The Debate about "algorithms"

James H. Davenport

Departments of Mathematical Sciences and Computer Science
University of Bath, Bath BA2 7AY

June 25, 2017

Algorithms, in both the broad and narrow senses of the word, affect practically every aspect of our daily lives. Algorithms decide which advertisements are shown on media such as Youtube (and sometimes get it wrong, causing producers of material to receive income from advertisements appearin with their material, when the advertisers would never consciously approve that material [Gua17], often causing public anger [Bir17, for example]), what prices we are quoted for airline flights [MPR09], and, in the USA, whether people get bail [ALMK16].

The House of Commons Select Committee on Science & Technology launched an inquiry into "algorithms in decision-making" on 28 February 2017. Both the Institute for Mathematics and its Applications and BCS, the Chartered Institute for IT, submitted evidence, with the writing process coordinated by the author. The vote on April 19th for an early General Election terminated the inquiry: all submissions can be viewed at `http://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/inquiries/parliament-2015/inquiry9/publications/`. The aim of this article is to draw together the evidence and general points made in these submissions.

1 Background on Algorithms

The word algorithm has, particularly in American but creeping into British English, acquired a wider connotation than the OED definition: "a precisely defined set of mathematical or logical operations for the performance of a particular task". For example, Merriam–Webster has "set of rules a machine (and especially a computer) follows to achieve a particular goal". Hence we drew various distinctions to help the Committee, and the reader.

1.1 Proclaimed or Inferred (from data)

A classic example of a proclaimed algorithm is Income Tax: if (for a given set of circumstances) your net income is X, you pay $f(X)$ in income tax, where f is

laid down in the Finance Act. A classic example of an inferred algorithm is the "wind chill", which is an algorithm with a precise mathematical statement, but was experimentally determined by the multinational JAG/TI group of scientists. Much of the current trend in "machine learning" is for a computer, sometimes using a well-understood meta-algorithm to produce an "algorithm" without a precise mathematical statement, but based on a large amount of experimental data, generally referred to as "training data". Most of these algorithms operate on data, often training or other background data (database) as well as foreground data (the question). They are at most as good as the background data they operate with, and bad data can lead to disastrously wrong results, as when the London Ambulance Service did not know where the Velodrome was. [Gua16]

1.2   Published or secretly understood, or not understood

Income tax, and wind chill, are published algorithms. Many companies operating in regulated industries, generally in finance, have secret algorithms for credit scoring, loan approval etc. These algorithms are part of their competitive edge. Yet, they need to be able to explain them to regulators, and to justify decisions if required. An example might be a car insurance company, whose algorithm might include the step "To the base premium, we add a sum depending on the insured's occupation, from this table", where the table is computed based on past claims data. The table, the precise definition of occupations, and indeed whether it's adding a sum, multiplying by a ratio, or both, are part of the insurer's trade secrets. This example in insurance long predates the use of computers, never mind machine learning algorithms, but the advent of technology has permitted much more analysis and the use of many more factors. Equally, it has enabled the use of much more precise calculations, rather than "gut feel" to populate these tables. Conversely, many of the algorithms produced by machine learning, notably those based on "deep learning", are not understood at all.

1.3   Advisory or Determinative

The income tax algorithm determines the amount payable: it is not a suggestion to the tax inspector how much to charge. University degree classifications have become more determinative over the years. A recent adoption of a determinative algorithm is the Duckworth-Lewis algorithm [DL98] in cricket. The algorithm is public, and understood by experts, but not by the general public, who just accept it.

Advisory algorithms produce a piece of advice to a human being who makes an ultimate decision. This may consist of evaluating several different scenarios, or may just be a simple "answer". These answers may or may not (in practice, far too often do not) have some measure of confidence attached. Of course, an algorithm may in principle only be advisory but the human beings using it may in practice just rubber-stamp its "advice", so in practice it's determinative.

In the US case of Paul Zilly [ALMK16], defence and prosecution had agreed a plea bargain of a year in prison, but the judge looked at a recidivism score produced by a proprietary (and at least secret, probably not understood), over-turned the plea deal, and imposed a two-year sentence. In theory that was an advisory algorithm, but in practice it was being used to overrule the agreement which would have been rubber-stamped by the judge.

The GDPR (Article 22.1) applies to a "decision based solely on automated processing", i.e. a determinative algorithm. The relevant Recital (71) is less clear, and one might expect a lot of litigation where humans are rubber-stamping a decision that is in theory only advisory[1].

1.4   Continuous or Discrete

Many algorithms are used that yield numerical results varying continuously with their input. So if you earn £10 more, you pay £2 (or £4 or . . .) in income tax, and £20 more is twice that (unless you cross a threshold) and so on. If the wind speed increases by 1 km/h, then the wind chill changes by a certain amount. Other algorithms, particularly machine learning "classifier" algorithms give one of a small number of discrete results, e.g. degree classification. In many cases the answer is binary: bail or no bail; (referral for) melanoma or not. For a continuous algorithm, a small error in the inputs should result in a small error in the answer, whereas for a discrete algorithm, even the smallest error in the inputs may result in a different answer. If we understand the algorithm, we can consider the question "how close are we to a boundary", and possibly act on this information. For example, a university department might adopt (and some do) a rule that if a student fails by 1% or less, the scripts should be checked again. But if all we have is a black box that outputs decisions, we cannot ask this question.

1.5   Effect on people

It is a truism that every decision has consequences. The main concern is determinative (either *de facto* or *de jure*) algorithms making decisions that significantly affect people's lives. Examples would include medical diagnosis, mortgage approvals and [in the USA] granting or otherwise of bail, and possibly sentencing. However, there are also effects on people's prospects (being shortlisted for jobs), finances, e.g. gender discrimination in insurance, which in theory is illegal, but can be perpetuated through use of gender-correlated data, and other important aspects.

---

[1]Post-submission note: The Alan Turing Institute submission (number 73) claims that there is no "Right of Explanation", but that was not the consensus at the NYU Conference on Algorithms and Explanations: `http://www.law.nyu.edu/centers/ili/events/algorithms-and-explanations`.

1.6 Data

Any inferred algorithm is based on data. In the case of a scientific experiment, where the algorithm is often called a Law, as in Boyle's Law, one expects the experiment to the reproduced, and these repeat experiments to confirm the algorithms. In the case of other fields, the nature of the data is more variable. The term "ground truth" is used to refer to absolute facts, such as "this skin lesion is a melanoma", or "this person did not offend while on bail", whereas we use "proxy data" to refer to what is actually used as the training data, "did a dermatologist diagnose skin cancer" or "did a bail judge refuse bail [on the grounds of likelihood of re-offending]". Clearly ground truth is better, but it may be too difficult, or even impossible, to obtain. The study in [GKW$^+$90] (which deduces that dermatologists have a 64% accuracy rate, but a sensitivity of over 80%) was very rare in actually conducting biopsies on patients with a negative diagnosis by the dermatologist, so that we have ground truth. In many circumstances it may be impossible, e.g. we can't obtain "did this person [who was bailed] offend while on bail", merely "was this person [who was bailed] caught offending while on bail". This may seem pedantic, but note that we can't obtain the opposite datum "would this person [who was not bailed] have been caught offending while on bail had he been released.

This use of background data to derive an algorithm has two important consequences. The first is that any biases in the background data will be perpetuated in the algorithm. The second is that we have no guarantee at all that the algorithm produced will remain valid outside the range of those data. This means that every such algorithm needs to be traceable back to the background data that produced it.

2 Uses of Algorithms

Although well-known to many readers, we thought it important to draw Parliament's attention to various existing uses.

2.1 QE aircraft carrier

The business case for the QE Carrier drew heavily on the Cost Capability Trade Off Model, which uses a variety of modern algorithmic techniques such as Hybrid Model of Non-Linear Regression, Optimisation, Monte-Carlo Simulation and Design of Experiment in order to forecast the optimum performance within budgetary constraints. It is a mixture of proclaimed and experimental, is the company's intellectual property, but was explained to the customer, and was advisory, exploring a variety of scenarios. See [CP16] and `http://www.telegraph.co.uk/education/stem-awards/stem-hq/the-beauty-of-maths/`.

2.2 Scheduling etc.

Linear Programming, and its generalisation Mixed-Integer Programming, are used, generally as determinative algorithms, in practically every sector [Bix15, slide 18] The specification of the algorithm is public, even though the details of what makes it fast (not what the answer is, which is defined by the specification) are trade secrets. It is also an area in which the algorithmic advances are even greater than hardware advances, to the point where you would be better off running todays software on 1991 computers than vice versa [Bix15, slide 37]. For example, a supermarket chain would use Mixed-Integer Programming to decide how many lorries to send from each warehouse to each (delivery run of) supermarket(s), carrying which goods, to meet the order plan. If it used an independent haulage firm, that firm would use Mixed-Integer Programming to schedule the drivers and vehicles, and so on, and again this would be a determinative algorithm.

2.3 Online Advertising

Nowhere is commerce using opaque machine-learning algorithms more than in the Internet itself: algorithms determine what search engines return, what advertisements are shown to human beings, what advertisements are shown next to which content (which has caused controversy recently) and so on. It is probable that none of these activities are intentionally (on the part of the designers of the algorithms) biased, but in practice they are [DTD15, §6.1]. These algorithms are determinative.

3 "biases or discrimination"?

The Committee asked, and it's a very natural question, whether the use of "algorithms" can eliminate, introduce or amplify biases or discrimination, and whether this can be detected, and in a transparent and accountable way, and what the implications of increased transparency are.

Firstly, it must be noted that 'biases or discrimination' are human perceptions of a process, and what is, or is not, discrimination depends on society, and changes over time. Consider, for example, the recent ruling about gender bias in insurance, where what had been a common, and public, industry practice, was suddenly declared to be a bias. It's also a very nebulous question when pushed to precise specification, as doing something algorithmically requires.

3.1 Eliminate Bias

If the algorithm is known and fully understood, it is possible to assert that it is no more biased than the data fed into it, and should produce the same output on *identical* input. New algorithms for combining panelists' scores [MKLP17] could be used for many decision-making processes, e.g. the Research Excellence

Framework, and would improve the openness of the decision-making. If the algorithm is not known and fully understood, it is only possible to eliminate those biases which are known about and checked for, and then only if the algorithm has the appropriate mathematical properties, e.g. linearity.

Most such "machine learning algorithms" are not understood at all: no human being can say "why" the algorithm does what it does, nor can predict what it will do on data which are not the training data [Kni17]. Even strong advocates of these admit this major weakness. "Although D[eep] N[eural] N[etwork]s have demonstrated tremendous effectiveness at a wide range of tasks, when they fail, they often fail spectacularly, producing unexplainable and incoherent results that can leave one to wonder what caused the DNN to make such decisions. The lack of transparency in the decision-making process of DNNs is a significant bottleneck in their widespread adoption in industry, such as healthcare, defence, cybersecurity, etc., where the error tolerance is very low and the ability to interpret, understand, and trust decisions is critical." [KWT17]. We noted that the research in that paper, though good, is only looking at specific mistakes, and answering questions like "why did the DNN misclassify this Chihuahua as a Shih-Tzu", and not "how does the DNN recognise Chihuahuas".

3.2   Introduce Bias

It is certainly possible to introduce biases through an algorithmic process, just as it is possible to do so through a manual process. As pointed out [ALMK16], it is possible for such processes to be biased by considering things that human beings would refuse to consider publicly: the system used in Broward County Florida asks "Was one of your parents ever sent to jail or prison?", while it's hard to imagine a judge accepting the prosecution argument "the defendant deserves a harsher sentence because his father went to prison".

Consider, again, insurance, more specifically car insurance. It is no longer legal to discriminate on the basis of gender. It is, currently, legal to discriminate on the basis of occupation, even though some occupations are predominantly occupied by one gender, giving rise to indirect discrimination [McD15].

3.3   Amplify Bias

If one can introduce a bias, one can certainly amplify it. But there is an effect ("uncertainty bias") by which an unbiased algorithm can become biased by considering factors that are not uniformly distributed, and this bias can grow over time as an "active learning" algorithm learns and reinforces its bias [GF16].

3.4   Detection of Bias

Currently, detecting bias relies on human effort, sometimes aided itself by machine learning. We have quoted various instances of bias, but these are merely those that researchers have chosen to investigate. Direct bias, e.g. highly-paid jobs being offered to explicitly male searchers [DTD15] can be obviated by search

engines never taking gender into account (but this is highly improbable), but even so this would not eliminate indirect bias [McD15].

3.5   Transparent and Accountable Way

Many of the most efficient algorithms for certain tasks are not transparent or accountable. It is, with the current state of technology, impossible to understand how a typical neural net reaches its decisions. Though the state of technology may change, it is currently the case that only algorithms which can be reduced to a (possibly complex) formula, can be understood, and even then possibly only by experts.

3.6   The implications of increased transparency

The background data should have been thoroughly anonymised (easier said than done). The foreground data should be subject to the usual data protection rules. There is a real challenge with "active learning" algorithms, where today's foreground data becomes tomorrow's background data. To the best of our knowledge, no satisfactory research has been carried out here.

An issue that has received very little attention (but see [MKWH16]) is the interaction between "the right to be forgotten" and machine learning. If X exercises his right to be forgotten, does/should Y's insurance premium change?

4   Bad practice

We were actually asked for examples of good practice, but couldn't really find any in the machine learning area. We did want to warn about examples of bad practice.

4.1   Correlation versus Causation

Though this error is warned about in most statistics courses, confusing 'correlation', which is what algorithms, machine learning or otherwise, can detect, with 'causality', which is what people are generally interested in, is still common. Consider [Gri16, page 16], the example of "Screen time [at 14 years] was associated with lower academic performance [at 16 years]". Media reporting generally converted correlation into causality, as in "programmes aimed at reducing screen time could have important benefits for teenagers' exam grades", whereas increased screen time might be an early indicator of poor performance, and the investment should be in coaching/catching up.

4.2   Training versus Testing Data

Many machine learning protocols require splitting the background data into training and testing data. However, it is bad practice to believe that this split is in any way significant. At the very least, this process should be repeated

several times, known as *crossvalidation*. The state of the art is not yet able to give good guidelines for "several".

5 Recommendations

Both societies made several recommendations to Government, many of which will be obvious to readers. The full set of recommendations are in the submissions. Key ones were the following.

1. The Government, and all its agencies and subcontractors, *including recruitment agencies*, need to review as a matter of urgency the use of all automatic processing to ensure it:

   (a) Is only based on legitimate personal data of the individual;

   (b) Carries a precise description of any training data it was built from;

   (c) Has been tested for indirect discrimination;

   (d) Is capable of being explained in line with GDPR 15.1(h). In particular, buzzphrases like "Artificial Intelligence", "deep learning algorithm", "data-based algorithm" should act as warning signs that the "algorithm" is in fact probably no more than unscientific reasoning by generalisation

2. No data should be fed to an algorithm, determinative or advisory, that would not be acceptable in an equivalent manual process: see the Zilly case above and [ALMK16].

3. In machine learning algorithms, the background data contribute to the decision, so every such algorithm should be prominently labelled with the data that created it, both as a statement the lay person can understand (e.g. "based on London traffic data 1982-2002") and question ("but that was before the congestion charge"), and ultimately such that an expert can analyse it. This is an algorithmic consequence of the 15.1(h) right.

6 Conclusions

The rôle of the "traditional" algorithm, whether proclaimed or inferred, is great and growing (possibly not as fast as it should), whether in government or elsewhere. However, there is a need to be cautious when endowing the "algorithms" produced by many forms of machine learning, which humans, even experts, do not understand, with determinative (either *de facto* or *de jure*) powers that affect people, individually or as society, in non-trivial ways.

to the NYU Privacy Research Group, especially Professor Strandberg, for their debates.

References

[ALMK16] J. Angwin, J. Larson, S. Mattu, and L. Kirchner. Machine Bias. There is software that is used across the county to predict future criminals. And it is biased against blacks. `https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing`, 2016.

[Bir17] Birmingham Mail. Anger over Youtube Burger Bar gang shooting boast videos. `http://www.birminghammail.co.uk/news/midlands-news/anger-over-youtube-burger-bar-12770609`, 2017.

[Bix15] R.E. Bixby. Computational Progress in Linear and Mixed Integer Programming. *Presentation at ICIAM 2015*, 2015. Available at `http://staff.bath.ac.uk/masjhd/Others/Bixby2015a.pdf`.

[CP16] N. Chamberlain and E. Pinker. The Cost Capability Trade Off Model (presentation at ORS 58). `http://staff.bath.ac.uk/masjhd/Others/ChamberlainPinker2016a.pdf`, 2016.

[DL98] F. Duckworth and A. Lewis. A fair method for resetting the target in interrupted one-day cricket matches. *J. Oper. Res. Soc.*, 49:220–227, 1998.

[DTD15] A. Datta, M.C. Tschantz, and A. Datta. Automated Experiments on Ad Privacy Settings. *Proceedings on Privacy Enhancing Technologies,*, 2015:92–112, 2015.

[GF16] B. Goodman and S. Flaxman. European Union regulations on algorithmic decision-making and a "right to explanation". `https://arxiv.org/pdf/1606.08813.pdf`, 2016.

[GKW+90] C.M. Grin, A.W. Kopf, B. Welkovich, R.S. Bart, and M.J. Levenstein. Accuracy in the Clinical Diagnosis of Malignant Melanoma. *Arch. Dermatol.*, 126:763–766, 1990.

[Gri16] P. Grindrod. Beyond privacy and exposure: ethical issues within citizen-facing analytics. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 374(2083), 2016.

[Gua16] The Guardian. Cyclist died after three ambulances could not find Olympic velodrome. `https://www.theguardian.com/uk-news/2016/jun/02/cyclist-died-after-three-ambulances-could-not-find-olympic-velodrome`, 2016.

[Gua17]   The Guardian. Google ad controversy: what the row is all about. `https://www.theguardian.com/technology/2017/mar/17/youtube-and-google-search-for-answers`, 2017.

[Kni17]   W. Knight. The Dark Secret at the Heart of AI. `https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/`, 2017.

[KWT17]   D. Kumar, A. Wong, and G.W. Taylor. Explaining the Unexplained: A CLass-Enhanced Attentive Response (CLEAR) Approach to Understanding Deep Neural Networks. `https://arxiv.org/abs/1704.04133`, 2017.

[McD15]   S. McDonald. Indirect Gender Discrimination and the Test-Achats Ruling: An Examination of the UK Motor Insurance Market (presentation to Royal Economic Society, April 2015). `https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=RES2015&paper_id=791`, 2015.

[MKLP17]   R.S. MacKay, R. Kenna, R.J. Low, and S. Parker. Calibration with confidence: A principled method for panel assessment. *R. Soc. Open Sci. 160760*, 4, 2017.

[MKWH16]   B. Malle, P. Kieseberg, E. Weippl, and A. Holzinger. The right to be forgotten: towards machine learning on perturbed knowledge bases. *International Conference on Availability, Reliability, and Security, CD-ARES 2016*, Springer, pp. 251–266, 2016.

[MPR09]   P. Malighetti, S. Paleari, and R. Redondi. Pricing strategies of low-cost airlines: The Ryanair case study. *Journal of Air Transport Management*, 15(4):195–203, 2009.