring Symp. Series*, 2016, pp. 310–313.

[15] A. Badawy, E. Ferrara, and K. Lerman, "Analyzing the digital traces of political manipulation: The 2016 Russian interference Twitter campaign," in *Proc. IEEE/ACM Int. Conf. Advances Soc. Netw. Anal. Mining*, 2018, pp. 258–265.

[16] Y. Wang, Y. Li, and J. Luo, "Deciphering the 2016 U.S. presidential campaign in the Twitter sphere: A comparison of the trumpists and clintonists," in *Proc. Int. Conf. Weblogs Soc. Media*, 2016, pp. 723–726.

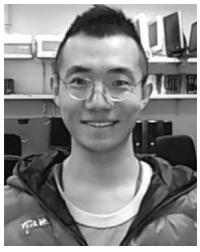
[17] Y. Wang, Y. Feng, and J. Luo, "How polarized have we become? A multimodal classification of trump followers and clinton followers," in *Proc. Int. Conf. Soc. Informat.*, 2017, pp. 440–456.

[18] Y. Wang and J. Luo, "Gender politics in the 2016 presidential election: A computer vision approach," in *Proc. Int. Conf. Soc. Comput. Behavioral-Cultural Model. Prediction Behav. Representation Model. Simul.*, 2017, pp. 35–45.
[19] Y. Wang, J. Luo, R. G. Niemi, Y. Li, and T. Hu, "Catching fire via 'Likes': Inferring topic preferences of trump followers on Twitter," in *Proc. Int. Conf. Weblogs Soc. Media*, 2016, pp. 719–722.
[20] A. Gupta, A. Joshi, and P. Kumaraguru, "Identifying and characterizing user communities on Twitter during crisis events," in *Proc. Workshop Data-Driven User Behavioral Modelling Mining From Soc. Media*. 2012, pp. 23–26.
[21] A. A. Díaz-Faes, T. D. Bowman, and R. Costas, "Towards a second generation of 'social media metrics': Characterizing Twitter communities of attention around science," *PloS One*, vol. 14, no. 5, 2019.
[22] T. Wang *et al.*, "Detecting and characterizing eating disorder communities on social media," in *Proc. 10th ACM Int. Conf. Web Search Data Mining*, 2017, pp. 91–100.
[23] R. Murimi, "Online social networks for meaningful social reform," in *Proc. World Eng. Educ. Forum-Global Eng. Deans Council*, 2018, pp. 1–6.
[24] S. Palmer, "Characterizing university library use of social media: A case study of Twitter and Facebook from Australia," *The J. Academic Librarianship*, vol. 40, no. 6, pp. 611–619, 2014.
[25] E. Omodei, M. De De Domenico, and A. Arenas, "Characterizing interactions in online social networks during exceptional events," *Front. Phys.*, vol. 3, 2015, Art. no. 59.
[26] J.-P. Allem *et al.*, "Characterizing JUUL-related posts on Twitter," *Drug Alcohol Dependence*, vol. 190, pp. 1–5, 2018.
[27] K. Lee, S. Webb, and H. Ge, "Characterizing and automatically detecting crowdturfing in Fiverr and Twitter," *Soc. Netw. Anal. Mining* vol. 5, no. 1, 2015, Art. no. 2.
[28] P. Bhattacharya and N. Ganguly, "Characterizing deleted tweets and their authors," in *Proc. 10th Int. AAAI Conf. Web Soc. Media*, 2016.
[29] P. Cavazos-Rehg *et al.*, "Characterizing the followers and tweets of a marijuana-focused Twitter handle," *J. Med. Internet Res.*, vol. 16, no. 6, 2014, Art. no. e157.
[30] Z. Waseem and D. Hovy, "Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter," in *Proc. NAACL Student Res. Workshop*, 2016, pp. 88–93.
[31] I. Gagliardone *et al.*, *Countering Online Hate Speech*. Paris, France: Unesco Publishing, 2015.
[32] R. Magu, K. Joshi and J. Luo, "Decoding the hate code on social media," in *Proc. 11th Int. Conf. Web Soc. Media*, 2017.
[33] B. Z. Erdogan, "Celebrity endorsement: A literature review," *J. Marketing Manage.*, vol. 15, no. 4, pp. 291–314, 1999.
[34] D. J. Scala and K. M. Johnson, "Political polarization along the rural-urban continuum? The geography of the presidential vote, 2000–2016," *The Ann. Amer. Acad. Political Soc. Sci.*, vol. 672, no. 1, pp. 162–184, 2017.
[35] J. Rocklöv and H. Sjödin, "High population densities catalyze the spread of COVID-19," *J. Travel Med.*, vol. 27, 2020, Art. no. taaa038.
[36] C. C. Marshall and F. M. Shipman, "Social media ownership: Using Twitter as a window onto current attitudes and beliefs," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2011, pp. 1081–1090.



Hanjia Lyu received the BA degree in economics from Fudan University, China, in 2017. He is currently working toward the MS degree in data science at the University of Rochester, Rochester, New York. His research interests include statistical machine learning, and data mining.

Long Chen received the BS degree in computer science and the BA degree in economics from the University of Rochester, Rochester, New York, in 2020. He is currently working toward the MS degree with the Center of Data Science at New York University, New York. His research interests include data mining on social media, natural language processing and applied machine learning.



Yu Wang received the PhD degree in political science and computer science from the University of Rochester, Rochester, New York. His first-authored and single-authored papers have appeared in *EMNLP, Political Analysis, Chinese Journal of International Politics, ICWSM*, and International Conference on Big Data. He is interested in international relations with a regional focus on China.

Jiebo Luo (Fellow, IEEE) is currently a professor of computer science at the University of Rochester, Rochester, New York which he joined, in 2011 after a prolific career of fifteen years at Kodak Research Laboratories. He has authored more than 400 technical papers and holds more than 90 US patents. His research interests include computer vision, NLP, machine learning, data mining, computational social science, and digital health. He has been involved in numerous technical conferences, including serving as a program co-chair of ACM Multimedia 2010, IEEE CVPR 2012, ACM ICMR 2016, and IEEE ICIP 2017, as well as a general co-chair of ACM Multimedia 2018. He has served as the editorial boards of the *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, *IEEE Transactions on Multimedia (TMM)*, *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, *IEEE Transactions on Big Data (TBD)*, *ACM Transactions on Intelligent Systems and Technology (TIST)*, *Pattern Recognition*, *Knowledge and Information Systems (KAIS)*, *Machine Vision and Applications*, and *Journal of Electronic Imaging*. He is the current editor-in-chief of the *IEEE Transactions on Multimedia*. He is also a fellow of ACM, AAAI, SPIE, and IAPR.