Data Bias and Algorithmic Discrimination

Tong Chang

Artificial intelligence is affecting people's lives, both online and in the real world. Algorithms transform people's online footprints and activities, such as their online habits, shopping history and GPS location data, into scores and predictions of people. These ratings and predictions, in turn, shape decisions that affect people's lives, making discrimination and inequity a significant issue whether people are aware of discrimination or not. This article mainly studies some typical examples of data discrimination in real life, as well as observations and ideas for correction of data discrimination

1 Introduction

With big data, machine learning, artificial intelligence, algorithms and so on as the core of the application of automatic decision-making system is becoming more and more widely, from shopping to recommend, personalized content recommendation, accurate advertisement to loans, insurance assessment, employee appraisal to the crime in the judicial process of risk assessment, decision-making more and more of the replaced by machines, algorithms and artificial intelligence, Algorithms can bring complete objectivity to all kinds of affairs and decision-making work in human society. But this is wishful thinking. In any case, the design of algorithms is the subjective choice and judgment of programmers, and it is questionable whether they can impartially write existing legal or ethical rules into the program. Algorithmic Bias thus becomes a problem that needs to be addressed. The problems of opacity, inaccuracy, unfairness and difficulty to review caused by rule codification need careful consideration and research.

2 Real World Examples Of Algorithmic Bias And Social Impact

In the era of artificial intelligence, algorithms have strong independent learning ability and can continuously learn and extract rules from massive data, thus forming automatic decisions, greatly improving efficiency, meeting personalized needs, and bringing convenience to people's work and life. But algorithms seem to naturally hide in black boxes. The algorithm of an enterprise or website is often completed by tens of hundreds or even thousands of engineers, but the algorithm is not completely generated in accordance with the code written by the engineer, but the machine self-learning training, continuous adjustment and optimization. Its complexity and opacity have caused many problems.

More and more practical experience shows that seemingly neutral algorithms can also produce discriminatory results. When the algorithm produces discrimination or causes infringement, it is difficult to be detected by people, and even if detected, people are difficult to protect their rights.

2.1 Price Discrimination

In recent years, price discrimination has frequently attracted attention in China. The price of the same product for the old users who recharge the membership is higher than that for the new users; Take-out from the same store increases in price after several meals; Using several mobile phones to book air tickets and hotels in the travel software at the same time, there are price differences, and the booking interface of "regular guest" will mark "will be sold out", "room shortage" and other false information. Users reported that they encountered price discrimination, involving Tmall, Meituan, Didi and other major e-commerce platforms.

Price discrimination also happen in coupons. Coupons are used in retail in order to differentiate customers by reserve price. The hypothesis is that those who go through the trouble of collecting coupons are more price sensitive than those who don't. Thus, for example, offering coupons allows breakfast cereal makers to charge less price-sensitive customers more, while still making some profit from more price-sensitive customers.

Understanding different consumers' willingness to pay is one of the important prerequisites for enterprises to implement price discrimination, which is not difficult for Internet platforms in the era of big data. Through data mining and other technical means, the platform can acquire a large amount of user data, including shopping history, browsing history, geographic location, mobile phone model, etc., and make accurate portraits of users, so as to make differentiated product recommendations and pricing.

From the point of view of economics, it is a differentiated treatment made by businesses in order to maximize their interests, which is in line with economic rationality. However, as e-commerce platforms are in a strong position with technological and capital advantages, while ordinary users are in a weak position, there is a serious imbalance between the two sides.

In addition, due to the black-box nature of the algorithm, the decision of the algorithm is not transparent to users and there is information asymmetry. The platform makes differentiated pricing without consumers' knowledge, which infringes consumers' legitimate rights and interests such as the right to know and the right to fair trade and overdraws consumers' trust.

2.2 Educational Discrimination

According to the *Big Data Report: Algorithmic Systems, Opportunity, and Civil Rights* released by the White House in 2016, higher education institutions, by collecting and analyzing data on a large number of applicants, may predict their likelihood of graduation before they enroll, and then decide whether to admit them. Family income is one of the important factors in making predictions, which may lead universities to consider it not worthwhile to provide additional financial support to some applicants, such as students from poor families or facing special challenges in graduation, in order to cut costs, thus making them face barriers to admission and leading to discrimination.

Besides, according to the Washington Post, admissions officers at the University of Wisconsin used a little-known but increasingly common practice: They installed tracking software on the school's website that automatically identifies who a student is when he or she visits the site based on a cookie code and generates a student's profile, This includes name, contact information, race, high school, detailed web browsing information (content of all pages viewed and how long each page was viewed), geographic location, etc., to assess his/her level of interest in the college. Colleges are gathering more data about their prospective students, with at least 44 public and private colleges across the country working with outside consulting firms to measure each student's likelihood of enrolling by tracking their online activity or developing predictive scores.

In another example of academic discrimination, A-level exams have been cancelled in the UK, and the British government has developed A smart grading system to rate students' performance based on the overall performance of school districts. The distribution of students' scores from 2017 to 2019, the ranking of students' scores in the test area, and the examinees' personal scores in the past are important criteria for scoring by the algorithm. It eventually ranks past test scores in the test area to determine a candidate's final score. Under the algorithm's system, the "school district" of a candidate becomes the key to the final score. Many of the graduates' grades have been downgraded by the algorithmic system and are lower than the teachers' expectations.

Some students who had consistently scored A+ in math on mock tests in the past had their math scores dropped one grade because their public high school was not rated well overall. Many see the system as a system of discrimination against children from less well-off backgrounds.

It can be seen that the risk of discrimination is hidden in the algorithm, and the decisions it makes are likely to solidify social discrimination and aggravate social injustice. Algorithmic discrimination should be taken seriously.

3 Thoughts For Correction Of Data Discrimination

Because algorithmic bias is unknowable and untraceable, it makes anti-bias work tricky. Under the existing response system, whether it is policy institutions, technological breakthroughs or innovative countermeasures, we are trying to solve this problem that has gone beyond technology from different perspectives.

3.1 Solution 1: Build More Impartial Data Sets

Unfair data sets are fertile ground for bias -- if the data sets used to train machine learning algorithms are not representative of objective reality, the results of the application of that algorithm are often biased against certain groups of people. Therefore, the most direct solution to algorithm bias is to adjust the original unbalanced data set.

Corrected data ratios: Use more equitable data sources to ensure fairness in decision making. Many companies are already doing this with remarkable results. In June 2018, Microsoft worked with experts to modify and extend the data set used to train the Face API. The Face API is an API in Microsoft Azure that provides pre-training algorithms to detect, recognize, and analyze properties in Face images. The new data reduced the identification error rate between dark-skinned men and women by 20 times, and for women by nine times, by adjusting for the proportion of skin, gender and age. There are also companies trying to optimize data sets by building global communities. By pooling whatever information an organization might be looking for on a large scale through global communities, and doing it in such a combination of breadth and depth, it makes it possible to bring in radically different data to train AI systems to help overcome problems such as algorithm bias.

Combining "big data" with "small data" does produce better results, ensuring accuracy based on the volume of data. Data sets should not be limited to extensive collection, but to precise grasp. Merely focusing on the volume of data often fails to bring about more fair results, because big data analysis focuses on correlation, which leads to errors in deducing causality. The introduction of small data can partly solve this problem. Small data refers to the data form that focuses on individual users. It pays more attention to details and differences, and can present more accurate data and avoid errors when deducing causality. Therefore, the combination of big data with rich information and small data with accurate information can avoid errors to a certain extent.

Building a more fair data set is undoubtedly one of the fundamental solutions to algorithm bias, and it is also the direction of efforts of many enterprises and scholars. At present, breakthroughs have indeed been made in this field.

3.2 Solution 2: Improve "Algorithm Transparency"

Although algorithm models are written by engineers, many times humans do not understand what process a computer goes through to arrive at a particular result. This is the "algorithm black box" problem in machine learning. Therefore, enterprises are required to improve the transparency of the algorithm model and find out the "cause" of bias, which has become one of the ways to solve the "black box" dilemma. Whether it is through policy and "heteronomy" of provisions, or through ethical "self-discipline" or technological exploration, companies continue to focus on opening the "black box" when fighting algorithmic bias.

3.2.1 Self-Discipline: The Ethical Proposition Of Companies

Over the past two years, a number of large tech companies have published principles for the application of AI, all of which involve bias governance. These principles can be seen as statements of position and a starting point for self-regulation. Microsoft, Google and IBM all emphasize transparency and fairness in algorithms. Microsoft has set up an ARTIFICIAL Intelligence and Ethical Standards (AETHER) committee to enforce its principles, and plans to have every AI product launched in the future undergo an AI ethical review. Google's Model Cards feature is also a response to greater transparency by going outside of committee. Model Cards is similar to the algorithm manual, which explains the adopted algorithm, informs its advantages and limitations, and even the calculation results in different data sets.

3.2.2 Heteronomy: Monitoring Process Transparency And Outcome Justice

The *GENERAL Data Protection Regulation of the European Union (GDPR)*, which came into effect in 2018, and the *Data Ethics Framework* updated by the British government require algorithms to have certain openness, transparency and interpretability. On April 10, 2019, members of both houses of Congress introduced the Algorithmic Accountability Act, which would require large tech companies to assess the impact of their automated decision-making systems and eliminate bias based on race, color, religion, political belief, gender, or other characteristics.

3.2.3 Transparent Algorithm Still Has Disadvantages

However, there are still some limitations to improving algorithm transparency through policy regulations and ethical guidelines. First of all, the requirement that the algorithm be interpretable conflicts strongly with the interests of the possible enterprise. Rayid Ghani, director of the Center for Data Science and Public Policy at the University of Chicago, argues that simply publishing all the parameters of a model does not provide an explanation of how it works. In some cases, giving away too much information about how an algorithm works could allow malicious people to attack the system. A December 2019 paper also noted that variants of LIME and SHAP, the two main technologies used to explain black-box algorithms, could be hacked, meaning that "explanations made by AI could be deliberately tampered with, leading to a loss of trust in the model and the explanations it gives".

Second, the core of the Accountability Act is to encourage companies to examine themselves. However, this top-down system has undoubtedly increased a huge amount of work for enterprises. In the round of review and evaluation, the progress of technology will be limited, and the innovation of enterprises will be affected.

3.3 Solution Three: Technological Innovation Against Prejudice

When biases were hidden in countless pieces of code, engineers wanted to solve technical problems with technology itself. This approach is not to start from the source of bias, but to creatively use technology to detect and remove bias.

Bias Detection Tool: In 2018, Google launched what-If, a tool for detecting bias in TensorBoard. With this tool, developers can explore the feature importance of machine learning models, find out the cause of misclassification, determine decision boundaries, and detect algorithm fairness through interactive visual interfaces and counterfactual reasoning.

Technology itself, being used to combat prejudice, is a highly maneuverable approach because engineers tend to be good at using technology to solve practical problems. However, from the current results, most of the technological breakthroughs are still at the initial stage, remaining in the detection of bias, and eliminating bias may still need to be worked on in the next stage.

4 With Big Data, Will Discrimination Disappear?

It might be the opposite. Big data and machine learning algorithms may be more prone to bias than we are.

Why does discrimination exist? Broadly speaking, discrimination stems from a disapproval of a way of thinking or acting. Discrimination does not mean that the person being discriminated against necessarily has a problem, only that there are differences. Then what is the core of the big data era? More "information sharing". Information sharing means that more personal information is made public and available to the public. This means that the impression of a person or group will be further formed.

Take one example: Big data tells us that people in xyz place love to drink. For those who oppose alcoholism, because of their displeasure with such behavior, thinking, and expression. This leads to "discrimination". Therefore, I believe that in the era of big data, discrimination will further expand. To put it simply, how would us feel about people we know who are doing things we don't approve of every day through big data?

5 Conclusion

More fair data sets, more timely error detection, more transparent algorithmic processes... A concerted effort by technology companies, research institutions, regulators and third parties to combat algorithmic bias. Such measures may not eliminate prejudice completely, but they do a great deal to prevent technology from amplifying prejudices inherent in society. Algorithm prejudice than will completely to blame technology, more important is realized, the technology as a tool and the application should have boundaries. It penetrates into the depth of the daily life, and prudent decision-making is needed.

REFERENCES

[1] A Survey on Bias and Fairness in Machine Learning
NINAREH MEHRABI, FRED MORSTATTER, NRIPSUTA SAXENA, KRISTINA LERMAN, and ARAM GALSTYAN, USC-ISI

[2] RDP Binns. 2018. Fairness in machine learning: Lessons from political philosophy. Journal of Machine Learning Research (2018).

[3] A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle, Harini Suresh, John V. Guttag

[4] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In Advances in neural information processing systems. 4349–4357

[5] Alexandra Chouldechova and Aaron Roth. 2018. The frontiers of fairness in machine learning. arXiv preprint, arXiv:1810.08810 (2018).