Trust and Ethics in Algorithmic Crime Analysis

Shion Guha
Department of Mathematics,
Statistics and Computer
Science, Marquette University
Milwaukee WI 53201
shion.guha@marquette.edu

ABSTRACT

We face applications and outcomes of algorithms everyday. There has been intense criticism [7] about the application and appropriateness of these algorithms in various spheres of life. In addition, sometimes this has also led to a chilling effect where researchers, developers and data scientists have been loath to study certain phenomena [2] due to its potential, future controversy. Crime analysis is no exception to this effect. In this whitepaper, I summarize the main literature in algorithmic crime analysis and outline specific research directions that make algorithmic ethics within crime analysis a unique, interesting and unsolved problem to study for designing trustworthy algorithms.

ALGORITHMIC CRIME ANALYSIS

Algorithms already govern, curate, manage and inform [13] many of our everyday life decisions. Automated decision-making through algorithms has significant impacts on the everyday lives of people. This is especially true in the area of crime analysis or predictive policing.

Police departments mainly analyze crime through crime mapping, a process that uncovers high density crime areas[8]. Crime mapping is used to make decisions on how to distribute law enforcement across an area, seeking maximum efficiency with commonly budget-strapped resources[18]. Past crime mapping tools were rudimentary and provided little more than crime points on a map, but with the access to technology modern police departments have, these datasets have begun to get bigger and encompassing more variables[3]. Spatial crime analysis tools are not new in law enforcement, but previously these tools were not intended for widespread, general use[5].

Crime data mining and crime data fusion have opened up many avenues for crime analysis[5]. Data now streams from a variety of sources providing a variety of data types that analysts need to stitch together, including unstructured data such as video, text and audio[6]. Police departments are now combining their databases with geographical information systems (GIS) in order to identify spatial trends. One example of the type of tool provided to law enforcement crime analysts is ReCAP, an analysis tool used to understand and predict crime activity in an area[5]. ReCAP attempts to implement spatial analysis tools into traditional crime analysis frameworks. This software allows law enforcement agencies across the country to plug local data into their models and begin analyzing spatial and temporal results.

Clustering methods are one of the most widely used approaches[1, 6] in algorithmic crime analysis. While there are many approaches, the density and graph based approaches [1, 12] and its variants are extremely popular for easy computation, compatibility with spatial data and perceived ease of understanding to the layperson.

ETHICS AND ALGORITHMS

Many scholars [10, 17, 19, 16] have pointed out that attempts to evaluate the ethical implications of algorithms are frequently hampered by black boxes that prevent a full "code audit", i.e. an investigation of the source code for potentially unethical behavior [19]. Sandvig et al. [19] laid out several other methods to audit algorithms despite and perhaps in active defiance of black boxes. While algorithmic audits can reveal important behavior, they have their own ethical concerns. Krafft, Macy, and Pentland [15], point out that tactics employing "bots" or automation themselves have ethical risks, as they frequently run afoul of terms of service agreements and depend on deception of normal users. Mainstream ethics research [11] focuses on the interrelationship between people and trust as necessary for developing ethical frameworks.

Identifying that bias exists is only half of the task; however, as discovering *why* it exists is equally important. Significant work [4, 16, 10, 14] has already been done to identify various ways in which algorithms can produce bias. Bias can just as easily be introduced unintentionally through transformational effects—such as the order of search listings—or emergent effects that occur when underlying assumptions change over the long term [10, 16]. Since emergent bias develops post hoc and in relation to unanticipated interactions, it is often difficult to isolate the source, either in the algorithm itself or the data [4]. In order to address these concerns, the way the data were constructed and the design of the algorithm must be considered at the outset.

Many scholars have criticized the idea that datasets are objective and comprehensive [4, 9]. Any dataset is given inher-

Every submission will be assigned their own unique DOI string to be included here.

ent biases by the way in which it is constructed [17, 9]. In essence, the act of collecting data is also an act of design. Decisions must be made about what data are relevant to collect, what constitutes anomalous data, and what methods are available to make the measurements [9].

A RESEARCH AGENDA

Summarizing the debate in this area, I propose the following directions for research. This is also part of my current research program and part of my attending this workshop is to collaborate with potential, similar minded researchers.

1. We must advocate for a human-centered and socio-technical understanding of algorithmic analysis i.e. it is not only important how the technical machinations of algorithms operate but also *how* they are imagined and applied by their human operators.

2. There is a rich body of ethics research literature [11] that is not often tapped in the design of algorithms for crime analysis. We must investigate these prevalent ethical frameworks that are often derived from fundamental research in philosophy and theology and overlay them when inspecting or designing algorithms.

3. Deconstructing popularly applied crime analysis algorithms to understand the difference between statistically appropriate norms and those that are actually used can be an important avenue for exploration. At Marquette, we have embarked upon such a project to understand biases in crime analysis in the city of Milwaukee.[1]

4. The final chapter of such a research agenda must constitute the design of human-centered and ethically minded algorithms that analyze crime by minimizing harm and discrimination. In addition, they must foster trust within the community that is being analyzed as well by the analysts themselves. This is broader conversation involving all stakeholders as well as implications for education and training in higher education.

REFERENCES

1. Agarwal, J., Cse, M., Nagpal, R., and Sehgal, R. *Crime Analysis using K-Means Clustering*.
2. Amy Wesolowski, Caroline O. Buckee, L. B. E. W. X. L. A. J. T. Commentary: Containing the ebola outbreak - the potential and challenge of mobile network data. *PLoS Currents 6* (sep 2014).
3. Boba, R. Introductory guide to crime analysis and mapping. *Community Oriented Policing Services. USA* (2001).
4. Boyd, D., and Crawford, K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society 15*, 5 (2012), 662–679.
5. Brown, D. E. The regional crime analysis program (recap): a framework for mining data to catch criminals. In *Systems, Man, and Cybernetics, 1998. 1998 IEEE International Conference on*, vol. 3, IEEE (1998), 2848–2853.
6. Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., and Chau, M. Crime data mining: a general framework and some examples. *computer 37*, 4 (2004), 50–56.
7. danah boyd, K. C. Critical questions for big data. *Information, Communication & Society 15*, 5 (jun 2012), 662–679.
8. Eck, J., Chainey, S., Cameron, J., and Wilson, R. Mapping crime: Understanding hotspots.
9. Feinberg, M. A design perspective on data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ACM (2017), 2952–2963.
10. Friedman, B., and Nissenbaum, H. Bias in computer systems. *ACM Transactions on Information Systems (TOIS) 14*, 3 (1996), 330–347.
11. HALL, M. A. The importance of trust for ethics, law, and public policy. *Cambridge Quarterly of Healthcare Ethics 14*, 2 (2005), 156167.
12. Hartigan, J. A., and Wong, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 28*, 1 (1979), 100–108.
13. Kate Crawford, J. S. Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review 55* (jun 2014), 93.
14. Kraemer, F., Van Overveld, K., and Peterson, M. Is there an ethics of algorithms? *Ethics and Information Technology 13*, 3 (2011), 251–260.
15. Krafft, P. M., Macy, M., and Pentland, A. Bots as virtual confederates: Design and ethics. *arXiv preprint arXiv:1611.00447* (2016).
16. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., and Floridi, L. The ethics of algorithms: Mapping the debate. *Big Data & Society 3*, 2 (2016), 2053951716679679.
17. O'Neil, C. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Books, 2016.
18. Roth, R. E., Ross, K. S., Finch, B. G., Luo, W., and MacEachren, A. M. A user-centered approach for designing and developing spatiotemporal crime analysis tools. In *Proceedings of GIScience*, vol. 15 (2010).
19. Sandvig, C., Hamilton, K., Karahalios, K., and Langbort, C. Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry* (2014).

[1] https://marquettecomputationalsocialscience.github.io/clusteredcrimemaps/