Machine Learning in Policing and Justice System

By Feng Wang

Machine learning has been one of the most heated topics globally for the past decade, and a wide range of industries are utilizing machine learning to improve efficiency in achieving tasks and understanding targeted consumers to maximize profits. Among all these industries, machine learning is taking a more critical role in the policing and justice system. The US has the highest incarceration rate globally, with nearly 2.2 million adults in prison/jail and 4.5 million in other forms of incarceration facilities [2]. To reduce the incarceration rate and maximize the police force's efficiencies, AI and machine learning algorithms are trained with historical data to achieve that [2].

Machine learning is adopted for policing for various reasons. The main reason is that by showcasing police officers in predicted crime hotspots in the city, the policing efficiency is significantly improved compared with merely strolling in the cities [1,2]. Besides, a facial recognition system can identify frequent offenders, which improved the arrest's quality and significantly [1,2]. However, the algorism for policing is shown to produce biased results causing increased racial profiling, possibly due to the quality of the data and integrity of the uses [1]. There are a couple of reasons which influence the quality of the data—the main reason being that the crime data itself is a Blackbox [7]. Crimes are often under-reported for specific populations, while others are over-reported [7,12]. Police officers are almost always sent to a minority neighborhood, resulting in over-policing in these areas [3]. Like carrying graffiti stencils and spray paint, dark-skinned individuals are much more likely to be arrested for the same minor crime, while others are let loose under the radar with simple verbal warnings [12]. Further, the face recognition system for identifying suspects failed repeated, especially when it comes to dark-skinned individuals [2]. PredPol, being the most widely used software, disclosed that they only use location, time, crime type into consideration when training the algorism [3]. However, with the location and crime type being racially discriminating historically, the data used to train the algorism becomes inherently discriminating towards certain groups. As a result, the algorism is inevitably producing biased results.

While the utilization of machine learning in policing is controversial, the more significant controversy lies in criminal risk assessment algorisms, which is an estimation score for someone's likelihood of reoffending [2]. It is used to determine whether someone is qualified for a pretrial release, and the judge used it to determine defendants' sentences and redhibition services [2]. By theory, this is aimed to reduce prison overpopulation by releasing defendants pre-trail without increasing crime rates and allocate resources efficiently [2]. The criminal risk assessment score also aims to give an unbiased score to assist just ruling from judges [2].

However, the seemingly unbiased criminal risk assessment score is far from neutral. In one of the used systems named Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), black defendants are 77 percent more likely to receive a high criminal risk assessment score regarding violent crime and 45 percent more likely regarding any future crime [3]. Contrary to PredPol, 137 features are used to determine the final score, which produced a more personalized score with enforced stereotypes towards racial, gender, and socioeconomic [3]. To make matters worse, the algorism is proprietary software, resulting in substantial elimination of defendants' abilities to dispute the software's scientific accuracy [3]. The supreme court refused an appeal regarding the risk assessment score, Loomis vs. Wisconsin 2017 [3,7]. The reason for bias here is the same as PredPol in the policing case. The criminal data is like a black box; the data itself is biased due to the history of over-policing and racial profiling; when those partial data are fed into the algorism, the outcome inevitably mirrored the biased past in American history.

In general, such algorithms are designed from a market-oriented standpoint, and the details of the algorithms are usually not understood by people using them [4,6]. These algorisms are considered the trade secret and are protected by the companies under great effort while being marketed to the public as a cutting-edge product utilizing the most recent algorisms [6,9]. On top of the lack of transparency from these companies, it is also unclear how the criminal data are collected, cleaned, and prepared pre training the algorism [7]. However, due to the protection from policies and companies, scientists are unable to examine these algorisms and their training data closely, while the result of such algorisms being used to make substantial decisions.

Several fairness tools are aiming to detect and adjust the bias, and many more in development. Here are several outstanding examples. Pymetrix developed an audit AI tool named Audit AI, which measures and mitigates the potential bias in the training data set [10]. Google developed a What-If Tool that offers a visual interface for exploring model results, which helps understand different features and their effects on outcomes [10]. IBM developed an AI fairness 360 toolkits aiming to test and mitigate biases in the algorism. Other fairness tool includes Teach & Test AI Framework by Accenture, FairML, Lime, SHAP [10,11]. However, it is unclear how those tools are utilized during the algorism designing phase for these companies or used at all. Even when companies are using these fairness tools, it is confusing how to utilize such a tool to reach a fair result. The past data reflected a past in American with conflicts and biases, which still exist today. While society is making progress as a whole, it is impossible to reflect the degree of advances or complete 'clean' the data using the fairness tools. It is proceeding into an ethical grey zone, where the developer of the algorism is given too much power in a territory where no clear answers can be given. The following link contains an interactive article which allow the user to examine the notion of fairness when determine thresholds for such algorisms with visual aids. https://www.technologyreview.com/2019/10/17/75285/ai-fairer-than-judge-criminal-risk-assessment-algorithm/

As Karen Hao stated in MIT Technology Review, there is something inherently concerning regarding such a scoring system: "they turn correlative insights into causal scoring mechanisms." The logic is straightforward; if someone is from a neighborhood with a high crime rate, the system will give the person a higher risk assessment score for that category. While this correlative insight is correct, individuals in the neighborhood are inherently different. Some people in the neighborhood may never even think of committing any crime, while others could frequent the police station. Their risk assessment score should be different, but they will receive the same high-risk scores under the location system category. The risk assessment score is causal in this case because someone lives in a neighborhood with a high crime rate; he/she will have a high-risk assessment score. Such a causal scoring mechanism is dangerous because the biases, which are faced by historically disproportionally targeted populations like low income and minority communities, are being further amplified and justified by the criminal risk assessment score [7]. This idea further resonates with the fundamental ethical concern regarding risk assessment score, which is its purpose [7]. When people see data, mainly if it is the results from the most cutting-edge algorism with fancy names, they forget that the number represents probabilities instead of causal relations [7]. The correlations produced by the algorism can either be real or ridiculous [7]. We need to be very careful about how we interpret those correlations and avoid bias and prejudice in the process because any false positives results can severely interfere with one's fundamental liberty rights and change a person's life path [7].

While it is easy for people to misinterpret such data, the people using them are merely given enough knowledge of the algorism to use these results wisely. The outcome of the score is a correlative insight that could be true or only nonsense. In no case should it be used as a standard to evaluate a person, which overlooks the person's whole background stories and his/her defense. However, in reality, such a score is used to determine whether someone could be released in pretrial. In several cases, being used as a reference when sentence a person [6]. I believe much more strict procedures need to be undertaken before anyone attempts to use the algorism in any way. The algorism itself needs to be carefully examined by exports outside the company regarding reliability, instead of remain secret to the public. Reports need to be produced to educate the people using them and how to use the results unbiased and wisely.

The algorism in policing and the justice system aims to improve policing efficiency and reduce incarceration. However, I believe an algorism is not the solution to the problem. The problem resides in education, health care, and the social safety net, which call for more significant changes. I think attention needs to be moved to society as a whole and aim to reduce the crime rate in neighborhoods by providing better education for children, improved rehabilitation programs, a social network that merit all people, etc. That being said, I do believe that machine learning is a powerful and useful tool. However, society needs to be very careful and more educated when utilizing such a tool and make wise and well-articulated decisions to avoid mistakes with severe consequences.

Work Cited:

[1] Simon, K. (2019, October 22). Pros and Cons of Predictive Policing. Retrieved December 03, 2020, from https://www.profolus.com/topics/pros-and-cons-of-predictive-policing/

[2] Sowards, A. (2020, March 10). Key Pros and Cons of Predictive Policing. Retrieved December 03, 2020, from https://infinigeek.com/key-pros-and-cons-of-predictive-policing/

[3] Hao, K. (2020, April 02). AI is sending people to jail-and getting it wrong. Retrieved December 03, 2020, from https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/

[4] Russell, C. (2019, May 10). Machine Learning in Crime Prevention and Criminal Justice. Retrieved December 03, 2020, from https://medium.com/@camrussell215/machine-learning-in-crime-prevention-and-criminal-justice-47662291eeb3

[6] Farid, H. (2018, October 2). The danger of predictive algorithms in criminal justice | Hany Farid | TEDxAmoskeagMillyard. Retrieved December 3, 2020, from https://www.youtube.com/watch?v=p-82YeUPQh0&feature=emb_title

[7] Završnik, A. (n.d.). Algorithmic justice: Algorithms and big data in criminal justice settings - Aleš Završnik, 2019. Retrieved December 03, 2020, from https://journals.sagepub.com/doi/full/10.1177/1477370819876762

[8] Rosenberg, J (2016) Only humans, not computers, can learn or predict. TechCrunch, 5 May

[9] Automation in the Criminal Justice System: Pros and Cons. (2017, August 14). Retrieved December 03, 2020, from https://www.eldenlawgroup.net/automation-criminal-justice-system-pros-cons/

[10] Pennington, M. (2018, December 30). Five tools for detecting Algorithmic Bias in AI - Technomancers - LegalTech Blog. Retrieved December 03, 2020, from https://www.technomancers.co.uk/2018/10/13/five-tools-for-detecting-algorithmic-bias-in-ai/

[11] Badr, W. (2019, March 13). Evaluating Machine Learning Models Fairness and Bias. Retrieved December 03, 2020, from https://towardsdatascience.com/evaluating-machine-learning-models-fairness-and-bias-4ec82512f7c3

[12] Kleinman, Z. (2018, September 19). IBM launches tool aimed at detecting AI bias. Retrieved December 03, 2020, from https://www.bbc.com/news/technology-45561955