Detecting and classifying anomalous behavior in spatiotemporal network data*

William Chad Young
University of Washington
Department of Statistics
wmchad@uw.edu

Joshua E. Blumenstock
University of Washington
Information School
joshblum@uw.edu

Emily B. Fox
University of Washington
Department of Statistics
ebfox@uw.edu

Tyler H. McCormick
University of Washington
Department of Statistics
tylermc@uw.edu

ABSTRACT

We investigate different models for detecting and classifying important geopolitical events in high-frequency spatiotemporal network data. Building on previous empirical work on the network response to real-world events, our goal is to develop a generative model that can identify the time, location, and nature of different emergency and non-emergency events. As a testbed for these models, we use a large dataset containing billions of anonymized mobile phone calls and text messages from Afghanistan, and associated metadata on several known important geopolitical events. We find that simple and scalable time-series models of geographically aggregated call volume can accurately identify the onset of major events when the approximate time and location of the event is known. However, such models ignore the network structure in the data, and are not well suited to spatial localization. Preliminary results from dynamic matrix factorization models, which generatively model network structure, indicate a promising area for future work.

Categories and Subject Descriptors

G.3 [**Mathematics of Computing**]: Probability and Statistics; J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Sociology*

Keywords

Emergencies, anomaly detection, call detail records (CDR)

1. INTRODUCTION

As mobile phones and other sources of network data proliferate across the globe, there exists the possibility of using these data to better model and understand the changing real-world environment in which the data are generated [7]. In this paper, we investigate the promise of using generative modeling frameworks to distinguish between emergency and non-emergency events localized in space and time based on mobile phone interaction data. An important question, which is left primarily open, is how to accomplish this task efficiently at scale.

We perform this analysis using a large mobile phone network dataset from Afghanistan that contains billions of interpersonal communications with spatial and temporal markers. We calibrate these methods using a handful of important violent and non-violent geopolitical events. In this context, for instance, we want to be able to detect a soccer match based on how the call patterns of the users change, and determine that what occurred was a soccer match rather than a bombing or an earthquake. This approach builds on recent empirical work that has shown that different types events produce heterogeneous responses in such data [3, 10, 5].

Building on these insights, our goal is to develop a generative model which differentiates between emergency and non-emergency events. Given a specific temporal and spatial location for an event, we show that straightforward modifications to existing time series and network methods recover sufficient signal to differentiate between events. We use a Markov switching autoregressive model to evaluate changes in tower volume and a time-varying latent factor model as a means of recovering signal from graph data.

Such methods perform well when the spatial and temporal context are approximately known, but perform poorly absent a pre-specified context. We contend that the remaining open modeling challenge is spatiotemporal localization. Spatiotemporal localization here refers to the ability of a method to differentiate signal from a spatially, temporally, and network dependent noise distribution. In the case of time series models, for example, localization requires identifying which cell tower is likely closest to an event from among a series of towers which have correlated volume profiles. The complex dependence differentiates the localization problem from so-called "rare event" models where there is limited signal, but typically unstructured noise.

Our work relates closely to a series of recent studies that

KDD-LESI 2014, August 24, 2014, New York City, USA.

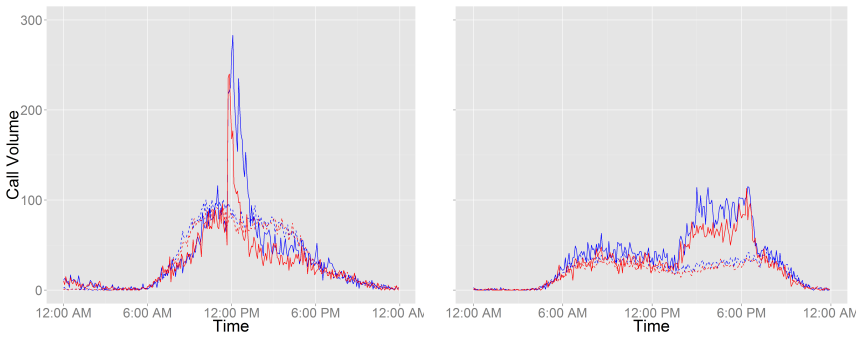

Figure 1(a): Call volume at the tower nearest a major bombing (left) and a stadium inauguration (right). Red indicates incoming call volume and blue shows outgoing call volume from the tower. The dashed lines indicates average volume for the tower on the same day of the week.

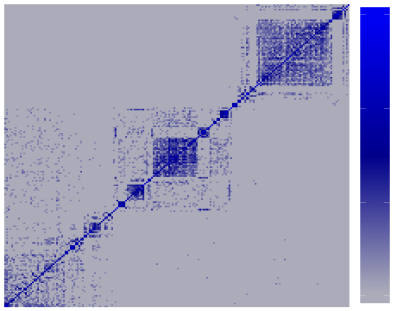

Figure 1(b): Heatmap of (log) call volume between pairs of 200 towers during the 30 minutes following a major bombing. Towers are ordered closest to furthest starting at the bottom left.

have used similar high-frequency network data to explore the impact of unexpected events on network traffic. In work closely related to our own, [2] provide a scalable framework for detecting events when analysis of the full network is prohibitive by focusing more narrowly on activity in subnetworks and at network hubs. Similarly, by focusing on features of the users' local networks over time, [1] detect events by pinpointing times resulting in the greatest deviations. Finally, [5] provide a method that locates an earthquake in Rwanda based on anomalous call volume at nearby towers, and assess continuing need in the area by monitoring persistence of deviations in tower traffic. Our work extends these efforts by seeking a generative model of network activity that can be used to identify and classify the type, timing, and location of important events.

The remainder of the paper is organized as follows: Section 2 details the data and the Afghanistan context. Section 3 describes and presents results from the autoregressive model, and Section 4 develops the matrix factorization approach. In Section 5 we conclude with an outline for future work.

2. DATA

Our data comes from a large telecommunications operator in Afghanistan. The full dataset contains detailed information on all communication events that occurred on the nation's primary mobile phone network in 2011. This captures billions of calls and text messages made by millions of subscribers. For each such event, we observe the time and date of the event, as well as anonymized identifiers for each party on the transaction. We additionally observe the identifier of the mobile phone tower that was used to route the event, which allows us to infer the approximate location of each subscriber at the time of the call or text message.

Since our goal is to capture events which affect large numbers of users simultaneously, we initially focus on activity aggregated at the level of the mobile phone tower. Our dataset includes over 1,000 towers spread across Afghanistan, and which cover all major population centers.

One event that exemplifies the impact an event can have on cell phone usage is a major bombing in the capital city of Kabul on December 6, 2011. The left plot in Figure 1(a) shows the incoming and outgoing call volume at the tower nearest the bombing for that day. There is an immediate spike in the call volume at the time of the bombing, common to many nearby towers. Notably, the incoming call volume spikes slightly later and lasts longer than the outgoing call volume. This is reasonable as there is likely some delay as people learn about the bombing and call in to check on those they know were in the area.

In comparison, the right plot in Figure 1(a) provides a similar depiction of traffic at the time a new stadium is inaugurated in a nearby region of Kabul. Instead of the quick spike and return to normal that we see in the bombing, there is prolonged period of increased activity, similar to that observed by [3] for other non-emergency events.

These descriptive figures focus on changes in call volume at the single tower closest to the event, and ignore the networked structure in the data. A more natural representation, and one that we will draw in Sec. 4, treats each tower as a localized node in a network, and calls between towers as edges. A simple representation of this time-varying network is given in Figure 1(b), which shows a 30-minute snapshot of the call volume between towers following the bombing in Kabul.

3. TIME SERIES MODELS

We start by considering the differential structure of events in terms of call volume alone, as in [3]. To examine the ability of generative models to distinguish between events given a known context (e.g., time window and location), we first look at techniques which analyze call volume from the single closest tower. Let y_t be the observed volume at time t, $t \in \{1, \ldots, T\}$. The goal, then, is to identify the set of time points which are associated with an event at the tower. To this end, we consider a straightforward autoregressive hidden Markov model (AR-HMM):

$$\begin{aligned} y_t = a^{(z_t)} y_{t-1} + \mu^{(z_t)} + \epsilon_t, \qquad & \epsilon_t \sim N(0, \sigma^{2(z_t)}), \\ \Pr(z_t = j | z_{t-1} = i) \quad = \quad & \pi_{ij}. \end{aligned}$$

Here, z_t is a Markov process indexing the hidden state of the tower at time t with with π_{ij} indicating the probability of

transitioning from state i to state j. This formulation allows for flexibility in modeling the dynamics of different types of events. One state may capture the immediate spike caused by an emergency, while another state captures the slower, prolonged build associated with a concert or festival.

We take a Bayesian approach and specify priors on this model. For K hidden states, we have

$$y_t \sim N(a^{(z_t)} y_{t-1} + \mu^{(z_t)}, \sigma^{2(z_t)}),$$
$$\pi_i \sim \text{Dirichlet}(1/K), \qquad \mu^{(i)} \sim N(0, \sigma_\mu^2),$$
$$a^{(i)} \sim N(0, \sigma_a^2), \qquad \sigma^{2(i)} \sim \text{Inv-Gamma}(a, b).$$

This sparse Dirichlet formulation allows for a data-driven adaptive technique that encourages the use of a subset of the available states. Based on the conjugate prior specification, performing inference is straightforward using Gibbs sampling. In addition, we can sample the entire sequence $z_{\{1:T\}}$ using the forward-backward algorithm [9].

To assess performance, we looked at four different events: two emergency events and two stadium inaugurations. All the events occurred in different locations. To localize the data, we use the tower closest to each event. In order to avoid focusing on daily trends in call volume, the data were de-trended by subtracting the mean for that time period and day of week using data from the rest of the month for the tower. The data were then normalized to have maximum volume equal to 1. We used diffuse prior settings $\sigma_\mu^2 = \sigma_a^2 = 1$, $a = 0.1$ and $b = 0.001$ and ran the sampler for 40,000 iterations, discarding the first 20,000 as burn-in. To examine the inferred states, we looked at the best sample as determined by joint model probability. To identify events, we look specifically at the rare states in the sample. This reflects our expectation that events occur suddenly and are short-lived. This should result in a jump to a unique 'event' state and a quick return to the 'normal' or baseline states.

As shown in Figure 2, the three-state model ($K = 3$) is able to detect the emergencies as a unique state but does not have enough other states to adequately capture the variability in the baseline noise process. As such, the $K = 3$ setting fails to uniquely identify the stadium inaugurations. As we move from three to five and then to ten states, we see that the background states are increasingly differentiated from the inauguration state, with the ten-state model finally having enough states to adequately model the baseline noise as separate from the inaugurations.

Figure 3 compares the ten-state model with a Bayesian changepoint model [4]. The changepoint model partitions a time series into a set of independent segments, each with its own mean and variance. These segments are similar to the states in the AR-HMM model, with the major difference that in the changepoint formulation, the system can not return to a previous state. Both models use multiple states to capture baseline noise, but the changepoint model cannot classify the two emergencies or non-emergencies as the same type.

Our results indicate that there is indeed signal that can be modeled generatively to detect the onset of major events. Clearly, however, the multiple baseline states indicate that a more sophisticated dynamic model is important for the task of temporal localization of events, especially as more and more 'null' data (non-event days) are considered. Possibilities include models based on hierarchical HMMs or hidden semi-Markov models [8], though both present greater computational complexity. However, perhaps the greatest challenge is in spatial localization from the massive number of tower-specific time series. So far, our AR-HMM has considered the closest tower only. Assuming a collection of independent towers does not capture the complex correlation structure between tower series. Alternatively, the multivariate switching vector autoregressive (VAR) processes fails to scale to high dimensions (i.e., large numbers of towers) and likewise does not capture the (potentially non-Euclidean) spatial diffusion process.

4. NETWORK MODELS

To analyze all towers jointly and take advantage of the network structure of our data, we consider a matrix factorization approach. Matrix factorization [6] takes an $n \times n$ matrix X and decomposes it into two smaller $n \times k$ matrices U and V such that $X \approx UV^T$. In our case, X is the matrix of tower-tower calls, and X_{ij} is the number of calls from tower i to tower j. The rows of the U and V matrices can then be seen as an embedding of the towers into a low-dimensional latent space, with U and V modeling incoming and outgoing volume separately. These latent locations give us information about the similarity of the towers. Importantly, however, in our application we have a matrix of tower-to-tower call volume $X^{(t)}$ at *each time step* t.

As exploratory data analysis to ascertain whether changes in network structure differ, even at a coarse level, between event types, we first look at matrix factorization on snapshots of the data separately, comparing communication patterns before and after the event. For each of a bombing and a stadium inauguration, we used a 30-minute snapshot of tower-to-tower call volume directly before the event as well as one after the event started. We used $k = 5$ latent factors. From the resulting latent positions of the towers, we computed a distance matrix where the ij'th entry is the distance between tower i and tower j in the latent space. We then subtracted the distance in the post-event snapshot from the distance in the pre-event snapshot to see how the towers moved relative to each other over this time period.

Figure 4 shows 200 towers with high volume, sorted by distance from the event. Blue indicates that a pair of towers is further apart after the bombing than before, and red indicates that they have moved closer. For the bombing, the most important feature is the blue line along the bottom margin of the plot, indicating that the towers nearest the event moved away from the other towers. Also of interest is the set of towers in the middle of the plot which move away from the other towers. These towers are near a concurrent bombing that took place in another city 300 km away from Kabul. The stadium inauguration looks visually distinct from the bombing, indicating that a dynamic matrix factorization approach might provide information that can be used in differentiating between events.

Motivated by our exploratory data analysis, as a first cut at a model integrating the temporal and network aspects of our data, we explore a dynamic matrix factorization formulation

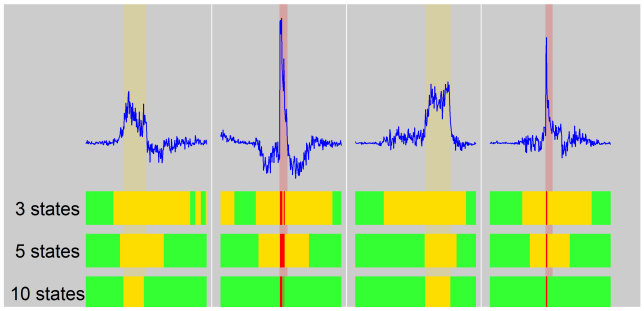

Figure 2: Results of the AR-HMM on two emergency events and two non-emergency events. The de-trended call volume is shown in blue and the actual events are highlighted in yellow (non-emergency) and red (emergency). The bars below show the inferred event states from the sampler where K varies. Green is the set of baseline states, while yellow and red indicate the rare states picked during the true non-emergency and emergency events, respectively.

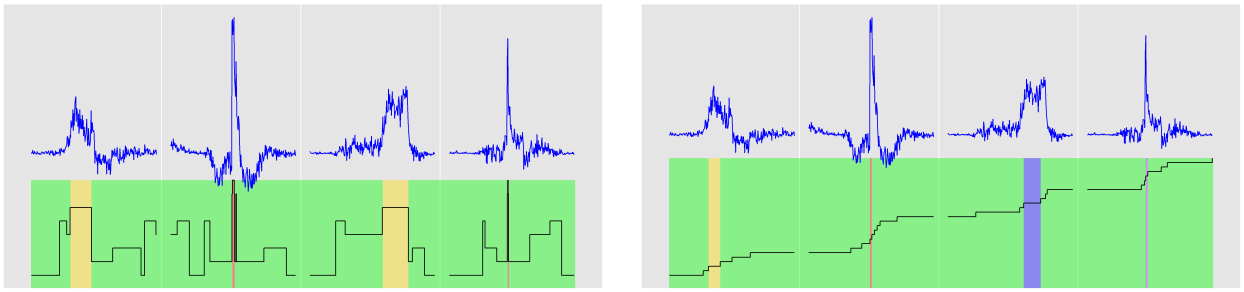

Figure 3: Comparison of AR-HMM (left) and Bayesian changepoint detection (right) on the events of Figure 2. For the changepoint method, the non-green colored bars show the state at the beginning of each event.

similar to that in [11]. In this formulation, given $n \times k$ matrices $U^{(t)}$ and $V^{(t)}$, we model $X^{(t)}$ via a latent random walk as

$$\begin{aligned} X_{ij}^{(t)} | U^{(t)}, V^{(t)} &\sim N(U_i^{(t)} V_j^{(t)T}, \sigma^2), \\ U_i^{(t)} \sim N(U_i^{(t-1)}, \sigma_U^2 \mathbf{I}), &\quad V_j^{(t)} \sim N(V_j^{(t-1)}, \sigma_V^2 \mathbf{I}), \\ U_i^{(0)} \sim N(0, \sigma_0^2 \mathbf{I}), &\quad V_j^{(0)} \sim N(0, \sigma_0^2 \mathbf{I}). \end{aligned}$$

The coupling of the $U^{(t)}$ and $V^{(t)}$ matrices over successive time points allows sharing of information about towers across time. In the absence of significant events, we expect the locations of towers in the latent space will move smoothly. Events which have a large impact on towers in the network will be reflected by larger moves in the latent space. A Bayesian approach would put priors on the hyperparameters $\sigma^2, \sigma_0^2, \sigma_U^2, \sigma_V^2$.

Our formulation does not explicitly model event-driven dynamics as in Sec. 3. Instead, event detection could be a post-processing step. However, this model affords scalable inference, which is key given the size of our dataset. In particular, finding the maximum a posteriori (MAP) estimate of the matrices $\{U^{(t)}\}$ and $\{V^{(t)}\}$ is equivalent to minimizing

$$\sum_{t=1}^{T} \|X^{(t)} - U^{(t)} V^{(t)T}\|_F^2 + \lambda_U \sum_{t=1}^{T} \|U^{(t)} - U^{(t-1)}\|_F^2 + \lambda_V \sum_{t=1}^{T} \|V^{(t)} - V^{(t-1)}\|_F^2 + \lambda_0 (\|U^{(0)}\|_F^2 + \|V^{(0)}\|_F^2).$$

Although this objective is not convex, a local optimum can be efficiently computed using stochastic gradient descent or block coordinate descent, with opportunities for parallelization making the model highly scalable.

As a test of the dynamic matrix factorization model, we applied it to the entire network in a time period around the major bombing as well as to a time period around a stadium inauguration. We again used $k = 5$ with $\lambda_U = \lambda_V = \lambda_0 = 10$ and ran a block coordinate descent algorithm for 1,000 iterations. To evaluate our results, we looked at how the distance between the closest tower and other towers changes over time. For comparison, we also look at the same plot for a random tower distant from the event. Figure 5 visualizes the motion of the towers of interest in the latent space. Here, the green dot indicates the time of the event of interest. Reflecting the results of the independent matrix factorization of Figure 4, the closest tower to the bombing (top left quadrant of Figure 5) does move away from the other towers quickly at the time of the event, returning to a closer position relatively quickly. In contrast, there is no such distinct shift for the distant tower. For the stadium inauguration, there is no such significant move for the *closest* tower, also keeping with the results of Figure 4. In particular, the lower margin of the inauguration plot is mostly grey, interestingly indicating that the nearby towers did not move significantly in relation to most of the other towers.

Our results indicate that there is indeed signal in our spatiotemporal network data that can be modeled generatively to distinguish between events. However, as emphasized in

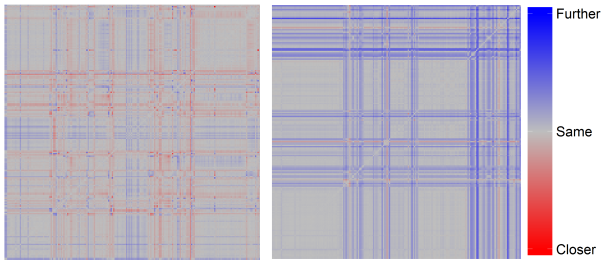

Figure 4: Heatmap showing changes in distance between towers in the latent space from independent matrix factorizations using a 30-minute snapshot of tower-to-tower volume before and after an event for the bombing (left) and the stadium inauguration (right). The towers are ordered by geographic distance to the event with the closest towers in the lower left.

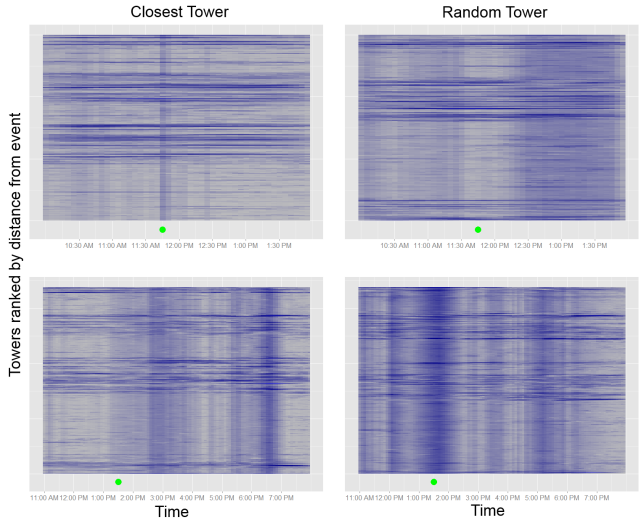

Figure 5: Dynamic matrix factorization results. Heatmaps showing distance in the latent space between a selected tower and other towers across time. The top figures are from the bombing, and the bottom are from the stadium inauguration. The green dot indicates the beginning of the event.

Section 3, it is difficult to find a scalable model that can account for the full course of the data, when neither the time, location, or type of the event is specified. We believe this localization problem is again at the heart of the challenge, and an important open task for the community.

5. CONCLUSION

We use mobile phone interaction data to detect and differentiate between events. When given a general time and location, existing generative models effectively differentiate between emergency and non-emergency events. Performance deteriorates rapidly without this information. Given these results, we contend that there are two primary open challenges in this literature. First, localizing events in time and space requires modeling a background noise process with a high degree of spatial, temporal, and network structure. Information on both time and location of an event was critical in both the call volume data and network data. Second, scaleability remains a substantial open issue. For our network results the matrix factorization model provides a scaleable representation of changing structure in the graph. We could not, however, scaleably incorporate a generative model which detects and classifies events using existing models. Scaleable implementations of such models remains an open area of research.

6. ACKNOWLEDGMENTS

We gratefully acknowledge DARPA Grant FA9550-12-1-0406 negotiated by AFOSR, Google Faculty Research Award 2013_R1_307, and U.S. Army Research Office 62389-CS-YIP.

7. REFERENCES

[1] L. Akoglu and C. Faloutsos. Event detection in time series of mobile communication graphs. In *Army Science Conference*, pages 77–79, 2010.

[2] Y. Altshuler, M. Fire, E. Shmueli, Y. Elovici, A. Bruckstein, A. S. Pentland, and D. Lazer. Detecting anomalous behaviors using structural properties of social networks. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 433–440. Springer, 2013.

[3] J. P. Bagrow, D. Wang, and A.-L. Barabasi. Collective response of human populations to large-scale emergencies. *PloS one*, 6(3):e17680, 2011.

[4] C. Erdman and J. W. Emerson. bcp: An R package for performing a bayesian analysis of change point problems. *Journal of Statistical Software*, 23(3):1–13, 2007.

[5] A. Kapoor, N. Eagle, and E. Horvitz. People, quakes, and communications: Inferences from call dynamics about a seismic event and its influences on a population. In *AAAI Spring Symposium: Artificial Intelligence for Development*, 2010.

[6] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[7] D. Lazer, A. S. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, et al. Life in the network: the coming age of computational social science. *Science (New York, NY)*, 323(5915):721, 2009.

[8] K. Murphy. Hidden semi-Markov models (segment models). Technical Report URL http://www.cs.ubc.ca/˜murphyk/Papers/segment.pdf, 2002.

[9] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[10] P. Sundsoy, J. Bjelland, G. Canright, K. Engo-Monsen, and R. Ling. The activation of core social networks in the wake of the 22 july oslo bombing. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 586–590, Aug. 2012.

[11] L. Xiong, X. Chen, T.-K. Huang, J. G. Schneider, and J. G. Carbonell. Temporal collaborative filtering with bayesian probabilistic tensor factorization. In *SDM*, volume 10, pages 211–222. SIAM, 2010.