How to cite this article:

Islam, M. K., Ahmed, M. M., Zamli, K. Z., & Mehbub, S. (2020). An online framework for civil unrest prediction using tweet stream based on tweet weight and event diffusion. *Journal of Information and Communication Technology,* 19*(1)*, 65-101. https://doi.org/10.32890/jict2020.19.1.4

AN ONLINE FRAMEWORK FOR CIVIL UNREST PREDICTION USING TWEET STREAM BASED ON TWEET WEIGHT AND EVENT DIFFUSION

[1]Md Kamrul Islam, [2]Md Manjur Ahmed, [2]Kamal Zuhairi Zamli & [3]Salman Mehbub

[1]Lorraine Research Laboratory in Computer Science and its Applications, University of Lorraine, France

[2]Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Malaysia.

[3]Department of Computer Science & Engineering, Jessore University of Science & Technology, Bangladesh.

kamrul.islam@loria.fr; manjur@ump.edu.my; kamalz@ump.edu.my; s.a.mehbub@gmail.com

ABSTRACT

Twitter is one of most popular Internet-based social networking platform to share feelings, views, and opinions. In recent years, many researchers have utilized the social dynamic property of posted messages or tweets to predict civil unrest in advance. However, existing frameworks fail to describe the low granularity level of tweets and how they work in offline mode. Moreover, most of them do not deal with cases where enough tweet information is not available. To overcome these limitations, this article proposes an online framework for analyzing tweet stream in predicting future civil unrest events. The framework filters tweet stream and classifies tweets using linear Support Vector Machine (SVM) classifier. After that, the weight of the tweet is measured and distributed among extracted locations to update the overall weight in each location in a day in a fully online manner. The weight history is then used to predict the status of civil unrest in a

Received: 17/11/2018 **Revised:** 7/5/2019 **Accepted:** 22/05/2019 **Published:** 23/12/2019

location. The significant contributions of this article are (i) A new keyword dictionary with keyword score to quantify sentiment in extracting the low granularity level of knowledge (ii) A new diffusion model for extracting locations of interest and distributing the sentiment among the locations utilizing the concept of information diffusion and location graph to handle locations with insufficient information (iii) Estimating the probability of civil unrest and determining the stages of unrest in upcoming days. The performance of the proposed framework has been measured and compared with existing logistic regression based predictive framework. The results showed that the proposed framework outperformed the existing framework in terms of F1 score, accuracy, balanced accuracy, false acceptance rate, false rejection rate, and Matthews correlation coefficient.

Keywords: Text classification, information diffusion, sentiment analysis, polynomial regression, connected graph.

INTRODUCTION

Social networking platforms allow sharing of information, ideas, feelings and events among individuals, communities, and organizations. These platforms work as a tool for people in a positive or negative way (Chumwatana, 2018). The ubiquitous use of these platforms has enabled its users to communicate with thousands of other users within a few seconds.The users build a virtual community by connecting each other and transmitting information (Olanrewaju & Ahmad, 2018). Azpeitia, Ochoa-Zezzatti, & Cavazos, (2017) noted that the socio-cultural aspect of individuals can be used to characterize the behaviour of a community. A lot of such platforms exist in the world and among them Twitter has drawn the notable attention of scholars as it has an impressive number of users and open read characteristics (Oh, Eom, & Rao, 2015; Valenzuela, 2013). Due to its high reachability and popularity, in recent years, researchers have investigated the role of Twitter in motivating, planning and mobilizing civil unrest events (Muthiah et al., 2015; Oh et al., 2015; Van Dyke & Amos, 2017). Civil unrest is a kind of social problem that includes riots, demonstrations, marches, protests, barricades, sit-ins and strikes. Sometimes it can cause a significant amount of economic and political loss (Dermisi, 2017; Passarelli & Tabellini, 2017). Several studies have found that most of today's civil unrest events are planned and mobilized in advance on social networking platforms (Filchenkov, Azarov, & Abramov, 2014; Muthiah

et al., 2015). This social dynamic feature of social networking platforms like Twitter has motivated social scientists in analyzing tweet streams to detect and forecast unrest events while they are unfolding (Korolov et al., 2016; Watts, 2013).

There are many unrest predictions and forecasting frameworks that exist based on social network streams (Galla & Burke, 2018; Kang et al., 2017; Wu & Gerber, 2018) or based on Short Message Service (SMS) or texting via Smartphone (Chiluwa, 2018). However most of these frameworks are missing two critical issues. Firstly, they lack details. In other words, they fail to describe the low granularity level of sentiment inside a tweet. There are several methods for extracting hidden sentiment in tweets (Parlar, Özel, & Song, 2018). However, they are mostly frequency based (e.g. word frequency, inverse document frequency, word presence, and relevance frequency) (Parlar et al., 2018) and frequency based sentiment weighting is not by itself a good measure of sentiment as it fails to look more in depth inside the tweet. For an illustration, consider these two tweets, "Today, protestors leave the road as police build *barricades*" and "Today, protestors leave the road as police started to *fire*". Though, both tweets have equal negative sentiment (equal frequency), but in practice, the second tweet bears more negative weight than the first tweet. Secondly, the distribution of Twitter users is not symmetric in all cities or countries in the world (O'Leary, 2015).

Some countries have large numbers of tweet users (e.g. USA, Mexico and Saudi Arabia) while some countries have very low numbers of users (e.g. Bangladesh, Afghanistan). This fact results in a situation where some countries have insufficient numbers of tweet messages to predict civil unrest events. Almost, all frameworks fail to handle locations where enough tweet information is not available. However, a partial solution is provided using heterogeneous and multiple data sources for prediction (Korkmaz et al., 2015, 2016). But collecting information from multiple sources is a time-consuming and costly task. Alternatively, a multi-tasking framework has been introduced based on sharing the information among all locations (Zhao et al., 2015, 2017). However, some locations may have no relation at all with the information and thus sharing all information among all locations is not a feasible task. The problem is reflected in the unstable performance over different locations. Thus, the scarcity of an online framework still exists to predict civil unrest events. The framework should be able to analyze the tweet in more granularities for in-depth sentiment analysis. The framework will also be able to predict and forecast unrest events in regions where sufficient tweet information is available as well as in regions where sufficient information is not available. It is hoped that the prediction framework will improve performance and show stable performance.

In this study, we propose an online framework to predict civil unrest occurrence by analyzing real-time tweet stream utilizing the concept of sentiment analysis based on word weight. The following are the main contributions of this study:

i) This study proposes a new keyword dictionary with scores to measure word weight and extract sentiments in depth. The weight of keywords describes the importance of words in a text. To the best of our knowledge, the three existing methods to analyze tweets at a more granularity level are: AFFIN (Nielsen, 2011), MPQA (Raina, 2013) and NRC (Mohammad, Kiritchenko, & Zhu, 2013). However, all of these methods only extract the positive or negative polarity of sentiments in text documents. Unlike these methods, we are interested in extracting the level of sentiment of tweets which are reflected in their weight. In our framework, we assign a positive score to every word in the keyword dictionary. Moreover, we also consider the effect of negative words and influencing words (Table 3) which are not included in other prediction frameworks (Cadena et al., 2015; Kang et al., 2017; Korkmaz et al., 2015; Wu & Gerber, 2018). This method of sentiment analysis enables the proposed framework to handle the first issue (as mentioned above) by analyzing tweets in depth.

ii) The revolution of social media enables the globalization of information that helps the world to know quickly what is happening in a city (Berestycki, Nadal, & Rodriguez, 2015; Zhao et al., 2017). The spread of information about any event or activity to other locations using communication networks is termed as information diffusion (Ferrara, 2018; Liang & Kee, 2018). Several studies discussed the importance of information diffusion in influencing and spreading unrest activities to geographical proximity locations (Huang, Boranbay-Akan, & Huang, 2016; Lang & De Sterck, 2014; Zhao et al., 2017). Utilizing the concept of information diffusion, we propose a new diffusion model of information related to civil unrest. In the proposed framework, the locations are extracted from tweet messages and users' profile. The weight of the tweets are distributed among extracted locations. The locations of study are represented by a location graph (Figure 4) which reflects the vicinity of the locations. The unrest sentiment of any location is represented by its measured weight using tweets which belong to that location. The location with insufficient tweets, gets the diffused weight from its neighbour locations to predict future civil unrest events. This dynamics of unrest events solves the second problem (as mentioned above).

iii) Henslin (2011) noted that a planned unrest event goes through a sequence of five distinct stages and they are: observe, agitation, mobilization, organization and occurrence. The unrest process starts from the observe stage when people are actually not happy in their present social behaviour. They express their unhappiness in social media like Twitter. During the agitation stage, people share their thoughts with known people about an issue in society and they also want to change it according to their view point. In the mobilization stage, some people agree and use blogs, and tweets for a movement. In the organization stage, people are organized, select leaders, policies and tactics to keep the movement alive. Finally, the unrest events happen at the occurrence stage. The participants actively take part in demonstrations, road marches, strikes, etc. We utilize this concept to identify the stages of civil unrest events based on tweet weight. We measure the probability of unrest events in future and distribute the probability over the event stages to forecast future unrest events occurring.

Recent literature in the domain of civil unrest is reviewed in the 'related work' section and a proposed framework is described in detail in the section that follows. After that, the performance of the proposed framework is measured and compared with existing prediction frameworks. The measured performance shows that the proposed framework has excellent capability/ potential to predict future civil unrest events and has outperformed existing prediction frameworks. Finally, the article summarizes and describes future research directions.

RELATED WORK

The quantification of social interactions on social media enables us to formulate social-psychology for objectively discovering social dynamics (Boonstra, Larsen, & Christensen, 2015). Social scientists and policymakers are especially interested in predicting civil unrest events as it can break geographical stability (El-Katiri, Fattouh, & Mallinson, 2014). The current study is the application of computational intelligence in the field of social science for forecasting civil unrest events based on the quantification of messages on Twitter. Several frameworks for extracting useful information from Twitter platform are extensively reviewed (Imran, Elbassuoni, Castillo, Diaz, & Meier, 2013; Kumar, Morstatter, & Liu, 2014). Based on these works, researchers have succeeded in varying degrees in predicting civil unrest

events. We reviewed articles from the recent five years from 2013 to 2017 in the domain of civil unrest prediction based on social media contents to describe their features, advantages and weaknesses.

Table 1

Dimensions for annotating related literature articles

Letter	Dimension	Category	Description
A	Classification method		Text classification method.
B	Prediction method		Future event prediction method.
C	Data source	Single (S)	Only one source to collect test dataset/ single social media.
		Multiple (M)	More than one source to collect test dataset/ multiple social media.
D	Analysis	Content (D1)	Mine only textual content for feature extraction and event forecasting.
		User profile (D2)	Mine user profiles and networks for extracting event related information.
E	Features	News count (NC)	Total related news frequency.
		Bag-of-words (BOW)	Related keywords and key phrases in dictionary.
		Sentiment (SM)	Ultimate intention of user.
		Temporal expression (TE)	Present time-related information.
F	Region(s)	Single (S)	Conduct experiments on protests happening in a country/region.
		Multiple (M)	Perform event forecasting which can be extended to more than one region.
G	Language(s)	Single (S)	Conduct experiments only on English Language tweets.
		Multiple (M)	Testbed consisting of multiple language content.
H	Dealing with insufficient event information		Handle locations which fail to provide enough unrest related tweets.
I	Gold Standard Report (GSR) Dataset		Ground truth for framework evaluation.

Table 1, illustrates several criteria in depth for comparing the reviewed civil unrest frameworks. From Table 1, the classification method criterion (A) describes the classifier used to classify the tweets on Twitter and prediction method criterion (B) presents the method used for predicting future unrest

events. Data source dimension (C) describes whether the framework uses single (S) or multiple (M) sources to generate its test dataset in completing prediction tasks. The common metadata for analyzing in extracting features are message content and user profile which are denoted by D1 and D2, respectively. The extracted features are categorized into news count (NC), bag-of-words (BOW), user's sentiment (SM), temporal or time-related information (TE) and the dimension is denoted as features (E). Whether the framework is applicable to multiple regions (M) or limited to a single region (S) is indicated by the region(s) criterion (F). The language dimension (G) describes the ability of the framework to process single (S) or multiple (M) languages. The next aspect (H) explains the fact as to whether or not the framework can predict unrest events in those locations where sufficient information is not available. The GSR (Gold Standard Report) dimension (I) is used to describe the ground truth dataset for framework evaluation. GSR contains the actual occurrences of civil unrest events.

Table 2 summarizes featured frameworks and compares reviewed articles. According to column A of Table 2, most of the articles use keyword dictionary for filtering and classifying contents into informative or uninformative classes (Cadena et al., 2015; Korkmaz et al., 2015; Manrique et al., 2013; Wu & Gerber, 2018; Xu, Lu, Compton, & Allen, 2014; Zhao et al., 2017). Some articles criticized this method of content classification as keywords alone cannot decide the relativity of content to civil unrest. Instead, they used keyword dictionary for filtering and the popular supervised classifiers like SVM (Chen & Neill, 2014; Korolov et al., 2016; Qiao & Wang, 2015), Naïve Bayes (Qiao et al., 2017; Ramakrishnan et al., 2014; van Noord, Kunneman, & van den Bosch, 2016) and n-gram classifier (Compton et al., 2014) for classification of content. An exception, Ranganath et al. (2016) used the latent discriminant classifier based on extracted latent dimension representation of content to identify related content.

Column B of Table 2, illustrates that nearly one-fourth of the reviewed articles search for future dates in the content for predicting unrest events occurring (Compton et al., 2014; Muthiah et al., 2015; van Noord et al., 2016; Xu et al., 2014). However, these frameworks suffer from the fact that nowadays protestors are smart and rarely specify dates of their planned events directly in their posts. In view of this limitation, other authors use regression techniques like logistics and Least Absolute Shrinkage and Selection Operator (LASSO) regression (Cadena et al., 2015; Korkmaz et al., 2015; Korolov et al., 2016; Qiao & Wang, 2015; Ramakrishnan et al., 2014; Wu & Gerber, 2018), HMM (Hidden Markov Model) (Qiao et al., 2017), random forest (Singh & Pal, 2018) and Long Short-Term Memory Networks (LSTM) (Galla & Burke, 2018) in predicting tasks that use historical event information.

Table 2

Taxonomy of Civil Unrest Event Prediction Framework

Article	A	B	C	D1	D2	E	F	G	H	I
Benkhelifa et al. (2014)	SA & ASR, Aggarwal frame.	IWARS	M	√		BOW, SM	M	M	×	-
Cadena et al. (2015)	Key word based	Logistic regression	S	√	√	BOW, NC	M	M	×	IARPA
Chen and Neill (2014)	SVM	NPHGS	M	√	√	SM, NC, BOW	M	S	×	MSMLN
Compton et al. (2014)	Logistic regression	Future dates	M	√	√	TE, BOW	S	M	×	MSMLN
Galla and Burke (2018)	ADA Boost with Random Forest	LSTM	M			GDELT features	M	M		GDELT
Hoegh et al. (2015)	Key word based	Bayesian model fusion framework	M	√		BOW, NC, TE	M	M	√	MSMLN
Hossny and Mitchell (2018)	Key word based	Naïve Bayes	S	√	√	BOW, NC, TE	M	M		-
Kang et al. (2017)	DBSCAN	2-Phase PPM	M	√		BOW, NC	M	S		MSMLN
Korkmaz et al. (2015)	Key word based	Logistic regression	M	√		BOW, KC	M	M	√	MSMLN
Korkmaz et al. (2016)	Key word based	Logistic regression with Lasso	M	√		BOW, NC	M	M	√	MSMLN
Korolov et al. (2016)	SVM	Logistic regression	S	√		BOW, TC	S	S	×	MSMLN
Manrique et al. (2013)	Key word based	Key terms momentum	S			BOW	M	S	×	GT
Muthiah et al. (2015)	RosetteLinguistics Platform (RLP)	Future dates	M	√	√	TE, BOW	S	M	×	MSMLN
Qiao and Wang (2015)	SVM	Logistic regression	S			GDELT features	M	M		GDELT
Qiao et al. (2017)	Bayes decision theory	HMM	S	√		GDELT features	M	M		GDELT
Ramakrishnan et al. (2014)	Naïve Bayes	Logistic regression	M	√	√	BOW, TE	S	M	×	MITRE
Ranganath et al. (2016)	Linear discriminant classifier	Brownian motion theory	S	√		BOW, NC, TE	S	S	×	MSMLN
Singh and Pal (2018)	Naïve Bayes	Random Forest	S	√		BOW, NC	S	S	×	MSMLN
van Noord et al. (2016)	Naive Bayes	Future dates.	S	√		NC, SM BOW	S	S	×	TwiNL
Wu and Gerber (2018)	Key word based	Logistic regression	S	√		NC, TE, BOW	M	M	×	GDELT
Xu et al. (2014)	Key word based	Future dates.	S	√		BOW, TE	S	M	×	GAP
Zhao et al. (2017)	Key word based	MTFL	S	√		BOW, NC	M	M	√	MITRE

*Table 1 contains the description of all symbols.

The rest of the authors introduce their own prediction models like data mining frameworks based indication and warning assessment, recognition system (IWARS) (Benkhelifa, Rowe, Kinmond, Adedugbe, & Welsh, 2014), graph-based non-parametric heterogeneous graph scan (NPHGS) model (Chen & Neill, 2014) and multi-task learning framework based Multi-Task Feature Learning (MTFL) model (Zhao et al., 2017), Bayesian model fusion framework (Hoegh, Leman, Saraf, & Ramakrishnan, 2015), Naïve Bayes (Hossny & Mitchell, 2018), and keyword frequency-based model (Manrique et al., 2013).

Column C of Table 2 shows that more than 50% of the articles use single social media like Twitter, Tumblr while less than 50% use multiple social media combing social networking platforms with news sources for collecting, analyzing and predicting civil unrest occurrences.

Based on column D of Table 2, most of the reviewed articles analyze media contents in their framework. With the exception of Manrique et al. (2013) who use Google Trends (GT) and two other frameworks which use Global Data on Events, Location, and Tone (GDELT) database features (Qiao et al., 2017; Qiao & Wang, 2015) in their prediction task instead of accessing media content or user profiles. Besides analyzing contextual metadata, some researchers (Cadena et al., 2015; Chen & Neill, 2014; Compton et al., 2014; Muthiah et al., 2015; Ramakrishnan et al., 2014) elicited several types of information from user profiles such as follower lists and home locations from their frameworks.

Based on column E of Table 2, all articles generated their own keywords and key-phrases dictionary in their respective prediction framework. Some of the articles also extracted temporal expressions like tomorrow, January 1, 7.00 a.m., etc., from news contents (Compton et al., 2014; Muthiah et al., 2015; Ramakrishnan et al., 2014; Wu & Gerber, 2018; Xu et al., 2014), while some of the articles used the frequency of news in their framework (Cadena et al., 2015; Chen & Neill, 2014; van Noord et al., 2016; Wu & Gerber, 2018; Zhao et al., 2017). Another important feature termed as the sentiment hidden in a post was extracted from a few articles for a more in-depth analysis of social media content to improve prediction performance (Benkhelifa et al., 2014; Chen & Neill, 2014; van Noord et al., 2016).

From column F of Table 2, it can be seen that the articles are classified into two classes (single, or multiple) based on the region(s) covered. About 43% of them are focused on a specific region such as the Netherlands, Latin America, and Africa whereas more than 50% of the reviewed articles are applicable to multiple regions.

Column G of Table 2 illustrates that nearly all popular social media like networking platforms and news media supply the features of providing

content in multiple languages. However, all the articles are not able to process any language in the domain of civil unrest prediction. We have categorized the articles into two categories (single or multiple). The first category can process single languages like English, Dutch, Hindi or Spanish (Chen & Neill, 2014; Korolov et al., 2016; Manrique et al., 2013; Qiao et al., 2017; van Noord et al., 2016) and the rest of the articles can process multiple languages.

It is also important to handle cases when a location has insufficient social media content to predict the occurrence of civil unrest. Column H of Table 2 noted that most of the articles missed this vital issue. It has been proven that civil unrest is diffused to neighbouring locations by communication networks (Ferrara, 2018; Huang et al., 2016; Lang & De Sterck, 2014; Zhao et al., 2017). Thus, considering this fact, two efforts have been found so far. The first effort comes from Korkmaz et al. (2015) and Korkmaz et al. (2016) where heterogeneous data sources are used, and the second effort comes from Zhao et al. (2015) and Zhao et al. (2017) where information is equally diffused to all locations to predict unrest events.

Finally, column I of Table 2, gives the Gold Standard Dataset (GSR) which is used by each article to measure performance. More than, one-third of them generate their own truth data set by MSMLN (Manually Searching Major Local Newspapers) to use as ground truth dataset. The other datasets include MITRE, Govt. Agencies Provided (GAP), Cross-National Time Series Dataset (CNTS), Intelligence Advanced Research Projects Activity (IARPA), GDELT (Global Data on Events, Location, and Tone), TwiNL, Google Trends (GT).

From the discussion on existing frameworks, most of the articles used their own keyword dictionary to filter tweet data sets. They also used the existing classifier to classify tweets and term frequency method to extract tweet sentiments. But, there was no existing work to measure the level of sentiment. In other words; they did not provide the low granularity level of tweet stream. Regression technique remains popular for the prediction of unrest tasks. Most of the frameworks processed tweet streams to predict unrest events in small set locations and generated their own ground truth data sets. The frameworks worked mostly in offline mode where whole datasets were required during the execution of the frameworks. Moreover, most of the existing frameworks were not able to predict civil unrest events where sufficient tweet information was available.

METHODOLOGY

An online framework has been proposed as in Figure 1 in this study, to analyze tweet streams for predicting civil unrest events in multiple regions

of the world. Overall, the framework starts with a 2-stage filtering (keyword and location based) of noisy tweet stream as will be explained later in this section. The stream then passes through a trained SVM classifier to classify each tweet into related or unrelated class. Using the proposed keywords scale (examples in Table 3), the weight of each tweet is measured and distributed to corresponding locations. The overall weight of each location is updated recursively. The weight of a location with an insufficient number of tweets is determined using the connected location graph. The weight indicates the negative sentiment of a citizen in a particular area towards an unrest event. This weight history is used to predict and forecast future unrest possibility in a specific area.

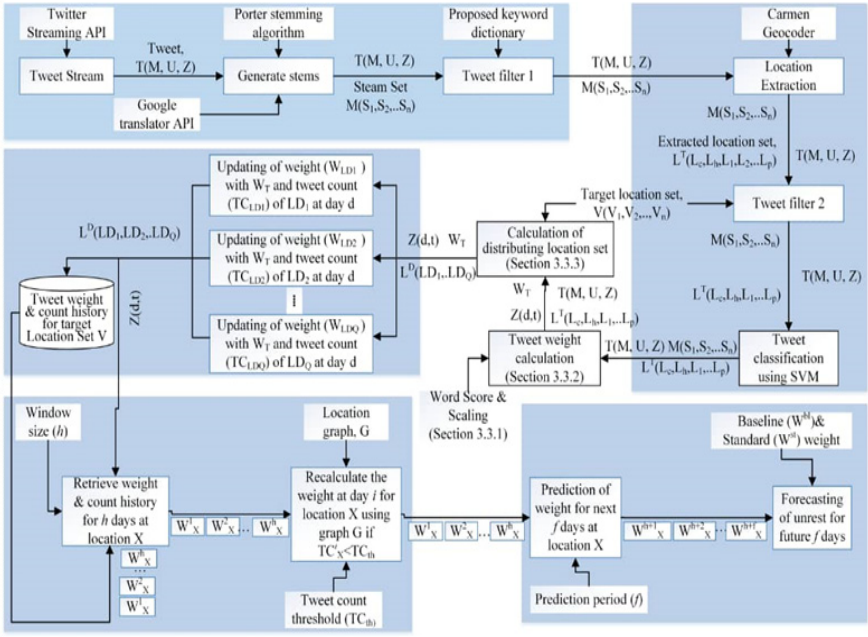


Figure 1. Proposed civil unrest prediction framework.

Data Acquisition, Cleaning and Keyword-based Filtering

Twitter platform provides two APIs for capturing tweets from Twitter. The first one is REST API that allows the downloading of 100 tweets per queries and 450 queries per 15 minutes (Ochoa-Zezatti, 2016). This limitation of REST API is overcome by another API called Streaming API that allows the streaming of tweets using a persistent HTTP connection without separating

user request. The proposed framework collects civil unrest related tweet stream using Streaming API. After acquiring the tweet stream, it is cleaned by deleting re-tweets as they carry existing information. Emojis, html links and punctuations are also removed. The raw tweet is then split into its words using classical tokenization and all non-alphanumeric characters and web links are also removed. A word may be in different parts of speech. For simplification, Porter stemming algorithm (Porter, 1980) is used to represent the derived words in their stem or root form.

A keyword-based filter is used to remove unnecessary tweets from the tweet stream. The filter is developed based on a keyword dictionary. Civil unrest related tweets on Twitter for 30 days are analyzed and a dictionary is developed with the help of a domain expert who is basically a social scientist to determine the keywords. The keyword identification method in the proposed framework is similar to some other efforts (Cadena et al., 2015; Chen & Neill, 2014; Korkmaz et al., 2015; Ranganath et al., 2016; Xu et al., 2014). The tweet which is written in English language is considered for keyword identification. The identified keywords form a dictionary called "keyword dictionary". Figure 2 represents the developed keyword cloud that contains 200 keywords.

Figure 2. Keyword cloud for proposed framework.

The proposed keyword dictionary is used to filter out unusable tweets. Suppose, a tweet message is represented by a tuple, T(M, U, Z) where M is the tweet message, U is the user profile and Z is the temporal information. The user profile, U is a metadata that includes name, permanent/home and current location of user. The temporal variable, Z stores the date and time of tweet posting. Tweet message, M is represented using its N stems that is, $M = \{S_1, S_2, S_3, ..., S_{n-1}, S_n\}$. Then, the relativity ($R_M$) of M to unrest is a binary variable (0 means related and 1 means unrelated) which is calculated based on its n stem's existence in keyword dictionary, D using the following logical expression in Equations 1 and 2.

$$R_M = E_{S_1} \vee E_{S_2} \vee E_{S_3} \vee ..... \vee E_{S_{n-1}} \vee E_{S_n} \quad (1)$$

Where, the existence of i^{th} stem (S_i) in dictionary D, $E_{S_i} = \begin{cases} 1, S_i \in D \\ 0, S_i \notin D \end{cases}$ (2)

The tweet, T is identified as unrelated if $R_M = 0$ and it is filtered out. Otherwise, T is passed though a second filter for further processing.

Location based filtering and classification

After keyword based filtering, the tweet is passed though a second filter called location based filter. The tweet with no location of interest is filtered out. This filtering prevents the processing of uninformative tweets. For this purpose, spatial information is extracted from the user profile (U) and tweet message (M). The current (L_M) and permanent/ home location (L_h) of the user are extracted using Carmen geocoding technique from the user's profile, U (Cadena et al., 2015). Besides, a tweet message may contain one or more explicit locations and the extracted location set is $L_M = \{L_1, L_2, L_3, ..., L_{p-1}, L_p\}$. The total extracted location set is $L^T = L_c \cup L_h \cup L^M = \{L_c, L_h, L_1, L_2, ..., L_{p-1}, L_p\}$ Suppose the target location set of the framework is for R locations. Then, the tweet, T is filtered out if the following filter (Equation 3) FL returns an empty location set (i.e. $FL = \phi$).

$$FL = V \cap L^T \quad (3)$$

If $FL \neq \phi$, then tweet T, belongs to at least one location from the target location set V and T is classified into two classes (informative or uninformative) using the linear support vector machine (SVM) classifier. Linear SVM has gained

popularity in the field of text classification because it not only minimizes errors but also has low complexity (Dilrukshi, De Zoysa, & Caldera, 2013) The basic principle of the linear SVM is to create a hyperplane that separates the samples into two classes (Zou & Jin, 2018). The closest samples to the hyperplane are computed from both classes. These samples are called support vectors and the distance between the line and the support vectors is called the margin. The hyperplane with the maximum margin is the optimal hyperplane. Before tweet classification, the linear SVM classifier is trained with training tweet sets. We downloaded tweets from Twitter platforms and manually labelled the tweets into either informative or uninformative classes. The annotated tweets are used as training data set for the classifier. In this stage, Count Vectorizer module of Scikit-learn Toolkit (Pedregosa et al., 2011) is used for tokenizing and extracting feature vectors from tweet datasets.

Recursive Solution to Measure Tweet Weight

The overall sentiment of people in a particular location in any specific day towards an unrest event is measured based on keyword scoring and tweet weight. The overall weight in a location is computed recursively so that the framework provides an online solution for unrest prediction. The following sections describe the scoring of keywords and measuring of individual tweet weights. The recursive method for computing the overall tweet weight is described in later sections.

Scaling of Keywords

The keyword dictionary contains a total of 200 keywords (Figure 2). But all of them do not carry the same level of information related to civil unrest events. Motivated by the work of Nielsen (2011), who scored the words in a sentence for sentiment analysis; we conducted scaling of the keywords to weigh the impact on an event. For example, the word 'murder' carries more weight than the word 'fight', and also the latter word carries more weight than the word 'trouble'. Moreover, the presence of adjectives, adverbs and determiners (influencing words) before a keyword affects the overall weight of sentiment of the keyword. For example, the appearance of 'huge' before 'protest' increases the weight of 'protest' while the word 'minor' before 'protest' decreases the weight. Also, the presence of a negation word before a keyword alters the polarity of the following word (Hasan, Sabuj, & Afrin, 2015). For example, 'no strike' opposes the meaning of 'strike'. In the proposed framework, the keywords are categorized into five categories and the influencing words in three special categories (Table 3). The keywords from first category describe

the highest negative impact of an unrest event on civil life while keywords from fifth category have the lowest negative impact of an event. Influencing words from the first special category increases the impact of keywords that come next while words from the second special category decreases the impact. The third special category is for negation words; these are extracted from the work of Hasan et al. (2015). Words which are not in the keyword dictionary or in a special category word list are termed as uncategorized words. Examples of words from these eight categories are given in Table 3 with their scale.

Table 3

Examples of keywords with its categories and scales

Category	**Example of keywords**	**Summation Scale (A)**	**Multiplication Scale (B)**
Category 1	kill, attack, bomb, terror, riot, clash, murder, etc.	5.0	1.0
Category 2	smash, blast, shot, fire, brutal, tyranny, etc.	4.0	1.0
Category 3	strike, damage, protest, fight, march, hijack, etc.	3.0	1.0
Category 4	rob, harass, assail, loot, trouble, defeat, etc.	2.0	1.0
Category 5	disorder, against, panic, apostasy, confusion, etc.	1.0	1.0
Special Category 1	huge, much, terrible, serious, dangerous, major, etc.	0	1.20 (Increase preceding keyword weight by 20%)
Special Category 2	few, near, slight, minor, etc.	0	0.80 (Decrease preceding keyword weight by 20%)
Special Category 3	not, no, never, n't.	0	-1.0 (Inverse preceding keyword weight)
Uncategorized	Other words	0	1.0

Furthermore, in Table 3, the proposed framework sets two kinds of scaling (summation and multiplication scale) for each keyword and influencing word. Keywords from the first category have the highest negative impact on society, so the summation score is 5.00 while the minimum summation scale of 1.00 is assigned to the fifth category of keywords as they have the lowest negative impact. This framework also assigns the multiplication scale value

of 1.20 for words in the first special category i.e. those that increase preceding keyword weight by 20%. Words from the second special category have a multiplication scale of 0.80 which decreases the preceding keyword weight by 20%. Words from the third special category have a multiplication scale value of -1.00 i.e. those that inverse preceding keyword weight.

Measure Weight of Tweets

After classification, each tweet (T) is processed further to measure its weight to determine whether the tweet (T) falls into "informative" class. The weight of T (M, U, Z) is computed based on its stem score and scaling (Table 3). The temporal information, Z of tweet T, includes the data (d) and time (t) of tweet posting. As described earlier, the message M, is represented by its stem set, $M = \{S_1, S_2, S_3, ..., S_{n-1}, S_n\}$ where n is the number of stems (root form of each word). Then, there exists a stem set P such that $P = M \cap D$ where D represents the keyword dictionary. The weight $W(t)$ of tweet T at time t, can be calculated using Equation 4.

$$W(t) = \frac{1}{|P|} \sum_{i=1}^{n} A_i B_{i-1} \quad where, B_0 = 1.0 \tag{4}$$

Where, A_i= Summation score of stem S_i and B_{i-1}= Multiplication scale of stem S_{i-1}

Weight Distribution

In the proposed framework, we distributed the weight of tweets instead of distributing raw tweets. This operation saves cost on computing the weight of the same tweet multiple times. After measuring the weight of tweet (T) at time t, the weight is distributed among the extracted locations of interest. The location set of interest is extracted from the extracted location set $L^T = \{L_c, L_h, L_1, L_2, ..., L_{P-1}, L_P\}$ using user profile U and tweet message (M) and the target location set is $V = \{V_1, V_2, V_3, ..., V_{R-1}, V_R\}$ for R locations. Whether the tweet (T) and its weight, $W(t)$ at time t, belongs to a target location V_j or not, is determined based on the distributor function Equation 5.

$$f(V_j, L^T) = \begin{cases} 1, V_j \in L^T \\ 0, V_j \notin L^T \end{cases} \tag{5}$$

The model of tweet (or its weight) distributor function is shown in Figure 3 as follows.

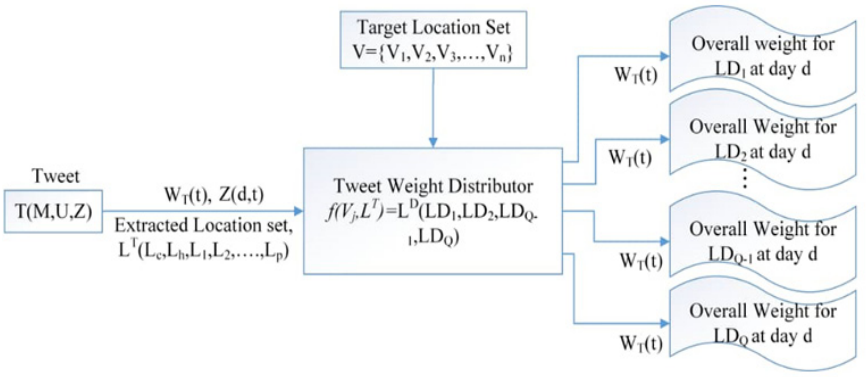


Figure 3. Tweet weight distributor.

If for any location V_j, the distributor function Equation 5 returns true for a tweet T, then tweet T belongs to the location V_j. In this case, V_j is added to the weight distributing location set, L^D, as stated in Equation 6.

$$L^D = L^D \cup V, \text{ if } f(V_j, T) = 1 \tag{6}$$

Suppose, Equation (6) generates the weight distribution location set $L^D = \{LD_1, LD_2, ..., LD_Q\}$ with Q locations where $0 < Q \leq R$ and $L^D \in V$. Finally, the weight, $W(t)$ of tweet (T) is distributed to every location from location set L^D. The overall weight and tweet count of every location in L^D is location updated using $W(t)$.

Recursive Weight Updating

The overall weight of any target location from location set $V = \{V_1, V_2, ..., V_{R-1}, V_R\}$ at any day d is updated in an online manner. The weight (W(t)) of a newly arrived tweet is used to update the overall weight, $O^d_{LD_i}$ and tweet count, $C^d_{LD_i}$ to every extracted location LD_i in distributing location set $L^D = \{LD_1, LD_2, ..., LD_Q\}$ where $L^D = \{LD_1, LD_2, ..., LD_Q\}$. If at time instant t, there are $C^d_{LD_i}(t)$ tweets from the tweet stream for location LD_i, then the overall weight $O^d_{LD_i}$ for LD_i is represented using Equation 7.

$$O^d_{LD_i}(t) = \frac{\sum_{k=1}^{t} W_{LD_j}(k)}{C^d_{LD_i}(t)} \tag{7}$$

To enable the online tweet stream processing, the computation of the overall weight for the next tweet will be updated recursively. If at time t, the weight is in a target location LD_i is $O^d_{LD_i}(t)$, then at (t+1)th time instant the weight $O^d_{LD_i}$ (t+1) is calculated using the recursive Equation 8.

$$O^d_{LD_i}(t+1) = \frac{[O^d_{LD_i}(t) \times C^d_{LD_i}(t)] + W_{LD_j}(t+1)}{C^d_{LD_i}(t)+1} \quad (8)$$

where, at time (t+1) the weight of new tweet is, $W_{LD_i}(t+1)$.

Remarks: The above Equation 8, for overall weight computation for any location D_i, works well for large sets of tweets in a target location. A low number of tweets may create a false positive weight where every single tweet sample's weight significantly affects the overall weight. For example, if 2 tweets exist in a target location and the overall weight is high, then generally it is interpreted as a high level of civil unrest polarity. However, at a practical level, the overall weight may not reflect the actual situation of the target location. Thus, the framework sets a threshold value TC_{th} of tweet counter to minimize this bias.

Predicting Future Weights

The weights of future days (prediction period) are predicted based on the weight history (window size) at location V_i. The weights are imported from the storage of weights. For a location (V_i), if the overall weight and tweet count at day d, are $O^d_{V_i}$ and $C^d_{V_i}$ respectively, then the weight is set using Equation 9.

$$WH^d_{V_i} = \begin{cases} O^d_{V_i}, C^d_{V_i} >= TC_{th} \\ 0 \quad , C^d_{V_i} < TC_{th} \end{cases} \quad (9)$$

In the field of time series prediction, multiple regression is one of the most popular techniques used in various applications like traffic flow prediction (Priambodo & Ahmad, 2018). For future weight prediction tasks, we will use the 2nd order polynomial regression for weight estimation as the weights of tweet stream have non-linear properties.

Equation 9 states that the weight is 0 for any location with tweet count ($C^d_{V_i}$) less than the count threshold TC_{th}. Therefore, the case of zero weight or information insufficiency is handled with connected location graph. The location graph $G=(V, E)$, is one kind of connected graph that visualizes the

connectivity among target locations. The set of L target locations $V = \{V_1, V_2, ..., V_{L-1}, V_L\}$ works as vertices and neighbouring set $E = \{e_1, e_2, ..., e_{n-1}, e_n\}$ works as the edge between two vertices. In graph G, $v_i(WH_{V_i}^d)$ represents the vertice v_i with weight $WH_{V_i}^d$. If a location V_i has weight average 0 and its adjacency vertex set is R, then the diffused weight of V_i is computed using Equation 10.

$$WH_{V_i}^d = \frac{1}{|Y|} \sum_{\forall v_j \in Y} WH_{V_j}^d \tag{10}$$

An illustration of the location graph is is shown in Figure 4 as follows.

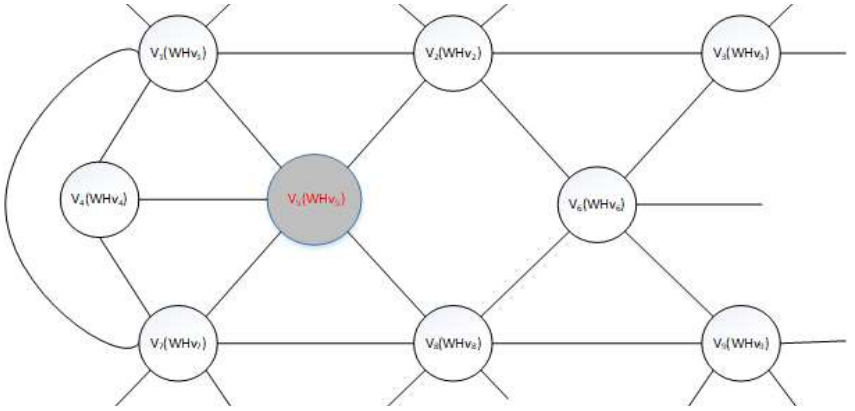


Figure 4. Connected location graph.

In Figure 4, the weight of the vertex/location, v_5 is 0.0 and its adjacency vertex set, A = $\{V_1, V_2, V_4, V_7, V_8\}$. Thus the diffused weight of the target location v_5 can be calculated as Equation 11.

$$DW_{V_5}^d = \frac{WH_{V_1}^d + WH_{V_2}^d + WH_{V_4}^d + WH_{V_7}^d + WH_{V_8}^d}{5} \tag{11}$$

Qiao et al. (2017) proved that a planned unrest event goes through a series of five stages. This fact implies that the overall weight at any location in a specific day not only depends on diffused weight, but also depends on the weight history. Thus, at day d, the weight ($WH_{V_i}^d$) of a location V_i is measured based on both diffused weight ($DW_{V_i}^d$) and weight on previous day ($WH_{V_i}^{d-1}$) as Equation 12.

$$WH_{V_i}^{d} = \rho WH_{V_i}^{d-1} + (1-\rho)DW_{V_i}^{d} \quad (12)$$

Where, the parameter ρ negotiates between the diffused weight ($DW_{V_i}^{d}$) and weight ($WH_{V_i}^{d-1}$) history to determine ($WH_{V_i}^{d}$). With ρ=0 setting implies that the weight history and diffused weight contribute equally to measure the current weight.

Forecasting Future Civil Unrest

An increase in the recent collective level of knowledge of a future protest is correlated with the occurrence of the onset of a protest in the near future (Wu & Gerber, 2018). Benkhelifa et al. (2014) also proved that future unrest events can also be predicted by comparing the current record with historical records of activities during normal and unrest periods in social networking platforms. Similarly, the proposed framework estimates the weight ($WH_{V_i}^{d+1}, WH_{V_i}^{d+2}.., WH_{V_i}^{d+F}$) in future dates (prediction period, F) using historical (window size, H) weights ($WH_{V_i}^{d}, WH_{V_i}^{d-1}.., WH_{V_i}^{d-H}$). We have used the 2nd order polynomial regression for weight estimation as the weights of tweet stream have non-linear properties. The estimated weight is used to forecast future unrest events with the help two other framework parameters, baseline weight (W^{bl}) and standard weight (W^{std}). Baseline weight (W^{bl}) is the minimum weight during times of peace and the standard weight (W^{std}) is the maximum weight during times of unrest. If in a location (V^{i}), the predicted weight (WH_{vj}^{d+j}) at jth the prediction day from current day d, then the probability (P_{vj}^{d+j}) of unrest event occurring as Equation 13.

$$P_{V_j}^{d+j} = \frac{WH_{V_j}^{d+j} - W^{bl}}{W^{std} - W^{bl}} \quad \text{for j=1,2,3, ……..,F} \quad (13)$$

This probability (P_{vj}^{d+j}) is used to determine the current stage of civil unrest at location (V_i) on jth future day. The measured probability is equally distributed over five unrest stages (observe, agitation, mobilization, organization, and occurrence) as in Figure 5. The final stage is the occurrence stage where unrest breaks out.

Thus in distributing a probability equally, if the probability falls between 0.80 and 1.0, then in this case the framework forecasts the occurrence of an unrest event at location V_i on day, d.

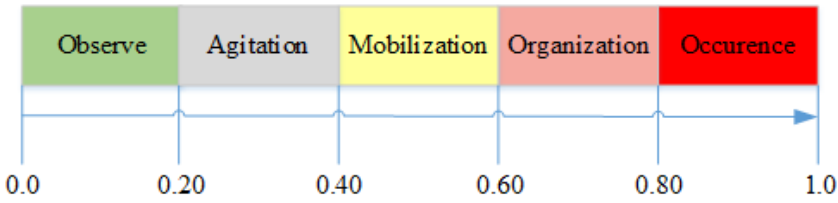


Figure 5. Probability distribution of unrest stages.

RESULTS AND DISCUSSION

Prior to the execution of the framework, three parameters namely tweet count threshold (TC^{th}) standard weight (W^{std}) and baseline weight (W^{bl}) were set. To set the parameters, we collected tweet streams from all target locations (178 countries) in a duration of 180 days from April 05, 2017 to October 01, 2017. The tweets from the stream were classified and the overall weights were measured each day for a duration of 180 days from April 05, 2017 to October 01, 2017. During this period, significant civil unrest events around the world were searched in the GlobalData on Events, Location, and Tone (GDELT) database (Leetaru & Schrodt, 2013). In this experiment, significant unrest events referred to those events which had international news coverage and significant negative impact on society. The standard weight (W^{std}) was set to the maximum weight that was found among all unrest events occurring day in all countries. The experiment found that the maximum of such weight was 6.96 and the standard weight was set as $W^{std} = 7.0$ by rounding the value. On the other hand, the baseline weight (W^{bl}) was set to the minimum weight that was found among all peace days in all countries. The experiment set found the minimum weight as 4.54 and the baseline weight was set as $W^{bl} = 4.5$. The experiment set the count threshold as $TC^{th} = 50$.

The experiment executed the proposed framework from November 2017 to June 2018 in 178 countries around the world. The Twitter data streams for the period from November 1, 2017 to November 25, 2017 were used to train SVM, while the data streams from November 26, 2017 to June 25, 2018 were used to evaluate performance. Similar to the effort by Korolov et al. (2016), the performance of the proposed framework was compared with the mainly used logistic regression based framework. The existing logistic regression-based method uses Gold Standard Report (GSR) event stream which are coded using the binary variable for predicting future events. To identify the true and false prediction, the predicted events of both frameworks were compared with the labelled events set, known as the GSR. The experiment used the GDELT

(Leetaru & Schrodt, 2013) dataset as GSR, with its tremendous amount of event records which were more than any other event datasets (Qiao et al., 2017). The dataset was downloaded using the GDELT provided event exporter service. According to the GDELT data format codebook (Gerner, Schrodt, Yilmaz, & Abu-Jabr, 2002), any data row (event) in the database with 'Event Root Code' 14 is decoded as an unrest event, 'NumArticles' is the total number of source documents containing the event and 'AvgTone' is the average tone range from -100 (extremely negative) to +100 (extremely positive) to express the total impact of an event. As for each country, we were interested to predict civil unrest events which had a significant negative impact on society, so the GDELT was further filtered with a minimum number of articles with 10 to reduce biases and the average tone with less than -5.00 to identify events which had significant negative impact.

To measure the performance of SVM classifier, we selected 10,500 tweets randomly from the testing data stream. The tweets were manually classified as 'informative' or 'uninformative' with the help of domain expert. The tweets were then classified using trained SVM and then compared with the manually labelled class. The comparisons generated four results; they were true positive (manual=informative, SVM=informative), true negative (manual=uninformative, SVM=uninformative), false positive (manual=uninformative, SVM=informative) and false negative (manual=informative, SVM=uninformative). Based on these class results, the performance SVM classifier is illustrated in Figure 6.

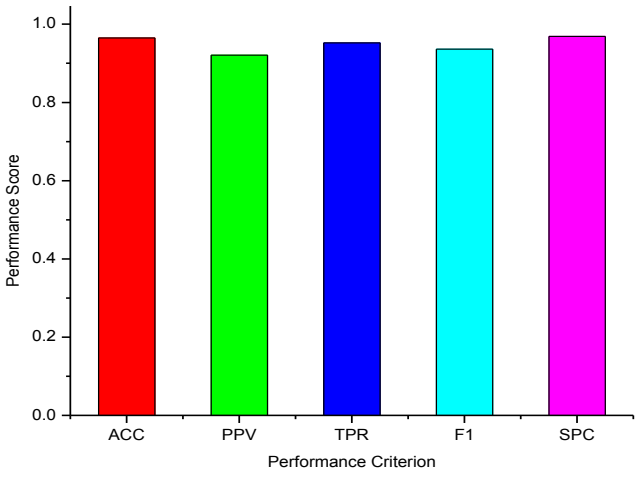


Figure 6. Performance of SVM classifier.

In Figure 6, the performance of SVM is described in terms of five parameters. They are accuracy (ACC), F1 score (F1), true negative rate or specificity (SPC), recall or true positive rate (TPR) and precision (PPV). The figure shows that the trained SVM classifier indicates more than 95% accuracy and recall. Precision and specificity are also nearly 95%. The overall F1 score is 0.936 which proves the excellent performance of SVM. However, accuracy and F1 score are below 1 as they fail to identify some truly informative tweets (false negative). This occurs when tweets contain abbreviations and informal words.

Muthiah et al. (2015) reported that the best prediction period for civil unrest using Twitter was 2.82 days. Kallus (2014) also predicted significant unrest which occurred in three days that followed. Similarly the proposed framework forecasted the occurrence of civil unrest in three days that followed (prediction period, F=1st, 2nd, and 3rd day) from December 03, 2017 to June 28, 2018 in a seven-day window (H=7) for four selected countries namely, Afghanistan, Argentina, Australia and Bangladesh. The predicted probabilities are presented in Figures 7(a-d) in a similar manner as presented by Azpeitia, Ochoa-Zezzatti, and Cavazos (2017).

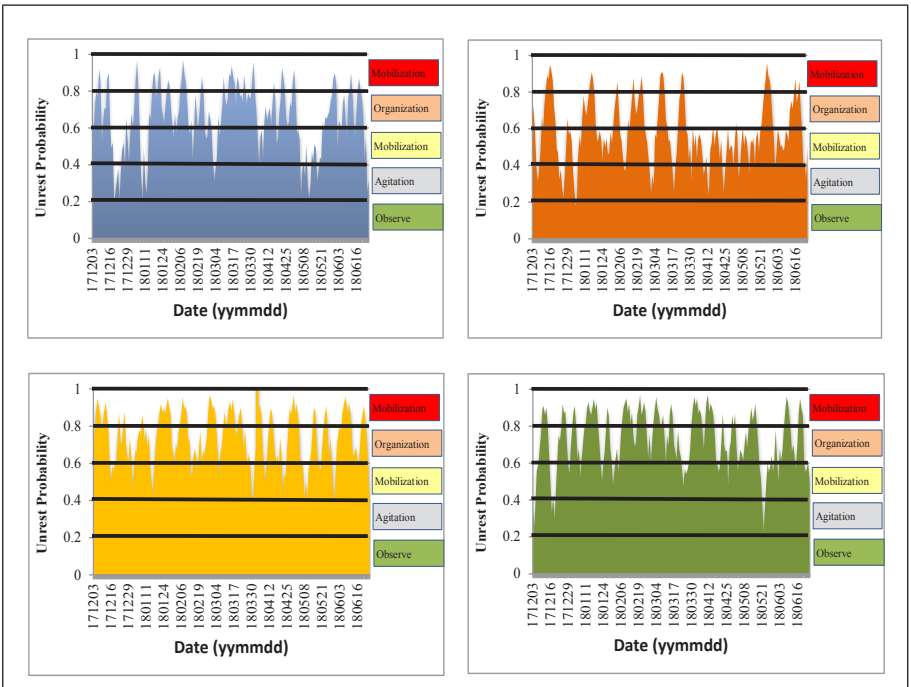


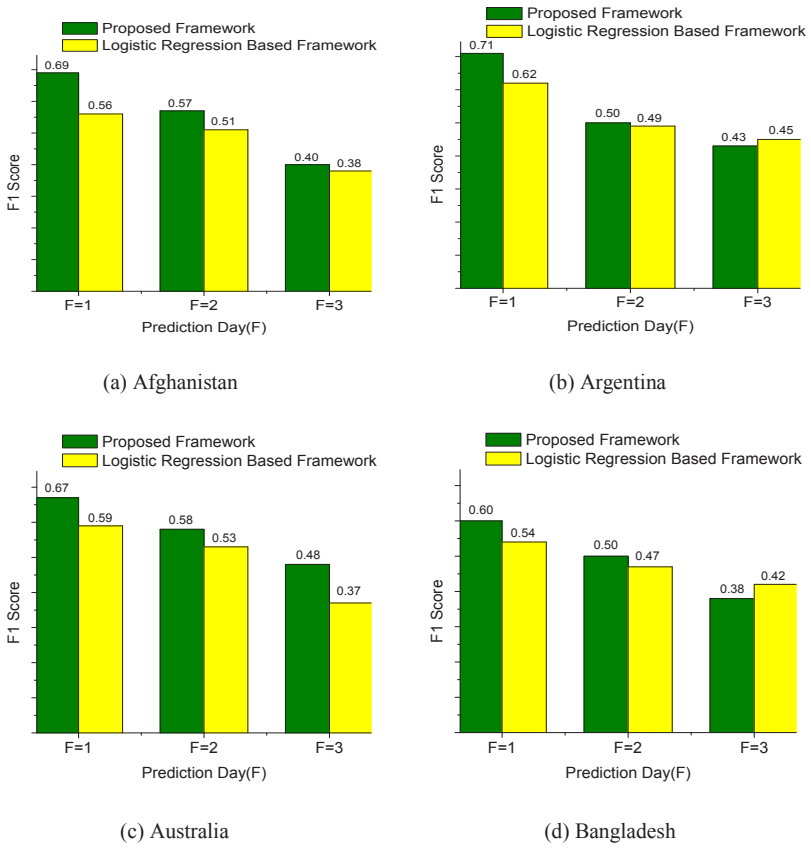
Figure 7. Predicted civil unrest probability by proposed framework for four countries.

Based on Figure 7, Bangladesh shows special behaviour where for most days it has insufficient tweet information, and thus uses diffused and historical weights for predicting future unrest probability. The unrest stages are, 'observe' and 'agitation' in a few number of days. The 'mobilization' stage is the most frequent stage found in all four countries. Among all the countries in Figure 7, Bangladesh is seen to be the country with the most potential for civil unrest occurrence. Most importantly, it can be seen that the probability for civil unrest increases and decreases sequentially on most days. For example, on 6th December 2017 the unrest stage was 'agitation' in Argentina and it turned into the 'mobilization' stage on 9th December and progressed to 'organization' stage on 12th December. Finally civil unrest broke out on 16th December in Argentina.

To measure the performance of any forecasting framework, it is necessary to determine how many instances are correctly identified and how many instances are missed or incorrectly identified. If for any country c, the true events set and predicted events sets for *l* prediction days are, $T_l^c = (t_{l,1}^c t_{l,2}^c .....t_{l,n}^c)$ and $P_l^c = (p_{l,1}^c p_{l,2}^c .....p_{l,n}^c)$ respectively where $t_{l,i}^c \in \{0,1\}$ for i = 1,2,3,....,n test days then true positive (TP) is the number of truly identified significant unrest event i.e. $t_{l,j}^c = p_{l,j}^c = 1$ and true negative (TN) is the case where the framework correctly identifies that no unrest events occur i.e. $t_{l,j}^c = p_{l,j}^c = 0$. False positive (FP) and false negative (FN) is defined as the number of incorrectly identified and missed unrest events. The performance of the framework is compared with logistic regression based prediction framework using its F1 score. F1 score considers both precision and recall to measure the accuracy of forecasting framework using Equation 14.

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{14}$$

Figures 8 (a-d) illustrate the F1 score values for three different prediction periods in four different countries. From Figures 8(a-d), the F1 scores are the highest and best for the proposed framework in all countries on the first (F=1) and second (F=2) prediction day. The scores are very close to each other for both frameworks in all the four countries on the third prediction day (F=3). The F1 scores of the proposed framework are better in Afghanistan and Australia than the existing logistic regression based framework while the results altered for the other two countries. Overall, the F1 score is nearly between 0.6 and 0.7 on the first prediction day, a 10% decrement on the second prediction day for both frameworks in the countries mentioned. The scores remain close to each other in both frameworks in all countries on the third prediction day.



(a) Afghanistan (b) Argentina

(c) Australia (d) Bangladesh

Figure 8. F1 scores for different prediction periods in different countries.

Thus, the performance of the proposed framework is better than the existing framework in terms of F1 scores and the best result is found on the first prediction day.

The performance of the proposed framework is further measured based on their balanced accuracy (BACC) that is defined as the mean of true positive rate (TPR) and true negative rate (TNR) (Qiao et al., 2017). TPR is the fraction of TP to positive instances and TNR is the fraction of TN to total negative instances. BACC is computed as Equation 15.

$$BACC = \frac{TPR + TNR}{2} \tag{15}$$

The BACC for three prediction periods in four different countries is presented in Figures 9(a-d).

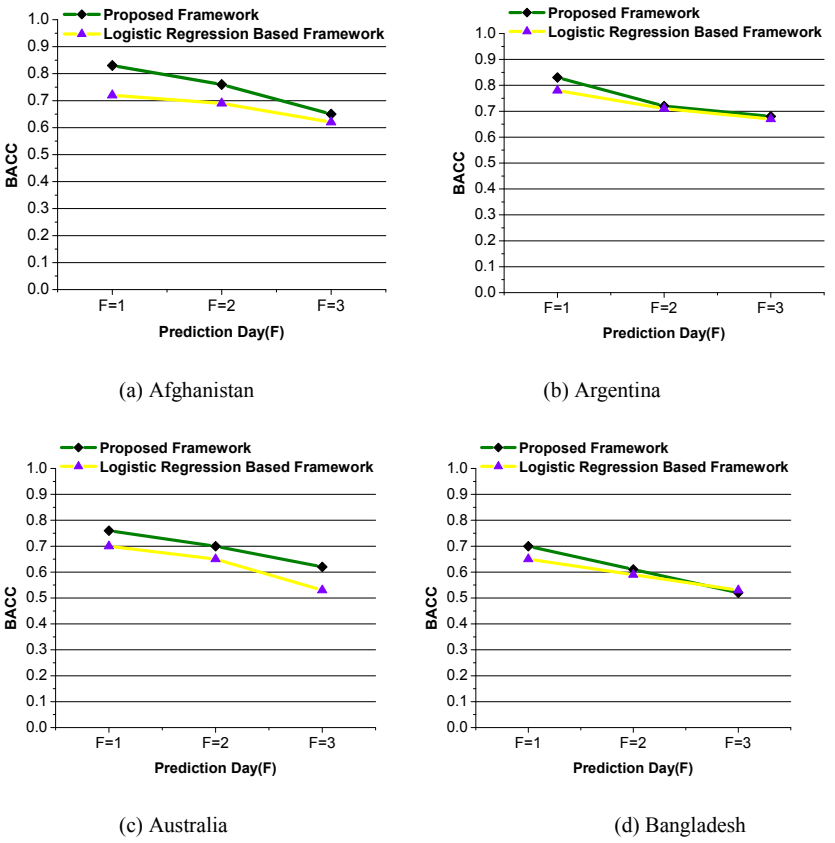


Figure 9. BACC for different prediction periods in different countries.

According to Figures 9(a-d), on the first prediction day (F=1), the proposed framework outperforms the logistic regression based framework in all countries in terms of BACC. On the second prediction day (F=2), the BACC values of the proposed framework are better than the existing framework in Afghanistan and Australia whereas they are nearly equal for both frameworks in Argentina and Bangladesh. But in the case of the third prediction day (F=3), BACC of the proposed framework is below the logistic regression based framework and for the other three countries, the proposed framework is still better than the existing one. It is also found that the BACC value in Bangladesh is lower than the other three countries. The derived tweet weight is hoped to be deviated by a small amount which contributed to this depreciation in performance.

To evaluate the falsely produced result, two metrics were used in this experiment; the FAR and FRR (Hernández, Ortiz, Andaverde, & Burlak,

2008). FAR is defined as the ratio of falsely accepted unrest events to the total number of unrest events occurring. This indicates the likelihood that an event may be falsely accepted and must be minimized in high-performing prediction framework. FAR is measured as Equation 16.

$$FAR = \frac{FP}{FP + TP} \tag{16}$$

The measured FAR for three prediction days (F=1, 2, 3) in four countries is shown in Figures 10(a-d).

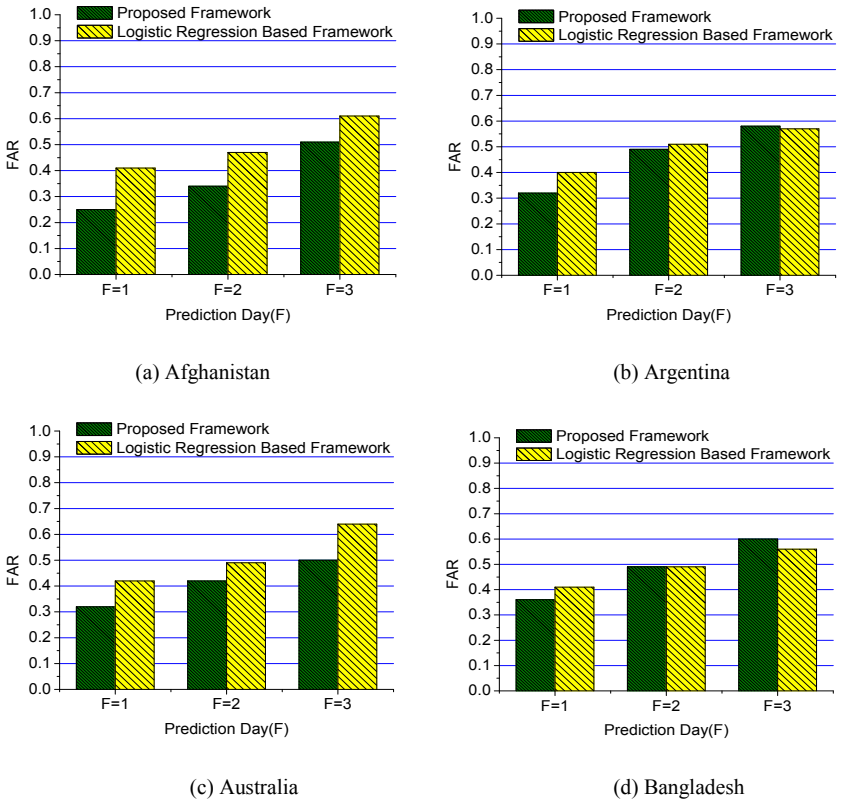


Figure 10. FAR for different prediction periods in different countries.

From Figures 10(a-d), on the first prediction day (F=1), the proposed framework show less FAR than the logistic regression based framework in all countries. On the second prediction day (F=2), the value of the framework is better than the existing framework in Afghanistan and Australia whereas

they are nearly equal for both frameworks in Argentina and Bangladesh. But in the case of the third prediction day (F=3), the proposed framework finds less falsely accepted events than other frameworks in Bangladesh and for the other three countries; the proposed framework is still better than the existing one. It is also found that the FAR value in Bangladesh is lower than the other three countries.

FRR is defined as the ratio of falsely rejected unrest events to the total number of non-events occurring. FRR describes the probability that a true event may be rejected as a non-event and measured as Equation 17.

$$FRR = \frac{FN}{FN + TN} \tag{17}$$

Where FN is False Negative and TN is True Negative.

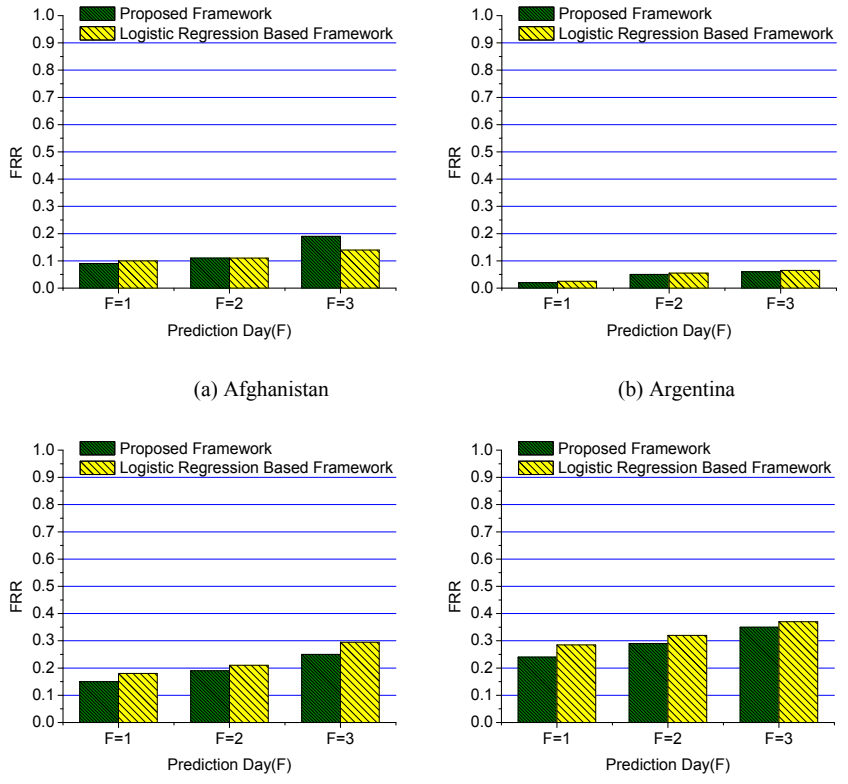


Figure 11. FRR for different prediction periods (F=1, 2, 3) in different countries.

Figures 11(a-d) illustrates that FRR follows approximately the same characteristics as FAR, Figures 10(a-d). The FRR is below in the proposed framework than the existing logistic regression framework in Australia and Bangladesh. It remains nearly unchanged for Argentina but in Afghanistan the proposed framework shows more falsely rejected cases on the third prediction day (F=3).

Matthews correlation coefficient (MCC) was first introduced by B.W. Matthews to assess the performance of a prediction framework (Matthews, 1975). MCC is measured using Equation 18.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \quad (18)$$

Figures 12(a-d), illustrates the MCC value for different prediction periods where on the first (F=1) and second (F=2) prediction day, the coefficient value is best for the proposed framework in all countries.

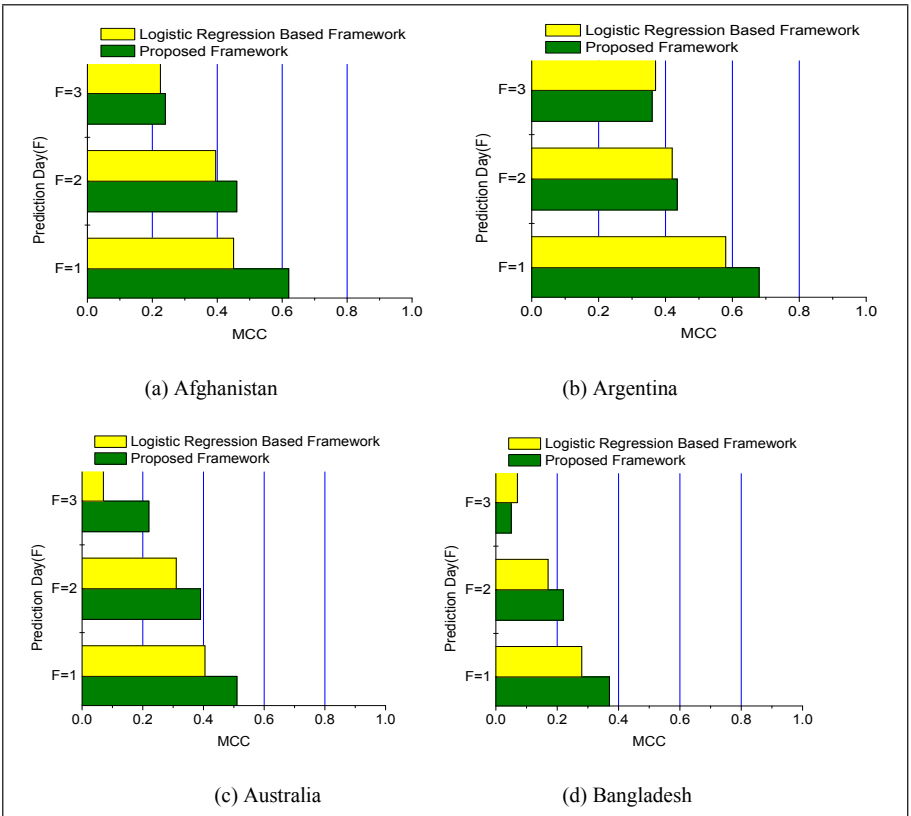


Figure 12. MCC for different prediction periods in different countries.

According to Figures 12(a-d), on the first prediction day (F=1), the proposed framework shows better MCC value than the logistic regression based framework in all the four countries. On the second prediction day (F=2), the MCC value of the proposed framework is more than the existing framework in Afghanistan, Australia and Bangladesh whereas they are nearly equal for both frameworks in Argentina. But in case of the third prediction day, the MCC of the proposed framework is below the existing logistic regression based framework and for the other three countries, the proposed framework is still better than the existing one. Similar to the F1 score and BACC, the experiment found that the MCC value in Bangladesh is lower than the other three countries.

In this experiment, the performance of the proposed framework is analyzed for three prediction days (F=1st day, F=2nd day and F=3rd day) based on three different historical time granularity (H=7 days, H=14 days and H=21 days) in terms of accuracy for determining the effective window size. Accuracy is the proportion of correctly identified instances (Visa, Ramsay, Ralescu, & Van Der Knaap, 2011) as Equation 19.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{19}$$

The measured ACCs for three prediction periods (F=1, 2, 3) in four different countries are represented in Figures 13(a-d). From Figures 13(a-d), it can be seen that on the first prediction day (F=1) the accuracy is highest for the window size of seven days (H=7) in all the four countries. By increasing the prediction day (F=2, 3), the accuracy decreases in all the four countries. It can also be seen from Figures 13(a-d) that for the window size of 14 days (H=14), the accuracies in all countries are lower than the window size of seven days (H=7) for all prediction days (F=1, 2, 3). When the window size is further increased to 21 days (H=21) days, for the first prediction day (F=1), the accuracy performance decreases for all countries.

The prediction accuracies is highest for the window size of seven days in all the following countries for all prediction periods except Afghanistan where the accuracy for the window size of 21 days crosses the accuracy for the window size of seven days. It is also illustrated in the figures that accuracies for the window size of 21 days are better than the window size of 14 days.

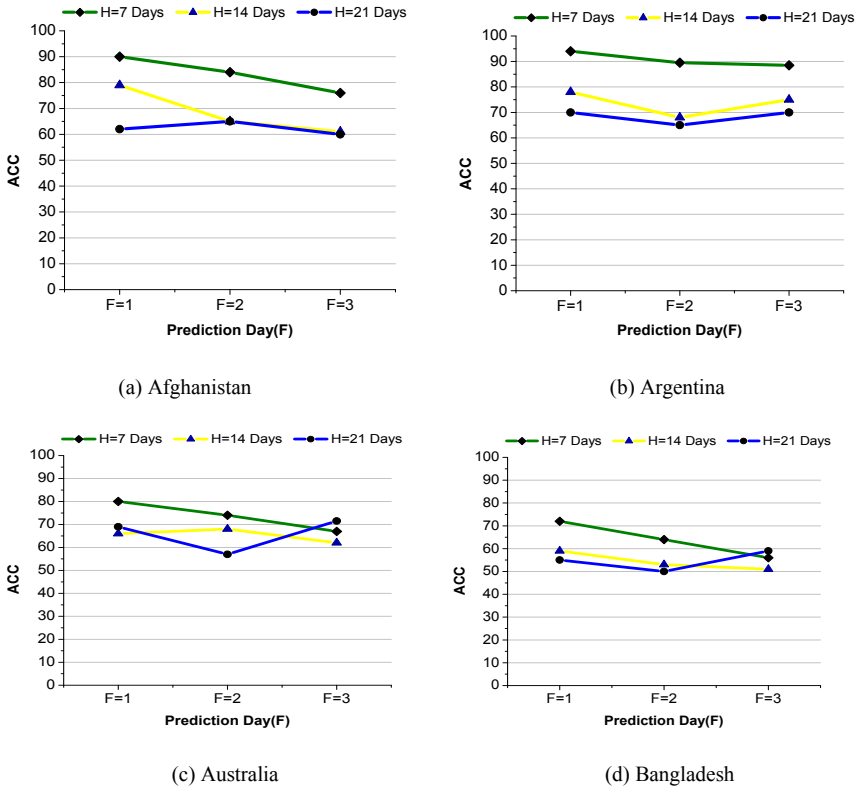


Figure 13. Accuracy for different prediction periods (F) and different window sizes (H).

CONCLUSION

In this study, a granularity framework for tweet stream processing has been proposed which enables us to predict the probability of civil unrest events occurring in any location. While almost all existing frameworks are designed simply based on term frequency, the proposed framework analyzes tweet stream at a more granularity level by weighting the term along with the frequency. The quantitative value of unrest related sentiments in a tweet is measured based on keywords and influencing words scaling. The recursive nature of calculating the overall weight of tweet stream in the proposed framework enables us to analyze the tweet in real time. This feature supports the proposed framework as an online analysis framework. The proposed framework also utilizes the diffusion property of civil unrest events and connected location

graph to handle locations where sufficient information related to unrest is not available. This feature of the framework circumvents the effort and cost of depending on alternative sources of required information.

To the best of our knowledge, this is the first quantitative framework to analyze tweet stream for predicting and forecasting the occurrence of civil unrest events in the future. The forecasting performance of the proposed framework outperforms the mainly used logistic regression-based framework on the first prediction day in all experimental locations. In addition, accuracy on the first prediction day is improved by about 10% if the proposed framework uses the tweet stream weight for the seven days window instead of 14 or 21 days. Though, the proposed framework successfully predicts the occurrence of significant civil unrest events at around 85% of cases with some exceptions on the first day of prediction, the accuracy decreases with the increasing prediction period. This fact leaves future research on improving performance. Besides, the proposed keyword dictionary contains only English words; even though, a lot of users use language specific keywords in their tweets which are not translatable, including language specific keywords which could improve forecasting accuracy. Furthermore, forecasting unrest periods, groups of participants in civil unrests and the economic losses incurred could also serve as directions for future research from the current study.

ACKNOWLEDGEMENT

This research work was supported by the Ministry of Higher Education, Malaysia through the Fundamental Research Grant Scheme (Grant no. RDU 170395 and RDU 190184). The author would like to acknowledge Universiti Malaysia Pahang for its partial support from another research grant (Grant no. RDU 170103).

REFERENCES

Azpeitia, D., Ochoa-Zezzatti, A., & Cavazos, J. (2017). Viral Analysis on Virtual Communities: A Comparative of Tweet Measurement Systems. In P. Melin, O. Castillo & J. Kacprzyk (Eds.), *Nature-Inspired Design of Hybrid Intelligent Systems* (pp. 801-808). Cham: Springer International Publishing.

Benkhelifa, E., Rowe, E., Kinmond, R., Adedugbe, O. A., & Welsh, T. (2014, August). *Exploiting social networks for the prediction of social and civil unrest: A cloud based framework.* Paper presented at the 2014

International Conference on the Future Internet of Things and Cloud (FiCloud), Barcelona, Spain.

Berestycki, H., Nadal, J.-P., & Rodriguez, N. (2015). A model of riots dynamics: Shocks, diffusion and thresholds. *arXiv preprint arXiv:1502.04725*.

Boonstra, T. W., Larsen, M. E., & Christensen, H. (2015). Mapping dynamic social networks in real life using participants' own smartphones. *Heliyon, 1*(3), e00037.

Cadena, J., Korkmaz, G., Kuhlman, C. J., Marathe, A., Ramakrishnan, N., & Vullikanti, A. (2015). Forecasting social unrest using activity cascades. *PloS one, 10*(6), e0128879.

Chen, F., & Neill, D. B. (2014, August). *Non-parametric scan statistics for event detection and forecasting in heterogeneous social media graphs.* Paper presented at the Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, USA.

Chiluwa, I. (2018). SMS & civil unrest *Encyclopedia of Information Science and Technology* (4th ed.). (pp. 6275-6285): IGI Global.

Chumwatana, T. (2018). Comment analysis for product and service satisfaction from Thai customers' review in social network. *Journal of Information and Communication Technology, 18*(2), 271-289.

Compton, R., Lee, C., Xu, J., Artieda-Moncada, L., Lu, T.-C., De Silva, L., & Macy, M. (2014). Using publicly visible social media to build detailed forecasts of civil unrest. *Security informatics, 3*(1), 4.

Dermisi, S. (2017). Social media, civil unrest and fallout for cities and hotels: European Real Estate Society (ERES).

Dilrukshi, I., De Zoysa, K., & Caldera, A. (2013, April). *Twitter news classification using SVM.* Paper presented at the 2013 8th International Conference on Computer Science & Education (ICCSE), Colombo, Sri Lanka.

El-Katiri, L., Fattouh, B., & Mallinson, R. G. (2014). *The Arab uprisings and MENA political instability: Implications for oil & gas markets.* Oxford Institute for Energy Studies, UK.

Ferrara, E. (2018). Measuring social spam and the effect of bots on information diffusion in social media *Complex Spreading Phenomena in Social Systems* (pp. 229-255). Place of publication?: Springer.

Filchenkov, A. A., Azarov, A. A., & Abramov, M. V. (2014, November). *What is more predictable in social media: Election outcome or protest action?* Paper presented at the Proceedings of the 2014 Conference on Electronic Governance and Open Society: Challenges in Eurasia, New York, USA.

Galla, D., & Burke, J. (2018, 2018//). *Predicting Social Unrest Using GDELT.* Paper presented at the Machine Learning and Data Mining in Pattern Recognition, Cham, Switzerland.

Gerner, D. J., Schrodt, P. A., Yilmaz, O., & Abu-Jabr, R. (2002). Conflict and mediation event observations (CAMEO): A new event data framework for the analysis of foreign policy interactions. *International Studies Association, New Orleans*.

Hasan, K. A., Sabuj, M. S., & Afrin, Z. (2015, December). *Opinion mining using naive bayes.* Paper presented at the 2015 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), Dhaka, Bangladesh.

Henslin, J. (2011). *Essentials of Sociology* (9th ed.). New York, USA: Pearson.

Hernández, J. A., Ortiz, A. O., Andaverde, J., & Burlak, G. (2008). *Biometrics in online assessments: A study case in high school students.* Paper presented at the 18th International Conference on Electronics, Communications and Computers, 2008 (CONIELECOMP 2008).

Hoegh, A., Leman, S., Saraf, P., & Ramakrishnan, N. (2015). Bayesian model fusion for forecasting civil unrest. *Technometrics, 57*(3), 332-340.

Hossny, A. H., & Mitchell, L. (2018). *Event detection in Twitter: A keyword volume approach.* Paper presented at the 2018 IEEE International Conference on Data Mining Workshops (ICDMW).

Huang, H., Boranbay-Akan, S., & Huang, L. (2016). Media, protest diffusion, and authoritarian resilience. *Political Science Research and Methods*, 1-20.

Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., & Meier, P. (2013, May). *Extracting information nuggets from disaster-related messages in social media.* Paper presented at the Proceedings of the 10th International Conference on Information Systems for Crisis Response & Management, Baden- Baden, Germany.

Kang, W., Chen, J., Li, J., Liu, J., Liu, L., Osborne, G., . . . Neale, G. (2017). *Carbon: Forecasting Civil Unrest Events by Monitoring News and Social Media*, Cham.

Korkmaz, G., Cadena, J., Kuhlman, C. J., Marathe, A., Vullikanti, A., & Ramakrishnan, N. (2015, August). *Combining heterogeneous data sources for civil unrest forecasting.* Paper presented at the Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, Paris, France.

Korkmaz, G., Cadena, J., Kuhlman, C. J., Marathe, A., Vullikanti, A., & Ramakrishnan, N. (2016). Multi-source models for civil unrest forecasting. *Social network analysis and mining, 6*(1), 50.

Korolov, R., Lu, D., Wang, J., Zhou, G., Bonial, C., Voss, C., . . . Ji, H. (2016). *On predicting social unrest using social media.* Paper presented at the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

Kumar, S., Morstatter, F., & Liu, H. (2014). *Twitter data analytics*. New York, USA: Springer.

Lang, J. C., & De Sterck, H. (2014). The Arab Spring: A simple compartmental model for the dynamics of a revolution. *Mathematical Social Sciences, 69*, 12-21.

Leetaru, K., & Schrodt, P. A. (2013). *Gdelt: Global data on events, location, and tone, 1979–2012.* Paper presented at the ISA Annual Convention.

Liang, Y., & Kee, K. F. (2018). Developing and validating the ABC framework of information diffusion on social media. *New Media & Society, 20*(1), 272-292.

Manrique, P., Qi, H., Morgenstern, A., Velasquez, N., Lu, T.-C., & Johnson, N. (2013, June). *Context matters: improving the uses of big data for forecasting civil unrest: Emerging phenomena and big data.* Paper presented at the 2013 IEEE International Conference on Intelligence and Security Informatics (ISI), Seattle, Washington, USA.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure, 405*(2), 442-451.

Mohammad, S. M., Kiritchenko, S., & Zhu, X. (2013, June). *NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets.* Paper presented at the Proceedings of the Seventh International Workshop on Semantic Evaluation Exercises, Atlanta, Georgia, USA.

Muthiah, S., Huang, B., Arredondo, J., Mares, D., Getoor, L., Katz, G., & Ramakrishnan, N. (2015, January). *Planned Protest Modeling in News and Social Media.* Paper presented at the Proceedings of the Twenty-Seventh Conference on Innovative Applications of Artificial Intelligence Austin, Texas, USA.

Nielsen, F. (2011). *Afinn, informatics and mathematical modelling, Technical University of Denmark.* Technical University of Denmark, Denmark.

O'Leary, D. E. (2015). Twitter mining for discovery, prediction and causality: Applications and methodologies. *Intelligent Systems in Accounting, Finance and Management, 22*(3), 227-247.

Ochoa-Zezatti., C. A. (2016). *Identifying consumption patterns in Twitter using text mining to classify trends in shopping*. Paper presented at the CICLing 2016, Konya, Turkey.

Oh, O., Eom, C., & Rao, H. R. (2015). Research note—Role of social media in social change: An analysis of collective sense making during the 2011 Egypt revolution. *Information Systems Research, 26*(1), 210-223.

Olanrewaju, A.-S. T., & Ahmad, R. (2018). Examining the information dissemination process on social media during the Malaysia 2014 floods using social network analysis (SNA). *Journal of Information and Communication Technology, 17*(1), 141-166.

Parlar, T., Özel, S. A., & Song, F. (2018). *Interactions Between Term Weighting and Feature Selection Methods on the Sentiment Analysis of Turkish Reviews*, Cham.

Passarelli, F., & Tabellini, G. (2017). Emotions and political unrest. *Journal of Political Economy, 125*(3), 903-946.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research, 12*(Oct), 2825-2830.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program, 14*(3), 130-137.

Priambodo, B., & Ahmad, A. (2018). Traffic flow prediction model based on neighbouring roads using neural network and multiple regression. *Journal of Information and Communication Technology, 17*(4), 513-535.

Qiao, F., Li, P., Zhang, X., Ding, Z., Cheng, J., & Wang, H. (2017). Predicting social unrest events with hidden Markov models using GDELT. *Discrete Dynamics in Nature and Society, 2017*.

Qiao, F., & Wang, H. (2015). *Computational approach to detecting and predicting occupy protest events.* Paper presented at the 2015 International Conference on Identification, Information, and Knowledge in the Internet of Things (IIKI).

Raina, P. (2013). *Sentiment analysis in news articles using sentic computing.* Paper presented at the 2013 IEEE 13th International Conference on Data Mining Workshops (ICDMW).

Ramakrishnan, N., Butler, P., Muthiah, S., Self, N., Khandpur, R., Saraf, P., . . . Korkmaz, G. (2014, August). *'Beating the news' with EMBERS: Forecasting civil unrest using open source indicators.* Paper presented at the Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, USA.

Ranganath, S., Morstatter, F., Hu, X., Tang, J., Wang, S., & Liu, H. (2016, February). *Predicting Online Protest Participation of Social Media Users*. Paper presented at the Proceedings of the 13th AAAI Conference on Artificial Intelligence, Phoenix, Arizona, USA.

Singh, S., & Pal, R. (2018, 2018//). *Characterizing and Detecting Social Outrage on Twitter: Patel Reservation in Gujarat.* Paper presented at the Data Science and Analytics, Singapore.

Valenzuela, S. (2013). Unpacking the use of social media for protest behavior: The roles of information, opinion expression, and activism. *American Behavioral Scientist, 57*(7), 920-942.

Van Dyke, N., & Amos, B. (2017). Social movement coalitions: Formation, longevity, and success. *Sociology Compass, 11*(7), e12489.

van Noord, R., Kunneman, F. A., & van den Bosch, A. (2016, November). *Predicting civil unrest by categorizing Dutch Twitter events.* Paper presented at the Benelux Conference on Artificial Intelligence, Amsterdam, The Netherlands.

Visa, S., Ramsay, B., Ralescu, A. L., & Van Der Knaap, E. (2011, April). *Confusion Matrix-based Feature Selection.* Paper presented at the 22nd Midwest Conference on Artificial Intelligence and Cognitive Science, Cincinnati, OH, USA. .

Watts, D. J. (2013). Computational social science: Exciting progress and future directions. *The Bridge on Frontiers of Engineering, 43*(4), 5-10.

Wu, C., & Gerber, M. S. (2018). Forecasting Civil Unrest Using Social Media and Protest Participation Theory. *IEEE Transactions on Computational Social Systems, 5*(1), 82-94.

Xu, J., Lu, T.-C., Compton, R., & Allen, D. (2014, April). *Civil unrest prediction: A tumblr-based exploration.* Paper presented at the International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, Washington, DC, USA. .

Zhao, L., Sun, Q., Ye, J., Chen, F., Lu, C.-T., & Ramakrishnan, N. (2015). *Multi-task learning for spatio-temporal event forecasting.* Paper presented at the Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Zhao, L., Sun, Q., Ye, J., Chen, F., Lu, C.-T., & Ramakrishnan, N. (2017). Feature constrained multi-task learning models for spatiotemporal event forecasting. *IEEE Transactions on Knowledge and Data Engineering, 29*(5), 1059-1072.

Zou, H., & Jin, Z. (2018, July). *Comparative Study of Big Data Classification Algorithm Based on SVM.* Paper presented at the 2018 Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC).