Crime prediction using self-exciting point processes and image features as covariates

Anonymous Author(s)
Affiliation
Address
email

Abstract

State-of-the-art crime prediction models exploit the spatio-temporal clustering patterns and the self-exciting nature of criminality to predict vulnerable crime areas. However, omitting covariates correlated with the crime occurrence potentially bias the estimated parameters. This research combines self-exciting point processes and environmental covariates extracted through convolutional networks from street-level images to predict hotspots across Bogota, using a methodology recently proposed by [6]. Our model using image features as covariates outperforms a standard self-exciting point process.

1 Introduction

State-of-the-art crime prediction models exploit the spatio-temporal clustering patterns and the self-exciting nature of criminality to predict vulnerable crime areas. For instance, the model proposed in [5] is constructed under three assumptions: criminality concentrates in specific areas, there is higher incidence of crime at certain times and days of the week, and crime spreads from one place to another like seismic activity. Moreover, this model has been successfully implemented in numerous cities in the world for predictive policing, including Bogotá [2]. In this work, we include street-level images as covariates to predict criminal hotspots with the aim of explicitly take into account the effect of the urban environment on crime occurrence, using the methodology recently proposed by [6].

The motivation behind the use of covariates in a self-exciting point process model is that ignoring the spatial characteristics of a place potentially bias the estimated parameters that capture the spreading effect of crime. In the economic literature this is documented as endogeneity by omitted variables and was well studied for the particular case of crime prediction models in [6]. Moreover, combining spatial covariates with the self-exciting nature in a (semi-)parametric model allows us to study the effect of environmental features on crime occurrence, shedding light on hotspots dynamics' and giving insights to the design of public policies.

2 Methodological approximation

Previous studies have used images and transfer learning to predict socioeconomic characteristics such as consumption, expenditure and asset wealth. In [4], a convolutional neural network is trained to extract image features that can explain up to 75% of the variation in local-level economic outcomes. In particular, [3] tests the ability of environmental variables, through the use of street-level images, to predict criminal hotspots in Chicago. Employing transfer learning techniques, they extract features from the street-level images using the pre-trained deep neural network AlexNet, followed by another DNN where the extracted features and other spatial covariates are combined to predict crime occurrence.

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.

With this in mind, and assuming crime as a self-exciting point process, we use the methodology proposed by [6] extending [5], and include street-level images that capture spatial features as covariates:

$$\lambda(s,t) = \exp(\beta X_{C(s)}) + \sum_{i:\, t_i<t} g(s - s_i, t - t_i). \quad (1)$$

In this setting, crimes may occur as background events given their spatial characteristics, $\exp(\beta X_{C(s)})$, or as aftershock events triggered by past crimes nearby, $g(s_j - s_i, t_j - t_i)$. Variables X correspond to street-level images gathered by [1] over the locality of Chapinero in Bogotá, and $C(s)$ refers to the index of the nearest image to point s.

The images dataset was obtained by building a city street-level image crawler using the Google Street View API V3.0. The dataset is composed by 5,505 images after filtering the data set by their SIFT local descriptors. Spatial covariates are obtained exactly as in [1] using a VGG19 for feature extraction. A VGG19 is an image classification pre-trained convolutional neural network with 19 deep layers developed by [7]. In [1] this network's fully connected top layer was removed and loaded with the weights trained on ImageNet, hence obtaining a 512 valued vector.

3 Implementation and preliminary results

We use geo-referenced and time-stamped crimes occurring in Chapinero locality in Bogotá during March and April, 2018, provided by the District Security Office of the city. To train the proposed model we adapted the Expectation-Maximization setting used in [6] and compare its predictive power against the standard self-exciting point process model [5]. Finally, the models were tested on unseen observations and we study the Hit Rate achieved by each of them when varying the percentage of cells flagged as hotspots. Results of the test setting are shown in Figure 1.

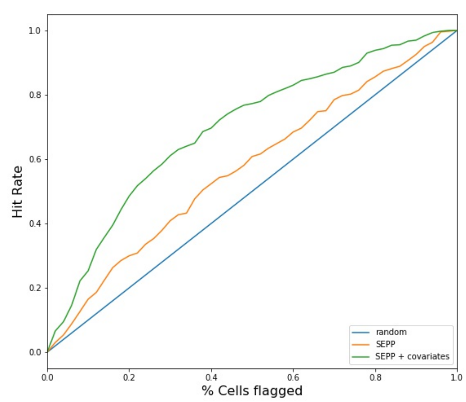

Figure 1: Hit Rate vs. % Cells flagged as hotspots

Our model using image features as covariates to capture environmental variables outperforms a standard self-exciting point process. Specifically, the AUC of the proposed model is 12 points higher that the one of the self-exciting point process. For instance, our model captures correctly 25% of the crimes in the test dataset with 10% of the cells flagged as hotspots, against 16% captured by the SEPP model.

Finally, to test the robustness and sensibility of our model, we are currently comparing its performance, and the effect of introducing the street level image features as static covariates, against a multivariate model that explicitly incorporate additional variables to the crimes itself, including unstructured data. These models have previously been studied in [8] by introducing demographic variables and others that account for past crime behaviour in a delimited area.

References

[1] Acosta, S. A. and Camargo, J. E. Predicting City Cafety Perception Based on Visual Image Content. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, 177–185, 2019.

[2] Dulce, M., Ramírez-Amaya, S. and Riascos, Á. Efficient allocation of law enforcement resources using predictive police patrolling. *arXiv preprint arXiv:1811.12880*. 2018.

[3] Kang, HW. and Kang, HB. Prediction of crime occurrence from multimodal data using deep learning. *PloS one, 12(4), e0176244*. 2017.

[4] Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B. and Ermon, S. Combining satellite imagery and machine learning to predict poverty. *American Association for the Advancement of Science*, volume 353, 790–794, 2016. DOI: 10.1126/science.aaf7894

[5] Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P. and Tita, G. E. Self-exciting point process modeling of crime. *Journal of the American Statistical Association Volume 106, Issue 493*, 2011.

[6] Reinhart, A. and Greenhouse, J. Self-exciting Point Processes with Spatial Covariates: Modeling the Dynamics of Crime. *arXiv preprint arXiv:1708.03579v2*, 2019.

[7] Simonyan, K. and Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[8] Wang, X., Brown, D. and Gerber, M. (2012). Spatio-Temporal Modeling of Criminal Incidents Using Geographic, Demographic, and Twitter-derived Information. *IEEE International Conference on Intelligence and Security Informatics* (pp. 36-41). IEEE.