In: Emre Bayamlıoğlu, Irina Baraluic, Liisa Janssens and Mireille Hildebrandt (eds), BEING PROFILED:COGITAS ERGO SUM. 10 Years of Profiling the European Citizen, 2018: Amsterdam University Press, 66-70. DOI 10.5117/9789463722124/CH11.

The geometric rationality of innocence in algorithmic decisions

*Tobias Blanke**

Abstract

In this provocation, I would like to develop what I call the geometric rationality of algorithmic decisions, which measures social relations using the distance of data points in abstract geometric spaces. This analysis follows on to the work by Claudia Aradau and myself, where we introduce the concept of 'in-betweenness' in abstract information spaces as a foundation of algorithmic prediction. In this paper, I elaborate how algorithmic innocence, i.e. innocence before an algorithm, is (pre-)decided by a geometric rationality of algorithms. I show how (non)-innocent subjects are created proactively and to be acted upon pre-emptively by algorithmic manipulation of an abstract feature space.

Keywords: algorithmic decisions, geometric rationality of algorithms, presumption of innocence

Introduction

In this provocation, I would like to develop what I call the geometric rationality of algorithmic decisions, which measures social relations using the distance of data points in abstract geometric spaces. This analysis follows on to the work by Claudia Aradau and myself (Aradau and Blanke 2017), where we introduce the concept of 'in-betweenness' in abstract information spaces as a foundation of algorithmic prediction. In this paper, I elaborate how algorithmic innocence, i.e. innocence before an algorithm, is (pre-)decided by a geometric rationality of algorithms. I show how (non)-innocent subjects are created proactively and to be acted upon pre-emptively by algorithmic manipulation of an abstract feature space.

Abstract geometric information spaces

Computational decision-making techniques generally operate with the spatial metaphor of abstract geometric spaces. AI has set off with the idea of abstract information spaces, as an MIT website from the 1990s reveals: 'An information space is a type of information design in which representations of information objects are situated in a principled space. In a principled space location and direction have meaning, so that mapping and navigation become possible' (MIT Artificial Intelligence Laboratory 1998). In the world of AI, we are interested in meaningful information spaces that do not count all available information but only information, which can 'feature' in the calculation of a problem. These are the features describing to algorithms us and all other things in the world. Together these features span an abstract information space using 'vectors' of features. For instance, for people we might think about gender, height, weight and age as features, each of which is a dimension of the problem to be modelled. In this case, we have a four-dimensional feature space. Machine learning techniques that have propagated across different fields can tell us how 'people are materialised as a bundle of features' (Mackenzie 2013).

Decision-making algorithms plot data as points/dots in feature spaces, which thus become a geometrical representation of all the data available to them. Each dot in this space is defined by how much abstract space is in-between it and the other dots in the same space or how distant they are from each other. For practitioners who operate decision-making algorithms, '[d]ata is in some feature space where a notion of "distance" makes sense' (Schutt and O'Neil 2013, 81). In principle, there is no limit to the number of features that can be used to build such an artificial space. Feature spaces can have hundreds, thousands or hundreds of thousands of features/dimensions, depending on how much a computer can process. Machine learning algorithms manipulate this feature space in order to create labels for each example that can already be found in the feature space or that might be found in the future in the feature space. They 'partition' the feature space into zones of comparable features. Each data points in these zones is labelled the same way. Labelling is the materialisation of decision-making by machine learning.

It is this feature space, which drives the (big) data needs in machine learning: 'How many data points would you need to maintain the same minimum distance to the nearest point as you increase the number of inputs of the data? As the number of inputs increases, the number of data points needed to fill the space comparably increases exponentially' (Abbott 2014, 153), because the number of inputs corresponds to the number of features. The more feature dimensions an abstract information space has the more space there is to fill in this space. The big data drive is a result of the attempt to fill the feature space. In a famous paper Banko and Brill (2001) set the agenda for the big data hype and its rationale. They demonstrated that digital reasoning of all kinds gets more accurate by throwing more data at it, as the feature space gets filled with data points.

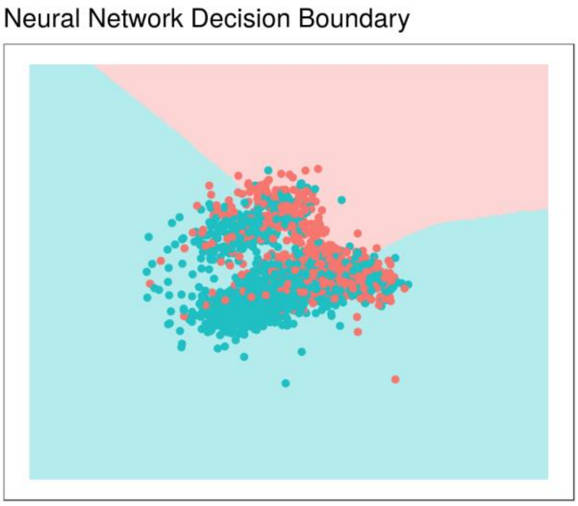

Figure 1. Decision Boundaries

Geometric decision-making

As analysed by Anrig et al. (2008), there are many examples of decision-making algorithms and how they work the feature space. They all 'partition' the space into zones to generate meaning for all data points in the space. Partitioning is geometric decision-making by algorithms. 'Decision trees', e.g., partition the space into decisions made over a subset of features. This way, they collect all points in the space that are, for instance, of gender female, taller than 1.70m, weigh more than 65 kg, etc. 'Clustering' using nearest neighbours partition the feature space into zones of points that share similar feature dimensions and are defined by their border to other zones in the high-dimensional space. 'Regression analysis' as well as '(deep) neural networks' can learn more complex boundaries between zones of similar features. 'Often, different methods are used and the quality of their results compared in order to select the "best one".' (Anrig et al. 2008, 79).

We generated the worked example in Figure 1 to demonstrate a feature space with data points in two classes (red and blue dots). The space is partitioned into two zones by a complex non-linear boundary generated by a neural network. While complex decision boundaries can generate highly accurate zones and partitions, they are known to be difficult to understand. Neural networks are unintelligible compared to the example of decision trees. Cathy O'Neil likens it to the godlike unintelligibility: 'Like gods, these mathematical models were opaque, their working invisible to all but the highest priests in their domain: mathematicians and computer scientists.' (O'Neil 2016, 7). As these models and algorithms are integrated within complex artificial systems, they risk becoming 'black boxes', unintelligible even to the 'high priests' of the digital world. 'In the era of machine intelligence', O'Neil cautions, 'most of the variables will remain a mystery. (…). No one will understand their logic or be able to explain it' (O'Neil 2016, 157).

Such 'unintelligibility' prevents human observers from understanding how well algorithmic geometric rationality works. We cannot be reassured by following the rules of evaluation of the 'high priests' either, as they are made to make the algorithms perform computationally and not socially. This is expressed in what counts for right and wrong decisions in the feature space. Errors and error rates are key sites of the transformation of knowledge, but also sites of controversy with regards to innocence and non-innocence, as decisions by algorithms are contested. In 2018, e.g., the media reported that the company ASI Data Science had developed an extremism blocking tool with government funding of £600,000, which could automatically detect 94% of Isis propaganda with 99.99% accuracy (Lomas 2018). As reported, this is at best confusing information, as nothing else is known about the experiments that led to these error rates. 94% will still be concerning for the security analyst dealing with a system like Facebook and billions of new items a day. 6% missed content can then mean 1,000s of items. We do not know how the accuracy is measured either but 'the government says' for 'one million "randomly selected videos" only 50 of them would require "additional human review".' (Lomas 2018). This means in our Facebook example a block of '50,000 pieces of content daily'. Finally, the tool is also single-minded and does not partition the feature space for all terrorist content but only for ISIS data of a particular time. Complex digital reasoning tends to be single-minded in this way because each feature space is a unique geometry. High accuracy figures for decision-making algorithms should never be enough to reassure us that these algorithms are correct and make wise pre-emptive decisions.

The reader might have also noticed a red dot in the Figure 1 far away in the bottom-right non-innocence blue corner. This is called an outlier and is as such suspicious/interesting, because we are not just innocent in the feature space by association with other innocent dots close by but also by dissociation with other dots. 'The outliers are determined from the "closeness" (…) using some suitable distance metric.' (Hodge and Austin 2004, 91). We investigate the relations of digital selves and others implied by outlier detection in (Aradau and Blanke, 2018), where we present the real-life security impact outlier methods have. Outliers predetermine innocence in feature spaces as much as closeness does.

The data scientist McCue specialises in outlier-detections in predictive policing. She gives an example from security analytics that demonstrates the power of outlier detection using a cluster analysis (McCue 2014, 102). They monitored conference calls to find clusters of numbers based on geographies and regions. '[I]t would not have been possible to analyze these data without the application of data mining.' (McCue 2014, 104). The resulting two-dimensional feature space exhibits three clusters including one outlying cluster in the bottom-left corner. Features included 'the conference IDs (a unique number assigned by the conference call company), the participants' telephone numbers, the duration of the calls, and the dates' (McCue 2014, 104). '[T]hree groups or clusters of similar calls were identified based on the day of the month that the conference occurred and the number of participants involved in a particular call.' (McCue 2014, 106). The outlier cluster correctly identified a professional criminal network. For McCue, this approach has various advantages. Firstly, one can literally 'see' in the feature space why one cluster is different from the others and an outlier. Secondly, the information used to cluster the participants is not necessarily based on detailed information of individuals in the cluster as it summarises their existence into features, and surveillance can take place without much attention to privacy limitations. Finally, the clusters that are not outliers build a dynamic, algorithmic model of normality. Non-suspicion or innocence is determined by declaring some cluster to be not outliers, while anomalies are outside any cluster. The geometrical distance in the feature space makes outlier dots stand out as outliers.

Conclusion

This short provocation presented ideas on how innocence through algorithms is pre-determined by the position of dots in an abstract feature space and an underlying geometric rationality of distances between dots. We examined the foundations of this geometric rationality, its need for more and more data as well as issues preventing reasoning about errors critically. To be finally counted as innocent, a dot should be close enough to the innocent dots in the abstract space and also not too close or too far away in order not to be suspicious again.

* Tobias Blanke is a Professor of Social and Cultural Informatics in the Department of Digital Humanities and current Head of Department at King's College London .

References

Abbott, Dean. 2014. Applied predictive analytics: Principles and techniques for the professional data analyst. New Jersey: John Wiley & Sons.

Anrig, Bernhard, Will Browne, and Mark Gasson. 2008. “The Role of Algorithms in Profiling.” In Profiling the European Citizen: Cross-disciplinary Perspectives, edited by Mireille Hildebrandt and Serge Gutwirth, 65-87. Dordrecht: Springer.
Aradau, Claudia, and Tobias Blanke. 2017. “Politics of prediction: Security and the time/space of governmentality in the age of big data.” European Journal of Social Theory 20(3): 373-91.
Aradau, Claudia, and Tobias Blanke. 2018. “Governing others: Anomaly and the algorithmic subject of security.” European Journal of International Security 3(1): 1-21.
Banko, Michele, and Eric Brill. 2001. “Scaling to very very large corpora for natural language disambiguation.” Proceedings of the 39th annual meeting on association for computational linguistics.
Hodge, Victoria J., and Jim Austin. 2004. “A survey of outlier detection methodologies.” Artificial Intelligence Review 22(2): 85-126.
Lomas, Natasha. 2018. “UK outs extremism blocking tool and could force tech firms to use it.” https://techcrunch.com/2018/02/13/uk-outs-extremism-blocking-tool-and-could-force-tech-firms-to-use-it/, accessed at 31/8/2018.
Mackenzie, Adrian. 2013. “Programming subjects in the regime of anticipation: Software studies and subjectivity.” Subjectivity 6(4): 391-405.
McCue, Colleen. 2014. Data mining and predictive analysis: Intelligence gathering and crime analysis. Oxford: Butterworth-Heinemann.
MIT Artificial Intelligence Laboratory. 1998. The JAIR Information Space. http://www.ai.mit.edu/projects/infoarch/jair/jair-space.html, accessed at 31/8/2018.
O'Neil, Cathy. 2016. Weapons of math destruction: How big data increases inequality and threatens democracy. New York: Crown.
Schutt, Rachel, and Cathy O'Neil (2013). Doing data science: Straight talk from the frontline. New York: O'Reilly Media, Inc.