Predicting Political Instability and Social Conflicts Using Multimodal Data

Nishith Khandwala
nishith@stanford.edu

Amanda Chow
amdchow@stanford.edu

Cheng Chen
cchen91@stanford.edu

Abstract

We propose a modular framework for predicting local violence with high temporal and spacial resolution. First, the region is divided into a grid. Then, given historical data on conflicts, weather, etc. in each cell over time, the model predicts the likelihood of conflict at every cell in the next time step. It captures spatial and temporal patterns in conflict occurrences using convolutional and recurrent neural networks. To account for a lack of available time-variant data, the model also allows for time-invariant data. This modularity allows for easy incorporation of different kinds of data. Our model ultimately achieves a significantly higher F1 score compared to existing models.

1. Introduction

The history of mankind is riddled with wars, violence, and protests over injustices. Areas recently afflicted with such violence include Egypt, Syria, and Qatar, to name a few. These conflicts have disastrous effects on civilian lives, and many have wondered whether it is possible to foresee these conflicts in advance.

In recent years, technology has provided a virtual space for people in areas of conflict to express their opinions publicly and raise attention to local violence. Technology has also improved record-keeping on past conflicts, with more comprehensive coverage of local violence. These conflicts do not simply occur overnight; they are usually preceded by periods of unrest accumulating over time. Building on this premise, we ask — given the increasing amount of historical records, is it possible predict the onset of an uprising?

Our primary goal is to design a model that is able to identify preceding factors into societal uprisings, and make short-sighted predictions with good temporal and spacial resolution. Constructing reliable prediction models for social instability is absolutely crucial; it would enable policymakers to detect trends and respond with the right political, economic, and developmental sanctions. Doing so can potentially stop unnecessary and violent military interventions, and prevent conflicts and government collapse.

2. Related Work

There are several recent works that apply machine learning techniques to conflict prediction [1, 2, 3, 4]. The majority of these methods operate on national-level data [1, 2, 4], losing all higher-resolution, sub-national spatial information. One method feeds conflict maps and economic statistics (e.g. GDP) into Naive Bayes and random forests [1], while another uses Latent Dirichlet allocation on newspaper articles to find patterns in trending topic drift preceding conflicts [2]. The most spatially granular and detailed data set was at the city-level (e.g. demographics), collected via survey and used by Blair et. al [3]. Unfortunately, the collected data is not open to the public. Additionally, the temporal resolution of their forecast is not very high; predictions on a yearly basis are not very useful.

These conflict prediction models do not provide useful predictions, mainly due to their low spatio-temporal granularities. In general, they are missing several important features:

- Combining both spatial and temporal data at granular levels
- Accurate predictions of conflict at these granular (sub-national) levels
- Robust predictions without detailed surveys
- Expandability to allow for different data types

These are ideal features for a useful conflict prediction model to have.

3. Approach

We aim to design a model that solves all of the above issues. In short, we aim for a highly modular framework that achieves high prediction accuracy at granular spatial and temporal levels.

3.1. Problem Statement

To study conflict at higher spatial resolution, we focus on a single country with sufficient data. We chose Uganda, due to its relatively large numbers of conflict (hence, its data set is less sparse). There is also other publicly available information for Uganda. We grid the area of Uganda into 0.5 degree latitude/longitude square cells (each approximately 50km×50km), to create an 11×11 grid covering Uganda. Each grid cell will contain its own conflict prediction.

To study conflict at a higher temporal resolution, we aim for a reasonable time scale of one month. That is, we take the last several months of data and use it to make predictions in the next month.

The question now becomes — given a snapshot of the grid in each of the last four months, can we predict the likelihood of conflict in each grid cell at the next month?

3.2. Desired Predictive Powers of Model

Spatio-temporal Data
We would like our model to incorporate both spatial and temporal factors into predictions, because the likelihood of conflict in an area is likely influenced by (1) the area's past conflict history and (2) the conflict histories of neighboring areas.

A natural way to account for the spatial influence from neighboring areas is a technique commonly used in computer vision called the Convolutional Neural Network (CNN). CNN's capture local patterns in two-dimensional grids and can providing us a spatially-aware encoding of the grid at a single time-step (one month).

Given these spatially-aware encodings at each time step, a natural way to account for such history is to use a Recurrent Neural Network (RNN) variation called the Long Short-Term Memory network (LSTM). The encoding inputs at each time-step are spatially-aware, and the RNN adds temporal awareness. Thus, the RNN output accounts for both spatial and temporal factors. The final output of our model is a prediction grid, with each cell containing the probability of conflict occurring at the next month.

Time-Invariant Data
This model is very ideal, and is contingent on the availability of spatially- and temporally-comprehensive data sets with points at every grid cell at every time step. In reality, very few data sets are at sub-national levels with precise geo-coordinates, let alone spanning across time. We want our framework to be highly modular and allow the inclusion of spatially granular data that does not vary over time.

To incorporate such time-invariant data, we still feed the grid of data to a CNN to capture spatial patterns. However, there is no need to feed the only grid (the only time-step) to an LSTM. Thus, we can concatenate the time-invariant data from this CNN output with the time-variant data from the RNN output. This is the final encoding that is used to obtain the grid of conflict probabilities.

4. Data

As mentioned, this framework allows for both time-variant and time-invariant data, as long as there is sufficient spatial resolution. The most directly related data set to is a historical record of conflicts, which is our main data set.

The following data sets, though not direct records of conflict history, are closely related to conflict occurrences.

- Climate history
- Poverty data
- Newspaper articles [2]
- Public Facebook and Twitter activity (around 3 million tweets for the Arab Spring)
- The Fragile States Index (FSI) list [3]
- The World Bank data set: 1300 indicator variables for each country mapped in FSI [3]
- The High-risk countries database [4]
- Inflation, in particular, changes in food prices

However, despite this plethora of related data sets, only climate and poverty information are currently available at the sub-national level, with accurate latitude and longitude geo-tags. The other data are either not geo-tagged (e.g. Facebook post), or by convention usually reported as a whole country (e.g. High-risk countries database).

As our goal is to design a model with higher temporal and spatial resolution, we will only consider data sets with sufficient resolution. Therefore, we will use a conflict history record as the main data set, while incorporating climate data and poverty data into our model.

4.1. Conflict History (Time-Variant)

The historical record of conflicts is our main data set, which we obtain from the Armed Conflicts and Event Data Project (ACEDP) [5]. This project collects the latitude/longitude coordinates, date of occurrence, casualty count, etc. of recorded conflicts in African and Asian countries from 1997 to 2015. Again, our country of focus is Uganda, so we filter this data to only include past conflicts in Uganda. In this time frame, Uganda witnessed a total of 4611 conflicts. Figure 1 shows an aggregate map of conflicts over all 19 years of the dataset, overlaid on the grid.

The ACEDP offers conflict data over 19 years, and our time scale is one month. At each time step, conflicts are bucketed into their appropriate grid cell within the 11×11 grid. For each cell, we accumulate a total number of conflicts and total number of deaths at the particular time step. Thus, the conflict-data input at each time step consists of an 11×11×2 grid.

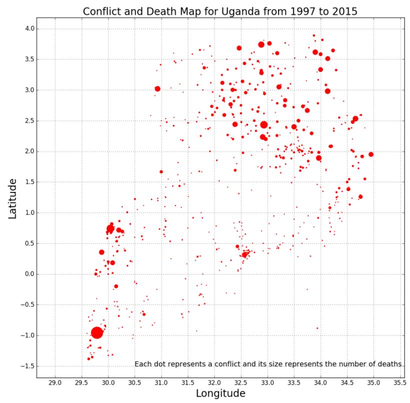

Figure 1. Conflict and death map of Uganda between 1997 and 2015: each red dot represents a conflict, and its radius is proportional to number of deaths during the conflict.

4.2. Climate data (Time-Variant)

The next data set with sufficient spatial granularity is climate information, which we obtained from the University of Delaware Air Temperature & Precipitation [6]. This data set includes very high resolution climate information (e.g. such as temperature and precipitation) for the entire earth.

We grid the information to align with our spatial scale (the 11×11 grid) and our time scale (for every month from 1997 to 2014). This data set cuts off at 2014, so we truncate our other data there as well, for a total of 216 time-steps. At each time step, each grid cell will contain an average temperature and precipitation. Thus, our climate-data input at each time step is a 11×11×2 grid. Figure 2 visualizes the average temperatures of each grid cell in Uganda over the time span of the data set.

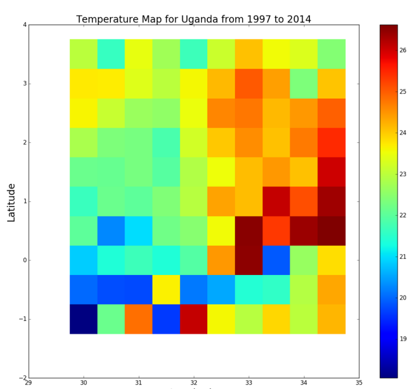

Figure 2. Average temperature map of Uganda between 1997 and 2014

4.3. Poverty data (Time-Invariant)

The final data set with sufficient spatial granularity was taken from a poverty mapping project covering several countries in Africa [7]. The data set offers information at many geo-coordinates, including:

- Number of households
- Urban index (indicator variable)
- Average nighttime intensity
- Average household consumption
- Average number of rooms per household
- Average household asset index
- Elevation
- Distance to local administrative headquarter
- Distance to nearest market
- Distance to major road
- Distance to population center of larger than 20,000 people

Once again, we geo-bucket these points into our 11×11 grid. and average the above features for each cell. This data set does not span across time, so it is a time-invariant data set. It is ultimately fed into our model as a 11×11×11 grid (the third dimension is for the above features of each grid cell).

5. Model

5.1. Architecture

In this section, we propose an end-to-end model [1] that leverages deep learning to extract spatial and temporal features from the input datasets and output conflict predictions across time.

We have three sets of data - past conflicts, weather maps and poverty indicators. The first two are time-variant i.e. the data sets includes samples from multiple time steps. The last data set, poverty indicators, is time-invariant. Our proposed model deals with each of the data sets separately to fully exploit the limited data we were able to find and enhance robustness of its predictions.

[1]Code is publicly available at
`https://github.com/nishithbsk/ConflictPrediction`

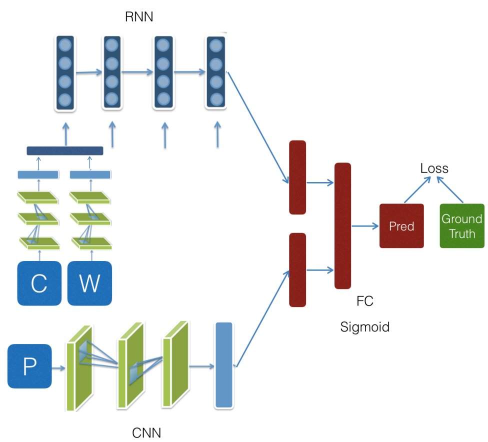

Figure 3. Proposed Model

For time varying data, we use a LSTM network, known for working well with sequences, to capture high level representations across time steps. While each time varying dataset could be modelled using a separate LSTM, we chose to use a single one-layer LSTM for all of them in the interest of decreasing model complexity. Thus, at each time step, the LSTM is fed with concatenation of features from the conflicts history and weather datasets. In our model, the number of time steps is fixed at 4, each representing a month and the hidden size of the LSTM is 128. We use Rectified Linear Units (ReLU) between time steps. The final hidden state of the LSTM is considered to be the output of this sub-model.

We pass the the time varying grids through distinct CNNs before feeding them into the LSTM. This enables the model to use spatial context to understand the spread of conflicts and make stronger predictions. Note that these CNNs do not share weights among themselves.

For the time-invariant poverty indicators dataset, we use a CNN (again, with the same architecture as the above CNNs) to obtain a high level representation, which is then used in conjunction with the LSTM output.

In all these cases, the CNN consists of two convolutional (conv) layers followed by a fully connected (FC) layer. The first conv layer has a filter size of 3x3 and depth 2 and is used with stride 1. The second has a filter size of 5x5 and depth 5 and is also used with stride 1. ReLU is applied after each layer. The size of the fully connected layer is different among datasets. The CNN corresponding to the conflicts history dataset has a FC layer with 128 hidden units. For the weather and poverty datasets, the FC layer has 64 units each. No non linearity is applied to FC layer, making the output of the CNN an encoding of the input data.

Finally, given the outputs of the time-variant LSTM and time-invariant CNN, we concatenate them and feed them into a last fully connected network of 2 hidden layers of sizes 256 and 121. Note that the output of this network is of the same shape as our ground truth grid (11×11) but flattened. While we use ReLU after the first layer, we use the sigmoid function on the second and final layer. We do so because we would like our output 'grid' to consist of real values between 0 and 1 representing the probability of a conflict.

Batch normalization is applied wherever applicable and unless specified otherwise. We discuss training details in the next section.

5.2. Optimization

We use backpropagation to train our model. The error signal is propagated through the entire model, starting at the fully connected network, going through the LSTM and time-invariant CNN and finally through the CNNs applied on time-variant data.

Given that our dataset is rather small compared to those used traditionally in deep learning, we opted for stochastic gradient descent (SGD) over mini-batch SGD to increase the number of updates per epoch. This optimization algorithm was used to minimize L2 norm loss between the prediction and ground truth grids. We also experimented with other losses, including the softmax cross entropy loss and found similar results (detailed in the next section).

We note here that we applied masks to zero-out the backpropagation error signal from any grid cell that shows zero activity in the conflicts history and poverty dataset. This helps the model to weed out any data points that might induce a strong irrelevant bias. In the case of Uganda, which is bordered by water on a side, anything learnt from conflict prediction on grid cells covering the sea is ignored.

We trained our network for 200 epochs with a learning rate of 1e-4 on a NVIDIA Tesla K80 GPU using the TensorFlow framework.

6. Results and Analysis

Since this paper presents a new study on a ensemble of datasets, we benchmark our results against logistic regression. For logistic regression, the input conflict grid is flattened and fed into the model cell by cell. Note that this baseline only considers the conflicts history dataset due to its incapacity to work with multimodal data, giving our method a significant modelling advantage. We also mention results from a model that predicts conflict and no-conflict with equal probability to establish an absolute benchmark.

6.1. Evaluation Metrics

Given the indeterminate nature of this problem, it is crucial to find appropriate metrics to evaluate our model. The output of our model is an 11×11 grid of binary grid of predictions, which is compared to an 11×11 binary grid of ground truths.

Accuracy
A traditional metric is to do an element-wise comparison of the predictions and ground truth, with accuracy as the number of correct predictions. However, because the ground truth matrix is very sparse, even a model that predicts all 0's will have 90% accuracy.

Precision
Precision is calculated as the number of true positives divided by the total number of positive predictions. In this context, it measures the model's ability to predict no-conflict regions as 0's instead of 1's. However, precision alone is not a sufficient metric, as it focuses is on no-conflict regions, rather than conflict regions.

Recall
Recall is calculated as the number of true positives divided by the number of ground truth positives (true positives and false negatives). It measures the ability of our model to predict regions of conflict as 1's, solving the problem with using precision as a metric. However, recall alone does not suffice, as a model that predicts all 1's achieves a perfect precision of 1.

F1 Score
Precision and recall both have their values, but neither metric alone provides a thorough evaluation. The F1 score is a metric that tries to balance precision and recall via a harmonic mean (P = precision, R = recall):

$$F1 = \frac{2PR}{P + R}$$

For this reason, F1 score is an appropriate metric for evaluating our model.

6.2. Results

A comparison of our model's accuracy is shown in Figure 4, relative to the random and baseline predictors. Our model has a higher accuracy than logistic regression, but only slightly. At first, it was surprising to see logistic regression performing well on such a hard task. However, a closer inspection revealed that the baseline model predicted almost all 0s, exploiting the sparse nature of our ground truth and input data. This highlights the flaws of using accuracy as a metric for our problem.

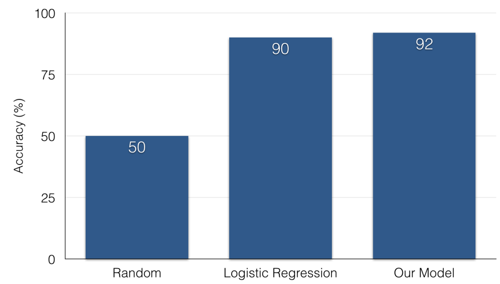

Figure 4. Accuracy Comparison

Nevertheless, accuracy served as a good first pass for any prediction model - it showed that the model learnt. It just was not a useful metric for comparing models.

Using the F1 score as our final metric, a comparison is shown in Figure 5. our model significantly outperforms logistic regression.

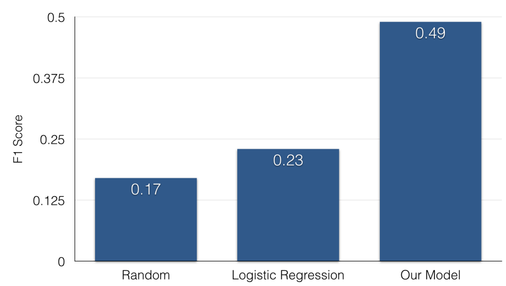

Figure 5. F1 Score Comparison

The bar graph above shows that F1 is a more meaningful metric to distinguish between models. Our proposed method obtains an F1 score of 0.49 overshadowing logistic regression at 0.23.

6.3. Analysis

While the previous section established that our model was sophisticated enough to capture spatio-temporal patterns in these extremely sparse datasets, we have not yet analyzed which of the available predictive indicators were more correlated with conflicts.

We trained subsets of our model under the same training details and found that both poverty and weather datasets were beneficial to our predictive power. Without them, our model obtained a F1 score of 0.41. While the poverty dataset was more helpful than the weather dataset on a standalone addition to the conflicts history dataset (0.47 vs. 0.45 F1 score), the three datasets together made the best model - achieving a F1 score of 0.49.

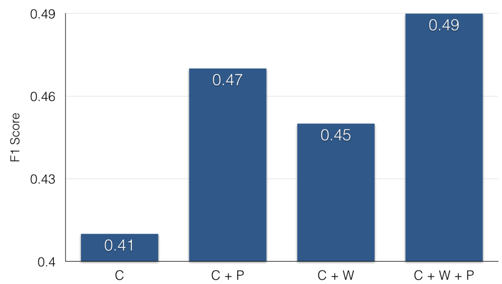

Figure 6. Feature Comparison

In the above bar graph, C refers to the conflicts history dataset, P refers to poverty indicators dataset and W refers to the weather dataset. The $+$ sign means using the operand datasets in conjunction.

7. Conclusion

In this paper, we presented a robust sub-national conflict prediction model that is capable of working with multi-modal input data, leveraging spatio-temporal features and is country agnostic. We compared our model against baseline models and observed that our model performs significantly better in terms of F1 score. Finally, we conclude by presenting an analysis that showed that the poverty indicators are useful in predicting conflicts and together with weather data, they make a powerful model.

8. Future Work

Although our proposed model has a good predictive power, it is by no means perfect. There still exists enormous scope for including more data sources, increasing the model robustness and performing further analysis on different countries.

We would like to see how our model performs on transfer learning. An interesting test would be to test our current trained model on different countries and check the robustness of our architecture in multiple settings.

Also, we would like to explore ways to incorporate text data such as newspapers [2] and social media (Twitter/Facebook posts) at a sub-national scale in our model. This would help us in predicting the scale of a future conflict. We leave this for future work.

Acknowledgements

We would like to thank Prof. Stefano Ermon and Prof. Marshall Burke for their incredible support and guidance for the project. We would also like to extend gratitude to Prof. Silvio Savarese and his group, Computational Vision and Geometry Lab (CVGL) for allowing us to use their compute power.

References

[1] C. Perry, "Machine Learning and Conflict Prediction: A Use Case," *Stability: International Journal of Security & Development*, vol. 2, no. 3, pp. 1–18, 2013.

[2] H. Mueller and C. Rauh, "Reading Between the Lines: Prediction of Political Violence Using Newspaper Text," *Cambridge-INET Working Paper Series*, 2016.

[3] R. Blair, C. Blattman, and A. Hartman, "Predicting Local Violence," *Available at SSRN 2497153*, 2014.

[4] B. Huffman, E. Marriott, and A. Yu, "Predicting High-Risk Countries for Political Instability and Conflict," *Stanford University CS 229 Project*, 2014.

[5] P. Wallensteen and T. Pettersson, "UCDP/PRIO Armed Conflict Dataset."

[6] C. J. Willmott and K. Matsuura, "Terrestrial air temperature and precipitation: Monthly and annual time series," 2015.

[7] M. Xie, N. Jean, M. Burke, D. Lobell, and S. Ermon, "Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping," *In Proc. 30th AAAI Conference on Artificial Intelligence*, 2016.