

Does Machine Learning Reduce Racial Disparities in Policing?

By Jyotishka Datta and Grant Drawve,
University of Arkansas, Fayetteville

Since the mid-1990s, there have been strategic reforms across the United States to improve criminal justice policies to control the unchecked growth of mass incarceration and reduce the racial disparities in policing. Yet reports of systematic biases in law enforcement e.g. police brutality against specific race-groups more than the others, continue to trouble us. In a 2016 study, Ashley Nellis[1] documented these racial disparities: the rate of incarceration of African Americans across state prisons is 5.1 times more than that of whites, crossing 10:1 in five states. A key driver of such inequality, as Kristian Lum points out in a recent *Nature* paper, is demographically disparate treatment by the courts [2], which in turn is caused by "implicit" racial biases that affect even the most well-intended judges' discretion and can act independently of their explicit beliefs about racial inequality [3]. To counteract this, law enforcement agencies are relying more and more on machine learning tools to build an objective and fair criminal justice system, as we see the rise of "predictive policing". Perry et al. (2013) [4] define predictive policing as "the application of analytical techniques - particularly quantitative techniques - to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions."

Predictive policing is becoming more common in law enforcement agencies, similar to how hot spot techniques spread across agencies. Kelly (2015) [5] indicates an issue law enforcement agencies face is the ability to define a "high-crime area", especially with the U.S. Supreme Court's decision in *Illinois v. Wardlow* (528 U.S. 119, 124 (2000)) allowing for reasonable suspicion to be a plus-factor for neighborhoods classified as high-crime. "The courts' failure to require law enforcement agencies to present concrete evidence demonstrating that a neighborhood has a heightened propensity for crime raises significant constitutional concerns" (pp. 304, Kelly, 2015) [5]. In particular, safe-guarding Fourth Amendment protections against unreasonable searches and seizures without probable cause and general reasonable suspicion. As Ferguson (2012) [6] discusses in relation to predictive policing, "...this predictive information will be used to justify stops under existing Fourth Amendment precedent."

But what is a high-crime area? The proverbial "officer gut-instinct" does not hold as much merit with statistical techniques capable of operationalizing "high-crime areas", however are predictive policing algorithms unbiased?

"The way to stop discriminating on the basis of race is to stop discriminating on the basis of race."

-Chief Justice of the U.S. Supreme Court John Roberts in 2007.

Perry et al. (2013) compare conventional methods with full-scale predictive analysis techniques, e.g. whereas the former would look for “hot spots” from a small amount historical crime data, the latter promises to harness the power of “big data” and sophisticated mathematical modeling, yielding risk terrains using regression, classification and clustering. The underlying assumption is that these models, based on large amounts of data rather than human judgment, will bring fairness and objectivity to decision making. As a statistician, one would naturally wonder, do these machine learning tools, powered with “big data”, really help reduce the inequalities? Any graduate student of statistics would tell you, it depends on the data that the models are fed. The rand.org guidebook by Perry et al. (2013) warns its users, “Predictive policing has been so hyped that the reality cannot live up to the hyperbole.” Such a “crystal ball” cannot exist as “predictions are only as good as the underlying data used to make them.” Lum and Isaac (2016) [7] warn of the negative consequences, “if biased data is used to train these predictive models, the models will reproduce and in some cases, amplify those same biases. At best, this renders the predictive models ineffective. At worst, it results in discriminatory policing.”

Lum and Isaac (2016) further investigate the impact of predictive policing by investigating the “hot spots” yielded by PredPol [8-11], one of the biggest vendors of predictive policing software, that applies a sliding window approach to forecast crime using only data on type of crime and time and places of past crimes. Their case study shows that PredPol reinforces the apparent biases in existing police records, disproportionately targeting communities of color and low-income. While there is no “ground truth” to act as a reference frame, Lum and Isaac (2016) combine a demographically representative synthetic population data with National Survey of Drug Use and Health (NSDUH) data that produces a map of drug crimes more evenly distributed than police records, where some areas are significantly over-represented. The authors conclude, “This creates a feedback loop where the model becomes increasingly confident that the locations most likely to experience further criminal activity are exactly the locations they had previously believed to be high in crime: selection bias meets confirmation bias.”

“Even though it's not me that's the yellow guy, your

officers are going to treat whoever comes out of that house in his boxer shorts as the yellow guy,” Jouvonal, 2016 [12].

A hidden danger is the “tautological obscurity” that leads to shifting accountability from human decision makers to machines that are treated as black-boxes. While sophisticated algorithms like PredPol claim to produce a forecasting system that is race-neutral, its mathematical underpinnings are beyond the reach of its users.

For example, PredPol uses a method inspired by seismology, and posits that like aftershocks following an earthquake, probability of reoccurrence of the same event in a similar place would increase after the first occurrence of an event [11]. Hunchlab [13] claims to sharpen PredPol's method by adding risk terrain modeling (RTM) [14-15] based on classifying geographical landmarks as crime attractors or generators. The statistical methodology for the risk terrain model [14] would appear familiar to any graduate student of statistics: use penalized regression to select only a few of 192 variables, where most of the coefficients are forced to be zero to favor sparsity. Indeed, the utility uses an elastic net regularized regression with Poisson distributed events, with further model simplification using a bidirectional stepwise regression, using Bayesian Information Criterion (BIC).

A statistician might also tell you that an “elastic net” is unstable in high dimensions, especially when predictors are correlated (Zou, 2006) [16], and question the “bet on sparsity” principle, i.e. is the truth necessarily sparse? Model selection methods inherently act like “Occam's razor” or favor parsimonious models. Philosophically pleasing as they may be, their influence on crime forecasting must be questioned: a simpler model of crime is not necessarily the best model for crime.

More important is perhaps the underlying technological barrier: the agency deploying this model would not challenge the hidden assumptions, even if they understand the operational aspects.

The second category of predictive policing is the more disturbing, albeit not as widely adopted, strategy of predicting offenders from their digital footprints. Chicago's “heat-list” compiles names of individuals likely to be involved in major crimes (Papachristos, 2009) [18]. Another method, “Beware” [17], claims to

“Even though it's not me that's the yellow guy, your officers are going to treat whoever comes out of that house in his boxer shorts as the yellow guy,”

-Jouvonal, 2016 [12]

use data from social media to calculate individuals' threat scores although, "neither are people made aware of the score that is assigned to them, nor does the police department have any insight into how the score is calculated [18]." The future of predictive policing looks disturbingly close to the movie *Minority Report* (<http://www.imdb.com/title/tt0181689/>).

How accurate are these models in real life prediction? On a 400x400 square feet area, hot spot analyses with RTM provides an accuracy of 25%, increasing to 68% in an 800x800 square foot cell [16]. A statistician must interpret these accuracies in the light of the heavily unbalanced lengths of positive and negative examples and issues of overfitting and sparsity. Although the true positives far outweigh the false positives [19] negative effects of the wrongful assumptions are not scarce: ranging from false positives in the heat-list leading to false accusations and financial harm to assigning threat scores based solely on the address of a house.

What can we do as statisticians? Most importantly, we can educate the community at large about the potential negative consequences of naively applying deep learning methods, not just in the context to

arrest records to predict hotspots, but also caution that algorithms can be biased when they ignore the socio-technical context. One key step is reducing the obfuscation of machine learning techniques: knowing the assumptions behind the methods and their inductive biases can help the agencies critically evaluate the algorithms deployed. Equally important is taking a proactive role in this rapidly evolving process that affects our society, contribute in building accurate and interpretable methods for crime forecasting keeping the human cost in mind.

Teach our future generations of statisticians, "let the data speak for themselves," but more importantly, teach them, "garbage in, garbage out"!

Envoi: India is not far behind: a tool called CMAPS (Crime Mapping, Analytics and Predictive System) is being developed by Delhi Police in collaboration with Indian Space Research Organization to forecast crime activities [20]. Although CMAPS is reported to design predictive policing algorithms based on stored records of criminal data, it stands out from their offshore counterparts by using "space technologies" (details of the method were not available at the time of writing this article).

References:

[1] Nellis, Ashley. The color of justice: Racial and ethnic disparity in state prisons. The Sentencing Project (2016). (<http://www.sentencingproject.org/publications/color-of-justice-racial-and-ethnic-disparity-in-state-prisons/>)

[2] Lum, Kristian. "Limitations of mitigating judicial bias with machine learning." *Nature Human Behaviour* 1.7 (2017): s41562-017.

[3] Clemons, John Tyler. "Blind injustice: The Supreme Court, implicit racial bias, and the racial disparity in the criminal justice system." *Am. Crim. L. Rev.* 51 (2014): 689.

[4] Perry, Walt L. Predictive policing: The role of crime forecasting in law enforcement operations. Rand Corporation (2013).

[5] Koss, Kelly K. "Leveraging predictive policing algorithms to restore fourth amendment protections in high-crime areas in a post-Wardlow world." *Chi.-Kent L. Rev.* 90 (2015): 301.

[6] Ferguson, Andrew G. (2012). Predictive Policing and Reasonable Suspicion. *Emory Law Journal*, 259-325.

[7] Lum, Kristian, and William Isaac. "To predict and serve?" *Significance* 13.5 (2016): 14-19.

[8] Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. *J. Amer. Statist. Assoc.*, 106(493), 100-108.

[9] Mohler, G. (2013). Modeling and estimation of multi-source clustering in crime and security data. *Ann. of Appl. Stat.*, 7(3), 1525-1539.

[10] Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2015). Randomized controlled field trials of predictive policing. *J. Amer. Statist. Assoc.*, 110(512), 1399-1411.

[11] <http://predpol.com>

[12] Jouvenal J (2016) The new way police are surveilling you: calculating your threat 'score'. *The Washington Post*, January 10. https://www.washingtonpost.com/local/public-safety/the-new-way-police-are-surveilling-you-calculating-your-threat-score/2016/01/10/e42bccac-8e15-11e5-baf4-bdf37355da0c_story.html

[13] <https://www.hunchlab.com>

[14] Caplan, J. M., Kennedy, L. W., & Miller, J. (2011). Risk terrain modeling: brokering criminological theory and GIS methods for crime forecasting. *Justice Quarterly*, 28(2), 360-381.

[15] Caplan, J. M., Kennedy, L. W., Barnum, J. D., & Piza, E. L. (2015). Risk terrain modeling for spatial risk assessment. *Cityscape*, 17(1), 7. Papachristos, A. V. (2009). Murder by structure: Dominance relations and the social structure of gang homicide. *Amer. J. of Soc.*, 115(1), 74-128.

[16] Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* 101 1418-1429. MR2279469.

[17] <https://www.west.com/safety-services/public-safety/powerdata/>.

[18] Degeling, M., & Berendt, B. (2017). What is wrong about Robocops as consultants? A technology-centric critique of predictive policing. *AI & SOCIETY*, 1-10.

[19] Drawve, G. (2016). A metric comparison of predictive hot spot techniques and RTM. *Justice Quarterly*, 33(3), 369-397.

[20] <https://timesofindia.indiatimes.com/city/delhi/Predictive-policing-to-nail-criminal-minds/articleshow/52824404.cms>