Machine Learning in Justice System and Policing

Zihang Jiang
jzh15@bu.edu

Machine learning algorithms is growing more and more popular in justice system and policing to help people decide who can be bailed and who should be arrested. People use these algorithms not only because they are efficient and fast which helps officers avoid spending lots of time on making decisions, but also as a consequence of people believing that these algorithms are more accurate, more objective and more dispassionate, which prevents bias in human's intuition and opinions.

How can machine learning algorithms be applied to these areas? The principle is that machine learning algorithms can also make predictions after they are trained with a variety of data. Compared to human decision, one of the biggest advantages of machine learning algorithms is that they are far faster than human. Based on experience and intuition, a human may make a decision upon the defendant in several seconds, while predicting model can make thousands of decisions at the same time with great accuracy, which greatly shortens the working time and reduce the heavy burdens on judges and police officers. Another advantage lies in that machine learning algorithms can have a larger view and more considerate analysis than human. Take police patrol area as an example, without predicting model, police officers may decide where to patrol only based on their previous experience and intuition. But a machine learning algorithm can analyze all crime data happened in the whole city in past years and predict which area has a larger possibility to commit crimes and needs more patrol wagons, which is impossible for human to do that. There is also a good wish from human that machine learning algorithms can eliminate the bias that exist in human's minds and make decisions more fairly. In people's opinion, machines are reliable because they are emotionless and impersonal. They only make decisions based on cold data, so without emotion, they also don't have bias, which then make predictions more objective and accurate. However, as things stand now, it seems that those machine learning algorithms still have bias on human's race, human's sex, etc, and sometimes they even amplify those bias.

How the bias happen and what factors impact the fairness? The predicting algorithms of machine learning have a basic pattern that they make predictions based on the data fed into them including age, sex, race and what defendants have done in the history. The biggest reason leading to bias lies in here: what if the data fed into those predicting models are unequal? On that situation, the predicting model may even amplify the bias in the data. The most horrible thing is that people even don't realize that the data used for model training are unequal. The experiment done by Dartmouth professor Dr. Hany Farid shows that there still exist differences in false positive and false negative rates between black and white even though the data don't include race information. What people are unconscious of is that the data themselves mirror the inequality in our society and reflect the bias among different people. Researchers may think that the data they choose aren't related to any aspect that have bias, which can be only basic information like age and sex. But the results getting from these data are still unfair and with prejudice, because the data they use come from the society, and there exists lots of bias in the society, leading to inequality in data. In this way, inequality is hidden behind algorithms, making it harder to be found and realized. Another reason behind the fairness is that the users like judges and

police officers usually don't know how the predicting algorithms work and the principles behind them, which may lead to wrong decisions. What they only know is that the predicting results come from a predicting model using latest and most advanced artificial intelligence technologies. In that scenario, users may overvalue the predicting results from machine learning models. However, in the same experiment done by Dr. Hany Farid, the predicting accuracy of simple two-factor model is similar to that of nowadays commercial predicting models, and also similar to that from human judgement. The experiment show at least one thing that the principle behind those fancy commercial models may be quite simple and they probably are not able to carry out sophisticated inferences, especially in those complex areas like bailing and arresting people. The results from predicting models also follow human intuition. For example, people that are older and committed less crime in the past are predicted to be less likely to commit crime in the future while people younger and committing more crime before are predicted to be more likely to do that. In this case the intuition behind the predicting results is quite simple. If you tell the judge your prediction based on your intuition about age and crime history, the judge is most likely to ignore your advice. But if the predicting model tell the judge the same result following the same intuition logic, the judge will cautiously consider its advice, in which the problem lies.

Due to the fact that there exists bias in predicting model, fairness tools are developed to detect and adjust bias. Take AI Fairness 360 developed by IBM as an example, the basic working flow includes checking for bias in the initial training data, mitigate the bias, and recheck, which can be divided into five steps. Firstly, importing necessary packages needed including aif360, numpy and sys. Secondly, setting necessary parameters including bias detection options and processing the dataset. Thirdly, computing fairness metric on original training dataset. A common used detecting method called mean_difference is to compare the percentage of favorable results for the privileged and unprivileged groups, subtracting the former percentage from the latter. Fourthly, mitigating bias by transforming the original dataset. There are many bias mitigation algorithms to choose from, which includes reweighting, optimized pre-processing, adversarial debiasing and reject option biased classification. Finally, computing fairness metric on transformed training dataset to test the effectiveness of bias mitigation algorithm.

The bias in the machine learning algorithms can lead to serious consequences even disasters, which making it not only a technical problem but also a social issue. Just imaging a scenario that one day a version of wrong predicting model with heavy bias is updated over judge and police system, where innocent people are arrested and bad guys are bailed due to wrong prediction, you can find how unacceptable the bias are. What's more, I think from some prospective machine learning is also a ethic problem. A hidden fact lies behind machine learning is that what you have done in the past decide what you will do in the future. If you buy something on the Internet, the similar goods will be recommended to you as you are predicted to be more likely buy those things. If you did something like buying a knife, you are more likely to be arrested because the chance of machine learning algorithms predicting you to commit a crime in the future increase. But for our humans, it sounds very horrible if anything you will do in the future can be predicted by machine learning algorithm. Those algorithms should be strictly regulated because not only we human are quite complex and it's easy to make mistakes in future action prediction, but also we don't want any past actions are fully analyzed and any future steps are predicted by machines. Another thing is that a real person is behind each statistic in prediction model. Unlike in other field such as image classification and semantic segmentation where the model can make mistakes, in justice and policing system, the prediction models shouldn't make any mistakes as each mistake leads to an innocent person suffering. However, all current commercial predicting models can't reach 100% accuracy, which means there are many wrong predictions made. Just as Dr. Hany

Farid said, a technical and ethic panel should be set up to consider and deal with those problems. The technology should follow the rules of society and regulations should be developed to supervise it.

Reference

[1] A Survey on Bias and Fairness in Machine Learning. NINAREH MEHRABI, FRED MORSTATTER, NRIPSUTA SAXENA, KRISTINA LERMAN, and ARAM GALSTYAN. https://arxiv.org/pdf/1908.09635.pdf

[2] How AI Could Reinforce Biases In The Criminal Justice System. https://www.youtube.com/watch?v=ZMsSc_utZ40

[3] The danger of predictive algorithms in criminal justice | Hany Farid. https://www.youtube.com/watch?v=p-82YeUPQh0

[4] IBM Fairness Tool. https://aif360.mybluemix.net/