User-Driven Predictive Visual Analytics on Multivariate, Spatio-Temporal Incident Reports

Jillian Aurisano*, Miriam Cha† and Michael Snyder†

Abstract— In this paper, we develop and implement approaches to user-driven predictive visual analytics on multivariate, spatio-temporal incident reporting data on the Lords Resistance Army (LRA) activity in Central Africa. We concentrate on specific predictive questions that pivot on LRA movement and cause-and-effect patterns. Our approach represents the output of simple models, such as a movement model and a local-related-incidents model. In addition, we provide users with interactive features to explore, which allows for the isolation of appropriate subsets of the data and visually separates signal from noise when addressing predictive questions. We implement this approach and use it to develop predictive hypotheses around variations in movement patterns of the LRA and to identify potential cause-and-effect patterns around the activity of commanders and solider defections.

Index Terms—Visual analytics, prediction, spatio-temporal data

1 Introduction

The proliferation of diverse open source intelligence (OSINT) enables many new capabilities for geospatial analysis from a military as well as a humanitarian perspective. These capabilities include locating enemy encampments, vulnerability assessment, and determination of safe refugee areas. With the increase in sensing resources and ability to "crowd source" geospatial data, the need for spatio-temporal visualization tools that can support analysts and decision makers is as relevant as ever.

There have been multiple organizations willing to share spatio-temporal humanitarian data for further exploration [14, 10, 6]. As a recent example of a real world effort to find Joseph Kony, an insurgent leader of the Lords Resistance Army (LRA), Invisible Children and Resolve published reports of LRA attacks that were gathered through Invisible Children's Early Warning Network of radios implemented in the affected region, in addition to research conducted by the United Nations and international organizations working on the ground. These reports were made available to decision makers and the general public on an interactive mapping platform, the LRA Crisis Tracker [9]. The LRA Crisis Tracker map illustrates the basic tradecraft of spatio-temporal intelligence analysis: large spatio-temporal data is queried based upon set of criteria (such as incident type and time) in order to compare quantitative measures of different incidents in the affected region. Although this tool effectively communicates the gravity of the situation on the ground, and serves to persuade the public to get involved, it is not designed to enable analysis that facilitates planning or post-event responses for affected individuals.

While prediction is a priority for groups in these regions, predictive analysis is complicated by several factors. Conflict zones are chaotic, and aggressive actors behave unpredictably to evade capture and surprise victims. Information about the group may be limited and it can be difficult to capture the relevant data needed to generate predictive models. This problem is therefore analogous to the task of modeling multiple mixed signals. Since there are an unknown and dynamically changing number of signals, inconsistently sampled and dominated by irregularity, predictive modeling using automated methods alone is extremely difficult in this domain. Finally, since decisions made in these contexts may have life-or-death consequences, keeping humans in the loop is vital.

In this paper, we describe an approach to user-driven predictive visual analytics on multi-variate, spatio-temporal conflict reporting data. We implemented this approach using open-source reporting data on the activity of the LRA in Uganda, South Sudan, the Democratic Republic of the Congo (DRC) and the Central African Republic (CAR), collected by Invisible Children and Resolve from December 2007 to May 2014. This violent rebel group has had a significant impact on the region, both killing and abducting thousands of civilians as well as disrupting and displacing communities in the region. Our data included dates and locations of the incidents, a note about the incident, categorization of the incident type (eg. abduction, looting, violence), as well as quantitative information about who was affected.

We concentrate on specific predictive questions that pivot on LRA movement and cause-and-effect patterns. The goal is to represent the output of simple models and provide users with the ability to apply filters to isolate appropriate subsets of the data, thus visually separating signal from noise. In addition, to enable better decisions, our approach is designed to facilitate interactive exploration and drill-down into the data, to help users develop internal predictive reasoning, such as intuition about actions and outcomes, which may supply answers to questions out of reach of algorithmic approaches.

2 Related Work

Spatial data visualization tools allow for representing data on a map, via heat maps or chloropleth maps (GeoDa, ArcView) along with statistical analysis frameworks ([4, 15]). Geospatial and temporal visualization approaches have been studied extensively, especially in the context of data analysis and predictive modeling ([1, 13, 3]). However, many of these approaches were developed on data with trajectories or unique identifiers for discrete individuals, allowing for the estimation of paths and movement. Additionally, a variety of tools have been created to visualize crime data or the output of predictive models ([12, 5, 8, 7, 11]). Unlike our domain, this type of data encompasses a much larger number of incidents, which results in different options for data analysis and representation. In addition, none of these approaches represent movement prediction models or related incident models, as in our tool.

3 Modeling, visualizing and exploring movement

Deciding where to direct resources following an incident is a pressing question that depends on using past movement data to estimate where

** Jillian Aurisano is with the Electronic Visualization Laboratory at the University of Illinois at Chicago, 851 S Morgan, Room 1120, Chicago, IL 60607. E-mail: jauris2@uic.edu*

† Miriam Cha and Michael Snyder are with MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420. E-mail: {miriam.cha, michael.snyder}@ll.mit.edu

This work is sponsored by the Assistant Secretary of Defense for Research and Engineering under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the United States Government.

Proceedings of the 2014 IEEE VIS Workshop on Visualization for Predictive Analytics. Available online at: http://predictive-workshop.github.io

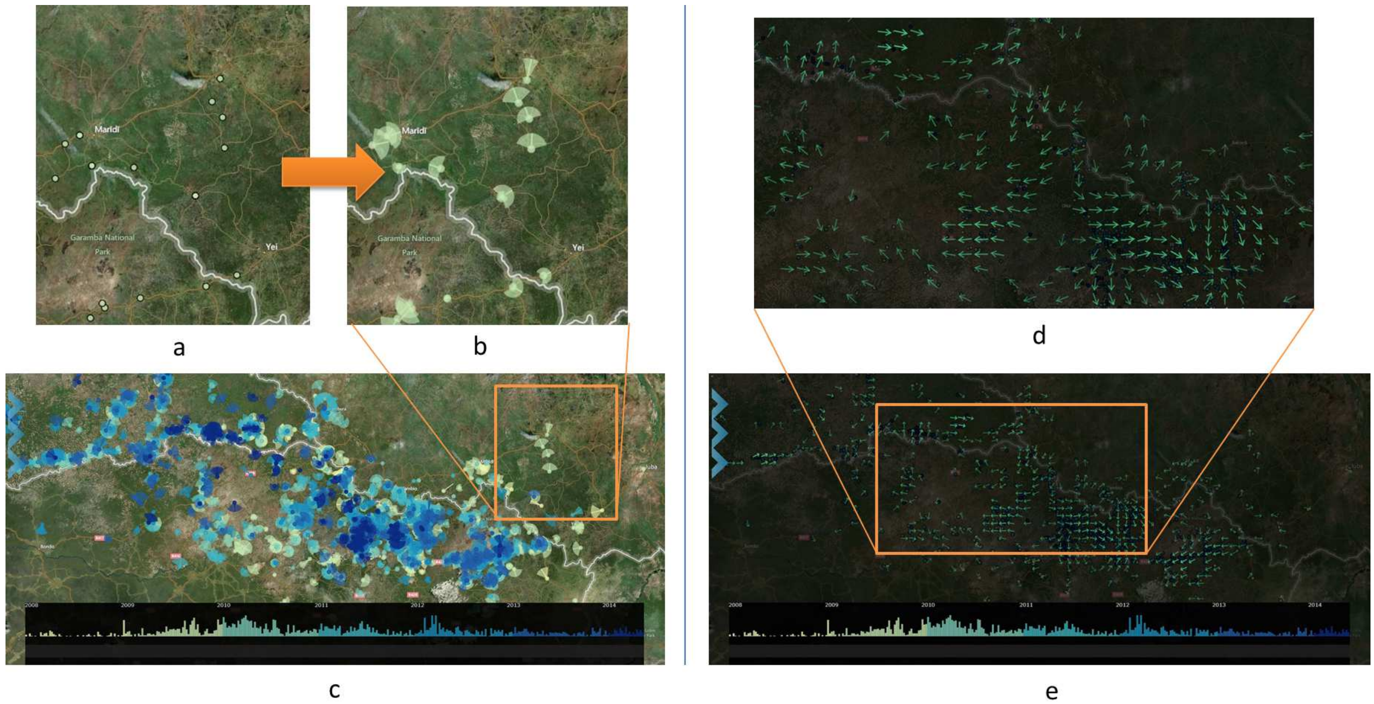


Fig. 1. a) Incidents plotted without vectors. b) Incidents, from the greater picture in subfigure c, transformed to show movement through a vector indicating net direction from temporally-local incidents, and an arc indicating average deviation from the computed vector. d) A subset of a flow field, from the greater picture in subfigure e, that presents incident progression over time.

and when future incidents may take place. Ideally, this analysis could be supported visually through an interface that would show data about past LRA movement. However, the raw data contains only incident locations in time and space, without information to definitively connect incidents together and determine movement patterns. Initially, we developed approaches that represented the raw data, in the form of points, and allowed users to interactively select subsets of the data or animate through different time periods to visually estimate movement patterns. Unfortunately, this approach alone made estimations of movement difficult, particularly when attempting to form generalizations over the entire region and time period. This observation is consistent with prior research indicating that animation is effective for observing general trends, but is less effective for in-depth investigations over multiple time points ([2]).

3.1 Movement modeling

To gain movement estimations from past incident data, we adopted an approach that first assumed that proximate sets of incidents in space and time may be considered related. Using this assumption, we model movement by computing a vector for each incident, weighted by distance in space and time along a Guassian distribution, using the following equation:

$$\vec{v}_i = w_i \sum_{j=0,i\neq j}^{n} \frac{\vec{x}_j - \vec{x}_i}{\left|(\vec{x}_j - \vec{x}_i)\right|} \exp\left(-\frac{\left|\vec{x}_j - \vec{x}_i\right|^2}{\sigma_d^2 \Delta t} - \frac{\Delta t^2}{\sigma_t^2}\right) \tag{1}$$

where

$$w_i = \sum_{j=0,i\neq j}^{n} \exp\left(-\frac{\left|\vec{x}_j - \vec{x}_i\right|^2}{\sigma_d^2 \Delta t} - \frac{\Delta t^2}{\sigma_t^2}\right) \tag{2}$$

Here $\vec{x}_i$ and $\vec{x}_j$ are the positions of incidents and Δt is the time between incidents in days, with σ_d set to 50 miles and σ_t set to 2 weeks.

The computed incident movement vector represents an answer to the following questions about past data: given an incident with close and proximate future incidents, where did the next incident take place, and second, given an incident with close and proximate past incidents, from where did previous incidents come?

We also determined the spread of subsequent incidents, by computing the average angle (Θ_i) of deviation from the computed vector ($\vec{v}_i$), where $\vec{d}_{ij}$ is a direction vector between incidents and w_i is the weight computed in (2):

$$\theta_i = \frac{2}{w_i} \sum_{j=0,i\neq j}^{n} \arccos\left(\frac{\vec{v}_i \cdot \vec{d}_{ij}}{\left|\vec{v}_i\right| \left|\vec{d}_{ij}\right|}\right) \tag{3}$$

The result is a large angle for incidents with distributed movement patterns and a narrow angle for incidents with tightly directed movement patterns.

3.2 Representation of movement model output

To represent this information initially, we plotted the vectors and arcs alongside the data. Users can toggle the vectors and arcs, can select subsets of time and can animate to see events unfolding (Figure 1).

Users noted that it was easier to infer movement when, instead of interactively exploring, each incident was represented with a vector and an arc in a static timeframe. In addition, users noted that animations were easier to follow with this information displayed. We believe this may occur because the positions of subsequent incidents are indicated through the vectors, which effectively directs the user's attention.

3.3 Aggregation

While representing past and future potential movement for each incident was helpful, to make a predictive determinations based on the locations of new incidents, aggregation across the data was needed.

Mapping each incident to a regular grid, we summed incident vectors for each point in the grid and smoothed using a Guassian kernel. The model generates an incident flow field which aggregates the data such that given a point in space, a direction estimate is supplied, with a magnitude corresponding to the strength of that estimate.

The result is an aggregated estimate of the net direction of incidents such that, given a point in space, the model supplies a directional vector with the magnitude corresponding to the strength of that estimate.

3.4 Interactive subset selection for prediction refinement

To represent the flow field we used an arrow to indicate direction and color intensity to represent the magnitude of the computed vector. Rather than represent the entirety of the field, with each position in the grid showing a vector, we opted to only represent grid positions containing incident points. (Figure 1). For grid positions with no supporting incident vectors, users can make a determination of movement visually, with the knowledge that no past data directly supports prediction at that location.

Initially all data points are used as input to compute the aggregated data. However, for some predictive questions users may want to consider a subset of the data. For instance, movement patterns following looting events may be different than movement patterns following abductions. Or, movement patterns may vary during different seasons, depending on weather conditions. To allow users to ask these targeted predictive questions, we provided filter options including incident type (abduction, return, sighting, displacement, capture, clash, violence, sexual violence and looting), month of the year and location classification (in a field, in a town, along a road, near a river, in a home). Results are updated on the fly, so users can immediately assess the consequences of the filter.

4 Enabling cause-and-effect analysis for prediction of outcomes

In making a decision, humans tend to rely on predictive intuition, learned from past experience, to estimate likely outcomes to particular actions. For chaotic reporting data, forming this predictive intuition can be difficult using data alone.

Our visualization aims to help users examine cause and effect patterns by finding incidents that fit specific profiles, perhaps reflecting a particular predictive question. Users can then explore the potentially related incidents immediately preceding and following the targets of the query (Figure 2). For instance, users might want to examine incidents in which LRA members defected to see whether specific actions led to or followed from defection.

4.1 Modeling and representing local related incidents

To allow users to perform queries on this data to capture cause-and-effect relationships around target incidents, we implemented search and filter features, such as a search box to find query words in the incident note or incident type filters to focus on specific events.

To capture cause-and-effect on target queries, we needed to implement an approach to identify potentially related incidents for each entry in the dataset. We used a simple approximation, assuming incidents within a specific distance (50 miles) and timeframe (2 weeks) could be potentially related. These incidents were grouped in a set and stored with the target for fast access.

After searching for sets of incidents fitting a specific query, users can hover over an incident and view the associated note along with the computed related incidents. These related incidents are displayed within a circle, along with the movement vectors and arcs described in Section 3. In addition, a simple timeline is provided to indicate the temporal distribution of the related incidents.

4.2 Interactive, incident exploration

One significant challenge in representing this data and facilitating interactive exploration is relating spatial and temporal information together, at multiple scales. While a timeline selection feature provides users with the ability to explore changes in the spatial distribution of events over a number of weeks or years, it is difficult to simultaneously provide time-filtering at a finer scale to enable temporal filtering around target incidents.

Our approach to this challenge was to provide an interactive feature where the user can scroll through related incidents, ordered by time, around selected targets. As incidents are selected, they are highlighted in a timeline with a note describing the sequence of events. Incidents are also highlighted on the map and in the mini timeline, helping the user relate spatial and temporal information together (Figure 2).

Over-plotting, where multiple incidents are mapped to the same position on-screen, presents a further challenge to incident exploration. Our approach to this problem was to allow a user to scroll through incidents in the same position, thus exposing incidents of interest.

4.3 Cause-and-effect regional analysis

After isolating incidents related to specific predictive queries, users may wish to look at regional patterns around target dates to see broader impacts of incidents or to detect variations with an impact on forecasting future events. We concentrated on enabling three types of analyses: looking for annual and weekly variations within a region and timeframe, finding variations in incident type within a region and timeframe, and examining the impact of incidents over time in a region and timeframe.

Using two coupled and interactive lenses, one geo-spatial and one temporal, users can select regions and timeframes of interest, such as around specific incidents. By dragging the geo-spatial lens around the map, users can review histograms to see regional variations in incident-type frequency as well as annual and weekly variations. By repositioning the temporal lens, different ranges of time can be viewed. Additionally, we supply several stacked quantitative data plots showing numbers of individuals affected by incidents within the selected region and timeframe. By coupling regional and temporal lenses with plots and histograms, users can examine wider periods and regions, or find patterns with potential consequences for predictive questions (Figure 2).

5 Implementation

This visualization was implemented in JavaScript using the HTML5 canvas. ModestMaps JavaScript API was used for map visualization. Color brewer qualitative data ramps were used for coloring incidents by year.

6 Use Cases

6.1 Movement prediction variations through subset selection

Users exploring movement data found several promising patterns using this interactive visual approach that would have been extremely difficult to derive through modeling alone. First, we noted border crossing variations in the aggregated flow field data that showed specific places where crosses tended to occur or be avoided. In addition, we noted that incidents tended to flow from the South Sudan to the DRC and from the CAR to South Sudan, but not from the DRC into the CAR.

Second, we noted significant differences in movement patterns in specific regions when input data was filtered by different incident types. For instance, preceding and following incidents tended to occur in slightly different directions when examining incidents involving looting versus those involving abductions. Given new data with incidents of different types in this region, decision makers could decide to deploy resources based on the complete aggregated data or based on aggregations across a specific incident profile (Figure 3).

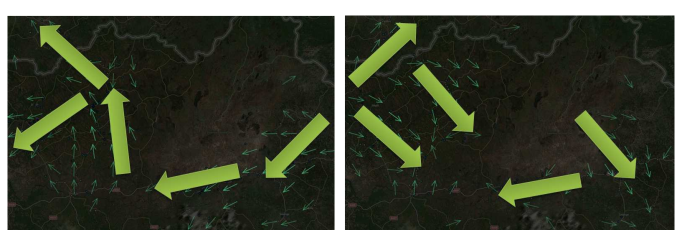
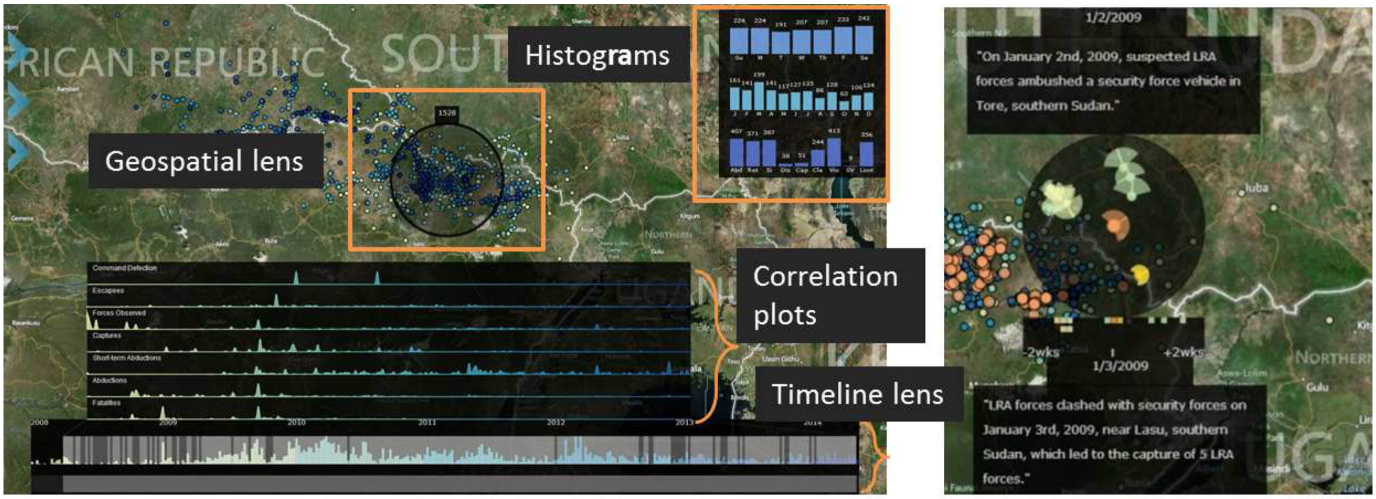


Fig. 3. Differing incident flow is observed when examining different incident types, as annotated in the images. Looting incidents are pictured on the left and abduction ones are on the right.



a b

Fig. 2. a) Visual analytics features in the tool are coupled to a geospatial lens and a timeline lens. b) When exploring incidents of interest, such as the orange incident found by searching for 'ambush', related incidents are highlighted and can be explored.

6.2 Cause and effect analysis of defections

Our approaches to representing a locally-related-incidents model and enabling interactive exploration of these related incidents, coupled with filter and search features, allowed users to identify potentially predictive patterns surrounding captures and defections of LRA commanders.

We used our application to filter for specific incident types, such as clashes, which led to the discovery that incidents involving large numbers of observed forces tended to also be associated, through our locally-related-incidents model, with sightings of commanders. We then targeted these specific commanders and noted that after the capture of Commander Achellam in May of 2012, a series of LRA solider defections were noted in the region, likely in response to the capture. Using the geospatial and temporal lenses, we found that rates of abduction decreased dramatically in the region where Achellam was operating prior to his capture. Using this same approach of filtering and incident exploration, we then identified a commander previously operating in the DRC who moved into the region occupied by Achellam a few months after his capture. Interestingly, we noted, using the geospatial and temporal lenses, no increase in abductions in the region following the change in leadership.

Given new data, such as new incidents involving clashes or new defections, individuals on the ground can use this tool to isolate similar incidents in past data and make decisions based on cause-and-effect patterns observed through the locally-related incident exploration and trend examination using geospatial and temporal lenses.

7 Conclusion

Given the complexity of the problem and data limitations, as well as the need to retain a human in the loop to authorize decisions, we believe a visual analytics predictive approach better accommodates prediction in chaotic and critical conflict zones. Future work will focus on adding an imminence prediction model and on refining temporal correlation analysis methods. We also plan to provide interactive control over parameters to the model, thus providing better control over the exploration process.

References

[1] N. Adrienko and G. Adrienko. Spatial generalization and aggregation of massive movement data. *Visualization and Computer Graphics, IEEE Transactions on*, 17(2):205–219, 2011.

[2] N. Andrienko and G. Andrienko. Interactive visual tools to explore spatio-temporal variation. In *Proceedings of the working conference on Advanced visual interfaces*, pages 417–420. ACM, 2004.

[3] N. Andrienko, G. Andrienko, and P. Gatalsky. Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages & Computing*, 14(6):503–541, 2003.

[4] L. Anselin, I. Syabri, and Y. Kho. Geoda: an introduction to spatial data analysis. *Geographical analysis*, 38(1):5–22, 2006.

[5] C. C. Calhoun, C. E. Stobbart, D. M. Thomas, J. A. Villarrubia, D. E. Brown, and J. H. Conklin. Improving crime data sharing and analysis tools for a web-based crime analysis toolkit: Webcat 2.2. In *Systems and Information Engineering Design Symposium, 2008. SIEDS 2008. IEEE*, pages 40–45. IEEE, 2008.

[6] Innovative Support for Emergencies Diseases and Disasters. *[Online]*. `http://instedd.org/`.

[7] S. Kim, R. Maciejewski, K. Ostmo, E. J. Delp, T. F. Collins, and D. S. Ebert. Mobile analytics for emergency response and training. *Information Visualization*, 7(1):77–88, 2008.

[8] N. Levine. Crime mapping and the crimestat program. *Geographical Analysis*, 38(1):41–56, 2006.

[9] LRA Crisis Tracker. *Modern Weaponry: Technology Targets Elusive War Criminal*.

[10] LRA Crisis Tracker. *[Online]*. `http://www.lracrisistracker.com/`.

[11] R. Maciejewski, R. Hafen, S. Rudolph, S. G. Larew, M. A. Mitchell, W. S. Cleveland, and D. S. Ebert. Forecasting hotspotsa predictive analytics approach. *Visualization and Computer Graphics, IEEE Transactions on*, 17(4):440–453, 2011.

[12] A. Malik, R. Maciejewski, T. F. Collins, and D. S. Ebert. Visual analytics law enforcement toolkit. In *Technologies for Homeland Security (HST), 2010 IEEE International Conference on*, pages 222–228. IEEE, 2010.

[13] S. Rinzivillo, D. Pedreschi, M. Nanni, F. Giannotti, N. Andrienko, and G. Andrienko. Visually driven analysis of movement data by progressive clustering. *Information Visualization*, 7(3-4):225–239, 2008.

[14] Ushahidi. *[Online]*. `http://www.ushahidi.com/`.

[15] D. W.-S. Wong and J. Lee. *Statistical analysis of geographic information with ArcView GIS and ArcGIS*. John Wiley & Sons Hoboken, NJ, USA, 2005.