New York University

Big Data: Critical Perspectives

Project Proposal

Predictive Policing - Opaqueness, Impact and Legal Standards

Author:
Kevin Gallagher

Professor:
Helen Nissenbaum

May 5, 2016

Contents

1 Introduction

One Friday morning in July of 2011, two women were arrested in Santa Cruz, California after being caught peering into cars parked in a local parking garage. In addition to the suspicious nature of their activities, one of the women arrested had outstanding warrants, and the other was in possession of drugs. Though arrests of this nature are rarely reported on the news, this particular arrest was an exception due to its peculiar circumstances. These arrests were made after officers were dispatched beause a computer algorithm predicted the area as likely to be effected be crime. [13]

This technology, called Predictive Policing, is definied as the application of analytical techniques, particularly quantitative techniques, to identify promising targets for police intervention and prevent or solve crime. [18] As such, predictive policing can be seen as a branch of "Big Data" or "Data Science," the collection and analysis of large quantities of data to form predictive tools. The purpose of this paper is to study the effect of predictive policing and big data on due process in libral democracies.

The remainer of this paper will be organized as follows, Section 1.1 will define predictive policing. Section 1.2 will discuss the role of due process in liberal democracies and why it is important. In Section 2 will review the current relevant literature and address the relevant questions posed in the introduction. Finally, in Section 3 we conclude.

1.1 Predictive Policing: Definition and Taxonomy

As mentioned in Section 1, predictive policing is by Perry et al. to be the application of analytical techniques, particularly quantitative techniques, to identify promising targets for police intervention and prevent or solve crime. In the book *Predictive Policing: The role of crime forcasting in law enforcement operations*, Perry et al. present a taxonomy of predictive policing methods. This taxonomy breaks predictive policing into four categories: [18]

1. Methods for predicting future crimes
2. Methods for predicting future offenders
3. Methods for predicting perpetrator's identities
4. Methods for predicting victims of crimes

Category one, methods for predicting future crimes, focuses on predicting the times and places in which crimes are anticipated to occur. Category two, methods for predicting future offenders, focuses on predicting individuals and groups that are "at risk of offending in the future." Category three, methods for predicting perpetrator's identities, focuses on identifying individuals who have committed past crimes. Lastly category four, methods for predicting victims of crimes, focuses on predicting individuals and groups that are "likely to become victims of crime."

1.2 Due Process

Due Process is defined by the Oxford English Dictionary (US Edition) to be "[f]air treatment through the normal judicial system, especially as a citizens entitlement." In modern liberal democracies (such as the United States), due process serves an important role; Due process protects citizens from unfair treatment under the law.

Liberal democracies have at the core of their philosophies the idea that every individual has equal protection of human rights, civil rights, civil liberties, and other political freedoms. As such, due process of law is necessary to protect the core tenants of this philosophy, and to protect the rights of citizens from governments that seek to violate the rights of its citizens. It ensures that the weak, poor and powerless have the same legal protections as the strong, wealthy and powerful, creating equal treatment under the law.

In the United States, due process of law is seperated into two groups, proceedural due process and substantive due process. Procedural due process focuses on rights that apply to government proceedings that may result in the denial of an individual's right to life, liberty or property [4]. As part of procedural due process, it is well accepted that the following rights apply: [11]

1. An unbiased tribunal.
2. Notice of the proposed action and the grounds asserted for it.
3. Opportunity to present reasons why the proposed action should not be taken.
4. The right to present evidence, including the right to call witnesses.
5. The right to know opposing evidence.
6. The right to cross-examine adverse witnesses.
7. A decision based exclusively on the evidence presented.
8. Opportunity to be represented by counsel.
9. Requirement that the tribunal prepare a record of the evidence presented.
10. Requirement that the tribunal prepare written findings of fact and reasons for its decision.

Substantive due process is "[a] doctrine holding that the 5th and 14th Amendments require all governmental intrusions into fundamental rights and liberties be fair and reasonable and in furtherance of a legitimate governmental interest." [6] One part of substantive due process is the application of the Bill of Rights, which contains as part of it the Fourth Amendment, or the protection against unreasonable searches and seizures by the government.

In modern policing, two legal standards keep the actions of law enforcement agencies in check: reasonable suspicion and probable cause. Reasonable suspicion is the standard that must be met in order for law enforcement officials to justify brief stops and detentions, but not full searches. [5] Reasonable suspicion must be based on "specific and articulable facts... taken together with rational inferences from those facts," which must be associated with a particular individual. [7] In order to receive a warrant for a search or to make an arrest, a law enforcement agent must meet the standard of probable cause. Probable cause requires a "reasonable basis for believing that a crime may have been committed (for an arrest) or when evidence of the crime is present in the place to be searched (for a search)." [3] These standards keep the police from arbitrarily targeting individuals for investigation of crimes, instead requiring reasons for the infringement on their rights.

Until the era of big data began, actions taken by police were based on specific facts and tips provided by the circumstances of the crime or by witnesses of the crime the law enforcement agents were investigating. Because of this, police were able to articulate specific facts and circumstances that lead to the reasonable suspicion for brief stops or detentions and probable cause in order to obtain a warrant. However, the current use of big data challenges these well established standards and overall due process. This paper will attempt to address how big data and predictive policing change due process and the standards of reasonable suspicion and probable cause.

2 Due Process and Predictive Policing

As previously mentioned, predictive policing is an extention of big data into the field of policing and law enforcement. As such, it inherits many of the problems that big data exhitbits, such as a lack of transparency and issues of bias. In addition to these issues, the application of big data to the realm of policing creates whole new questions, such as the question of how effectiveness is measured, and what role probability and statistics plays in the standards of due process, such as reasonable suspicion and probable cause. In addition, mistakes in law enforcement action taken in response to the predictions made carry potentially high costs to law enforcement and private citizens alike. In this section, these issues will be explored and addressed, and will be tied back to the cases mentioned in Section 1.

2.1 Transparency

One historical problem that big data struggles with is the problem of transparency. In many realms in which big data is used, such as credit reporting, advertising, data brokers, employment, housing and others, decisions made based on prediction algorithms affect the lives of countless individuals. However, due to claims of intellectual property and trade

secrests individuals have no ability to peer into the machines that dictate the directions of certain aspects of their lives. Worse, even when individuals are not blocked by legal and business challenges, often times peering into an algorithm does nothing to help an individual to understand why a particular classification was made. Part of this lies in the fact that the prediction algorithms are hard to understand by a layperson, but part of it also lies in the technology itself.

One large part of the big data aparatus is the application of machine learning, or an area of statistics and computer science in which computers determine correlations between input features and predicted labels. One of the great strengths of machine learning is to find correlations that human experts cannot. This is done by the pure computational power that a machine possesses and people do not. However, this strength is also a great weakness. Due to the fact that a machine can predict based on input patterns that may escape a human expert, a large amount of opacity enters the system, even when the engineers and statisticians themselves are involved. Even if a domain expert is able to determine why a prediction algorithm output a certain prediction, this solution is not scalable. The entire purpose of big data is to deal with quantities of data that are impossible for human teams to sift through, making it difficult for companies and agencies that employ big data techniquest to verify the accuracy and fairness of a prediction or classification.

This opaquness poses a unique challenge for law enforcement purposes. As seen in the breakdown of procedural due process mentioned above, the justice system of liberal democracies requires a large amount of transparency. Individuals who have had an action taken against them must be able to defend themselves, and as such they need to know the reasons that lead to a particular action being taken. As predictive policing features more prominently in our justice system, it will begin to affect more and more police actions. As such, the police will need to be able to justify their actions and provide individuals with the evidence that was collected that allowed the police to do so. However, due to the opaque nature of big data such disclosures will be exceedingly difficult, if not impossible. This leads one to ask how much, if at all, should opaque big data classifications be used to guide police actions? By the breakdown of procedural due process given above, it would seem that the lack of transparency in predictive policing methods may threaten the liberties of an accused individual, and may even place the justification of law enforcement's actions into question.

2.2 Biases

Another historical problem of big data is the introduction of biases. Though many proponents of big data claim that the introduction of statistical algorithms leaves little to no room for human bias, this is simply not the case.[14] As mentioned previously, machine learning, which finds the correlations between inputs and classification labels,

plays a large role in the aparatus of big data. Machine learning is split into the following phases:

1. Data collection.
2. Feature creation.
3. Feature selection and/or extraction.
4. Data labeling.
5. Algorithm training.
6. Algorithm testing and validation.
7. Production and classification.

Though many proponents claim that raw data is discovered and fed into the algorithm, the steps above show that data goes through a lot of processing and transformation before the classification step is reached. As shown in step 2, human intervention is usually need to decide what aspects of the data become features. Though domain experts are usually consulted in this step, this interaction has the capability to introduce a large amount of human bias. For example, in Perry's report on predictive policing [18], the idea of using quantitative measures such as the LSI-R in predictive policing algorithms is discussed. Though this could be seen as a positive step to introduce quantitative measures into the predictive policing, there are examples of bias even in the LSI-R. For example, in the South Dekota Department of Human Services LSI-R training slides, one of the included predictors of future crime is "belonging to some minority groups." [21] If the LSI-R scoring is based on racial bias, and the LSI-R numerical output is included as a feature in predictive policing, the output of the policing algorithm is at least in part based on the race of the individual, which is commonly accepted as wrong and unjust.

In addition to being introduced at the feature creation and selection/extraciton phases, bias can also be introduced at the label phase. As shown in the series of steps above, machine learning algorithms must be trained. In many cases, supervised training algorithms are used, meaning that many instances of data and the "correct" output label are fed to the learning algorithm and a classification model is built based on this data. These labels, however, are not "discovered." They are frequently assigned by domain experts and heuristic methods. This is another area in which human bias can find its way into big data.

NOTE: In the remainder of this section I will tie in these issues with predictive policing and due process.

2.3 Future Discussion Areas

I ran out of time and space while writing this paper. In addition to the issues mentioned above, I plan on touching on the problems of effectiveness and accuracy measurement, the role of probability in due process and the introduction of human error to policing practices.

3 Conclusion

This paper has analyzed and discussed the application of predictive policing and its effect on the realm of due process. As part of this, it has discussed what predictive policing is and how it is used, what due process is and why it is important in liberal democracies, and finally the drawbacks of big data and predictive policing, and how these drawbacks effect due process.

References

[1] Definition of due process in english:.

[2] Due process.

[3] Probable cause.

[4] Procedural due process.

[5] Reasonable suspicion.

[6] Substantive due process.

[7] Terry v. ohio, 1968.

[8] L. Amoore. Data derivatives on the emergence of a security risk calculus for our times. *Theory, Culture & Society*, 28(6):24–43, 2011.

[9] D. K. Citron and F. A. Pasquale. The scored society: due process for automated predictions. *Washington Law Review*, 89, 2014.

[10] A. G. Ferguson. Predictive policing and reasonable suspicion. 2012.

[11] H. J. Friendly. Some kind of hearing. *University of Pennsylvania Law Review*, 123(6):1267–1317, 1975.

[12] L. Gitelman. *Raw data is an oxymoron*. MIT Press, 2013.

[13] E. Goode. Sending the police before theres a crime, Aug 2011.

[14] M. Hardt. How big data is unfair: Understanding sources of unfairness in data driven decision making, 2014.

[15] M. Hildebrandt and K. De Vries. *Privacy, due process and the computational turn: the philosophy of law meets the philosophy of technology*. Routledge, 2013.

[16] E. E. Joh. Policing by numbers: Big data and the fourth amendment. 2014.

[17] W. J. Perry and C. M. Vest. Protecting individual privacy in the struggle against terrorists a framework for program assessment: Executive summary. *Journal of Privacy and Confidentiality*, 2(1):5, 2010.

[18] W. L. Perry, B. McInnis, C. C. Price, S. C. Smith, and J. S. Hollywood. *Predictive policing: The role of crime forecasting in law enforcement operations*. Rand Corporation, 2013.

[19] D. Robinson, H. Yu, and A. Rieke. Civil rights, big data, and our algorithmic future. *Robinson & Yu*, 2014.

[20] J. Scahill and G. Greenwald. The nsa's secret role in the u.s. assassination program, Feb 2014.

[21] D. o. R. S. South Dekota Department of Human Services. Lsi-r: Level of service inventory-revised introductory training course, Apr 2012.

NOTES

- Metrics - how to measure?
- Schultz and Crawford - definitions
- Why due process is important in a liberal democracy - Look at this
- Incursions of data science into the methodologies and procedures and how they may challenge the meaning/foundation of these.
- Need to fix and find citations