Proceedings of the Twenty-Seventh Conference on Innovative Applications of Artificial Intelligence

Planned Protest Modeling in News and Social Media

Sathappan Muthiah
Virginia Tech
Arlington, VA 22203
sathap1@cs.vt.edu

Bert Huang
University of Maryland
College Park, MD 20742
bert@cs.umd.edu

Jaime Arredondo
University of California
San Diego, CA 92093
jarredon@ucsd.edu

David Mares
University of California
San Diego, CA 92093
dmares@ucsd.edu

Lise Getoor
University of California
Santa Cruz, CA 95064
getoor@soe.ucsc.edu

Graham Katz
CACI Inc.
Lanham, MD 20706
ekatz@caci.com

Naren Ramakrishnan
Virginia Tech
Arlington, VA 22203
naren@cs.vt.edu

Abstract

Civil unrest (protests, strikes, and "occupy" events) is a common occurrence in both democracies and authoritarian regimes. The study of civil unrest is a key topic for political scientists as it helps capture an important mechanism by which citizenry express themselves. In countries where civil unrest is lawful, qualitative analysis has revealed that more than 75% of the protests are planned, organized, and/or announced in advance; therefore detecting future time mentions in relevant news and social media is a direct way to develop a protest forecasting system. We develop such a system in this paper, using a combination of key phrase learning to identify what to look for, probabilistic soft logic to reason about location occurrences in extracted results, and time normalization to resolve future tense mentions. We illustrate the application of our system to 10 countries in Latin America, viz. Argentina, Brazil, Chile, Colombia, Ecuador, El Salvador, Mexico, Paraguay, Uruguay, and Venezuela. Results demonstrate our successes in capturing significant societal unrest in these countries with an average lead time of 4.08 days. We also study the selective superiorities of news media versus social media (Twitter, Facebook) to identify relevant tradeoffs.

Civil unrest (protests, strikes, and "occupy" events) is a common happening in both democracies and authoritarian regimes. Although we typically associate civil unrest with disruptions and instability, for a social scientist civil unrest reflects the democratic process by which citizenry communicate their views and preferences to those in authority. The advent of social media has afforded citizenry new mechanisms for organization and mobilization, and online news sources and social networking sites like Facebook and Twitter can provide a window into civil unrest happenings in a particular country.

Why study and forecast protests? Our region of interest is Latin America and protest is an important topic of study here, as many countries here are democracies struggling to consolidate themselves. The combination of weak channels of communication between citizen and government, and a citizenry that still has not grasped the desirability of elections as the means to affect politics means that public protest will be an especially attractive option. To illustrate the power of protest in Latin America we need only recall that between 1985 and 2011, 17 presidents resigned or were impeached under pressure from demonstrations, usually violent, in the streets. Protests have also resulted in the rollback of price increases for public services, such as during the 'Brazilian Spring' of June 2013.

Forecasting protests is an important capability in many domains. For the tourism industry, forecasting protests can support the issuance of travel warnings. For law enforcement, forecasting protests can aid in preparedness. For the social scientist, protests forecasts will provide insight into how citizens express themselves. For the government, a protest forecasting system can help prioritize citizen grievances. Finally, protests can have a debilitating effect on multiple industries (esp. those that rely on worldwide supply chain management) and thus a protest forecasting system can aid in planning and design of alternative travel and shipping routes.

Planned protests Our basic hypothesis is that protests that are larger will be more disruptive and communicate support for its cause better than smaller protests. Mobilizing large numbers of people is more likely to occur if a protest is organized and the time and place announced in advance. Because protest is costly and more likely to succeed if it is large, we should expect planned, rather than spontaneous, protests to be the norm. Indeed, in a sample of 288 events from our study selected for qualitative review of their antecedents, for 225 we located communications regarding the upcoming occurrence of the event in media, and only 49 could be classified as spontaneous (we could not determine whether communications had or had not occurred in the remaining 14 cases).

EMBERS We are an industry-university partnership charged with developing an automatic protest forecasting

system for 10 countries in Latin America. Our system, called EMBERS, has been deployed since Nov 2012 and has been generating forecasts (called warnings or alerts) automatically, without a human-in-the-loop. These forecasts are emailed to a third party (MITRE) for evaluation. Analysts at MITRE organize a reference database of protests (called the Gold Standard Report, or GSR) by surveying newspapers for reportings of protests, and compare our warnings against the GSR to generate a scoring report (evaluation criteria described later).

The full EMBERS system has been described elsewhere (Ramakrishnan et al. 2014), including the overall system architecture, data sources used for analysis, and the various forecasting models in EMBERS. EMBERS adopts a multi-model approach, wherein different models are leveraged for their selective superiorities to generate a fused set of alerts. Arguably, one of the best performing models in EMBERS is the planned protest model that detects ongoing organizational activity and generates warnings accordingly. This paper is the first to present this model in detail, including the research issues involved, and how we addressed them in EMBERS.

Capturing mentions of protest planning and organization is not as easy as it might appear. First, articles of interest are written in different languages (Spanish, Portugese, French, Dutch, and English). Second, multiple locations are often mentioned in a given article, leading to (natural language) ambiguity about the intended location of the event. Significant reasoning is required to discern the correct protest location. Finally, dates are often described in relative terms, e.g., 'Sunday' and thus these vague references need to be resolved into absolute temporal information.

Our detection approach combines shallow linguistic analysis to identify a corpus of relevant documents (articles, tweets) which are then subject to targeted deep semantic analysis. Despite its simplicity, we are able to detect indicators of event planning with surprisingly high accuracy. Our contributions are:

1. We present a protest forecasting system that couples three key technical ideas: key phrase learning to identify what to look for, probabilistic soft logic to reason about location occurrences in extracted results, and date normalization to resolve future tense mentions. We demonstrate how the integration of these ideas achieves objectives in precision, recall, and quality (accuracy).
2. We illustrate the application of our system to 10 countries in Latin America, viz. Argentina, Brazil, Chile, Colombia, Ecuador, El Salvador, Mexico, Paraguay, Uruguay, and Venezuela. Our system predicts the *when* of the protest as well as *where* of the protest (down to a city level granularity). We conduct ablation studies to identify the relative contributions of news media (news + blogs) versus social media (Twitter, Facebook) to identify future happenings of civil unrest. Through these studies we illustrate the selective superiorities of different sources for specific countries.
3. Unlike many studies of retrospective forecasting of protests, our system has been **deployed and in operation for nearly two years.** The end consumers of our alerts are analysts studying Latin America. Our results demonstrate that we are able to capture significant societal unrest in the above countries with an average lead time of 4.08 days.

1 Related Work

Five categories of related work are briefly discussed here (see also Table 1).

Event detection via text extractions is an extensively studied topic in the literature. Document clustering techniques are used in, e.g., (Gabrilovich, Dumais, and Horvitz 2004) to identify events retrospectively or as the stories arrive. Works like (Banko et al. 2007; Chambers and Jurafsky 2011; Riloff and Wiebe 2003) focus on extraction patterns (templates) to extract information from text.

Temporal information extraction is another well studied topic. The TempEval challenge (Verhagen et al. 2009) led to a significant amount of algorithmic development for temporal NLP. For instance, a specification language for temporal and event expressions in natural language text is described in (Pustejovsky et al. 2003). Refs. (Llorens et al. 2012) and (Mani and Wilson 2000) provide methods to resolve temporal expressions in text (our own work here uses the TIMEN package (Llorens et al. 2012)).

Event forecasting: Radinsky and Horvitz (Radinsky and Horvitz 2013) find event sequences from a corpora and then use these sequences to determine if an event of interest (e.g., a disease outbreak, or a riot) will occur sometime in the future. This work predicts only if a potential event will happen given a historical event sequence but does not geolocate the event to a city-level resolution, as we do here. Kallus (Kallus 2014) makes use of event data from RecordedFuture (Truvé 2011) to determine if a significant protest event will occur in the subsequent three days and casts this as a classification problem. This work only focuses on prediction of significant events (suitably defined) and the forecast is limited to a fixed time window of the next three days.

Future retrieval: Baeza-Yates (Baeza-Yates 2005) provides one of the earliest discussions of this topic; here future temporal information in text is found and used to retrieve content from search queries that combine both text and time with a simple ranking scheme. Kawai et al. (Kawai et al. 2010) present a search engine (ChronoSeeker) for searching future and past events. RecordedFuture (Truvé 2011), introduced earlier, conducts real-time analysis of news and tweets to identify mentions of events along with associated times. Anectodally it is estimated that approximately (only) 5–7% of events extracted by RecordedFuture are about the future.

Planned protest detection: Two publications—Compton et al. (Compton et al. 2013) and Xu et al. (Xu et al. 2014)—align very closely to our own work as their emphasis is on protest forecasting. Both works are aimed at forecasting protests but emphasize different data sources and different methodologies. For instance, the work in (Compton et al. 2013) filters the Twitter stream for keywords of interest and searches for future date mentions in only absolute terms, i.e., explicit mentions of a month name and a number (date) less than 31. Such an approach will not be capable of extracting the more common way in which future dates are ref-

Table 1: Comparison of our approach against other techniques.

	Relative date resolution	Ingest multiple sources?	Reasoning about location	Learning word/phrase filters
'Future' Search Engines (2010; 2011; 2005)	✓			
Time-to-Event Recognition (2013; 2013)	✓			
Planned Protest Detection (2014; 2013)		✓		
This paper	✓	✓	✓	✓

erenced, e.g., phrases like "tomorrow," "next tuesday." The work in (Xu et al. 2014) by the same group of authors uses the Tumblr feed with a smaller set of keywords but again is restricted to the use of absolute time identifiers. Furthermore, in this work, location is restricted to membership in a whitelist of 1022 entries which would not be able to capture the diversity of location identifiers at the city level; the system is also unable to reason about the presence of multiple location names.

2 Probabilistic Soft Logic

In this section, we briefly describe probabilistic soft logic (PSL) (Kimmig et al. 2012), a key component of our geocoding strategy described later. PSL is a framework for collective probabilistic reasoning on relational domains. PSL uses first order logic rules to capture the dependency structure of the domain, based on which it builds a joint probabilistic model over all atoms. Instead of hard truth values of 0 (false) and 1 (true), PSL uses soft truth values relaxing the truth values to the interval $[0, 1]$. The logical connectives are adapted accordingly. User defined *predicates* are used to encode the relationships and attributes and *rules* capture the dependencies and constraints. The rules can also be labeled with non-negative weights which are used during the inference process. The set of predicates and weighted rules thus make up a PSL program where known truth values of ground atoms derived from observed data and unknown truth values for the remaining atoms are learned using the PSL inference.

Given a set of atoms $\ell = \{\ell_1, \ldots, \ell_n\}$, an interpretation defined as $I : \ell \to [0, 1]^n$ is a mapping from atoms to soft truth values. PSL defines a probability distribution over all such interpretations such that those that satisfy more ground rules are more probable. PSL uses the *Lukasiewicz t-norm* and its corresponding co-norm as relaxations of the logical AND and OR respectively to determine the degree to which a ground rule is satisfied. Given an interpretation I, these relaxations of logical conjunction ($\wedge$), disjunction ($\vee$), and negation ($\neg$) are

$$\ell_1 \tilde{\wedge} \ell_2 = \max\{0, I(\ell_1) + I(\ell_2) - 1\},$$
$$\ell_1 \tilde{\vee} \ell_2 = \min\{I(\ell_1) + I(\ell_2), 1\},$$
$$\tilde{\neg} l_1 = 1 - I(\ell_1),$$

The probability distribution over interpretations is log linear with features defined by *distances to satisfaction*, which follows the Lukasiewicz t-norm to relax the fact that a rule $r \equiv r_{body} \to r_{head}$ is satisfied if and only if the truth value of head is at least that of the body. The rule's distance to satisfaction measures the degree to which this condition is violated.

$$d_r(I) = \max\{0, I(r_{body}) - I(r_{head})\}$$

PSL then induces a probability distribution over possible interpretations I over the given set of ground atoms l in the domain. If R is the set of all ground rules that are instances of a rule from the system and uses only the atoms in I then, the probability density function f over I is defined as

$$f(I) = \frac{1}{Z} \exp\left(-\sum_{r \in R} \lambda_r (d_r(I))^p\right) \quad (1)$$

$$Z = \int_I \exp\left(-\sum_{r \in R} \lambda_r (d_r(I))^p\right) \quad (2)$$

where λ_r is the weight of the rule r, Z is the continuous version of the normalization constant used in discrete Markov random fields, and $p \in \{1, 2\}$ provides a choice between two different loss functions, linear and quadratic. The values of the atoms can be further restricted by providing linear equality and inequality constraints allowing one to encode functional constraints from the domain.

PSL enables fast inference of the most probable explanation, i.e., the interpretation that is most likely in the PSL distribution given a partial interpretation with grounded atoms based on observed evidence. Because of the structure of PSL rules and their distances to satisfaction, this inference is always solvable via a convex optimization. Further, the structure of the probabilistic model enables this convex optimization to be solved very quickly using the alternating direction method of multipliers (ADMM) (Boyd et al. 2011; Bach et al. 2012; 2013). Using this approach, PSL enables our system to easily define a rich probabilistic model and perform inference with a lightweight computational cost.

3 Approach

The general approach we adopt is to identify open-source documents that appear to indicate civil unrest event planning, extract relevant information from identified documents and use that as the basis for a structured warning about the planned event (see Fig. 1).

Linguistic Preprocessing

A wide array of textual documents—RSS feeds (news and blogs), mailing lists, URLs referenced in tweets, the contents of the tweets themselves, and Facebook event pages—are subjected to shallow linguistic processing prior to analysis. This involves identifying the language of the document, distinguishing the words (tokenization), normalizing words for inflection (lemmatization), and identifying expressions referring to people, places, dates and other entities and classifying them (named entity extraction). Since our region of

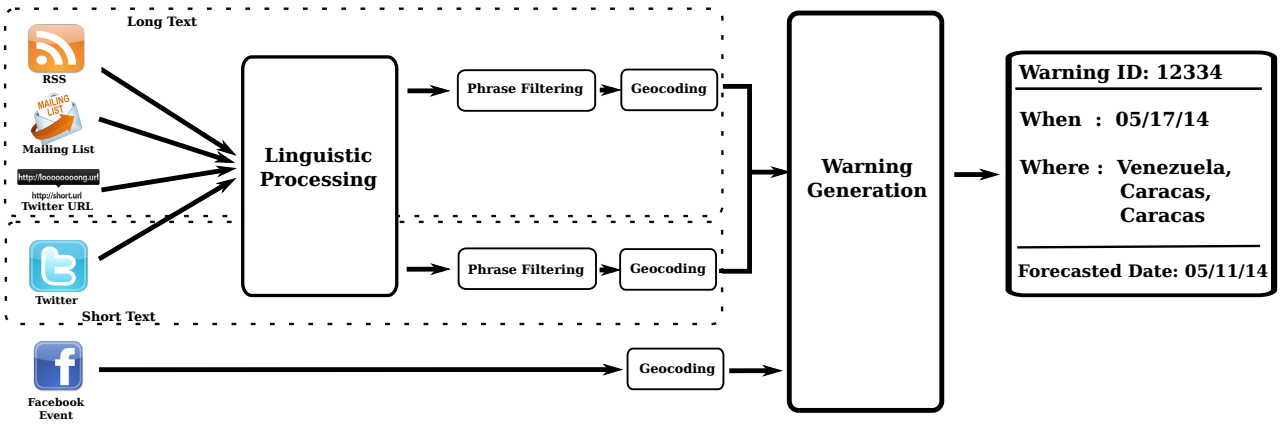


Figure 1: Schematic of the planned protest detector that ingests five different types of data sources.

interest is Latin America, the collection of text harvested is inherently multilingual, with Spanish, Portugese, and English as the dominating languages; we use Basis Technology's Rosette Linguistics Platform (RLP) suite of multilingual commercial tools (http://www.basistech.com/text-analytics/rosette/) for this stage. The output of linguistic preprocessing serves as input to subsequent deeper analysis in which date expressions are normalized and the geographic focus of the text identified.

Date processing is particularly crucial to the identification of future oriented statements. We use the TIMEN (Llorens et al. 2012) date normalization package to normalize and deindex expressions referring to days in English, Spanish and Portuguese. This system makes use of meta-data such as the day of publication, and other information about the linguistic context of the date expression to determine for each date expression, what day (or week, month or year) it refers to. For example in a tweet produced on June 10, 2014, the occurrence of the term *Friday* used in a future-tense sentence *We'll get together on Friday* will be interpreted as June 13, 2014. Each expression identified as a date by the RLP preprocessor is normalized in this way.

Phrase filtering

In order to identify relevant documents, input documents are filtered on a set of key phrases, i.e., the text of the document is searched for the presence of one or more key phrases in a list of phrases which are indicative of an article's focus being a planned civil unrest event. The list of key phrases indicating civil unrest planning was obtained in a semi-automatic manner, as detailed in the below section. Articles which do match are processed further, those that do not are ignored.

Phrase list development The set of key phrases was tailored (slightly) to the genre of the input. In particular different phrases were used to identify relevant news articles and blogs from those used to filter Tweets.

Initially, a few seed phrases were obtained manually with the help of subject matter experts. An analysis of news reports for planned protests in the print media helped create a minimum set of words to use in the query. We choose four nouns from the basic query that is used predominantly to

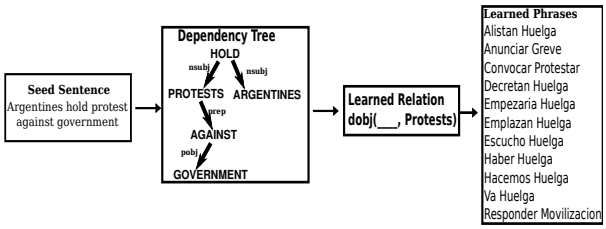


Figure 2: An example of phrase learning for detecting planned protests.

indicate a civil unrest in the print media - *demonstration, march, protest* and *strike*. We translated them into Spanish and Portuguese, including synonyms. We then combined these with future-oriented verbs, e.g., *to organize*, *to prepare*, *to plan*, and *to announce*. For twitter, shorter phrases were identified, and these had a more direct call for action, e.g., *marchar*, *manhã de mobilização*, *vamos protestar*, *huelga*.

To generalize this set of phrases, the phrases were then parsed using a dependency parser (Padró and Stanilovsky 2012) and the grammatical relationship between the core nominal focus word (e.g., *protest*, *manifestación*, *huelga*) and any accompanying word (e.g., *plan*, *call*, *anunciar*) was extracted. These grammatical relations were used as extraction patterns as in (Riloff and Wiebe 2003) to learn more phrases from a corpora of sentences extracted from the data stream of interest (either news/blogs or tweets). This corpus consists of sentences that contained any one of the nominal focus words and also had mentions of a future date. The separation threshold for a phrase was also learned, being set to the average number of words separating the nominal focus and the accompanying word.

The set of learned phrases is then reviewed by a subject matter expert for quality contraol. Using this approach, we learned 112 phrases for news articles and blogs and 156 for tweets. This phrase learning process is illustrated in Fig. 2.

Phrase matching Our key phrase matching is highly general and linguistically sophisticated. The phrases in our list are general rules for matching, rather than literal string sequences. Typically a phrase specification comprises: two or

more word lemmas, a language specification, and a separation threshold. This indicates that words—potentially inflected forms—in a given sequence potentially separated by one or more other words, should be taken to be a match. We determined that this kind of multi-word key phrases was more accurate than simple keywords for extracting events of interest from the data stream.

The presence of a keyphrase is checked by searching for the presence of individual lemmas of the keyphrase within the same sentence separated by at most a number of words that is fewer than the separation threshold. This method allows for linguistically sophisticated and flexible matching, so, for example, the keyphrase **[*plan protest*, 4, English]** would match the sentence *The students are planning a couple big protests tomorrow* in an input document.

Geocoding

After linguistic preprocessing and suitable phrase filtering, messages are geocoded with a specification of the geographical focus of the text—specified as a city, state, country triple—that indicates the locality that the text is about. We make use of different geocoding methodologies for Twitter messages, for Facebook Events pages, and for news articles and blogs. These are described below.

Twitter and Facebook For tweets, the geo-focus of the message is generated by a fairly simple set of heuristics involving i) the most reliable but least available source, i.e., the geotag (latitude, longitude) of the tweet itself, ii) Twitter *places* metadata, and iii) if the above are not available, the text fields contained in the user profile (location, description) as well as the tweet text itself to find mentions of relevant locations. Additional toponym disambiguiation heuristics are used to identify the actual referent of the mention. Similar methods are used to geocode event data extracted from Facebook event pages.

News and Blogs For longer articles such as news articles, the geo-focus of the message is identified using much more complex methods to extract the protest location from news articles, we use PSL to build probabilistic models that infer the intended location of a protest by weighing evidence coming from the Basis entity extractions and information in the World Gazeteer.

The primary rules in the model encode the effect that Basis-extracted location strings that match to gazatteer aliases are indicators of the article's location, whether they be country, state, or city aliases. Each of these implications is conjuncted with a prior for ambiguous, overloaded aliases that is proportional to the population of the gazetteer location. For example, if the string "Los Angeles" appears in the article, it could refer to either Los Angeles, California, or Los Ángeles in Argentina or Chile. Given no other information, our model would infer a higher truth value for the article referring to Los Angeles, California, because it has a much higher population than the other options.

$$ENTITY(L, location) \tilde{\wedge} REFERSTO(L, locID) \rightarrow FOCUSLOCATION(Article, locID)$$

$$ENTITY(C, location) \tilde{\wedge} IsCountry(C) \rightarrow ArticleCountry(Article, C)$$

$$ENTITY(S, location) \tilde{\wedge} IsState(S) \rightarrow ArticleState(Article, S)$$

(Note that the above are not deterministic rules; e.g., they do not use the logical conjunction $\wedge$ but rather the Lukasiewicz t-norm based relaxation $\tilde{\wedge}$. Further, these rules do not fire deterministically but are instead simultaneously solved for satisfying assignments as described in the Section 2)

The secondary rules, which are given half the weight of the primary rules, perform the same mapping of extracted strings to gazetteer aliases, but for extracted persons and organizations. Strings describing persons and organizations often include location clues (e.g., "mayor of Buenos Aires"), but intuition suggests the correlation between the article's location and these clues may be lower than with location strings.

$$ENTITY(O, organization) \tilde{\wedge} REFERSTO(O, locID) \rightarrow FOCUSLOCATION(Article, locID)$$

$$ENTITY(O, organization) \tilde{\wedge} IsCountry(O) \rightarrow ArticleCountry(Article, O)$$

$$ENTITY(O, organization) \tilde{\wedge} IsState(O) \rightarrow ArticleState(Article, O)$$

Finally, the model includes rules and constraints to require consistency between the different levels of geolocation, making the model place higher probability on states with its city contained in its state, which is contained in its country. As a post-processing step, we enforce this consistency explicitly by using the inferred city and its enclosing state and country, but adding these rules into the model makes the probabilistic inference prefer consistent predictions, enabling it to combine evidence at all levels.

$$FOCUSLOCATION(Article, locID) \tilde{\wedge} Country(locID, C) \rightarrow ArticleCountry(Article, C)$$

$$FOCUSLOCATION(Article, locID) \tilde{\wedge} Admin1(locID, S) \rightarrow ArticleState(Article, S)$$

4 Experiments

We evaluate our planned protest detection system using metrics similar to those described in Ramakrishnan et al. (Ramakrishnan et al. 2014). Given a set of alerts issued by the system and the GSR comprising actual protest incidents, we aim to identify a correspondence between the two sets via a bipartite matching. An alert can be matched to a GSR event only if i) they are both issued for the same country, ii) the alert's predicted location and the event's reported location are within 300km of each other (the distance offset), and iii) the forecasted event date is within a given interval of the true event date (the date offset). Once these inclusion criteria apply, the quality score (QS) of the match is defined as a combination of the location score (LS) and date score (DS):

$$\text{QS} = (LS + DS) * 2 \tag{3}$$

where

$$\text{LS} = 1 - \frac{\min(\text{distance offset}, 300)}{300} \tag{4}$$

and

$$\text{DS} = 1 - \frac{\min(\text{date offset}, \text{INTERVAL})}{\text{INTERVAL}} \tag{5}$$

Here, we explore INTERVAL values from 0 to 7. if an alert (conversely, GSR event) cannot be matched to any GSR event (alert, respectively), these unmatched alerts (and

events) will negatively impact the precision (and recall) of the system. The lead time, for a matched alert-event pair, is calculated as the difference between the date on which the forecast was made and the date on which the event was reported (this should not be confused with the date score, which is the difference between the predicted event date and the actual event date). Lead time concerns itself with reporting and forecasting, whereas the date score is concerned with quality or accuracy. On average the forecasted event date is within 1.19 days of the true event date and the forecasted location is within 45km of the true location.

We conduct a series of experiments to evaluate the performance of our system.

How does the distribution of protests detected by the system compare with the actual distribution of protests in the GSR? Fig. 3 reveals pie charts of both distributions. As shown, Mexico, Brazil, and Venezuela experience the lion's share of protests in our region of interest, and the protests detected also match these modes although not the specific percentages. The smaller countries like Ecuador, El Salvador, and Uruguay do experience protests but which are not as prominently detected as those for other countries; we attribute this to their smaller social media footprint (relative to countries like Brazil and Venezuela).

Are there country-specific selective superiorities for the different data sources considered here? Table 2 presents a breakdown of perfomance, country-wise and source-wise, of our approach for a recent month, viz. March 2014. It is clear that the multiple data sources are necessary to achieve a high recall and that by and large these sources are providing mutually exclusive alerts. (Note also that some data sources do not produce alerts for specific countries.) Between Twitter and Facebook, the former is a better source of alerts for countries like Chile and the latter is a better source for Argentina, Brazil, Colombia, and Mexico. News and blogs achieve higher recall than social media sources indicating that most plans for protests are announced in established media. They are also higher quality sources for alerts in countries like El Salvador, Paraguay, and Uruguay. Finally, note that news and blogs offer a much higher lead time (4.57 days) as compared to that for Facebook (2.44 days) or for Twitter (2.82 days). The quality scores are further broken down in Table 3 into their date and location components. A longitudunal perspective on quality scores is given in Fig. 4a. Note that in general Twitter tends to have a higher quality score as multiple re-tweets of future event mentions is a direct indicator of the popularity of an event as well as the intent of people to join an event. In contrast, mentions of future events in news do not directly shed any insight into popularity or people's support for the event's causes.

How did our system fare in detecting key country-wide protests? The recent Venezuelan protests against President Nicolas Maduro and the Brazilian Protests during June 2013 against bus fare hike were two significant protests during our period of evaluation. Fig. 4b and Fig. 4f describe our performance under these two situations illustrating the count of protests detected against the GSR. Notice that our system was able to identify the Venezuelan protests much better than the Brazilian protests. This is because there was a significant amount of spontaineity to the Brazilian protests; they arose as discontent about bus fare increases but later morphed into a broader set of protests against government and most of these subsequent protests were not planned.

What is the tradeoff between lead time and quality? Fig. 4c shows that the QS of the planned protest model decreases (as expected) with lead time, initially, but later rises again. The higher quality scores toward the right of Fig. 4c are primarily due to Facebook event pages.

How does the method perform under stringent matching criteria? Fig. 4d shows the perfomance of the model when the matching window is varied from 7 to 1 in steps. We can see that the performance degrades quite gracefully even under the strict matching interval of a 1-day difference.

What is the distribution of quality scores? The clear mode toward the right side of the Fig. 4e signifies that a majority of the planned protest alerts are of high quality. Further, the quality score distribution is unimodal suggesting that the careful reasoning of locations and date normalization are crucial to achieving high quality.

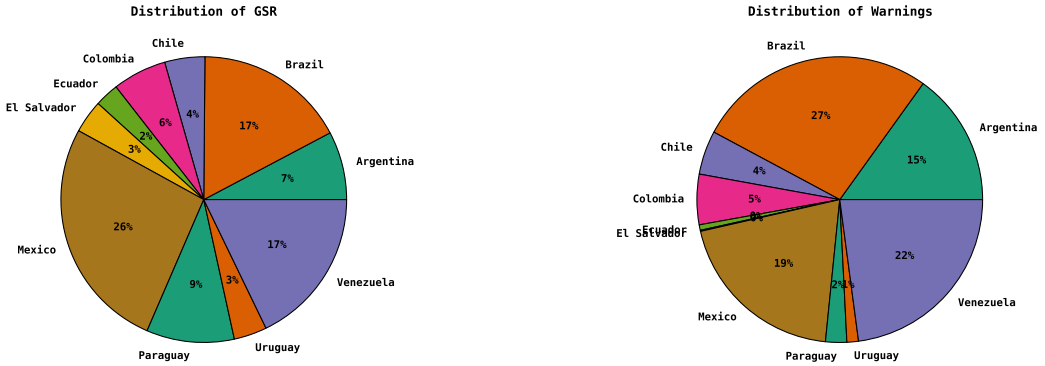
5 Development and Maintenance

The core algorithms behind the planned protest detector were implemented in Python. The PSL geocoder was implemented in Java. The Basis Rosette Linguistic Platform is the key external library utilized. The development process took 3 months (June 2012 to Aug 2012) and was primarily led by the first author with contributions from the other authors. After two months of testing (Sep 2012 and Oct 2012), the system was deployed in Nov 2012. Since there is not an explicit training phase, the system has required minimal reengineering over time. Key changes made to the system over time was to increase the sources used for data ingest and supporting the inclusion of additional phrases. Agile software engineering methods were used for project management.

6 Discussion

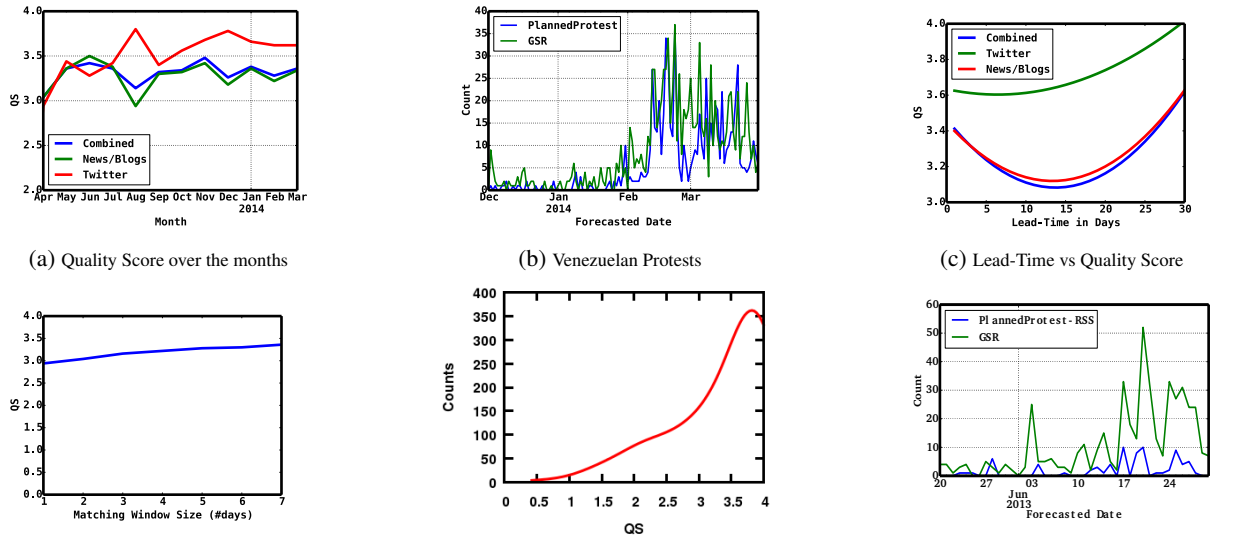
We have described an approach to forecasting protests by detecting mentions of future events in news and social media. The two twin issues of i) resolving the date and ii) resolving the location have been addressed satisfactorily to realize an effective protest forecasting system. As different forms of communication media gain usage, systems like ours will be crucial to understanding the concerns of citizenry.As stated in the introduction, there are many legitimate applications of a protest forecasting system such as ours; our goal is to provide advance warnings of protest occurrences to all relevant stakeholders.

Our future work is aimed at three aspects. First, to address situations such as nationwide protests and systems of



(a) GSR distribution from 2012-11 to 2014-03. (b) Alerts distribution from 2012-11 to 2014-03.

Figure 3: Distribution of alerts and GSR events across the countries studied in this paper.



(a) Quality Score over the months

(b) Venezuelan Protests

(c) Lead-Time vs Quality Score

(d) QS vs Matching Interval Trade-Off

(e) Quality Score Distribution

(f) System Performance during Brazilian Spring

Figure 4: Evaluation of planned protest forecasting system

Table 2: Country-wise breakdown of forecasting performance for different data sources. QS=Quality Score; Pr=Precision; Rec=Recall; LT=Lead Time. AR=Argentina; BR=Brazil; CL=Chile; CO=Colombia; EC=Ecuador;SV=El Salvador; MX=Mexico; PY=Paraguay; UY=Uruguay; VE=Venezuela. A – indicates that the source did not produce any warnings for that country in the studied period.

	News/Blogs				Twitter				Facebook				Combined			
	QS	Pr.	Rec.	LT	QS	Pr.	Rec.	LT	QS	Pr.	Rec.	LT	QS	Pr.	Rec.	LT
AR	3.14	0.32	0.69	3.94	3.52	**0.78**	0.14	3.14	**3.70**	0.50	0.04	3.00	3.02	0.36	**0.80**	**4.50**
BR	3.14	0.48	0.54	**5.85**	-	-	-	-	**3.62**	**0.76**	0.18	2.46	3.28	0.49	**0.65**	5.15
CL	3.06	0.91	0.67	5.40	**3.52**	**1.00**	0.23	4.29	-	-	-	-	3.16	0.83	**0.80**	**5.92**
CO	2.74	0.90	0.56	**7.44**	3.30	**1.00**	0.15	2.43	**4.00**	**1.00**	0.02	2.00	2.88	0.84	**0.65**	6.47
EC	-	-	-	-	**2.32**	**1.00**	**0.06**	**17.00**	-	-	-	-	**2.32**	**0.50**	**0.06**	**17.00**
MX	2.96	0.88	0.25	**3.69**	3.14	**1.00**	0.02	1.43	**3.72**	0.67	0.01	2.00	3.00	0.87	**0.27**	3.51
SV	**3.22**	**1.00**	**0.03**	**1.0**	-	-	-	-	-	-	-	-	**3.22**	**1.0**	**0.03**	**1.0**
PY	3.38	**1.00**	**0.16**	9.11	3.84	**1.00**	0.04	**11.40**	3.96	**1.00**	0.01	2.00	3.60	0.96	**0.20**	9.35
UY	**3.24**	**1.00**	**0.29**	**2.40**	-	-	-	-	-	-	-	-	3.24	**1.00**	**0.29**	3.24
VE	**3.80**	**1.00**	0.36	**3.27**	3.68	0.97	0.33	2.39	-	-	-	-	3.64	0.99	**0.69**	2.88
ALL	3.34	0.69	0.35	**4.57**	3.62	**0.97**	0.15	2.82	3.66	0.74	0.03	2.44	3.36	0.73	**0.51**	4.08

Table 3: Comparing the location and date scores of different sources in specific countries. AR=Argentina; BR=Brazil; CL=Chile; CO=Colombia; EC=Ecuador;SV=El Salvador; MX=Mexico; PY=Paraguay; UY=Uruguay; VE=Venezuela. A − indicates that the source did not produce any warnings for that country in the studied period.

Source		AR	BR	CL	CO	EC	SV	MX	PY	UY	VE	All
News/Blogs	LS	0.82	0.76	0.75	0.60	-	**0.75**	0.66	0.79	**0.79**	**0.95**	0.81
	DS	0.75	0.81	0.78	0.77	-	**0.86**	0.82	0.90	**0.83**	**0.95**	0.86
Facebook	LS	**1.0**	**0.92**	-	**1.00**	-	-	**0.86**	**0.98**	-	-	**0.93**
	DS	0.85	**0.89**	-	**1.00**	-	-	**1.00**	**1.00**	-	-	0.90
Twitter	LS	0.88	-	**0.84**	0.81	**0.45**	-	0.71	**0.98**	-	0.91	0.89
	DS	**0.88**	-	**0.92**	0.84	**0.71**	-	0.86	0.94	-	0.93	**0.92**

protests, we must generalize our system from generating protests at a single article level to digesting groups of articles. This will require more sophisticated reasoning using PSL programs. Second, we would like to generalize our approach that currently does detection of overt plans for protest to not-so-explicitly stated expressions of discontent. Finally, we plan to consider other population-level events of interest than just civil unrest, e.g., domestic political crises, and design detectors to recognize the imminence of such events.

Acknowledgments

Supported by the Intelligence Advanced Research Projects Activity (IARPA) via DoI/NBC contract number D12PC000337, the US Government is authorized to reproduce and distribute reprints of this work for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the US Government.

References

Bach, S.; Broecheler, M.; Getoor, L.; and O'leary, D. 2012. Scaling mpe inference for constrained continuous markov random fields with consensus optimization. In *Advances in Neural Information Processing Systems*.

Bach, S. H.; Huang, B.; London, B.; and Getoor, L. 2013. Hinge-loss Markov random fields: Convex inference for structured prediction. In *UAI*.

Baeza-Yates, R. 2005. Searching the future. In *SIGIR Workshop on Mathematical/Formal Methods in Information Retrieval*.

Banko, M.; Cafarella, M. J.; Soderland, S.; Broadhead, M.; and Etzioni, O. 2007. Open information extraction from the web. In *IJCAI*.

Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; and Eckstein, J. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*.

Chambers, N., and Jurafsky, D. 2011. Template-based information extraction without the templates. In *HLT*.

Compton, R.; Lee, C.; Lu, T.-C.; De Silva, L.; and Macy, M. 2013. Detecting future social unrest in unprocessed twitter data:"emerging phenomena and big data". In *ISI*.

Gabrilovich, E.; Dumais, S.; and Horvitz, E. 2004. Newsjunkie: Providing personalized newsfeeds via analysis of information novelty. In *WWW*.

Hurriyetoglu, A. H.; Kunneman, F.; and van den Bosch, A. 2013. Estimating the time between twitter messages and future events. In *Dutch-Belgian Workshop on Information Retrieval*.

Jatowt, A., and Au Yeung, C. m. 2011. Extracting collective expectations about the future from large text collections. In *CIKM*.

Kallus, N. 2014. Predicting crowd behavior with big public data. In *WWW*.

Kawai, H.; Jatowt, A.; Tanaka, K.; Kunieda, K.; and Yamada, K. 2010. Chronoseeker: Search engine for future and past events. In *ICUIMC*.

Kimmig, A.; Bach, S.; Broecheler, M.; Huang, B.; and Getoor, L. 2012. A short introduction to probabilistic soft logic. In *NIPS Workshop on Probabilistic Programming: Foundations and Applications*.

Llorens, H.; Derczynski, L.; Gaizauskas, R. J.; and Saquete, E. 2012. TIMEN: An open temporal expression normalisation resource. In *LREC*.

Mani, I., and Wilson, G. 2000. Robust temporal processing of news. In *ACL*.

Padró, L., and Stanilovsky, E. 2012. Freeling 3.0: Towards wider multilinguality. In *LREC*.

Pustejovsky, J.; Castano, J. M.; Ingria, R.; Sauri, R.; Gaizauskas; et al. 2003. Timeml: Robust specification of event and temporal expressions in text. *New directions in question answering*.

Radinsky, K., and Horvitz, E. 2013. Mining the web to predict future events. In *WSDM*.

Ramakrishnan, N.; Butler, P.; Muthiah, S.; Self, N.; Khandpur, R.; Saraf, P.; et al. 2014. 'beating the news' with embers: Forecasting civil unrest using open source indicators. In *SIGKDD*.

Riloff, E., and Wiebe, J. 2003. Learning extraction patterns for subjective expressions. In *EMNLP*.

Tops, H.; van den Bosch, A.; and Kunneman, F. 2013. Predicting time-to-event from twitter messages. In *BNAIC*.

Truvé, S. 2011. Big data for the future: Unlocking the predictive power of the web. *Recorded Future, Cambridge, MA, Tech. Rep*.

Verhagen, M.; Gaizauskas, R.; Schilder, F.; Hepple, M.; Moszkowicz, J.; and Pustejovsky, J. 2009. The tempeval challenge: identifying temporal relations in text. *Language Resources and Evaluation*.

Xu, J.; Lu, T.-C.; Compton, R.; and Allen, D. 2014. Civil unrest prediction: A tumblr-based exploration. In *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer.