2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

When do Crowds turn Violent? Uncovering Triggers from Media

Yue Ning*, Sathappan Muthiah*, Huzefa Rangwala†, David Mares‡, and Naren Ramakrishnan*
*Discovery Analytics Center, Department of Computer Science, Virginia Tech
Email: {yning, sathap1, naren}@vt.edu
†Department of Computer Science, George Mason University
Email: rangwala@cs.gmu.edu
‡University of California, San Diego
Email: dmares@mail.ucsd.edu

***Abstract*—Mass gatherings often underlie civil disobedience activities and as such run the risk of turning violent, causing damage to both property and people. While civil unrest is a rather common phenomenon, only a small subset of them involve crowds turning violent. How can we distinguish which events are likely to lead to violence? Using articles gathered from thousands of online news sources, we study a two-level multi-instance learning formulation, `CrowdForecaster`, tailored to forecast violent crowd behavior, specifically violent protests. Using data from five countries in Latin America, we demonstrate not just the predictive utility of our approach, but also its effectiveness in discovering triggering factors, especially in uncovering how and when crowd behavior begets violence.**

I. Introduction

Large public crowd gatherings are common in all forms of society and some of them can lead to violence, involving damage to both property and people. Examples of such crowd gatherings include political rallies, protests, and commemorative events. When a crowd turns violent it generates economic, political, and social costs, in addition to the emotional and physical (including death) consequences for individuals directly involved in the violence. Each of the parties who support the right to peaceful gatherings (e.g., government and police officials, community organizations) seek to develop better insights into the triggers that can foment violence in hopes of reducing the risk of violence. Efforts to decrease the probability of a violent gathering without understanding the dynamics that differentiate violent from non-violent events can lead to measures that instead increase that probability. For example, deploying a significant show of force with police and the military at the first sign of a protest can make the protesters feel intimidated and frustrated rather than protected. That frustration can, as we suggest, build into anger and increase the likelihood of violence during the next such gathering.

The outbreak of violent crowd behavior in public is usually a culmination of a stream of preceding unresolved public issues or events. As such, we hypothesize that there would be an underlying progression of events that have occurred in the past that may cause outrage and action in violence. Figure 1 demonstrates the average number of public gatherings that have occurred before both violent and non-violent events in five South American countries. Before the occurrence of a violent event, more protests occur on average in the prior week in comparison to a non-violent event for all the countries. Inspired by this observation, we leverage recent work in multi-instance learning [1] to develop new methods that forecast the occurrence of violent crowd behavior in advance. In particular, by integrating the correlation between the past protest events and future violent protest events, the model forecasts outbreaks of crowd violence using historical web data in both spatial and temporal aspects. Moreover, this approach identifies the precursors in the days preceding the violent protest. Our key contributions are summarized as follows:

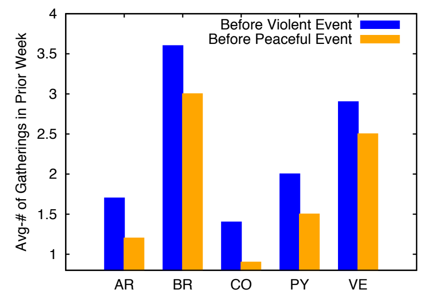


Fig. 1: Average number of gatherings (in prior week) that precede violent and non-violent events. AR: Argentina, BR: Brazil, CO:Colombia, PY: Paraguay, VE:Venezuela.

- We develop a framework based on multi-instance learning for forecasting violent protest events. The framework is built on the hypothesis that violent crowd behavior tends to have a qualitatively different set of trigger events signaling their occurrence in the future. The framework is significantly advantageous over computer vision techniques (e.g. [2]) that only detect events (not forecast them) and which require the first images of violence to be published.
- In addition to forecasting violent events better, our approach leads to explainable predictions by identifying related documents in the past that can be viewed as precursors for violence. Such evidence helps policy-makers and social scientists to better understand the processes governing the formation of collective identities that turn to violence.

IEEE/ACM ASONAM 2018, August 28-31, 2018, Barcelona, Spain
978-1-5386-6051-5/18/$31.00 ©2018 IEEE

- We conduct extensive experiments comparing our approach with existing state-of-the-art models on open datasets from five different countries. In particular, the proposed model outperforms a deployed online system (EMBERS [3]), in terms of quality of forecasts and other baseline methods in terms of accuracy and AUC scores.

II. Related Work

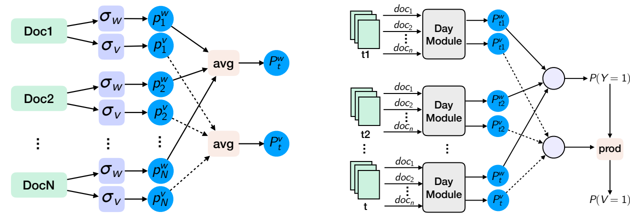
Event forecasting and detection with open source feeds such as Twitter and news articles has been an active area of research in the past decade [4], [1]. With large scale multi-source datasets, forecasting these events with spatial and temporal specificity has been explored using both supervised and unsupervised approaches. Established techniques use a combination of content-based features, such as topic related keywords, as input to methods such as SVM, LASSO, and multi-task learning algorithms [5], [6], [3], [7]. Real-time violence detection has been previously studied in the field of computer vision [2], [8], [9].

Multi-instance learning (MIL) has been exploited in various applications [10], [11]. Within the MIL paradigm, labels are associated with groups of instances commonly referred by *bags* instead of individual instances. Recent work [12] has focused on instance-level predictions from group labels (GICF) and allows for general aggregation functions to detect sentences associated with sentiments. A nested multi-instance learning (nMIL) has been proposed [1] to forecast civil unrest events. However, the nMIL model does not explore the violence problem within civil unrest events and does not exploit the partial labels from the instances regarding the number of civil unrest events that have occurred in the week prior to the event of interest. In this paper, we build upon this prior work by extending the nested multiple instance learning formulation to handle temporal associations across bags and using the relationships between bags and instances to analyze how gatherings beget violence.

III. Methodology

Formally, we are given a set of training examples $\mathcal{D} = \{\mathbb{S}, Y, V\}_{r=1}^{N}$, where $\mathbb{S} = \{X_i\}_{i=1}^{H}$ is a bag of history days and $Y \in \{0,1\}, V \in \{0,1\}$ when $Y = 1$. Y is the label for one training example indicating if a protest event occurs on day $t+k$. $k \in [1, 2, ...]$ is the lead time of prediction that can be tuned in the experiment to evaluate how early the model can predict. V is the label for a violent protest event on day $t+k$. Each day $X_i = \{\mathbf{x}_{ij}\}_{j=1}^{n_i}$ is a set of documents. The overall architecture of the estimated probabilities are demonstrated in Fig. 2b and Fig. 2. In the nested multi-instance learning model [1], for a news article $\mathbf{x}_{ij}$(j is the document index and i is the day index), the probability of it being associated with a protest event is modeled as a logistic function: $p_{ij} = \sigma_w(\mathbf{w}\mathbf{x}_{ij})$. Here $\mathbf{w}$ is the model parameter that is to be optimized and it has the same dimension as $\mathbf{x}_{ij}$.

In our problem, there are two categories for the target events: *violent* and *non-violent* events. We introduce a new model parameter, $\mathbf{v}$, for violent crowd events, with the same



(a) Day Module (b) Overall architecture

Fig. 2: The proposed method

dimension as $\mathbf{x}_{ij}$. Likewise, the probability of a news article related to a violent event is defined as:

$$p_{ij}^{v} = \sigma_v(\mathbf{v}^T\mathbf{x}_{ij}) = \frac{1}{1 + e^{-\mathbf{v}^T\mathbf{x}_{ij}}} \tag{1}$$

Given a protest event, the historic daily probability (P_i) for day $t-i$ indicates how likely a protest event is going to happen on day $t+k$. The probability of this event occurring, $P(Y=1)$, on day $t+k$ is estimated as an average vote from H history days:

$$P(Y=1) = \frac{1}{H}\sum_{i=t-H}^{t} P_i = \frac{1}{H}\sum_{i=t-H}^{t}\frac{1}{n_i}\sum_{j}^{n_i} p_{ij} \tag{2}$$

The probability of a violent protest event is modeled as a joint probability of violence and protest. Applying the Bayes rule, we get:

$$\begin{aligned}\gamma &= P(V=1, Y=1) = P(V=1|Y=1) * P(Y=1)\\ &= \Big(\frac{1}{H}\sum_{i=t-H}^{t}\frac{1}{n_i}\sum_{j}^{n_i} p_{ij}^{v}\Big)\Big(\frac{1}{H}\sum_{i=t-H}^{t}\frac{1}{n_i}\sum_{j}^{n_i} p_{ij}\Big)\end{aligned} \tag{3}$$

Given a set of true labels Y (protest) and V (violent protest) for the super bags, we also know if any protest event occurs ($Y_i = 1/0$) on each history day i in the same city before the target event. We can train our model by minimizing the following cost function with respect to $\mathbf{w}$ and $\mathbf{v}$ as:

$$\begin{aligned}\min_{\mathbf{w},\mathbf{v}} \frac{\alpha}{n}\sum_{\mathbb{S}\in\mathcal{S}} \Big[&\mathcal{L}(\mathbf{w},\mathbf{v}) + (\gamma - \frac{1}{H}\sum_{i=t-H}^{t} Y_i)^2\\ &+ \frac{1}{H}\sum_{i=t-H}^{t} g(\mathbf{w}, X_i, X_{i-1})\\ &+ \frac{1}{H}\sum_{i=t-H}^{t}\frac{1}{n_i}\sum_{j=1}^{n_i} h(\mathbf{x}_{ij}, \mathbf{w})\Big] + R(\mathbf{w},\mathbf{v})\end{aligned} \tag{4}$$

where $\mathcal{L}(*)$ represents the negative log-likelihood loss, $g(*) = (P_i - P_{i-1})^2$ is a squared loss function for two consecutive days requiring two days' probabilities to be close, and $h(*) = \max(0, m_0 - sgn(p_{ij} - p_0)\mathbf{w}^T\mathbf{x}_{ij})$ is the hinge loss function at the instance level where p_0 and m_0 are hyper-parameters. $R(\mathbf{w},\mathbf{v}) = \frac{\beta_1}{2}||\mathbf{w}||^2 + \frac{\beta_2}{2}||\mathbf{v}||^2$ is the regularization for the model parameters. In particular, violence is an attribute or a result of protest events. Protest events can end either peacefully or violently. The probability of a violent protest

Algorithm 1 `CrowdForecaster` algorithm

Input: $\mathcal{S} = \{(\mathbb{S}_r, Y_r, V_r)\}$,
Output: $\{(\mathbf{w}, \mathbf{v})\}$
Pre-compute $\{\sum Y_i | (i = t - H, ..., t)\}$ for each event.
Initialize $\mathbf{w}, \mathbf{v}$
for $\tau = 1$ to T **do**
  for super bag (S_r, Y_r, V_r) **do**
    Fix $\mathbf{w}$, update $\mathbf{v}$ ▷ Solving Eq. 6
    Fix $\mathbf{v}$, update $\mathbf{w}$ ▷ Solving Eq. 7
return $\{(\mathbf{w}, \mathbf{v})\}$

event is conditional on the probability of a protest event. Thus the negative log-likelihood for violent events is calculated as follows:

$$\begin{aligned}\mathcal{L}(\mathbf{w}, \mathbf{v}) = &- I(Y = 0) \log P(Y = 0) - (Y = 1) \log P(Y = 1) \\ &- I(V = 1, Y = 1) \log P(V = 1, Y = 1) \\ &- I(V = 0, Y = 1) \log P(V = 0, Y = 1) \qquad (5)\end{aligned}$$

From our observations, violent protest events usually follow a sequence of protest events. Thus, the probabilities of violent protest events are assumed to be highly related with the number of protest events that occur before these violent events $((\gamma - \frac{1}{H}\sum_{i=t-H}^{t} Y_i)^2)$. For instance, if there are seven protest events in the same city in one week, it is highly probable that some protest will turn to violence because the growing anger and dissatisfaction tend to make people resort to violence. We develop an alternating minimization algorithm that can be applied to achieve a solution of Eq. 4.

Update $\mathbf{v}$ when $\mathbf{w}$ is fixed. All $\mathbf{v}$'s can be solved by a stochastic gradient descent algorithm as:

$$g(v) = \alpha \Big[- \frac{V - \gamma}{\gamma(1-\gamma)} \frac{\partial \gamma}{\partial \mathbf{v}} + 2\big(\gamma - \frac{1}{H}\sum_{i=t-H}^{t} Y_i\big) \frac{\partial \gamma}{\partial \mathbf{v}} \Big] + \beta_2 \mathbf{v} \qquad (6)$$

Update $\mathbf{w}$ when $\mathbf{v}$ is fixed. Likewise, when $\mathbf{v}$ is fixed, all $\mathbf{w}$'s derivatives have a formulation as follows:

$$\begin{aligned} g(w) = &\alpha \Big[- \frac{Y - P}{P(1-P)} \frac{\partial P}{\partial \mathbf{w}} - \frac{V - \gamma}{\gamma(1-\gamma)} \frac{\partial \gamma}{\partial \mathbf{w}} \\ &+ 2\big(\gamma - \frac{1}{H}\sum_{i=t-H}^{t} Y_i\big) \frac{\partial \gamma}{\partial \mathbf{w}} \\ &+ \frac{1}{H}\sum_{i=t-H}^{t} \Big(2(P_i - P_{i-1})(\frac{\partial P_i}{\mathbf{w}} - \frac{\partial P_{i-1}}{\mathbf{w}}) \\ &- \frac{1}{n_i}\sum_{j=1}^{n_i} sgn(p_{ij} - p_0)\mathbf{x}_{ij} I_{ij} \Big) \Big] + \beta_1 \mathbf{w} \qquad (7) \end{aligned}$$

where I_{ij} is the indicator function when $sgn(p_{ij} - p_0)\mathbf{w}^T\mathbf{x}_{ij} \leq m_0$, it returns 1. We alternatively updating $\mathbf{w}$ and $\mathbf{v}$ via gradient descent toward convergence. A complete algorithm is described in Algorithm 1. After learning the model parameters $(\mathbf{w}, \mathbf{v})$ from training examples, test examples are evaluated in terms of the metrics described later in Section IV. One issue in the problem is the imbalanced class distribution due to fewer examples of violent protest in the real world datasets. We apply one of the traditional techniques, *OverSampling*, to adjust the class distribution. It samples the smaller class at random with replacement until it has as many samples as the majority class.

In order to discover the historical related documents for violent protest events, we apply the algorithm described in [1]. It selects instances $(\mathbf{x}_{ij})$ as the precursor documents for each violent event based on their estimated probabilities (p_{ij}^v) if the estimated p_{ij}^v is beyond a threshold.

IV. Experimental Evaluation

A. Experimental Design.

*1) **Datasets**:* We collected news articles from top news sources in different countries including *Argentina, Brazil, Colombia, Paraguay, Venezuela* from January 2014 to April 2015. Each dataset contains about 9400 to 11,000 news and 600 to 2000 events. Among these events, 6% to 30% are violent events.

*2) **Ground Truth**:* The civil unrest forecasting results were validated against a labeled set called Gold Standard Report (GSR) that was exclusively provided by MITRE (see [3] for more details). The GSR is a manually curated dataset that records the occurrence of a civil unrest events reports from the ten most significant news outlets as ranked by International Media and Newspapers in each of the countries studied here. An example of a ground truth GSR recorded event is given by a tuple: (Location= "Argentina, Fortaleza, Ceara", DATE = "2014-01-20", Protest = "True", Violence="True"). Here we use the "Violence" attribute as our violence label V and "Protest" as our event label Y. These GSR reports are the target events that are used for validation of our algorithm.

*3) **Experimental Setup**:* To evaluate the MIL-based violent event forecasting, for each violent protest event in a city, we download all the published news articles in that city for up to 14 days before the occurrence of the specific event. This ordered collection of per-day news documents up to the violent protest day are considered as positive super bags. We generate negative samples in two ways: (i) For each location (city) we identify a period of three consecutive days where no identified event of interest (as reported by GSR) occurs. The ordered collection of per-day news documents not leading to a protest event are labeled as negative super bags. (ii) Similarly for each location (city) the ordered collection of per-day news documents leading to a protest event but of *non-violent* nature are labeled as negative super bags as well. We split our datasets into training and testing (held-out) partitions and perform 3-fold cross-validation on the training set to tune the parameters of the proposed models. We represent documents by word and document embedding generated by paragraph vector models [13]. For each document, we learn its representation with dimension of 300 for training. According to the `Nested` model, doc2vec has better performance compared to TFIDF representations.

*4) **Comparative Methods**:* 1). **MI-SVM** [14] (`MI-SVM`): The MI-SVM model extends the notion of a margin from individual patterns to bags. In our case, we collapse the news articles from the r historical days into one bag and each bag has two labels indicating the occurrence of a protest event and a violent protest event on the $t+k$-th day. The MISVM iterates each example in each bag to determine the most positive instance and least negative instance.

2). **Relaxed-MIL** [15] (`Relaxed`): This proposed model uses a Noisy-OR function ($P_i = 1 - \prod_{j=1}^{m_i}(1-p_{ij})$) to estimate the probability of a bag being positive. In our dataset, each bag (day) has more than 10 instances (news articles). Noisy-OR function tends to generate a positive probability when the number of instances in a bag is large. Thus, we estimate the probability of a bag being positive by applying an average function of each instance in the bag.

3). **Nested MIL** (`Nested`): This approach is proposed by Ning et. al. [1]. Instead of general protest event prediction, we use violent protest events as the positive examples in this method.

4). **Nested MIL-MC** (`Multi-Class`): This is the multi-class classifier of nMIL model. In this experiment, we divide the examples into three classes: *violent protest, non-violent protest*, and *no protest*.

5). **EMBERS**[3] (`EMBERS`): This is an automated, 24x7 continuous system for forecasting civil unrest across 10 countries in Latin America using open source indicators such as tweets, news sources, blogs, economic indicators, and other data sources.

*5) **Parameter Settings**:* For `CrowdForecaster`, the parameters were set as follows: The supervised hyper parameter α is set to 0.6. The hyper parameter β_1 and β_2 were set to 0.5 according to the nMIL paper [1]. The learning rate is adaptively set to $\frac{1}{(t+1)*\lambda}$ where λ is 0.05 according to the rMIL paper [15]. m_0 and p_0 are set to be 0.5. Lead time is used to evaluate how early the model forecasts. For instance, if we use data from day 1 to day 5 and forecast if there is a violent event on day 6, the lead time is 1 day. In the experiment, we set lead time as 1 day by default for offline evaluation and 4 days for online evaluation. We vary the lead time from 1 day to 4 days to compare the early forecasting power of the different models.

*6) **Performance Metrics**:* The forecasting alerts generated by the model and the real events are structured records including: date, event type (violent or not), location(city, state, country). The quality score for a forecast involved evaluations based on time and location given by:

$$\mathtt{QS} = 2 * (\mathtt{DS} + \mathtt{LS})$$

where $\mathtt{DS}, \mathtt{LS}$ denote the date score and location score respectively.

$$\mathtt{DS} = 1 - \min(|d_e - d_p|, 7)/7$$

where d_e is the event date and d_p is the predicted date for the event. If the predicted date of the event is the same as the actual date of the event, then $\mathtt{DS}$ is 1. Location score has many ways of definition. In our problem, location is in terms of a triples of (country, state, city). Comparing a true event with a predicted event, we obtain a score at these three levels:

$$\mathtt{LS} = \frac{1}{3}l_1 + \frac{1}{3}l_1 l_2 + \frac{1}{3}l_1 l_2 l_3$$

where l_1 is the country-level score, l_2 is the state level score, and l_3 is the city level score. We selected a set of cities based on their scales. Then we built training and testing examples for these cities and the location score is only calculated for the selected cities.

Other typical evaluation metrics for classification include: accuracy (ACC) and area under curve (AUC) score. True positive examples are the true violent events and the model predicts correctly. True negative examples are the true non-violent events or no-event and model predicts correctly.

B. Experimental Results.

We introduce the results of forecasting violent crowd behavior in several parts. First, we show the offline performance comparison with models `MI-SVM`, `Relaxed`, `Nested`, `Multi-Class` based on accuracy (ACC) and AUC score for five countries in South America. Second, we present the online evaluation of quality scores (QS, LS, DS) for the proposed model and `EMBERS` that delivers warnings 24/7 for these countries. Next, we study a few cases of precursor discovery for violent and non-violent events. We also calculate and analyze the computation time of the proposed model with other models. Finally, we present the sensitivity of model parameters on the performance of the proposed method.

*1) **The overall accuracy for violence prediction**:* Table I lists the comparative performance in terms of Accuracy (ACC) and AUC scores for five countries in South America. In the bottom part of the table, the lift/drop percentage of the proposed model is presented comparing to the `MI-SVM` as a baseline. The accuracy computes the fraction of correct predictions for both positive and negative classes. The AUC is a common evaluation metric for binary classification problems with imbalance. It will be close to 1 when the true positive rate increases quickly. The proposed model outperforms other state-of-the-art methods for all datasets with lead time equal to one day. With a portion of training samples changing from 10% to 100%, Figure 3a shows the AUC scores for the `CrowdForecaster` model and other state-of-the-art models for Argentina dataset. Given the space limitation, we only show this result on Argentina. In general, the AUC scores increase when the number of training samples is increased. With the full set of training samples, the proposed `CrowdForecaster` model outperforms `MI-SVM`, `Relaxed`and `Nested` methods by $26\%, 25\%$ and 22%, respectively.

*2) **Leadtime Evaluation**:* Table II shows the AUC performance of the proposed method `CrowdForecaster` in comparison to the best baseline method, `Nested`, with lead time varying from 1 to 4. For each value of the lead time we train a model where X_r is a super bag containing t historical days and Y_r indicates if a violent protest event happened on day $t+k$. Notice that lead time k indicates the model predicts k days in advance. For Brazil, Colombia, Paraguay

TABLE I: Violent event forecasting performance comparison based on Accuracy (Acc) and AUC score w.r.t to state-of-the-art methods. The proposed `CrowdForecaster` method outperforms state-of-the-art methods across the five countries with 2 weeks historical data.

	Argentina		Brazil		Colombia		Paraguay		Venezuela	
	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
`MI-SVM`	0.307	0.572	0.221	0.515	0.317	0.564	0.281	0.509	0.306	0.518
`Relaxed`	0.631	0.552	0.462	0.518	0.631	0.552	0.569	0.597	0.438	0.489
`Nested`	0.669	0.568	0.552	0.522	0.680	0.645	0.650	0.589	0.560	0.570
`Multi-Class`	0.544	0.551	0.295	0.523	0.626	0.664	0.531	0.587	0.234	0.457
`CrowdForecaster`	**0.804**	**0.712**	**0.762**	**0.540**	**0.791**	**0.681**	**0.892**	**0.661**	**0.584**	**0.594**

TABLE II: AUC scores of the proposed method and the best baseline with lead time from 1 to 4 for violent events.

	Method	AUC			
Dataset	Leadtime(→)	1	2	3	4
Argentina	`Nested`	0.568	0.608	0.610	0.656
	`CrowdForecaster`	**0.712**	**0.674**	**0.646**	**0.689**
Brazil	`Nested`	0.522	0.519	0.507	0.573
	`CrowdForecaster`	**0.540**	**0.584**	**0.540**	**0.613**
Colombia	`Nested`	0.645	0.549	0.693	0.627
	`CrowdForecaster`	**0.681**	**0.619**	**0.735**	0.614
Paraguay	`Nested`	0.589	0.670	0.596	0.593
	`CrowdForecaster`	**0.661**	**0.758**	**0.635**	**0.692**
Venezuela	`Nested`	0.570	0.597	0.609	0.563
	`CrowdForecaster`	**0.594**	**0.628**	**0.642**	**0.588**

TABLE III: Quality scores for the proposed method and the delivery from online system, EMBERS [3]. DS and LS are over 1; QS is over 4.

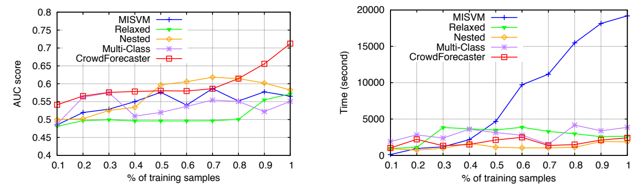
Dataset	Methods	DS	LS	QS
Argentina	`EMBERS`	0.83	0.69	3.04
	`CrowdForecaster`	**0.99**	**0.93**	**3.84**
Brazil	`EMBERS`	0.85	0.81	3.32
	`CrowdForecaster`	**0.99**	**0.99**	**3.96**
Colombia	`EMBERS`	0.82	0.75	3.14
	`CrowdForecaster`	**0.94**	**0.99**	**3.86**
Paraguay	`EMBERS`	0.89	0.76	3.3
	`CrowdForecaster`	**0.95**	**1**	**3.9**
Venezuela	`EMBERS`	0.82	0.8	3.24
	`CrowdForecaster`	**0.93**	**0.99**	**3.84**

and Venezuela, it is noted that a shorter lead time (k=1) does not necessarily imply a better predictive performance compared to a longer lead time (k=3, 4).

*3) **Comparison to the online system in production**:* Table III presents the performance with respect to quality scores for `EMBERS` delivered system and the proposed methods. The proposed model, `CrowdForecaster`, outperforms the online system in event data and event location scores for all the five datasets with an average performance improvement of 21%.

*4) **Trigger Analysis from Results**:* Table IV presents case studies on two violent protests and one non-violent protest. The detected precursor events are reported before these protest events. We select each precursor event by setting a threshold for its probabiliy $p_{ij}^{v} > 0.5$. The top words are selected from the precursor document of that day based on their frequency. We can make three observations about these results.

First, the occurrence of keywords such as "police", "teargas" in the precursors for violent protests suggest that greater forms of control and authority often beget violence. This reflects ongoing understanding that violence often has at its roots distrust between authorities and citizens. Second, we observe that before a violent protest event there have been other protests and strike events (even if peaceful) among the precursors. Words such as "protest", "block", "march", "strike" tend to appear more frequently in news articles preceding a violent protest than a non-violent protest. These findings suggest that a rapidly rising sequence of peaceful events can produce an emotional state among protesters that increases the likelihood of violence in the next protest. We suggest that emotional state is frustration. Given the myriad of challenges confronting civil society in Latin America today, fortunately frustration is not sufficient to produce violence. But the pattern of an increase in a set of words and events is intriguing. Third, analysis of precursor events indicates that they are occurring across a few cities or states, not just in the locale that will subsequently experience violence. One might have thought that protesters would be most affected by what happens locally but this data suggests that those protesters prone to violence reflect upon their national and not just local experiences when formulating their grievances and developing their frustrations. This correlation needs to be explored carefully because it both limits the responsibility of local authorities for potential violence and suggests that locally focused tactics to lower the risk of violence will be of little value.



(a) AUC scores.

(b) Computation time.

Fig. 3: Evaluation for Argentina

*5) **Computational Complexity**:* Figure 3b presents the computation time of the proposed method and state-of-the-art methods on a Dell server with Intel Xeon CPU, 80-core, 504 GB memory based Ubuntu 12.04.5 operating system. The `MI-SVM` algorithm is computationally more expensive when the number of training examples is large. Other probability based MIL methods are relatively stable when the training set is varied from 10% to 100%.

TABLE IV: Precursor events and word distributions in past seven days for violent protest events and one non-violent protest event. Related keywords are manually highlighted (from Google Translate).

Day	Precursor Events	Top Words
Day T-8	**P1**. The rain of Sao Paulo didn't prevent women march on street for equal rights	street
Day T-5	**P1**. Protest in front of the central railway station of Brazil. **P2**. About 25 students and their fathers participate a protest	protest, students
Day T-3	**P1**. Thousands of protesters gather together in the Cinelndia, downtown Rio.	protest
Day T-2	**P1**.The demonstration against the Dilma government and corruption in Belo Horizonte. **P2**. About 3000 people, according to estimate of the military police, participate in the protest	government, police, national
2015-03-17	**Violent Protest**:A group of protesters close the runway in the marginal Tiete, burn tires and garbage bags.	
Day	**Precursor Events**	**Top Words**
Day T-8	**P1**. A young guy died from an attempt of kidnapping.	criminal, justice
Day T-7	**P1**. Senate start debate on the reform of the political penal code and the criminal justice commission matter **P2** Deputy of the New Alliance Party arrive to the session of the plenary of the local congress,	economic, maintain, human
Day T-5	**P1**. The workers against Congress gather in May Square.	organization
Day T-2	**P1**. The judge rejected the proposition on one of the accused criminals who took the property of Lugano	financial, property
Day T-1	**P1**. Men in police disguise assaulted the house of a doctor. **P2**. After two hours of chaos, workers lead a protest	medical, police
2014-05-07	**Non-Violent Protest**: About 500 people gathered last night at the Plaza Santo Martin to demand the authority to enforce the security of the citizens.	

V. Conclusion

We have introduced a framework based on mutli-instance learning for forecasting violent crowd behavior, and identifying precursor events that lead to violence. We empirically evaluated the strengths of our developed method on open source news datasets from five Latin American countries and we conducted a semantic analysis using social media for violent events. Through extensive evaluation and analysis, we illustrate the strong forecasting performance of the proposed methods for violence prediction. We also show qualitatively, via several case studies, the characteristics of identified precursors for both violent and non-violent protest events. In the future, we plan to study the patterns of change in protest events that turn to violence and other societal factors that contribute to the evolution of violent protest events.

Acknowledgements

This work was partially funded by the Office of Naval Research under contract N00014-16-C-1054 and by the U.S. Department of Homeland Security under Grant Award Number 2017-ST-061-CINA01.

References

[1] Y. Ning, S. Muthiah, H. Rangwala, and N. Ramakrishnan, "Modeling precursors for event forecasting via nested multi-instance learning," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1095–1104, 2016.

[2] T. Hassner, Y. Itcher, and O. Kliper-Gross, "Violent flows: Real-time detection of violent crowd behavior," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1–6, 2012.

[3] N. Ramakrishnan, P. Butler, S. Muthiah, and 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, 2014.

[4] N. Kallus, "Predicting crowd behavior with big public data," in *Proceedings of the 23rd International Conference on World Wide Web*, pp. 625–630, 2014.

[5] X. Wang, M. S. Gerber, and D. E. Brown, "Automatic crime prediction using events extracted from twitter posts," in *2012 International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction*, pp. 231–238, 2012.

[6] L. Zhao, Q. Sun, J. Ye, F. Chen, C.-T. Lu, and N. Ramakrishnan, "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, 2015.

[7] S. Muthiah, P. Butler, *et al.*, "EMBERS at 4 Years: Experiences Operating an Open Source Indicators Forecasting System," in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, (New York, NY, USA), pp. 205–214, 2016.

[8] A. Hendel, D. Weinshall, and S. Peleg, "Identifying surprising events in videos using bayesian topic models," *ACCV*, pp. 448–459, 2011.

[9] J. Lin and W. Wang, *Weakly-Supervised Violence Detection in Movies with Audio and Video Based Co-training*, pp. 930–935. 2009.

[10] Z.-H. Zhou and J.-M. Xu, "On the relation between multi-instance learning and semi-supervised learning.," in *Proceedings of the 24th International Conference on Machine Learning* (Z. Ghahramani, ed.), vol. 227, pp. 1167–1174, 2007.

[11] Y.-F. Li, J.-H. Hu, Y. Jiang, and Z.-H. Zhou, "Towards discovering what patterns trigger what labels," in *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, pp. 1012–1018, 2012.

[12] D. Kotzias, M. Denil, N. de Freitas, and P. Smyth, "From group to individual labels using deep features," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 597–606, 2015.

[13] Q. V. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proceedings of the 31st International Conference on International Conference on Machine Learning*, pp. 1188–1196, 2014.

[14] S. Andrews, I. Tsochantaridis, and T. Hofmann, "Support vector machines for multiple-instance learning," in *Proceedings of Advances in Neural Information Processing Systems 15*, pp. 561–568, 2002.

[15] X. Wang, Z. Zhu, C. Yao, and X. Bai, "Relaxed multiple-instance SVM with application to object discovery," in *Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1224–1232, 2015.