Leveraging Heterogeneous Data Sources for Civil Unrest Prediction

Lu Meng[1] and Rohini K. Srihari[1]

State University of New York at Buffalo, Buffalo NY 14260, USA
{lumeng,rohini}@buffalo.edu

Abstract. Predicting significant societal events and generating early warnings is a challenging and critical problem since it involves multiple societal factors, including economics, politics, environment, and culture. Civil unrest prediction tasks have thus far relied on manually curated data sources and heuristic features determined by human expertise. Furthermore, current research focuses on machine learning approaches which do not effectively use heterogeneous data sources reflecting sequential data. In this paper, we propose a novel predictive model which effectively exploits such data sources through LSTM networks. Extensive experiments have been conducted on 2 different datasets related to 2 countries to illustrate the effectiveness of our model. Our results show that 1) LSTM networks demonstrate outstanding performance for such tasks, outperforming baserate models and other machine learning approaches, 2) historical data is useful for predicting the near future (the next day), but other, dynamic data is necessary to accurately predict events with increased lead time (3, 5 or 7 days).

Keywords: Civil unrest events prediction · Multiple data sources · Recurrent neural network · Predictive analytics.

1 Introduction

Non-violent Civil Unrest (CU) events are physical acts that occur in public venues such as demonstrations and protests. Such events have critical impact on the stability of a society and daily lives of the public. For example, the protest demonstrations against communism in 1982 played a role in the fall of the Berlin Wall, and the student revolt and labor strikes that ground France to a halt in 1968. Having the ability to produce predictions on CU events has significant impact in multiple areas. It helps the government to make decisions of national security plans and mobilize resources, industries to manage resource allocation, especially the companies with supply chains nationwide or worldwide. It also allows the public to be aware of the situation ahead of time and avoid hostile regions. Therefore, government and research organizations are paying great attention to this area.

In recognition of the importance of this task, IARPA is running a competition, the Mercury Challenge[1], which looks for innovative solutions for the au-

[1] https://www.iarpa.gov/challenges/mercury.html

tomation of societal events forecasts in the Middle East. Topics of the challenge include Military Activity in Egypt, Saudi Arabia, etc., Non-violent Civil Unrest in Egypt and Jordan, and Infectious disease in Saudi Arabia. Participants are required to provide warnings at least 3 days in advance, and all kinds of data source are encouraged to be included.

In this paper, we are trying to gather data which reflects day-to-day life, this falls into economic, general social sentiment, environment, and political areas. Our goal is to learn a predictive model that can provide forecast about CU events for the given countries, and determine which of the factors are directly correlated to the events. An earlier version of this work, which covers multiple cities in countries including India, Kenya, etc., has been deployed in a system for early warning and social disruption developed at a leading NGO, PeaceTech Lab [2]. The system described in this paper reflects a new approach including new data sources.

The main contributions of our study are summarized as follows. 1) Unlike most of the existing social science efforts which have used annual or semi-annual macroeconomic data in predicting disruptions, our work uses the proxies for these macroeconomic indicators which are updated daily, weekly, and monthly. 2) We combine these proxies organically by introducing the deep learning model, which learns the representations and features of sequential data automatically, in contrast of the traditional ways which rely heavily on human expertise and efforts. 3) In participation of the IARPA Mercury Challenge, we are able to forecast the event counts for a specific day or a week over a country, which poses more challenges and significance than only giving whether there will be an event in such country. 4) Our model has been evaluated and proven the effectiveness by using 2 separate datasets, the Gold Standard Reports (GSR) data provided by the challenge and the Armed Conflict Location & Event Data Project (ACLED) data [8], for 2 countries, Egypt and Jordan, against the base-rate in terms of the "Mercury score" and accuracy.

2 Related Work

Many kinds of data sources have been used for forecasting events. Publicly available online information, such as Recorded Future, which collects open content web sources from Twitter, blogs, online forums, and government publications, are used for making predictions about event occurrences [3]. Such data sources are also used for planned protests prediction by extracting relevant information according to features learned [6]. Information extracted form the Global Database of Events, Language, and Tone (GDELT) [4] is widely used by research work on predicting protest events, such as the temporal burst pattern of GDELT [7]. Some work focuses on using external information to predict conflicts, for example, future climate scenarios and variables like population, political rights are used to forecast conflicts in sub-Saharan African [10]. Interest in introducing social media data into the prediction tasks has been growing in recent years, and

[2] https://www.peacetechlab.org

the significant positive correlation between the volume of future protests descriptions on social media and protest onsets has already been proved [11]. EMBERS combines historical data and social media data to make predictions of protests by using classical machine learning models, such as logistic regression [9].

In the work described above, traditional machine learning models are used, such as logistic regression models [11], LASSO regression model [9], and Dynamic Query Expansion (DQE) [13] [14]. This poses the limitation in combining heterogenous data sources and addressing the importance of time sensitive sequential data. Researches have been done in predicting significant protests in only the next few days (within 3 day) [3], which lacks of lead time and prediction granularity when comparing with our work. Others have conducted similar work or benchmarked on event datasets that are static, and focus on answering questions such as "will there be a protest in the next few days". This motivates our work presented in this paper to study the potential drivers of civil unrests by using deep learning model which overcomes the shortcomings of traditional machine learning models, and to generate predictions on the counts of future events.

3 Data Sources

Historical Data. We use a large volume of data generated by IARPA for the Mercury Challenge as the ground truth for our evaluation. The Gold Standard Reports (GSR) data sets contain details on more than 120,000 significant events in the areas of Military Activity, Disease, and Non-violent Civil Unrest, which is the type of events we focus on in this paper. The CU events in the data set are encoded for Egypt and Jordan for over 44 months since May 01, 2015, and are updated every two weeks with events from the previous weeks while the challenge is proceeding. Our focus is forecasting the counts of CU events for each country and generating corresponding warnings.

Table 1 shows a single CU event from the GSR about "Other Government Issues" happened on April 01, 2016 in Alexandria, Egypt, which involved a large crowd of people. Additional information not shown in the table includes links of the news where the event is first reported, news sources, and revision date. Figure 1 shows the frequency of events in each country in each month.

Table 1: A sample NCU event entry in GSR

Event ID	**Event Date**	**Event Type**	**Country**	**City**
CU51468	2016-04-01	Civil Unrest	Egypt	Alexandria
Crowd Size	**Earliest Reported Date**	**Reason**		...
Large	2016-04-01	Other Government Issues		...

We also apply the proposed model on the ACLED data to validate our observations and experiment at a finer level, say generating city specific predictions.

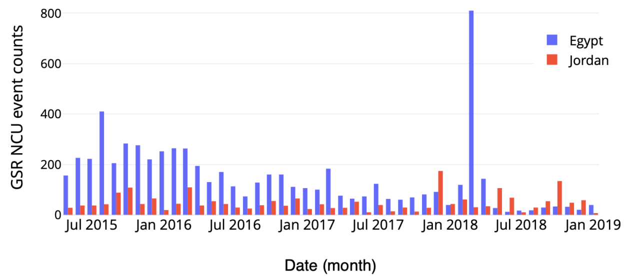

Fig. 1: Monthly frequencies of CU events in GSR data in Egypt and Jordan from May 2015 to January 2019.

ACLED data is a publicly available dataset which contains date, actors, locations, and fatalities of all reported violence and protest events across Africa, South Asia, and so on. Comparing to the GSR data, ACLED data has better coverage of countries, and richer description of the data also allow us to generate audit trails and summaries of events. One problem of using ACLED for forecasting future events is that there is a lag of one week or more for availability of data. The CU events, which are categorized as “Riots/Protests” in ACLED, are extracted from Cairo, Egypt within the same time period as the GSR data. Events in Amman, Jordan cover the timespan from January, 2016 to January, 2019, since they start to be recorded by ACLED only until then.

Economic Indicators. Studies show that exogenous political and economic shocks can serve as the necessary underlying drivers of social unrest [2]. Therefore we studied the relationship between the CU event counts in GSR data and commodity prices[3], unemployment rate, inflation rate[4], etc. for each country. Table 2 shows the covariance of GSR counts and some of the important indicators according to Wikipedia[5][6]. We notice that the importance or relativeness of indicators vary from country to country. For example, the price of Copper is strongly correlated with the CU events in Jordan, while it is not so important in Egypt, as the covariance value is quite negative. Different input are then used to train models for different counties.

Social Media Data. Social media has transformed the traditional relationship between government authority and its citizens by providing the people with an innovative and powerful means to harmonize their efforts in expressing their political and social concerns as well as getting to know the political trends [12].

[3] Source: http://www.worldbank.org
[4] Source: https://tradingeconomics.com
[5] Source: https://en.wikipedia.org/wiki/Economy_of_Egypt
[6] Source: https://en.wikipedia.org/wiki/Economy_of_Jordan

Table 2: Covariance between economic indicators and GSR CU event counts.

Indicator	Egypt	Jordan	Indicator	Egypt	Jordan
Cotton	-9.83	-0.29	Gold	-2202.15	-228.08
Rice	-741.70	-1.82	Natural gas	-21.12	**7.22**
Wheat	-423.98	**56.12**	Iron ore	-431.48	**1.17**
Maize	**362.73**	**25.43**	Copper	-26373.92	**1470.69**
Sugar	-0.61	-0.24	Unemployment rate	**58.64**	-0.41

Political tweets from May 2015 to January 2019 are collected for a list of Twitter accounts who are politicians and journalists in Egypt and Jordan. Daily volume and sentiment are then calculated by filtering and aggregating the above mentioned tweets by using a dictionary of keywords which are determined by political scientists and CIA The World Factbook[7]. The keyword set contains English words and the corresponding Arabic translation that denote civil unrest, for example, "protest", "protester", "riot", "strike".

Open Source Indicators. Open Source Indicators such as Google Trends[8] have the potential to reveal the dynamics of social behavior that precede episodes of civil unrest [5]. Google Trends is a real-time daily and weekly index of the volume of queries that users enter into Google search [1]. We gathered weekly Google Trends data in Egypt and Jordan for the same time period of GSR data by providing a set of keywords, which contains English words, synonyms and translations into Arabic.

4 Methodology

4.1 Problem definition

Our goal is to forecast CU events counts by ingesting multiple data sources, as well as trying to find out what are the leading factors to predict CU events in the countries that we study. We define this task as a classification problem: suppose $\mathbf{x}_t \in \mathbb{R}^n$ is a feature vector corresponding to a day t, where each entry $x_i \in \mathbb{R}^d$ of the vector contains the historical data from the previous d days, and $y_{t+\Delta t} \in \{0, ..., k\}$ refers to the category indicating the number of events on the day of $t + \Delta t$. We are trying to produce the function $f : \mathbb{R}^n \to \{0, ..., k\}$, such that $y_{t+\Delta t} = f(\mathbf{x}_t)$.

4.2 Models

Clear shortcomings of traditional feedforward neural networks have been proven when being applied to time series related tasks, because time sequences are hard

[7] Source: https://www.cia.gov/library/publications/resources/the-world-factbook/
[8] Source: https://www.google.com/trends

to be captured by these models. Recurrent Neural Network (RNN) addresses this issue and uses the internal state to process the sequences of input. LSTM Networks are a special kind of RNN which has the capability to learn long term dependencies in the input sequences. Inspired by the Tensorflow Quick, Draw! tutorial [9], we believe societal events are like the drawings, where historical statistics of social indicators are the strokes of the picture. We propose an RNN-based model, Cov-LSTM, which uses a combination of convolution layers and LSTM layers, to predict the CU events and learn what are the critical "strokes".

Figure 2 shows the architecture of our proposed model. Input layer takes in vectors of indicator values within the input window. Values are then normalized, and differences between consecutive days are calculated. A series of 1 dimensional convolutional layers are added as well as the dropout layers. Output of the convolutional layer are input into a stack of bidirectional LSTM layers. Finally a softmax layer is used as the output layer for the classification task.

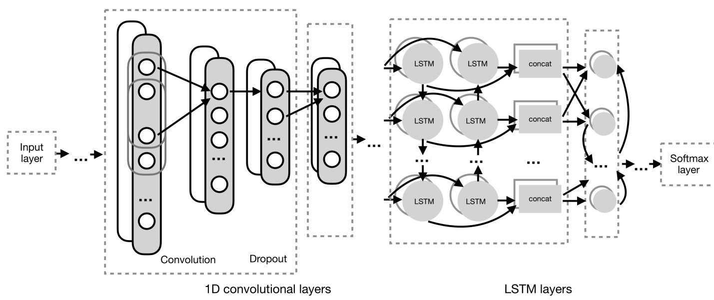

Fig. 2: The framework of the proposed model.

5 Experiments

Our work involves more than 1340 data points from over 3 years timespan from May 01, 2015 to January 08, 2019 (ACLED data for Amman, Jordan starts in January, 2016). We conduct the experiments on 2 datasets - GSR data and ACLED data, and focus on 2 countries, Egypt and Jordan. 70% of the data is used for training, and the rest 30% is used as testing. Each data point corresponds to a day and is labeled by the number of events happened on that day. Data points are then categorized into the buckets according to the event counts on that day. We choose the volume of political tweets and the sentiment, i.e., the percentage of angry posts, as proxies for government and civil unrest related activities on social media. Indicators that are updated on weekly or monthly basis are precessed to have daily values by assigning the same value to the days within the time period.

[9] Source: https://www.tensorflow.org/tutorials/sequences/recurrent_quickdraw

5.1 Evaluation

We conducted the evaluation of our proposed model for each country by using the GSR data against the base rate provided by the Mercury Challenge. Performance is judged by taking into account of both the lead time and the Quality Score (QS). Lead time is defined as the average number of days between the date the prediction is generated and the reported event date. It is required that the lead time should be greater than or equal to 3 days. Quality Score measures the difference between the predicted counts and actual counts, and is calculated as Equation 1. Average is taken among all the data points being tested.

$$Quality\ Score = 1 - \frac{abs(Predicted - Actual)}{max(Predicted, Actual, 4)} \tag{1}$$

The Mercury Challenge uses the Mercury Score for ranking, which is calculated as follows:

$$Mercury\ Score = 1,000,000 * QS \tag{2}$$

Table 3 summarizes the Mercury Score of our proposed model in comparing with the base rate and the scores of other leading groups that are ranked on the top of the leaderboard, for example, "rekcahd" and "valilenk". Experiment settings include: bucket size of 1 is used for categorizing the data, days with more than 10 events are classified into one class, i.e., there are 12 classes in total; social media indicators and economic indicators are used; feature values from the previous 30 days are ingested to forecast the event counts for 3 days later, i.e., lead time equals to 3 days; 3 layers of 1-dimensional convolutional layer and 3 layers of bidirectional LSTM layer are used with 0.3 dropout rate.

Table 3: Comparisons of the Mercury Scores calculated for each evaluation period of our proposed model (Cov-LSTM), base rate model, and top-ranking groups.

	2018-08-01 2018-10-31	2018-11-01 2018-11-13	2018-11-01 2018-11-27	2018-11-01 2018-12-11	2018-11-01 2018-12-25	2018-11-01 2019-01-08
Egypt						
base rate	316479	269231	388889	390244	372727	362319
Cov-LSTM	798007	**846153**	787037	**829268**	**831818**	**829710**
rekcahd	**835015**	836152	**789545**	772238	802162	806661
valilenk	823034	788462	731481	762195	768182	778986
Jordan						
base rate	634114	**1000000**	**888889**	**758608**	**730986**	**699858**
Cov-LSTM	**658532**	647435	698765	680081	691060	699348
rekcahd	521188	768555	679934	571972	615693	647649
valilenk	378845	500000	515873	595238	612670	572817

The results shown in Table 3 demonstrate that by using social media and selective economic indicators, our proposed model outperforms the other competing groups as well as the benchmark, which uses the ARIMA model trained based on the pre-challenge GSR historical data. Specifically, for predictions made for Egypt, our proposed model beats the base rate model in every evaluation period. This also reveals the fact that the ability of historical data for predicting time sensitive events is limited. Models will need to be retrained with new data in order to learn the most up-to-date information for tasks involving time series data. Though the performance of our model trained for Jordan is better than the base rate only for the first period, our model shows strong advantages over the other competing groups for all evaluation periods. As shown in Figure 1, records for CU events in the GSR data for Jordan is much fewer than the ones recorded for Egypt. This indicates less information could be learned from the data by the model trained for Jordan, which may lead to the result that it is hard to beat the base rate performance.

We also conducted experiments under different system settings to study the impact of different indicators. Table 4 illustrates the model performance on the GSR data in terms of prediction accuracy. Results listed show that in general, the model accuracy for predicting the near future is higher than predicting events that will happen far away from the "present", when the same amount of historical data is used. This indicates that the CU events are most likely to be relevant to more recent statistics. When comparing the performance between historical GSR data based model and social media plus indicators based model, we find that historical data tends to be more helpful for the near future. When looking across the different input window size and LT for using the same kind of input, we can also find that including more data from previous days is not necessarily helpful for improving the prediction accuracy. This also proves that the up-to-date information is important for such tasks.

Similar observations can be made when conducting evaluation on the ACLED dataset at city level. Results for Cairo, Egypt are shown as AUC-ROC graphs in Figure 3. One challenge of forecasting city level events is the lack of data points. Therefore, among all the predicted counts for the 2 cities, we find that more false negative cases (the predicted count is 0 while there was 1 or more events happened on that day) than using the models trained by the GSR data at country level.

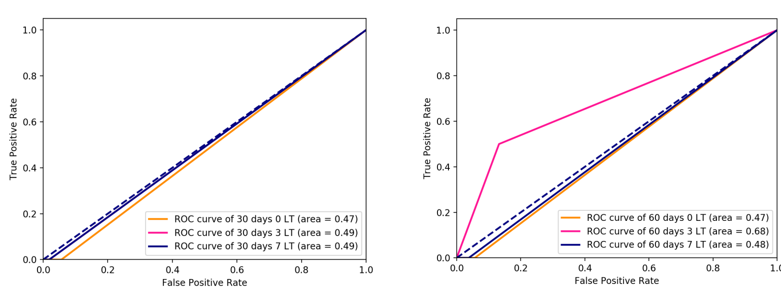

Fig. 3: ROC curves of historical data based model on ACLED data for Cairo, Egypt.

Table 4: Prediction accuracies on GSR data under different system settings.

	Egypt		Jordan	
Input data window and Lead Time (LT)	Historical GSR	Social media + Indicators	Historical GSR	Social media + Indicators
3 days 0 LT	0.255	0.206	0.442	0.426
3 days 3 LT	0.216	0.183	0.343	0.412
3 days 5 LT	0.275	0.162	0.317	0.375
3 days 7 LT	0.325	0.157	0.305	0.287
5 days 0 LT	0.287	0.125	0.275	0.463
5 days 3 LT	0.250	0.262	0.387	0.287
5 days 5 LT	0.200	0.208	0.326	0.462
5 days 7 LT	0.287	0.198	0.275	0.435
30 days 0 LT	0.137	0.150	0.370	0.343
30 days 3 LT	0.087	0.171	0.462	0.423
30 days 5 LT	0.137	0.237	0.439	0.413
30 days 7 LT	0.125	0.217	0.362	0.372
60 days 0 LT	0.175	0.137	0.372	0.323
60 days 3 LT	0.187	0.471	0.230	0.362
60 days 5 LT	0.162	0.250	0.319	0.337
60 days 7 LT	0.137	0.224	0.302	0.324

If the bucket size for categorizing data is 3 instead of 1, which means a prediction is considered as valid if the difference between the predicted value and the actual value is smaller than 3, our model can reach the accuracy of 0.6. We also attempted to include Google Trends data as input, results show that the performance is worse than using historical data or other indicators.

6 Conclusions and Future Work

Civil Unrest events reflect a complexity due to multiple instigating factors and can rarely be attributed to a single reason. Clues to such events can be discerned in advance through indicators of economic, political, or social conditions. Such indicators often change rapidly and vary from country to country. We see promising results by including heterogenous data sources for predicting civil unrests and using LSTM models that effectively exploit sequential data. Future work includes the incorporation of a multitude of new data sources including infrastructure data (electricity, telecom, transportation), video, satellite imagery and other streaming data in a quest to further improve accuracy of prediction. We also are in the process of evaluating our models on other countries and cities using the ACLED data set. This work shows the possibility of leveraging existing data sets to provide predictions of civil unrest with sufficient lead time and granularity to be used in a deployed early warning system for effective resource allocation, safety and security planning and other decision making tasks.

References

1. Choi, H., Varian, H.: Predicting the present with google trends. Economic Record **88**, 2–9 (2012)
2. Dewey, T., Kaden, J., Marks, M., Matsushima, S., Zhu, B.: The impact of social media on social unrest in the arab spring. International Policy Program **5** (2012)
3. Kallus, N.: Predicting crowd behavior with big public data. In: Proceedings of the 23rd International Conference on World Wide Web. pp. 625–630. ACM (2014)
4. Leetaru, K., Schrodt, P.A.: Gdelt: Global data on events, location, and tone, 1979–2012. In: ISA annual convention. vol. 2, pp. 1–49. Citeseer (2013)
5. Manrique, P., Qi, H., Morgenstern, A., Velasquez, N., Lu, T.C., Johnson, N.: Context matters: improving the uses of big data for forecasting civil unrest: emerging phenomena and big data. In: Intelligence and Security Informatics (ISI), 2013 IEEE International Conference on. pp. 169–172. IEEE (2013)
6. Muthiah, S., Huang, B., Arredondo, J., Mares, D., Getoor, L., Katz, G., Ramakrishnan, N.: Planned protest modeling in news and social media. In: AAAI. pp. 3920–3927 (2015)
7. Qiao, F., Chen, K.: Predicting protest events with hidden markov models. In: Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2016 International Conference on. pp. 109–114. IEEE (2016)
8. Raleigh, C., Linke, A., Hegre, H., Karlsen, J.: Introducing acled: An armed conflict location and event dataset: Special data feature. Journal of peace research **47**(5), 651–660 (2010)
9. Ramakrishnan, N., Butler, P., Muthiah, S., Self, N., Khandpur, R., Saraf, P., Wang, W., Cadena, J., Vullikanti, A., Korkmaz, G., et al.: 'beating the news' with embers: forecasting civil unrest using open source indicators. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1799–1808. ACM (2014)
10. Witmer, F.D., Linke, A.M., O?Loughlin, J., Gettelman, A., Laing, A.: Subnational violent conflict forecasts for sub-saharan africa, 2015–65, using climate-sensitive models. Journal of Peace Research **54**(2), 175–192 (2017)
11. Wu, C., Gerber, M.S.: Forecasting civil unrest using social media and protest participation theory. IEEE Transactions on Computational Social Systems (2017)
12. Yang, M.: The collision of social media and social unrest: Why shutting down social media is the wrong response. Nw. J. Tech. & Intell. Prop. **11**, xix (2012)
13. Zhao, L., Chen, F., Dai, J., Hua, T., Lu, C.T., Ramakrishnan, N.: Unsupervised spatial event detection in targeted domains with applications to civil unrest modeling. PloS one **9**(10), e110206 (2014)
14. Zhao, L., Sun, Q., Ye, J., Chen, F., Lu, C.T., Ramakrishnan, N.: Multi-task learning for spatio-temporal event forecasting. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1503–1512. ACM (2015)