

Spatio-Temporal Modeling of Criminal Incidents Using Geographic, Demographic, and Twitter-derived Information

Xiaofeng Wang
Department of Systems and
Information Engineering
University of Virginia
Charlottesville, Virginia 22904
Email: xw4u@virginia.edu

Donald E. Brown
Department of Systems and
Information Engineering
University of Virginia
Charlottesville, Virginia 22904
Email: brown@virginia.edu

Matthew S. Gerber
Department of Systems and
Information Engineering
University of Virginia
Charlottesville, Virginia 22904
Email: msg8u@virginia.edu

Abstract—Personal and property crimes create large economic losses within the United States. To prevent crimes, law enforcement agencies model the spatio-temporal pattern of criminal incidents. In this paper, we present a new modeling process that combines two of our recently developed approaches for modeling criminal incidents. The first component of the process is the spatio-temporal generalized additive model (STGAM), which predicts the probability of criminal activity at a given location and time using a feature-based approach. The second component involves textual analysis. In our experiments, we automatically analyzed Twitter posts, which provide a rich, event-based context for criminal incidents. In addition, we describe a new feature selection method to identify important features. We applied our new model to actual criminal incidents in Charlottesville, Virginia. Our results indicate that the STGAM/Twitter model outperforms our previous STGAM model, which did not use Twitter information. The STGAM/Twitter model can be generalized to other applications of event modeling where unstructured text is available.

I. INTRODUCTION

As reported by the Bureau of Justice and Statistics [1], there were 22,879,720 personal and property crimes with an estimated economic loss of 18 billion dollars in the United States in 2008. One approach to preventing crimes is to model the spatio-temporal patterns of criminal incidents mathematically. With a well developed spatio-temporal model of crimes, law enforcement agencies can study the possible underlying patterns of crimes, predict the location and time of future criminal activity, and efficiently deploy limited resources such as walking and driving patrols to areas with the highest risk.

Researchers have developed and applied several spatio-temporal prediction models. The widely applied spatial hot spot model [2] clusters historical criminal incidents into hot spots using statistical methods such as mixture models and density estimation. Such models assume that future crimes are more likely to happen within these hot spots. Although popular, these models do not account for environmental factors (e.g., the proximity of a police station) that influence criminal behavior. Furthermore, these models cannot be generalized to areas without historical data. More sophisticated statistical

models have been studied by Brown and his colleagues. Liu and Brown [3] applied a conditional point pattern density model to relate criminal incidents to geographic features, demographic features, and consumer expenditure features. Xue and Brown [4] and Smith and Brown [5] developed a spatial choice model by assuming that criminals make rational decisions to maximize their gains and minimize the risk of being caught. Brown, Dalton, and Hoyle [6] applied generalized linear models (GLM) over spatial grids to predict terrorist events.

One problem with the above models is that they do not account for the temporal patterns of crimes. The models typically estimate parameters using the most recent criminal reports. To better account for temporal effects, we developed a spatio-temporal generalized additive model [7], [8]. This model utilized a variety of data types (e.g., spatial, temporal, geographic, and demographic data) to make predictions. The model achieved better prediction performance than the hot spot and spatial GLM models and had good interpretability. The model did not, however, incorporate information derived from textual sources such as Twitter, which are abundant and freely available. In more recent work [9], we developed a model that predicts criminal incidents using information extracted from Twitter¹ posts. Our Twitter-based model did not, however, incorporate the spatial, temporal, geographic, and demographic data used by the generalized additive model. In this paper, we present research that combines our generalized additive model with unstructured textual information from Twitter. As we show, this hybrid model offers improved capabilities for criminal incident modeling, overcoming limitations of each individual approach.

This paper is organized as follows. Section II reviews the spatio-temporal generalized additive model and discusses how to incorporate textual information into it. We then describe our textual analysis technique, which derives event information from unstructured text. A new feature selection method based

¹<http://www.twitter.com>

on least angle regression is also described in this section. Section III evaluates the model developed in this paper using breaking and entering incidents from Charlottesville, Virginia. Section IV provides concluding remarks and suggestions for future work.

II. MODELING

This section describes our application of spatio-temporal generalized additive modeling (STGAM) to criminal incidents. The first subsection reviews the overall STGAM model, and the second subsection discusses how to represent text numerically. The numerical representation of text is incorporated into STGAM. The third subsection describes a randomized feature selection algorithm based on least angle regression, which is used to select features for STGAM.

A. The Spatio-Temporal Generalized Additive Model

The spatio-temporal generalized additive model (STGAM) is a generalized additive model building on spatial grids with the temporal information encoded by dummy variables [7]. Mathematically, the model is defined as follows:

$$\text{logit}(p(\text{inci}_{s_i, t_j} = 1)) = \sum_{n=1}^N f_n(x_{n, s_i, t_j}) + \kappa_{s_i, t_j - t^0} \quad (1)$$

In Equation 1, $p(\text{inci}_{s_i, t_j} = 1)$ is the probability of at least one incident occurring in the spatial grid s_i at time t_j ; $\text{logit}(p) = \log(\frac{p}{1-p})$ is a logit link function; N is the total number of features; x_{n, s_i, t_j} is the n^{th} feature associated with location s_i and time t_j ; f_n is a smooth function of the n^{th} feature to be estimated from data; and $\kappa_{s_i, t_j - t^0}$ is a dummy variable indicating the length of the continuous zeros (no incident occurring) that precede the current observation at location s_i and time t_j .

STGAM has the form of a regular GAM, and thus it can be estimated using well developed methods like the penalized iteratively re-weighted least squares method. Such methods are available in standard statistical software such as R and S-plus. For example, we use the packages “mgcv” and “gam” in R [10] to estimate the parameters of our STGAM. A possible difficulty with the estimation of STGAM is that the training set might be very large. As shown in [7], for a land area of 16 square miles, a grid size of 0.02 miles \times 0.02 miles, and a time interval of one month, there are 4.8×10^5 records for a year. To solve this problem, sub-sampling is suggested. To generate a sample from the records, all the records with a response of 1 as well as a random sample from the records with a response of 0 are combined. This sampling is biased. The effect of the bias can be approximately corrected by adding an offset term $\log(\frac{\text{sample size}}{\text{total number of records}})$ in the estimation process.

To evaluate the performance of the spatio-temporal prediction, we developed a high-risk percentage (HRP) versus true incident percentage (TIP) plot. The HRP-TIP plot measures the performance of a model based on the following two criteria: (1) a good model should predict high probabilities for the locations and times when incidents actually happen, and (2)

the total area of the locations with high probabilities should be small at a given time. To generate a HRP-TIP plot, we first compute the following variables:

$$HRP_\delta = \frac{\|\{s_i | p(\text{inci}_{s_i, t_j} = 1) > \delta\}\|}{\|\{s_i\}\|} \quad (2)$$

$$TIP_\delta = \frac{\|\{\text{inci}_{s_i, t_j} = 1 | s_i \in \{s_i | p(\text{inci}_{s_i, t_j} = 1) > \delta\}\}\|}{\|\{\text{inci}_{s_i, t_j} = 1\}\|} \quad (3)$$

In the equations above, $\|\cdot\|$ is the size of a set, and δ is a threshold. HRP represents the percentage of high-risk area predicted by a model, and TIP represents the percentage of incidents (from the test set) that occur within the predicted high-risk area. After computing two vectors of HRP and TIP with different thresholds, we plot TIP against HRP . The curve from a good model should be close to the upper left corner.

The STGAM can incorporate various features as long as they can be represented numerically as a vector or matrix. For example, suppose we have the textual information D_{s_i, t_j} associated with the location s_i and time t_j . If we can represent D_{s_i, t_j} using a numerical vector $X_{s_i, t_j} = (x_{d_1, s_i, t_j}, \dots, x_{d_m, s_i, t_j})$, we can add $x_{d_1, s_i, t_j}, \dots, x_{d_m, s_i, t_j}$ into Equation 1. Doing so adds the textual information to the standard spatio-temporal model. The following section discusses how to represent text using numerical vectors for use within the STGAM.

B. Extracting Textual Information by Semantic Role Labeling-Based Latent Dirichlet Allocation (SRL-LDA)

Extracting information from text and structuring textual data as numerical vectors are fundamental text mining tasks. A widely applied method to structure text is the vector space model, where documents are represented by term-document matrices. Weighting schemes such as term-frequency/inverse-document-frequency are often used in conjunction with vector space models [11]. This type of model represents a collection of documents within a high-dimensional feature space. Recent text mining research has developed probabilistic topic models such as latent Dirichlet allocation (LDA), which reduce a text’s feature space dimensionality [12]. LDA performs well for different tasks like document organization, prediction, and image labeling [13], [14].

The standard LDA model can be represented graphically as shown in Figure 1a. It is a hierarchical Bayesian model that extracts latent variables from a collection of documents. The generative process for each document d in a collection D can be described as follows. First, draw T topics from a Dirichlet distribution $\beta_t \sim \text{Dir}_V(\eta)$, where a topic is a distribution over V words. Second, for each document d , draw topic proportions from another Dirichlet distribution: $\theta_d \sim \text{Dir}_K(\alpha)$. Third, for each word $w_{d,n}$ in the document d , draw a topic $z_{d,n} | \theta_d \sim \text{Multinomial}(\theta_d)$ and then draw a word $w_{d,n} | z_{d,n}, \beta_{1:K} \sim \text{Multinomial}(\beta_{z_{d,n}})$.

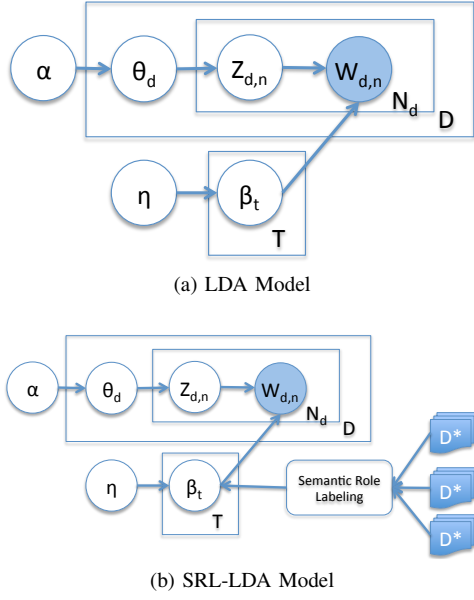


Fig. 1: LDA and SRL-LDA Models

As shown, the generative LDA model regards each document as a bag of words that contains no information beyond the words themselves. We used semantic role labeling (SRL) to incorporate additional semantic information into the LDA process [9]. SRL extracts events mentioned a document, the entities associated with the events, and the roles of the entities with respect to the events. With the use of SRL, we have developed a semantic role labeling-based latent Dirichlet allocation (SRL-LDA) model that can be represented graphically as shown in Figure 1b.² In the SRL-LDA model, the topic distribution β_t is defined in terms of the events extracted from unstructured text. To estimate SRL-LDA, documents are first processed for SRL structure. Non-event words are filtered out and the standard LDA estimation algorithm is applied to the filtered documents to estimate the LDA parameters. With an estimated SRL-LDA model, the numerical vector θ_d can be used to represent the document d . This vector is meaningful: it describes the event-based topics contained in the document d . In aggregate, the numerical vectors θ_d capture the hidden structure of events described in documents. In our previous work, we showed that such vectors can be used to predict future criminal activity [9].

C. Feature Selection by Randomized Least Angle Regression

Feature selection is critical in modeling, especially with many features. Well-selected features not only simplify models, but also improve prediction accuracy. Among available feature selection methods, least angle regression (LAR) [17] is a fast method to select features for linear regression models.

For a problem with n features, LAR computes the order of features entering the regression model and chooses the best number $m(< n)$ of features based on criteria like mean squared error.³ It can select a small subset of features effectively and efficiently for linear regression models. However, if predictors are nonlinearly correlated with the response, it might not be able to select features correctly. For instance, suppose the response variable y is determined as:

$$y = (x_1 - 0.5)^2 + 0 \cdot x_2 \quad (4)$$

where $x_1, x_2 \sim U(0, 1)$ and $U(0, 1)$ denotes the uniform distribution ranging from 0 to 1. Given $D_s = \{x_1, x_2, y\}$, a good feature selection method should rank x_1 with higher priority than x_2 . However, applying LAR to D_s , x_2 is ordered before x_1 .

LAR does not work in the above example because y is quadratically related to x_1 . The linear correlation between these variables is close to zero as shown by the horizontal black line in Figure 2a. Because y is independent of x_2 , the linear correlation between them is also close to zero. Due to the randomness of x_1 and x_2 , the linear correlation between y and x_2 is slightly higher. Since LAR measures the linear correlation between predictors and the response, x_2 is more important than x_1 . This problem might be solved if we sample subsets from D_s randomly. For example, the grey dashed lines in Figures 2a and 2b measure the linear relationships between variables using samples of 10% of the original dataset. As shown, the slopes of the grey lines are generally greater than the slopes of black lines. Also, the variance of the coefficients of the linear models for x_1 and y is greater than the variance for x_2 and y .

We propose a randomized LAR (RLAR) method to introduce randomness into the dataset to improve LAR when the relationships between variables might be nonlinear. The method is described as follows. Instead of applying LAR directly to the dataset $D_{n \times p}$ (n is the number of observations and p is the number of predictors), the method first samples S datasets $D_{m \times q}^s$ from $D_{n \times p}$, where $m < n$ and $q < p$. LAR is then applied to each D^s to rank features. The feature priority for D^s is $rank_s = \langle feature_{s1}, \dots, feature_{sq} \rangle$. In the last step, feature priorities are voted on by $\{rank_s\}$. The complete algorithm is shown in Algorithm 1.⁴ By applying RLAR to the example given in Equation 4, we get the desired ranking of x_1 higher than x_2 .

We compared RLAR with LAR using the Communities and

³If m is the best number, all features entering the regression model before the m^{th} step are selected. Thus, the order of features entering the regression model can be considered as the priority of features. The earlier a feature enters the regression model, the more important it is.

⁴LAR solves the problem with quantitative response variables. If the response variable is binary, the L_1 -regularization path algorithm [18] can be used for generalized linear models. Therefore, if response variables are qualitative, the above RLAR algorithm can still be used by substituting LAR with the L_1 -regularization path algorithm. For simplicity, we still call it RLAR.

²In our study, we used the SRL systems developed by Punyakanok et al. [15] and Gerber and Chai [16].

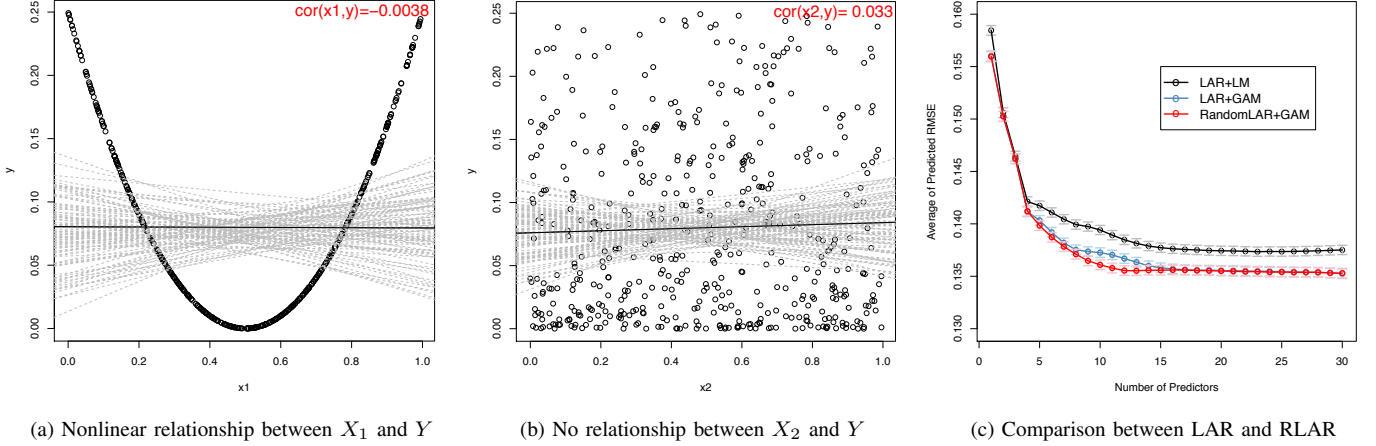


Fig. 2: Randomized LAR

Algorithm 1 Randomized LAR

```

1:  $D = \text{dataset } \{x, y\}$ 
2: for  $s$  from 1 to  $S$  do
3:   sample  $D^s$  from  $D$ 
4:   apply LAR on  $D^s$  to get ranked features  $rank_s = \langle feature_{s1}, \dots, feature_{sp} \rangle$ ; where  $feature_{si}$  is added in the  $i^{th}$  step of LAR algorithm
5: end for
6: for  $i$  in  $c(1:p)$  do
7:   for  $s$  in  $c(1:S)$  do
8:     if  $feature_i \in rank_s$  then
9:        $vote_i = vote_i + r_{si}$ , where  $rank_s[r_{si}] == feature_i$ 
10:       $sample.time_i = sample.time_i + 1$ 
11:     end if
12:   end for
13:    $vote_i = \frac{vote_i}{sample.time_i}$ 
14: end for
15: the smaller  $vote_i$  is, the more important  $feature_i$  is.

```

Crime dataset [19].⁵ The cleaned data include 99 features describing incidents of violent crimes. We used two-thirds of the data for training and the rest as testing to compute root mean squared errors (RMSE). In the comparison, RLAR and LAR were applied to the training set to rank features. Next, the linear and additive models with different numbers of features were built and the RMSE of each model was calculated using the test set. Figure 2c shows the result. The x-axis is the number of top-ranked predictors used to predict the violent crimes. The y-axis is the average predicted RMSE.

⁵The data were cleaned by removing columns with “NA”. The first five columns about community information and the fold number were also removed.

The black line is the linear model using features selected by LAR; the blue line is the additive model using features selected by LAR; the red line is the additive model using features selected by RLAR. As shown in the plot, the additive models performed better than the linear model. Given any number of predictors, RLAR performed the same or better than LAR. RLAR achieved the lowest RMSE with 12 predictors whereas LAR achieved the lowest RMSE with 17 predictors.

III. MODEL APPLICATION AND EVALUATION

This section applies the modeling process described in Section II to breaking and entering crimes in Charlottesville, Virginia. Textual information about the area was extracted from Twitter posts (known as tweets). We evaluated STGAM’s prediction performance with textual information and compared it to STGAM without textual information.

A. Data

We used four datasets in our study. The first dataset contained breaking and entering crimes that occurred in Charlottesville, Virginia during the period of March 1st - October 31st, 2011. In total, there were 88 incidents. The dataset was obtained from local law enforcement agencies.⁶ Each incident was associated with a street address and time. Each street address was mapped to geographic coordinates using the Mapquest API.⁷ The second dataset was a collection of Twitter posts downloaded using Twitter’s publicly accessible API. We used all tweets posted by the CBS19 news agency.⁸ On average, there were 15 tweets per day. The last two datasets contained the geographic and demographic information used in our prior work [8]. The geographic dataset contained information layers such as locations of roads, small businesses, and schools. The demographic dataset measured features of Charlottesville in census block groups, including population and race.

⁶<http://www.charlottesville.org/index.aspx?page=257>

⁷<http://developer.mapquest.com>

⁸<http://www.newsplex.com>

To build the spatio-temporal models, we used a grid size of 0.02 miles \times 0.02 miles and a time interval of 24 hours. There were 23,089 grids within the area of Charlottesville and 5,656,805 spatio-temporal records. Each record had a response variable indicating whether at least one incident occurred within the grid and time interval. Each record also had three types of features describing the characteristics of the space and time. The first feature type captured the minimum distances between the centroid of a grid and certain geographic landmarks. The second feature type captured demographic properties of grids' neighborhoods. The last feature type contained textual information describing the day's news in Charlottesville. All the records located on the same day were associated with the same textual feature. In total, there were 14 distance features, 20 demographic features, and 1 textual feature. A subset of the distance and demographic features is shown in Table I. Features with names ending in “_dist” are distance features and the rest are demographic features. The textual feature contained Twitter posts from CBS19 grouped by date and analyzed using the SRL-LDA method described earlier. To test the model's prediction performance, we used the incident data between October 1st and October 31st. We used the remaining data to estimate the model's parameters.

B. Modeling and Results

We first applied the feature selection method developed in Section II-C to select distance and demographic features. Because of the large size of the training set, we used the sub-sampling technique described in Section II-A with RLAR. The top 20 features with ranks are shown in Table I. To decide the number of features to use, we built STGAM for each different number of features and computed the restricted maximum likelihood (REML) score of each model. The smaller the score, the better the model was. Figure 3 shows the result. The x-axis indicates the number of features used in STGAM. For example, “5” means the STGAM used the five top-ranked features in Table I. The y-axis is the REML score of the corresponding STGAM. The variances of REML scores were estimated by 100 replicates of the model. The grey vertical bars show the 95% confidence intervals of REML scores based on the estimated variances. Based on this figure, we chose 18 features to model the criminal incidents within Charlottesville.

Next, we extracted textual information from the Twitter posts using SRL-LDA, which was described in Section II-B. We grouped Twitter posts by date. Thus, each day was associated with a “document” containing the day's news tweets. We processed all tweets with the verb-based SRL system of Punyakanok et al. [15] and the noun-based SRL system of Gerber and Chai [16]. We filtered out all twitter words that were not events, as indicated by the SRL systems. After the SRL analysis, we trained a 10-topic LDA model using the “topicmodels” package in R. We used the output of topic distributions as the numerical representation of the textual information.

Using textual features as well as distance and demographic features, we built the STGAM to model the criminal incidents

TABLE I: Ranked Features

Rank	Feature	Description
1	small_business_dist	distance to the nearest small business
2	nursehomes_dist	distance to the nearest nurse home
3	nevermarry	number of people who are never married
4	roads_dist	distance to the nearest road
5	vacant	count of vacant houses
6	rivers_dist	distance to the nearest river
7	renter_occ	count of renter-occupied households
8	married	number of people who are married
9	it_hardware_dist	distance to the nearest IT hardware
10	roads_interstates_dist	distance to the nearest interstate highway
11	telecom_services_dist	distance to the nearest telecom services
12	medianrent	median rent charged for all housing units that are rented
13	separated	number of people who are separated
14	telecom_products_dist	distance to the nearest telecom products
15	females	number of females
16	black	number of African American persons
17	hispanic	number of hispanic persons
18	elect_trans_dist	distance to the nearest electricity transmission lines
19	it_services_dist	distance to the nearest IT services
20	owner_occ	count of owner-occupied households

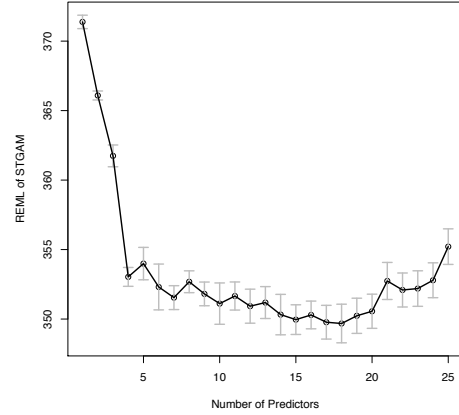


Fig. 3: REML of STGAM with Different Predictors

and predict on the test set. To see whether including text information can help to improve the prediction performance, we also built a STGAM without using textual features. To compare the two models, we used the HRP-TIP plot described in Section II-A. Figure 4 shows the result. In the figure, HRP and TIP are the percentage of high-risk area and the percentage of incidents that occurred within the high-risk area, respectively. Overall, we can see that STGAM using textual information performed better than STGAM without using textual information. More than 60% of actual incidents occurred within the top 20% of the area predicted using STGAM and textual information. This is compared to the 50% of actual incidents captured by STGAM in the same area when not using textual information. We tested the significance of this

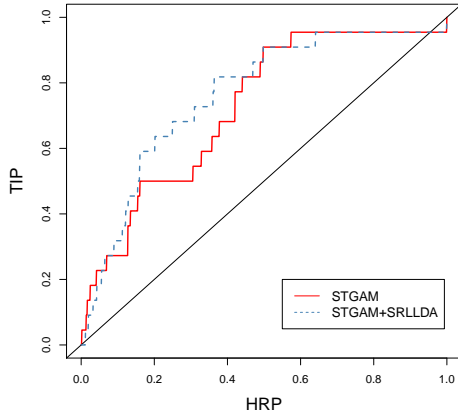


Fig. 4: Predictions for October 2011

difference by computing the differences between points on the blue and red curves at 100 different *HRP* values. A one-sided paired Wilcoxon significance test indicated an exact *p*-value of 2.234×10^{-6} (significant at $p < 0.05$).

IV. CONCLUSION

This paper presents work that brings together two prior approaches to criminal incident modeling. The first approach (STGAM) used numerical features describing the geographic and demographic properties of a region. The second approach used textual information mined from Twitter posts. We evaluated our hybrid model using actual criminal incident data for Charlottesville, Virginia. Our results indicate that the hybrid model exhibits improved prediction performance versus the standard STGAM model. The hybrid model can be generalized to other application areas where unstructured textual information contains indicators relevant to the spatio-temporal properties of events. In addition, this paper has described a new feature selection algorithm. Tests with simulated data and real data showed the algorithm performed better than a classic penalized linear regression model. This algorithm can be applied independently to choose features for nonlinear models.

There are several ways in which this work can be extended. First, one could take advantage of the spatial and temporal extraction capabilities of the SRL systems. In our SRL-LDA model, we ignored textual information describing an event's spatial and temporal location. This information could be used to map tweets to particular spatio-temporal grid locations. This would improve the model's ability to identify textual information that correlates with spatio-temporal criminal incident patterns. Second, instead of only using the words labeled as events to build LDA models, one could use other entities extracted by the SRL systems. For example, a Twitter post describing a shooting should have different modeling

implications depending on whether the shooter is a police officer or a gang member. This type of information is provided by SRL systems; however, we did not include it in our SRL-LDA model. Third, we plan to collect additional data to test the model in different geographical locations and for longer periods of time.

REFERENCES

- [1] C. Maston and US Dept of Justice and Bureau of Justice Statistics, "Criminal victimization in the united states, 2007—statistical tables," *Bureau of Justice Statistics (NCJ 227669)*, <http://bjs.ojp.usdoj.gov/index.cfm?ty=pbdetail&iid=1743>, p. Table 82, 2010.
- [2] C. Block, "STAC hot-spot areas: A statistical tool for law enforcement decisions," in *Crime analysis through computer mapping*. Washington, DC: Police Executive Research Forum. Citeseer, 1995, pp. 15–32.
- [3] H. Liu and D. Brown, "Criminal incident prediction using a point-pattern-based density model," *International journal of forecasting*, vol. 19, no. 4, pp. 603–622, 2003.
- [4] Y. Xue and D. Brown, "Spatial analysis with preference specification of latent decision makers for criminal event prediction," *Decision support systems*, vol. 41, no. 3, pp. 560–573, 2006.
- [5] M. A. Smith and D. E. Brown, "Application of discrete choice analysis to attack point patterns," *Information Systems and e-Business Management*, vol. 5, no. 3, pp. 255–274, June 2007.
- [6] D. Brown, J. Dalton, and H. Hoyle, "Spatial forecast methods for terrorist events in urban environments," *Intelligence and Security Informatics*, pp. 426–435, 2004.
- [7] X. Wang and D. E. Brown, "The spatio-temporal generalized additive model for criminal incidents," in *Intelligence and Security Informatics (ISI), 2011 IEEE International Conference on*. IEEE, 2011, pp. 42–47.
- [8] X. Wang and D. Brown, "The spatio-temporal modeling for criminal incidents," *Security Informatics*, vol. 1, 02/2012 2012. [Online]. Available: <http://www.security-informatics.com/content/1/1/2/>
- [9] X. Wang, M. Gerber, and D. Brown, "Automatic crime prediction using events extracted from Twitter posts," in *Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, College Park, Maryland, April 2012.
- [10] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011, ISBN 3-900051-07-0. [Online]. Available: <http://www.R-project.org>
- [11] G. SALTON, "Term Weighting Approaches in Automatic Text Retrieval," *Information Processing and Management*, vol. 24, no. 5, pp. 513–523, 1988.
- [12] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.
- [13] M. Steyvers and T. Griffiths, "Probabilistic topic models," *Handbook of latent semantic analysis*, vol. 427, 2007.
- [14] D. Blei, L. Carin, and D. Dunson, "Probabilistic Topic Models," *Signal Processing Magazine, IEEE*, vol. 27, no. 6, pp. 55–65, 2010.
- [15] V. Punyakanok, D. Roth, and W.-t. Yih, "The importance of syntactic parsing and inference in semantic role labeling," *Comput. Linguist.*, vol. 34, no. 2, pp. 257–287, 2008.
- [16] M. Gerber, J. Chai, and A. Meyers, "The role of implicit argumentation in nominal SRL," in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Boulder, Colorado: Association for Computational Linguistics, June 2009, pp. 146–154. [Online]. Available: <http://www.aclweb.org/anthology/N/N09/N09-1017>
- [17] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [18] M. Park and T. Hastie, "L1-regularization path algorithm for generalized linear models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 4, pp. 659–677, 2007.
- [19] M. Redmond, "Communities and crime data set." Last accessed: February 2012. [Online]. Available: <http://archive.ics.uci.edu/ml/datasets/Communities+and+Crime>