Deep Convolutional Neural Networks for Spatiotemporal Crime Prediction

Lian Duan[1, 2], Tao Hu[3], En Cheng[4], Jianfeng Zhu[5], Chao Gao[6*]
[1]Geography Science and Planning School, Guangxi Teachers Education University, Nanning, China
[2.] Education Ministry Key Laboratory of Environment Evolution and Resources Utilization in Beibu Bay, Guangxi Teachers Education University, Nanning, China
[3]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China
[4] College of Arts and Sciences, University of Akron,, Akron, Ohio, USA
[5] Digital of School, Kent State University. Kent, Ohio, USA
[6] Key Laboratory of Police Geographic Information Technology, Ministry of Public Security, Changzhou Municipal Public Security Bureau, Changzhou, China
*Corresponding author, e-mail: gaoch_ren@163.com

Abstract - *Crime, as a long-term global problem, has been showing the complex interactions with space, time and environments. Extracting effective features to reveal such entangled relationships to predict where and when crimes will occur, is becoming a hot topic and also a bottleneck for researchers. We, therefore, proposed a novel Spatiotemporal Crime Network (STCN), in an attempt to apply deep Convolutional Neural Networks (CNNs) for automatically crime-referenced feature extraction. This model can forecast the crime risk of each region in the urban area for the next day from the retrospective volume of high-dimension data. We evaluated the STCN using felony and 311 datasets in New York City from 2010 to 2015. The results showed STCN achieved 88% and 92% on F1 and AUC respectively, confirming the performances of STCN exceeded those of four baselines. Finally, the predicted results was visualized to help people understanding its linking with the ground truth.*

Keywords: Crime Spatiotemporal Prediction; Crime Risk Estimation; CNN (Convolution Neural Networks); Deep Learning; Urban Computing

1 Introduction

There has been increased concern about spatiotemporal predictive policing that forecasts the future crime risks for each community in the urban area, since it provides the significant aids for law enforcements to better identify the underlying patterns behind the crime propagations, efficiently deploy the limited police resources and improve the public safety [3] . However, the task of forecasting where and when the crimes will occur is inherently difficult because it is sensitive to the highly complex distributions of crimes in space and time. Whereas, recent literature recognized that crime incidents tend to exhibit spatial and temporal dependencies with the dynamical social environments [1] [2] . The spatial dependencies denote the crime risks of a region are affected by the crime-related events or environment factors in its spatial proximate regions as well as distant regions. For example, empirical evidence identified that burglary offenders, bars, incomes, race populations [4] and traffics [13] were statistically related to the spatial concentration of crime. On the other side, the temporal dependencies mean the crime risks of a region are influenced by the crime-related events or environment factors at recent, near and even distant time intervals. For example, the near-repeated patterns found in crimes indicated the recent frequently crimes are considered to be the powerful variable for predicting local crime risks in the immediate future [4] . Moreover, dynamical social events such as 311 or 911 incidents, may imply high crime risks in the near spatiotemporal scope [8] . In summary, the challenge in crime prediction centered on identifying the effective spatiotemporal dependencies from the dynamic interplay of crimes between space, time and environmental factors.

In order to address this challenge, scholars have incorporated spatiotemporal point processes [9] and random space–time distribution hypothesis [10] for crime predictions by modeling the crime spatial propagations. The risk terrain analysis [11] , geography regressive models and Bayesian models [12] were also developed to assess how multiple environment factors contribute to future crime risks. Taking the dynamical correlations between crimes and other social activities into consideration, recently studies have explored various feature engineering approaches to characterize the crime-related features regarding Foursquare data [6] , Twitter data [7] , 911 events [8] and taxi trajectories [13] , for enhancing the predictive power. However, most of the studies required profound crime knowledge and resorted to complicated spatiotemporal analysis or cumbersome feature engineering processes. Furthermore, once a bit of data changed, many features must be re-analysis and re-engineered by hand again. Thus, it is not only a challenging to capture valid spatiotemporal dependencies from geo-crime data automatically and efficiently, but the performances of these models were also significantly affected by artificial features.

Recently, deep convolution neural networks (CNNs), which used multiple-layer architectures to extract the inherent local correlations of data from the lowest to the highest levels, were developed to address the feature extraction issues from high-dimension data. They have led to marked improvements in many domains, especially of computer vision [14] and natural language processing [15] . Similar to the pixels in an image or words in a sentence, social incidents such as crimes incidents and traffic flows, exhibit multiple local spatial dependencies. Accordingly, studies of handling big geo-datasets using deep CNNs for poverty prediction [16] , traffic congestion prediction [17] , crowd flows prediction [23] and air pollution prediction [18] , have become available.

Hence, based on deep CNNs, this paper presented a novel Spatiotemporal Crime Network (STCN), drawing on crime data — along with 311 data, to provide insight into the automatically spatiotemporal dependencies abstraction that was tied to the crime risk prediction. The main contributions of this paper are the follow: (1) A novel deep CNNs framework is firstly proposed for the end-to-end crime spatiotemporal prediction, which makes it particularly flexible and minimizes feature engineering bias; (2) The combination of inception networks and fractal networks in a unified framework made it be able to automatically learn the nearby and distant spatiotemporal dependencies between crime and 311 incidents. (3) We evaluated STCN using crime and 311 datasets of New York City during 2010 and 2015, with the results showing that our model outperformed four well-known baselines on F1 and AUC.

The remainder of this paper is structured as follows. Section 2 illustrates the architecture and learning procedure of proposed STCN model. Section 3 validates the effectiveness of the proposed algorithm through the numerical experiments. Finally, we present the concluding remarks in section 4.

2 Preliminaries

This section described the datasets and basic definitions, then gave the formal problem.

2.1 Datasets

The raw datasets[1] came from New York City (NYC) over a span of Jan 1st 2010 and Dec 31th 2015, including felony data and 311 data.

Felony dataset: There were 653,447 incidents. Each incident had the properties of coordinates, offense type and time. We removed 8,000 incidents without the borough information and 8 incidents with abnormal coordinates. The left 645,439 incidents were used in our experiment.

311 Dataset: This dataset had information about 10 million complaint records. About 0.12 million complaint records were omitted as they did not contain coordinates or fall out of the sector boundaries. Therefore, the final 311 dataset contained 9.75 million records.

2.2 Formulation of Crime Prediction Problem

Definition 1 (**Region**): As shown in Figure 1(a) and (b), we devided the study area into disjoint 120×100 grids, $G = \{g_1, g_2, \cdots, g_{120\times 100}\}$. Each grid in G is considered as a *region*. The colored grids denote at least one crime occurred during 2010 to 2015 in NYC. The area of each region is 0.18 km^2 (0.47 km ×0.38 km). Such fine spatial scale is well-aligned with the police resource deploying [19] .

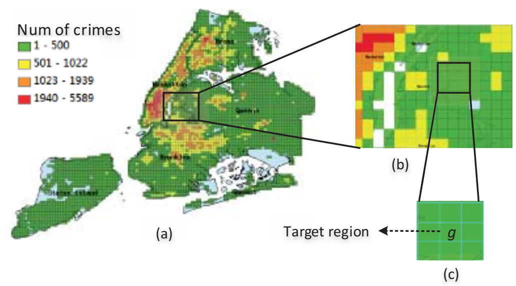


Figure 1 Grid-based map

Definition 2 (**Spatial neighbor set**): As shown in **Figure 1**(c), a *target region g* is where we'd like to forecast whether a crime is most likely to occur. The 3×3 regions geographically surrounding g are called the *spatial neighbor set* of g.

Definition 3 (**Time window**): The duration of M days is regarded as a *time window*.

Definition 4 **(Crime feature** and **311 feature)** $\mathrm{x}_{crime}^{g} \in \mathbb{R}^{M\times N}$ is used to present the crime feature in the form of a 2-D data structure, in which M indicates the length of the time window and N denotes the total region counts in spatial neighbor set . Each entry in x_{crime}^{g} indicates the crime number. As shown in Figure 2, the three grid-nets in the 3-D coordinate system express the crime numbers of regions during the three continuous days (a time window with M=3). We obtain $\mathrm{x}_{crime}^{g} \in \mathbb{R}^{9\times 3}$ by mapping the spatial neighbors set of g along with three continuous days.

[1] NYC Open Data (https://data.cityofnewyork.us/)

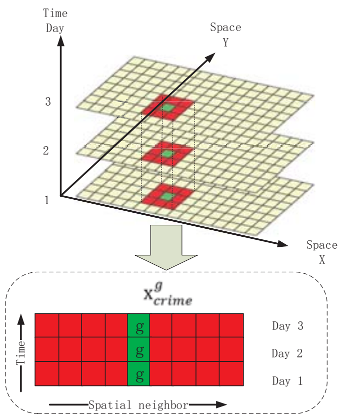


Figure 2 x^{g}_{crime} constructed from raw data

Also, we define the **311 feature** x^{g}_{311} in the same way.

Definition 5 **(Observation)** For the target region g at day t, the observation can be represented as $\mathbf{X}^{g}_{t} \in \mathbb{R}^{M\times N\times 2}$. It composes of x^{g}_{crime} and x^{g}_{311}, as shown in Figure 3.

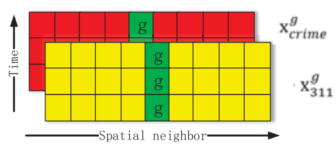


Figure 3 Observation structure

Definition 6 **(Crime label)** y^{g}_{t} is the crime numbers for target region g at day t, which is represented as crime label.

Each sample, $\{X^{g}_{t}, y^{g}_{t+1}\}$, is generated by combining the observation and the crime label. However, we found 93% of total samples were labeled with 0, resulting the class imbalance problem. Two approaches were hence utilized to address it. The first is to collapse the number of *Crime label* categories to two, that means y^{g}_{t} is either 0 or 1. Secondly, we under-sample [8] 0 class for 50% and over-sample 1 class according to SMOTE [26] .

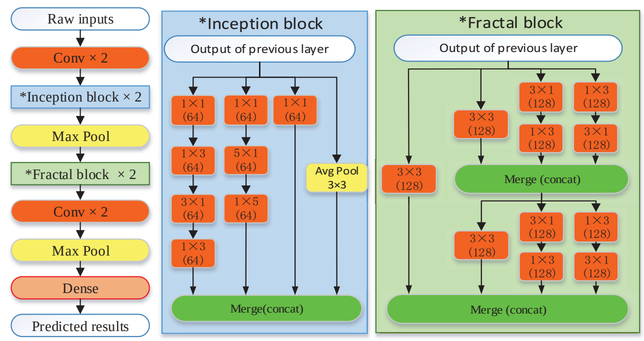
Problem Given the observation $\mathbf{X}^{g}_{t}$, the goal is to estimate whether y^{g}_{t+1} is 0 or 1.

3 Methodology

This section elaborated how the deep CNN architecture resolves the problem.

3.1 Architecture

Figure 4 shows an overview of the proposed STCN model, with its parameters in Table 1. Particularly, The crime and 311 data are converted into two 2D image-like arrays (Definition 5) as input feature maps. By passing them into a sequence of two convolution layers, the model is able to capture the low-level spatiotemporal dependencies for crimes incidents. As the output feature maps flowing deep in the networks, the model began to abstract the high-level spatiotemporal features mainly relying on inception blocks and fractal blocks, which leverage branches of convolutional layers and merge layer to fuse the crime-related features with different abstract levels. Finally, the highest-level crime-related features were aggregated into the dense layer, which acts as the classifier in the framework, to realize the crime risk prediction. The number of each type of layer /block and the corresponding position (depth) in the networks referenced architectures described in [21] [22] and were further determined by the experiments.



(a) STCN (b) Inception (c) Fractal

Figure 4 framework. Red layer is convolution layer. In (a), "×" means the number of components stacked, "*" denotes it as an abbreviation of the structures depicted in (b) or (c), where the numbers denote the size and counts of kernels.

Table 1 parameters of STCN

Position	Layer (Stacked num)	Kernel size (num)	Output size
0	Raw input (1)	-	60×9×2
1-2	Conv (2)	3×3 (32)	60×9×32
3-13	*Inception (2)	-	60×9×64
14	Max Pool (1)	3×3 (1)	30×4×64
15-27	*Fractal (2)	-	30×4×128
28	Conv (2)	3×3 (64)	30×4×64
29	Max Pool (1)	5×5 (1)	64
30	Dense(1)	64	2

3.2 Spatiotemporal dependencies learning within CNNs

3.2.1 Convolution Operation

A convolution operation can naturally captures a close spatiotemporal dependency for crime incidents, as shown in formula (1) ,

$$f_l(x) = f\left(\sum_{j=1}^{N_l} w_{lj}x_j + b_l\right) = f(\mathrm{w}_l^T \mathrm{x}_l + b_l)\quad(1)$$

where l indicates the l_{th} layer, N_l indicates the size of the kernel, x_j indicates the features of crime (or 311) of regions during the period both covered by the kernel, $w_{l.}$ is the learnable weight in the kernel, and b_l is the linear bias. The $f(.)$ is a ReLU activation function [14] which decides whether a neuron in this layer can be activated, thus a particular spatiotemporal feature is captured. Furthermore, a stack of convolution operations catches the mutual influences of crimes among distant spatiotemporal scope [23] automatically. Figure 5 demonstrated these self-learning processes. A neuron in layer 1 merely capture a small-

ISBN: 1-60132-463-4, CSREA Press ©

scale spatiotemporal dependencies across 3 × 3 range in layer 0 . Compared to it, a neuron in layer 2 is able to describe a large-scale spatiotemporal dependency across 6 × 6 range in layer 0. By this manner, more convolutions can capture much farther spatiotemporal dependency..

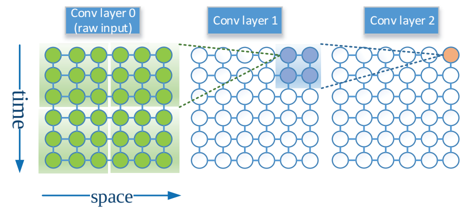


Figure 5 Stacking of conv layers. A neuron in conv layer 0 represents a spatiotemporal region.

3.2.2 Inception block and fractal block

Existing studies found that only by straightforwardly stacking more conv layers together or simply widen the networks make the extended network more prone to overfitting and become dramatically inefficient [21] . To avoid such drawbacks, this study developed the *inception block* and *fractal block*. Through stacking of them, our model is armed with the deep network structure to improve the crime prediction performance as well as keeping the networks efficient to train.

Inception block. The crime risks are usually affected by various spatiotemporal patterns among crime-related incidents. For example, the crime risk of a particular region may has not only the positive correlations to the historical crime intensities of some surrounding areas, but also has the negative associations to other nearby areas. What's more, it is may further affect by its temporal patterns such as seasons, and the impacts of 311 incident evolution patterns. In order to extract such complicated features, We designed this block, which stacks asymmetric convolutions in terms of the sequences of 1×3, 3×1 and 1×3 conv layers [21] , as shown in Figure 4 (b). At the bottom of the block, using element-wised methods, the merge layer (green rectangle) is leveraged to collect multiple spatiotemporal dependencies information together from different branches into a "thick" feature map.

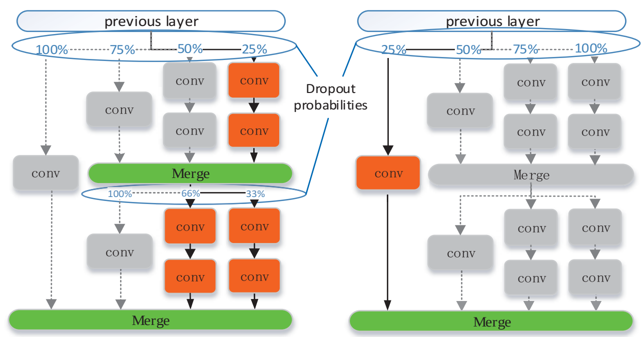
Fractal Block. The design of the fractal block was adopted from FRACTALNET [22] with the core idea of fractal stacking architecture and multiple drop-paths, as shown in Figure 4 (c). However, there are three major differences in the proposed fractal block with FRACTALNET.

(1) Conv layers were factorized into 3×1 and 1×3 asymmetric ones, maintaining the expression ability with fewer parementers.

(2) The merge layer applied concated function which overlays the outputs of different branches to capture multiple crime-related features, instead of using the element-wised average function.

(3) The *deep drop-path* and *shallow drop-path* were formulated to build the complex substructure and simple substructure, respectively. Fractal blocks selected either of them by the auto-parameter searching method [24] . This way help to handle the diversities of relation complexities in the data, thereby improving the prediction performance.

In the *Deep drop-path* or *shallow drop-path*, the layer with multiple branches drops each of them with decreasing (Figure 6 (a)) or increasing (Figure 6 (b)) probabilities from left to right, but making sure at least one survives. The decreasing or increasing method is applied to obtain the dropout probability of each branch. For example, Figure 6 (a) depicted the *Deep drop-path* applied in fractal block with seven floors of layers. Compared to the *shallow drop-path* in Figure 6 (b), there is merely three floors of layers. Since the "previous layer" of the block owns four branches, the dropout probability set is $\boldsymbol{\rho_d}=\{\frac{4}{4},\frac{3}{4},\frac{2}{4},\frac{1}{4}\}$ from left to right for *deep drop-path* in Figure 6 (a), or $\boldsymbol{\rho_s}=\{\frac{1}{4},\frac{2}{4},\frac{3}{4},\frac{4}{4}\}$ for *shallow drop-path* in Figure 6 (b).



(a) Deep Drop-path (b) Shallow Drop-path
Figure 6 Drop-paths. Gray color means disabled.

4 Experiments

4.1 Baselines

STCN was evaluated by comparing with the other four models, which are conventional time series prediction approach and classifiers employed on spatiotemporal predictions published recently.

Rolling weight average (RWA): The crime risk is determined by the average value of previous crime events along a time window (e.g., the last 60 days) with the values of more recent days being given the bigger weights.

$$y_{t+1}=\frac{ny_t+(n-1)y_{t-1}+\cdots+2y_{t-n+2}+y_{t-n+1}}{n+(n-1)+\cdots+2+1} \quad (1)$$

where n is the size of the time window, y_t is the number of crime incidents at day t. A threshold $\varepsilon = 0.1$ is used as a cut-off to separate 1 or 0 classes.

Support Vector Machines (SVM) [5] : in this model, we used the Gaussian (RBF) kernel, and other two parameters c and γ, which were set 3 and 0.16 respectively to the maximum its F1.

Random Forests [5] [6] [8] : In our implementation, the number of trees was 200, and the maximum depth of each tree was 20.

Shallow fully connected neural networks (SFCNN**)** [23] **:** An artificial neural network composed of 3 full connecting hidden layers, each of them contains 64 neurons. The model was built to compare with deep learning structure that whether it owns the ability to capture the complex spatiotemporal dependencies among crime and 311 incidents.

4.2 Evaluation metrics

This paper used F1 and AUC (for ROC) as the performance metrics for our imbalance dataset since the two metrics are not biased towards the minority class [26] .

In the training stage of each model, we repeated ten cross-validations, in which the whole dataset was split into 90% randomly to training and 10% for validation using stratified sampling technology. In the testing stage, the validity of each model is assessed by the metrics mentioned above.

4.3 Experimental Setup

The experiment used TensorFlow1.01 and CUDA8.0 to build STCN and SFCNN, which were executed on a desktop computer with Intel i7 3.4GHz CPU, 32GB memory and NVIDIA GeForce GTX650 GPU (2GB RAM). The learnable parameters are initialized using a Gaussian distribution with the mean and standard deviation (0, 0.02). These neuron networks models were trained on the full training data for 100 epochs with the mini-batch size of 128. Besides, all of the conv layers use the ReLU activation function with the batch normalization technology. Then the model was trained by gradient descent optimization method and Adam [26] with a weight decay of 0.0001 and a momentum of 0.9, to minimize cross-entropy error. Other baseline models were developed through Scikit package. The over-sampling processing for imbalance dataset drawn support from imbalanced-learn package [26] .

4.4 Performance Comparison

4.4.1 Performance studies on 311 data

The experiment has been carried out to evaluate the effectiveness of 311 data to the crime prediction. The size of the spatial neighbor set is fixed to 3×3 with the length of time window as 60 days. In Table 2, we found the F1 and AUC of the proposed STCN were highest without the 311 data. Further, after accommodating the 311 data, the F1 and AUC increased by 3% and 6% respectively, expressing the most improvement of our model. This fact demonstrated not only the 311 incidents have certain spatiotemporal correlations with crimes, but also the better capability of our model on mining such relationships.

Table 2 Performance with and without 311 data

Model	F1 (no 311)	F1	AUC (no 311)	AUC
SVM	**0.70**	0.68	**0.76**	0.74
RandomForest	0.81	0.81	0.82	0.82
SFCNN	0.80	**0.81**	0.79	**0.82**
STCN	*0.85*	**0.88**	*0.86*	**0.92**

4.4.2 Performance studies on time windows

It was evident from Figure 7 and Figure 8 that the proposed STCN approach outperformed the other baselines with increasing time windows. There were more than 6% and 10% improvements on F1 and AUC when compared to the SFCNN, which was the best classifiers among baselines. The other models, SVM and RWA, gained suboptimal performances. The curves in the two figures also demonstrate that the prediction performances of all models were improved as the time window increasing, due to more temporal patterns or trends can be discovered. The F1 and AUC of SVM began to decline when the length of time window beyond 40, while the performances of our STCN continued to grow until the length of time window surpassed 60. This situation implies that the STCN benefited from its capability to capture more valid spatiotemporal features hidden in crime and 311 incidents. However, when the length of the time window beyond 60, all of the models produced more errors, implying that a further remote historical data probably has less correlation with current time.

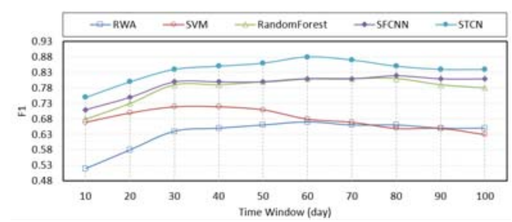


Figure 7 F1 on time window

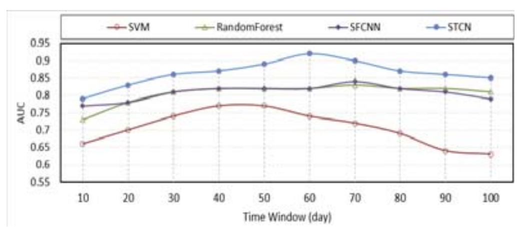


Figure 8 AUC on time window

It was proved that our proposed model had the best generalization performance since both F1 and AUC of the proposed STCN were still higher than other baselines even the length of time window reached 100.

4.5 Results Visualization

The predicted result acts as an important indicator for identifying high-risk locations that require further attention and police place-based intervention. As an example, Figure 9 gave the comparison of predicted result and the ground truth of crimes in NYC on 2nd Nov 2015. By observing the spatial distribution of red grids in Figure 9 we can know where crimes would happen in the future. In addition, by the spatial adjacent analysis on Figure 9, we found 86% yellow grids were located within 2 km of corresponding nearest red grids, with the left 10 ones out of 2 km. Therefore, just by expanding the police resources (e.g., patrol cars) from the predicted red grids to extra 2 km, we could further remove opportunities for committing crimes from 76% (71/(225+71)) to 97% ((225+71-10)/(225+71)) of the high-risk locations. That could be used to answer the question “how would crime be prevented?”.

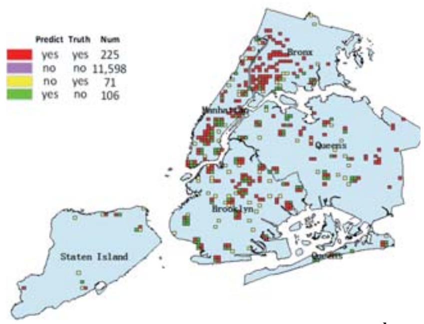


Figure 9 Comparison of crime distribution on 2nd Nov 2015

5 Conclusions

In this paper, we proposed a novel deep CNNs-based model for crime predicting. By comparing with 4 baseline methods based on the crime and 311 data, the proposed model demonstrated its superiority in spatiotemporal crime risks forecasting task. However, CNN is a “black-box”, in which the neuron connection structures are not readily interpretable [25] . Whereas, rather than waiting for the day when the “black box” is open, this study is eager to take a quick pace to apply deep CNNs on the spatiotemporal crimes analysis for preventing crimes now. We believe this work just touches the surface of what is possible in this direction and there are many avenues for further exploration such as modeling different categories of crimes risks, identifying significantly socioeconomic features to prevent crimes etc. In addition, more data types should be taken into account to increase prediction performances in the future.

Acknowledgements

This project has primarily been funded by: National Natural Science Foundation of China, “Research on House-breaking Suspects Location Spatiotemporal Prediction Base on Semantic Correlation” (#41401524); Guangxi Natural Science Foundation, “Research on Crime Events Spatial Correlation Evolution and its Location Prediction Based on Spatiotemporal Near-Repeating Network Pattern Analysis” (#2015GXNSFBA139191); Open Research Program of Key Laboratory of Police Geographic Information Technology, Ministry of Public Security, “House-breaking Suspects Location Spatiotemporal Prediction Base on Semantic Correlation” (#2016LPGIT03); Scientific Project of Guangxi Education Department, “Research on Crime Spatiotemporal hotspot Prediction Based on Spatiotemporal Near-Repeating Network Analysis” (#KY2015YB189); Open Research Program of Key Laboratory of Environment Change and Resources Use in Beibu Gulf (Guangxi Teachers Education University), Ministry of Education, “Research on Intelligent Activity-route Planning Based on Multi-source Mobiling Trajectory Pattern Mining” (#2014BGERLXT14); Open Research Program of Key Laboratory of Mine Spatial Information Technologies of National Administration of Surveying, Mapping and Geoinformation, “Intelligent Activity-route Planning Research Based on Multi-source Mobiling Trajectory Pattern Mining in ‘Intelligent City’” (#KLM201409).

References

[1] Kelvin Leong, Anna Sung. “A review of spatio-temporal pattern analysis approaches on crime analysis”. International e-Journal of Criminal Science, vol. 9, 1—33, 2015.

[2] Xiaofeng Wang, Donald E. Brown, Matthew S. Gerber, “Spatio-Temporal Modeling of Criminal Incidents Using Geographic, Demographic, and Twitter-derived Information”. IEEE International Conference on Intelligence and Security Informatics (ISI), 11-14, June 2012.

[3] Spencer Chainey. “The Crime Prediction Framework- a spatial temporal framework for targeting patrols, crime prevention and strategic policy”. http://proceedings.esri.com/library/userconf/nss15/papers/nss_03.pdf, 2015.

[4] Mohsen Kalantari, Bamshad Yaghmaei, Somaye Ghezelbash (). Spatio-temporal analysis of crime by developing a method to detect critical distances for the Knox test, International Journal of Geographical Information Science, vol. 30, 11, 2302-2320, 2016.

[5] Addarsh Chandrasekar, Abhilash Sunder Raj, Poorna Kumar. “Crime Prediction and Classification in San Francisco City”. http://cs229.stanford.edu/proj2015/228_report.pdf, 2015.

[6] Kadar Cristina, Iria José, Pletikosa Cvijikj, Irena, “Exploring Foursquare-derived features for crime prediction in New York City”. The 5th International Workshop on Urban Computing, San Francisco, California USA, 2016.

[7] M. S. Gerber. “Predicting crime using Twitter and kernel density estimation. Decision Support Systems”. Vol.61, 1, 115–125, 2014

[8] Alex Chohlas-Wood, Aliya Merali, Warren Reed, Theodoros Damoulas. “Mining 911 Calls in New York City- Temporal Patterns, Detection and Forecasting”. AAAI Workshop: AI for Cities, 2015.

[9] Monsuru Adepeju, Gabriel Rosser, Tao Cheng. “Novel Evaluation Metrics for Sparse Spatio-temporal Point Process Hotspot Predictions: A Crime Case Study”, International Journal of Geographical Information Science, vol. 30, 11, 2133-2154, 2016

[10] William Wells, Ling Wu, Xinyue Ye. “Patterns of Near-Repeat Gun Assaults in Houston”. Journal of Research in Crime and Delinquency, vol. 49, 2, 49-186, 2012.

[11] Joel M. Caplan, Leslie W. Kennedy, Joel Miller, “Risk Terrain Modeling: Brokering Criminological Theory and GIS Methods for Crime Forecasting”. Justice Quarterly, Vol. 28, 2, 360-381, April 2011.

[12] Jane Law, Matthew Quick and Ping Chan, “Bayesian Spatio-Temporal Modeling for Analysing Local Patterns of Crime Over Time at Small Area-level”. Journal of Quantitative Criminology, vol. 30, 1, 57–78, 2014.

[13] Hongjian Wang, Zhenhui Li, Daniel Kifer, Corina Graif, “Crime Rate Inference with Big Data”. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, 635-644, August 2016.

[14] K. Simonyan, A. Zisserman. “Very deep convolutional networks for largescale visual recognition”. https://arxiv.org/abs/1409.1556, 2015.

[15] Wang, P., Xu, J., Xu, B., Liu, C., Zhang, H., Wang, F., & Hao, H. . “Semantic Clustering and Convolutional Neural Network for Short Text Categorization”. Proceedings ACL, 352–357, Beijing China, July 2015.

[16] Neal Jean, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, Stefano Ermon, “Combining satellite imagery and machine learning to predict poverty”. Science, vol. 353, 790-794, 2016.

[17] Ma X., Yu H., Wang Y., Wang Y.. “Large-Scale Transportation Network Congestion Evolution Prediction Using Deep Learning Theory”. PLoS ONE, vol. 10, 3, 1-17, 2015.

[18] Li X., Peng L., Hu Y., Shao J., Chi T.. “Deep learning architecture for air quality predictions”. Environ Sci Pollut Res Int, vol. 23, 22, 22408-22417, 2016.

[19] Wilpen L. Gorr, YongJei Lee. “Early Warning System for Temporary Crime Hot Spots”. J Quant Criminol, vol. 31, 25–47, 2015

[20] Y. Movshovitz-Attias, Q. Yu, M. C. Stumpe, V. Shet, S. Arnoud, L. Yatziv. “Ontological Supervision for Fine Grained Classification of Street View Storefronts”. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1693–1702, Boston MA USA， June 2015.

[21] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens and Zbigniew Wojna, “Rethinking the Inception Architecture for Computer Vision”. https://arxiv.org/abs/1512.00567, 2015

[22] Gustav Larsson, Michael Maire, Gregory Shakhnarovich. “FractalNet: Ultra-Deep Neaural Networks without Residuals”. https://arxiv.org/abs/1605.07648, 2017.

[23] Junbo Zhang, Yu Zheng, Dekang Qi. “Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction”. https://arxiv.org/abs/1610.00081, 2017.

[24] Bolme, D. S., Beveridge, J. R., Draper, B. A., Phillips, P. J., Lui, Y. M.. “Automatically Searching for Optimal Parameter Settings Using A Genetic Algorithm”. In Lecture Notes in Computer Science, 6962, 213–222, 2011.

[25] Ravid Shwartz-Ziv, Naftali Tishby. “Opening the Black Box of Deep Neural Networks via Information”. arXiv:1703.00810, 2017.

[26] Kingma, D., Ba, J.. “Adam: A method for stochastic optimization”. arXiv:1412.6980, 2017.

[27] Guillaume Lematre, Fernando Nogueira, Christos K. Aridas. “Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning”. Journal of Machine Learning Research, vol. 18, 17, 1-5, 2017.