Runaway Feedback Loops in Predictive Policing

Danielle Ensign
University of Utah

Sorelle A. Friedler
Haverford College

Scott Neville
University of Utah

Carlos Scheidegger
University of Arizona

Suresh Venkatasubramanian
University of Utah

ABSTRACT

Predictive policing systems are increasingly used to determine how to allocate police across a city in order to best prevent crime. Observed crime data (arrest counts) are used to update the model, and the process is repeated. Such systems have been shown susceptible to runaway feedback loops, where police are repeatedly sent back to the same neighborhoods regardless of the true crime rate. In response, we develop a model of predictive policing that shows why this feedback loop occurs, show empirically that this model exhibits such problems, and demonstrate how to change the inputs to a predictive policing system (in a black-box manner) so the runaway feedback loop does not occur, allowing the true crime rate to be learned.

CCS CONCEPTS

• **Theory of computation** → **Reinforcement learning**; • **Computing methodologies** → *Machine learning algorithms*;

KEYWORDS

Feedback loops, predictive policing, urn models

ACM Reference format:
Danielle Ensign, Sorelle A. Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. 2017. Runaway Feedback Loops in Predictive Policing. *Presented as a talk at FAT/ML 2017, Halifax, Nova Scotia, Canada, August 2017*, 5 pages.

1 INTRODUCTION

Predictive policing is increasingly employed to determine where to send police, who to target for surveillance, and even who may be a future crime victim [12]. We focus on the most popular of these forms of predictive policing (with PredPol, HunchLab, IBM, and other companies entering the market) which attempts to determine how to deploy police given historical crime data.

Definition 1.1 (Predictive Policing). Given historical crime data for a collection of regions, decide how to allocate patrol officers to areas to detect crime.

Once police are deployed based on these predictions, arrest data from the neighborhood is then used to further update the model. Since arrests only occur in neighborhoods that police have been sent to *by the predictive policing algorithm itself*, there is the potential for this sampling bias to be compounded, causing a runaway feedback loop. Indeed, Lum and Isaac [7] have recently shown that this can happen.

Lum and Isaac's work focused on PredPol [8], a predictive policing system in use by the LAPD and other cities across the U.S. Lum and Isaac [7] model what would happen if PredPol were used in Oakland to distribute police to find drug crime by using police department arrests[1] as the historical data and a synthetic population of likely drug users based on public health data [10, 9]; they find that increasing policing efforts based on arrest rates causes PredPol's prediction to substantially diverge from the true crime rate, repeatedly sending back police to the same neighborhoods.

In addition to its importance in the criminal justice pipeline, predictive policing serves as an archetypal problem, through which we can better understand issues which arise out of deploying batch-mode machine learning algorithms in an online setting, where they essentially see results that are influenced by their own predictions. Other such algorithms include recidivism prediction, hiring algorithms, college admissions, and distribution of loans. In all of these contexts, the outcome of the prediction (e.g., who to hire) determines what feedback the algorithm receives (e.g., who performs well on the job).

1.1 Results

We use the theory of urns (a common framework in reinforcement learning) to analyze existing methods for predictive policing. We show formally as well as empirically why these methods will not work. Subsequently, we provide remedies that can be used directly with these methods in a black-box fashion that improve their behavior, and provide theoretical justification for these remedies.

2 RELATED WORK

Our work builds most strongly on the work of Lum and Isaac [7] described above. In addition, it relates to the narrower question of defining notions of fairness in *sequential* learning settings such as the setting of bandits (regular, contextual and linear) and Markov decision processes[6, 5, 3, 4]. There, the primary goal is to understand how to define fairness in such a

This research is funded in parts by the NSF under grants IIS-1633387, IIS-1633724 and IIS-1513651.

FAT/ML 2017, August 2017, Halifax, Nova Scotia, Canada

[1] Although the Lum and Isaac data (that we will also use for simulation) consists of crime reports (that include both arrest data and reported crimes that did not result in arrests), for clarity of the source of the feedback loop, we'll refer to the data as "arrest data" throughout.

process, and how ensuring fairness might affect the ability to learn an accurate model.

Uneven policing can be seen as a probability distribution over crime samples in a survey of crime. With this perspective, the solution we propose in Section 3.4 is a rejection-sampling variant of the Horvitz-Thompson estimator [2].

3 PREDICTIVE POLICING WITH URNS

We start our investigation with an overview of PREDPOL. PREDPOL [8] assumes that crimes follow an earthquake aftershock model, so that regions that previously experienced crime are likely to experience crime again, with some decay. Mohler et al. [8] model the crime rate $\lambda_r(t)$ in region r at time t as follows: $\lambda_r(t) = \mu_r + {}_{t_r^i < t}\, \theta\omega e^{-\omega(t - t_r^i)}$ where t_n^i represents the time of an event in region r, ω quantifies the time decay of a shock, and θ captures the degree to which aftershocks are generated from an initial event. They use an expectation-maximization procedure to determine the parameters of the model.

Note that this model only uses event (arrest) observations per region to determine the crime rate[2] and does not use any context in the form of demographics, arrest profiles and so on. PREDPOL, in essence, is predicting where arrests will happen (since that's all it sees), not where *crime* will happen. Each day officers are sent to the areas with highest predicted intensity and the resulting arrest data is fed back into the system.

3.1 Urn Models

We will model the predictive policing process by a series of urn models with increasing complexity. Urn models (especially the Pólya-Eggenberger urns) have a long history in machine learning, but notably also in reinforcement learning [11], where they have been used, starting with the work of Erev and Roth [1], as a way to model how bounded-rationality players in a game might interact with each other. Studying the dynamics of urn models allows us to understand the convergent behavior of reinforcement learning in such settings.

We will use a *generalized* Pólya urn model [11] containing balls of two colors (red and black). At each time step, one ball is drawn from it, the color is noted, and the ball is replaced. Then the following replacement matrix is used to decide how to update the urn contents:

	Red replacements	Black replacements
Red ball sampled	a	b
Black ball sampled	c	d

This matrix says that if we initially sampled a red ball, then we replace it and add a more red balls and b more black balls to the urn. We refer to the *standard Pólya urn* as a generalized urn with $a = d = 1$ and $b = c = 0$.

3.2 A simple predictive policing model

In the simplest predictive policing setting, a precinct has a single police officer and polices two regions A and B. Every day the police officer is sent to one neighborhood where they may or may not observe crime; if crime is observed, it is reported. The goal is to build a predictive model for where to send the officer on each day. Specifically, the goal is to distribute the police officers in proportion to the crime in each area.[3]

GOAL 3.1 (EFFECTIVE POLICING). *A region with Λ percent of the crime in the precinct should receive Λ percent of the police.*

Achieving this goal requires learning the relative crime rates of the regions.

To understand the behavior of predictive models, we will make some simplifying assumptions. We will firstly assume that the predictive model only uses current statistics (in some form) to make predictions.

ASSUMPTION 3.1 (PREDICTIVE MODEL). *The officer tosses a coin based on current statistics to decide where to go next.*

To fully specify a predictive model, we also need to understand *context* – what information is collected during policing – and *ground truth* – what assumptions we make about underlying crime rates. In the subsections below, we explore what happens as we vary these factors. To start, we make the following assumptions.

ASSUMPTION 3.2 (CONTEXT). *The only information retained about a crime is a count.*

ASSUMPTION 3.3 (CRIME OBSERVATION). *If an officer goes to a location with an underlying ground truth crime rate of λ, the officer observes crime at a rate of λ.*

Note that the Crime Observation assumption allows the predictive policing system to operate in a generous context. There are many reasons to believe that this assumption does not hold. We will show that even in this optimistic setting problems occur.

3.2.1 Uniform crime rates. Let us start by assuming that the crime rate is uniform between areas.

ASSUMPTION 3.4 (CRIME RATE). *If an officer goes to a location, crime happens with probability λ.*

Consider an urn that contains red and black balls, where the proportion of red and black balls represent the current observed statistics of crime in areas A and B respectively. Visiting area A corresponds to picking a red ball and visiting area B corresponds to picking a black ball. Observing crime (which happens with probability λ) causes a new ball of the same color to be placed in the urn. The initial balls are always returned to the urn. The long-term distribution of red and black balls in the urn corresponds to the long-term belief about crime prevalence in the two areas.

In general, we can describe the evolution of this process as the following urn. We toss a coin that returns 1 with probability λ. If the coin returns 1, we simulate one time step of

[2] PREDPOL — critically — conflates amount of crime and number of arrests.

[3] Why should this be the goal? Suppose there are exactly enough police officers to stop all the crime and no more, then a deployment according to the true crime rates will perfectly police all regions.

a standard Pólya urn, and if 0, we merely replace the sampled ball. This corresponds to a standard Pólya urn "slowed" down by a factor λ. As such, its long-term convergence is well-characterized.

LEMMA 3.1 ([13]). *Assume the urn starts off with n_r red balls and n_b black balls. Then the limiting probability of seeing a red ball is a draw from the beta distribution $\beta(n_r, n_b)$.*

Significance. The long-term probability of seeing red is the long-term estimate of crime in area A. The above result shows that this probability is a random draw governed only by the parameters n_r, n_b, which represents the prior *belief* of the system. In other words, the prior belief coupled with the lack of feedback about the unobserved region *prevents the system from learning that the two areas are in fact identical with respect to crime rates.*

On the contrary, consider how this process would work *without* feedback. The officer could be sent to an area chosen uniformly at random each day, and this process would clearly converge to a uniform crime rate for each area. Indeed, such a process resembles the standard update for the bias of a coin where the prior distribution on the bias is governed by a β distribution.

3.2.2 Nonuniform crime rates. Let us now drop the assumption of uniformity in crime rates, replacing Assumption 3.4 by

ASSUMPTION 3.5. *A visit to area A has probability λ_A of encountering a crime, and a visit to area B has probability λ_B of encountering a crime.*

The resulting Pólya urn can be represented by the stochastic addition matrix $\begin{pmatrix} \lambda_A & 0 \\ 0 & \lambda_B \end{pmatrix}$. By limiting ourselves to the subsequence of events when some ball is added to the urn, and using a general lemma characterizing the asymptotics of *deterministic* urn updates from Renlund [13], we can prove the following lemma about the urn under this new assumption.

LEMMA 3.2. *In the urn with addition matrix given above, the asymptotic probability of sampling a red ball is* 1 *if $\lambda_A > \lambda_B$ and is zero if $\lambda_B > \lambda_A$.*

Significance. In this scenario, the update process will view one region as having much more crime than the other, *even if crime rates are similar*. In particular, if region A has a crime rate of 10% and region B has a crime rate of 11%, the update process will settle on region B with probability 1. This is a classic "go with the winner" problem where feedback causes a runaway effect on the estimated probabilities.

3.2.3 Observational decay. Finally, we reexamine Assumption 3.2. Thus far, our urn models have captured some key elements of the model used by PREDPOL – the idea of differential crime rates as well as the updates based on crime observed. PREDPOL includes a notion of *limited memory*, both by incorporating time decay into crime aftershocks, and by using a limited time window for training. We model limited memory in the urn setting by adding a simple notion of decay. After every round, each ball disappears from the urn independently with a fixed probability p_d. This can be thought of as a relaxation of Assumption 3.2. Varying p_d is analogous to varying the size of the PREDPOL training window.

3.3 Evaluation of the urn model for PREDPOL

The combination of non-uniform crime rates and observational decay models the behaviour of PREDPOL. However, to the best of our knowledge there is no theoretical characterization of the asymptotic distributions in this model. We present empirical evidence illustrating the problems with using this model to learn crime rates. In our experiments, $p_d = 0.01$.

Using the Lum and Isaac [7] data, we consider a two neighborhood police deployment scenario using, first, the two regions of Oakland with the most historical arrests ($Top1$ and $Top2$) and, second, the Oakland neighborhood with the most arrests as compared to a randomly chosen region with many fewer arrests (*Random*). We simulate the effect of the historical arrest data on the prior for the system by determining the number of balls for each region in our urn based on the past number of arrests. We use the full number of arrests (609, 379, and 7 for regions $Top1$, $Top2$ and *Random* respectively) as the starting number of balls in the urn from each region. The urns are then updated based on the estimated number of daily drug use incidents, i.e., $\lambda_{Top1} = 3.69$, $\lambda_{Top2} = 2.82$, and $\lambda_{Random} = 2.36$.

The results, shown in the left of Figure 1, demonstrate that even if police sent to a neighborhood arrest people according to the true crime rate, the urn model will converge to *only* sending police to the neighborhood with the most crime. This replicates (within our urn model) the feedback loop problems with PREDPOL found by Lum and Isaac [7]. Recall, from Lemma 3.2, that skew occurs even if the difference in crime rates between the two neighborhoods is small. Note that while we included a notion of decay in our urn model in order to more closely model PREDPOL, we found similar results under the urn model without decay.

3.4 Modifying the urn model to account for feedback

In order to learn the crime rate, we want the Pólya urn to contain balls in proportion to the relative probability of crime occurrence. As we have seen, a standard Pólya urn with stochastic update rates will converge to a distribution that has no relation to the true crime rates. Here, we present a simple change to the urn process which *does* guarantee that the urn proportion will converge to the ratio of replacement (i.e crime) rates.

Consider the standard Pólya urn update rule: the probabilities λ_A and λ_B model the probability of an additional ball being added to the urn, *conditional* on a ball of the respective color having been sampled. This means that the probability of a ball being added is not λ_A, but $\lambda_A n_A^{(t)}(n_A^{(t)} + n_B^{(t)})^{-1}$. As a result, the expected ratio of balls being added to the urn after one step of the process is not $\lambda_A(\lambda_A + \lambda_B)^{-1}$.

This immediately suggests a fix: instead of always adding the new balls, we *first sample another ball from the urn, and only add the new balls if the colors are different*. With this fix,

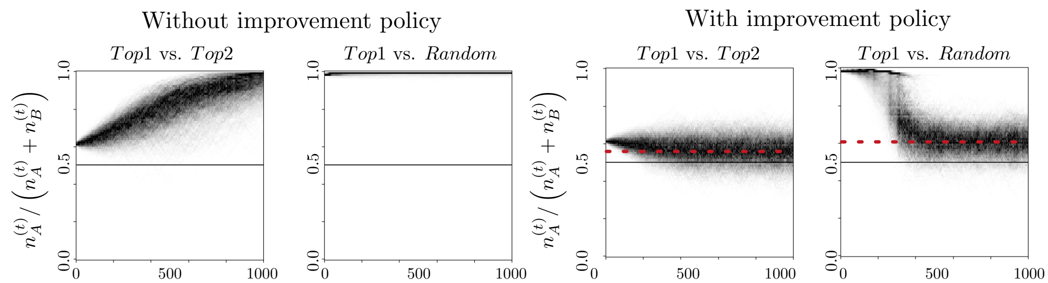

Figure 1: The distribution over 1000 days versus percentage of balls from region $Top1$ **in the urn over 1000 runs. A police force deployed based on the underlying crime rates would send** 56.7% **of the force to** $Top1$ **instead of** $Top2$ **and** 61.0% **of the force to** $Top1$ **instead of** *Random*. **Left: both charts converge to sending** 100% **of the force to** $Top1$. **Right (improvement policy): the charts appear to converge to the correct crime rates.**

the probability of adding a ball with label 1 is $n_A^{(t)}(n_A^{(t)} + n_B^{(t)})^{-1}\lambda_A n_B^{(t)}(n_A^{(t)} + n_B^{(t)})^{-1}$, while the probability of adding a ball with label B is $n_B^{(t)}(n_A^{(t)} + n_B^{(t)})^{-1}\lambda_B n_A^{(t)}(n_A^{(t)} + n_B^{(t)})^{-1}$. Crucially, these two expressions are proportional to λ_A and λ_B, except for a constant factor that is a function of the current state of the urn.

The intuition behind this fix is that if our decision procedure sends police to region A 90% of the time, we should not be surprised that arrests in region A happen at a rate of nine to one, even if the crime rate is the same across both regions. Interpreting our fix as a form of rejection sampling, then the importance-sampling variant (in this nine-to-one example adding a ball with weight 0.1) is precisely the weighting scheme of the Thompson-Horvitz estimator, used in survey designs with unequal probability distributions [2].

3.4.1 Evaluating the modified urn. Using this improvement policy to determine when to replace balls, we can now determine if the urns can learn the true crime rate despite the issue of observational bias. Again, using the estimated daily drug usage per region as the underlying true crime rate and the historical arrest data as the prior for the urn color distribution, we simulate the urn's ability to find the relative crime rates in two regions, the $Top1$ and $Top2$ arrest regions and a *Random* region. As shown in the right of Figure 1, urns under this improvement policy converge to a distribution of colors that represents the true crime rate.

4 FIXING PREDPOL

The urn models we explore provide a justification for the observed feedback loop failures of PredPol. But can we remedy PredPol itself using the improvements described in Section 3.4? We first demonstrate how asymmetric feedback affects PredPol by simulating the decisions a precinct might take after running it. We run PredPol's prediction model (using the Lum and Isaac [7] data and implementation), trained on Oakland historical crime data, and generate crime according to the drug usage rates described above.

At each simulation day, PredPol trains on the previous 180 days of arrest data, and produces predicted crime rates r_A and r_B. The decision of where to send police is made probabilistically, by a Bernoulli trial with $p = r_A(r_A + r_B)^{-1}$. This models the targeting effect of sending more police where more crime is expected, simulating a typical use of PredPol [8].

To counteract the effects of the feedback, we can employ the same strategy as in Section 3.4. The key insight is that we need only filter the inputs presented to PredPol rather than trying to modify its internal workings. Specifically, once we obtain crime report data from the system, we conduct another Bernoulli trial with $p = r_O(r_A + r_B)^{-1}$, where r_O is the predicted rate of the district we did *not* police that day, and *only add the arrests to the training set if the trial succeeds*. In other words, the more likely it is that police are sent to a given district, the less likely it is that we should incorporate those arrests.

4.1 Evaluating the PredPol simulation and its repair

Simulating the effects of PredPol on policing as described above, both before and after our improvement policy is applied, we compare the policing rates of region $Top1$ to $Top2$ and $Top1$ to *Random* as before. Each simulation is repeated 300 times and run for one year. As can be seen in Figure 2, regular PredPol rates fluctuate wildly over different runs, and do not converge to the appropriate crime rates (marked with the red dashed line). However, when the inputs to PredPol are changed according to our improvement policy, PredPol's prediction rates appear to fluctuate around the correct crime ratio. Note that the process is still quite noisy, a further indication that PredPol generates crime rate predictions that are still somewhat unreliable.

5 CONCLUSIONS

In this paper we show that urn models can be used to formally model predictive policing as well as indicate remedies for problems with feedback. We demonstrate this both formally and empirically. Our solution also suggests a black-box method to counteract runaway feedback in predictive policing by appropriately filtering the inputs fed to the system. Note however that while we provide a strategy to counter

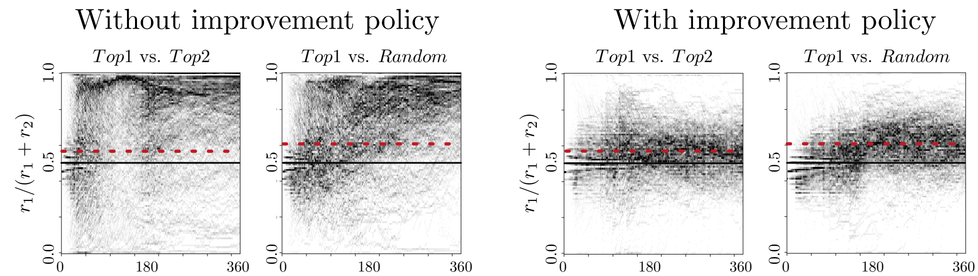

Figure 2: PREDPOL's decisions under targeted policing. Left: PREDPOL operating as usual. Right: arrest observations modified using our improvement policy. Police deployment based on underlying crime rates would send 56.7% **of the force to** *Top*1 **instead of** *Top*2 **and** 61.0% **of the force to** *Top*1 **instead of** *Random*. **These correct crime rates appear to be what PREDPOL converges to under the improvement policy.**

the runaway feedback loop, a full solution to other problems predictive policing systems face would need to additionally address bias in the observation and reporting of crime. We hope future work will continue to carefully study this problem.

6 ACKNOWLEDGEMENTS

This paper would not have been possible without Kristian Lum and William Isaac's generosity in sharing the code and data developed for [7].

REFERENCES

[1] Ido Erev and Alvin E Roth. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American economic review*, pages 848–881, 1998.

[2] Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47 (260):663–685, 1952.

[3] Shahin Jabbari, Matthew Joseph, Michael Kearns, Jamie Morgenstern, and Aaron Roth. Fair learning in markovian environments. *CoRR*, abs/1611.03071, 2016. URL http://arxiv.org/abs/1611.03071.

[4] Matthew Joseph, Michael Kearns, Jamie Morgenstern, Seth Neel, and Aaron Roth. Rawlsian fairness for machine learning. *arXiv preprint arXiv:1610.09559*, 2016.

[5] Matthew Joseph, Michael Kearns, Jamie H Morgenstern, and Aaron Roth. Fairness in learning: Classic and contextual bandits. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 325–333. Curran Associates, Inc., 2016. URL http://papers.nips.cc/paper/6355-fairness-in-learning-classic-and-contextual-bandits.pdf.

[6] Sampath Kannan, Michael Kearns, Jamie Morgenstern, Mallesh Pai, Aaron Roth, Rakesh Vohra, and Z Steven Wu. Fairness incentives for myopic agents. *arXiv preprint arXiv:1705.02321*, 2017.

[7] Kristian Lum and William Isaac. To predict and serve? *Significance*, pages 14 – 18, October 2016.

[8] George O. Mohler, Martin B. Short, Sean Malinowski, Mark Johnson, George E. Tita, Andrea L. Bertozzi, and P. Jeffrey Brantingham. Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association*, 110(512):1399 – 1411, 2015.

[9] United States Department of Health, Human Services. Substance Abuse, Mental Health Services Administration. Center for Behavioral Health Statistics, and Quality. National survey on drug use and health, 2011. Inter-university Consortium for Political and Social Research (ICPSR) [distributor], 2015. URL http://doi.org/10.3886/ICPSR34481.v4.

[10] United States Department of Justice. Office of Justice Programs. Bureau of Justice Statistics. National crime victimization survey, 2014. Inter-university Consortium for Political and Social Research (ICPSR) [distributor], 2015-08-27. URL http://doi.org/10.3886/ICPSR36142.v1.

[11] Robin Pemantle. A survey of random processes with reinforcement. *Probab. Surveys*, 4:1–79, 2007. doi: 10.1214/07-PS094. URL http://dx.doi.org/10.1214/07-PS094.

[12] Walter L. Perry, Brian McInnis, Carter C. Price, Susan Smith, and John S. Hollywood. Predictive policing. RAND Corporation, 2013.

[13] Henrik Renlund. Generalized pólya urns via stochastic approximation. *arXiv preprint arXiv:1002.3716*, 2010.