Teargas, Water Cannons and Twitter: A case study on detecting protest repression events in Turkey 2013

Fatma Elsafoury
Computing Science Department
University of the West of Scotland
fatma.elsafoury@uws.ac.uk

Abstract

Since the Arab spring in 2011, protests have been spreading around the world for different reasons, often these protests are faced with violent repression. Studying protest repression requires appropriate datasets. Existing datasets like GDELT focus mainly on events reported in news media. However, news media reports have issues including censorship and coverage bias. Recently, social scientists have started using Machine Learning (ML) to detect political events, but it is costly and time consuming to hand label data for training ML models. This paper proposes using ML and crowdsourcing to detect protest repression events from Twitter. Our case study is the Turkish Gezi Park protest in 2013. Our results show that Twitter is a reliable source reflecting events happening on the ground as soon as they happen. Moreover, training conventional ML models on crowdsourced labelled data gave good results with an AUC score of 0.896 to detect protest events and 0.8189 to detect repression events.

1 Introduction

In Venezuela 2017, 50 protesters were killed and 2700 were arrested by the police forces and pro-government militias [Gen17]. Similar events happened during the Arab Spring in 2011 and the Ukrainian Euromaidan protests in 2013. Ortz shows that between 2006 and 2013, 843 protests took place worldwide and the majority of them were faced with repression from state agents [OBBC13]. This type of violence against protesters is called *protest repression*. Protest repression is defined as “Any action by another group which raises the contender’s cost of collective action” [Til78]. According to literature, there are different types of protest repression events based on a three key dimensional typology [Ear03]. In this paper, we are interested in automatically detecting observable, coercive actions carried out by state agents against protesters. For social scientists to study protest repressions, they need two important tools: 1. A dataset containing events of protest repression. 2. A ML model that automatically detect the occurrence of protest repression events as quickly as possible. To build these tools, we need to solve two challenges. The first challenge is choosing a good data source to collect the protest repression events. Currently, protest repression events are collected as part of political conflict datasets

In: R. Campos, A. Jorge, A. Jatowt, S. Bhatia (eds.): Proceedings of the Text2Story’20 Workshop, Lisbon, Portugal, 14-April-2020, published at http://ceur-ws.org

like the Global Database of Events, Language, and Tone (GDELT), the Integrated Crisis Early Warning System (ICEWS) and the Social, Political and Economic Event Database Project (SPEED). These datasets collect data from traditional news reports. [DB02] and [ESM03] argue that depending on traditional news stories results in problems with coverage bias, accuracy issues and censorship. The second challenge is to hand label a training dataset to train the ML model as it is costly and time consuming.

In this paper, we investigate addressing those challenges by using Twitter as a data source and by using corwdsourcing to avoid hand-labeling a training dataset for the ML model. We use tweets sent during the Turkish Gezi Park protest in 2013. The contributions of this paper are:

1. Investigating the performance of crowdsouring workers when identifying the subjective topic of repression in a collection of tweets sent during the Gezi Park protest.

2. Demonstrating that Twitter is a reliable and timely source for detecting protest repression events during the Turkish Gezi Park protest in 2013 using ML models.

2 Related Work

There are already political conflict datasets that contain protest and protest repression events among others like the Uppsala Conflict Data Program (UCDP) and the Social Conflict Analysis Dataset (SCAD). These datasets collect events from news media reporters like BBC, Reuters, Associated Press (AP) and Agency French Press (AFP) [SHH+12, Gel18]. [ESM03] uses daily editions of the New York Times as a source to build their dataset collection of 1905 protest repression events and [Bis15] uses independent newspapers and recorded TV interviews with some of the participants in the protests in Egypt during the period of 2007-2011. On the other hand, social media possesses characteristics like scale and diversity that qualifies it to be a good source for event detection and overcoming the limitations of news reports [ZP19]. Twitter is a dynamic social media platform where people share news and events [LNSL17]. There exists a body of research on using Twitter to detect disasters like earthquakes [SOM10, SOM12, EBG12] and social events [RCE11, APM+13, SHM09, ASG13]. Social media has been widely used during the Arab Spring, the Occupy movement and the Euromedian protests in Ukraine [SVC17, PB12, Ani16]. By studying the role that Twitter plays during protests compared to other social media platforms like Facebook. [EMHMM+13], analyzed a dataset of 30,296 tweets collected during the protests against the G20 meetings held in Pittsburgh in 2009. The protests were faced with arresting of over a 100 participants, smoke canisters and rubber bullets. Earl and coauthors found that Twitter was used during the protest more than other social media platforms like Facebook which might be used prior to the protest to spread information and collect people while Twitter was used to report information during the protest about police activity and protest repression [EMHMM+13]. The same findings are supported by Poell and Borra in their study of using Twitter, YouTube and Flicker as alternative news sources during the same G20 protest. They found that 57% of top retweets about the G20 reported police activity. They also argue that Twitter is the most promising social media platform as an alternative news media [PB12]. Even though sometimes authoritarian regimes block or ban social media like Egypt during the 2011 revolution, people managed to find ways around the blockage using VPN networks to still connect to he internet and spread their message to the world [SA+11]. To the best of our knowledge, we are the first to use Twitter to detect repression events during protests.

3 Case Study: Gezi Park Protest in Turkey 2013

In summer 2013, protests emerged in Istanbul against the removal of the Gezi Park and the building of a shopping mall instead. The protests started on the 31st of May and lasted for a month [OV17, BW15]. In this paper, we used tweets collected during the Gezi park protest from 31/05/2013 to 30/06/2013 by researchers at the Social Media and political participation lab[1] at the New York university [TNM+16]. They collected 1,290,451 English tweets and we randomly selected 6693 tweets to be labelled using crowdsourcing for training ML models in our study. We will first look into building a labelled training dataset using crowdsourcing in the next section, before using it to train ML model to detect protest repression events in section 5.

4 Building Training Dataset Using Crowdsourcing

Crowdsourcing has been used in many studies to obtain labels for unlabelled data to train ML models because it is an easy, cheap and fast way of labelling data [SOJN08]. It gives access to larger and more diverse workers

[1] https://wp.nyu.edu/smapp/

which according to [SBDS14] leads to less biased data and more labelled data for the same cost of hiring smaller number of experts to do the same task with similar results [CBD10, MBG$^+$10, HS10, FAC11, WHK13]. There are some known challenges when using crowdsourcing. In our case, we were faced with two challenges. The workers' integrity (We need to make sure that the workers understand the task and choose the right answers to the questions) and the subjectivity of violence (disagreement on what is considered violence). We chose the following measures to address these two challenges: the worker's trust score and the agreement score between the workers. Trust score is the performance of each worker on "Test questions" which are a set of 116 tweets that we hand labelled and compared our labels to the worker's labels to evaluate his performance. Figure-Eight is a crowdsourcing platform that combines these two measures, trust score and agreement score, together in one accuracy measure. This is why we used Figure-Eight to host our crowdsourcing task. In the next sections, we describe our crowdsourcing task design and the the outcomes of the task.

4.1 Crowdsourcing Task Design

We designed the task to ask two questions about each tweets in our dataset: 1. *Protest question*: Is this tweet related to the Gezi Park protest in Turkey 2013? 2. *Violence question*: Does this tweet report/discuss a violent incident? We hired three workers to answer the two questions for each tweet. They were given two options: "Yes" and "No". We limited the task to only workers from Turkey and with medium level of experience (this level is determined based on their performance on previous tasks on Figure-Eight). We used the 116 test questions to evaluate the performance of the workers on the task before starting labelling the 6693 tweets in our dataset and we removed workers whose trust score was less than 0.75. We ended up with 835 trusted workers. This task costed us less than 1000 pounds.

4.2 Crowdsourcing Task Results

After the trusted workers finished labelling the 6693 tweets, we got the following results: for the protest question, we got low inter-agreement Krippendorff-alpha score of 0.428 between the three workers. This might be because some workers do not read the full tweet and just read hashtags which might be protest related while the actual content of the tweet is not. For the violence question, The Krippendorff-alpha score between the three workers is 0.64. This score is higher because the number of tweets that report violence is very low and most of the agreement was on the tweets that do not report violence. Some tweets were also not clear if they report violence or not. For example, in this tweet "RT @cey-lanozbudak:152.000 new flowers, 30 new trees were planted to #GeziPark #Taksim after it was cleaned of the #protestors", it is not clear if Taksim square was cleaned of protests with violence or not. This is why this tweet received disagreement on the violence question. Similarly, most of the disagreements happened because of clarity issues with the message and the subjectivity of violence. For that reason, after the task was finished on Figure-Eight, to aggregate all the answers from the three workers for each tweet, we used only the tweets that received the same answer from all three workers. The crowdsourced data is available here[2]. This resulted in two datasets: 1. **Protest dataset** containing 3860 tweets with fully agreed answers for the protest question with 1518 (39.3%) tweets received a"Yes" to the protest question (protest-positive) and 2342 (60.67%) tweets that received a"No" to the protest question (protest-negative). 2. **violence dataset** containing 5247 tweets with fully agreed answers for the violence questions with 323 (6.15%) tweets received a "Yes" to the violence question (violence-positive) and 4925 (93.84%) tweets received a"No" to the violence question (violence-negative). It is clear that the violence dataset is hugely imbalanced with only 6% positive examples. There are two methods like oversampling the minor class or under-sampling the major class to train the model on balanced dataset. But because in real life the number of positive examples is smaller than the negative examples, we decided not to employ any of these method. Figure 1 shows a visualization of the two datasets using t-SNE which projects high dimensional data like text into two dimensions [MH08].

4.3 Summary

In this section, we discussed using crowdsourcing for labelling training data from Twitter. We found a low agreement between the workers on the protest and the violence questions. The disagreement comes from the lack of clarity of some tweets and the subjective nature of violence. Our investigation suggests that hiring more crowd workers would result in higher agreement score. Also in the case of low inter-agreement score, it is better to use only tweets that received full agreement even if the number of training dataset is lower but we know that

[2] `https://github.com/efatmae/crowdflower_basic_analytics`

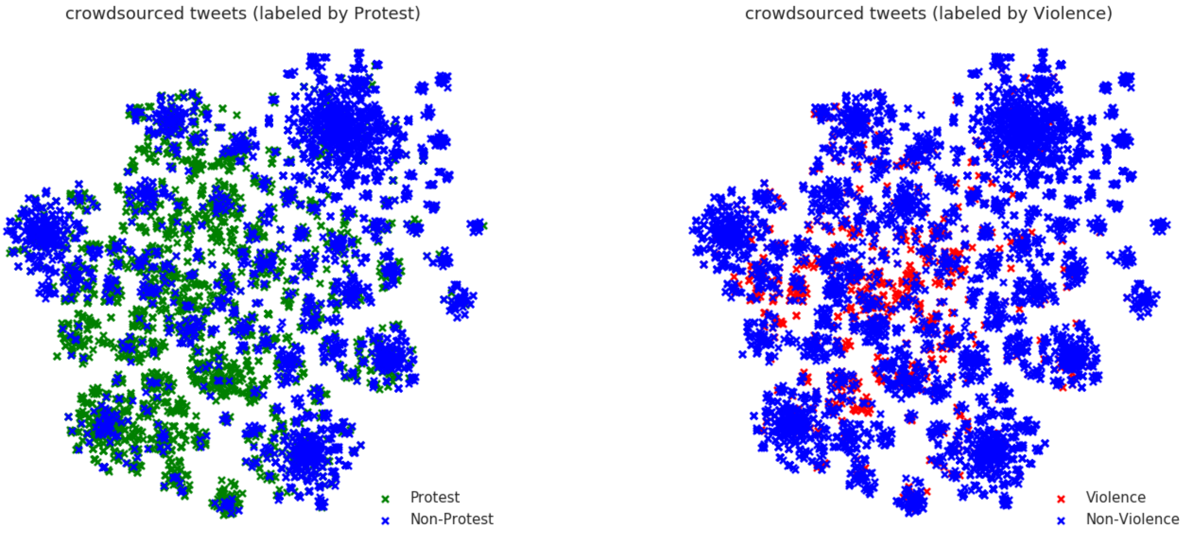


Figure 1: Words features in tweets collection protest dataset (left) and violence dataset (right)

the model's training will be better. Next, we use the gathered training dataset to build a ML model for detecting protest violence events form Twitter for the Gezi park protest.

5 Protest Repression Event Detection

After building a training dataset labelled by the crowdsourcing task. We are ready to develop a ML model to detect protest repression events by training two ML models: one to detect protest related tweets (Protest event detection model) and another to detect violent events (violence event detection model).

5.1 Preprocessing

We cleaned the tweets from user mentions, http addresses, hashtags, digits, stop words and words that are less than two characters long. We then removed duplicated tweets and tweets that are less than three words long. After the processing step, we ended up with 3666 tweets in the protest dataset with 1497 (49%) protest-positive tweets and 2169 (51%) protest-negative tweets. For the violence dataset, we have 4975 tweets with 322(6%) violence-positive tweets and 4653 (94%) violence-negative tweets.

5.2 Models Training

For each event detection task, we compared two models that proved to be effective in text classification tasks: Support Vector Machine (SVM)and Multinomial Naive Bayes (MNB) [Sar16, Joa02]. We used the protest dataset to train the protest detection model and the violence dataset to train the violence detection model.We randomly split the dataset into 2 equal sizes: a training dataset and a test dataset using the train-test-split[3] function in the sklearn package. We train the different models on a held-out fold in 10 folds cross-validation on the training set. We used SVM [4] function and MNB [5] implementation using Python package Sklearn package. We ran a grid search to choose the best parameters for each model that fit the data for the two tasks. For the **protest event detection task**, we found that linear SVM model with margin value $C = 1$ and MNB with smoothing parameter $\alpha = 2$ and uniform class prior gave the best results. For the **violence event detection task**, we chose a linear SVM with margin values $C = 10$ and MNB with smoothing parameter $\alpha = 0.1$ and uniform

[3] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html
[4] https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html
[5] https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html

Table 1: The AUC scores of SVM (kernel = linear, $C = 1$) with TF-IDF and MNB ($\alpha = 2$ and uniform class prior) with WC on the protest test set.

Feature	linear SVM ($C = 1$)	MNB ($\alpha = 2$)
WC	0.895	**0.856**
TF-IDF	**0.896**	0.846
WE	0.872	-

Table 2: The AUC scores of SVM (kernel = linear, $C = 10$) with TF-IDF and MNB ($\alpha = 0.1$ and uniform class prior) with TF-IDF on the test set of the violence dataset.

Feature	linear SVM ($C = 10$)	MNB ($\alpha = 0.1$)
Count	0.8018	0.816
TF-IDF	**0.8127**	**0.8189**
WE	0.7689	-

class prior. For feature extraction, we used Word Count (WC), TF-IDF and Word Embedding (WE) using the word2vector model with window size = 3 and vector dimension size = 800. We chose those values based on a grid search to choose between different values suggested by [YMO18] on a similar task. We fed each of the three features to the SVM and the MNB models except for WE which could contain negative weights and is not compatible with MNB.

Protest Event Detection Model Table 1 displays the AUC scores of SVM and MNB using each of the feature extraction models on the test set of the protest dataset. The table shows that SVM with TF-IDF outperforms MNB with WC features. To statistically compare the performance of SVM and MNB, we used McNemar test on the AUC score of each model. The test result is $P-value < 0.01$ which means, SVM with TF-IDF is significantly better.

Violence Event Detection Model Table 2 displays the AUC scores of SVM and MNB using each feature extraction on the test set of the violence dataset. The table shows that MNB with TF-IDF outperforms SVM with TF-IDF. We also used McNemar test to statistically compare the performance of the two models. The test result is $P - value < 0.16$ which means that MNB with TF-IDF is not significantly better than SVM with TF-IDF features.

After comparing the results of the different models and the different feature extraction models, we decided to choose a linear SVM ($C = 1$) with TF-IDF for the protest event detection model and MNB ($\alpha = 0.1$ and uniform class prior) with TF-IDF for the violence event detection model. We applied these models to the remaining 1,283,758 unlabelled tweets in our dataset. We used the same preprocessing steps we used before and this resulting in 854,313 tweets. When we applied the protest event detection model to the data, the model classified 36% of the tweet as protest-positive and 64% as protest-negative. The violence event detection model classified 15% of the tweets as violence-positive and 85% as violence-negative. 33% of the protest-positive tweets are violence-positive (protest repression tweets) and 67% as violence-negative. On the other hand, only 5% of the protest-negative tweets are violence-positive which are tweets that describe violence but not related to the protest.

5.3 Tweets Timeline

During the Gezi park protest in 2013, there were three days: 31/05/2013, 11/06/2013 and 15/06/2013 when the police was the most violent towards the protesters [TNM+16]. When we plot the number of tweets sent between 31/05 and 30/06 as in Figure 2, we see a spike on the most violent days. The spike happens in "all tweets", "protest tweets" and "repression tweet". There are other cases when the spikes of the three groups do not match. For example, on the 13th of June, There was a spike in "all tweet" and "protest tweets" but not in the "repression tweets" showing that people tweeted about the protest but not about protest repression events. On the 16th of June, the "all tweets" count spiked while the "protest tweets" and the "repression tweets" stayed the same. We speculated that this is due to people using protest hashtags like #gezipark in spam tweets. When we extracted the most frequent 27 words in the tweet, the resulted words like "ga", "tear", "attack", "water", "riot", "clash", "cannon", "violenc", "fire", "peac", "forc", "brutal", "report", "injur", "street", "demonstr", "right", "kill", "bomb", "tearga", "stop", "violent", "continu", "hotel", "chemic", "arrest", show the type of violence that was used during the protest. The words are stemmed (suffixes from words like es, s, ed , ing) that

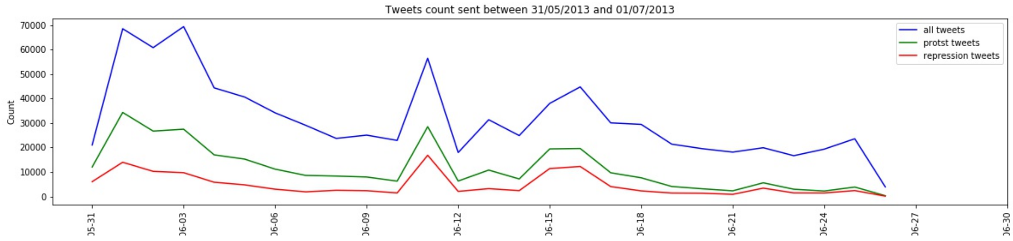
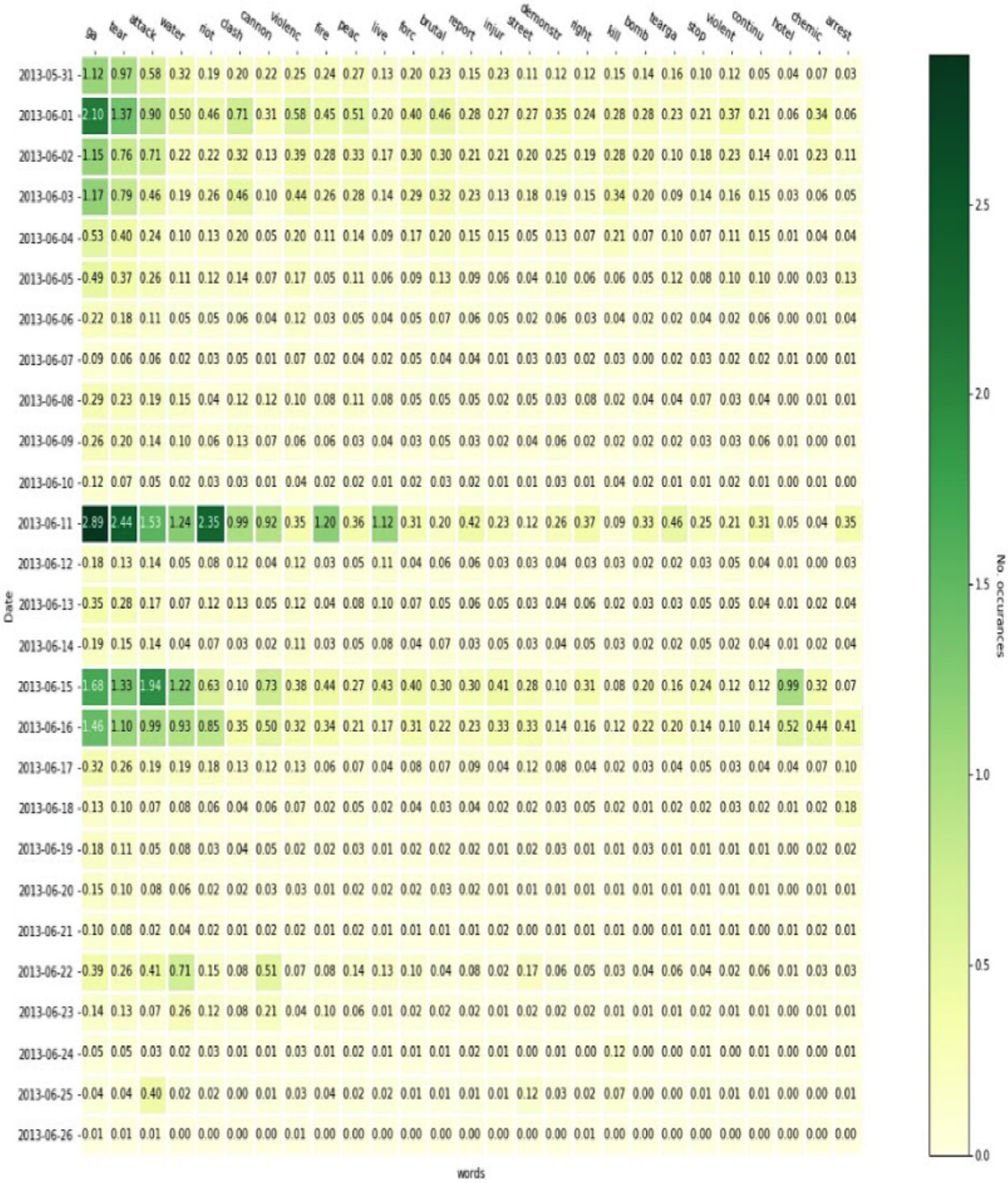


Figure 2: Timeline of the number of tweets, protest tweets and protest repression tweets sent during the Gezi protest period from 31/05/2013 to 27/06/2013



	ga	tear	attack	water	riot	clash	cannon	violenc	fire	peac	live	forc	brutal	report	injur	street	demonstr	fight	kill	bomb	tearga	stop	violent	continu	hotel	chemic	arrest
2013-05-31	1.12	0.97	0.58	0.32	0.19	0.20	0.22	0.25	0.24	0.27	0.13	0.20	0.23	0.15	0.23	0.11	0.12	0.12	0.15	0.14	0.16	0.10	0.12	0.05	0.04	0.07	0.03
2013-06-01	2.10	1.37	0.90	0.50	0.46	0.71	0.31	0.58	0.45	0.51	0.20	0.40	0.46	0.28	0.27	0.27	0.35	0.24	0.28	0.28	0.23	0.21	0.37	0.21	0.06	0.34	0.06
2013-06-02	1.15	0.76	0.71	0.22	0.22	0.32	0.13	0.39	0.28	0.33	0.17	0.30	0.30	0.21	0.21	0.20	0.25	0.19	0.28	0.20	0.10	0.18	0.23	0.14	0.01	0.23	0.11
2013-06-03	1.17	0.79	0.46	0.19	0.26	0.46	0.10	0.44	0.26	0.28	0.14	0.29	0.32	0.23	0.13	0.18	0.19	0.15	0.34	0.20	0.09	0.14	0.16	0.15	0.03	0.06	0.05
2013-06-04	0.53	0.40	0.24	0.10	0.13	0.20	0.05	0.20	0.11	0.14	0.09	0.17	0.20	0.15	0.15	0.05	0.13	0.07	0.21	0.07	0.10	0.07	0.11	0.15	0.01	0.04	0.04
2013-06-05	0.49	0.37	0.26	0.11	0.12	0.14	0.07	0.17	0.05	0.11	0.06	0.09	0.13	0.09	0.06	0.04	0.10	0.06	0.06	0.05	0.12	0.08	0.10	0.10	0.00	0.03	0.13
2013-06-06	0.22	0.18	0.11	0.05	0.05	0.06	0.04	0.12	0.03	0.05	0.04	0.05	0.07	0.06	0.05	0.02	0.06	0.03	0.04	0.02	0.02	0.04	0.02	0.06	0.00	0.01	0.04
2013-06-07	0.09	0.06	0.06	0.02	0.03	0.05	0.01	0.07	0.02	0.04	0.02	0.05	0.04	0.04	0.01	0.03	0.03	0.02	0.03	0.00	0.02	0.03	0.02	0.02	0.01	0.00	0.01
2013-06-08	0.29	0.23	0.19	0.15	0.04	0.12	0.12	0.10	0.08	0.11	0.08	0.05	0.05	0.05	0.02	0.05	0.03	0.08	0.02	0.04	0.04	0.07	0.03	0.04	0.00	0.01	0.01
2013-06-09	0.26	0.20	0.14	0.10	0.06	0.13	0.07	0.06	0.06	0.03	0.04	0.03	0.05	0.03	0.02	0.04	0.06	0.02	0.02	0.02	0.02	0.03	0.03	0.06	0.01	0.00	0.01
2013-06-10	0.12	0.07	0.05	0.02	0.03	0.03	0.01	0.04	0.02	0.02	0.01	0.02	0.03	0.02	0.01	0.01	0.03	0.01	0.04	0.02	0.01	0.01	0.02	0.01	0.00	0.01	0.00
2013-06-11	2.89	2.44	1.53	1.24	2.35	0.99	0.92	0.35	1.20	0.36	1.12	0.31	0.20	0.42	0.23	0.12	0.26	0.37	0.09	0.33	0.46	0.25	0.21	0.31	0.05	0.04	0.35
2013-06-12	0.18	0.13	0.14	0.05	0.08	0.12	0.04	0.12	0.03	0.05	0.11	0.04	0.06	0.06	0.03	0.03	0.04	0.03	0.03	0.02	0.02	0.03	0.05	0.04	0.01	0.00	0.03
2013-06-13	0.35	0.28	0.17	0.07	0.12	0.13	0.05	0.12	0.04	0.08	0.10	0.07	0.05	0.06	0.05	0.03	0.04	0.06	0.02	0.03	0.03	0.05	0.05	0.04	0.01	0.02	0.04
2013-06-14	0.19	0.15	0.14	0.04	0.07	0.03	0.02	0.11	0.03	0.05	0.08	0.04	0.07	0.03	0.05	0.03	0.04	0.05	0.03	0.02	0.02	0.05	0.02	0.04	0.01	0.02	0.04
2013-06-15	1.68	1.33	1.94	1.22	0.63	0.10	0.73	0.38	0.44	0.27	0.43	0.40	0.30	0.30	0.41	0.28	0.10	0.31	0.08	0.20	0.16	0.24	0.12	0.12	0.99	0.32	0.07
2013-06-16	1.46	1.10	0.99	0.93	0.85	0.35	0.50	0.32	0.34	0.21	0.17	0.31	0.22	0.23	0.33	0.33	0.14	0.16	0.12	0.22	0.20	0.14	0.10	0.14	0.52	0.44	0.41
2013-06-17	0.32	0.26	0.19	0.19	0.18	0.13	0.12	0.13	0.06	0.07	0.04	0.08	0.07	0.09	0.04	0.12	0.08	0.04	0.02	0.03	0.04	0.05	0.03	0.04	0.04	0.07	0.10
2013-06-18	0.13	0.10	0.07	0.08	0.06	0.04	0.06	0.07	0.02	0.05	0.02	0.04	0.03	0.04	0.02	0.02	0.03	0.05	0.02	0.01	0.02	0.02	0.03	0.02	0.01	0.02	0.18
2013-06-19	0.18	0.11	0.05	0.08	0.03	0.04	0.05	0.02	0.02	0.03	0.01	0.02	0.02	0.02	0.01	0.02	0.03	0.01	0.01	0.03	0.01	0.01	0.01	0.01	0.00	0.02	0.02
2013-06-20	0.15	0.10	0.08	0.06	0.02	0.02	0.03	0.03	0.01	0.02	0.02	0.02	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.01
2013-06-21	0.10	0.08	0.02	0.04	0.02	0.01	0.02	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.02	0.00	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.02	0.01
2013-06-22	0.39	0.26	0.41	0.71	0.15	0.08	0.51	0.07	0.08	0.14	0.13	0.10	0.04	0.08	0.02	0.17	0.06	0.05	0.03	0.04	0.06	0.04	0.02	0.06	0.01	0.03	0.03
2013-06-23	0.14	0.13	0.07	0.26	0.12	0.08	0.21	0.04	0.10	0.06	0.01	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.02	0.01	0.01	0.00	0.01	0.01
2013-06-24	0.05	0.05	0.03	0.02	0.03	0.01	0.01	0.03	0.01	0.02	0.01	0.01	0.01	0.02	0.01	0.00	0.01	0.00	0.12	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.01
2013-06-25	0.04	0.04	0.40	0.02	0.02	0.00	0.01	0.03	0.04	0.02	0.02	0.01	0.01	0.01	0.01	0.12	0.03	0.02	0.07	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.01
2013-06-26	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Figure 3: The number of occurrences of the most frequent words during the protest days from 31/05/2013 to 27/06/2013

is why words like “ga” means “gas”. Then, we showed the number of occurrences of these words in each day of the 27 days in Figure 3. The figure shows that the number of occurrences of the words on the most violent days of the protest (the same days of the spikes in Figures 2 and 3).

6 Discussion and Conclusion

In this paper, we investigated using ML models trained on crowdsourced labelled data to detect protest repression events from tweets. We used a case study from the Turkish Gezi Park protest 2013 and our results show a good potential in using Twitter as a source of information to detect protest repression events quickly and reliably. There are also challenges as tweets often contain spelling and grammatical mistakes. We overcame these by using Twitter specific preprocessing steps. We found two challenges in the crowdsourcing task: the subjectivity of violence and the accuracy of the workers. We might overcome these challenge by hiring more crowd workers with a good experience level and by providing test questions to evaluate the performance of the workers. The model presented in this paper is trained only on data from one protest which limits its general-ability. Collecting tweets from other protests from other countries is part of future work. As well as training the model on text and images like in [ZP19, ALR15]. Using conventional ML models gave good AUC scores but deep learning models like LSTM or CNN could give better results especially with bigger dataset. Also with the huge leap in Natural Language Processing with the development of BERT, this will be used in the future work to represent each word with regarding to the context of the text [DCLT18].

Acknowledgment

Thanks to Dr.Michel Steuwer for giving valuable feedback on this paper.

References

[ALR15] Samar M Alqhtani, Suhuai Luo, and Brian Regan. Fusing text and image for event detection in twitter. *arXiv preprint arXiv:1503.03920*, 2015.

[Ani16] Alexei Anisin. Repression, spontaneity, and collective action: the 2013 Turkish Gezi protests. *Journal of Civil Society*, 12(4):411–429, 2016.

[APM+13] Luca Maria Aiello, Georgios Petkos, Carlos Martin, David Corney, Symeon Papadopoulos, Ryan Skraba, Ayse Göker, Ioannis Kompatsiaris, and Alejandro Jaimes. Sensing trending topics in twitter. *IEEE Transactions on Multimedia*, 15(6):1268–1282, 2013.

[ASG13] Hamed Abdelhaq, Christian Sengstock, and Michael Gertz. Eventweet: Online localized event detection from twitter. *Proceedings of the VLDB Endowment*, 6(12):1326–1329, 2013.

[Bis15] Dina Bishara. The politics of ignoring: Protest dynamics in late Mubarak Egypt. *Perspectives on Politics*, 13(4):958–975, 2015.

[BW15] Ceren Budak and Duncan J Watts. Dissecting the spirit of gezi: Influence vs. selection in the occupy gezi movement. *Sociological Science*, 2:370–397, 2015.

[CBD10] Chris Callison-Burch and Mark Dredze. Creating speech and language data with Amazon’s Mechanical Turk. pages 1–12. Association for Computational Linguistics, 2010.

[DB02] Christian Davenport and Patrick Ball. Views to a kill: Exploring the implications of source selection in the case of Guatemalan state terror, 1977-1995. *Journal of conflict resolution*, 46(3):427–450, 2002.

[DCLT18] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[Ear03] Jennifer Earl. Tanks, tear gas, and taxes: Toward a theory of movement repression. *Sociological theory*, 21(1):44–68, 2003.

[EBG12] Paul S Earle, Daniel C Bowden, and Michelle Guy. Twitter earthquake detection: earthquake monitoring in a social world. *Annals of Geophysics*, 54(6), 2012.

[EMHMM$^+$13] Jennifer Earl, Heather McKee Hurwitz, Analicia Mejia Mesinas, Margaret Tolan, and Ashley Arlotti. This protest will be tweeted: Twitter and protest policing during the Pittsburgh G20. *Information, Communication & Society*, 16(4):459–478, 2013.

[ESM03] Jennifer Earl, Sarah A Soule, and John D McCarthy. Protest under fire? Explaining the policing of protest. *American sociological review*, pages 581–606, 2003.

[FAC11] Karën Fort, Gilles Adda, and K Bretonnel Cohen. Amazon mechanical turk: Gold mine or coal mine? *Computational Linguistics*, 37(2):413–420, 2011.

[Gel18] Johanna Gellerman. Repression and Protests: A Comparative Case Study on the Causes of Protest Violence. 2018.

[Gen17] Jared Genser. Venezuela Needs International Intervention. Now. *The New York Times.[online]*, 30, 2017.

[HS10] Michael Heilman and Noah A Smith. Rating computer-generated questions with Mechanical Turk. pages 35–40. Association for Computational Linguistics, 2010.

[Joa02] Thorsten Joachims. Text Classification. In *Learning to Classify Text Using Support Vector Machines*, volume 668 of *The Springer International Series in Engineering and Computer Science*, pages 7–33. Springer US, January 2002.

[LNSL17] Quanzhi Li, Armineh Nourbakhsh, Sameena Shah, and Xiaomo Liu. Real-time novel event detection from social media. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pages 1129–1139. IEEE, 2017.

[MBG$^+$10] Bart Mellebeek, Francesc Benavent, Jens Grivolla, Joan Codina, Marta R Costa-Jussa, and Rafael Banchs. Opinion mining of spanish customer comments with non-expert annotations on mechanical turk. pages 114–121. Association for Computational Linguistics, 2010.

[MH08] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[OBBC13] Isabel Ortiz, Sara L Burke, Mohamed Berrada, and Hernán Cortés. World Protests 2006-2013. 2013.

[OV17] Christine Ogan and Onur Varol. What is gained and what is left to be done when content analysis is added to network analysis in the study of a social movement: Twitter use during gezi park. *Information, Communication & Society*, 20(8):1220–1238, 2017.

[PB12] Thomas Poell and Erik Borra. Twitter, YouTube, and Flickr as platforms of alternative journalism: The social media account of the 2010 Toronto G20 protests. *Journalism*, 13(6):695–713, 2012.

[RCE11] Alan Ritter, Sam Clark, and Oren Etzioni. Named entity recognition in tweets: an experimental study. pages 1524–1534. Association for Computational Linguistics, 2011.

[SA$^+$11] Tarek S Sobh, Yasser Aly, et al. Effective and extensive virtual private network. *Journal of Information Security*, 2(01):39, 2011.

[Sar16] Dipanjan Sarkar. Text Classification. In Dipanjan Sarkar, editor, *Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from your Data*, pages 167–215. Apress, Berkeley, CA, 2016.

[SBDS14] Marta Sabou, Kalina Bontcheva, Leon Derczynski, and Arno Scharl. Corpus Annotation through Crowdsourcing: Towards Best Practice Guidelines. pages 859–866, 2014.

[SHH$^+$12] Idean Salehyan, Cullen S Hendrix, Jesse Hamner, Christina Case, Christopher Linebarger, Emily Stull, and Jennifer Williams. Social conflict in Africa: A new database. *International Interactions*, 38(4):503–511, 2012.

[SHM09] Hassan Sayyadi, Matthew Hurst, and Alexey Maykov. Event detection and tracking in social streams. In *Third International AAAI Conference on Weblogs and Social Media*, 2009.

[SOJN08] Rion Snow, Brendan O'Connor, Daniel Jurafsky, and Andrew Y Ng. Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. pages 254–263. Association for Computational Linguistics, 2008.

[SOM10] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes Twitter users: real-time event detection by social sensors. pages 851–860. ACM, 2010.

[SOM12] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Tweet analysis for real-time event detection and earthquake reporting system development. *IEEE Transactions on Knowledge and Data Engineering*, 25(4):919–931, 2012.

[SVC17] Chan S Suh, Ion Bogdan Vasi, and Paul Y Chang. How social media matter: Repression and the diffusion of the Occupy Wall Street movement. *Social science research*, 65:282–293, 2017.

[Til78] Charles Tilly. Collective violence in European perspective. 1978.

[TNM$^+$16] Joshua A Tucker, Jonathan Nagler, Megan MacDuffee, Pablo Barbera Metzger, Duncan Penfold-Brown, and Richard Bonneau. Big Data, Social Media, and Protest. *Computational Social Science*, page 199, 2016.

[WHK13] Aobo Wang, Cong Duy Vu Hoang, and Min-Yen Kan. Perspectives on crowdsourcing annotations for natural language processing. *Language resources and evaluation*, 47(1):9–31, 2013.

[YMO18] Xiao Yang, Craig Macdonald, and Iadh Ounis. Using word embeddings in twitter election classification. *Information Retrieval Journal*, 21(2-3):183–207, 2018.

[ZP19] Han Zhang and Jennifer Pan. Casm: A deep-learning approach for identifying collective action events with text and image data from social media. *Sociological Methodology*, 49(1):1–57, 2019.