2016 15th IEEE International Conference on Machine Learning and Applications

Area-Specific Crime Prediction Models

Mohammad Al Boni and Matthew S. Gerber
Department of Systems and Information Engineering, University of Virginia, Charlottesville, Virginia, USA
Email: {ma2sm, msg8u}@virginia.edu

***Abstract*—The convergence of public data and statistical modeling has created opportunities for public safety officials to prioritize the deployment of scarce resources on the basis of predicted crime patterns. Current crime prediction methods are trained using observed crime and information describing various criminogenic factors. Researchers have favored global models (e.g., of entire cities) due to a lack of observations at finer resolutions (e.g., ZIP codes). These global models and their assumptions are at odds with evidence that the relationship between crime and criminogenic factors is not homogeneous across space. In response to this gap, we present area-specific crime prediction models based on hierarchical and multi-task statistical learning. Our models mitigate sparseness by sharing information across ZIP codes, yet they retain the advantages of localized models in addressing non-homogeneous crime patterns. Out-of-sample testing on real crime data indicates predictive advantages over multiple state-of-the-art global models.**

I. Introduction

Crime varies from neighborhood to neighborhood [6, 9]; however, current crime prediction models have ignored this variation, favoring global models due to a lack of observations at finer resolutions [1–3, 8, 10–13]. For example, a global model might estimate the correlation between spatial density of bus stops and the occurrence and non-occurrence of thefts. Public safety officials might use a positive correlation between these variables to justify the allocation of resources (e.g., police patrols) to areas with the highest density of bus stops. However, this global correlation might not hold within each sub-area. The true correlations might be positive in one sub-area and negative in another. Area-specific models can reveal such insights, but building such models is complicated by a lack of training data at finer resolutions.

In this work, we develop and test two area-specific crime prediction models. In the first approach, we use hierarchical models [7]. In the second approach, we use regularized multi-task learning models [4]. Both approaches address the sparsity issue by sharing information across smaller areas during the learning process. Using real crime records from Chicago, Illinois, USA, we test the hypothesis that our area-specific models will outperform global models. Our experiments support our hypothesis: The area-specific models achieve better prediction performance on 12 out of 17 crime types. The following sections present our area-specific models and experiments in the context of prior work on statistical crime prediction. To the best of our knowledge, this is the first reported use and comparison of these methods.

II. Related Work

Researchers have developed several statistical methods for crime analysis and prediction [2, 3]. Kernel density estimation (KDE) has been used to compute retrospective crime visualizations, often called hot-spot maps [3]. As a predictive method, KDE assumes that areas with historically high crime rates are more likely to be victimized in the future than areas with historically low crime rates. The advantages of KDE are that it is simple to implement and applicable in any area with a recorded history of crime locations. However, the KDE model does not consider criminogenic factors (e.g., bus stop density in our example above), and it does not provide explanatory information for the incidence of crime. To improve upon KDE models, researchers have proposed risk terrain modeling (RTM) [2]. RTM models account for a variety of criminogenic factors that vary across crime types. For example, the risk factors of sexual assaults may include proximity to bars, clubs, and schools as well as the distribution of age, gender and wealth [2]. These models improve upon the performance and interpretability of KDE-based methods. Further research has investigated the relationship between crime and social media, demonstrating improved predictive performance when salient topics of discussion from social media are integrated with historical crime density patterns [1, 8, 13].

The methods above estimate global models using information aggregated from many smaller areas. The models assume that the smaller areas are homogeneous with respect to crime occurrence and the various criminogenic factors studied. These models and their assumptions are at odds with evidence that crime is not homogeneous across space [6]. This gap raises the following research questions: (1) Is it possible to build area-specific predictive models using sparse data? (2) Will area-specific models outperform global models, supporting hypotheses about non-homogeneity of crime? We address these questions in the remainder of the paper. Our experiments provide an affirmative answer to each question.

III. Data Collection And Preprocessing

In order to test and evaluate our crime prediction models, we collected historical crime data from the City of Chicago data portal.[1] Each recorded criminal incident includes a case number, a timestamp of occurrence, geo-spatial coordinates of the crime location at block-level resolution, an offense type, and a textual description of the incident.

[1]City of Chicago Data Portal: https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2

978-1-5090-6167-9/16 $31.00 © 2016 IEEE
DOI 10.1109/ICMLA.2016.43

Crime Type	Frequency(%)
THEFT	46,767 (23.82%)
BATTERY	34,509 (17.58%)
NARCOTICS	22,209 (11.31%)
CRIMINAL DAMAGE	19,588 (9.98%)
OTHER OFFENSE	11,731 (5.97%)
BURGLARY	11,642 (5.93%)
ASSAULT	11,357 (5.78%)
DECEPTIVE PRACTICE	9,278 (4.73%)
MOTOR VEHICLE THEFT	7,643 (3.89%)
ROBBERY	7,602 (3.87%)
CRIMINAL TRESPASS	5,441 (2.77%)
WEAPONS VIOLATION	1,987 (1.01%)
PUBLIC PEACE VIOLATION	1,886 (0.96%)
OFFENSE INVOLVING CHILDREN	1,555 (0.79%)
SEX OFFENSE	1,379 (0.70%)
PROSTITUTION	937 (0.48%)
INTERFERENCE WITH PUBLIC OFFICER	836 (0.43%)
Total	196,347

TABLE I: Frequency of historical crime records in Chicago.

Description	Type
Major Streets	Distance
Police Stations	Distance
Police Stations	Density
Hospitals	Distance
Hospitals	Density
Bus stops	Distance
Bus stops	Density
Parks	Distance
Parks	Density
Water ways	Distance
Pedestrian ways	Distance

TABLE II: Spatial features included in our crime prediction models. Distance features indicate linear distance from an analysis point to the spatial entities, whereas density features quantify the spatial density of those entities at that point.

We retained incidents between 2013-07-28 and 2014-04-14. During this window there were 196,347 incidents divided into 17 types. Table I shows the per-crime frequencies. Next, we extracted a number of spatial features from the City of Chicago data portal, shown in Table II. Given a spatial analysis point p, we calculated the linear distance from p to each Distance feature in Table II. We also calculated the spatial density of each Density feature at point p (more details below). Finally, in order to create area-specific models, we segmented Chicago into its 61 existing ZIP codes and grouped the crime data accordingly (Figure 1).

IV. Crime Prediction Models

In this section, we present the mathematical formulation of two global models and three area-specific models.

A. Global Model

The global model estimates the relative risk of crime type T at point p using a set of predictor features as follows:

$$Pr\left(Label_p = T | f_1(\theta_p), ..., f_{12}(\theta_p)\right) = F(f_1(\theta_p), ..., f_{12}(\theta_p)) \quad (1)$$

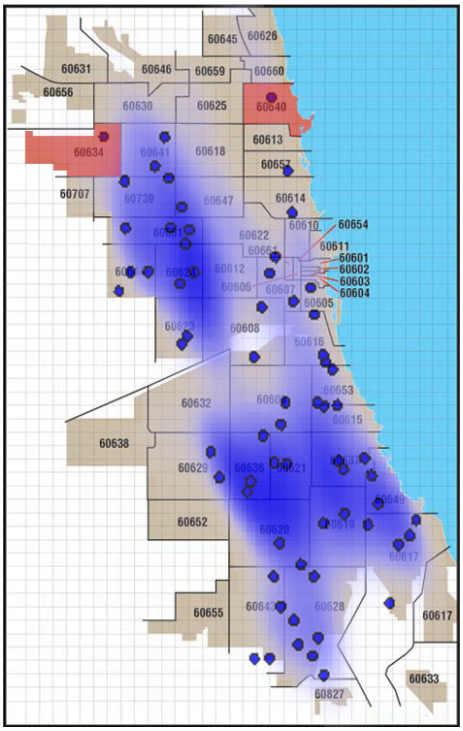

Fig. 1: The 61 ZIP code-based modeling areas within Chicago, overlaid with a KDE of assault locations in blue. Note the sparsity of incidents in areas 60634 and 60640, which are highlighted in the top left and right, respectively.

where $f_1(\theta_p), f_2(\theta_p), ..., f_{12}(\theta_p)$ are features describing point p with parameters θ: $f_1(\theta_p)$ is a kernel density estimate that quantifies the historical crime density. The historical crime KDE at point p is parametrized by $\theta_p = \{p, t_1, t_2\}$ such that

$$f_1(p, t_1, t_2) = k(p, h) = \frac{1}{Ph} \sum_{j=1}^{P} K\left(\frac{||p - p_j||}{h}\right), \quad (2)$$

where P is the total number of crimes of type T that occurred between times t_1 and t_2, h is a smoothing parameter, p is the point at which the density estimate is calculated, $||\cdot||$ is the L-2 norm, and K is a standard normal probability density function; $f_2(\theta_p)$ to $f_{12}(\theta_p)$ are the spatial features from Table II, which are parametrized by $\theta_p = \{p\}$. These spatial features either estimate the linear distance from p to the entity (e.g., linear distance to a major street), or they estimate the density of the physical entity at p, as measured by yet another KDE (e.g., the value at p of the KDE built from police station locations). Finally, F is a link function relating the predictors to the response. We used the logistic function:

$$F(f_1(\theta_p), ..., f_{12}(\theta_p)) = \frac{1}{1 + e^{-\left(\beta_0 + \sum_{i=1}^{12} \beta_i * f_i(\theta_p)\right)}} \quad (3)$$

In words, the global model contains coefficients representing the relationship between (1) crime occurrence and non-occurrence at point p and (2) the historical crime density

and various spatial features of p. These coefficients apply uniformly to the entire city. This is a standard model configuration, and we use it as a baseline.

B. Global Model with Area Indicator (Global-A)

Our second baseline augments the above global model with an additional categorical feature, $f_{13}(\theta_p = \{p\})$, which takes on the value of the ZIP code that covers point p.

C. Pooled Model

Our third baseline model is a pooled model that contains one independently fitted model for each of the 61 ZIP codes covering Chicago. The structure of each ZIP code-specific model is identical to the global model, but fitting is restricted to data in each area. Given infinite training observations, one would expect the pooled model to outperform the global model under the hypothesis that crime is not homogeneous across spatial regions [6]. In practice, many areas have sparse data, thus complicating the direct application of pooled models. In the next two sections we introduce area-specific models that mitigate sparsity by sharing information across different areas during the learning process.

D. Hierarchical Model

Hierarchical models, also known as multilevel models, provide a framework for explicitly capturing the structure that may exist in categorical variables [7]. For example, one may wish to measure the performance of students in different classes within various schools, which in turn reside in districts. This structure allows model parameters to vary across levels in the model hierarchy. In the present work, we used a varying-intercept hierarchical model given by

$$F(f_1(\theta_p), ..., f_{12}(\theta_p)) = \frac{1}{1+e^{-\left(\beta_{0[j]} + \prod_{i=1}^{12} \beta_i * f_i(\theta_p)\right)}}$$
$$\beta_{0[j]} \sim N\left(U_j, \sigma^2_{\beta_0}\right) \quad (4)$$

where $\beta_{0[j]}$ is the intercept of area j. Herein, the variability of the intercepts between different areas is assumed to be Gaussian with a mean U_j given by each area's data, and the within-area variance $\sigma^2_{\beta_0}$.

E. Multi-Task Model

Lastly, we implemented a regularized multi-task model [4]. This approach uses a kernel function with a task-coupling parameter to model each area as a separate task. The proposed kernel function overcomes the sparsity of area-specific models by encoding the relationship between tasks. The kernel is defined as

$$\phi\left((x, j)\right) = \left(\frac{x}{\sqrt{\mu}}, \underbrace{0, ..., 0}_{j-1}, x, \underbrace{0, ..., 0}_{J-j}\right) \quad (5)$$

where $j = 1, ..., J$ is the area index such that J is the total number of areas (61 in the case of Chicago), x is a vector of values $\langle f_1(\theta_p), ..., f_{12}(\theta_p)\rangle$, and μ is a trade-off parameter that balances the shared contribution of the tasks. Under this representation, each observation will fill in its values at the appropriate ZIP code location and in a shared location weighted by $\frac{1}{\sqrt{\mu}}$.

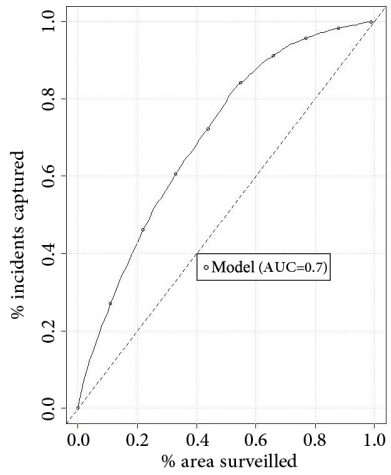

Fig. 2: An example surveillance plot, which shows the proportion of true future crimes (y-axis) that occur within the $x\%$ most threatened area predicted by the model (x-axis). This plot is summarized by the area under the curve (AUC) value.

V. Experimental Design and Evaluation

We evaluated the above models by creating training and testing sets. First, we built a grid of 200-meter squares that covered Chicago. We then labeled the center point of each square with NULL, forming a set of negative training observations. For the training phase, we combined the NULL points with true incident points for a crime type *T* to form the final training data, and we fit the models from Section IV on that training data. In cases where a negative point coincided with a positive one, we removed the negative point from the training data set. During testing, we used the fitted models to make predictions at each point in the 200-meter grid as well as at the crime points observed during the training phase. The rationale for making predictions at historically observed crime points is that we would like finer-grained predictions in areas of historically high crime. For the pooled model, we further divided both training and testing data points into 61 subsets based on their geo-location within the different ZIP code areas. We set the initial training window to the first month (7/27/2013 to 8/27/2013) and the initial prediction window to the first day following the training window. This design reflects a practical configuration where police officials use recent crime records to predict the next day's crime. At each time step, the training and testing windows are shifted one day into the future, the models are refit, a prediction for the new testing day is made, and the testing results are accumulated into an overall result.

We evaluated all predictions using surveillance plots, an example of which is shown in Figure 2. A surveillance plot shows the proportion of true future crimes (y-axis) that occur within the $x\%$ most threatened area predicted by the

model (x-axis). Figure 2 shows that around 60% of future crime incidents (y-axis) are captured by the top 35% most threatened area (x-axis), as predicted by the model. The curves for optimal predictions approach the top-left corner, indicating that the predicted crime risks reflect the true spatial distribution of crime. We summarize surveillance plots with scalar area under the curve (AUC) values. Surveillance plots offer two advantages over traditional ROC curves for the crime prediction task. First, surveillance plots provide decision makers with a better visualization of the underlying physical environment and crime process, as the x-axis is tied directly to the physical space being modeled. Second, traditional ROC curves assume that the set of positive and negative instances is known with certainty. In the crime domain, the set of positive points is well defined: it is the set of points at which a crime occurred and was recorded by the police. The set of negative crime points is problematic. If a crime was reported at a specific latitude-longitude location, should a point 5 meters away from this location be given a ground-truth negative (NULL) label? What about 10 meters or 100 meters? The absence of a justifiable distance threshold for negative points renders the negative ground-truth indeterminate and precludes the application of ROC plots. Surveillance plots capture the same intuition as ROC plots (i.e., assign positive points higher probabilities than negative points) but rely only on the set of ground-truth positive points.

Since we ran our prediction models on multiple testing days, we required a method for aggregating the daily surveillance plots and estimating the overall performance of the models. Gerber proposed a micro-level aggregation method for surveillance plots [8]. In this approach, one sums the number of true crimes of type T occurring within the $x\%$ most threatened area for each day, according to the model's predictions. For example, assume that we want to aggregate the surveillance plots for two testing days d_1 and d_2. Also assume the following: (1) the prediction area contains 100 square prediction cells; (2) the prediction for d_1 ranks cell c_{41} as most threatened, and the prediction for d_2 ranks cell c_{15} as most threatened; (3) 50 actual crimes occurred in c_{41} on day d_1 and 20 actual crimes occurred in c_{15} on d_2; and (4) there were 100 actual crimes on day d_1 across the city and there were 25 actual crimes on day d_2 across the city. The resulting y-value for $x = 1\%$ in the aggregated surveillance plot is calculated as $\frac{50+20}{100+25}$. In words, this y-value is the fraction of crime occurring in the aggregated, most-threatened 1% of the area as predicted by the model.

VI. Experimental Results and Evaluation

We evaluated our models in the following way: (1) we chose 70 random testing days between 2013-08-28 and 2014-04-14; (2) we set the training window size to one month; and (3) we trained Global, Global-A, and Pooled models using LibLinear's L2-regularized logistic regression solver (L2R_LR) [5]; multi-task models using LibLinear's L2-regularized SVM (L2R_L2LOSS_SVC) [5]; and hierarchical models using Glmer from the Lme4 package in R. Table III

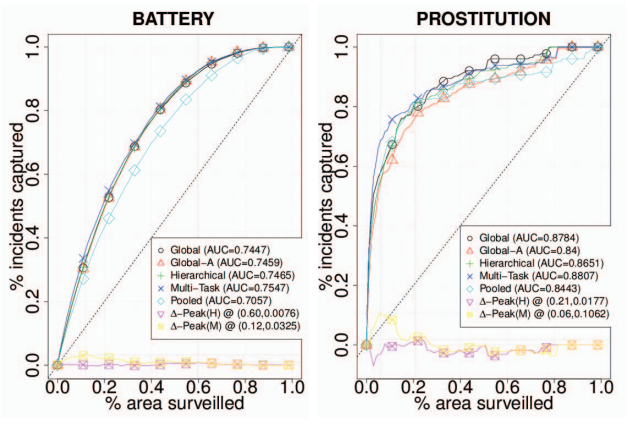

Fig. 3: Aggregated surveillance plots of battery and prostitution crimes. Gains from using the hierarchical model are calculated as Hierarchical - Global, and this gain peaks at the location indicated by $\Delta - Peak(H)$. The gains from using the multi-task model are calculated as Multi-Task - Global, and this gain peaks at the location indicated by $\Delta - Peak(M)$.

shows the aggregated AUC of the surveillance plots from the 70 prediction days. Next, we compared the performance of the Global, Global-A and Pooled models in order to select a single baseline for comparison with our two area-specific models. Table III shows that Pooled models were outperformed by Global and Global-A on all crime types, Global outperformed Global-A on 8 crime types, and Global-A outperformed Global on the remaining 9 crime types. However, we noticed that in most of these 9 crime types, the gain of Global-A over Global is not significant. To further explore the difference between these two models, we investigated the peak gains in the surveillance plots (points of maximal difference between the models). We found that the Global model had higher peak gains than the Global-A model in 10 out of 17 crime types. As a result, we selected the Global model as our baseline in the remaining analyses. Our area-specific models exhibit improved performance over the Global model for 12 out of 17 crime types (see Table III).

In practice, police departments cannot cover more than a small fraction of the city, so we emphasize gains over global models that are achieved for small values on the x-axis of our aggregate surveillance plots. For example, Table III indicates that the multi-task model gains 0.01 over the global model for battery and 0.0023 over the global model for prostitution. This might suggest that the gains were larger for battery; however, the gains are calculated assuming 100% surveillance of the study region, which is infeasible given limited resources. If we instead inspect the x-axis location of peak gains achieved by the multi-task model over the global model (Figure 3), we see that, for battery, peak gain of 3.25% is achieved at 12% surveillance versus peak gain of 10.6% being achieved at 6% surveillance for prostitution. Thus, multi-task modeling

Crime Type	Global	Global-A	Pooled	Hierarchical	Multi-Task
THEFT	0.7433	0.7500	0.7088	0.7513	**0.7603**
BATTERY	0.7447	0.7459	0.7057	0.7465	**0.7547**
NARCOTICS	0.8402	**0.8484**	0.7846	0.8482	0.8469
CRIMINAL DAMAGE	0.6886	0.6935	0.6515	0.6954	**0.6989**
OTHER OFFENSE	0.7076	0.7088	0.6678	0.7102	**0.7138**
BURGLARY	0.7031	0.7092	0.6768	0.7098	**0.7157**
ASSAULT	0.7382	0.7386	0.6901	0.7405	**0.7439**
DECEPTIVE PRACTICE	0.7723	0.7752	0.7269	0.7771	**0.7796**
MOTOR VEHICLE THEFT	0.7023	0.6992	0.6520	0.7022	**0.7051**
ROBBERY	0.7786	0.7747	0.7395	0.7810	**0.7832**
CRIMINAL TRESPASS	0.7874	0.7893	0.7620	0.7933	**0.8050**
WEAPONS VIOLATION	**0.7853**	0.7726	0.7421	0.7843	0.7744
PUBLIC PEACE VIOLATION	**0.7719**	0.7461	0.7121	0.7675	0.7614
OFFENSE INVOLVING CHILDREN	**0.6997**	0.6674	0.6557	0.6934	0.6825
SEX OFFENSE	**0.7080**	0.6681	0.6016	0.7006	0.6978
PROSTITUTION	0.8784	0.8400	0.8443	0.8651	**0.8807**
INTERFERENCE WITH PUBLIC OFFICER	0.8169	0.7823	0.7621	**0.8187**	0.7756

TABLE III: Performance on 70 random prediction days with respect to the 17 crime types.

Crime Type	Δ-Peak (H-G)		Δ-Peak (M-G)		Δ-Peak (G-GA)		Δ-Peak (G-P)	
	X	Y	X	Y	X	Y	X	Y
THEFT	0.43	0.0144	0.35	0.0394	0.01	0.0042	0.03	0.0689
BATTERY	0.60	0.0076	0.12	0.0325	0.18	0.0055	0.30	0.0785
NARCOTICS	0.12	0.0155	0.37	0.0141	0.04	0.0057	0.05	0.1926
CRIMINAL DAMAGE	0.19	0.0289	0.19	0.0360	0.54	0.0033	0.42	0.0684
OTHER OFFENSE	0.11	0.0190	0.06	0.0356	0.55	0.0130	0.39	0.0752
BURGLARY	0.28	0.0256	0.37	0.0332	0.65	0.0068	0.46	0.0506
ASSAULT	0.35	0.0111	0.12	0.0202	0.30	0.0077	0.28	0.0956
DECEPTIVE PRACTICE	0.02	0.0287	0.02	0.0353	0.14	0.0107	0.15	0.1050
MOTOR VEHICLE THEFT	0.02	0.0157	0.44	0.0141	0.35	0.0204	0.36	0.1027
ROBBERY	0.21	0.0219	0.21	0.0225	0.08	0.0209	0.13	0.0945
CRIMINAL TRESPASS	0.27	0.0189	0.12	0.0447	0.23	0.0161	0.15	0.0692
WEAPONS VIOLATION	0.65	0.0174	0.56	0.0282	0.16	0.0521	0.15	0.1323
PUBLIC PEACE VIOLATION	0.19	0.0216	0.11	0.0275	0.13	0.0589	0.32	0.1139
OFFENSE INVOLVING CHILDREN	0.04	0.0167	0.16	0.0335	0.37	0.0766	0.52	0.0885
SEX OFFENSE	0.65	0.0172	0.18	0.0172	0.36	0.1043	0.31	0.2546
PROSTITUTION	0.21	0.0177	0.06	0.1062	0.01	0.1062	0.79	0.0841
INTERFERENCE WITH PUBLIC OFFICER	0.07	0.0534	0.05	0.0243	0.45	0.0777	0.09	0.1505

TABLE IV: Peak gains in the aggregate surveillance plot's y-value when using Hierarchical (H) and Multi-task (M) models instead of the global (G) model. We chose to include the peak gains over the Global model instead of the Pooled (P) and Global-A (GA) models since the latter were outperformed by other models for most crime types. The x-values indicate percentage area surveilled according to model predictions, and the y-values indicate gains achieved by surveilling this $x\%$ according to the H or M models instead of the G model. In words, the first row indicates (1) that, when surveilling 43% of the city, one can capture 1.4% more crime by prioritizing according to H rather than G, and (2) that, when surveilling 35% of the city, one can capture 3.9% more crime by prioritizing according to M rather than G.

produces greater returns on investment for prostitution versus battery (assuming uniform costs across crime types). Table IV shows the peak gain for hierarchical and multi-task models over the Global baseline: 10 out of 17 and 11 out of 17 crime types have their peak gains within 25% surveillance using hierarchical and multi-task models respectively. Table IV also shows the peak gain for the Global model over the Global-A and Pooled models. The latter two baselines offer straightforward solutions for area-specific modeling; however, because of crime sparseness, both area-specific baselines failed to outperform the simple Global baseline. This emphasizes the robustness of our area-specific hierarchical and multi-task models against sparse observations.

To better understand the impact of using area-specific models, we computed the number of crimes that could be targeted when patrolling according the various models (Table V). For example, when patrolling 5% of the city following the multi-task model instead of the Global model, police would have been in the vicinity of 94 additional thefts and 224 additional batteries. The improved patrol focus could have increased deterrence and decreased response time for undeterred crimes. It is also important to consider the potential social impact of improvements due to area-specific modeling. As a measure of severity, we considered prison sentences for the various crime types. For example, Class-1 felonies such as sexual assaults, residential burglaries, and possession of heroin or cocaine have sentencing periods of 4 to 15 years, while Class-4 felonies such as aggravated assault have sentencing periods of 1 to 3

Crime Type	5%			10%			Δ-Peak (H-G)		Δ-Peak (M-G)	
	Global	Hierarchical	Multi-Task	Global	Hierarchical	Multi-Task	Global	Hierarchical	Global	Multi-Task
THEFT	3112	-11 (↓ 0.4%)	**94 (↑ 3.0%)**	4277	77 (↑1.8%)	**224(↑ 5.2%)**	9199	173(↑1.9%)	8388	453(↑5.4%)
BATTERY	1682	0 (↑ 0.0%)	**224 (↑13.3%)**	2803	9 (↑0.3%)	**272(↑ 9.7%)**	8395	70(↑0.8%)	3001	298(↑9.9%)
NARCOTICS	2655	**21 (↑ 0.8%)**	-22 (↓ 0.8%)	3383	**77 (↑2.3%)**	39(↑ 1.2%)	3495	90(↑2.6%)	4984	80(↑1.6%)
CRIMINAL DAMAGE	653	**65 (↑10.0%)**	64 (↑ 9.8%)	1135	58 (↑5.1%)	**110(↑ 9.7%)**	1874	130(↑6.9%)	1874	173(↑9.2%)
OTHER OFFENSE	481	27 (↑ 5.6%)	**107 (↑22.2%)**	809	57 (↑7.0%)	**82(↑10.1%)**	876	37(↑4.2%)	553	105(↑19.0%)
BURGLARY	447	40 (↑ 8.9%)	**42 (↑ 9.4%)**	737	39 (↑5.3%)	**64(↑ 8.7%)**	1573	66(↑4.2%)	1918	90(↑4.7%)
ASSAULT	561	-24 (↓ 4.3%)	**-1 (↓ 0.2%)**	892	13 (↑1.5%)	**54(↑ 6.1%)**	2096	27(↑1.3%)	956	60(↑6.3%)
DECEPTIVE PRACTICE	853	16 (↑ 1.9%)	**34 (↑ 4.0%)**	1096	16 (↑1.5%)	**18(↑ 1.6%)**	528	70(↑13.3%)	528	86(↑16.3%)
MOTOR VEHICLE THEFT	268	**-4 (↓ 1.5%)**	-9 (↓ 3.4%)	438	**19 (↑4.3%)**	0(↑ 0.0%)	84	30(↑35.7%)	1443	22(↑1.5%)
ROBBERY	465	-3 (↓ 0.6%)	**20 (↑ 4.3%)**	728	21 (↑2.9%)	**32(↑ 4.4%)**	1120	34(↑3.0%)	1120	39(↑3.5%)
CRIMINAL TRESPASS	443	13 (↑ 2.9%)	**60 (↑13.5%)**	629	21 (↑3.3%)	**53(↑ 8.4%)**	982	27(↑2.7%)	656	64(↑9.8%)
WEAPONS VIOLATION	106	**-3 (↓ 2.8%)**	-14 (↓13.2%)	180	**-6 (↓-3.3%)**	-15(↓-8.3%)	438	8(↑1.8%)	424	11(↑2.6%)
PUBLIC PEACE VIOLATION	109	-9 (↓ 8.3%)	**5 (↑ 4.6%)**	174	-5 (↓-2.9%)	**14(↑ 8.0%)**	278	6(↑2.2%)	190	12(↑6.3%)
OIC	69	-2 (↓ 2.9%)	**0 (↑ 0.0%)**	109	**6 (↑5.5%)**	1(↑ 0.9%)	58	4(↑6.9%)	143	14(↑9.8%)
SEX OFFENSE	47	0 (↑ 0.0%)	0 (↑ 0.0%)	85	-7 (↓-8.2%)	**0(↑ 0.0%)**	318	4(↑1.3%)	124	6(↑4.8%)
PROSTITUTION	131	-7 (↓ 5.3%)	**24 (↑18.3%)**	152	-1 (↓-0.7%)	**19(↑12.5%)**	181	3(↑1.7%)	135	22(↑16.3%)
IWPO	101	**3 (↑ 3.0%)**	-6 (↓ 5.9%)	152	**1 (↑0.7%)**	-22(↓-14.5%)	81	8(↑9.9%)	65	3(↑4.6%)

TABLE V: Number of criminal incidents captured by global, hierarchical and multi-task models at different surveillance levels: 5%, 10%, and the percentage at which the peak gain is achieved. We show the gain in true counts as well as in percentages of increase or decrease. Note, OIC is offense involving children, and IWPO is interference with public officer.

years.[2] Our area-specific models offer significant social impact through improved prediction of the most severe crimes within small surveillance areas (5% to 25% most-threatened areas versus the Global baseline, see Table IV).

VII. Conclusions

The sparsity of crime in many areas complicates the application of area-specific predictive modeling. In this work, we developed and tested area-specific crime prediction models using hierarchical and multi-task learning. These approaches mitigate sparseness by sharing information across different areas, yet they retain the advantages of localized models in addressing non-homogeneous crime patterns. We tested our models on crime data from the city of Chicago, Illinois, and we observed gains in many surveillance-feasible regions of the study area. In the future, we plan to investigate the use of area-grouping approaches to build spatial hierarchies. This would help in guiding the sharing of information across areas within the hierarchical and multi-task models. This would also highlight similar areas for uniform intervention, thus reducing resource expenditures.

References

[1] M. Al Boni and M. S. Gerber. Predicting crime with routine activity patterns inferred from social media. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 2016.

[2] J. M. Caplan and L. W. Kennedy. Risk terrain modeling compendium. *Rutgers Center on Public Security, Newark*, 2011.

[3] J. Eck, S. Chainey, J. Cameron, and R. Wilson. Mapping crime: Understanding hotspots. Technical report, National Institute of Justice, 2005.

[4] T. Evgeniou and M. Pontil. Regularized multi–task learning. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 109–117. ACM, 2004.

[5] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. Liblinear: A library for large linear classification. *The Journal of Machine Learning Research*, 9:1871–1874, 2008.

[6] M. Felson. Linking criminal choices, routine activities, informal control, and criminal outcomes. In D. B. Cornish and R. V. Clarke, editors, *The reasoning criminal: Rational choice perspectives on offending*, chapter 8. Transaction Publishers, 1986.

[7] A. Gelman and J. Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, 2006.

[8] M. S. Gerber. Predicting crime using twitter and kernel density estimation. *Decision Support Systems*, 61:115–125, 2014.

[9] S.-Y. Kuo, S. J. Cuvelier, C.-J. Sheu, and J. S. Zhao. The concentration of criminal victimization and patterns of routine activities. *International journal of offender therapy and comparative criminology*, 56(4):573–598, 2012.

[10] H. Liu and D. E. Brown. A new point process transition density model for space-time event prediction. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 34(3):310–324, 2004.

[11] M. A. Smith and D. E. Brown. Discrete choice analysis of spatial attack sites. *Information Systems and e-Business Management*, 5(3):255–274, 2007.

[12] X. Wang and D. Brown. The spatio-temporal modeling for criminal incidents. *Security Informatics*, 02 2012.

[13] X. Wang, M. S. Gerber, and D. E. Brown. Automatic crime prediction using events extracted from twitter posts. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 231–238. Springer, 2012.

[2] Illinois prison talk:http://www.illinoisprisontalk.org/index.php?topic=23719.0