Applied Soft Computing Journal 97 (2020) 106743

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

Predicting political sentiments of voters from Twitter in multi-party contexts

Aparup Khatua [a,1], Apalak Khatua [b], Erik Cambria [c,*]

[a] L3S Research Center, Leibniz Universität Hannover, Hannover, Germany
[b] XLRI – Xavier School of Management, Jamshedpur, India
[c] School of Computer Science and Engineering, Nanyang Technological University, Singapore

ARTICLE INFO

Article history:
Received 9 April 2020
Received in revised form 28 August 2020
Accepted 16 September 2020
Available online 25 September 2020

Keywords:
Sentiment analysis
Political leaning
Multi-party context
Mix tweeting pattern

ABSTRACT

Prior Twitter-based electoral research has mostly ignored multi-party contexts and 'mix tweets' that jointly mention more than one party. Hence, we investigate the complex nature of these mix tweets in a multi-party context, and we argue mix tweeting patterns of users implicitly capture their political opinions. We predict the political leaning of users based on their mix tweeting patterns in the context of the 2014 Indian General Election. We have agglomerated 2.4 million tweets from 0.15 million unique users. Next, we employ a multinomial logit regression model to test the hypothesized causal relation between mix tweeting patterns and the political leaning of users. Additionally, we also employ neural network-based algorithms to predict political leaning. Our study demonstrates that user-level mix-tweeting patterns can reveal the political opinions of Twitter users.

1. Introduction

Twitter, a microblogging platform, allows users to post short text messages known as tweets. Thus, millions of users express and share their sentiments and opinions on these social media platforms [1–6]. Extant literature has employed sentiment-mining techniques of this opinionated text in the context of product reviews or brand perceptions [6,7]. However, opinion mining can be also employed to other wider topics. Thus, existing literature has explored this opinionated text data for analyzing various socio-economic events, and electoral study is a dominant theme in this domain. Nowadays, the Twitter platform allows voters to get engaged in political discourse by contesting or supporting a political statement [8]. This trend has allowed information science researchers to analyze the linguistic content of tweets to understand political sentiments. The pioneering work, on election forecasting by using Twitter data, demonstrated that the 'mere number of messages mentioning a party reflects the election result' [9]. A series of follow-up papers attempted to predict election results and challenges associated with the same in the context of the USA [10–15], the Netherlands [16,17], Singapore [18], the UK [17,19,20], India [21,22] and Venezuela [23].

In brief, Twitter has emerged as a popular platform for political discourse. Hence, a plethora of studies explored Twitter data, but empirical evidence is mixed at best [8]. Consequently, a study claimed that 'electoral predictions using the published research methods on Twitter data are not better than chance' and listed 'how (not) to predict elections' [24]. Similarly, [25] also argued that the findings of [9] were nothing but 'contingent on arbitrary choices of the authors.' [26] claimed that his attempt to forecast election results ended up with a 'lousy paper.' Twitter-based research is fraught with the pitfalls of self-section biases and over-representation of tech-savvy younger generation electorate on social media platforms [27].

More importantly, a politically active voter can post an innumerable number of tweets, but he will be able to cast *'only one vote.'* Hence, volumetric analysis of twitter trends for electoral forecasting has some inherent limitations. For instance, tweet volume can be a function of 'competitiveness and money spent in the race' [28]. Twitter bots can potentially manipulate the political discourse in an electoral context [29–31]. A few core users can propagate fake and extremely biased news on the Twitter platform [10,11]. A simplistic volumetric analysis mostly ignores these finer nuances. Consequently, volumetric analysis at the aggregate-level might fail to forecast the electoral outcome accurately. To overcome this challenge of multiple posts by a single user, we need to probe the political leaning of individual voters [32]. Existing literature has mostly formulated this problem as user-classification task [33–35].

Interpreting user's political leaning on the Twitter platform is a challenging task [36]. Elementary text mining techniques might

* Corresponding author.
E-mail addresses: khatua@L3S.de (A. Khatua), apalak@xlri.ac.in (A. Khatua), cambria@ntu.edu.sg (E. Cambria).

1 First author is now at L3S Research Center, but did a significant portion of this work earlier at Nanyang Technological University, Singapore.

https://doi.org/10.1016/j.asoc.2020.106743
1568-4946/

be efficient at 'retrieving texts' or counting 'mere number of messages' but have limited capabilities for interpreting the contextual meaning to extract useful information in the political domain [2,3,9,37–39]. Opinions and sentiments on social media platforms are mostly conveyed through latent semantics. Consequently, purely syntactical approaches fail to decipher the contextual meaning of a text [40,41]. Additionally, 'users may choose not to publicly post about their political preference for various social goals' [42].

We find that existing literature has mostly ignored multi-party contexts as well as tweets that jointly mention more than one party. Hence, our study has attempted to address these research gaps by predicting the political leaning of individual users based on their mix tweeting patterns in a multi-party context. Prior studies, mostly in a two-party context, have formulated the user-classification task as a binary classification problem. For instance, in the context of the USA, it is a binary classification between republicans and democrats [33–35]. On the contrary, the political sentiments of users in a multi-party context, such as India, is not a binary classification task [19,22].

We find that social media users frequently mention more than one party within a tweet. Thus, we define a tweet as a mix tweet if it jointly refers more than one party. We demonstrate that mix tweets can reveal the political leaning of users in a multi-party context. Syntactical approaches will fail to extract useful information from these mix tweets, and thus, prior studies mostly ignored them. For instance, two sentences, with resembling lexical behavior, might look similar from the bag-of-words perspective, such as 'iPhone 12 is expensive but nice' and 'iPhone 12 is nice but expensive'. However, in reality, they bear opposite polarity [40,43]. Similarly, in political contexts, a tweet 'Party A is better than Party B' is similar to 'Party B is better than Party A' from the bag-of-words perspective. However, the political leanings of these two sentences are diagonally opposite. The former (latter) tweet conveys positive (negative) sentiments about Party A (Party B) and negative (positive) sentiments about Party B (Party A). We explore these mix tweets and hypothesize the relationship between user-level mix tweeting pattern and their political orientation. On the methodological front, we have employed regression modeling as well as deep learning-based algorithms to decipher the user's political leaning. Our modeling has considered user-level mix tweeting patterns as input variables. We set our study in India. To the best of our knowledge, none of the prior studies has considered mix tweeting pattern of individual users to predict their political sentiments. Considering the user-level mix tweeting pattern as an input variable in a multi-party context is the core contribution of our study.

The rest of the paper is organized as follows: Section 2 provides the literature review of electoral studies and identifies the research gap in the existing literature. Section 3 elucidates the complexities of the Indian political landscape, major political parties, and the nature of Twitter data in a multi-party context. Sections 4 and 5 offer an exhaustive analysis of mix tweets both at the aggregate and at the user level, respectively. Section 6 reports our proposed hypotheses. Sections 7 and 8 discuss our regression-based approach and neural network-based approach, respectively, and subsequently, the findings from these two-pronged approaches. Section 9 suggests a few managerial and social implications of our findings and concludes by identifying a few potential avenues for future research.

2. Electoral studies using social media data: A literature review

In recent times, Twitter has emerged as a dominant platform for political deliberations not only for candidates or political parties but also for ordinary voters to voice their opinions and express their sentiments [8,28]. For instance, during the 2016 USA Presidential Election, both Clinton and Trump have used the Twitter platform 'to express their positions, to attack each other, to retweet endorsements, to encourage people to vote, to give news pre-views, and a lot more' [44]. Twitter-based Electoral research can be classified into two streams, namely election forecasting by using Twitter data and predicting the political orientation of social media users. [45] and [8] have done a systematic literature review of this domain. In this section, we are elucidating the challenges associated with the election forecasting studies, and subsequently, we explore how some of these challenges can be addressed. In brief, we identify the research gap in the existing literature and offer our proposed solution.

Prior electoral studies have explored Twitter data across a wide range of countries. A significant portion of the prior studies is in the context of the UK [17,19,20,46] or the USA [10,11,14,24,32]. In other words, prior studies were mostly in the context of advanced nations — expect a few studies in the context of India [21,22] or Venezuela [23]. Existing literature has employed volumetric analysis, sentiment analysis, and also a combined approach to predict the election results [9,14,21,22,47]. A few studies also probed how political candidates use the Twitter platform during the election. For instance, [17] noted that Dutch politicians embraced the interactive potential of Twitter in comparison to U.K. candidates. Similarly, in the context of 2016 U.S. Presidential election, a study probed the tweeting pattern and frequencies of Clinton vis-à-vis Trump [44]. Another study explained how Trump dominated the unpaid media market [48]. However, except these few exceptions, like [17,44,48], mostly the literature focused on election forecasting by probing the political deliberations on the Twitter platform. As we noted earlier, one of the initial electoral studies in the German context concluded that 'mere number of messages mentioning a party reflects the election result' [9]. Accordingly, a series of initial studies followed this approach and performed volumetric analysis [16,18,22]. In addition to volumetric analysis, a few studies also consider a sentiment-based approach. For instance, a study in the USA context claimed that 'simple sentiment detector based on Twitter data replicates consumer confidence and presidential job approval polls' [14]. Another study in the context of the Irish General election argued that combining volume-based analysis and sentiment analysis by using supervised learning can predict election results [47]. Some of these studies found encouraging results, whereas others were not conclusive [8,45]. For instance, in the Singapore General election, [18] observe that the efficiency of Twitter data for election forecasting is higher at the national level in comparison to the constituency level.

We argue that one of the potential reasons for the inconsistencies in the election forecasting studies can be due to disproportionate participation by users. For example, [11] noted that only 0.1% of social media users accounted for nearly 80% of fake news sources shared during the 2016 U.S. Presidential election. Other studies also confirm the presence of core or influential users who aggressively post rumors or extremely biased news to support their favored candidate during the U.S. Presidential elections [10,13,32]. Similarly, a study in the context of the French Presidential election observed 'profiles that are currently deleted from Twitter were among the most active and central nodes … suggesting the presence of social bots' [31]. Interestingly, [9] noted that 'only 4% of all users accounted for more than 40%' of the social media discussion in the German context. This pattern leads to the puzzling question — whether the 'mere number of messages' was an indicator of 40% or 4% of the users? In other words, the volumetric analysis of party-wise tweets might not be an accurate indicator of electoral forecasting.

We find that this issue of multiple tweets from the same user is prevalent beyond the USA context, such as the Dutch

Senate election [16], German Federal election [9,30], Singapore election [18], and so on. For example, [18] considered 110,815 political tweets from a set of 20,000 unique users. Similarly, [32] considered 250,000 tweets from 20,000 users, that is, roughly 12 tweets per user. This pattern strongly suggests that some users have tweeted multiple times. In this context, [32] cautioned that 'predictions were not to be made by aggregating Twitter data,' but we need to probe the political orientation of these unique users. Otherwise, the accuracy of election forecasting will be 'contingent on arbitrary choices of the authors' [25]. For instance, during the French Presidential election, right-wing voters were more active on the Twitter platform [31]. Hence, to nullify the effects of these core and active users, we need to explore the political opinion of individual voters instead of aggregate-level volumetric analysis [49].

Some of the prior studies have investigated the political orientation of the voters. [33] tried to classify the political orientations of the USA voters by manually building a list of hashtags and applied a stacked-SVM-based binary classifier. Follow-up studies, such as [34] and [35], also emphasized the lexical usage of users. For instance, [34] found that the SVM-based model predicted the political affiliation of the USA voters with 91% accuracy. However, [50] found that Bayesian volume-based predictions outperformed SVM in the context of the U.K. election. In addition to SVM or Naive Bayes classifier [27,33,50], prior studies have also considered boosted decision trees or latent Dirichlet allocation (LDA) [36,51]. Broadly these studies proposed a machine learning framework by using user-centric features for large-scale classification of social media users [35].

Prior studies also probed the social linkages of a user [52] or retweeting pattern of users [53] to predict the political orientation of voters. [52] suggested that considering multiple and heterogeneous types of social links, instead of a single type of link, will be a better predictor of one's political ideology. In other words, network affiliation influences the polarization of voter preference [12]. Similarly, [49] investigated the interaction patterns of users to predict the political preference, and found that this approach is on par with the human annotators in the context of Albertan and Pakistani General Election.

Another set of studies employed a topic-based or theme-based approach. For instance, [54] focused on polarized topics, which 'require the user to side exclusively with one position' and employed the same to identify 'polarized users'. Subsequently, they detected 'polarized keywords' 'by monitoring the activities of previously classified users' in the context of Italian and European Parliament Election. Similarly, [55] analyzed the Twitter activity of 32 U.S. politicians around 2016 U.S. presidential election and employed a weakly supervised method for understanding the orientation of prominent U.S. politicians on issues such as abortion, gun control, immigration, and so on. Extant literature mostly probed the linguistic content of Twitter deliberations. However, an interesting study predicted the political opinion of users by considering non-lexical features, such as 'users' discourse patterns (proportion of Tweets that are retweets or replies) and their rate of use of capitalization and punctuation' [56]. [56] argued that these non-lexical features could reveal the political alignment of users. To sum up, Twitter data allows us to investigate the political orientation of voters, and the accuracy level of predictions from Twitter data is comparable to traditional methods of opinion polling [57].

2.1. Research gap

Our review of Twitter-based electoral studies elucidates some of the inherent limitations of Twitter data in the electoral context. One of the significant limitations is the overwhelming presence of active users. This disproportionate participation by social media users can be addressed by investigating the political opinion of individual users, which is also known as the user-classification task in the literature. Our literature review reveals two research gaps in the existing literature.

First, 'the vast majority of the existing methods on ideology detection on social media have oversimplified the problem as a binary classification problem (i.e., liberal vs. conservative)' [52]. Extant literature on user-classification is mostly in the context of the U.S., where users are predominantly from two opposing camps. These USA-based studies have conceptualized this problem as a binary classification task [33,35]. For example, [33] considered the 'user-property classification tasks as *binary classification problems* and built separate binary classifiers for each attribute' in the USA context. However, it is worth noting that user-classification is not a binary classification task in a multi-party context. There are hardly any electoral studies, except a few like [19,22], which considered the multi-party context. [19] highlighted the limitations of using Twitter in a multi-party context, such as the U.K., in the presence of regional parties. Similarly, [22] also considered the multi-party context in their econometric modeling. However, [19,22] explored political deliberations on the Twitter platform from the perspective of electoral forecasting. To the best of our knowledge, none of the prior studies in political domain probed user-classification in a multi-party context.

Second, existing literature mostly ignores mix tweets. For example, [16] argued that 'the quality of the data collection' can be improved 'by removing … tweets mentioning more than one party name.' In other words, prior studies have mostly considered a simplistic tweet (e.g. the leaders of *Party A* are efficient) and ignored a complex tweet that jointly mentioned more than one party (e.g. the leaders of *Party A* are more efficient than *Party B*). However, a significant portion of political deliberations on the Twitter platform jointly mention more than one party for comparative purpose — especially in a multi-party context.

Our paper attempts to address both the research gaps mentioned above by predicting the political ideology of a user by investigating user-level mix tweeting patterns in a multi-party context. Therefore, our research question is: *Can we predict the political leanings of users based on their mix tweeting patterns?*

We demonstrate that a mix tweeting pattern of an individual user can reveal her political ideology in a multi-party context. A study in the context of the 2015 cricket world has considered the joint mentioning of two cricket teams for user-classification [58]. However, extant literature has not investigated the relationship between mix tweets and political ideology. [9] carried out some elementary analysis of joint mentioning (of two parties) at the aggregate level but not at the user-level. Our corpus revealed that joint mentioning of two, as well as more than two parties, by voters is quite frequent in a multi-party context. Thus, we attempt to predict the political leaning by probing the user-level mix tweeting pattern, and considering mix tweeting patterns as the input variable is the novelty of our study.

On the methodology front, existing literature has mostly used machine learning-based classification algorithms like support vector machine (SVM) or boosted decision trees (BDT), and a few recent studies have employed convolutional neural network for the classification task [23,35,36]. We are applying regression modeling as well as neural network-based classification algorithms for our analysis. Neural network models, such as deep-learning techniques, are efficient for complex tasks, but they are difficult to interpret. On the contrary, regression modeling allows us to understand the causality between input variables and the predicted output. Thus, our two methodologies complement each other. We hardly find any study in the user-classification domain that has employed both regression analysis as well as deep learning-based classification.

3. Research context: India's 2014 general election

Twitter is a popular social media platform in India. Consequently, prior studies considered the Indian context for various socio-economic issues [21,59–61]. Thus, we consider the 2014 Indian General Election for our study. Considering India's vast electorate, the 2014 General Election was conducted in nine phases (from April 7, 2014, to May 12, 2014) for 543 parliamentary constituencies [22]. The political landscape of India displays an intricate relationship between different national and regional parties. Two major alliances, namely, the National Democratic Alliance (NDA) led by the *Bhartiya Janata Party (BJP)* and the United Progressive Alliance (UPA) led by the *India National Congress (INC),* were the frontrunners in this election [22]. Smaller regional parties form electoral alliances with leading national parties to remain politically relevant in the national context [22]. For example, the NDA was an alliance between the *BJP* (which was the main opposition party) and other smaller regional parties. Similarly, the UPA was also an alliance between the *INC* (which was the ruling party in the previous term) and a few other regional parties. The presence of multiple parties in the Indian context provided us with a natural setting to explore our research question. Interestingly, the *Aam Aadmi Party* (AAP) was formed in 2013 by a few social activists who created a significant buzz on social media [22]. However, the strong social media presence of AAP was not an accurate indicator of their electoral performance [22]. Ex-ante, this evidence confirms that the volumetric analysis might be misleading because the social media presence can be a function of an aggressive social media campaign by a few active users [28]. We have also extracted the data for a few other prominent regional parties like the All India Trinamool Congress, Communist Party of India, Samajwadi Party, YSR Congress Party, etc. However, we put all these parties, with a scarce presence on social media, into the 'Other Parties' (OTH) category.

We have also probed the Twitter data of the 2019 General election, but incidentally, the Twitter deliberation was mostly bi-polar in 2019. NDA ensured a landslide victory in the 2019 election. Apart from the two major alliances, namely NDA and UPA, the presence of other regional parties on social media platforms was scant for a large scale analysis. For instance, AAP lost its political relevance as well as its presence on social media platforms. Lack of sufficient mix tweet data did not allow us to consider the 2019 election in our analysis.

3.1. Data description

We extracted our data by using the Twitter search API during the period March 15, 2014, to May 12, 2014 (i.e., polling date for the last phase). The election results were declared only on May 16, 2014. Hence, we did not consider data beyond May 12, 2014. Our initial set of hundred-odd crawling keywords included the names of political parties, names of prominent leaders, and Twitter handles of politically influential users. Some of the prominent keywords from this list were *'Arvind'* (a leading political personality from AAP), *'BJP'*, *'Cong'* (an abbreviation of Congress or INC), *'Manmohan'*(the first name of the previous prime minister), *'Modi'* (the surname of the prime ministerial candidate from BJP), *'Rahul'* (the first name of the prime ministerial candidate from INC), *'NDA'*, *'UPA'* and so on. However, this set of static keywords could not capture various temporal trends during our study period. Thus, we regularly updated our list of crawling keywords by incorporating temporal trends and hashtags during our entire study period. Some of the popular hashtags which emerged during the electoral campaigns were '*#AAPPositive*', '*#MyVoteForCongress*', '*#WeWantModi*', and so on. Additionally, there were another set of hashtags, which were very temporal. For instance, when Narendra Modi visited the Varanasi parliamentary constituency, then the hashtag '*#ModiInVaranasi*' became viral for a few days around his visit. In the data preprocessing stage, we have eliminated duplicate tweets, punctuation marks, non-alphanumeric characters, and URL links for our subsequent analysis.

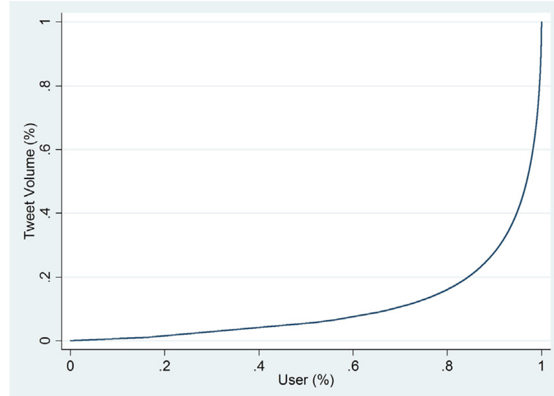

Fig. 1. Disproportionate participation of users. (Y-axis Twitter Volume % and X-axis User %).

Finally, for the analysis purpose, we have considered 2.4 million tweets from 0.15 million unique users. As we pointed out earlier, Twitter deliberations in the political domain are often characterized by disproportionate participation by social media users [9,11,27]. Accordingly, our Indian research context is no different. Fig. 1 reports the user-wise distribution of our data where the X-axis and Y-axis denote the user-wise and volume-wise distribution, respectively. We have noted an exponential nature of participation by social media users. For instance, 80% of users in our corpus accounted for less than 20% of the tweet volume. Interestingly, just the top 5% of users accounted for 58.9% of the total tweets. Our pattern of skewed participation is comparable to prior electoral studies [9,27].

Following the prior election-related studies [9,22,35], we focused on context-specific prototypical words, which are mostly polarized keywords, for party-wise classification of the corpus. Initially, we generated the list of most frequent words in our corpus. Next, we identified and labeled the unambiguous party-specific keywords from that list. In our exhaustive party-wise context-specific corpus, we considered all possible inflectional forms as well as context-specific synonyms of a word. For instance, India National Congress (INC) was the main opposition party, and social media users have used multiple inflectional forms such as Congress, Cong., and INC. We consider all of them in our list of party-specific keywords. A standard lemmatization process or a standard vocabulary will not be able to capture this diverse and complex nature of our data. So, we rely on our proprietary context-specific corpus.

We consider the linguistic content of each tweet based on our corpus of party-related 'prototypical words' [35]. For instance, if a tweet contains only AAP-related keywords (and keywords related to the UPA, NDA, and OTH are absent), then it is classified as an exclusive AAP-related tweet, and similarly for other parties [22]. However, if a tweet contains keywords which are related to more than one party, then it is classified as a MIX tweet. Table 1 elucidates this process by using different color codes. Our classification is a function of the occurrences of party-wise keywords within a tweet where each tweet should be labeled either as an AAP, UPA, NDA, OTH (Others), or MIX tweet (refer to the column 'Mix Tweet: Yes or No?' in Table 1). Subsequently, we are also labeling a MIX tweet based on parties mentioned within the tweet (refer to the column 'Final Category' in Table 1).

Table 1
Illustrative examples of tweet classifications.

Sl. No.	Sample tweets (from multiple users)	Mix tweet: Yes or No?	Final category
1	Pl give one chance to **aap**, Delhi govt powers r limited, but they did enough, Pl think frm ur own mind not frm others mind.be Indian	Not a mix tweet	AAP
2	**BJP** PM candidate **Narendra Modi** will address a massive rally in Wardha, Maharashtra today.	Not a mix tweet	NDA
3	**BJP** releases comprehensive chargesheet against **Congress** led **UPA** government.	YES	NDA-UPA (NU)
4	**BJP**'s much hyped "protest" against **AAP** was a big fiasco! Even their own councilors and MLAs did not attend.	YES	AAP-NDA (AN)
5	**AIADMK, DMK** need support of **Congress** to get things done: P **Chidambaram** - The Economic Times	YES	UPA-OTH (UO)
6	**BSP, AIADMK, BJD, RJD** will win more seats then *(sic)* **AAP** that to without complaining media or using it. #*Intolerant***AAP**	YES	AAP-OTH (AO)
7	**CPI(M)** MANIFESTO APPEALS TO ALL Reject **Congress**, Defeat **BJP**- Vote for **CPI(M)**. Strengthen the Left For a Secular and Democratic Alternative	YES	NDA-UPA-OTH (NAO)
8	While in **AAP** it is a tough job to find a bad candidate, in **BJP** and **Congress** tough to locate good candidates.	YES	AAP-NDA-UPA (ANU)

Note: Relevant party-wise keywords (within a tweet) are color-coded as follows, **AAP-Related Keywords**, **NDA-Related Keywords**, **UPA-Related Keywords** and **Keywords for Other Parties**.

4. Prevalence of mix tweets in Indian multi-party context

Fig. 2 plots the volumetric trend of our data. The vertical axis reports the tweet counts, and it reveals that there are roughly 40,000 tweets/day in our corpus. This figure also reveals that the social media presence of NDA (i.e. the orange area) was significantly higher than the rest of the parties, and twitter deliberation about OTH parties (i.e. red area) was not so significant. This trend justifies our classification of tweet corpus into four political categories, namely AAP, UPA, NDA, and OTH. The gray area reports the presence of MIX tweets in our corpus (refer to Table 1). Interestingly, mix tweets represent a significant portion of our corpus. Hence, Fig. 2 highlights *the prevalence and relevance of mix tweets* in our multi-party research context.

Next, we probe the number of unique users per party per day. In Fig. 3, the dotted black line indicates the total numbers of unique users on a specific day, whereas the area chart represents the party-wise break-up of these unique users. For example, if one user posts, say, n numbers of UPA-related tweets (and nothing else) in a day, then we consider her as 'one unique user' for the UPA Party on that particular day. On the same day, if another user posts one exclusive tweet for the UPA, and one mix tweet (where she compares UPA with NDA), then we have counted her twice (for two different categories – UPA and MIX) as 'unique user.' In other words, this second user would feature two times (in two different categories) in the area chart, but she would feature only once in the dotted black line. Similarly, a supporter of NDA can appreciate NDA (by posting a positive tweet) and criticizes UPA (which will have a negative sentiment) on the same day. This active user is posting two tweets for two different categories (i.e. NDA and UPA) on the same day. Fig. 3 highlights that the total number of unique users (i.e. the dotted black line) is consistently lower than the summation of category-wise unique users (i.e. the outer edge of the area chart). Hence, Figs. 1 and 3 jointly suggest that active users have not only tweeted multiple times but also they have posted different types of tweets according to our proposed classification. Interestingly, mix tweet (i.e. gray area) volume is broadly higher than AAP (i.e. blue area), UPA (i.e. green area) and OTH (i.e. red area). This strongly suggests that a significant portion of our active users are posting mix tweets. Hence, Fig. 3 highlights the importance of probing *mix tweets and the political orientation of users.*

To explore this further, we have extracted structured information of all 0.15 million users by collating all tweets posted by them during our entire study period, and it reveals an intricate pattern. One user can hypothetically post five types of tweets, namely, NDA, AAP, UPA, OTH, and MIX Tweets. Moreover, mix tweets can be of 11 classes (mentioning any 2 parties from the set of 4 parties + mentioning any 3 parties from the set of 4 parties). Refer to Figs. 4b and 4c for a detailed discussion.

Table 2
Frequency of mix tweets which jointly mention two parties.

	UPA	AAP	Other Parties
NDA	11.5%	6.5%	5.9%
UPA	–	2.1%	2.3%
AAP	–	–	0.5%

We find that roughly 28.8% of our corpus jointly mentioned two parties, and 6.0% tweets mentioned three parties (see Table 2 for the detailed break-up of 28.8%). We also observe that an insignificant 0.5% tweets mentioned all four parties. However, most of them are junk tweets, which simply used multiple trending political hashtags. Hence, we have not considered them for our analysis. Table 2 strongly indicates that mix tweeting is prevalent in a multi-party context, but these mix tweets are not suitable for automatic processing.

In such situations, we can employ computational intelligence techniques which combines commonsense computing and linguistics to decipher the political sentiments of social media users. Commonsense computing emphasizes the obvious things that people usually know, and most of the times leave unstated [4,39]. When social media users communicate with each other on the Twitter platform, then they have a common understanding of the background knowledge — how one object/concept is related to another object/concept [4]. Thus, we also need to consider how political parties relate to each other. As we mentioned earlier, the Indian context is a 'multi-party system with relatively high fragmentation (regional parties have a strong presence only in certain parts of the country)' [22]. For example, in the 2014 General election, 464 political parties contested, and around 300 parties participated in less than 5 seats out of a total of 543 seats [22]. Electoral alliances between political parties in India are formed well in advance to remain politically relevant in the national context. As we already noted, NDA was an alliance between the BJP and other regional parties. Similarly, the UPA was also an alliance between the INC and a few other regional parties. This fine-grained understanding of the Indian context enables us to assimilate how one political orientation/ideology is related to another orientation/ideology [4].

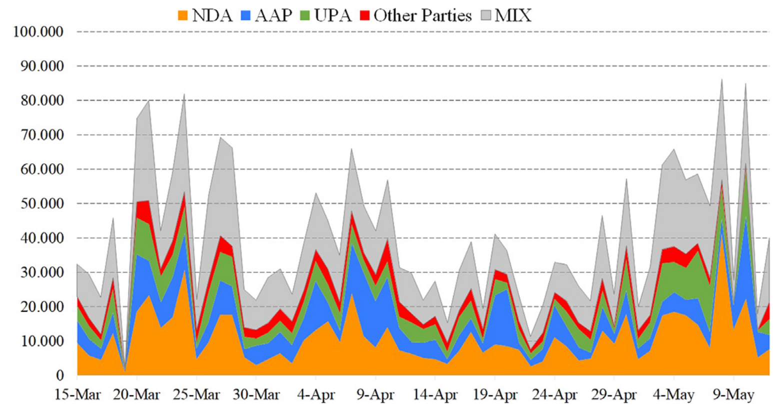

Fig. 2. Party-wise tweet distribution.

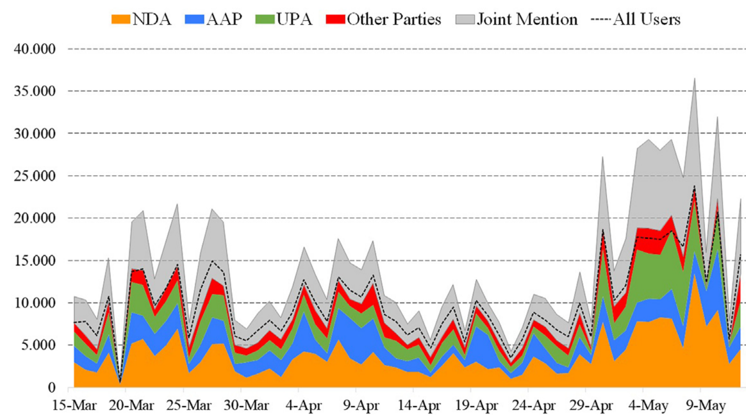

Fig. 3. User-wise tweet distribution.

5. User-level mix tweeting pattern

This section probes the complexity of the mix tweeting patterns at the user level. For the sake of brevity and clarity, we are reporting the tweeting patterns of 10 randomly selected users. The identities of these users are masked, and they are labeled as *User_1* to *User_10*. Figs. 4a, 4b, and 4c report the user-level tweeting pattern of these 10 randomly selected users through the Chord diagram. Chord diagram allows representing complex inter-relationships between entities graphically. Entities are arranged circularly, and relationships between entities are depicted through connecting arcs.

A unique color depicts each entity along the circumference of the circle. We are reporting two types of entities, namely users (i.e. 10 representative users) and their tweeting patterns. Our chord diagrams are delineating the tweet distribution pattern of these 10 representative users. The length of the archs, in these diagrams, is proportional to the tweet volume of different categories. For instance, in Fig. 4a, the length of the arc of *User_4* (depicted by the color green) is longer than *User_6* (depicted by the color red). This indicates *User_4* has tweeted more number of tweets than *User_6* during our study period.

Similarly, the length of the MIX tweet arc (depicted by the color black) is longer than the remaining 4 categories, namely NDA, UPA, AAP, and OTH. Hence, the cumulative volume of mix tweets by these 10 representative users is higher than the remaining 4 categories. As we noted earlier, mix tweets can be of 11 types, but reporting all of them in one diagram will make the graphics incomprehensible. Thus, the first chord diagram, i.e. Fig. 4a, reports all (i.e., 11 types of) mix tweets under one category (MIX). From Fig. 4a, we can also interpret that *User_4* has mostly posted 4 types of tweets: *NDA*-related tweets, *UPA*-related tweets, *MIX* tweets, and only a few tweets for *OTH* parties.

However, we want to probe the mix tweeting pattern of these 10 representative users. Hence, the second chord diagram, i.e. Fig. 4b, delineates the mix tweeting pattern of these users. In this chord diagram, we focus on joint mentioning of two parties (6 types) and reports joint mentioning of 3 or 4 parties under one category (mentioned as *3/4 PARTY*). Kindly note that Fig. 4b is not reporting the exclusive tweets posted by these users (for graphical clarity and ease of interpretability). For instance, Fig. 4b reveals that *User_4* has mostly posted MIX tweets, which mentioned NDA and UPA (i.e. *NU*). Similarly, *User_2* has mainly posted MIX tweets that jointly mentioned AAP and NDA (i.e. *AN*).

Subsequently, the third chord diagram, i.e. Fig. 4c, delineates the joint mentioning of 3 or 4 parties; and it is not reporting the exclusive tweets and joint mentioning of two parties (for visual clarity). For instance, *User_4* has mostly mentioned NDA, UPA, and AAP (i.e. *NUA*) in Fig. 4c. This is in accordance to her tweeting pattern in Fig. 4b, where *User 4* mostly mentioned *NU*. These simple graphical representations, especially Figs. 4b and 4c, elucidate the complexity of user-level mix tweeting patterns in a multi-party context. Hence, predicting the political sentiments of these users based on their mix tweeting pattern is a challenging task. We consider the underlying principles of computational intelligence techniques to hypothesize the relationship between political orientation and mix tweeting patterns in the following section.

6. Mix tweets and political sentiments: Hypotheses

Commonsense computing highlights the importance of shared understanding about the background knowledge [4,39]. In our

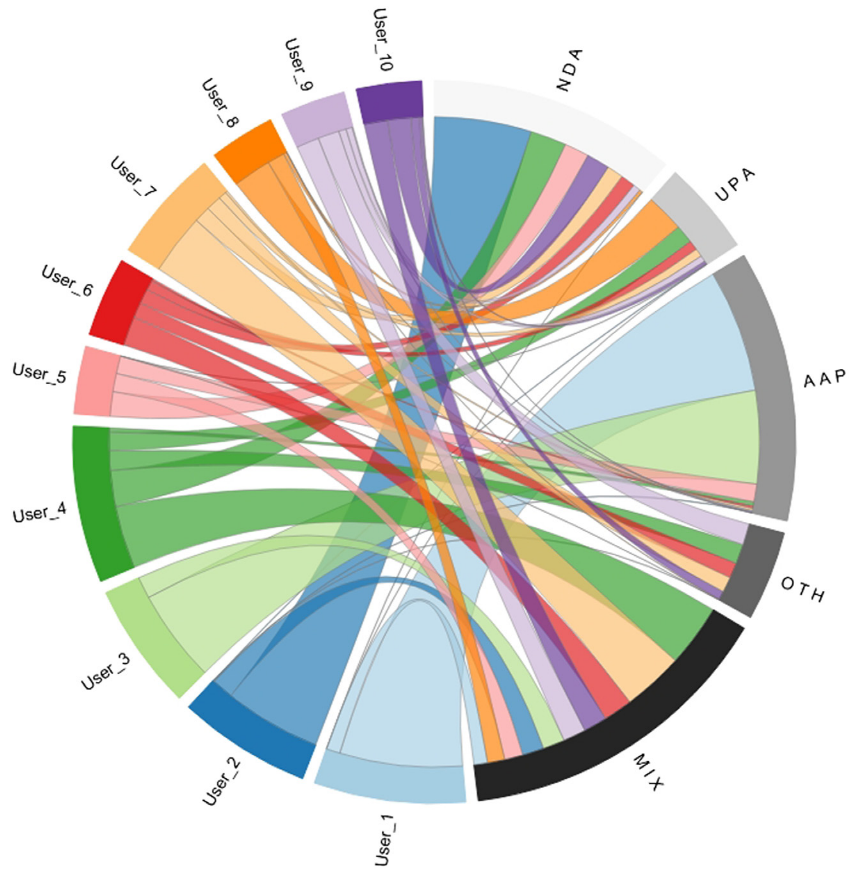

Fig. 4a. User-level tweet distribution pattern.

research context, this background knowledge is the political rivalry and ideologies of major parties that voters know, and most of the time, they do not mention it explicitly. For instance, NDA and UPA were two leading contenders in the 2014 Indian election, and they had ideological differences. The political analyst found that UPA mostly 'counts on the lower social orders as its most important voting bloc'. On the contrary, BJP and its coalition 'represent the socially privileged, the educated, and high-income groups [62]. The Hindu-nationalist ideology of BJP significantly differs from the political ideology of INC [62,63]. However, in the 2014 election, 'BJP for the most part kept quiet about Hindu nationalism' and focused on the 'Congress-led government's corruption and poor performance, particularly the slow growth, unemployment, and inflation' [63]. The BJP and INC 'faced each other in 189 head-to-head contests, and the BJP won 166 of these' parliamentary constituencies [63]. This trend is effectively getting captured in our distribution of mix tweets. For instance, 11.5% of our corpus jointly mentioned the NDA and UPA within a tweet, which reveals their relevance and competitive rivalry in Indian politics (see Table 2).

We observe that mix tweets are mostly a comparative evaluation of major political parties. For instance, a supporter of Party A would prefer to compare Party A with other political competitors such as *'the election manifesto of Party A is more inclusive than Party B (or more inclusive than both Party B and Party C)*. This will give a political advanatge to Party A. However, this supporter of Party A would not be interested in posting a tweet like *'the economic growth was higher during the regime of Party B in comparison to Party C'*. This second tweets compares the economic performance of Party B in comparison to Party C. Comparative evaluation between Party B and Party C will not give any political advanatge to Party A. Thus, there will be a negative likelihood of the joint mentioning of *Party B* and *Party C* from the supporter of *Party A*. This common background knowledge allows us to hypothesize the political opinion of users based on their mix tweeting pattern. Similar to [58], our notations and assumptions for our proposed hypotheses are as follows:

i = *Party* i *where* $i \in I$ *and* $I = \{1, \ldots, n\}$. In our context I = {AAP, NDA, UPA, OTH}. Thus, n = 4.

j = User j ($j = 1, \ldots, N$)

C (n, k) = the number of *k-combinations* from a given *set S* of *n elements*

$m_{i_1 i_2 j} = m^{th}$ mix tweet which jointly mentions Party i_1 and Party i_2 by User *j*. $m_{i_1 i_2 j}$ can be of *C (4,2)* in our context, i.e., joint mentioning of 2 parties from a set of 4 parties (*AAP, NDA, UPA, OTH*). Similarly, $m_{i_1 i_2 i_3 j} = m^{th}$ mix tweet which jointly mentions Party i_1, Party i_2, and Party i_3 by User *j*. $m_{i_1 i_2 i_3 j}$ can be of *C (4,3)* types in our context. Hence, we hypothesize,

Hypothesis 1 (*Regarding Joint Mentioning of Two Parties*)**.** A supporter of Party i_p will have

a. a higher propensity to post mix tweets such as $m_{i_p i_1 j}$, $m_{i_p i_2 j}$ … $m_{i_p i_{n-1} j}$ and
b. a negative propensity to post mix tweets such as $m_{i_1 i_2 j}$, $m_{i_2 i_3 j}$ … $m_{i_{n-2} i_{n-1} j}$ where $i \in I$ and $I = \{p, 1, 2, \ldots, n-1\}$ $j = 1, \ldots, N$.

Hypothesis 2 (*Regarding Joint Mentioning of Three Parties*)**.** A supporter of Party i_p will have

a. a higher propensity to post mix tweets such as $m_{i_p i_1 i_2 j}$, $m_{i_p i_1 i_3 j}$, …, $m_{i_p i_2 i_3 j}$ … $m_{i_p i_{n-2} i_{n-1} j}$ and

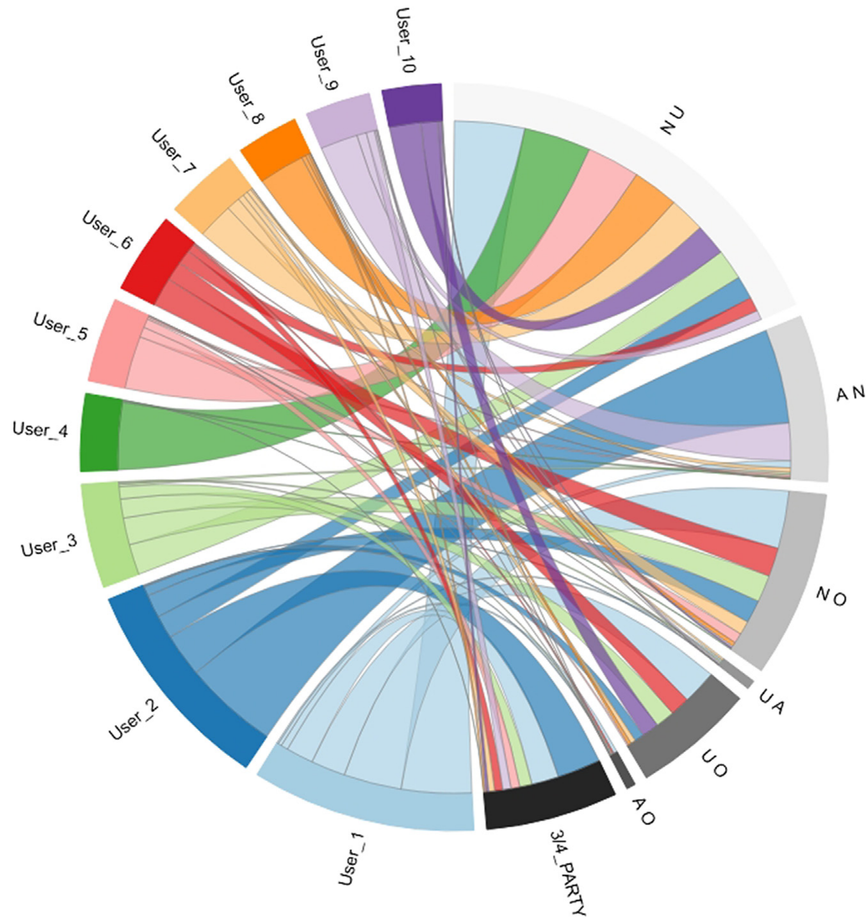

Fig. 4b. User-level mix tweet distribution pattern (excluded exclusive tweets). AN: AAP-NDA, AU: AAP-UPA, AO: AAP-Others, NU: NDA-UPA, NO: NDA-Others, UO: UPA-Others.

b. a negative propensity to post mix tweets such as $m_{i_1i_2i_3j}$, $m_{i_1i_2i_4j}$, … $m_{i_{n-3}i_{n-2}i_{n-1}j}$ where $i \in I$ and $I = \{p, 1, 2, \ldots, n-1\}$ $j = 1, \ldots, N$.

In the Indian context, these hypotheses can be presented in a tabular format (refer to Table 3).

7. Predicting political sentiment by using multinomial logit regressional

The Hypotheses 1 and 2 predict a causal relation between mix tweeting pattern and political opinion. We have employed a multinomial logit (MNL) regression model to test the above causal relations. MNL models can consider unordered categorical outcomes as the dependent variable. Here, the categorical outcomes for our econometric modeling are *NDA supporter*, *AAP supporter*, *UPA supporter,* and *Supporter of Other parties (OTH).*

We have coded these outcomes as 1, 2, 3, and 4 (these numerical values are arbitrary and unordered). Considering $i = Party\ i$ *where* $i \in I$ *and* $I = \{AAP, NDA, UPA, OTH\}$ and $j =$ User j ($j = 1, \ldots, N$) the explanatory variables for our MNL models are as follows,

$V_{i_1i_2.j}$ = Volume of mix tweets which jointly mention Party i_1 and Party $i_2$by User j

$V_{i_1i_2.i_3.j}$ = Volume of mix tweets which jointly mention Party i_1, Party $i_2$and Party $i_3$by User j

Our dependent variable DV_{PLj} is the political leaning of user j. Thus, our user-level estimation models to predict political leanings are as follows,

For Hypothesis 1 : $DV_{PLj} = \alpha + \sum \beta.V_{i_1i_2.j} + \text{error}\ \forall i, j$

For Hypotheses 1 & 2 : $DV_{PLj} = \alpha + \sum \beta_1.V_{i_1i_2.j}$
$+ \sum \beta_2.V_{i_1i_2.i_3.j} + \text{error}\ \forall i, j$

7.1. Dependent variable for multinomial logit models

Our dependent variable is the political opinion (or political leaning) of a voter. Following prior studies, such as [50,64], we consider the volume-based approach to classifying a user's political sentiment. [50] has noted that 'tweets referring to one name of the party or its leader' can effectively capture political views. Thus, [50] has counted 'the frequencies referencing parties (or party leaders) in a user's tweets and then assigned the most frequently referenced party to the user's political party.'

Similarly, [64] has also explored political leanings through the 'usage of manually selected, highly partisan hashtags.' Interestingly, the classification accuracies of [50,64] are comparable with the machine-learning-based approach of [34] and [35]. Our notations for the *volume-based classifier* are as follows,

i = Party i where $i \in I$ *and* $I = \{1, \ldots, n\}$
j = User j ($j = 1, \ldots, N$)
V_{ij} = Volume of the total number of tweets for Party i by User j
$m_{ij} = m^{th}$ tweet posted for Party i by User j ($m_{ij} = 1, \ldots, V_{ij}$)
SV_{ij} = Share/percentage of tweets for Party i by User j

$$SV_{ij} = \frac{V_{ij}}{\sum_{i\in I} V_{ij}} \forall i, j$$

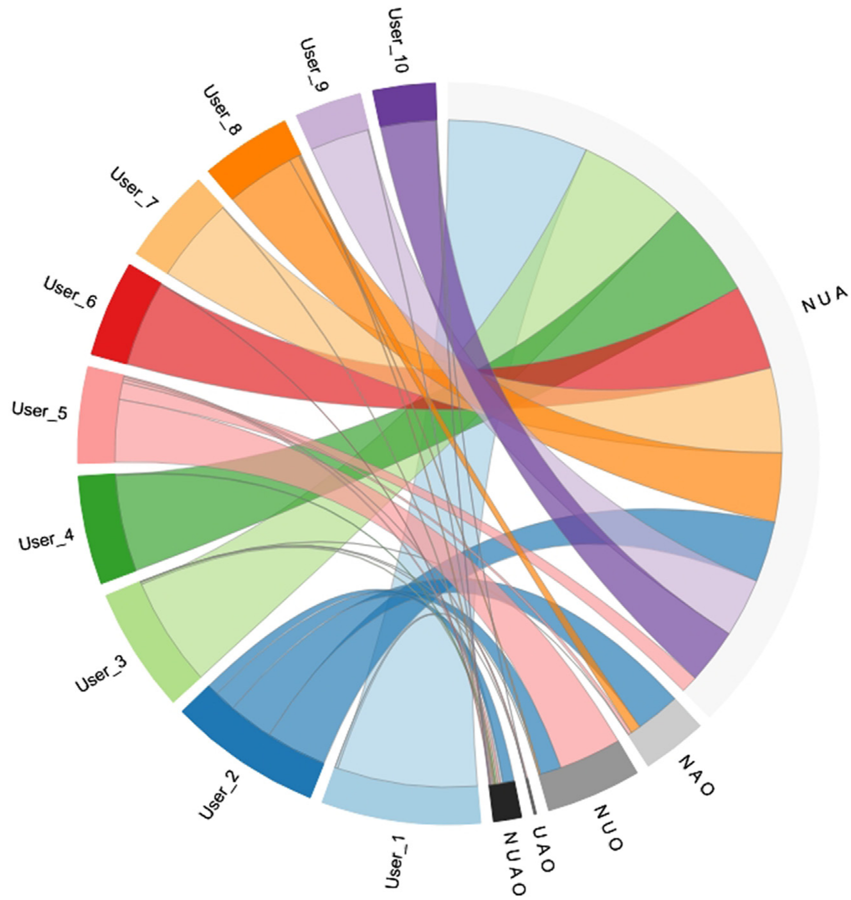

Fig. 4c. User-level mix tweet distribution pattern for 3 or more parties. ANU: AAP-NDA-UPA, ANO: AAP-NDA-Others, AUO: AAP-UPA-Others, NUO: NDA-UPA-Others.

Table 3
Tabular presentation of Hypotheses 1 and 2.

Political sentiment	Hypothesis 1a Higher propensity	Hypothesis 1b Negative propensity	Hypothesis 2a Higher propensity	Hypothesis 2b Negative propensity
AAP Supporter	**A**N, **A**U, **A**O	NU, NO, UO	**A**NU, **A**NO, **A**UO	NUO
NDA Supporter	A**N**, **N**U, **N**O	AU, AO, UO	A**N**U, A**N**O, **N**UO	AUO
UPA Supporter	A**U**, N**U**, **U**O	AN, NO, AO	AN**U**, A**U**O, N**U**O	ANO
Supporter of OTH	A**O**, N**O**, U**O**	AN, NU, AU	AN**O**, AU**O**, NU**O**	ANU

Note: AN: AAP-NDA, AU.: AAP-UPA, AO.: AAP-Others, NU: NDA-UPA, NO: NDA-Others, UO: UPA-Others, ANU: AAP-NDA-UPA, ANO: AAP-NDA-Others, AUO: AAP-UPA-Others, NUO: NDA-UPA-Others.

PV_j = Volume-based political leaning of User j

$$PV_j = \underset{i \in I}{\mathrm{Max}}[SV_{ij}] = \underset{i \in I}{\mathrm{Max}}[\frac{V_{ij}}{\sum_{i \in I} V_{ij}}]\forall j$$

Say, if a voter posts N tweets and n of them exclusively mention Party i_1, and m_1, m_2, m_3 tweets for Party i_2, Party i_3 and Party i_4, respectively (where $n > m_1, m_2, m_3$ and $N = n + m_1 + m_2 + m_3$), then we classify her as a supporter of Party i_1. However, if she posts n tweets for both the Party i_1 and Party i_2, the remaining $(N - 2n)$ tweets for the Party i_3 and Party i_4, and $n \geq (N - 2n)/2$, then we label her as a non-classifiable user. We have assumed that n is always the maximum (or jointly maximum) volume of tweets across 4 political parties. To avoid a potential endogeneity problem, we have considered only exclusive tweets (not MIX tweets) for our dependent variable.

7.2. Results and discussion

Our entire corpus has 0.15 million unique users, but our volume-based classifier was able to label (or classify) the political sentiment of around 0.12 million users (and the remaining users were non-classifiable — mostly due to an insufficient number of exclusive tweets). However, we note that a significant portion of these 0.12 million users did not tweet mix tweets. Hence, we did not consider these users in our regression analysis since these users are not appropriate for our analysis. Our final sample, thus, for MNL regression analysis, includes 51,091 users who have posted MIX tweets (in Table 4).

We use the statistical software STATA for our regression analysis - specifically the STATA command 'mlogit'. As we mentioned earlier, the dependent variable in our MNL models is the political orientation of users (i.e. *NDA supporter*, *AAP supporter*, *UPA supporter*, and *Supporter of OTH)*, and these are unordered categorical outcomes. We have coded these political leanings as 1, 2, 3, and 4. However, these numerical values are arbitrary and unordered because 'mlogit' estimates a set of coefficients, $\beta^{(1)}$, $\beta^{(2)}$, $\beta^{(3)}$ and $\beta^{(4)}$, corresponding to each outcome as follows,

$$\Pr(y = 1) = \frac{e^{X\beta^{(1)}}}{e^{X\beta^{(1)}} + e^{X\beta^{(2)}} + e^{X\beta^{(3)}} + e^{X\beta^{(4)}}}$$

Table 4
MNL regression analysis of mix tweets.

	Model 1	Model 2	Model 3	Model 4	*E.S.*
Panel AAP					
AN	0.159***	0.120***	0.372***	0.329***	+
AO	0.353***	0.344***	0.358***	0.329***	+
NO	−0.314***	−0.290***	−0.146***	−0.143***	-
NU	−0.226***	−0.235***	−0.246***	−0.262***	-
UA	0.364***	0.338***	0.218***	0.186***	+
UO	−0.321***	−0.272***	−0.453***	−0.432***	-
NAO		0.156***		0.362***	+
NUA		0.145***		0.182***	+
NUO		−0.302***		−0.168***	-
UAO		0.034		−0.031	+
Const.	−0.413***	−0.397***	0.348***	0.342***	
Panel NDA					
AN			0.213***	0.209***	+
AO			0.004	−0.016	-
NO			0.168***	0.147***	+
NU			−0.020***	−0.027***	+
UA			−0.147***	−0.152***	-
UO			−0.132***	−0.160***	-
NAO				0.205***	+
NUA				0.037***	+
NUO				0.134***	+
UAO				−0.064*	-
Const.			0.760***	0.738***	
Panel UPA					
AN	−0.213***	−0.209***			-
AO	−0.004	0.016			-
NO	−0.168***	−0.147***			-
NU	0.020***	0.027***			+
UA	0.147***	0.152***			+
UO	0.132***	0.160***			+
NAO		−0.205***			-
NUA		−0.037***			+
NUO		−0.134***			+
UAO		0.064*			+
Const.	−0.760***	−0.738***			
Panel OTH					
AN	−0.418***	−0.402***	−0.205***	−0.193***	-
AO	0.210***	0.244***	0.215***	0.228***	+
NO	0.029***	0.047***	0.197***	0.193***	+
NU	−0.237***	−0.222***	−0.256***	−0.249***	-
UA	−0.179***	−0.151***	−0.326***	−0.303***	-
UO	0.338***	0.361***	0.206***	0.201***	+
NAO		−0.331***		−0.126*	+
NUA		−0.253***		−0.217***	-
NUO		−0.087***		0.047**	+
UAO		−0.154*		−0.218**	+
Const.	−1.244***	−1.191***	−0.484***	−0.453***	
LR Chi^2	18 578.9***	19 727.2***	18 578.9***	19 727.9***	
Pseudo R^2	0.1506	0.1599	0.1506	0.1599	

N = 51 091; (two-tailed tests) | ES: Expected Sign of Coefficients.
***Indicate statistical significance at 0.1% level.
**Indicate statistical significance at 1% level.
*Indicate statistical significance at 5% level.

MNL estimation sets one of $\beta^{(1)}$, $\beta^{(2)}$, $\beta^{(3)}$, or $\beta^{(4)}$ to 0. If $\beta^{(1)} = 0$, the remaining coefficients $\beta^{(2)}$, $\beta^{(3)}$ and $\beta^{(4)}$ will measure the change relative to the y = 1 group. Similarly, if we set $\beta^{(2)} = 0$, the remaining coefficients will measure the change relative to the y = 2 group. The coefficients will differ because they have different interpretations, but the predicted probabilities for y = 1, 2, 3, and 4 will still be the same. So, it hardly matters whether we are setting $\beta^{(1)} = 0$ or $\beta^{(2)} = 0$. If we set $\beta^{(1)} = 0$, the equation becomes

$$\Pr(y = 1) = \frac{1}{1 + e^{X\beta^{(2)}} + e^{X\beta^{(3)}} + e^{X\beta^{(4)}}}$$

The relative probability of y = 2 to the base outcome is

$$\frac{Pr(y = 2)}{\Pr(y = 1)} = e^{X\beta^{(2)}}$$

For the sake of clarity, we report our explanatory variables as NU and NUA, respectively (where Party i_1 = NDA, Party i_2 = UPA, and Party i_3 = AAP). Likelihood Ratio Chi^2 for all four models is statistically significant at 0.1% level.

The four panels for the AAP, NDA, UPA, and OTH report the propensity of mix tweeting patterns for their respective supporters. In other words, model 1 (for Hypothesis 1) and model 2 (for both Hypotheses 1 and 2) in the AAP Panel (i.e., the first panel

Table 5
Sample tweets from our gold standard.

User #	Political leaning	Selected user-specific[a] MIX Tweets from our Gold Standard	Contextual interpretation
1	**AAP supporter**	– RT @XYZ: Completely reject #**Cong****BJP**ForeignFunded and choose honest **@ArvindKejriwal** – RT @XYZ: Criminal Candidates- **Shiv Sena** -80% **BJP** - 41.07% **BSP** - 28.57% **SP** - 27.27% **Cong** - 23.88% can they provide clean politics?	Arvind Kejriwal & AAP was against all other parties
2	**NDA Supporter**	– RT @XYZ: **NDA** is firmly committed to creating a Skilled India, breaking away from the Scam India under **UPA's** misrule – **BJP** releases Chargesheet against the misdeeds of the **UPA** Government **@narendramodi**	Narendra Modi was the Prime Ministerial candidate of NDA
3	**OTH Supporter**	– **AMMA** the ultimatum for Tamil Nadu & India!!! Drastic Achievements at short span despite noncooperation from looting **Congress** – **Jayalalithaa** ends campaigning; says she is better than **Narendra Modi**	A supporter of Jayalalithaa & AIADMK–a regional party
4	**OTH Supporter**	– **Mayawati** Is Our Leader told u 100 times, i dont give a damn on ur **Modi** okk! – If **#Mayawati** has her way, she is the only one capable of giving **#Modi** a tough fight at both state and national level	A supporter of *Mayawati* & BSP – a regional party
5	**UPA Supporter**	– RT @XYZ: **BJP**, **S.P.** and **BSP** has been in power in UP since 25 years, have they provided you Employment? **#RahulGandhi** – RT @XYX: **BJP** does politics of hatred and anger, on the other hand **Congress** believes in brotherhood and harmony: **#RahulGandhi**	Rahul Gandhi was the leading campaigner of Congress & UPA

[a]For brevity we have reported ONLY two MIX tweets for these 5 representative users (out of our sample of 1033 users), but most of them have tweeted more than 10/12 MIX tweets.

in Table 4) indicate the propensity of different categories of mix tweets by AAP supporters. We have reported the coefficients with their significance level in models 1 to 4. The signs and statistical significance of these coefficients would reveal the relationship between mix tweeting patterns and political leaning. Models 1 and 2 consider the NDA as a base category. Thus, the NDA Panel is blank for models 1 and 2. We also repeat our analysis by considering UPA as a base category in models 3 and 4. Hence, the UPA Panel is blank for models 3 and 4.

The column labeled *E.S.* (*expected signs*) indicates the hypothesized relationship between political sentiments and mix tweeting patterns of supporters (as reported in Table 3). For example, we hypothesized that APP supporters would have a higher propensity ('+' sign) to post mix tweets such as *AN*, *UA*, *AO* (Hypothesis 1a), *NUA*, *NAO* and *UAO* (Hypothesis 2a) and negative propensity ('−' sign) for other categories.

We observe that the coefficients (in Panel AAP) for categories like *AN*, *UA*, *AO*, (Hypothesis 1a) *NUA*, and *NAO* (Hypothesis 2a) are positive and statistically significant at the 0.1% level. In contrast, the coefficients of *NU*, *NO*, *UO* (Hypothesis 1b) and *NUO* (Hypothesis 2b) are negative and statistically significant at the 0.1% level. These results strongly support our predicted relationship for AAP supporters. Similarly, we hypothesized that NDA supporters would have a higher propensity ('+' sign) to post mix tweets such as *AN*, *NO*, *NU* (Hypothesis 1a), *NAO*, *NUA*, and *NAO* (Hypothesis 2a) and negative propensity ('−' sign) for other categories. Accordingly, the model 4 of Panel NDA reports that the coefficients of AN and NO are positive and significant at 0.1% level. Coefficients of UA and UO are negative and significant at the 0.1% level. AO is negative (as hypothesized) but not significant. We did not get support for NU. So, this pattern mostly confirms our Hypothesis 1a and 1b for NDA supporters. Subsequently, the coefficients of *NAO*, *NUA*, and *NAO* are positive and significant at the 0.1% level. Similarly, UAO is negative and statistically significant at 5% level. This strongly confirms the Hypothesis 2a and 2b for NDA supporters. Our MNL analysis broadly confirms our hypotheses. Hence, we argue that in a multi-party context *mix tweeting patterns of users can predict their political leaning.*

8. Predicting political sentiments by using neural network-based models

Deep neural network-based architecture allows us to solve complex problems. Consequently, neural network-based architectures are becoming popular for user-classification tasks that range from image processing to financial prediction. Prior studies have employed neural network-based analysis in the context of social media data [6]. In the domain of user identification, researchers have used neural network-based architecture, such as RNN, to explore mobile usage patterns of customers [65], aggressive behaviors of social media users [66], mobile user identification [67] and others. Hence, we have employed a neural network-based classification for our research problem. We would also like to probe whether our approach is more accurate for a specific set of political supporters or not.

Following this recent stream of research, we are employing a recurrent neural network (RNN), long short time memory (LSTM), and bidirectional LSTM (Bi-LSTM) for predicting the political leaning of users. Generally, in the classical neural network setup, all inputs and outputs interact with each other independently. However, to understand the underlying meaning of data, we need to consider models such as RNN, which holistically address this aspect. In addition to the input gate and output gate, LSTM also has an additional forget gate. This additional forget gate makes LSTM superior to other models. Bi-LSTMs are an extension of traditional LSTMs. In some particular domains, Bi-LSTM is better than classical LSTM. For our deep learning-based analysis, we have prepared a gold standard of 1033 users (385 AAP users, 290 NDA users, 133 OTH users, and 225 UPA users). All of these users have posted multiple mix tweets.

Similar to the previous section, our output variable is the labeled political sentiment of the user, and our input variables are user-level mix tweeting patterns i.e. $V_{i_1 i_2 .j}$ and $V_{i_1 i_2 .i_3 .j}$. To prepare our gold standard, we have carefully considered the linguistic content of the tweets of these users and manually annotated their political sentiments (refer to Table 5 for a few representative samples). Similar to [54], our annotation process considered polarized topics which 'require the user to side exclusively with one position' and 'polarized keywords' to identify the polarized users. In the context of the 2017 French Presidential Election, [31] considered the left-wing vis-à-vis right-wing hashtags in their analysis. Hence, we have also considered the usage of hashtags by our users. Our final set of users has shown consistency in their political orientation, i.e. their political sentiment was consistent towards only one political ideology in our corpus.

Table 6
Classification accuracies for neural-network models.

Method	Loss	Accuracy
RNN	0.44	0.86
LSTM	0.48	0.82
Bi-LSTM	0.46	0.87

8.1. Results and discussion

We have performed training in batch sizes of 32 for all three models. We have used 80% of our gold-standard data for training (out of which 20% data was used for validation), and the rest unexposed 20% data for the testing purpose. Our setting for the dropout layer is 0.5 for RNN, LSTM, and Bi-LSTM. The recurrent dropout for LSTM is also 0.5. For predicting the final class, we have considered softmax activation in our final classification layer. We use rmsprop as our optimizer for all these models. Results were broadly consistent across various model hyper-parameters. We report the best results here. Tables 6 and 7 report the accuracy and confusion matrix, respectively, for all these models. Interestingly, all our models have performed well in predicting the political leaning of users. Our accuracy is in the range of 82% to 87%. Bi-LSTM is our best performing model. Table 7 reveals that these models are consistent across parties.

9. Conclusion and discussion

Nowadays, Twitter data are attracting information science researchers for various socio-economic issues, such as electoral analysis [8]. A dominant theme of this stream of research is identifying the political sentiment of users. [4] also pointed out that existing sentiment analysis approaches are efficient for explicitly expressed opinions and emotions, but not for implicitly expressed opinions and emotions. Our literature review also reveals that extant literature mostly considered this problem in electoral contexts as a binary classification task and tested their proposed approach in a two-party context like republican versus democratic supporters in the USA. However, [19,22] argued that electoral analysis, in a multi-party context, such as India, is not a binary classification task. Hence, our paper has attempted to address this.

We set our study in the Indian electoral context. We have employed multinomial logistic regression as well as neural network-based models to probe the political opinions of users. Specifically, we have explored the mix tweet that has jointly mentioned more than one party. Our study demonstrates that user-level mix tweeting patterns can predict the political opinion of voters. Our research has elucidated the significance of mix tweets in a multi-party context. To the best of our knowledge, none of the prior studies considered mix tweets in the political domain. Hence, demonstrating the relevance of the mix-tweeting pattern as an input feature for user classification is the core contribution of our study.

9.1. Theoretical contributions

A supporter of Party A can post appreciative (i.e. positive sentiment) tweets for Party A and criticizes (i.e. negative tweets) Party B. Simple aggregate-level volumetric analysis will ignore these nuances. Our corpus also reveals that a significant portion of users post complicated comparative tweets where they evaluate one party in comparison to another party. However, prior studies have suggested to ignore tweets that have jointly mentioned more than one party [16]. We find that the political sentiments of users are not ‘explicitly’ expressed in her mix tweeting pattern. On the theoretical front, we have attempted to combine commonsense computing and linguistics to decipher the political opinions of social media users. Following [4,39], we used our contextual knowledge to investigate the ‘implicitly’ expressed political sentiment of a user through her mix tweeting pattern in the Indian context. In short, we have attempted ‘to predict which party, if any, the user is most likely to vote for’ from their mix tweeting pattern [49]. None of the prior studies considered the intricate user-level mix tweeting pattern as an input variable. Our paper has demonstrated the dominance of mix tweets in the political discourse and relevance of mix tweets in probing the political opinions of users. This is the core theoretical contribution and originality of our study. On the methodological front, we did not come across any study which has employed both multinomial logistic regression and neural network models for user-classification. Neural network models are incredibly efficient for user-classification tasks, but regression analysis allows us to interpret the causal relations between variables. Hence, future studies can also consider this combined approach to have a fine-grained understanding of the user’s sentiment. Understanding the causal relationship is extremely crucial for the practical application of our approach, which we elucidate in the following section.

9.2. Managerial and social implications

Our opinion mining approach can be generalized beyond the electoral context, and this can have several practical implications. For instance, within an industry, there can be multiple competitors, but all of them are not competing for the same set of customers. Hence, competing brands can be classified in different clusters, and these clusters are commonly known as strategic groups in management literature. Sometimes these strategic groups are apparent. For example, in the automobile industry, high-end cars such as Ferrari, Lamborghini, and Porsche will be in one cluster, whereas Mercedes and BMW will be in another group. Subsequently, popular cars such as Toyota, Ford, General Motors, Chrysler, Honda, etc. will come in another cluster. Users often compare two similar offerings in their product reviews and jointly mention them in the product review. We argue that investigating these mix reviews will help us to identify the strategic clusters within an industry. Probably, a user will not jointly mention Lamborghini and Honda. However, it will be challenging to predict ex-ante whether Toyota, Ford, General Motors,

Table 7
Confusion matrix for neural-network models.

		True Label				True Label				True Label			
		AAP	NDA	OTH	UPA	AAP	NDA	OTH	UPA	AAP	NDA	OTH	UPA
Predicted Label	AAP	37%	2%	0%	0%	32%	2%	1%	1%	33%	1%	0%	0%
	NDA	1%	18%	0%	2%	3%	21%	1%	1%	2%	24%	0%	3%
	OTH	0%	0%	10%	0%	1%	0%	9%	1%	2%	1%	10%	0%
	UPA	0%	6%	1%	21%	2%	4%	0%	20%	0%	1%	0%	19%
		RNN Model				LSTM Model				Bi-LSTM Model			

Chrysler, Honda, and Nissan are in the same cluster, or there are sub-clusters. Thus, investigating joint mentioning of brands in the context of product reviews, which can range from alcoholic beverages to auto manufactures, will help us to answer — who competes with whom? Whether Samsung mobiles are competing with iPhones or Xiaomi mobiles? Possibly, a rigorous analysis will reveal that high-end Samsung phones compete with iPhone, whereas Xiaomi is a threat to their low-end phones. Similarly, in the context of a sports tournament, supporters frequently refer to other competing teams on the Twitter platform [58]. Thus, mix-tweet analysis in a multi-team context can answer – who supports whom? Knowing this fanbase is crucial – especially if we consider the phenomenal sports merchandise-related business opportunities [58]. For instance, a significant portion of revenues of football clubs, such as F.C. Barcelona or Real Madrid C.F., comes from their sports merchandise such as the team's jerseys. Fans display their loyalties by wearing team jerseys, and they become walking billboards for corporate sponsors, such as Adidas or Nike. Identifying the potential supporter base would lead to a win–win situation for sports clubs and their corporate sponsors.

In the electoral context, analyzing mix tweets can aid political parties to identify their closest competitor. Accordingly, political parties can decide their campaign strategies. Additionally, this proposed approach will also help political parties to identify their support base. Efficient user-classification will help political parties to segment the electorate, and subsequently, these parties can employ targeted political campaign. However, it is also worth noting that user-classification can also have adverse social implications. For instance, the social media data breach issue during the 2016 U.S. Presidential election raised various ethical concerns. Thus, extracting publicly available tweets of voters, and subsequently analyzing their political orientation can potentially lead to violation of data privacy rights. So, information science researchers need to maintain the delicate balance between academic pursuits and ethical concerns. We have consciously masked the identity of all users in our reported findings to ensure data privacy concerns.

9.3. Limitations and future scope of work

As we noted earlier, one major limitation of prior electoral studies was disproportionate participation on the Twitter platform [27]. A few active users drive the entire discussion, and the majority prefers to remain quiet. For example, 77% of our users posted less than 10 tweets during our entire study period. In the context of disproportionate participation, [36] noted that predicting political opinions is a challenging task for 'those who rarely discuss politics'. [27] also pointed out that the 'validity of the generalizations that one can make from that potential data source is conditional on our ability to overcome the limitations arising from the fact that participation, at least in the political discussion on this platform, is not homogeneously distributed among users'. Similarly, our approach also requires multiple mix tweets from a specific user to predict her political orientation accurately. We have collected data for nearly two months. Thus, our study period allowed us to crawl multiple tweets from a particular user, and we have successfully analyzed their mix tweeting pattern. However, our proposed methodology will not be an appropriate approach for inactive users or for an event that is temporal.

In the domain of sentiment analysis, one stream of researchers is arguing that multiple modalities offer better insights in comparison to only text or visual data [68]. Multiple languages [69, 70], and modalities [71], such as texts, images, videos, and social links, are related to each other in the context of social media. Interestingly, in the 2016 U.S. Presidential election, both Clinton and Trump have aggressively used images and videos through the online mobile photo-sharing platform Instagram [44]. Thus, it would be interesting to consider multiple languages and modalities in the political domain. However, the scope of our paper, as well as lack of relevant data, did not allow us to probe this yet, but we leave it to future work.

CRediT authorship contribution statement

Aparup Khatua: Conceptualization, Data curation, Methodology, Writing - original draft, Software. **Apalak Khatua:** Writing - original draft, Investigation, Validation, Writing - review & editing. **Erik Cambria:** Supervision, Methodology, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] E. Cambria, T. Mazzocco, A. Hussain, C. Eckl, Sentic medoids: Organizing affective common sense knowledge in a multi-dimensional vector space, in: International Symposium on Neural Networks, Springer, Berlin, Heidelberg, 2011, pp. 601–610.

[2] C. Catal, M. Nangir, A sentiment classification model based on multiple classifiers, Appl. Soft Comput. 50 (2017) 135–141.

[3] V. García-Díaz, J.P. Espada, R.G. Crespo, B.C.P. G.-B.ustelo, J.M.C. Lovelle, An approach to improve the accuracy of probabilistic classifiers for decision support systems in sentiment analysis, Appl. Soft Comput. 67 (2018) 822–833.

[4] E. Cambria, A. Hussain, C. Havasi, C. Eckl, Common sense computing: From the society of mind to digital intuition and beyond, in: Biometric ID Management and Multimodal Communication, 2009, pp. 252–259.

[5] A. Hussain, E. Cambria, Semi-supervised learning for big social data analysis, Neurocomputing 275 (2018) 1662–1673.

[6] T.C. Wong, H.K. Chan, E. Lacka, An ANN-based approach of interpreting user-generated comments from social media, Appl. Soft Comput. 52 (2017) 1169–1180.

[7] M. Rathan, V.R. Hulipalled, K.R. Venugopal, L.M. Patnaik, Consumer insight mining: aspect based twitter opinion mining of mobile phone reviews, Appl. Soft Comput. 68 (2018) 765–773.

[8] A. Jungherr, Twitter use in election campaigns: A systematic literature review, J. Inform. Technol. Polit. 13 (1) (2016) 72–91.

[9] A. Tumasjan, T.O. Sprenger, P.G. Sandner, I.M. Welpe, Predicting elections with twitter: What 140 characters reveal about political sentiment, in: Proceedings of the International Conference on Weblogs and Social Media (ICWSM), IEEE, 2010.

[10] A. Bovet, H.A. Makse, Influence of fake news in twitter during the 2016 U.S. presidential election, Nature Commun. 10 (1) (2019) 7.

[11] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, D. Lazer, Fake news on twitter during the 2016 U.S. presidential election, Science 363 (6425) (2019) 374–378.

[12] P. Grover, A.K. Kar, Y.K. Dwivedi, M. Janssen, Polarization and acculturation in U.S. election 2016 outcomes–can twitter analytics predict changes in voting preferences, Technol. Forecast. Soc. Change 145 (2019) 438–460.

[13] Z. Jin, J. Cao, H. Guo, Y. Zhang, Y. Wang, J. Luo, Detection and analysis of 2016 us presidential election related rumors on twitter, in: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, Springer, 2017, pp. 14–24.

[14] B. O'Connor, R. Balasubramanyan, B.R. Routledge, N.A. Smith, From tweets to polls: Linking text sentiment to public opinion time series, in: Proceedings of the International Conference on Weblogs and Social Media (ICWSM), IEEE, 2010.

[15] J. Shin, L. Jian, K. Driscoll, F. Bar, Political rumoring on twitter during the 2012 U.S. presidential election: Rumor diffusion and correction, New Media Soc. 19 (8) (2017) 1214–1235.

[16] E.T.K. Sang, J. Bos, Predicting the 2011 Dutch Senate Election Results with Twitter, in: Proceedings of the Workshop on Semantic Analysis in Social Media, 2012.

[17] T. Graham, D. Jackson, M. Broersma, New platform, old habits? Candidates' use of twitter during the 2010 British and Dutch general election campaigns, New Media Soc. 18 (5) (2016) 765–783.

[18] M. Skoric, N. Poor, P. Achananuparp, E.P. Lim, J. Jiang, Tweets and votes: A study of the 2011 Singapore general election, in: Proceedings of the 45th Hawaii International Conference OnSystem Sciences (HICSS), IEEE, 2012.
[19] P. Burnap, R. Gibson, L. Sloan, R. Southern, M. Williams, 140 characters to victory? Using twitter to predict the U.K. 2015 general election, Elect. Stud. 41 (2016) 230–233.
[20] F. Franch, (Wisdom of the crowds) 2: 2010 U.K. election prediction with social media, J. Inform. Technol. Politi. 10 (1) (2013) 57–71.
[21] V. Kagan, A. Stevens, V.S. Subrahmanian, Using twitter sentiment to forecast the 2013 pakistani election and the 2014 indian election, IEEE Intell. Syst. 30 (1) (2015) 2–5.
[22] A. Khatua, A. Khatua, K. Ghosh, N. Chaki, Can #twitter_trends predict election results? Evidence from 2014 Indian general election, in: Proceedings of the 48th Hawaii International Conference OnSystem Sciences (HICSS), IEEE, 2015.
[23] X. Yang, C. Macdonald, I. Ounis, Using word embeddings in twitter election classification, Inform. Retr. J. 21 (2–3) (2018) 183–207.
[24] P.T. Metaxas, E. Mustafaraj, D. Gayo-Avello, How (not) to predict elections, in: Proceedings of the Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), IEEE, 2011.
[25] A. Jungherr, P. Jürgens, H. Schoen, Why the pirate party won the german election of 2009 or the trouble with predictions: A response to sprenger, a. and sander, t.o. and welpe, i.m. predicting elections with twitter: what 140 characters reveal about political sentiment, Social Science Computer Review 30 (2) (2012) 229–234.
[26] D. Gayo-Avello, I wanted to predict elections with twitter and all I got was this lousy paper A balanced survey on election prediction using twitter data, 2012, Available at: http://arxiv.org/pdf/1204.6441.pdf.
[27] P. Barberá, G. Rivero, Understanding the political representativeness of twitter users, Soc. Sci. Comput. Rev. 33 (6) (2014) 712–729.
[28] S.C. McGregor, R.R. Mourão, L. Molyneux, Twitter as a tool for and object of political and electoral activity: Considering electoral context and variance among actors, J. Inform. Technol. Polit. 14 (2) (2017) 154–167.
[29] S. Rossi, M. Rossi, B. Upreti, Y. Liu, Detecting Political Bots on Twitter during the 2019 Finnish Parliamentary Election, in: Proceedings of the 53rd Hawaii International Conference on System Sciences(HICSS), 2020, pp. 2430–2439.
[30] F. Morstatter, Y. Shao, A. Galstyan, S. Karunasekera, From alt-right to alt-rechts: Twitter analysis of the 2017 German federal election, in: Proceedings of the World Wide Web (WWW) Conference, 2018, pp. 621-628.
[31] K. Sokolova, C. Perez, Elections and the twitter community: The case of right-wing and left-wing primaries for the 2017 french presidential election, in: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2018, pp. 1021–1026.
[32] D. Gayo-Avello, Don't turn social media into another 'Literary Digest' poll, Commun. ACM 54 (10) (2011) 121–128.
[33] D. Rao, D. Yarowsky, A. Shreevats, M. Gupta, Classifying latent user attributes in twitter, in: Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents, ACM, 2010.
[34] M. Conover, J. Ratkiewicz, M.R. Francisco, B. Gonçalves, F. Menczer, A. Flammini, Political polarization on twitter, in: Proceedings of the International Conference on Weblogs and Social Media (ICWSM), IEEE, 2011.
[35] M. Pennacchiotti, A.M. Popescu, A machine learning approach to twitter user classification, in: Proceedings of the International Conference on Weblogs and Social Media (ICWSM), IEEE, 2011.
[36] R. Cohen, D. Ruths, Classifying political orientation on twitter: It's not easy!, in: Proceedings of the International Conference on Weblogs and Social Media (ICWSM), IEEE, 2013.
[37] E. Cambria, A. Hussain, C. Havasi, C. Eckl, Senticspace: visualizing opinions and sentiments in a multi-dimensional vector space, in: International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Springer, Berlin, Heidelberg, 2010, pp. 385–393.
[38] I. Chaturvedi, E. Ragusa, P. Gastaldo, R. Zunino, E. Cambria, Bayesian network based extreme learning machine for subjectivity detection, J. Franklin Inst. B 355 (4) (2018) 1780–1797.
[39] N. Howard, E. Cambria, Intention awareness: Improving upon situation awareness in human-centric environments, Human-Centric Comput. Inform. Sci. 3 (1) (2013) 1–17.
[40] E. Cambria, A. Hussain, C. Havasi, C. Eckl, Sentic computing: Exploitation of common sense for the development of emotion-sensitive systems, in: Development of Multimodal Interfaces: Active Listening and Synchrony, 2010, pp. 148–156.
[41] Y.Q. Xia, E. Cambria, A. Hussain, H. Zhao, Word polarity disambiguation using Bayesian model and opinion-level features, Cogn. Comput. 7 (3) (2015) 369–380.
[42] D. Preoţiuc-Pietro, Y. Liu, D. Hopkins, L. Ungar, Beyond binary labels: political ideology prediction of twitter users, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017, pp. 729–740.
[43] E. Cambria, M. Grassi, A. Hussain, C. Havasi, Sentic computing for social media marketing, Multimedia Tools Appl. 59 (2) (2012) 557–577.
[44] L. Buccoliero, E. Bellio, G. Crestini, A. Arkoudas, Twitter and politics: Evidence from the U.S. presidential elections 2016, J. Mark. Commun. 26 (1) (2020) 88–114.
[45] D. Gayo-Avello, A meta-analysis of state-of-the-art electoral prediction from twitter data, Soc. Sci. Comput. Rev. 31 (6) (2013) 649–679.
[46] A. Khatua, E. Cambria, A. Khatua, I. Chaturvedi, Let's Chat about Brexit! A Politically-Sensitive Dialog System Based on Twitter Data, IEEE International Conference on Data Mining Workshops (ICDMW) (2017) 393–398, http://dx.doi.org/10.1109/ICDMW.2017.57.
[47] A. Bermingham, A.F. Smeaton, On using Twitter to monitor political sentiment and predict election results, in: Proceeding of IJCNLP conference, 2011.
[48] P.L. Francia, Free media and twitter in the 2016 presidential election: The unconventional campaign of donald trump, Soc. Sci. Comput. Rev. 36 (4) (2018) 440–455.
[49] A. Makazhanov, D. Rafiei, M. Waqar, Predicting political preference of twitter users, Soc. Netw. Anal. Min. 4 (1) (2014) 1–15.
[50] A. Boutet, H. Kim, E. Yoneki, What's in your tweets? I know who you supported in the U.K. 2010 general election, in: Proceedings of the International Conference on Weblogs and Social Media (ICWSM), IEEE, 2012.
[51] S. Poria, I. Chaturvedi, E. Cambria, F. Bisio, Sentic LDA: Improving on LDA with Semantic Similarity for Aspect-Based Sentiment Analysis, in: Proceedings of IJCNN, 2016, pp. 4465–4473.
[52] Y. Gu, T. Chen, Y. Sun, B. Wang, Ideology detection for twitter users with heterogeneous types of links, 2016, Available at: arXiv preprint arXiv:1612.08207.
[53] F.M.F. Wong, C.W. Tan, S. Sen, M. Chiang, Quantifying political leaning from tweets, retweets, and retweeters, IEEE Trans. Knowl. Data Eng. 28 (8) (2016) 2158–2172.
[54] M. Coletto, C. Lucchese, S. Orlando, R. Perego, Polarized user and topic tracking in twitter, in: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2016, pp. 945–948.
[55] K. Johnson, D. Goldwasser, Identifying stance by analyzing political discourse on twitter, in: Proceedings of the First Workshop on NLP and Computational Social Science, 2016, pp. 66-75.
[56] R. Tatman, L. Stewart, A. Paullada, E. Spiro, Non-lexical features encode political affiliation on twitter, in: Proceedings of the Second Workshop on NLP and Computational Social Science, 2017, pp. 63–67.
[57] D.J.S. Oliveira, P.H.D.S. Bermejo, P.A. dos Santos, Can social media reveal the preferences of voters? A comparison between sentiment analysis and traditional opinion polls, J. Inform. Technol. Polit. 14 (1) (2017) 34–45.
[58] A. Khatua, A. Khatua, Cricket world cup 2015: Predicting user's orientation through mix tweets on twitter platform, in: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), ACM, 2017, pp. 948–951.
[59] A. Khatua, E. Cambria, K. Ghosh, N. Chaki, A. Khatua, Tweeting in Support of LGBT? A Deep Learning Approach, Proceedings of the ACM India Joint International Conference on Data Science and Management of Data, 2019, https://doi.org/10.1145/3297001.3297057.
[60] R. Basu, A. Khatua, A. Jana, S. Ghosh, Harnessing Twitter Data for Analyzing Public Reactions to Transportation Policies: Evidences from the Odd-Even Policy in Delhi, India, Proceedings of the Eastern Asia Society For Transportation Studies (EASTS), 2017.
[61] A. Khatua, A. Khatua, How Stock Market Reacts to Budget Announcement? Through the Lens of Social Media in Indian Context, Manag. Labour Stu. 40 (3-4) (2015) 239–251, http://dx.doi.org/10.1177/0258042X16634568.
[62] A. Varshney, India's democratic challenge, Foreign Aff. 8 (2007) 6–93.
[63] E. Sridharan, Behind modi's victory, J. Democr. 25 (4) (2014) 20–33.
[64] K. Dyagilev, E. Yom-Tov, Linguistic factors associated with propagation of political opinions in twitter, Soc. Sci. Comput. Rev. 32 (2) (2014) 195–204.
[65] S. Yao, S. Hu, Y. Zhao, A. Zhang, T. Abdelzaher, Deepsense: A unified deep learning framework for time-series mobile sensing data processing, in: Proceedings of the 26th International Conference on World Wide Web (WWW), 2017, pp. 351–360.
[66] J. Risch, R. Krestel, Aggression identification using deep learning and data augmentation, in: Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018), 2018, pp. 150–158.
[67] L. Sun, Y. Wang, B. Cao, S.Y. Philip, W. Srisa-An, A.D. Leow, Sequential keystroke behavioral biometrics for mobile user identification via multiview deep learning, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2017, pp. 228–240.
[68] I. Chaturvedi, R. Satapathy, S. Cavallari, E. Cambria, Fuzzy commonsense reasoning for multimodal sentiment analysis, Pattern Recognit. Lett. 125 (2019) 264–270.
[69] S.L. Lo, E. Cambria, R. Chiong, D. Cornforth, Multilingual sentiment analysis: From formal to informal and scarce resource languages, Artif. Intell. Rev. 48 (4) (2017) 499–527.
[70] H. Peng, E. Cambria, A. Hussain, A review of sentiment analysis research in chinese language, Cogn. Comput. 9 (4) (2017) 423–435.
[71] E. Cambria, A. Hussain, Sentic album: Content-, concept-, and context-based online personal photo management system, Cogn. Comput. 4 (4) (2012) 477–496.