Machine learning in Justice System and Policing

Yiqin Zhang
evezhang@bu.edu

With the applications of big data analytics, machine learning, and artificial intelligence, both the justice system and public policing are becoming increasingly technologically advanced. Public safety benefits from machine learning (ML) extensively. For example, transportation and traffic systems identify violations and implement rules of the road. In addition, crime forecasts benefiting from ML provide a more efficient allocation of policing resources. ML is also helping to identify 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

Video and image analysis has been around for some time now in criminal justice and law enforcement to obtain information about people, objects, and actions to support criminal investigations. However, the analysis is very time-consuming and labor-intensive, requiring a significant investment in personnel with subject matter expertise. In addition, traditional software is limited to predetermined eye shape, eye color, and eye distance for facial recognition or demographic information for pattern analysis. Instead, video and image ML algorithms learn multiple tasks and develop and determine their own independent complex facial recognition features to accomplish these tasks ahead of humans. As a result, these algorithms have the potential to match faces, identify weapons and other objects, and detect complicated events such as accidents and crimes.

In the National Artificial Intelligence Research and Development Strategic Plan[1], Facial Recognition establishes an individual's identity and whereabouts. Using computer vision in the Janus project, an Intelligence Advanced Research Projects Activity, analysts are performing trials on algorithms to distinguish one person from another by facial features in the same manner as a human analyst. [2]

Law enforcement often relies on footage of the cameras on streets or in businesses to review crimes after the fact and catch criminals. Police use facial recognition to identify criminals on the run and missing persons. Moreover, object identification is crucial for police officers monitoring significant events, such as music festivals or marathons. ML can apply facial

---

[1] 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.

[2] 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.

recognition to these images and identify objects and complex activities like car accidents. Since they cannot be in multiple places at once, officers can rely on ML in law enforcement to alert if someone in the area has a weapon or acts unusually and may be a perceived threat.

Object identification can be used to identify vehicles based on set characteristics by analyzing street footage. For instance, 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 footage of a given intersection in a period to get the result quickly. Institutes improve the speed, quality, and specificity of data collection, imaging, and analysis and improve contextual information. Meanwhile, 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 technologies to decipher a license plate or identify a person in highly low-quality images or video. Researchers at Dartmouth College systematically degrade high-quality images and compare them with low-quality ones to better recognize lower-quality images and video. Clear images of numbers and letters are degraded to emulate low-quality images and then expressed and cataloged as mathematical representations. These degraded mathematical representations can be compared with low-quality license plate images to help identify the license plate.[4]

Law enforcement agencies also work with drone cameras to explore more surface areas and engage in quicker search-and-rescue efforts. These drones are usually equipped with ML facial and object recognition capabilities.

1.2. Crime Forecasting Predictive Analysis

Crime forecasting Predictive analysis processes large amounts of data to forecast and formulate potential outcomes. [3]ML predictive policing refers to predicting where crimes will occur, the types of crime, the individuals who will commit them, and the victims. Companies and police officials have started testing predictive policing systems, eventually predicting and preventing crimes. For instance, ML algorithms can analyze crime rates across various areas and map crime hot spots when predicting crime locations. Then, it tells police to target these areas for extra patrolling and surveillance.

Predictive policing may be most helpful in identifying possible future victims of crimes. Large volumes of information on the law and legal precedence, social information, and media can be processed in ML algorithms. So it can suggest rulings, identify criminal enterprises, predict their crime, and reveal people at risk from criminal organizations. 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

---

[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.

collection and analysis of historical patterns. However, there is some controversy over ML involving predictive policing.

ML can analyze large volumes of criminal justice-related records to predict potential criminal recidivism. The Police Department of Durham and Anne Arundel County collaborate with the Research Triangle Institute to develop an automated warrant service triage tool for the North Carolina Statewide Warrant Repository. Researchers use algorithms to analyze datasets of beyond 340,000 warrant records. [5]The algorithms construct decision trees and perform survival analysis to determine the time until the next issue of interest and predict the risk of re-offending for escaping offenders. This model will help practitioners triage warrant service when backlogs exist. So practitioners can reference the tool geographically and pursue concentrations of high-risk absconders, along with others who have active warrants.[5]

ML can help determine potential elder victims of physical and financial abuse. Researchers at the Health Science Center from the University of Texas used ML algorithms to analyze elder victimization. The algorithms can determine the perpetrator, the victim, and environmental factors distinguishing between financial exploitation and other forms of elder abuse. They can also differentiate "pure" financial exploitation from "hybrid" financial exploitation. The researchers work to transform data algorithms into web-based applications so that practitioners can reliably determine the likelihood of financial exploitation and quickly intervene. [6]

Finally, ML can predict potential victims of violent crime regarding associations and behavior. The Illinois Institute of Technology used ML algorithms to collect information and form initial groupings that construct social networks and analyze potential high-risk individuals. Then it becomes a part of the Chicago Police Department's Violence Reduction Strategy.[7]

2. Factors Impacting Fairness of Machine Learning

Fairness has different definitions across disciplines. From a law perspective, fairness includes protecting individuals and groups from discrimination or mistreatment, focusing on prohibiting behaviors, biases, and basing decisions on certain protected factors or social group categories. [8] Meanwhile, in quantitative fields, like math, and computer science, questions of fairness are seen as mathematical problems. Fairness tends to match some criteria for a particular task or problem, such as equal or equitable allocation, representation, or error rates. [8] Within

---

[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.

Philosophy, ideas of fairness "rest on a sense that what is fair is also what is morally right." Furthermore, Political philosophy connects fairness to notions of justice and equity. [8]

In the context of decision-making, fairness is an absence of prejudice or favoritism to a group or an individual based on inherent or acquired characteristics. Thus, an unfair algorithm is with decisions skewed toward a particular group of people.[9] Biased predictions stem from hidden or neglected biases in data or algorithms. As a result, two potential sources of unfairness in machine learning outcomes arise from biases in the data and the algorithms.

2.1. Bias in Data

Bias in data can exist in various shapes and forms, leading to unfairness in different downstream learning tasks. Seven primary sources of bias in ML exist in different cycles from data origins to its collection and processing.

(1) Historical bias**:**
Historical bias is a misalignment between reality and the values encoded in a model. The bias derives from how the world is or was when the data was collected. It is the already existing bias and socio-technical issues in the world. It can seep into the data generation process, even with perfect sampling and feature selection.[10]

(2)Representation bias
Representation bias occurs when the development sample under-represents some portion of the population and subsequently fails to generalize well for a subset of the user population.

(3) Measurement bias
Measurement bias occurs throughout feature selection and labeling, proxies for desired labels or features. However, the chosen set of features or labels might leave out essential characteristics or "introduce group- or input-independent noise that leads to differential performance." [10] Generally, proxies may be ranked differently over groups with differential measurement errors. Since the measurement process differs across groups, the quality of the data varies across groups. The defined classification task is oversimplified, using simple labels as proxies for complex predictions.

(4)Aggregation bias
Aggregation bias occurs during the model construction phase "when distinct populations are inappropriately combined." [10]In other words, a one-size-fits-all model is designed for

---

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.

differential groups. When combined with representation bias, it can lead to models that fit only the dominant population or do not work on any subgroups.

(5)Evaluation bias
Evaluation bias arises in the model iteration and evaluation phases when "testing or external benchmark populations do not equally represent the various parts of the user population." [11]It can also be exacerbated r when the performance metrics are inappropriate for real-world use. Evaluation bias can lead practitioners to overfit particular benchmarks.

(6)Deployment bias
Deployment bias arises when a mismatch exists between the problem a model is intended to solve and how it is used. It often occurs when a system is built and evaluated as fully autonomous. In reality, it operates in a complicated socio-technical system moderated by institutional structures and human decision-makers. In some cases, systems produce results that human decision-makers must first interpret. As a result, despite good performance in isolation, they may end up causing harmful consequences regarding phenomena, like automation or confirmation bias.

(7) Learning bias
Learning bias occurs when modeling choices increase performance disparities over different examples in the data. For instance, the objective function of the ML algorithm is an essential modeling choice during training. Typically, objective functions encode some measure of accuracy on the task, such as mean squared error for regression and cross-entropy loss for classification problems. However, when prioritizing one objective, like overall accuracy, another may damage, like disparate impact.

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.[13]

An investigation into the COMPAS software found a bias against African-Americans.[14] COMPAS is more probably assigns a higher risk score to African-American offenders than Caucasians with

---

[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] 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.

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

the same profile. [15] 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, this proxy is differentially mismeasured. There is a different mapping from crime to arrest for minority communities. Many of the other features, like "rearrest" to measure "recidivism," [16]used in COMPAS were also differentially scaled proxies. The resulting model had a significantly higher false-positive rate for black defendants versus white defendants. In other words, it was more likely to predict that black defendants were at a high risk of re-offending when they were not. [17]Although COMPAS uses 137 features, only seven were presented to the public. Furthermore, the study found that COMPAS is not better than a simple logistic regression model when making decisions.[18]

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 misconceptions and fears that drive mass incarceration. Implicit biases may also influence court decisions. Therefore, users of ML in the justice system need to watch for potential negative feedback loops that cause an algorithm to become increasingly biased over time.

Recognizing that these tools are used in courts and make decisions that affect peoples' lives, we should consider fairness constraints crucial while designing and engineering 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, methods to mitigate biases in the algorithms fall into three categories.

---

[15] 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.

[16] 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

[17] 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

[18] 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

3.1. mitigate biases

(1) Pre-processing.
Pre-processing techniques try to transform the data to remove the underlying discriminationif the algorithm is allowed to modify the training data.[19]

(2) In-processing.
In-processing techniques modify state-of-the-art learning algorithms to remove discrimination during the model training process. [20] If the learning procedure can change for an ML model, in-processing can be used by combining changes in the objective function or implementing a constraint.[21]

(3) Post-processing.
Post-processing accesses a holdout set after the training of the model. If the algorithm cannot modify the training data or learning algorithm, it can only treat the learned model as a black box. In that case, the labels assigned by the black-box model initially get reassigned based on a function during the post-processing phase.[22]

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
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
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

---

[19] 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.

[20] 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.

[21] 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).

[22] 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

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

Every day has potential for new ML applications in criminal justice, paving the way for future possibilities to support the criminal justice system and ultimately enhance public safety. 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 has the potential to be a permanent component of our criminal justice ecosystem, providing investigative assistance and allowing criminal justice professionals to manage public safety better.