STAT 157 Predictive Policing Paper

Ellen Kulinsky SID 26126845

Jong Ha Lee SID 25344865

Evan Limanto SID 25557086

Vaibhav Ramamoorthy SID 25539060

Arun Ramamurthy SID 26440691

December 15, 2017

Abstract

Since its inception in 2011 from the LAPD, predictive policing – the usage of mathematical and analytical techniques in law enforcement to identify potential criminal activity – has proliferated in police departments across the United States. However, concerns have been raised regarding the efficacy, lack of transparency, and systemic bias in said systems. In this paper, we specifically address such concerns present in PredPol, a vanguard spearheading predictive policing. We evaluate PredPol's underlying assumptions through literature, and empirically build a simple yet intuitive model to compare the predictive and capture accuracy of crimes against that of PredPol. We find that our model performs comparably to PredPol's complex model in correctly predicting crime, demonstrating the disproportionately expansive and complicated nature of predictive policing models compared to their performances. Through this paper, we hope to motivate police departments and crime analysts of the steep monetary implications and negative externalities that may result from blindly trusting predictive policing systems without considering the latent model bias behind them.

1 Introduction

Since its inception in 2011 from the LAPD, predictive policing – the usage of mathematical and analytical techniques in law enforcement to identify potential criminal activity – has proliferated throughout police departments in states such as California, Washington, South Carolina, Arizona, Tennessee, and Illinois. Due to its data-driven and so-claimed statistically and mathematically sophisticated complexities, predictive policing systems have been very attractive for police departments to implement without fully understanding its implications, limitations, and processes. It has become a convenient system for police departments to blindly trust.

PredPol is one of such systems police departments extensively utilize. Using cloud software technology that identifies high risk areas in near real-time, PredPol allows police officers to anticipate crime in advance. According to published literature, PredPol utilizes the Epidemic Type Aftershock Sequences (ETAS) model for seismicity, which originated from efforts to predict earthquakes and its aftershocks in geology. ETAS models are based on self-exciting point processes, which are event patterns where ”parent” events produce “offspring” aftershock events. In the context of policing, these offsprings refer to crimes that rise due to preceding crime such as revenge gang activity, burgled houses with smashed in windows, etc. This process used in PredPol has been previously described in literature [4], but the algorithm itself has not been released to the public.

While its goals are well-founded and its model seems sophisticated, we were concerned with the strong assumptions of its ETAS model – primarily the assumption that crimes follow a self-propagating pattern that spreads via a contagion-like process – and whether it truly does achieve better predictive accuracy. Furthermore, we were unsure of whether the model circumvents the inherent bias historical crime data possess. In the next sections, we look at existing literature to evaluate the underlying assumptions behind PredPol's model, and the possible negative consequences that may arise due to blindly trusting predictive policing systems. Then, we construct a simple and intuitive model to test against the efficacy PredPol's predictive policing algorithm. To do so, we first define the capture rate of a model as the fraction of crime caught over the maximum crimes attainable on the given day. We evaluate the capture rate results, limitations of our model, and further research required to reduce systemic input bias and improve prediction accuracy.

2 Literature Review

Our research is concerned with validating the effectiveness of the ETAS model PredPol uses to inform law enforcement agencies. Previous studies have each deconstructed areas of bias and the low performances of ETAS models as applied to crime or earthquake data.

Before even taking the model into account, we first note that police data is known to suffer from systemic bias. This can be attributed to factors such as community trust in the police force or the fact that, as empirical evidence shows, “police officers – either implicitly or explicitly – consider race and ethnicity in their determination of which persons to detain and search and which neighbourhoods to patrol” [2]. Furthermore, Lum notes that, “the presence of bias in the initial training data can be further compounded as police departments use biased predictions to make tactical policing decisions” [2]. Because of the danger of a feedback loop which enforces preconceived bias, police departments are advised to work towards eliminating systemic bias in their datasets necessarily prior to selecting a predictive policing model so as to produce impartial results. In her evaluation of the outcomes of predictive policing models, Lum implements the published algorithm used in PredPol’s software[4], and runs this model on Oakland crime data, comparing the results with the 2011 National Survey on Drug Use and Health (NSDUH). Lum finds that PredPol’s model, as would any predictive policing model that does not explicitly account for biased data, “results in increasingly disproportionate policing of historically over-policed communities” [2]. This finding motivates our investigation into the use of “black-box machinery, such as PredPol’s software, that purports to be scientific, evidence-based and race-neutral” [2] as compared to simpler procedures that better reveal the presence of systemic bias in the data. Our analysis of PredPol’s model uses Lum’s implementation verbatim and the mentioned Oakland crime dataset.

ETAS models, which include complexities such as an EM algorithm for parameter estimation and assumption of a Poisson distribution for background events, themselves have been shown to be comparable in predictive power to much simpler methods. Luen evaluates the effectiveness of the ETAS model for earthquake prediction by deconstructing the areas of model bias. The ETAS model makes two main assumptions: 1) that earthquakes can be split into parent and offspring events through seismicity, and that 2) the arrival of parent earthquakes is Poisson-distributed. Since "there may or may not be physical differences between 'background' and 'offspring' events, all existing methods to differentiate between them based on times, locations, and magnitudes are arbitrary" [1]. Making this assumption can even cause predictive power to slacken, as "it is difficult to tell if [notable features] are of physical interest or are artifacts of the declustering procedure" [1]. Finally, key properties of the ETAS model are "very sensitive to distributional assumptions and parameter values" [1]. In the context of police data, which already suffers from systemic bias that can affect parameter values, a highly sensitive model like ETAS can skew even further from reality, further reinforcing biases from the inputted data. Luen bolsters his analysis of the ETAS model by comparing it with a simple automatic alarm model. Luen finds, "automatic alarm strategies have almost all of the predictive success of the ETAS model, while avoiding many of that model's drawbacks" [1]. This finding motivates a deeper investigation into the predictive power of simple models against PredPol's ETAS model, as implemented by Lum. If the ETAS model as applied to predictive policing suffers from the same drawbacks noted by Luen for earthquake data, in conjunction with the real-world effects of racially biased feedback loops, further research and policy improvements are required to rapidly implement improved models for predictive policing that systematically fix problems such as racial bias and invalid distributional assumptions.

In the initial developments of the PredPol algorithm, Mohler motivates the use of ETAS, a model paradigm traditionally applied to seismology, by noting that, "criminological research has shown that crime can spread through local environments via a contagion-like process" [3] resulting in clusters similar to those seen in earthquake data. Mohler explains the motivation for application of an ETAS model to gang and drug-related crimes in particular, but PredPol applies this predictive algorithm to other crime types for which there is no longer intuition for reliance on a self-exciting process. Furthermore, Luen previously had shown little to no improvement in predictive accuracy gained by utilizing ETAS models to predict earthquakes, its original intention, hence providing even further questioning into the appropriateness of an ETAS-based model to predict crimes. In the next section, we empirically calculate and compare the accuracy of ETAS-based PredPol model against our simple and rather rudimentary model, to evaluate and answer our skepticism.

3 Description of Data

To compute the capture rate of our model, we consider two datasets released by the Oakland Police Department to OpenOakland, providing open access data containing information about crimes that occurred in Oakland from 1 January 2009 to 31 December 2011. The reported crimes have been pre-segmented into geographical "bins" (150 m by 150 m blocks) based on their location of occurrence.

The first dataset, `drug_crimes_with_bins.csv` contains rows denoting individual occurrences of a crime on a certain date and geographical bin.

The second dataset, `oakland_grid_data.rds` contains the geographical coverage of each bin as a multipolygon. We pre-processed these geometry objects to obtain a dictionary mapping bins to their neighboring bins.

4 Methodology

To determine the optimal placement of police officers to geographic bins in Oakland on a given day, we utilized two features which we believed were most indicative of predicting drug crimes. Intuitively, our first analysis was based on historical data – bins with a greater number of recent past crimes were more likely to see future crime. Another factor we utilize is drug crime data from neighboring geographic bins, as there may be spillover effects that induce more crimes to occur due to the proximity to those neighboring bins. We combine the two sources of information to create a crime "score" for each geographic bin on a given day. These scores can then be used to see the areas which possess the highest likelihood of a drug crime occurring on a given day. We outline our methodology in generating the score below.

4.1 Historical Data

For each bin on a given date, we look at crime data up until one year before the given date. This constitutes our historical dataset for that given area. Then, we utilize a simple exponential decay function to determine weightage given to each historical crime data point. The details of the *historical crime score* function are shown below:

$$H_{i,\tau} = \sum_{t=1}^{365} W_t \times C_{i,\tau-t} \tag{1}$$

where W_t = weight given to crime data that occurred in bin i at time τ - t as shown below

$$W_t = e^{-rt} \tag{2}$$

and

$$\begin{aligned} C_{i,\tau-t} &= \text{Number of crimes that occurred in bin } i \text{ at time } \tau \text{ - } t \\ \tau &= \text{Today's date} \\ t &= \text{A time lag} \\ r &= \text{The exponential decay factor} \end{aligned}$$

In words, we weigh crime data less the further we go back in time from the given date we wish to predict crime occurrence for. If drug crimes occurred almost one year ago, that would have much less importance in our model than crimes which occurred only just a week ago. We also note that we use historical data ranging a year back maximum from the given date.

4.2 Neighbor Data

Next, we incorporate crime data from neighboring bins by calculating their own historical crime score. We define neighboring bins by all the adjacent and diagonal geographic bins as defined by the geographic segmentation of Oakland provided by Kristian Lum's Oakland grid dataset. Then we sum their historical crime scores to generate a *neighbor crime score*. The detailed equation form is shown below:

$$N_{i,\tau} = \sum_{n \in N_i} H_{n,\tau} \tag{3}$$

where

$$\begin{aligned} i &= \text{Bin number} \\ N_i &= \text{List of neighbors of bin } i \\ \tau &= \text{Today's date} \end{aligned}$$

4.3 Model

Combining the two equations formulated above, we arrive at the following model incorporating both information from historical and neighbor crime data:

$$score_{i,\tau} = (1-s) \times \sum_{t=1}^{365} e^{-rt} C_{i,\tau-t} + s \times \sum_{n \in N_i} \sum_{t=1}^{365} e^{-rt} C_{n,\tau-t} \tag{4}$$

where s = a *neighbor coefficient*, $s \in [0,1]$, that weights the sum of the *historical crime score* and *neighbor crime score*.

4.4 Parameter Tuning

Our model requires tuning for two parameters: r - the exponential decay rate, and s - neighbor coefficient. We choose optimal values of such parameters based on a parameter grid search in an effort to choose the best (r,s) combination which maximizes the proportion of total possible crimes captured by the model. In doing so, we run a parameter grid search, between $r \in [0, 0.2]$ in increments of 0.02, and $s \in [0, 0.5]$ in increments of 0.05. We chose such ranges of values for these parameters after initial testing with large numbers, and decreasing the size of increment until the model was sensitive enough to show discrepancies in percent of total possible crime captured.

Due to current computational issues, we ran a parameter grid search on 6 different samples of 50 dates each, averaging the capture rate across the 6 samples for each combination of parameter values. This allowed us to choose the parameters that were robust across multiple different sample dates, allowing less possibilities for overfitting issues.

5 Results

5.1 Parameter Tuning

Results from the parameter grid search are shown in Table 1. We include the top 10 parameter combinations by mean capture rate across the 6 samples. We also include PredPol's capture rates across the same 6 samples, in Table 2.

Surprisingly, we observe that the parameter combination which maximizes the average capture rate is $r = 0$ and $s = 0$, meaning no exponential decay weight as time lag increases, and no weight on neighbor crime data. In words, the model which gives the same importance to crime that happened a year ago and a day ago, and does not consider crimes occurring in neighborhood bins, captures the largest proportion of total crimes occurring on a given day. Such a result was definitely unexpected, but provides strong evidence against some of the clustering assumptions made by the ETAS model. Most critically, our simple approach performs comparably to PredPol's complex

r	s	sample1	sample2	sample3	sample4	sample5	sample6	mean_capture_rate
0.00	0.00	0.32057	0.25761	0.31989	0.27329	0.34510	0.29881	0.30254
0.00	0.05	0.31190	0.26082	0.30188	0.27408	0.32960	0.31174	0.29834
0.00	0.30	0.31095	0.25461	0.31759	0.25430	0.31922	0.32138	0.29634
0.00	0.20	0.30900	0.24309	0.32141	0.26148	0.32287	0.32001	0.29631
0.00	0.25	0.31500	0.24087	0.32141	0.25577	0.32065	0.32401	0.29628
0.00	0.35	0.31295	0.25128	0.31455	0.25096	0.31366	0.31816	0.29359
0.00	0.40	0.30628	0.25059	0.30955	0.25096	0.30297	0.31246	0.28880
0.02	0.35	0.29145	0.26159	0.27450	0.27232	0.31333	0.30934	0.28709
0.02	0.15	0.32449	0.23046	0.28300	0.26396	0.31549	0.30412	0.28692
0.02	0.40	0.28745	0.25687	0.27632	0.27051	0.31333	0.30934	0.28564

Table 1: Parameter Grid Search Results

sample1	sample2	sample3	sample4	sample5	sample6	mean_capture_rate
0.31972	0.26452	0.34697	0.28019	0.31485	0.28727	0.30226

Table 2: PredPol Accuracy across 6 Samples

model.

5.2 ROC Curve Comparisons

We also demonstrate the comparable performance of our model against PredPol's model through the ROC Curve below. We plot the % of crimes missed[1] against varying number of police officers deployed into unique geographic bins. We assume that we deploy one police officer per bin, as the bin areas are only roughly 150 by 150 meters, and that a deployment in a bin captures all crime in that bin. We compare our final chosen model with parameters $r = 0$ and $s = 0$ from the grid search, with PredPol's model, results of which are shown in Figure 5.2.

We see that in almost all cases, our simple model misses less crime – and hence captures more crime – than PredPol's model. Specifically, we see our model outperform PredPol by roughly 2 percentage points when 14 to 18, or 25 or more police officers are deployed to geographic bins. Our model performs comparably with PredPol at the lower tail and at 20 police officers deployed. The relative success of our model suggests that our model may be a better choice for both small and large police departments with varying number of police officers.

[1] 1 - % of crimes captured

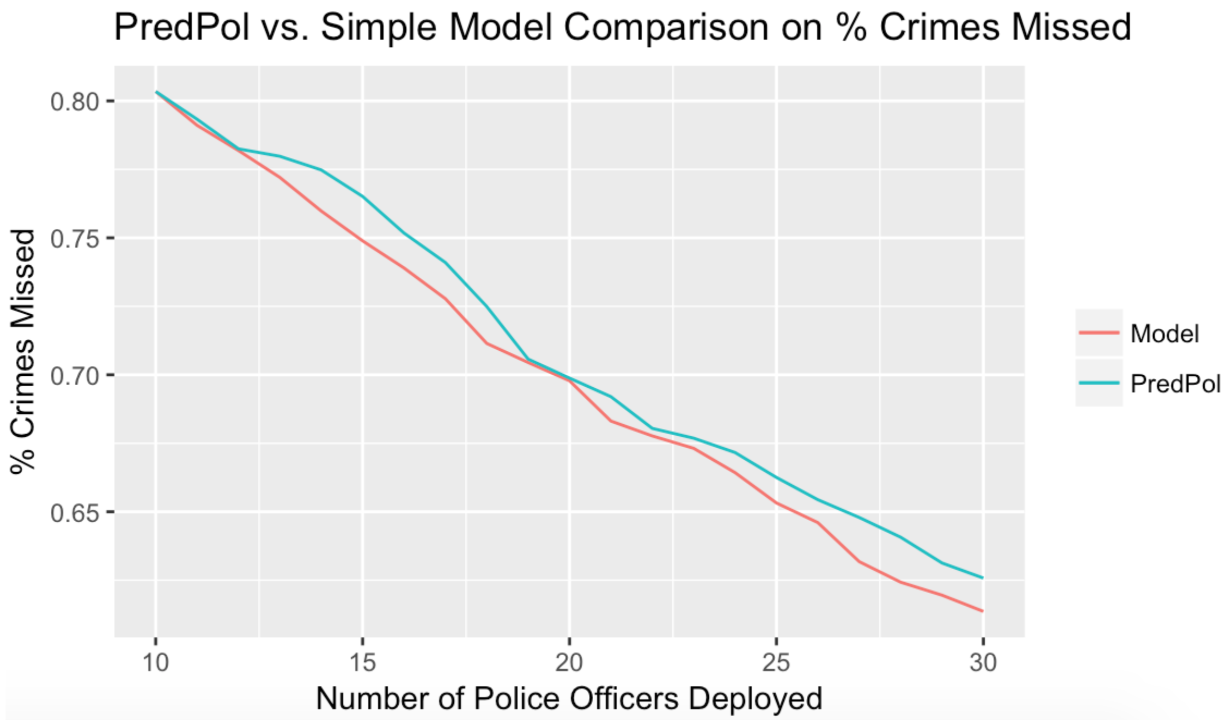

Figure 1: ROC Curve Comparing Our Model vs. PredPol

6 Discussion

In a similar vein to Luen's evaluation of the ETAS model [1], our analyses show that our simplistic model performed comparably to PredPol's complex ETAS models in proportion of true crimes captured on a given day, of roughly 30 percent. Despite the sophistication of PredPol's model in utilizng self-exciting processes, a Poisson-distributed background crime rate, and an EM procedure for parameter estimation, it performs almost equally, if not sometimes worse, than our barebones model based on linear weights and an exponential decay.

In the face of the exorbitant prices that PredPol charges police departments for use of its predictive model, and the public's vested interest in effective allocation of tax funds and effective policing strategies, our results indicate that the utility of PredPol's model is overstated and misused. Furthermore, when the systemic social consequences of using such a predictive policing model — highlighted in Lum's paper — are taken into account, PredPol's software might actually be harming communities in the United States and furthering police bias, while instituting a cost to police departments. While we don't claim that our model resolves the biases incident in PredPol that Lum discusses, we demonstrate that a free, simple solution performs as well or better and can save police departments significant tax payer dollars.

7 Limitations

One drawback of our methodology is that we compared our predictive policing model with Lum's recreation of PredPol rather than against the software itself. While Lum's implementation strictly follows the methodology outlayed by PredPol [4], we would like to test our model against the real software in future experiments. Our model was also trained on a narrow training set, as we used data on only one crime type (drug crimes) from one city's dataset – Oakland's open access crime history. Furthermore, the data breakdown used in both our algorithm and Lum's PredPol comparator do not account for per capita crime rates. Since bins with higher population density are naturally more likely to have more crime, both models are biased towards selecting bins with greater population, since they might naturally experience more crime as a byproduct of including more people. In order to produce a model not influenced by population density, we would have to reblock the city into bins by population rather than geographic area covered.

8 Conclusion

Given the simplicity of our model, which could be implemented very easily with basic open-source tools, we hope that our results would motivate police departments to consider exploring their own solutions for predictive policing, which corrects for bias in crime data, or to investigate models with better predictive power. Our model, in a much simpler vein, shows the same bias-inducing feedback loops that target specific racial or ethnic communities as PredPol, yet produces better results at no cost. The bias present in both models give police departments sufficient reason to investigate further the bias inherent in existing predictive policing solutions and explore corrective measures. In the meanwhile, police departments can implement cheaper alternatives and expect similar results.

9 Acknowledgements

First, we thank Professor Philip Stark for providing us with resources and ideas on predictive policing, which made this research project possible and interesting for all of us. We would also like to thank Kristian Lum for inspiring us to pursue this project with her in-class talk and for her helpful and frequent correspondence and collaboration throughout. Contributions made by each team member are briefly summarized below:

Jong Ha Lee helped generate results and ensure long, overnight processes were running smoothly when optimizing for parameters through grid search. After optimizing for parameters, he helped generate the ROC curve comparing PredPol and the paper's model in crime capture rate across varying number of police officers deployed. In writing the paper, he was primarily in charge of writing the "Methodology" and "Results" section, and overall initialized the setup of shareLaTeX document.

Vaibhav Ramamoorthy developed the code structure for our model and wrote functions to calculate bin scores and overall capture rate. He also wrote the script for the grid search, helping to generate the heatmap visualization. In writing the paper, he contributed to much of the "Methodology" section and formalized the model into mathematical notation for greater clarity. Vaibhav also handled group meeting time scheduling and some of the correspondence with Kristian Lum.

Arun Ramamurthy wrote the script to select neighboring bins given the Oakland grid data, but mostly focused on reviewing the current literary discourse surrounding the use of predictive policing models, ETAS as applied to crime data, and the limitations of ETAS models for earthquake prediction. In writing the paper, he was primarily in charge of writing the "Introduction", "Literature Review", "Discussion", and "Conclusion" sections. Finally, he initialized the setup of the Rpres document for the in-class presentation of the group's research.

Ellen Kulinsky performed a deep analysis of Kristian Lum's code to ensure that it correctly

implemented the algorithm outlined in Mohler's papers, describing PredPol's model. She was also critical in performing detailed literature review of all the papers we cited and in corresponding with Kristian Lum for clarifications. In writing the paper, Ellen contributed to the ”Introduction”, ”Literature Review”, ”Limitations”, and ”Conclusion” sections.

Evan Limanto wrote much of the code to calculate the capture rates of Kristian's model and to collect data on a larger array of days for testing. He also wrote unit tests for a sizable amount of the code used to run our experiments and model. In writing the paper, he contributed most to the ”Description of Data”, ”Abstract”, and ”Introduction” sections.

10 References

[1] Bradley Luen. “Earthquake prediction: Simple methods for complex phenomena”. In: (2010).

[2] Kristian Lum and William Isaac. “To Predict and Serve?” In: 13 (2016), pp. 14–19. DOI: 10.1111/j.1740-9713.2016.00960.x.

[3] G. O. Mohler et al. “Self-Exciting Point Process Modeling of Crime”. In: 106 (493 2011), pp. 100–108. DOI: 10.1198/jasa.2011.ap09546.

[4] G.O. Mohler et al. “Randomized Controlled Field Trials of Predictive Policing”. In: 110 (512 2015), pp. 1399–1411. DOI: 10.1080/01621459.2015.1077710.