Generative Models for Spatial-Temporal Processes with Applications to Predictive Criminology

Yoon-Sik Cho, Aram Galstyan
USC Information Sciences Institute
Marina del Rey, CA 90292
{yoonsik, galstyan}@isi.edu

Jeff Brantingham
UCLA
Los Angeles, CA 90095
pjb@anthoro.ucla.edu

George Tita
UCI
Irvine, CA 92697
gtita@uci.edu

Abstract

We present a generative model for spatial-temporal data that describes geographically distributed interactions between pairs of entities. We develop an efficient approximate algorithm to infer unknown participants in an event given the location and the time of the event. As a concrete application of the proposed approach, we focus on the problem of modeling inter–gang violence, where the objective is to infer the identities of participants in violent inter-gang attacks, based on the past observations of such attacks. We validate the model on synthetic as well as real–world data, and obtain very promising results on the identity–inference task. Furthermore, it is shown that combining both spatial and temporal information yields better accuracy than using either information separately.

1 Introduction

Spatial-temporal data describes processes or phenomena that are extended both in space and time. A classical problem associated with spatial-temporal data is tracking the trajectories of kinematic objects such as missiles and vehicles. More recently, call records have been used to map, track, and predict the movement of cell-phone users. In general, the emergence of various types of sensors has made it possible to gather spatial-temporal data for numerous real–world processes, which has made it imperative to have efficient computational models for modeling and predicting with spatial-temporal data.

In this paper, we consider a particular type of spatial-temporal process – the sequence of violent events occurring among rival street gangs. Specifically, we use Los Angeles Police Departmen's (LAPD) data on gang related violent crimes that covers a period from 1999 to 2002 in the Hollenbeck neighborhood of Los Angeles [8]. In this dataset, each entry describes a violent incident that includes the date and time of the incident as well as the latitudinal and longitudinal coordinates. Furthermore, each record might also contain information about the victim and/or the perpetrator of the attack. Often, the latter type of information is not available. On the other hand, knowing the identities of the gangs involved in an attack would be potentially helpful for predicting and/or preventing subsequent retaliatory attacks. Thus, the ability to infer the event participants would help the law-enforcement agencies to prioritize their resources efficiently for controlling outbursts of violent activities.

One of the main characteristics of gang–related activities is that different events are not statistically independent, but exhibit non–trivial correlations. In particular, recent observations suggest that events tend to cluster both temporally and spatially. To account for such correlations, here we suggest a generative model that is based on self-exciting Hawkes process. Due to the missing information about the participants, exact inference and learning is intractable for even moderately large datasets. Toward this end we develop an efficient algorithm for learning and inference based on the variational EM approach [1]. Our preliminary results on both synthetic and real-world data suggest that the model is able to recover missing data with reasonable accuracy.

2 Background

2.1 Predictive Models for Crime

The costs of crime can be substantial both in human and financial terms [5]. Not surprisingly, there has been significant effort among criminologists to develop approaches to predicting the spatial and temporal distribution of crime, with the goal that law enforcement resources may be directed to likely locations (and at likely times) to prevent crime. Crime hotspotting [2], capturing the fact that crime tends to concentrate in some places and not others, and repeat-victimization [3, 11], tracing the regular recurrence of crime against the same people or targets, have featured prominently in attempts to predict crime. However, many practical questions remain in these approaches. For the most part, predicting crime has primarily been backwards-looking, assigning law enforcement resources only to locations that have recently experienced crime. The past is a good guide to the present, but crime patterns evolve dynamically meaning that today's crime locations may not be identical to tomorrow's [4]. It has also been the case that predictive models have tended to use fixed model parameters based on mean system characteristics. We address both of these issue here specifically in the case of modeling inter-gang violence in Los Angeles.

3 Spatial-Temporal Model of Inter-Gang Violence

Here we describe our spatial-temporal model for characterizing interactions between a pair of entities. For the spatial component of our model, we assume that events occurring between two participants are likely to be clustered in space. In the case of the gang violence data, this assumption asserts that attacks between a pair of gangs are likely to be geographically confined.

As stated above, the existence of non-trivial temporal correlations between the events precludes the use of simple Poisson point process model. Instead, here we will use a Hawkes Process, which is a variant of a self-exciting process [6]. Self-exciting models have been used extensively in earthquake modeling, where a central event (earthquake) might excite other events (aftershocks) [6]. A justification of this choice for the inter-gang violence data is that an attack by a gang on its rival is likely to cause a retaliatory response from the victim.

3.1 Generative Model for Inter-Gang Violence

We assume we have a set of events (e.g., gang attacks) that are distributed in time and space. The set of observation is the time, and location of events occurring between the gangs. Further, we assume there are M gangs, and that each event involves one of the $M(M-1)/2$ pairs among those gangs.

We model the sequence of events between the pairs by a Hawkes Process [6]. Namely, we assume that the intensity of events between the gangs i and j has the following form:

$$\lambda_{ij}(t|\mathcal{H}_t) = \mu_{ij} + \sum_{p:t>t_p} g_{ij}(t-t_p) \quad (1)$$

where $\mathcal{H}_t$ denote the history of events up to time t as the set of all the events that have occurred before that time. For simplicity, we omit $\mathcal{H}_t$ in the expression of rate function. Here μ_{ij} describes the background rate of event occurrence that is time-independent, whereas the second term describes the self-excitation part, so that events in the past affect the probability of an event in the near future. The overall rate λ increases as each event p at t_p occurs, reflecting the retaliation to a previous attack. Here we will use a two-parameter family for the self-excitation term:

$$g_{ij}(t-t_p) = \beta_{ij}\omega_{ij}\exp\{-\omega_{ij}(t-t_p)\} \quad (2)$$

Here β_{ij} describes the weight of the self-excitation term (compared to the background rate), while ω_{ij} describes the decay rate of the excitation.

The generative process begins with the sampling of the first time of the incident.

1. For each pair, sample the first time of the incident using an exponential distribution with rate parameter μ.

2. For each pair, sample the duration of time until the next incident using Poisson thinning. Since we are dealing with non-homogeneous Poisson process, we use the so called thinning algorithm [7] to sample the next time of the event. By repeating step 2, we obtain the timestamps of incidents for each pair.

3. For every timestamp of a given pair we sample the location of the incident. We assume the location of incidents follows a Gaussian distribution.

Let us introduce binary indicator variables z_{pq}^n for each event n, so that $z_{pq}^n = 1$ when the gangs p and q are involved in the n-th event, and $z_{pq}^n = 0$ otherwise. To simplify notation, from this point on we denote z_{pq}^n as z_k^n, where k enumerates one of the possible pairs. For a sequence of N incidents, the joint probability of the location $\mathbf{r}^{1:N}$, time $t^{1:N}$ and the latent variables $z_k^{1:N}$ can be written in the following form :

$$p(\mathbf{r}^{1:N}, t^{1:N}, Z^{1:N}|\Theta) = \prod_k p_1(z_k^1|\lambda_k(t^1))p_3(\mathbf{r}^1|z_k^1) \times \prod_{k,n>1} p_2(t^n, z_k^n|\lambda_k(t^n), Z^{1:n-1})p_3(\mathbf{r}^n|z_k^n) \quad (3)$$

where Θ is the set of model parameters, p_1 is time-homogeneous exponential distribution, p_2 is non-homogeneous exponential, and $p_3(\cdot|z_k^n)$ is Gaussian when $z_k^n = 1$, and a constant otherwise:

$$p_3(\mathbf{r}^n|z_k^n) = \begin{cases} \mathcal{N}(\mathbf{m_k}, \Sigma_k) & z_k^n = 1 \\ \text{const.} & z_k^n = 0 \end{cases} \tag{4}$$

Here $\mathbf{m_k}$, Σ_k are the mean and covariance matrix of multivariate normal distribution that describes the spatial distribution of events involving the k-th pair. For simplicity, we assume that the off–diagonal elements of the covariance matrix are zero, thus discarding correlation between latitudinal longitudinal coordinates.

We focus on the temporal-only part in above equation 3 by ignoring the spatial component of the model. The first component of temporal part follows the exponential distribution with fixed rate, and solely determined by the background rate:

$$\prod_k p_1(t^1, z_k^1|\mu_k) = \prod_k \{\mu_k \exp(-\mu_k t^1)\}^{z_k^1} \tag{5}$$

The rate function from the 2nd event is determined by the previous events, and changes over time:

$$\prod_{k,n>1} p_2(t^n, z_k^n|Z^{1:n-1}) =$$
$$\prod_{k,n>1} \{\lambda_k(t^n)\}^{z_k^1} \exp\left\{-\int_{t_k(Z^{1:n-1})}^{t^n} \lambda_k(t)dt\right\} \tag{6}$$

where the lower limit of integration $t_k(Z^{1:n-1})$ is the time of the most recent event prior to the n-th event that involves the pair k.

Equations 3-6 complete the definition of the model. Due to the presence of latent variables z_k^n, there is no closed-form expression for the likelihood of the observed sequence of events. Instead, one has to resort to approximate techniques for learning and inference, which is described next.

4 Learning and Inference

The latent variables indicate assignments of pairs to particular incidents. If sufficient labeled data was available, then standard estimation techniques (such as maximum likelihood estimation) can be used. However, this is not feasible when there is a substantial amount of missing data, which is generally the case with police records. In this particular dataset, at least one of the participants is unknown for almost 70% of incidents. The EM approach presented here deals with this type of scenario. In our model, labeled data is used by clamping the corresponding latent variables to their true values. We rely on variational EM [1] by positing a simpler distribution $Q(X)$ over the latent variables with free parameters for learning and inference. The free parameters are adjusted so that the distribution is close to the true posterior in KL divergence.

$$D_{KL}(Q||P) = \int_X Q(X) \log \frac{Q(X)}{P(X,Y)} dX \tag{7}$$

where X is the hidden variables, and Y is the observed variables. In our case, X is the hidden identity of gangs involved in the incident, whereas Y describes the location and the time of the incident.

We introduce the following factorized variational multinomial distribution, which are independent of each other across steps and pairs:

$$Q(Z^{1:N}|\Phi) = \prod_{k,n} q(z_k^n|\vec{\phi}^n) \tag{8}$$

where free variational parameters ϕ_k^n describe the probability that the k-th pair is involved in the n-th event. Note that the present choice of the variational distribution discards correlations between past and future incidents, thus making the calculation tractable.

Minimizing the KL-divergence between $Q(X)$, and $P(X,Y)$ leads to the following approximate lower bound for the log-likelihood:

$$\begin{aligned} \mathcal{L}_\Phi(Q,\Theta) \quad = \quad & E_Q[\log \prod_{k,n} p_3(\mathbf{r}^n|z_k^n)] \\ + \quad & E_Q[\log \prod_{k,n>2} p_2(z_k^n|Z_,^{1:n-1}\Theta)] \\ + \quad & E_Q[\log \prod_k p_1(z_k^1|\Theta)] \\ - \quad & E_Q[\log \prod_{k,n} q(z_k^n|\vec{\phi}^n)] \end{aligned} \tag{9}$$

where Φ is the set of variational parameters, and Θ is the set of model parameters.

Variational EM algorithm works by iterating between the E–step of calculating the expectation value using the variational distribution, and the M–step of updating the model (hyper)parameters so that the data likelihood is locally maximized. The update equations for the parameters are rather cumbersome and will be provided elsewhere. The overall pseudo algorithm is shown in Algorithm 1. Next we provide our results.

4.1 Variational E-step

In the variational E-step, we minimize the KL distance over the variational parameters. Taking the derivative of KL distance with respect to each variational parameter and setting it to zero, we obtain a set of equations

Algorithm 1 Variational EM

Size: consider total of N events, K pairs
Input: data $\mathbf{r}^{1:N}$, $t^{1:N}$, Z of complete events
Start with initial guess of hyper parameters.
repeat
  Initialize all $\vec{\phi}^{1:N}$ to $\frac{1}{K}$ with unknown pairs
  repeat
    for $n = 1$ **to** N **do**
      if the pairs of n-th event is unknown **then**
        Update $\{\vec{\phi}\}^n = f(\{\vec{\phi}\}^{1:n-1}, \{\vec{\phi}\}^{n+1:N}, \mathbf{r}^n)$
      end if
    end for
  until convergence across all time steps
  Update hyper parameters.
until convergence in hyper parameters

that can be solved. For the k-th component of variational parameter $\vec{\phi}^\eta$ at time η,

$$\phi_k^\eta \propto \exp\left\{\frac{\partial}{\partial \phi_k^\eta} E_Q\left[\sum_{k,n} z_k^n \log \lambda_k(t^n)\right] - \int_{t_k(Z^{1:n-1})}^{t^n} \lambda_k(t)dt\right]\right\} p_3(\mathbf{r}^n|\mathbf{m}_k, \Sigma_k) \tag{10}$$

In the summation of expectation of $\log \lambda_k(t^n; Z^{1:n-1})$, ϕ_k^η is contained when $n > \eta$ as well as $n = \eta$. For certain time step when $n = \eta$, since we are taking the derivative with respect to ϕ_k^η, we need to compute the expectation of λ_k after time $t^{\eta-1}$. Hence, we solve the above equation using the equations below:

$$\frac{\partial}{\partial \phi_k^\eta} E_Q[\log \lambda_k(t^\eta)] = \sum_{z_k^{1:\eta-1}} \prod_{i=1}^{\eta-1} \phi_k^{i\, z_k^i} (1-\phi_k^i)^{(1-z_k^i)} \times \log\left[\mu_k + \sum_{i=1}^{\eta-1} z_k^i \beta_k \omega_k e^{-\omega \Delta t_i^\eta}\right] \tag{11}$$

Similarly, we can calculate the derivative with respect to ϕ_k^η for the terms with $n > \eta$.

$$\frac{\partial}{\partial \phi_k^\eta} E_Q[\log \lambda_k(t^n)] = \sum_{\tilde{\mathcal{Z}}_{k-1}^p} \phi_k^n \prod_{\substack{i=1\\ i\neq\eta}}^{n-1} \phi_k^{i\, z_k^i} (1-\phi_k^i)^{(1-z_k^i)} \times \log\left[\frac{\mu_k + \sum_{l=1, l\neq\eta}^{n-1} z_k^l g_k(t_k - t_l) + g_k(t_k - t_\eta)}{\mu_{ij} + \sum_{\substack{l=1\\ l\neq\eta}}^{n-1} z_k^l g_k(t_k - t_l)}\right] \tag{12}$$

4.2 Variational M-step

The M-step in the EM algorithm computes the parameters by maximizing the expected log-likelihood found in the E-step. The model parameters we are using are $\mathbf{m}_k$, $\mathbf{\Sigma}_k$, β_k,μ_k, and ω_k.

The re–estimation formulas for the spatial parameters, (i.e., the mean and the variance of Gaussian distribution), are straightforward:

$$\mathbf{m}_k = \frac{\sum_n \phi_k^n \mathbf{r}^n}{\sum_n \phi_k^n} \tag{13}$$

$$\sigma_{k,lat}^2 = \frac{\sum_n \phi_k^n (\mathbf{r}_{lat}^n - \mathbf{m}_{k,lat})^2}{\sum_n \phi_k^n} \tag{14}$$

$$\sigma_{k,long}^2 = \frac{\sum_n \phi_k^n (\mathbf{r}_{long}^n - \mathbf{m}_{k,long})^2}{\sum_n \phi_k^n} \tag{15}$$

The re–estimation of the temporal parameters are more involved. For instance, to estimate μ_k, we nullify the derivative of the likelihood with respect to μ_k, $\frac{\partial \mathcal{L}_\Phi}{\partial \mu_k} = 0$, which yields

$$\sum_n \sum_{z_k^{1:n-1}} \phi_k^n \frac{\prod_{i=1}^{n-1} \phi_k^{i\, z_k^i} (1-\phi_k^i)^{(1-z_k^i)}}{(\mu_k + \sum_{i=1}^{n-1} z_k^i \beta_k \omega_k e^{-\omega(t_n - t_i)})} = t^N - t^0$$

Similar equations for the parameters β_k and ω_k are omitted.

Unfortunately, the resulting equations do not allow closed form solutions, so they have to be solved using numerical methods, such as the Newton's method employed here.

5 Experiments

In this section, we report on our experiments using the generative model described above. Given a sequence of events, where some information about participants is missing, our goal is to reconstruct the missing information. Below we provide an evaluation of our algorithm on both synthetic and real-world data.

5.1 Synthetic data

Here we report our experiments with synthetically generated data. To compare the performance of our model with previous approaches, we follow the experimental set-up proposed in [9], were the authors used temporal-only information for reconstructing missing information in synthetically generated event data. In particular, [9] suggested an alternative to maximum likelihood (ML) estimation, by replacing the (combinatorial) ML objective by an appropriately defined energy function over relaxed continuous variables. They then perform constrained optimization of the new objective function using l^1, l^2 normalization for the relaxed variables. Although their method does not assign proper probabilities to the various timelines, it can provide a ranking of most likely participants.

Table 1 shows the overall performance of estimating 4 incomplete events (unknown actors) in 40 events of

Table 1: Total of N = 40 events between 6 pairs. Only 4 events have unknown participants. The parameters are $\mu = 10^{-2} days^{-1}$, $\omega = 10^{-1} days^{-1}$ and $\beta = 0.5$

Method	**Accuracy**
Exact ML	47.3 %
max l^1	47%
max l^2	47.1%
Variational EM	46.9%

6 pairs. For our algorithm the results are averaged over 1000 runs. The top method is the exact inference, while as the second, and third are from using the optimization problem with l^1, l^2-normalization respectively. Note that all four methods perform almost identically. Also we see that the result is remarkably higher than the random baseline 1/6, where each pair is selected randomly. We would like to emphasize that the above results rely on temporal information only. As we will see below, including spatial information will improve the results, often significantly.

Table 2: Total of N = 40 events between 6 pairs, with different fraction of unknown participants: {10%,25%,50%}. The parameters are $\mu = 10^{-2} days^{-1}$, $\omega = 10^{-1} days^{-1}$ and $\beta = 0.5$

Fraction of Unknown	**Accuracy**
10%	46.9 %
25 %	38.1%
50 %	33.7%

Next, we examine the impact of more missing information on the accuracy of inference using temporal-only information. The results are presented in Table 2, where we vary the fraction of events with unknown participants from 10% to 50%. Although the accuracy deteriorates, we note that even when the half of the events miss participant information, the accuracy of the model predictions are still well above the random baseline $\sim 16\%$.

In our next set of experiments we we examine the relative importance of spatial and temporal parts by comparing three variants of our algorithm that use 1. Temporal only data (T), 2. Spatial-only data (S); 3. Combined spatial and temporal data (ST). For the spatial component of the data, we use six multivariate normal distributions with the center of each on the vertex point of hexagon (for all 6 pairs). We fix the side length of the hexagon to 1, and analyzed how varying the standard deviation of the normal distribution affects the overall performance. Specifically, we varied the standard deviation σ from 0.2 to 4, and averaged results over 100 times for each case for spatial only and temporal+spatial. As expected, the relative importance of the spatial information decreases when increasing σ. In the limit when σ is very large, location of an event do not contain any useful information about the participants, so that the accuracy based on spatial information only should converge to the random baseline 1/6.

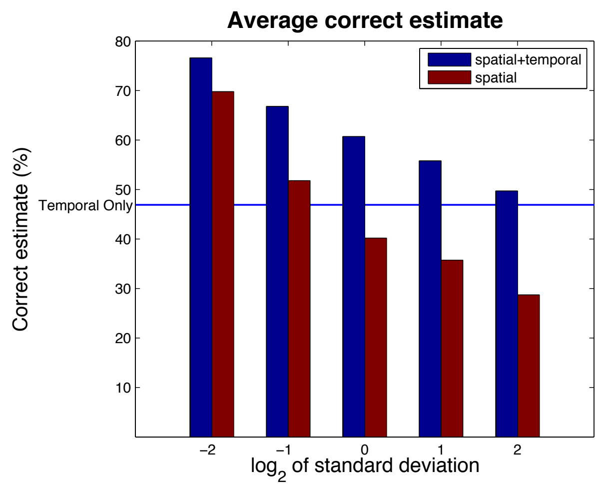


Figure 1: Average accuracy of 100 trials respect to various settings of standard deviations. Blue bar is from the inference based on spatial temporal data, and red bar is from the inference based on spatial data. Blue line is from the previous experiment which only used temporal data

On the other hand, for small values of σ, the spatial information helps to increase accuracy. Indeed, Figure 1 shows that inference using combined spatial-temporal data produces more accurate results for the whole range of the values of σ.

5.2 Real–World data

Next we describe our experiments on real-world dataset collected in the Hollenbeck division of Los Angeles from in 1999 to 2002.

5.2.1 Description of the data

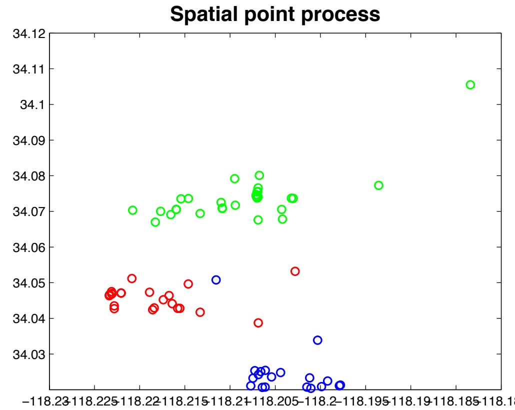
Hollenbeck is a 15.2 square mile (39.4 km2) policing division of the Los Angeles Police Department (LAPD), located on the eastern edge of the City of Los Angeles, with approximately 220,000 residents. Overall, 31 active criminal street gangs were identified in Hollenbeck between 1999-2002 [10]. Only 29 were still active by the end of 2002. These 29 gangs formed at least

66 unique rivalries, which are responsible for the vast majority of violent exchanges observed between gangs. Between November 14, 1999 and September 28, 2002 (1049 days), there were 1208 violent crimes attributed to criminal street gangs in the area. Of these, 1132 crimes explicitly identify the gang affiliation of the suspect, victim, or both. The remaining events include crimes such as shots fired which are known to be gang related, but the intended victim and suspect gang is not clear. For each violent crime, the collected information includes the street address where the crime occurred as well as the date and time of the event [10] , allowing examination of the spatial-temporal dynamics of gang violence. Due to computational complexity, we ignore the causal relation between two incidents which are separated from each other by more than 3 weeks. This assumption also agrees with the Hawkes intensity function which decays exponentially over time.

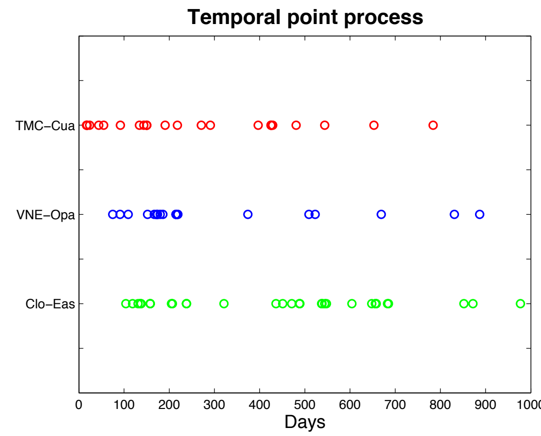
Experiments with most active pairs We first extract the events which are related to the three most active rivalries : (Eastlake-Clover), (OPAL ST-VNE), and (CUATRO FLATS-TMC). In Figure 2(a) we depict the locations of the incident color coded by one of three pairs, while Figure 2(b) depicts the temporal timeline of the attacks between different pairs.

Since each rivalry shows strong pattern in spatial domain, the inference only using spatial information is fairly good. But there were 4 events which was estimated incorrectly. We expect that adding temporal information will yield better estimate. Figure 2 (right) shows that the temporal data is highly clustered. Combining the temporal and spatial information, we were able to recover a single event to the correct rivalry. Hence, when the localization is strong, having temporal feature slightly increases the accuracy. This trend was shown in the result from the simulation on synthetic data. In the experiment using synthetic data, the improvement became smaller as standard deviation decreases(highly localized).

Experiments involving all gangs 31 active gangs which had more than 4 crime with in the period, and 40 pairs which had at least two crimes between each other were considered as candidates in the whole data. Of the data 7.33% misses the information of both gangs in a pair, 62.07% only knows one of two gang, which is mostly victim. Only 30.6% of the events contain information about both participants. Three locations which are isolated from clusters have been removed for estimating the means, and variance of Gaussian through our Experiment II. Our objective is to infer the unknown gangs with better estimates of the parameters. For this specific experiment we only use 30.6% of the whole data where we can compare our inference exactly to the original information. From this point, we denote this data as 'all known' data.



(a)



(b)

Figure 2: Location (a) and timeline (b) plot of the incidents involving three most active rivalries. Different colors represent different pairs.

With all the information of pairs and their locations and time for the all known data, we compute the mean and variance of locations for each pair. The inference using only spatial data, which is better than the inference using only temporal data is set as a baseline to compare. The most likely pairs for each incidents were picked based on the inference using spatial data, assuming the information of pairs are hidden. We compared our most likely pairs to the actual ones and the overall accuracy was 53.04%.

The top histogram in Figure 3 depicts the number of incidents between different pairs. It shows that inactive pairs which had only two incidents outnumber active pairs. Furthermore, the pairs that have only a few attacks between them, do not exhibit well–defined

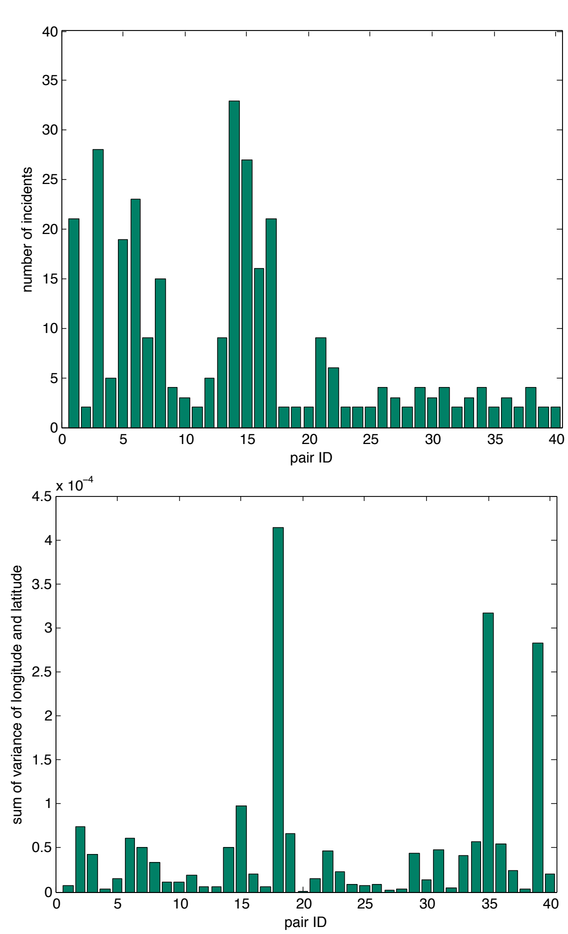


Figure 3: Number of known incidents for each pair (top) and its sum of variance of longitude and latitude (bottom)

spatial clustering, with higher variance in locations of incidents. This can be partly attributed to the fact that attacks between gangs that do not have intense rivalries are mostly random, as opposed to the patterns of attacks between gangs that have intense rivalries. In the latter case, self-excitation (i.e., retaliation) is likely to play a greater role.

To see how our method can recover missing data better than the inference based on spatial data, we compare our inference to the baseline. Further, we compare our result to the other inference of which the temporal process follows homogenous Poisson process while the spatial process is the same as ours. For this experiment, we assume the spatial parameters: mean and variance of Gaussian are given, which is the same setting of baseline, while as temporal parameters: parameters of Hawkes process are being estimated in the algorithm. We assume some of the portion of the data as hidden, and try to recover based on their location, and other incidents around the given time. By controlling the ration from 10, to 70%, we compare our most likely pairs to the actual pairs, and compute the average count of correct inference. Each experiment was repeated 20 times by selecting incidents randomly. Figure 4 shows that as we infuse more information, we obtain better estimate. Besides, over a wide range, our method shows better performance than the baseline.

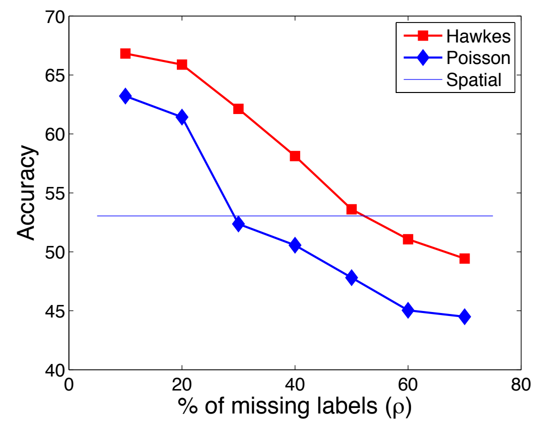


Figure 4: Average accuracy for different fraction of unknown participants. The horizontal line is the accuracy for the spatial-only model.

So far, we have limited our experiments to the 30.6% of data that contain events for which both participants are known. In our final set of experiments, we used the model parameters learned with the 30% of data as an input, and applied it to the dataset that also includes events that contain one unknown participant (about 60% of the data set). In particular, we assumed that all the participants in this 60% are unknown, and examined how well the model is able to infer those participants.

As a baseline, we again use the algorithm that uses spatial-only information. We compare our counting to the baseline which is shown in Table 3. The result shows our model achieves better accuracy than the baseline. However, the accuracy of this experiment cannot be compared to the accuracy of the previous experiment with selected whole data. In this experiment, when one of the pair contains known gang, it was counted as correct, while as in the previous experiment, both of the gangs in the pair should have been in the actual pair to be counted as correct.

In addition to finding the most likely participant, sometimes it can be useful to have a ranked list of possible participants. Thus, we examined whether the top 3 choices from the ranked list predicted by our algorithm contained the actual participants. The results are shown in Table 3. Remarkably, the model that uses combined spatial-temporal information, ranks the actual participant among the top 3 choices an impressive 90% of the time.

Table 3: Each row shows how often the partial observed events were inferred correct counting correct if the one of top n inferred pairs had the known gang. The accuracy in parentheses refer to the accuracy of random guess

Method	**Accuracy**
Spatial Only	68.03% (3.125%)
Top 1 (Spatial Temporal)	72.44% (3.125%)
Top 2 (Spatial Temporal)	84.72% (6.250%)
Top 3 (Spatial Temporal)	89.13% (9.375%)

6 Conclusion

Buoyed by the availability of large scale geo-coded data, an increasing number of agencies are adopting a predictive rather than purely reactive approaches to law enforcement. Thus, there is a growing need for efficient computational approaches for modeling complex spatial-temporal data. Here we have described a preliminary generative model and applied it to the problem of inferring participants and identities of perpetrators in violent inter-gang events, based on the past observations of such attacks. Our results on synthetic and real–world data show that by combining the temporal process with the spatial process, we achieve a better estimate. For the real–world data, we tested our model using the police data on gang crimes in Hollenbeck, and obtained reasonable results. Most prior crime models relied on maximum likelihood estimation on the model parameter with the complete set of data. However, in most cases, we are faced with incomplete data, without any knowledge of the actual parameters. The variational EM algorithm suggested here presents an efficient and reasonably accurate approach for dealing with incomplete data.

It will be interesting to examine other models that can account for the temporal correlations and clustering observed in the real-world data. One natural possibility is to use Hidden Markov Models (HMM) or switching Markov processes to characterize jumps in the intensity of inter-gang violence. Further work will also take into account possible impact of law-enforcement agents on the temporal characteristics of inter-gang violence. Finally, we would like to note that while the main focus here was crime prediction, the model presented here can be generalized broadly to other data sets that describe geographically distributed sequence of social interactions between different entities, where the event depends on the history of its interaction.

Acknowledgements

This research was supported in part by U.S. ARO MURI grant No. W911NF0610094, and US AFOSR MURI grant No. FA9550-10-1-0569.

References

[1] M. Beal and Z. Ghahramani. The variational bayesian em algorithm for incomplete data: with application to scoring graphical model structures. *Bayesian Statistics*, 7:453–464, 2003.

[2] K. J. Bowers, S. D. Johnson, K.Louise, and K. Pease. Prospective hot-spotting the future of crime mapping? *British Journal of Criminology*, 44(5):641–658, 2004.

[3] G. Farrell and K. Pease. *Once Bitten, Twice Bitten: Repeat Victimization and its Implications for Crime Prevention.* London; Home Office Police Department, 1993.

[4] G.Mohler, M.Short, P.Brantingham, F.Schoenberg, and G.Tita. Self-exciting point process modeling of crime. *JASA*, 2010.

[5] P. Heatonl. Hidden in plain sight: What costs-of-crime research can tell us about investing in police. Technical report, RAND, 2010.

[6] Y. Ogata. Space-time point process models for earthquake occurrences. *Ann.Inst.Statist.Math.*, 50:379–402, 1988.

[7] P.Lewis and G.Shedler. Simulation of non-homogenous poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, 1979.

[8] S. M. Radil, C. Flint, and G. E. Tita. Spatializing social networks: Using social network analysis to investigate geographies of gang rivalry, territoriality, and violence in los angeles. *Annals of the Association of American Geographers*, 100(2):307–326, 2010.

[9] A. Stomakhin, M. B. Short, and A. L. Bertozzi. Reconstruction of missing data in social networks based on temporal patterns of interactions. *Inverse Problems*, 27(11):115013, 2011.

[10] G. Tita, J. K. Riley, G. Ridgeway, C. Grammich, A. F. Abrahamse, and P. Greenwood. *Reducing Gun Violence: Results from an Intervention in East Los Angeles.* RAND Press, 2003.

[11] M. Townsley, R. Homel, and J. Chaseling. Infectious burglaries: A test of the near repeat hypothesis. *British Journal of Criminology*, 43(3):615–633, 2003.