Algorithmic Fairness – Fair Risk Assessment

Marcello Di Bello - ASU - Fall 2021 - Week #7

Our goal is twofold: one, to track people's ideas of fairness about data and risk assessment; two, to explore the history of risk assessment with a focus on banks and insurance companies.

(Un)fair Data

Earlier[1] we discussed how data can be biased: reliance on proxy variables that weakly track the outcome of interest; differences in sample size for data about socially relevant groups; data mirroring existing social biases; data not equally representative; etc.

    We now ask a different question. What sort of data would it be (un)fair to use in predictive algorithms, focusing on lending and insurance policy decisions? This question can be examined from different disciplinary angles. In the interest of interdisciplinarity, we examine this question by looking at Barbara Kiviat's research in economic sociology.[2]

[1] See Class Notes Week #3.

[2] Kiviat (2021), Which Data Fairly Differentiate? American Views on the Use of Personal Data in Two Market Settings, *Sociological Science* 8: 26-47

The survey

Kiviat constructed a representative sample of the US population and asked participants questions about what types of data they would consider fair to use in lending and insurance policy decisions. Data included things such as: accident history, speeding tickets, driving style, credit score, when and where a person drives, Zip code, rent payment on time, utility and TV bill payment on time, income.

Check out findings on pp. 34-35 of the paper.

Relatedness: statistical and logical

Many respondents maintained that fair data should be *related* to the outcome of interest, such as timely loan repayment in lending or no car accident in car insurance decisions. Relatedness is understood, first, as merely statistical. One respondent noted:

> "If there is a statistical correlation between how often you move and accident history, this seems like a fair thing to consider." (p. 37)

To beat competition and incraese profits, many lenders are now using alternative credit data—non-traditional data about creditworthiness. Check out reports about Alternative Credit Data by companies such as Experian or Transunion.

    Relatedness is understood, second, as *logical* in at least two distinct senses. For example, others respondents said:

> "If you don't pay other bills on time it is probably a good indicator if you would pay off a loan . . . any bill would be fair to use." (p. 37)

> "If a person is responsible in their credit, they're more likely to be a responsible driver." (p. 38)

In other words, logical relatedness is understood as: (a) the data describe a situation that is *similar* to the one to be predicted (e.g. paying bills is similar to paying mortgage); or (b) the data reveal *moral character traits* that are likely to persist across situations that are dissimilar (e.g. responsibility in credit repayment and in driving).

Kiviat notes an interesting—if not disturbing—implication of this distinction that many lenders or insurers might pick up.

> "...a focus on personal disposition, rather than situational similarly, enables data to reach further, into a broader range of new situations ... Using data to essentialize people helps make those data seem fair to use in more distant market settings." (p. 38)

When are generalization acceptable?

Making decisions about individuals based on group characteristics, averages, long-term tendencies seems objectionable in its own way:

> "predictions are about what is true on average for a group of people, and yet those who receive treatment are specific individuals." (p. 39)

Survey respondents, however, had a more discerning view. They distinguished between morally acceptable averages and morally unacceptable ones. Respondents said:

> "A person might have to work 3rd shift or be a drug dealer. Both work at night and shouldn't be judged on the same criteria."

So they made a distinction between groupings:

> "The problem is not that people have been grouped but that they have been *inappropriately* grouped. When the boundaries of a group capture too many different sorts of people and behavior, holding each individual responsible for the average is likely to assign blame where it does not belong, a sort of moral ecological fallacy." (p. 39)

History of risk assessment

Current debates about algorithmic fairness happened before in different ways under different terminologies in 1970s. Instead of *algorithmic fairness*, the more common terminology back then was *actuarial fairness*. Following the work of Ochigame[3], it is instructive to sketch a history of fair risk assessment, reaching back to the 17th century when the concepts of probability and chance were first theorized mathematically.

[3] Ochigame (2020), The Long History of Algorithmic Fairness, Phenomenal World. Check out Ochigame's alternative search engine at searchatlas.org

Fairness and gambling (17th century)

Notions of fairness were closely connected with questions of chance, uncertainty and the future. For example, consider the problem of points that prompted early developments in probability theory. Suppose we play a game in which if I get three heads first, I win and get all the money on the table. If you get three tails first, you win and get all the money. Suppose the game is interrupted. Say I got two heads, and you got one tail. How should we *fairly distribute* the money on the table? Each of us has contributed equal amounts of money.

Algorithms and mathematization were also seen as the solution to interminable disputes, war, social divisions and diplomatic intractability as in Leibniz' *Calculemus!*

The mathematization of jurisprudence (18th century)

Condorcet (1743-1794) and Laplace (1949-1827) set to mathematize the moral sciences, especially jurisprudence and legal decision-making. For example, Condorcet's jury theorem shows that the accuracy of jury decisions improves monotonically the more people are involved, assuming decisions are by majority voting and each juror is even slightly better than chance at tracking the truth. Hence, juries should decide by majority voting and consist of several people.

This quantitative and probabilistic approach to the justice system fell into disrepute in the 19th century, partly because of the fall of mechanistic assumptions about human nature. But this did not mean that probability itself feel into disrepute.

A turning point

Ochigame writes:

"At the end of the Napoleonic era, the sudden publication of large amounts of printed numbers, especially of crimes, suicides, and other phenomena of deviancy, revealed what philosopher Ian Hacking calls"law-like statistical regularities in large populations." These regularities in averages and dispersions gradually undermined previously-dominant beliefs in determinism"

Risk and actuarialism (19th and 20th century)

Life insurances in the 19th century relied on risk assessment. They were soon charged with racial bias:

"Risk classification soon surfaced controversies over racial discrimination: in 1881, life insurance corporations started to charge differential rates on the basis of race ... When civil rights activists challenged this

Here is Pascal in a letter to Fermat (1654): This is the way I go about it to know the value of each of the shares when two gamblers play, for example, in three throws, and when each has put 32 pistoles at stake. Let us suppose that the first of them has two (points) and the other one. They now play one throw of which the chances are such that if the first wins, he will win the entire wager that is at stake, that is to say 64 pistoles. If the other wins, they will be two to two and in consequence, if they wish to separate, it follows that each will take back his wager that is to say 32 pistoles. Consider then, Monsieur, that if the first wins, 64 will belong to him. If he loses, 32 will belong to him. Then if they do not wish to play this point, and separate without doing it, the first should say–I am sure of 32 pistoles, for even a loss gives them to me. As for the 32 others, perhaps I will have them and perhaps you will have them, the risk is equal. Therefore let us divide the 32 pistoles in half, and give me the 32 of which I am certain besides. He will then have 48 pistoles and the other will have 16.

From Hacking (1990), The Taming of Chance, Cambridge UP: The avalanche of numbers, the erosion of determinism, and the invention of normalcy are embedded in the grander topics of the Industrial Revolution.

policy, the corporations claimed differences in average mortality rates across races as justification."

Two views about "fair" pricing emerged, one closely tied to the accuracy of the prediction based on current available data, and another more aspirational:

> "...in 1884, Massachusetts state representative Julius C. Chappelle—an African American man born in antebellum South Carolina—challenged the fairness of the policy and proposed a bill to forbid it. The bill's opponents invoked statistics of deaths, but Chappelle and his allies reframed the issue in terms of the future prospects of African Americans, emphasizing their potential for achieving equality."

Can this disagreement about fair pricing of insurance policy be instructive to navigate current debates about algorithmic fairness?

Criminal justice, lending, education (1960s-80s)

Actuarial risk assessment became widespread in criminal justice (parole, sentencing) in 1980s, although the idea had been around for a while, at least as early as 1920s.[4] Credit scoring started to be widely adopted by financial institutions in 1960s. Standardized testing became widely adopted about the same time.[5]

[4] See Harcourt (2006), Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age, Chicago UP

[5] See e.g. Elementary and Secondary Education Act (1965) as well debates about the fairness of standardized tests in Hutchinson and Mitchell (2019), Translation Tutorial: A History of Quantitative Fairness in Testing

Discrimination / actuarial fairness

Here is how the concept of "actuarial fairness" came about, in response to charges of racial bias. Ochigame notes:

> "In the 1970s, at the height of controversies surrounding redlining, U.S. civil rights and feminist activists argued that risk classification in the pricing of insurance was unfair and discriminatory ... the insurance industry disseminated the concept of "actuarial fairness": the idea that each person should pay for her own risk."

What conception of algorithmic fairness is Actuarial Fairness more closely related to among the ones we examined in earlier classes?

Today

Ochigame describes the difference between the debates about actuarial fairness in the 1970s and today's debates about algorithmic fairness, as follows:

> "In the 1970s, proponents of "actuarial fairness" simply equated it with predictive accuracy; they posed fairness as equivalent to the optimization of utility in risk classification. Today, proponents of "algorithmic fairness" tend to define fairness and utility as distinct, often competing, considerations. Fairness is generally considered a complementary consideration or constraint to the optimization of utility, and proponents often speak of "trade-offs" between fairness and utility. This distinction responds to a widespread recognition that the conventional optimization of utility in actuarial systems—typically the maximization of profit or the minimization of risk—can be inherently unfair or discriminatory."