Machine learning in Justice System and Policing

Yiqin Zhang
evezhang@bu.edu

With the general applications of big data analytics, machine learning, and artificial intelligence, the public policing and justice system are becoming more and more technologically advanced. Public safety is benefiting from MLmachine learning (ML) extensively. For example, transportation and traffic systems utilized ML to identify violations and AI to implement rules of the road. In addition, crime forecasts integrated with ML provide more efficient distribution of policing resources. ML can also assist in assessing the potential for an individual under criminal justice supervision to re-offend.

1. Applications in Justice System and Policing

1.1. Public Safety Video and Image Analysis

Criminal justice and law enforcement agencies utilized video and image analysis to acquire information for crime investigation. However, video and image analysis on the people, objects, and actions involved in crimes takes an enormous amount of time and requires considerable engagement in personnel and expertise. In addition, traditional software for facial recognition remains restricted to preset eye shape, eye color, and eye distance. In addition, traditional software for facial recognition remains restricted to preset eye shape, eye color, and eye distance. And traditional pattern analysis is limited to demographic information.

Applying ML techniques to video and image analysis enables facial recognition via computer vision and pattern recognition, matching faces, and detecting weapons and other objects. Generally, it includes feature-based and image-based approaches. The feature-based method is more similar to the traditional one to learn the invariant features to detect faces but with more images and complicated external conditions. The image-based method uses ML to learn multiple tasks simultaneously and trains a model to recognize complex facial characteristics. Neural networks can learn the relevant features in the form of distribution models.
The public policing systems employ facial recognition to distinguish people through a trained facial features model,[1] so that they can determine individuals' identities and locations. [2]

Law enforcement often relies on footage of the cameras on streets or in business to review crimes after the fact and catch criminals. ML can apply facial recognition to these images and identify objects and complex events. Police use facial recognition to identify criminals on the

run and missing persons. Since they cannot be in multiple places at once, officers can rely on ML to alert if someone in the area has a weapon or acts abnormal and may be a perceived threat.ML in law enforcement to alert if someone in the area has a weapon or acts abnormal and may be a perceived threat.

Object identification can determine vehicles based on set characteristics by analyzing street footage. When officers look for a stolen vehicle or a criminal on the run in a specific type of vehicle, they can use ML to analyze the digital footage of a given intersection in a period to get the result quickly. Researchers develop ML algorithms to improve detection, recognition, and identification, even with images of poor resolution and low ambient light levels. [3]

Traffic safety systems use ML techniques to decipher a license plate or identify a person in highly low-quality images or video. Researchers degrade high-quality images and compare them with low-quality ones to better recognize lower-quality images and video. Clear plates images are degraded and expressed in mathematical expressions to simulate low-quality images. Experts can distinguish the license plate by comparing degraded images with the original, poor-quality license plate images from digital footage. [4] Law enforcement agencies also work with drone cameras to explore more surface areas and engage in quicker search-and-rescue efforts. They equipped the drones with ML facial and object recognition capabilities.

1.2. Crime Forecasting Predictive Analysis

Crime forecasting Predictive analysis processes considerable amounts of data to forecast possible results. [3] ML predictive policing includes predicting where crimes will occur, the types of crime, the individuals who will commit them, and the victims. Police have started testing predictive policing systems to predict and prevent crimes eventually. For instance, ML algorithms can analyze crime rates across various regions and map crime hot spots when predicting crime locations. Then, police target these spots for extra patrolling and guard.

Predictive policing may be most helpful in identifying possible future victims of crimes. ML models can process large volumes of information on the law, legal precedence, social information, and media. After training, models can output criminal organizations, predict probable criminality, and demonstrate people at risk from criminal organizations. ML predictive analysis can identify potential senior victims who suffer domestic violence. Specifically, the algorithms identify the types of domestic abuse, financial exploitation, physical abuse, or other forms of senior abuse. Meanwhile, models can also determine the lawbreakers, the victims, the violation interval, and the environmental aspects. [8] Moreover, ML can predict potential victims of violent crimes regarding associations and behavior. Experts create social networks on

categorizing the initially collected information and use algorithms to analyze potential high-risk people. Eventually, Chicago Police Department adopted this tool as a portion of the Violence Reduction Strategy. [9]

ML can also paint a clear picture of who is likely to commit a crime or re-offend once released from prison based upon data collection and analysis of historical patterns. However, there is some controversy over ML involving predictive policing. The Police Department creates an automated warrant service triage tool, a geographical reference to look for high-risk fugitives concentrations. [5] The algorithms construct decision trees on a dataset of 340,000 warrant records. The core part is to execute survival analysis to estimate the time period between the next issue of interest may appear.[5] The model also predicts the possibility of reconviction of escaping offenders.

2. Factors Inluencing Fairness

Fairness has different definitions across disciplines. From a law perspective, fairness includes protecting individuals and groups from discrimination or mistreatment. It focuses on prohibiting behaviors, biases. Fairness also involves decisions based on certain protected aspects or social group categories. [8] In quantitative fields, fairness problems are considered mathematical questions. Fairness usually matches some criteria, such as equal or equitable allocation, representation, or error rates. [8] In Philosophy, fairness is what is morally right. Fairness connects justice and equity in the field of Political philosophy. [8]

Regarding decision-making, fairness is an absence of prejudice or bias based on inherent or acquired characteristics of a group or someone. Therefore, an unfair ML algorithm produces outcomes skewed toward a particular group. [9] Biased predictions are likely to originate from hidden biases in data or ignored bias in algorithms. Two potential sources of unfairness in ML results arise from biases in the data and the algorithms.

2.1. Data Bias

Data bias derives from data origins to its collection, processing, and training, leading to unfairness in different learning tasks. Generally, Seven primary sources of bias in ML exist in different processes.

First, Historical bias is a misalignment between reality and the values encoded in a model. Even with perfect sampling and feature selection, the bias stems from the existing bias or social and technical problems already in the real world.[10] Representation bias occurs when the training

samples under-represents some portion of the population and consequently do not generalize well for a data subset. [10] Measurement bias occurs throughout feature selection, labeling, and proxies for desired labels or features. Aggregation bias occurs during the model construction phase when creating a one-size-fits-all model for differential groups with flawed population assumptions. [10] Aggregation bias occurs during the model construction phase when creating a one-size-fits-all model for differential groups with flawed population assumptions. [10] Aggregation bias may exaggerate with representation bias and cause models to fail to fit some groups. When test data does not equally represent the various user population, evaluation bias arises, especially in the model training and evaluation stages. [11] Evaluation bias can lead to overfitting specific benchmarks. Deployment bias arises when a system is built and evaluated as fully autonomous. Learning bias occurs when modeling choices increase performance disparities over different examples in the data.

2.2. Bias example

Risk Assessments in Criminal Justice System have been deployed at several points in criminal justice settings.[12] A well-known example is from a tool, Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), used by courts in the US to make parole decisions. The software measures a person's risk of recommitting another crime. Judges could use COMPAS to decide whether to release an offender or retain them in prison and make decisions around a pretrial release.[9]

An investigation into the COMPAS software found a bias against African-Americans.[13] COMPAS is more probably assigns a higher risk score to African-American offenders than Caucasians with the same profile.[9]The data for these models often include proxy variables such as "arrest" to measure "crime" or some underlying notion of "riskiness." Because minority communities are more policed, the model mismeasured the proxy differentially. There is a different mapping from crime to arrest for minority communities. Many of the other features, like "rearrest" to measure "recidivism," [14] used in COMPAS were also differentially scaled proxies. However, the false-positive rate for black suspects in the COMPAS model results is much higher than white suspects.[15] In other words, the resulting model was more likely to predict that black defendants were at a high risk of re-offending when they were not. [15] Another controversial point is that COMPAS only public seven features out of 137. Some researchers claim that COMPAS decision-making is no better than a simple logistic regression model.[14]

In reality, biases exist in using predictive tools in pretrial risk assessment. Risk assessment tools are driven by algorithms informed by historical crime data, using statistical methods to find

patterns and connections. Thus, it will detect patterns associated with crime, but patterns do not look at the root causes of crime. Often, these patterns represent existing issues in the justice system. Additionally, data can reflect social inequities, even if removing variables such as gender, race, or sexual orientation. As a result, the populations historically targeted by law enforcement are at risk of algorithmic scores that label them likely to commit crimes.

ML can also reinforce human biases. For example, some assessments may perpetuate people's incomprehension and fears that drive mass incarceration. In addition, implicit biases may also influence court decisions. Therefore, ML applications in the justice system need to watch for potential negative feedback loops and implied biases that cause an algorithm to accumulate bias progressively and influence court decisions.

Realizing that these tools are utilized in the justice system and employed to make decisions that affect peoples' lives, we should consider fairness a vital element when inventing and operating these sensitive tools.

3. Fairness Tools to detect and adjust bias

There have been plenty of trials to address bias in ML to achieve fairness. Generally, mitigate biases methods in the algorithms consist of three types.

3.1. mitigate biases

If the algorithm can modify the training data, pre-processing techniques can transform the data to remove the underlying discrimination.[16] If the learning procedure can modify, in-processing can adjust the objective functions or implement a constraint to remove bias during the model training process. [17] Suppose the algorithm cannot modify the training data or learning algorithm. In that case, post-processing can access a test set after training the model, so the appointed labels can be reassigned according to a function. [18]

3.2. mitigate biases tools

What-If and AI Fairness 360 are general tools that can be used to detect and mitigate bias in any machine learning model.

(1) Google What-If Tool (WIT)
Google What-If Tool (WIT) is an interactive tool that allows users to investigate the machine learning bias visually. It provides a way to analyze data sets in addition to trained TensorFlow

models. One example of WIT is the ability to manually edit examples from a data set and see the effect of those changes through the associated model. It can also generate partial dependence plots to illustrate how predictions change when a feature is changed. Once machine learning bias is detected, WIT can apply various fairness criteria to analyze the model's performance (optimizing for group unawareness or equal opportunity).

(2) IBM AI Fairness 360 Tool (AIF360)
AI Fairness 360 from IBM is an open-source toolkit for detecting and removing bias from machine learning models. AI Fairness 360 includes more than 70 fairness metrics and ten bias mitigation algorithms to help detect bias and eliminate it. For instance, Bias mitigation algorithms include optimized preprocessing, re-weighting, prejudice remover regularizer. Metrics include Euclidean and Manhattan distance, statistical parity difference. In addition, the toolkit permit researchers to add fairness metrics and migration algorithms.

(3) Subpopulation Analysis
Subpopulation analysis analyzes a target subpopulation from the whole dataset and calculates the model evaluation metrics for the peculiar population. This analysis can help identify if the model favors or discriminates against a particular section of the population. One way to perform subpopulation analysis is Pandas. It filters the target subpopulation as a new data frame and then calculates the metric for each of the data frames. Another more innovative way of sub-population analysis is Atoti leveraging the power of OLAP to slice and dice the model predictions.

4. Conclusion

New ML applications appear every day, lay the foundations for future utilization growth in the criminal justice system and policing. Bias is a concern in both humans and ML. However, much of the algorithmic bias originated from training data due to human biases. That is why ensuring unbiased training data should be the top priority when deploying risk assessment tools. Consequently, it is highly beneficial to consider all angles of the justice system and do our best to implement ML effectively and correctly. For ML to be used as effectively as possible in the justice system, we must solve ethical and social considerations. ML methods provide investigative assistance and will be a permanent part of our justice ecosystem. ML assists criminal justice specialists in managing law enforcement and policing systems and ultimately improving public safety.

5. Reference

[1]. The Intelligence Advanced Research Projects Activity, "Janus," Washington, DC: Office of the Director of National Intelligence, https://www.iarpa.gov/index.php/ research-programs/janus.

[2]. National Science and Technology Council and the Networking and Information Technology Research and Development Subcommittee, The National Artificial Intelligence Research and Development Strategic Plan, Washington, DC: Office of Science and Technology Policy, October 2016, https://www.nitrd.gov/PUBS/national_ai_rd_ strategic_plan.pdf.

[3]. Christopher Rigano, "Using Artificial Intelligence to Address Criminal Justice Needs," October 8, 2018, nij.ojp.gov: https://nij.ojp.gov/topics/articles/using-artificial-intelligence-address-criminal-justice-needs

[4]. "DeGrade It" at Dartmouth College, NIJ award number 2016-R2-CX-0012.

[5]. "Applying Data Science to Justice Systems: The North Carolina Statewide Warrant Repository (NCAWARE)" at RTI International, NIJ award number 2015-IJ-CX-K016.

[6]. "Exploring Elder Financial Exploitation Victimization" at the University of Texas Health Science Center at Houston, NIJ award number 2013-IJ-CX-0050.

[7]. "Chicago Police Predictive Policing Demonstration and Evaluation Project" at the Chicago Police Department and Illinois Institute of Technology, NIJ award number 2011-IJ-CX-K014.

[8]. Mulligan, D., Kroll, J., Kohli, N. & Wong, R. (2019). This thing called fairness: Disciplinary confusion realizing a value in technology. ACM Human-Computer Interaction, 3, 119. https://doi.org/10.1145/3359221.

[9]. Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A Survey on Bias and Fairness in Machine Learning. ACM Comput. Surv. 54, 6, Article 115 (July 2022), 35 pages. DOI:https://doi.org/10.1145/3457607

[10]. Suresh, H., & Guttag, J.V. (2021). A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. Equity and Access in Algorithms, Mechanisms, and Optimization.

[11]. Suresh, H., & Guttag, J.V. (2019). The Problem with "Biased Data". https://harinisuresh.medium.com/the-problem-with-biased-data-5700005e514c

[12]. Henry, Matt. 2019. "Risk Assessment: Explained." The Appeal, December 14, 2019.

[13]. HTTPS://WWW.PROPUBLICA.ORG/ARTICLE/MACHINE-BIAS-RISK-ASSESSMENTS-IN-CRIMINAL-SENTENCING

[14]. Julia Dressel and Hany Farid. 2018. The accuracy, fairness, and limits of predicting recidivism. Science Advances 4, 1 (2018). https://doi.org/10.1126/sciadv.aao5580 arXiv:https://advances.sciencemag.org/content/4/1/eaao5580.full.pdf

[15]. Suresh, H., & Guttag, J. (2021). Understanding Potential Sources of Harm throughout the Machine Learning Life Cycle. MIT Case Studies in Social and Ethical Responsibilities of Computing, (Summer 2021). https://doi.org/10.21428/2c646de5.c16a07bb

[16]. Brian d'Alessandro, Cathy O'Neil, and Tom LaGatta. 2017. Conscientious classification: A data scientist's guide to discrimination-aware classification. Big data 5, 2 (2017), 120–134.

[17]. Rachel KE Bellamy, Kuntal Dey, Michael Hind, Samuel C Hoffman, Stephanie Houde, Kalapriya Kannan, Pranay Lohia, Jacquelyn Martino, Sameep Mehta, Aleksandra Mojsilovic, et al. 2018. AI fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. arXiv preprint arXiv:1810.01943 (2018).

[18]. Richard Berk, Hoda Heidari, Shahin Jabbari, Matthew Joseph, Michael Kearns, Jamie Morgenstern, Seth Neel, and Aaron Roth. 2017. A Convex Framework for Fair Regression. (2017). arXiv:cs.LG/1706.02409