Yuxiang Wan

Prof. Osama Alshaykh

EC601 A2

Dec/28/2021

Machine learning in Justice System and Policing

Police nowadays increasingly apply advance in computer and statistic technique to attempt to aid their work. Predictive policing is becoming a new solution for crime prevention. Police have a great collection of digital data of many different sources, but historically lacked the technological capabilities needed to analyze this data to improve effectiveness and efficiency[1]. With the increasing complexity and volume of the data, applying machine learning tools to improve the decision-making process and service provision within police system is a great way to aid enforcement agencies. By using information technology, data, and analytical techniques to analyze and process the vast data, identifying the places and times of future crime or personnel having high risk of being future crime suspects or victims. Unlike human, machine learning software could be relatively more rational comparing with laborious data processing and decision making. Meanwhile in justice system, especially in country that their judgement is mostly based on previous cases, machine learning or artificial intelligence algorithm could be a possible way to decrease the rate of misjudged cases. To keep the fairness of police and justice system, using machine learning algorithm to validate analyses and determine the legality of their action.

However, to evaluate the performance of the computational tool of the crime forecasting is difficult. Most importantly, we still have little knowledge on how to systematically assess the work of software in data processing, model reality and crime prediction. Since crime prediction is not based on single dominant factor, but plurality ways. Hence, we need to figure out the use terms of the bog data in the context of policing prediction and evaluate the methods, and based on that, we should create the proper way to respond and governing the crimes[2].

Machine learning in predictive policing could be affected by different aspects. When using data in the past to predict events, predictive model is involved. This applies to crime forecasts as much as to any other prediction based on statistics or data. Algorithm

uses mathematical method to calculate the connection between variables and different entities.[2] However, the relationship between crime forecasting and factors is not statistic, while on the contrary, the connection between them is dynamic. Unlike concluding the previous risk of the crimes and then assuming that the crimes may happen only happens in the past, predicting crime is more like relying on current evidence to assess whether it crosses the border and gives the solution to these predictions[3].

While data itself covering biases and unfairness in ML systems often uses a “bias in, bias out” framing, emphasizing the central role of dataset quality. Algorithm decision making can overcome undesirable aspects of human decision making, biases in training data and modal making the data inaccuracy could lead to decision making unfairly among individuals base on biased character, such as race, gender, disability, political orientation, and other sensitive attributes[4]. ML farness is crucial in police and justice system using ML that in reality application should make sure the predicting results and model outputs will not be affected by the attributes which could be considered as unfair. For example, when a person followed by a person get in to a bank, the possibility of robbery should not depend on the races and gender, but the other attributes like acts and manners.

In recent years, researchers have developed a series way of detecting the unfairness in machine learning training dataset. Methods like preprocessing the input training data to remove the undesired biases, minimizing the impact of the sensitive attribute while maximizing the accuracy of predicting result and training a method to force the model have fair output thought the training dataset has some unfavored biases[5].

Fairlearn, an open-source toolkit, could allow researchers and developers to apply it in AI system in case to improve the fairness. The Fairlearn open-source package has two components: 1. Assessment Dashboard: A widget that could assess the effects of different groups on predictions, and it also allow to compare different models by using a kind of standard to reflect fairness and performance; 2. Mitigation Algorithms: A set of algorithms to mitigate unfairness in binary classification and regression[5].

Most methods obtain some models impose approximations on fairness essentials through constraints on lower-order moments or act correspondences to sensitive attribute. In Fairlearn open-source package, fairness is defined as group fairness, which could pick out the individuals that could do harm to the model output. To achieve the function, there is a vector or a matrix called sensitive_features in the open-source package. In assessment process, fairness is scored by variety of metrics, which include the comparation on model's prediction among different dataset. To mitigate unfairness in machine learning models, the Fairlearn open-source package includes a wide-range of unfairness mitigation algorithms. These algorithms could set constraints on prediction process, called parity constraints or criteria, which could require the prediction behavior become comparable among sensitive attributes. Meanwhile the algorithms provide postprocess, taking a black box machine learning estimator and generating a set of retrained models using a series of reweighted training dataset, and reduction unfairness mitigation algorithm, using sensitive features and existed dataset to produce a transformation of the classifier's predictions.

Unfairness in Machine Learning algorithm may present in many handcraft stages, such as the annotation process and manual classify process. Legal requirements often forbid the explicit use of sensitive attributes in the model. Be sure to avoid taking the sensitive attributes into model training process or applying fairness tool like Fairlearn, Aequitas and AI Fairness 360 to reduce the impact.

Citation:

[1] T. I. Cubitt, K. R. Wooden, and K. A. Roberts, "A machine learning analysis of serious misconduct among Australian police," *Crime Science,* vol. 9, no. 1, pp. 1-13, 2020.

[2] J. Hälterlein, "Epistemologies of predictive policing: Mathematical social science, social physics and machine learning," *Big Data & Society,* vol. 8, no. 1, p. 20539517211003118, 2021.

[3] W. L. Perry, *Predictive policing: The role of crime forecasting in law enforcement operations*. Rand Corporation, 2013.

[4] S. Tolan, M. Miron, E. Gómez, and C. Castillo, "Why machine learning may lead to unfairness: Evidence from risk assessment for juvenile justice in catalonia," in *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, 2019, pp. 83-92.

[5] S. Bird *et al.*, "Fairlearn: A toolkit for assessing and improving fairness in AI," *Microsoft, Tech. Rep. MSR-TR-2020-32,* 2020.